Systematic Botany (2012), 37(4): pp. 995–1005

DOI 10.1600/036364412X656482

Sequence Data from New Plastid and Nuclear COSII Regions Resolves Early Diverging Lineages in *Coffea* (Rubiaceae)

Michael D. Nowak,[1,4] Aaron P. Davis,[2] and Anne D. Yoder[3]

[1]Institute of Systematic Botany, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland.
[2]Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, U. K.
[3]Department of Biology, Box 90338, Duke University, Durham, North Carolina 27708, U. S. A.
[4]Author for correspondence (michael.nowak@systbot.uzh.ch)

Communicating Editor: Jennifer Tate

Abstract—Resolving phylogenetic relationships within the economically important genus *Coffea* (Rubiaceae) has proven to be difficult due to low levels of plastid and nuclear ITS sequence divergence. The recent identification of a large number of conserved, single-copy, putatively orthologous genes (COSII) provides a unique opportunity to apply regions of the nuclear genome to phylogenetic studies of Euasterid species. We expand a previously published plastid data set of *Coffea* with the addition of three new plastid regions and a single COSII nuclear locus. Significant phylogenetic incongruence is identified between the plastid and nuclear COSII data sets, and a combined analysis is performed after removal of incongruent taxa. Phylogenetic results from plastid, nuclear, and combined plastid/nuclear data sets allow the novel identification of early diverging lineages in *Coffea*. Specifically, the data show that a group of predominately Lower-Guinea/Congolian *Coffea* species form a clade that is sister to the rest of the genus, and that the dry adapted baracoffea alliance from western Madagascar falls outside of previously defined East Africa/Indian Ocean and Indian Ocean clades. Well-supported patterns of phylogenetic incongruence are observed between plastid and nuclear data in several species, suggesting a potential role for hybridization in their evolutionary history. The results also provide further evidence for the paraphyly of African *Coffea* and support the supposition that the evolutionary history of the genus is more complicated than previously proposed.

Keywords—*Coffea*, COSII, incongruence, phylogeny, plastid, Rubiaceae.

Coffea L. is best known as the source of the beverage coffee, which is produced from the roasted seeds of the species *C. arabica* L. (Arabica) and *C. canephora* A. Froehner (robusta). Until recently, *Coffea* comprised 104 mostly self-incompatible species distributed in tropical Africa, Madagascar, the Comoros Islands, and the Mascarene Islands (Davis et al. 2006, 2010; Nowak et al. 2011). Recent morphological and molecular data, however, show that the genus *Psilanthus* Hook. f. should be included within *Coffea*, adding around 20 further species and extending the distribution across tropical southern Asia and into southeastern Asia and northern Australia (Davis 2010, 2011; Davis et al. 2011). While generally restricted to humid evergreen forest habitats, several *Coffea* species are adapted to seasonally dry forest and shrublands (Davis et al. 2006). This is particularly true after the transfer of *Psilanthus* Hook. f. species to *Coffea*, as most of these species are from seasonally dry (deciduous) forest. Until recently, the genus was thought to be composed of two subgenera: *Coffea* subgenus *Coffea* distributed throughout the geographic range of the genus, and *Coffea* subgenus *Baracoffea* (nine species) endemic to seasonally dry forests in western Madagascar (Davis et al. 2005). Maurin et al. (2007), however, demonstrated that the recognition of subgenus *Baracoffea* cannot be upheld as it requires subgenus *Coffea* to be paraphyletic. Based on morphological and molecular data, the species that once composed subgenus *Baracoffea* are strongly supported as monophyletic (Maurin et al. 2007) and this clade is now referred to as the 'baracoffea alliance' (Davis and Rakotonasolo 2008). Accounting for around 50% of the species diversity of *Coffea*, the Malagasy species exhibit 100% endemism and generally have restricted local distributions. This pattern of diversity and local endemism exemplified by Malagasy *Coffea* is characteristic of other Malagasy Rubiaceae (Davis et al. 2009) and the Malagasy flora as a whole (Yoder and Nowak 2006).

Coffea, including the species formerly placed in *Psilanthus*, are among the 11 genera comprising the tribe Coffeeae. Molecular data show that *Coffea* (plus *Psilanthus*) is confidently placed sister to the rest of Coffeeae (Davis et al. 2007; Tosh et al. 2009). Morphologically, *Coffea* can be characterized by seed morphology. In each fruit there are two pyrenes (rarely one by abortion, e.g. "peaberry" coffee beans), and the seeds are covered in a thin to rather thick crustaceous parchment (the endocarp). Each seed is convex on the dorsal surface and usually flat on the ventral surface, where there is a distinct longitudinal groove curling inside the seed and separating the endosperm. In other words, the fruits of all *Coffea* species contain 'coffee beans.'

Resolving the phylogenetic relationships among species of *Coffea* has been the focus of a number of previous systematic studies that have employed data from diverse sources, including microsatellites (Cubry et al. 2008), RAPDs (Lashermes et al. 1996; Orozco-Castillo et al. 1996), morphological characters (Davis et al. 2005), long terminal repeat retrotransposons (Hamon et al. 2011) and DNA sequence data from both plastid and nuclear genomes (Cros et al. 1995; Lashermes et al. 1997; Cros et al. 1998; Maurin et al. 2007; Anthony et al. 2010; Davis et al. 2011). A persistent trend observed in all of these previous studies is the lack of phylogenetic resolution for the earliest divergences in the genus. Despite the existence of morphological and molecular support for the baracoffea alliance and lineages of species formerly included in *Psilanthus* (Davis et al. 2005, 2011), and molecular support for certain geographical clades (Maurin et al. 2007), there has been little insight into the relationships between the major lineages. The resolution of basal and lower-level nodes has important implications for reconstructing the biogeographic history of the genus. Presently, biogeographic hypotheses describing the evolutionary history of *Coffea* are speculative based on the data at hand.

Among the previous molecular phylogenetic studies of the genus, Maurin et al. (2007) is noteworthy due to remarkable sampling of species diversity (i.e. nearly 85% of described species) and the impressive molecular data set that included sequence data from both nuclear (ITS 1/5.8S/ITS 2) and plastid (*trnL–F* intron, *trnL–F* intergenic spacer (IGS), *rpl16* intron and *accD–psa1* IGS) genomes. Their results provide good

support for several geographically coherent African, Indian Ocean (Mascarene clade, baracoffea alliance clade) and Indian clades, and resolve many species level relationships in Africa. The study, however, failed to fully resolve the relationships between the major well-supported lineages mentioned above, and apart from the baracoffea alliance, the study found low levels of sequence divergence for the Malagasy *Coffea* species, resulting in poor topological resolution. Anthony et al. (2010) supplemented the plastid *trnL-F* sequence data of Maurin et al. (2007) with two additional non-coding chloroplast regions (intergenic spacers *trnT-L* and *atpB-rbcL*). Despite sampling a slightly different, but overall less diverse set of species, their results largely reinforced those of Maurin et al. (2007). Given the lack of phylogenetic support for early diverging clades in the genus, Anthony et al. (2010) conclude that *Coffea* experienced an adaptive radiation at this point in their history. Davis et al. (2011) used the same markers as Maurin et al. (2007) but added several new *Psilanthus* and *Coffea* species, updated the sequence alignments and conducted Bayesian analysis, where Maurin et al. (2007) had only used parsimony. The objective of their paper was to make a more detailed assessment of the relationship between *Psilanthus* and *Coffea*, and thus they reduced the sampling of Malagasy species. Their study added further insights into the evolutionary history of *Coffea*, most significantly that all (20 species) species formerly placed *Psilanthus* were deemed to belong to *Coffea*. Despite finding some support for an Indian Ocean clade, including Malagasy (including the baracoffea alliance) and Mascarene species, and retrieving the same major groupings of species as per Maurin et al. (2007), the relationship between the major groupings remained unresolved.

Here we present a molecular phylogenetic study in *Coffea* utilizing DNA sequence data originally published by Maurin et al. (2007) supplemented with sequence data from three new chloroplast regions developed with the aid of the sequenced *C. arabica* plastid genome (Samson et al. 2007). Additionally, taking advantage of the recently identified nuclear conserved ortholog set II (COSII; Wu et al. 2006) we complement these plastid sequence data with new sequence data from a single-copy nuclear COSII locus, which we call COS3. Our sampling scheme is focused on Indian Ocean *Coffea* (i.e. Madagascar, the Comoros Islands, and the Mascarene Islands). Our objective was to focus on the elucidation of relationships between the major well-supported lineages (Maurin et al. 2007), given that there has been little or no support for relationships at this level in previous analyses (Cros et al. 1995; Lashermes et al. 1997; Cros et al. 1998; Davis et al. 2005; Maurin et al. 2007; Anthony et al. 2010; Davis et al. 2011).

