Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on Coffea canephora (Robusta coffee)

Jean-Claude Charr, Andrea Garavito, Christophe Guyeux, Dominique Crouzillat, Patrick Descombes, Coralie Fournier, Serigne Ly, Eva Raharimalala, Jean-Jacques Rakotomalala, Piet Stoffelen, et al.

Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on *Coffea canephora* (Robusta coffee).

Short title: Complex evolutionary history of coffees

Jean-Claude Charr[a], Andrea Garavito[b], Christophe Guyeux[a], Dominique Crouzillat[c], Patrick Descombes[d], Coralie Fournier[d], Serigne N. LY[e], Eva N. Raharimalala[f], Jean-Jacques Rakotomalala[f], Piet Stoffelen[g], Steven Janssen[g], Perla Hamon[e] and Romain Guyot[eh*].

The two first authors contributed equally to this work.

[a] Femto-ST Institute, UMR 6174 CNRS, Université de Bourgogne Franche-Comté, France; christophe.guyeux@univ-fcomte.fr; jean-claude.charr@univ-fcomte.fr
[b] Departamento de Ciencias biológicas, Facultad de Ciencias Exactas y Naturales, Universidad de Caldas, Manizales, Colombia; current address UMR INRAE/UCA 1095 Génétique, Diversité et Ecophysiologie des Céréales (GDEC), Clermont-Ferrand, France; neagef@gmail.com
[c] Nestle Centre Tours, BP 49716, F-37097 Tours cedex 2, France; dominique.crouzillat@rdto.nestle.com
[d] Nestle Research Lausanne, Switzerland; patrick.descombes@rd.nestle.com; coralie.fournier@rd.nestle.com
[e] Institut de Recherche pour le Développement, UMR DIADE, CIRAD, Université de Montpellier, France; ndiawarly@hotmail.fr; perla.hamon@ird.fr; romain.guyot@ird.fr
[f] FOFIFA, BP 1444, Ambatobe, Antananarivo 101, Madagascar; evanathie.rah@gmail.com; rakotomalala.jjr@moov.mg
[g] Meise Botanic Garden, Nieuwelaan 38, BE-1860 Meise, Belgium; piet.stoffelen@plantentuinmeise.be; steven.janssens@plantentuinmeise.be
[h] Department of Electronics and Automatization, Universidad Autónoma de Manizales, Manizales, Colombia

*Corresponding author at: IRD, UMR DIADE, BP 64501, F-34394 Montpellier Cedex 5, France

E-mail address: romain.guyot@ird.fr (Romain Guyot)

Abstract

For decades coffees were associated with the genus *Coffea.* In 2011, the closely related genus *Psilanthus* was subsumed into *Coffea*. However, results obtained in 2017—based on 28,800 nuclear SNPs—indicated that there is not substantial phylogenetic support for this incorporation. In addition, a recent study of 16 plastid full-genome sequences highlighted an incongruous placement of *Coffea canephora* (Robusta coffee) between maternal and nuclear trees. In this study, similar global features of the plastid genomes of *Psilanthus* and *Coffea* are observed. In agreement with morphological and physiological traits, the nuclear phylogenetic tree clearly separates *Psilanthus* from *Coffea* (with exception to *C. rhamnifolia,* closer to *Psilanthus* than to *Coffea*). In contrast, the maternal molecular tree was incongruent with both morphological and nuclear differentiation, with four main clades observed, two of which include both *Psilanthus* and *Coffea* species, and two with either *Psilanthus* or *Coffea* species. Interestingly, *Coffea* and *Psilanthus* taxa sampled in West and Central Africa are members of the same group. Several mechanisms such as the retention of ancestral polymorphisms due to incomplete lineage sorting, hybridization leading to homoploidy (without chromosome doubling) and alloploidy (for *C. arabica*) are involved in the evolutionary history of the coffee species. While sharing similar morphological characteristics, the genetic relationships within *C. canephora* have shown that some populations are well differentiated and genetically isolated. Given the position of its closely-related species, we may also consider *C. canephora* to be undergoing a long process of speciation with an intermediate step of (sub-)speciation.