Materials and Methods

***Taxon Sampling and Plant Material*—**The taxa sampled in this study represent approximately half of the *Coffea* species described to date, including representatives of all major geographic and phylogenetic lineages identified in previous phylogenetic studies (Lashermes et al. 1997; Maurin et al. 2007; Davis et al. 2011). *Tricalysia perrieri* Homolle ex Randriamb. & De Block was chosen as an outgroup (Tosh et al. 2009). The samples used in this study, with accepted taxon names, voucher information and GenBank accession numbers for the newly generated sequences are given in Appendix 1. Many samples are of wild origin collected during field expeditions to Madagascar in 2008 and 2009. Other samples are from living material held in botanical gardens and coffee research stations, and some are from herbarium material (leaf samples or single seeds) taken from specimens held at K and BR (abbreviations after Thiers 2012). DNA from wild origin and cultivated material was extracted from silica gel dried leaf fragments. Voucher specimens are held in K, MO, TAN and at the Institut de recherche pour le développement (IRD) in Montpellier, France (Appendix 1). Many of the genomic DNA extractions used in this study have been deposited in the Kew DNA Bank (see Appendix 1). The DNA sequence alignments used in this study have been submitted to TreeBASE (study number 11927).

***Geographical Units*—**The terminology for area-based clades follows Maurin et al. (2007): Upper Guinea (UG) clade, Lower Guinea/Congolian (LG/C) clade, East-Central Africa (ECA) clade, East Africa (EA) clade; and Anthony et al. (2010) and Davis et al. (2011): Africa/Indian Ocean (A/IO) clade. The humid West and Central 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.

***DNA Extraction*—**Genomic DNA was extracted from approximately 100 mg (dry weight) of finely ground leaf tissue using a standard CTAB extraction procedure (Murray and Thompson 1980), and the resulting eluate was purified further using the Qiagen DNeasy plant mini kit (Qiagen, Valencia, California). The genomic DNA concentration (i.e. ng/μL) and purity of each sample extract were measured via 260/280 nm NanoDrop spectrophotometry (Thermo Scientific, Wilmington, Delaware).

***Primers, Amplification, and Sequencing*—**Primers for the three new chloroplast loci were designed from the published *C. arabica* chloroplast genome sequence (Samson et al. 2007; Table 1). The chloroplast region we refer to here as *atpF* is composed of the majority of the *atpF* alpha subunit and about 100 bp of exon 2 of the *atpF* beta subunit, including approximately 70 bases of intergenic sequence (Fig. 1). The chloroplast region we refer to here as *rpolb1* is composed of approximately half of RNA polymerase beta subunit and half of exon 2 of RNA polymerase beta' including approximately 150 bases of intergenic sequence (Fig. 1). The chloroplast region we refer to here as *rpolb2* is composed of approximately half of both exons 1 and 2 of RNA polymerase beta and the intervening intron sequence of approximately 700 nucleotides (Fig. 1).

Primers for the nuclear locus COS3 were designed by Wu et al. (2006), who used a bioinformatics approach to identify 2,869 single copy orthologous genes (COSII) in seven euasterid species of particular agricultural importance (tomato, potato, pepper, coffee, sunflower, and lettuce) to facilitate the utilization of nuclear non-coding regions in euasterid molecular systematics research. Their bioinformatic analysis performed BLAST searches of EST (expressed-sequence tag) libraries previously generated from *C. canephora* (Lin et al. 2005) against the *Arabidopsis thaliana* (L.) Heynh. genome to identify putative single copy

Table 1. Plastid and COSII primer sequences, amplicon lenth, and the distribution of parsimony informative characters in the data matrix. Primer sequences for *trnL-F* spacer, *rpl16* intron, and *accD-psaI* spacer are from Maurin et al. (2007). Parsimony informative characters are reported here as the sum of all parsimony informative sites across data partitions (exons and introns) including the outgroup.

Locus	Forward Primer (5′ – 3′)	Reverse Primer (5′ – 3′)	Characters	Parsimony Informative
atpF	TGAATAGATCGGCGGATAGG	GGCACTAATCGTACGTAAATGTAACTCG	1,082	5
rpolb1	GCTTTGAATGGGTAAATCAATCATTTGTCC	GCCGCGAGAAATAGCAATAG	1,386	11
rpolb2	CGCGAAATCTTCCTTCTTTG	GAAGGTATCAAATGGGATACATCAAACTCG	1,425	16
trnLF	GGTTCAAGTCCCTCTATCCC	ATTTGAACTGGTGACACGAG	848	15
rpl16	GCTATGCTTAGTGTGTGACTCGTTG	CGTACCCATATTTTTCCACCECGAC	994	29
accD-psaI	GGAAGTTTGAGCTTTATGCAAATGG	AGAAGCCATTGCAATTGCCGGAAA	1,064	29
COS3	CCCTTCGACACCATTGATTTGTCCG	CACCGGGTTCAAATGTCTTTGTAGATACC	787	54

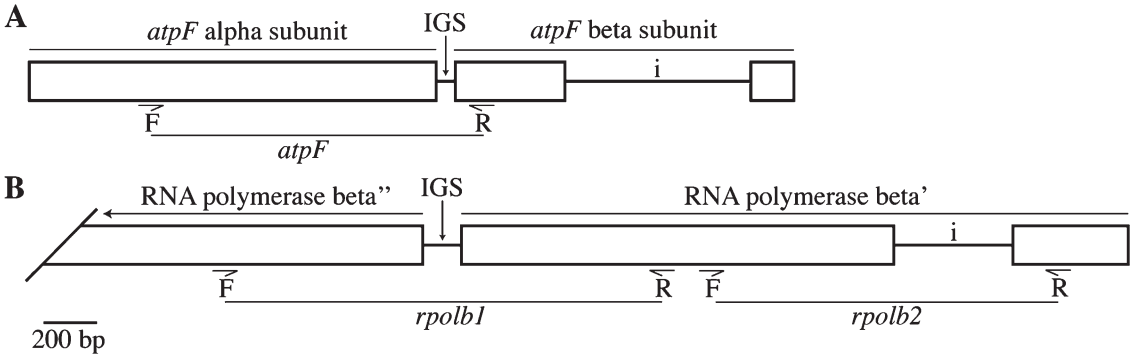

FIG. 1. Schematic representations of the three new plastid loci sampled in this study. A. Gene model for the *atpF* locus. B. Gene model for the *rpolb1* and *rpolb2* loci. Legend: F = forward priming site; R = reverse priming site; i = intron; IGS = non-coding intergenic sequence.

homologs. The *A. thaliana* gene model allowed identification of putative intron/exon boundaries, and they used a primer design algorithm to design exon-anchored primers up- and down-stream of the putative intron, which are sometimes referred to as EPIC markers (exon-primed intron-crossing; Bouck and Vision 2007). We redesigned the COS3 primers generated by Wu et al. (2006) to reduce degeneracy and more closely match the *C. canephora* exon sequence, which is accession number cgn_129888 on the Sol Genomics Network website (http://www.sgn.cornell.edu/; Mueller et al. 2005). The amplicon produced from the COS3 region is composed of two introns with a short intervening exon and two partial terminal exons (i.e. < 100 nucleotides each), resulting in five distinct regions that are treated as data partitions in all partitioned Bayesian phylogenetic analyses. The most likely *A. thaliana* homolog to COS3 is a gene (At1g09760; E value = 8e-13) that codes for a U2 small nuclear ribonucleoprotein A (U2A'), which is known to function in nuclear mRNA splicing. Based on this simple homology assessment, it appears clear that COS3 is a "housekeeping" gene, and thus there is no reason to believe that this locus experienced significant molecular evolution in response to strong directional selection, but we have not empirically tested this assertion.

The four loci (e.g. *atpF*, *rpolb1*, *rpolb2*, and COS3) were amplified from genomic DNA using standard PCR reagents and JumpStart *Taq* (Sigma Aldrich, St. Louis, Missouri) under the following cycling conditions: 95°C initial denaturation for 3 min, 30 cycles of 95°C for 30 sec, 56°C for 45 sec, and 72°C for 2 min, followed by a final extension at 72°C for 10 min. The resulting amplicon was treated with ExoSAPit (USB, Cleveland, Ohio) and employed as template in a standard dye-terminator sequencing reaction (i.e. using both the forward and reverse amplification primers), which was subsequently sequenced on an Applied Biosytems 3730xl capillary sequencer at the Duke University IGSP DNA Sequencing Core Facility. Primer sequences were removed and base calls were visually inspected in the program Sequencher (Gene Codes, Ann Arbor, Michigan).

Phylogenetic Analyses and Tests of Phylogenetic Congruence—Sequences from each locus were aligned by eye in MacClade v4.08 (Sinauer, Sunderland, Massachusetts). In the COS3 data set putative microsatellite repeat regions and characters that appeared heterozygous (i.e. double peaks in the chromatogram) and were not parsimony informative were excluded from all analyses (no putatively heterozygous sites were parsimony informative). The Akaike information criterion (AIC) in the program MrModeltest v2.0 (Nylander 2004) was used to identify the most appropriate model for each locus and data partition. Each locus was subjected to phylogenetic analysis using both likelihood (GARLI v1.0 bootstrap) and Bayesian (MrBayes v3.1.2; Ronquist and Huelsenbeck 2003) methods. Initial GARLI analyses were performed to find the tree with the highest likelihood using default settings and simultaneous estimation of the model parameters. A subsequent GARLI bootstrap analysis (100 replicates) provided branch support metrics for annotating the maximum likelihood phylogram using the Python script sumtrees.py, which is included in the Dendropy library v.3.3.0 (Sukumaran and Holder 2010). MrBayes analyses were performed in triplicate and each run was partitioned such that each locus and intron and exon sequences were assigned to a distinct partition, and model parameters for each partition were unlinked. The Markov chain was allowed to sample parameter space for ten million generations, with parameter sampling every one thousand generations. The parameters and likelihood scores of the three runs were visually analyzed for convergence and stationarity in the program Tracer v1.5 (Rambaut and Drummond 2007), and removal of one thousand trees as burn-in resulted in a posterior distribution of nine thousand trees from each run (i.e. a total posterior sample of 27,000 trees). Bremer support values (also known as decay indices; Bremer 1988) were estimated for well-supported nodes (posterior probability > 0.90) using the TreeRot v.3 software with default settings (Sorenson and Franzosa 2007) and PAUP* v.4.10b (Swofford 2002).

We tested phylogenetic congruence between the six plastid loci (*atpF*, *rpolb1*, *rpolb2*, *trnL-F*, *rpl16*, and *accD-psaI*) and between the concatenated plastid and COS3 regions with an incongruence length difference test (ILD, Farris et al. 1994; also known as the partition homogeneity test). ILD tests were performed with 100 replicates under the parsimony optimality criterion in PAUP* v.4.10b (Swofford 2002). A heuristic search was conducted for each ILD replicate with starting trees obtained by stepwise addition and ten random addition sequence replicates, saving no more than 1,000 trees per replicate.

RESULTS

Plastid Sequence Data—We sequenced three new chloroplast regions (*atpF*, *rpolb1*, and *rpolb2*) from specimens representing approximately one third of named Malagasy species and three African species (*C. arabica*, *C. canephora*, and *C. eugenoides*). The species *Tricalysia perrieri* was included as a representative outgroup. To assess the suitability of combining these data with previously published chloroplast sequence data from the *trnL-F*, *rpl16*, and *accD-psaI* regions (Maurin et al. 2007), we performed pair-wise ILD tests between the six loci. No significant incongruence was found between any of the loci (data not shown), and thus a combined plastid sequence alignment was constructed. The six loci were concatenated in parsimony and likelihood analyses, and for maximum likelihood analyses the concatenated data set was assigned a GTR + G + I model. Bayesian phylogenetic analyses were partitioned by locus and parameters of the best-fit substitution model for each partition were unlinked. For Bayesian analyses, each plastid locus was assigned a unique model as follows: *atpF* = HKY + I; *rpolb1* = GTR + I; *rpolb2* = GTR + G; *trnL-F* = HKY + G, *rpl16* = GTR + G; *accD-psaI* = GTR + G + I. The combined plastid sequence alignment contained 6,799 characters from 67 taxa (including the outgroup), where 6,411 characters were constant, 281 characters were variable but parsimony-uninformative, 107 characters were parsimony informative, and the alignment contained approximately 29% missing data (see Table 1).

The results of maximum likelihood and Bayesian phylogenetic analyses of the combined plastid data set as well as Bremer support values for well-supported clades are summarized in Fig. 2. Below we present the support values resulting from these phylogenetic analyses in square

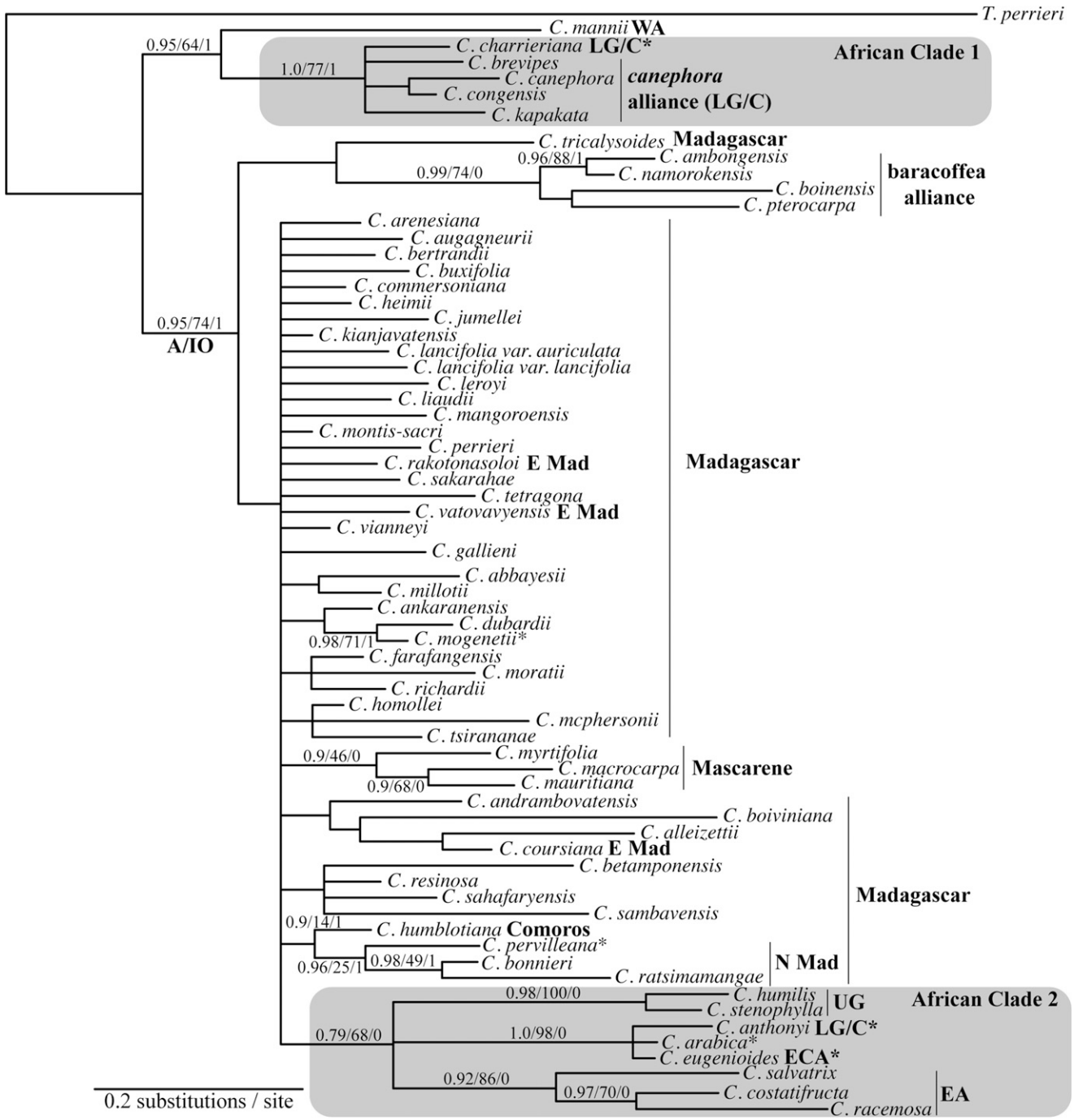

FIG. 2. The Bayesian majority rule tree showing all compatible groups (i.e. using the setting "allcompat" in MrBayes v3.1.2; Ronquist and Huelsenbeck 2003) of the combined plastid data set. Well-supported nodes are labeled with Bayesian posterior probabilities, maximum likelihood bootstrap proportions, and Bremer support indices, respectively. Clades discussed in the text are labeled: A/IO = African/Indian Ocean, African Clades 1 and 2, baracoffea alliance, *canephora* alliance. African species are highlighted in grey and species geographic distributions are labeled as follows: EA = East Africa, ECA = East Central Africa, LG/C = Lower Guinea/Congolian, UG = Upper Guinea, N Mad = Northern Madagascar, E Mad = Eastern Madagascar. Taxa marked with an asterisk exhibit an incongruent placement between the plastid and nuclear COS3 topologies.

brackets using the following convention: [Bayesian posterior probability = PP / maximum likelihood bootstrap = L / Bremer support = *b*]; and we use "BP" when referring to support values from previous publications that used maximum parsimony bootstrap. We designate clades receiving at least 0.95 PP as well supported. The African clade 1 is well supported as monophyletic [PP = 1.0/L = 77/*b* = 1] and placed sister [PP = 95/L = 64/*b* = 1] to the West African *C. mannii* (formerly *Psilanthus mannii*) at the base of the tree. African clade 1 is composed of "*canephora* alliance" species (*C. canephora*, *C. kapakata*, *C. congensis*, and *C. brevipes*) and the Lower Guinea species *C. charrieriana*. The relationships between the four species of the *canephora* alliance within the African clade 1 are unresolved. The UG [PP = 0.98/L = 100/*b* = 0] and EA clades [PP = 1/L = 98/*b* = 0] receive strong support for being monophyletic, as does a clade composed

of the ECA species *C. eugenioides*, *C. arabica*, and the LG/C species *C. anthonyi* [PP = 1/L = 98/*b* = 0]. These three clades compose a weakly supported [PP = 0.79/L = 68/*b* = 0] African clade 2 that is placed within a largely unresolved but well-supported clade [PP = 0.95/L = 74/*b*= 1] containing all Indian Ocean (i.e. Madagascar, Comoros, and Mascarene) species, which we refer to as the African/Indian Ocean (A/IO) clade in keeping with the terminology established by Anthony et al. (2010) and Davis et al. (2011). Despite a lack of resolution for the vast majority of relationships in the A/IO clade, we recover support for several smaller groupings, including a monophyletic baracoffea alliance from western Madagascar (*C. ambongensis*, *C. boinensis*, *C. pterocarpa*, and *C. namorokensis*,) [PP = 0.99/L = 74/*b* = 0], a monophyletic Mascarene *Coffea* (*C. myrtifolia*, *C. mauritiana*, and *C. macrocarpa*) [PP = 0.9/L = 46/*b* = 0], and a clade of species from northern Madagascar (*C. pervilleana*, *C. bonnieri*, and *C. ratsimamangae*; "N Mad" in Fig. 2) [PP = 0.96/L = 25/*b* = 1] sister to the Comoros Islands endemic species *C. humblotiana*. It is important to note here that *C. namorokensis* was originally identified as *C. decaryana* by Maurin et al. (2007).

***COS3 Nuclear Sequence Data*—**We sequenced a single copy COSII nuclear region, which we refer to here as COS3, from 62 taxa and the resulting sequence alignment consisted of 787 characters, 599 were constant, 134 were variable but parsimony uninformative, 54 were parsimony informative, and the alignment contained only 32 missing characters (<1%; Table 1). Parsimony informative characters (including the outgroup) were distributed among the five data partitions as follows: exon 1 = 1; exon 2 = 3; exon 3 = 2; intron 1 = 12; intron 2 = 36. The results of an ILD test suggested no significant incongruence between the five partitions (data not shown) representing one complete and two partial exons and two complete introns. Therefore, the five partitions were concatenated and treated as a single locus in parsimony analyses, and assigned the GTR + G + I model in likelihood analyses. For Bayesian analyses, each data partition in the COS3 locus was assigned a unique model as follows: exon 1 = K80; exon 2 = F81; exon 3 = JC; intron 1 = HKY; intron 2 = GTR + I. The results of maximum likelihood and partitioned Bayesian phylogenetic analyses of the COS3 alignment as well as Bremer support values for well-supported nodes are summarized in Fig. 3.

Consistent with the combined plastid data (Fig. 2), we find strong support for the paraphyly of African *Coffea*, but there are several differences in the composition of African clades 1 and 2 (Fig. 3). The main relationships at the base of the COS3 tree are not supported, and *C. mannii* is not confidently placed in the analysis. There are three well-supported clades: (1) A clade that is roughly equivalent to African Clade 1 composed of the four species of the *canephora* alliance (see above) together with the LG/C species *C. anthonyi* (Stoffelen et al. 2009), the ECA species *C. eugenioides* and *C. arabica* [PP = 1/L = 99/*b* = 4]. While this placement of *Coffea arabica* is not terribly surprising given that this species is an allotetraploid with known parentage of *C. eugenioides* and *C. canephora* (Raina et al. 1998; Lashermes et al. 1999; Maurin et al. 2007), it was somewhat unexpected that the amplicon of the COS3 locus from *C. arabica* does not appear to contain multiple copies, and thus was sequenced directly. (2) A clade composed of the three baracoffea alliance species (western Madagascar) and three species from eastern Madagascar (*C. rakotonasoloi*, *C. vatovavyensis*, and *C. coursiana*; "E Mad" in Figs. 1 and 2) [PP = 1.0/L = 90/*b* = 4]. (3) A clade composed of all Indian Ocean *Coffea* aside from the members of the baracoffea alliance/E Mad clade, plus six African species, which we refer to as "Clade A" in Fig. 3 [PP = 0.95/L = 54/*b* = 1]. The African species placed within Clade A are not resolved as monophyletic. The West African (Lower Guinea) *C. charrieriana* is in an unresolved position at the base of Clade A, and the remaining five species form a well-supported clade [PP = 1.0/L = 59/*b* = 2] roughly equivalent to African Clade 2 in Fig. 3 containing three East African species (*C. racemosa*, *C. costatifructa*, *C. salvatrix*) and the Upper Guinea species *C. humilis*. Also within Clade A, the three Mascarene *Coffea* are resolved as monophyletic [PP = 1/L = 92/*b* =2], and a clade composed of 34 Malagasy and a single Comoros species receives weak to moderate support [PP = 0.90/L = 19/*b* = 1]. The position of *C. charrieriana* at the base of Clade A in the COS3 analysis should be highlighted. In the plastid analysis it is placed in African Clade 1 with the *canephora* alliance species [PP = 1/L = 77/*b* = 1].

***Combined Plastid/COS3 Phylogenetic Analyses*—**The results of an ILD test showed significant phylogenetic incongruence between the combined plastid and COS3 data sets (p value < 0.05). To analyze the combined data set in the absence of phylogenetic incongruence, we removed all taxa showing significantly different placement (i.e. well-supported signal of incongruence) between the plastid and nuclear COS3 trees. This included *C. anthonyi*, *C. arabica*, *C. charrieriana*, *C. eugenioides* (see above), *C. pervilleana* and *C. mogenetii*. The resulting combined sequence alignment of 62 taxa contained 7,586 characters, of which 7,058 were constant, 398 were variable but parsimony uninformative, 130 were parsimony informative (i.e. 93 plastid characters and 42 nuclear characters), and the alignment contained approximately 24% missing data. The results of maximum likelihood and partitioned Bayesian phylogenetic analyses of these data as well as Bremer support indices are shown in Fig. 4. Strong support is recovered for a monophyletic African Clade 1 composed of the four species of the *canephora* alliance [PP = 1.0/L = 100/b = 10] and this clade is confidently placed sister to *C. mannii* [PP = 0.95/L = 71/b = 1]. The baracoffea alliance is monophyletic [PP = 1.0/L = 100/b = 15] and in a well-supported clade with the E Mad species [PP = 1.0/L = 86/b =1]. This clade is placed sister [PP = 1.0/L = 86/b = 2] to a monophyletic Clade A sensu Fig. 3 [PP = 1.0/L = 84/b = 2]. While the base of the Clade A is unresolved, it does contain a well-supported and fully resolved African Clade 2 [PP = 1.0/L = 94/b = 4] and the Mascarene *Coffea* are monophyletic [PP = 1.0/L = 81/b = 3]. Aside from *C. heimii* (northern Madagascar), in an unresolved position at the base of Clade A, the remaining Malagasy species and the Comoros endemic *C. humblotiana* are placed in a consistently retrieved but weakly supported clade.

Prominent results produced by the combined plastid-COS3 data set include (support values given above): (1) the supported relationship between *C. mannii* and the African Clade 1; (2) the well-supported relationship of the *C. mannii* + Clade 1 clade sister to the rest of *Coffea*; and (3) the well-supported sister relationship between the baracoffea alliance and Clade A. The last of these relationships conflicts with the findings of Davis et al. (2011), as they receive support for an A/IO clade (PP = 1.0/BP = < 50) including an Indian Ocean clade (IO) composed of Malagasy and Mascarene (Mauritius and Reunion) species (BP = 0.99/

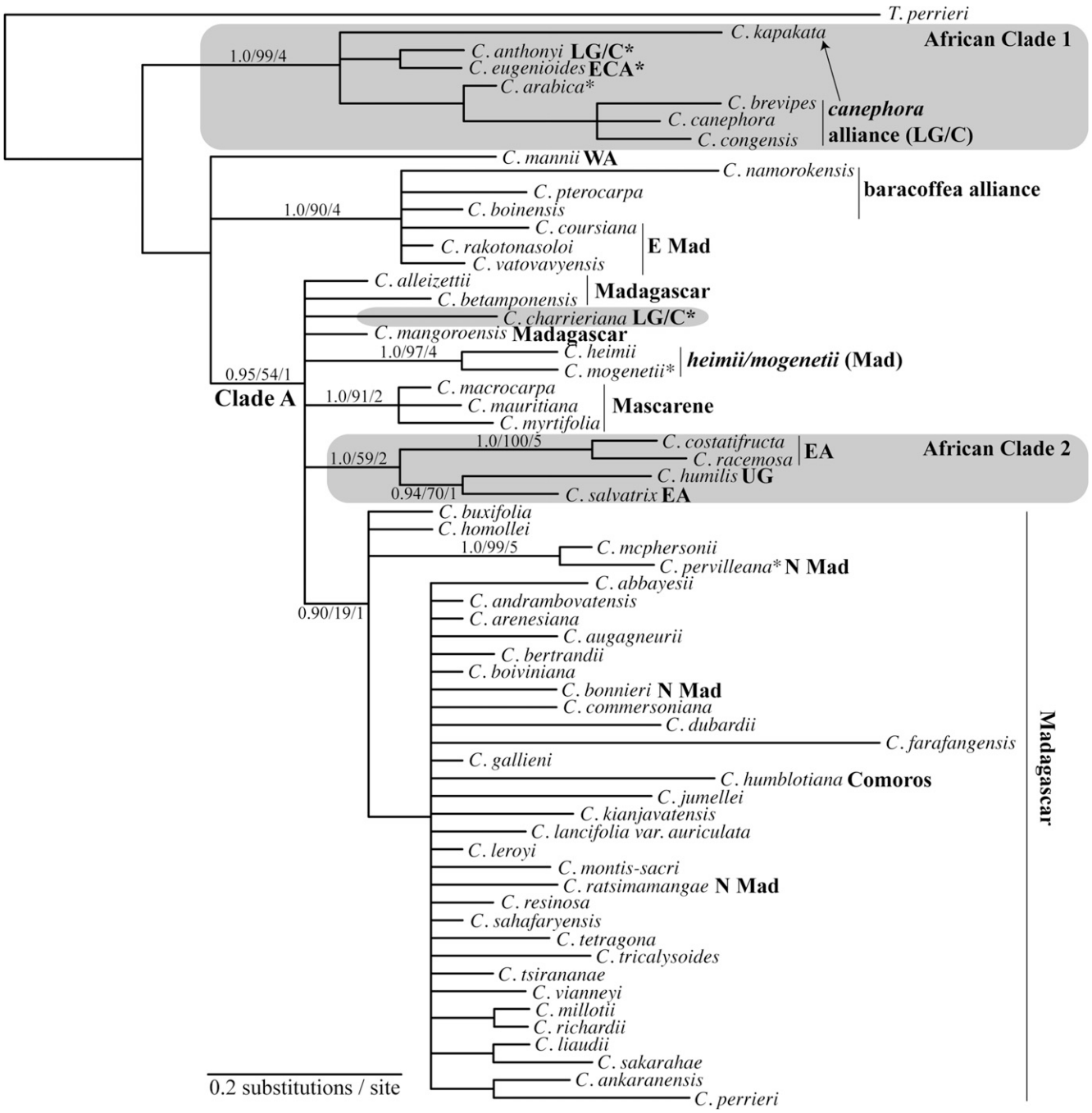

FIG. 3. The Bayesian majority rule tree showing all compatible partitions of the COS3 nuclear data set. Well-supported branches are labeled with support metrics and clades discussed in the text are labeled as in Fig. 2. The A/IO clade does not receive strong support, so instead the novel Clade A has been labeled. Taxa marked with an asterisk exhibit an incongruent placement between the plastid and nuclear COS3 topologies.

BP = 63); most significantly their IO clade includes the baracoffea alliance.

Discussion

In the present study we combine previously published plastid sequence data with data from three new chloroplast regions and a new single-copy COSII nuclear region to estimate phylogenetic relationships in the genus *Coffea*. While our results are largely congruent with those of previous studies by Maurin et al. (2007), Anthony et al. (2010) and Davis et al. (2011), several new findings are evident that allow novel interpretations of the evolutionary history of the genus. In the COS3 results, the placement of clades is generally not strongly incongruous with ITS data presented by Maurin et al. (2007) and Davis et al. (2011), although this could be due to the lack of support retrieved for ITS-based relationships, rather than agreement between these data sets. There is one obvious exception. In both the ITS analysis of Maurin et al. (2007) and Davis et al. (2011), the UG species are sister to the *canephora* alliance (PP = 71/*b* = 1; and PP = 1.0/BP = 80, respectively), rather than being sister to EA species in African Clade 2 as reported here. In the plastid analyses and combined plastid-ITS analyses (see below) of Maurin et al. (2007)

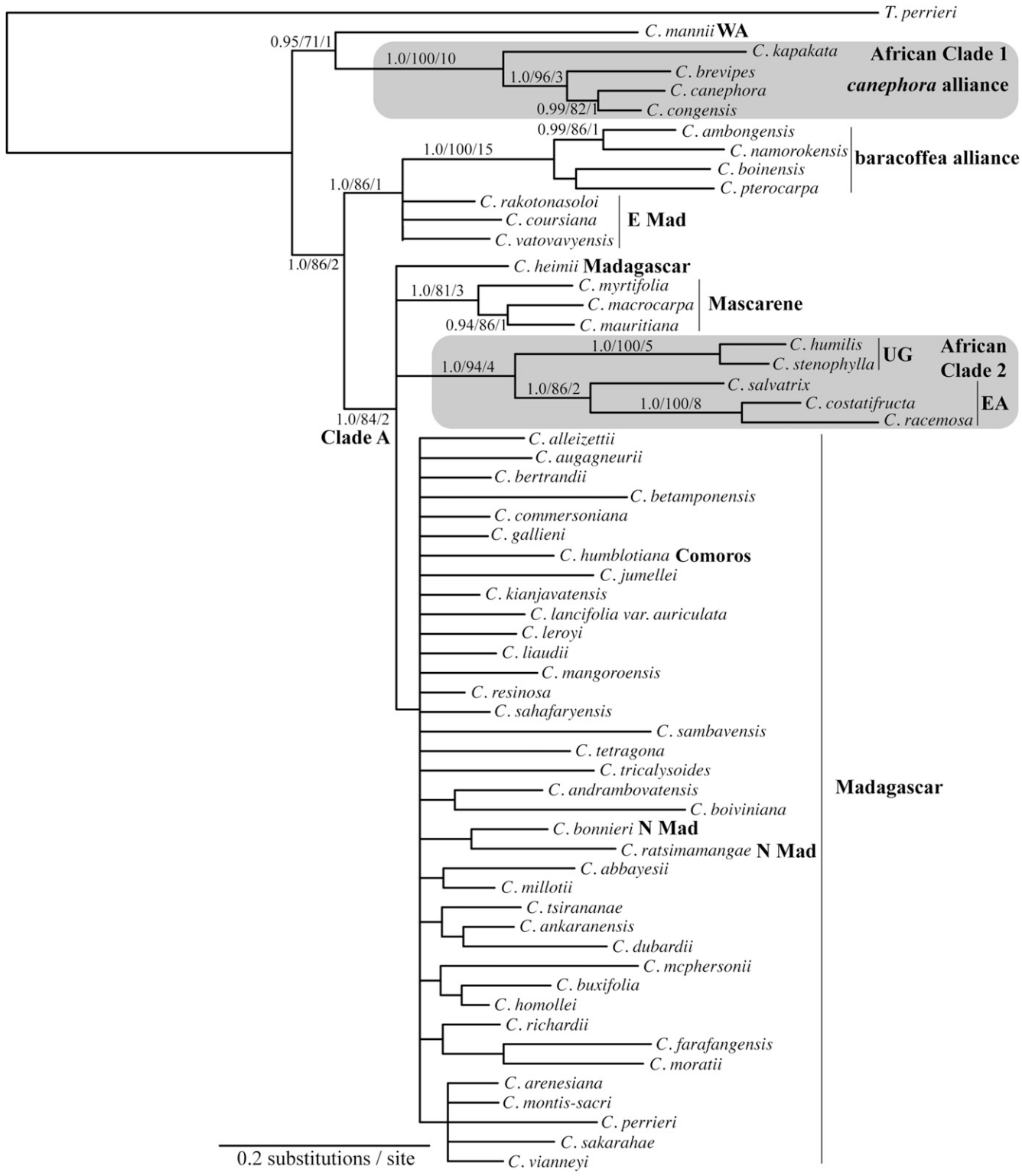

FIG. 4. The Bayesian majority rule tree showing all compatible partitions of the combined plastid and COS3 data set. Well-supported branches are labeled with support metrics and clades discussed in the text are labeled as in Fig. 2. The A/IO clade does not receive strong support, so instead the novel Clade A has been labeled. Note that taxa showing incongruent placement in Figs. 2 and 3 were excluded from these analyses.

and Davis et al. (2011), the UG species are placed with EA species. The ITS data of Maurin et al. (2007) and Davis et al. (2011) is concordant with the COS3 data in the support of the *canephora* alliance (BP = 86/b = 3; PP = 1.0/BP = 93, respectively) and the baracoffea alliance, less *C. pterocarpa* (BP <50/*b* = 1; PP = 1.0/BP = 65).

The results of our combined plastid/COS3 phylogenetic analyses also show significant congruence with previously published results. In the combined plastid-ITS studies of Maurin et al. (2007) and Davis et al. (2011), the LG/C clade (BP = 86/b = 4; PP 1.0/BP = 61, respectively), UG clade (BP = 100/b = 11; PP 1.0/BP = 100, respectively), EA clade

(consistently retrieved but lacking any significant support in both studies), baracoffea alliance clade (BP = 100/b = 6; PP 1.0/PP 99, respectively) and Mascarene clade (BP = 98/b = 7; PP 1.0/BP = 100, respectively) were also retrieved. The well-supported African Clade 2 in our tree is roughly equivalent to the UG/EA clade, which was also retrieved by Davis et al. (2011; PP = 1.0/BP = 64), with the absence of the ECA species that were removed in our analysis due to phylogenetic incongruence.

Patterns of Phylogenetic Incongruence Between Chloroplast and Nuclear Data—A comparison of the phylogenetic reconstructions resulting from the combined plastid and COS3 data sets reveals several examples of well-supported or "hard" phylogenetic incongruence. Phylogenetic incongruence between chloroplast and nuclear genomes is somewhat common in plant systematics, and while incongruent patterns in phylogenetic data tend to complicate interpretations of evolutionary history, the information provided by these patterns has been proposed as a "window into genome history" (Wendel and Doyle 1998). While various explanations have been proposed for hard phylogenetic incongruence, the most likely processes to be implicated in this study are hybridization and introgression, rapid diversification, and lineage sorting.

When considering the possibility of hybridization, it is important to note that *Coffea* trees grown in cultivation have been shown to exhibit considerable cross fertility, albeit with reduced seed set (Coulibaly et al. 2002; N'Diaye et al. 2007). Potential representatives of hybrid individuals have only rarely been reported in wild populations (Mahé et al. 2007; Gomez et al. 2009), but a hybridization event has been implicated in the origin of the allotetraploid species *C. arabica* (Raina et al. 1998; Lashermes et al. 1999; Maurin et al. 2007). The samples of *C. heterocalyx* used by Lashermes et al. (1997), Cros et al. (1998) and Maurin et al. (2007), most likely represent a wild hybrid (ITS homogenization is evident) between *C. liberica* and *C. eugeniodes*. Furthermore, controlled crosses of *C. arabica* and tetraploid individuals of *Coffea ebracteolatus* (formerly *Psilanthus ebracteolatus*) produced through colchicine treatment have been shown to produce a limited number of fertile offspring (Couturon et al. 1998). While these examples highlight the potential for hybridization and introgression among even distantly related species, it remains unclear how prevalent this may have been in the evolutionary history of the genus. In the sections that follow, we describe patterns of hard phylogenetic incongruence suggested by our results and discuss the implications of this incongruence with respect to interpreting the evolutionary history of the genus.

The most prominent patterns of phylogenetic incongruence in our analyses are observed in the clade membership and placement of African *Coffea* species. Among the previous phylogenetic studies in the genus *Coffea*, Maurin et al. (2007) and Davis et al. (2011) represent the only analyses utilizing sequence data from both the chloroplast and nuclear genomes. They found moderate support for incongruence between plastid and nuclear topologies in the placement of three species endemic to Upper Guinea (UG), which form a well-supported clade. While we were only able to sequence the COS3 locus from one of the two UG species sampled in this study (*C. humilis*), we find no evidence for incongruence between our plastid and nuclear COS3 data, both placing this species within African Clade 2, including East African species.

The results of our combined plastid analyses are largely consistent with the results of Maurin et al. (2007), Anthony et al. (2010) and Davis et al. (2011), but the placement of a few species markedly differs in our COS3 results. Consistent with the results of Anthony et al. (2010) and Davis et al. (2011), the plastid topology confidently places the Lower Guinea species *C. charrieriana* within African clade 1, clearly outside of the plastid A/IO clade (Fig. 2). In contrast, the nuclear COS3 topology places *C. charrieriana* within Clade A, but in an unresolved position near the base of this large clade (Fig. 3). Anthony et al. (2010) did not examine nuclear data and the ITS data of Davis et al. (2011) does not confidently place *C. charrieriana*. It is possible that the incongruent placement of *C. charrieriana* is the product of ancient hybridization that resulted in chloroplast capture, but given the unresolved placement of this taxon in the COS3 topology, we feel that this interpretation is speculative in the absence of more comprehensive sampling of African *Coffea* and well-supported relationships.

Our results also suggest hard phylogenetic incongruence in the placement of East-Central African (ECA) species *C. eugenioides*, *C. arabica*, and the LG/C species *C. anthonyi*. The plastid topology strongly supports a clade composed of *C. eugenioides*, *C. arabica*, and *C. anthonyi* that is confidently placed within the A/IO clade, consistent with the results of previous studies (Maurin et al. 2007; Anthony et al. 2010; Davis et al. 2011). The nuclear COS3 topology fails to find significant support for the monophyly of this clade (*C. arabica*, *C. eugenioides, C. anthonyi*), but confidently places these species in African Clade 1 with several LG/C species of the *canephora* alliance and clearly outside of Clade A, which is largely composed of Malagasy species. An incongruent placement for *C. arabica* might be expected given the allopolyploid origin of this species by putative hybridization between the marginally sympatric (see Fig. 1 in Maurin et al. 2007) *C. canephora* (primarily LG/C range; Maurin et al. 2007) and *C. eugenioides* (ECA; Raina et al. 1998; Lashermes et al. 1999), or recent ancestors of these species. But given that this incongruence in A/IO clade membership is also observed in *C. anthonyi* (LG/C) and *C. eugenioides*, this explanation is insufficient, and thus will require further validation with more sequence data from the nuclear genome and a more extensive sampling within these species.

Paraphyly of African and Indian Ocean Coffea—Phylogenetic analyses performed on the combined plastid alignment provide further evidence for the paraphyly of African *Coffea* species with respect to Indian Ocean *Coffea* (Fig. 3 in Maurin et al. 2007; Fig. 1 in Anthony et al. 2010; Fig. 3 in Davis et al. 2011). The phylogenetic results of our nuclear COS3 data are consistent with this conclusion, but differ from the plastid topology in the placement of several African species (see above). After removal of incongruent taxa for the combined analysis of the nuclear COS3 and plastid sequence data, we recover strong support for the paraphyletic placement of African *Coffea* species, represented by Clade 1 (*canephora* alliance) and Clade 2 (roughly equivalent to the EA/UG clade), consistent with the results of Maurin et al. (2007) and Davis et al. (2011). The COS3 topology supports a placement of the baracoffea alliance and E Mad species in a polytomy with *C. mannii* and Clade A composed of African Clade 2 and the remaining Malagasy *Coffea* (Fig. 3), although these relationships are weakly supported. In the combined analysis of plastid and COS3 data, however, this sister relationship

receives strong support (Fig. 4). By placing the baracoffea alliance/E Mad species outside of the A/IO clade, these results represent the first evidence for the paraphyly of Indian Ocean *Coffea* species and may have important implications for interpreting the evolutionary history of the genus. This relationship conflicts with the findings of Davis et al. (2011), as they receive support for both an A/IO clade (PP = 1.0/BP = < 50) and an Indian Ocean clade (IO) composed of all species from Madagascar (including the baracoffea alliance) and the Mascarenes (PP = 0.99/BP = 63).

Evolutionary History of Coffea and Identification of Early Diverging Lineages in Africa—Anthony et al. (2010) posit that the low plastid sequence divergence between Malagasy and African species provides evidence for a "rapid and radial mode of speciation in *Coffea* subgenus *Coffea*." Specifically, they suggest that *Coffea* species diversity is the product of a recent (< 1 Ma) adaptive radiation originating from an ancestral species in the LG/C region that colonized Upper Guinea to the west and East-Central and East Africa to the east. Their model then proposes that the colonization of Madagascar occurred from an ancestral East African species and an insular Malagasy radiation ensued. Based on our results and those of Davis et al. (2011) this hypothesis for the evolutionary history of *Coffea* is largely untenable. Our sequence data suggest that the evolutionary history of African *Coffea* is much more complicated than proposed by Anthony et al. (2010) in that there has been considerable interplay between different geographic regions of Africa (Upper Guinea, Lower Guinea/Congolia, East Central Africa, and East Africa), rather than unidirectional dispersal from the Lower Guinea region (see Fig. 4 in Anthony et al. 2010). Furthermore, the paraphyletic status of Indian Ocean *Coffea* (Figs. 1–3) presented here calls into question the feasibility of a single colonization of Madagascar.

The level of sequence divergence observed at the COS3 locus and in our plastid data allows the identification of early diverging lineages in the genus that have either not been retrieved or received significant phylogenetic support in previous studies of chloroplast and nuclear sequence data (Cros et al. 1995, 1998; Lashermes et al. 1997; Davis et al. 2005; Maurin et al. 2007; Anthony et al. 2010; Davis et al. 2011). Results from our combined data set (plastid plus COS3; Fig. 4) show that the earliest divergence in the genus is between a clade composed of species (African Clade 1) from the African LG/C regions and West Africa (*C. mannii*) and a large clade containing the rest of the genus *Coffea*. This is consistent with the idea of a center of origin for *Coffea* in West Africa as proposed by Anthony et al. (2010). The well-supported position of *C. mannii* (formerly *Psilanthus mannii*) as sister to the LG/C African Clade 1 (see also Fig. 2 in Maurin et al. 2007) indicates that this morphologically diverse group of species, which is now part of *Coffea* but previously comprising numerous (and potentially all) African *Psilanthus* species (see Davis et al. 2011), represents an early divergence within the genus. In particular, this group of frican species (e.g. *C. mannii*, *C. melanocarpa,* and *C. ebracteolatus*) all possess short styles, which are positioned near the base of long corolla tubes, well below the anthers. All other *Coffea* species, except the Asian representatives (which were until recently also included in *Psilanthus*; Davis 2010, 2011; Davis et al. 2011), have long styles that are positioned at or above the mouth of the corolla and level with the anthers. Sampling of Asian *Coffea* species (i.e. former Asian *Psilanthus*) and the enigmatic *C. rhamnifolia* (Davis et al. 2005, 2011) will be particularly interesting in terms of both resolving the most early-diverging branches and ascertaining the polarity of style length character state changes in the genus.

In combination with other findings presented here, we concur with Davis et al. (2011) that the evolutionary history of the genus is far more complex than is generally appreciated. It will be interesting to further explore the patterns of incongruence identified with additional COSII regions or other new molecular markers from the nuclear genome. Future work toward resolving the phylogenetic position of all major clades in the genus will require expanding the taxon sampling to include many more representatives of African *Coffea* as well as a geographically diverse sample of species formerly belonging to *Psilanthus*, which have extended the geographical range of *Coffea* into Asia and Australasia. Furthermore, the new phylogenetic placement of the baracoffea alliance and E Mad species with respect to the rest of the Malagasy *Coffea* will need to be tested with more independent nuclear loci. If this phylogenetic placement is consistently supported, it could suggest a biogeographic scenario in which Madagascar was colonized by two separate dispersal events from the African mainland.

Acknowledgments. This work was funded by a U. S. National Science Foundation Doctoral Dissertation Enhancement Project award to ADY and MDN, and graduate student research grants awarded to MDN from the American Society of Plant Taxonomists, the Society of Systematic Biologists, and the Duke University Graduate School. Field collections in Madagascar were carried out under collaborative agreements between the Parc Botanique et Zoologique de Tsimbazaza (PBZT), the University of Antananarivo (Madagascar), the Association Nationale de Gestion des Aires Protegées (ANGAP) and the Royal Botanic Gardens, Kew. Special thanks are extended to Dr. H. Ralimanana (K), Dr. F. Rakotonasolo (PBZT), Dr. Y. Jaufeerally-Fakim (University of Mauritius), the staff of the Kianjavato Coffee Research Station (FOFIFA), and the staff of the Centre ValBio (ICTE – SUNY Stony Brook) for help with fieldwork and specimen collection in Madagascar. The authors wish to thank Dr. F. Anthony, Dr. D. Lorence, and Dr. J. Horn for providing tissue samples that were used in this study.

Literature Cited

Anthony, F., L. Diniz, M. Combes, and P. Lashermes. 2010. Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. *Plant Systematics and Evolution* 285: 51–64.

Bouck, A. and T. Vision. 2007. The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* 16: 907–924.

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

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

Couturon, E., P. Lashermes, and A. Charrier. 1998. First intergeneric hybrids (*Psilanthus ebracteolatus* Hiern x *Coffea arabica* L.) in coffee trees. *Canadian Journal of Botany* 76: 542–546.

Cros, J., M. Combes, N. Chabrillange, C. Duperray, A. Angles, and S. Hamon. 1995. Nuclear DNA content in the subgenus *Coffea* (Rubiaceae): inter-and intra-specific variation in African species. *Canadian Journal of Botany* 73: 14–20.

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

Cubry, P., P. Musoli, H. Legnaté, D. Pot, F. de Bellis, V. Poncet, F. Anthony, M. Dufour, and T. Leroy. 2008. Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. *Genome* 51: 50–63.

Davis, A. P. 2010. Six species of *Psilanthus* transferred to *Coffea* (Coffeeae, Rubiaceae). *Phytotaxa* 10: 41–45.

Davis, A. P. 2011. *Psilanthus mannii* Hook. f., the type species of *Psilanthus* Hook. f., transferred to *Coffea* L. *Nordic Journal of Botany* 29: 471–472.

Davis, A. P. and F. Rakotonasolo. 2008. A taxonomic revision of the baracoffea alliance: nine remarkable *Coffea* species from western Madagascar. *Botanical Journal of the Linnean Society* 158: 355–390.

Davis, A. P., D. Bridson, and F. Rakotonasolo. 2005. A reexamination of *Coffea* subgenus *Baracoffea* and comments on the morphology and classification of *Coffea* and *Psilanthus* (Rubiaceae - Coffeeae). Pp. 399–420 in *A Festschrift for William G. D'Arcy: the legacy of a taxonomist*, eds. Keating R., V. Hollowell, and T. Croat. St. Louis: Missouri Botanical Garden.

Davis, A. P., R. Govaerts, D. Bridson, and P. Stoffelen. 2006. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* 152: 465–512.

Davis, A. P., M. Chester, O. Maurin, and M. Fay. 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, A. P., R. Govaerts, D. Bridson, R. Ruhsam, J. Moat, and N. Brummitt. 2009. A global assessment of distribution, diversity, endemism and taxonomic effort in the Rubiaceae. *Annals of the Missouri Botanical Garden* 96: 68–78.

Davis, A. P., F. Rakotonasolo, and P. De Block. 2010. *Coffea toshii* sp. nov. (Rubiaceae) from Madagascar. *Nordic Journal of Botany* 28: 134–136.

Davis, A. P., J. Tosh, N. Ruch, and M. Fay. 2011. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Botanical Journal of the Linnean Society* 167: 357–377.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing the significance of incongruence. *Cladistics* 10: 315–319.

Gomez, C., S. Dussert, P. Hamon, S. Hamon, A. de Kochko, and V. Poncet. 2009. Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evolutionary Biology* 9: 167.

Hamon, P., P. Duroy, C. Dubreuil-Tranchant, P. Mafra D'Almeida Costa, C. Duret, N. Razafinarivo, E. Couturon, S. Hamon, A. de Kochko, V. Poncet, and R. Guyot. 2011. Two novel Ty1-copia retrotransposons isolated from coffee trees can effectively reveal evolutionary relationships in the *Coffea* genus (Rubiaceae). *Molecular Genetics and Genomics* 285: 447–460.

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

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

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

Lin, C., L. Mueller, J. McCarthy, D. Crouzillat, V. Pétiard, and S. D. Tanksley. 2005. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theoretical and Applied Genetics* 112: 114–130.

Mahé, L., D. Le Pierrès, M. Combes, and P. Lashermes. 2007. Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors, *Coffea canephora*, in an exceptional sympatric zone in New Caledonia. *Genome* 50: 316–324.

Maurin, O., A. P. Davis, M. Chester, E. F. Mvungi, Y. Jaufeerally-Fakim, and M. F. Fay. 2007. Towards a phylogeny for *Coffea* (Rubiaceae): Identifying well-supported lineages based on nuclear and plastid DNA sequences. *Annals of Botany* 100: 1565–1583.

Mueller, L., T. H. Solow, N. Taylor, B. Skwarecki, R. Buels, J. Binns, C. Lin, M. Wright, R. Ahrens, Y. Wang, E. Herbst, E. Keyder, N. Menda, D. Zamir, and S. D. Tanksley. 2005. The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. *Plant Physiology* 138: 1310–1317.

Murray, M. G. and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321–4326.

N'Diaye, A., M. Noirot, S. Hamon, and V. Poncet. 2007. Genetic basis of species differentiation between *Coffea liberica* Hiern and *C. canephora* Pierre: analysis of an interspecific cross. *Genetic Resources and Crop Evolution* 54: 1011–1021.

Nowak, M. D., A. P. Davis, F. Anthony, and A. D. Yoder. 2011. Expression and trans-specific polymorphism of self-incompatibility RNases in *Coffea* (Rubiaceae). *PLoS ONE* 6: e21019, doi: 10.1371/journal.pone.0021019.

Nylander, J. A. A. 2004. MrModeltest v2. Program distributed by the author. Uppsala: Evolutionary Biology Centre, Uppsala University.

Orozco-Castillo, C., K. Chalmers, W. Powell, and R. Waugh. 1996. RAPD and organelle specific PCR re-affirms taxonomic relationships within the genus *Coffea*. *Plant Cell Reports* 15: 337–341.

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

Rambaut, A. and A. J. Drummond. 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

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

Samson, N., M. Bausher, S. Lee, R. Jansen, and H. Daniell. 2007. The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotechnology Journal* 5: 339–353.

Sorenson, M. D. and E. A. Franzosa. 2007. TreeRot, version 3. Boston: Boston University.

Sukumaran, J. and M. T. Holder. 2010. DendroPy: A Python library for phylogenetic computing. *Bioinformatics* 26: 1569–1571.

Stoffelen, P., M. Noirot, E. Couturon, S. Bontems, P. De Block, and F. Anthony. 2009. *Coffea anthonyi*, a new self-compatible Central African coffee species, closely related to an ancestor of *Coffea arabica*. *Taxon* 58: 133–140.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland: Sinauer Associates.

Thiers, B. 2012 [and continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. http://sweetgum.nybg.org/ih/.

Tosh, J., A. P. Davis, S. Dessein, P. De Block, S. Huysmans, M. Fay, E. Smets, and E. Robbrecht. 2009. Phylogeny of *Tricalysia* (Rubiaceae) and its relationships with allied genera based on plastid DNA data: resurrection of the genus *Empogona*. *Annals of the Missouri Botanical Garden* 96: 194–213.

Wendel, J. and J. Doyle. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. PP. 265–295 in *Molecular systematics of plants* 2, eds. D. E. Soltis, P. S. Soltis, and J. J. Doyle. Berlin: Kluwer Academic Publishers.

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.

Wu, F., L. Mueller, D. Crouzillat, V. Pétiard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the Euasterid plant clade. *Genetics* 174: 1407–1420.

Yoder, A. D. and M. D. Nowak. 2006. Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology Evolution and Systematics* 37: 405–431.

Appendix 1. Voucher information for specimens used in the current study is provided as follows: species, geographic distribution, voucher specimen (voucher location), [GenBank accession numbers] *atpF, rpolb1, rpolb2,* COS3, Kew DNA Bank accession number. Living voucher specimens are identified by their specimen numbers in cultivation at the Duke University teaching collection (Duke), the IRD collections in Montpellier (IRD), and the FOFIFA Kianjavato Coffee Research Station in Madagascar (Kianjavato).

Coffea abbayesii J.-F.Leroy, South East Madagascar, *A601* (Kianjavato), -, JQ856329, JQ856179, JQ856090, -. *C. alleizettii* Dubard, North West Madagascar, *A803* (Kianjavato), JQ856248, JQ856286, JQ856145, JQ856056, 33580. *C. ambongensis* J.-F. Leroy, West Madagascar, *Rakotonasolo 68* (K), JQ856223, JQ856276, JQ856129, -, 33420. *C. andrambovatensis* J.-F. Leroy, East Madagascar, *Davis 2322* (K), JQ856203, JQ856264, JQ856113, JQ856031, 12270. *C. ankaranensis* A. P. Davis & Rakotonas, North Madagascar, *Nowak 49d* (K), JQ856244, JQ856311, JQ856171, JQ856081, -. *C. anthonyi* Stoff. & F. Anthony, West Lower Guinea, *OD60* (IRD), -, -, -, JQ979042, -. *C. arabica* L., North East Central Africa, *66-37* (Duke) JQ856245, JQ856317, JQ856175, JQ856086,

-. *C. arenesiana* J.-F. Leroy, East Madagascar, *A403* (Kianjavato), -, JQ856282, JQ856139, JQ856050, 33573. *C. augagneurii* Dubard, North Madagascar, *A966* (Kianjavato), -, JQ856288, JQ856148, JQ856058, 33583. *C. bertrandii* A. Chev., South Madagascar, *A303* (Kianjavato), JQ856228, JQ856271, JQ856134, JQ856046, 33567. *C. betamponensis* Portères & J.-F. Leroy, East Madagascar, *A573* (Kianjavato), JQ856246, -, JQ856143, JQ856054, 33577. *C. boinensis* A. P. Davis & Rakotonas, West Madagascar, *Davis 2502* (K), -, JQ856262, JQ856109, JQ856028, 12234. *C. boiviniana* (Baill.) Drake in Grandid., North Madagascar, *Nowak 56b* (K), -, -, JQ856159, JQ856067, -. *C. bonnieri* Dubard, North Madagascar, *Nowak 32e* (K), JQ856235, JQ856284, JQ856154, JQ856062, -. *C. brevipes* Hiern, West Central Africa, *JB56* (IRD), -, -, -, JQ979039, -. *C. buxifolia* A. Chev., Central Madagascar, *Davis 1012* (K), JQ856224, JQ856277, JQ856130, JQ856042, 33421. *C. canephora* Pierre ex A. Froehner, West and South Africa, *BB60* (IRD), -, JQ856318, JQ856176, JQ856087, -. *C. charrieriana* Stoff. & F. Anthony, Central Africa & Cameroon, *OA1* (IRD), -, -, -, JQ979041, -. *C. commersoniana* (Baill.) A. Chev., South East Madagascar, *Davis 2708* (K), JQ856205, JQ856266, JQ856115, JQ856033, 12331. *C. congensis* A. Froehner, West Central Africa, *CB68* (IRD), -, -, -, JQ979034, -. *C. costatifructa* Bridson, East Africa, *OH* (IRD), -, -, -, JQ979043, -. *C. coursiana* J.-F. Leroy, East Madagascar, *A570* (Kianjavato), -, -, JQ856181, JQ856092, -. *C. dubardii* Jum., North West Madagascar, *Nowak 41-1e* (K), JQ856237, JQ856297, JQ856156, JQ856064, -. *C. eugenioides* S. Moore, West Central and East Africa, *A16-G17* (Kianjavato), -, JQ856319, -, JQ979035, -. *C. farafangensis* J.-F. Leroy, South East Madagascar, *A208-A* (Kianjavato), -, JQ856333, JQ856184, -, -. *C. gallieni* Dubard, North Madagascar, *Nowak 38b* (K), -, JQ856301, JQ856163, JQ856073, -. *C. heimii* J.-F. Leroy, North Madagascar, *A516* (Kianjavato), JQ856239, JQ856307, JQ856141, JQ856052, 33575. *C. homollei* J.-F. Leroy, East Madagascar, *A574* (Kianjavato), JQ856247, JQ856330, JQ856180, JQ856091, 33578. *C. humblotiana* Baill., Grande Comore Island, *A230* (Kianjavato), JQ856249, JQ856331, JQ856182, JQ856043, -. *C. humilis* A. Chev., West Africa, *F258* (IRD), -, -, -, JQ979038, -. *C. jumellei* J.-F. Leroy, North Madagascar, *A974* (Kianjavato), JQ856255, JQ856290, JQ856149, JQ856059, 33586. *C. kapakata* (A. Chev.) Bridson, South Central Africa & West Angola, *OK* (IRD), -, -, -, JQ979044, -. *C. kianjavatensis* J.-F. Leroy, East Madagascar, *A603* (Kianjavato), JQ856259, JQ856338, JQ856193, JQ856104, -. *C. lancifolia* var. *auriculata* J.-F. Leroy, East Madagascar, *A320-1* (Kianjavato), JQ856232, JQ856281, JQ856138, JQ856049, 33572. *C. lancifolia* var. *lancifolia* A. Chev., East Madagascar, *Davis 2310* (K), -, JQ856337, JQ856191, JQ856100, 12269. *C. leroyi* A. P. Davis, East Madagascar, *A315* (Kianjavato), JQ856257, JQ856339, JQ856194, JQ856102, 33570. *C. liaudii* A. P. Davis, East Madagascar, *Davis 3087* (K), JQ856210, -, -, JQ856036, 21270. *C. macrocarpa* A. Rich., Mauritius, *Nowak 62a* (K), -, JQ856303, JQ856165, JQ856075, -. *C. mangoroensis* Portères, East Madagascar, *A401* (Kianjavato), JQ856254, JQ856335, JQ856189, JQ856098, -. *C. mannii* (Hook. f.) A. P. Davis, West Africa, *F247* (IRD), -, -, -, JQ979036, -. *C. mauritiana* Lam., Mauritius and Reunion, *Nowak 68* (K), -, JQ856315, JQ856174, JQ856084, -. *C. mcphersonii* A. P. Davis & Rakotonas, North East Madagascar, *A977* (Kianjavato), -, JQ856291, JQ856150, JQ856060, 33587. *C. millotii* J.-F. Leroy, East Madagascar, *A206* (Kianjavato), JQ856251, -, JQ856186, JQ856095, -. *C. mogenetii* Dubard, North Madagascar, *Davis 4511* (K), JQ856222, JQ856275, JQ856128, JQ856041, 33419. *C. montis-sacri* A. P. Davis, East Madagascar, *A321* (Kianjavato), JQ856258, JQ856280, JQ856195, JQ856103, -. *C. moratii* A. P. Davis & Rakotonas, West Madagascar, *Maurin 76* (K), JQ856207, -, JQ856118, -, 19859. *C. myrtifolia* (A. Rich. ex DC.) J.-F. Leroy, Mauritius, *Nowak 63a* (K), JQ856241, JQ856305, JQ856166, JQ856076, -. *C. namorokensis* A. P. Davis & Rakotonas, West Madagascar, *Davis 2528* (K), JQ856211, -, JQ856121, JQ856037, 21271. *C. perrieri* Drake ex Jum. & H. Perrier, Madagascar, *A12-1* (Kianjavato), JQ856226, JQ856278, JQ856132, JQ856044, 33565. *C. pervilleana* (Baill.) Drake in Grandid., North Madagascar, *A540-1* (Kianjavato), JQ856250, JQ856334, JQ856185, JQ856094, -. *C. pterocarpa* A. P. Davis & Rakotonas, West Madagascar, *Davis 2538* (K), JQ856204, JQ856265, JQ856114, JQ856032, 12275. *C. racemosa* Lour., East and South Africa, *F253* (IRD), -, -, -, JQ979037, -. *C. rakotonasoloi* A. P. Davis, East Madagascar, *Davis 2284* (K), JQ856200, JQ856263, JQ856110, JQ856029, 12261. *C. ratsimamangae* J.-F. Leroy ex A. P. Davis & Rakotonas, North Madagascar, *Nowak 11a* (K), JQ856240, JQ856300, JQ856162, JQ856071, -. *C. resinosa* (Hook. f.) Radlk., East Madagascar, *A913* (Kianjavato), JQ856215, -, JQ856198, JQ856106, -. *C. richardii* J.-F. Leroy, East Madagascar, *A817* (Kianjavato), JQ856216, JQ856332, JQ856183, -, -. *C. sahafaryensis* J.-F. Leroy, North East Madagascar, *Nowak 03a* (K), -, JQ856292, JQ856158, JQ856065, -. *C. sakarahae* J.-F. Leroy, South Central Madagascar, *Davis 2167* (K), JQ856217, JQ856273, -, JQ856039, 33412. *C. salvatrix* Swynn. & Phillipson, East and South Africa, *LA* (IRD), -, -, -, JQ979040, 19445. *C. sambavensis* A. P. Davis & Rakotonas, North East Madagascar, *Tosh 399* (K), JQ856218, -, JQ856125, -, 33413. *C. stenophylla* G. Don, West Africa, *F232* (IRD)-, -, -, -, -. *C. tetragona* Jum. & H. Perrier, North West Madagascar, *A252* (Kianjavato), JQ856227, JQ856261, JQ856133, JQ856045, 33566. *C. tricalysioides* J.-F. Leroy, North Madagascar, *Nowak 39e* (K), JQ856243, JQ856310, JQ856170, JQ856079, -. *C. tsirananae* J.-F. Leroy, North Madagascar, *Nowak 19i* (K), JQ856233, JQ856293, JQ856151, -, -. *C. vatovavyensis* J.-F. Leroy, East Madagascar, *A308* (Kianjavato), JQ856229, JQ856279, JQ856135, JQ856047, 33568. *C. vianneyi* J.-F. Leroy, South East Madagascar, *A946* (Kianjavato), -, JQ856287, JQ856146, JQ856057, 33581. *Tricalysia perrieri* Homolle ex Randriamb. & De Block, West Madagascar, *A738* (Kianjavato), -, JQ856328, JQ856177, JQ856088, -.