Key words: Full plastid genome sequencing – Nuclear SNPs –Evolutionary history – Coffee species - Coffea canephora

1. Introduction

With the introduction of high-throughput sequencing technologies, additional levels in the field of molecular genetics have become available for a wide range of organisms, such as deep- and whole-genome sequencing data. For plants, and more specifically for crops and wild relatives of crops, several gigabases of sequence data have been generated, allowing researchers to reveal the evolutionary patterns and events of complex taxa such as *Citrus* (Wu et al., 2018), banana (D'Hont et al., 2012), peach (The International Peach Genome Initiative, 2013) or grapevine (Jaillon et al., 2007). For the last group, three distinct polyploidization events were observed, revealing the hexaploidal nature of the *Vitis vinifera* genome. In parallel, phylogenetic analyses have uncovered multiple incongruities existing between maternal and nuclear topologies for many more plant lineages than were initially accounted for: *Nothofagus* (Acosta and Premoli, 2010), *Lactuca* (Wang et al., 2013), Rosidae (Sun et al., 2015), *Osmorhiza* (Yi et al., 2015), Asian *Ixora* species, (Mouly et al., 2009), and *Coffea* (Guyeux et al., 2019). Part of these topological dissimilarities were caused by ancient hybridization events between two ancestral species followed by chromosome doubling, resulting in the emergence of new species (allopolyploid). This type of event has been reported for *Coffea arabica* (Lashermes et al., 1999), *Gossypium barbadense* (Liu et al., 2001), *Nicotiana tabaccum* (Chase et al., 2003) and *Brassica* (Osborn, 2004). In addition, hybridization did not always result in genome duplication but sometimes led to homoploid speciation which is often referred to as introgressive hybridization. Obviously, such events are difficult to detect and to distinguish from introgression. In their study, Yakimowski and Rieseberg (2014) considered 17 species to be putatively homoploid hybrids (e.g. *Senecio squalidus* (James and Abbott, 2005; Brennan et al., 2012), *Iris nelsonii* (Arnold, 1993; Taylor et al., 2013), *Pinus densata* (Wang et al., 2001; Gao et al., 2012), *Penstemon clevelandii* (Wolfe et al., 1998), *Paeonia anomala* (Pan et al., 2007), and three *Helianthus* (*H. anomalus*,

H. deserticola and *H. paradoxus*, Rieseberg, 1991; Lai et al., 2005). Homoploids and allopolyploids need to be reproductively isolated in order to avoid gene dilution from back-crosses with one of the parents. In these new hybrids, the change in mating preferences as well as the acquisition of new adaptive traits that increase the capability of colonizing new habitats can sometimes extremely facilitate the process of speciation (Chase et al., 2010; Servedio et al., 2011; Marques et al., 2016). Besides introgression and homoploidy, the retention of ancestral polymorphism due to incomplete lineage sorting (ILS) can also account for nuclear/cytoplasmic discrepancies as reported in pines trees (Zhou et al., 2017). At present, genome-scale approaches can resolve a significant number of phylogenetic incongruities and can help reconstruct a more accurate evolutionary history (Som 2015).

The genus *Coffea* belongs to the Rubiaceae, a large family within the Euasterid clade together with its sister genus *Psilanthus* (Robbrecht and Manen, 2006; Davis et al., 2007; Ferreira et al., 2019. Later, Davis et al. (2011) proposed the merging of the genus *Psilanthus* and sub-genus leading to the actual *Coffea* genus. With the inclusion of Psilanthus, the Coffea genus comprises 124 species (Hamon et al., 2017, Davis et al., 2019). Wild species of *Coffea s.s* (including only species of the former genus *Coffea*) are exclusively distributed throughout the African continent and the Western Indian Ocean Islands with approximately 50% of the species endemic to Madagascar (Davis et al., 2006, http://publish.plantnet-project.org/project/wildcofdb; Guyot et al., 2020). In the wild, the distribution of the *Coffea* species is species-specific with only a few cases of sympatry (growing side by side), and few putative hybrids reported (Berthaud and Charrier, 1988; Charrier, 1978; Noirot et al., 2016).

Psilanthus and *Coffea* species are diploids (Bouharmont, 1963; Louarn, 1972) with a chromosome number of x=11 and a total genome size ranging from 460 to 900 Mbp (Noirot et al., 2003, Razafinarivo et al., 2012), with the exception of *Coffea arabica*. To date, the only known natural polyploid species is *C. arabica,* an allotetraploid recently originated (Yu et al.,

2011) from two diploid progenitors: *C. canephora* and *C. eugenioides* as the maternal progenitor (Lashermes et al., 1999). *Coffea arabica* is self-fertile while all diploids are self-sterile with the exception of two species: *C. heterocalyx* (Stoffelen et al., 1996) and *C. anthonyi* (Stoffelen et al., 2009). The presence of short-styled included anthers in *Psilanthus* lead to possible self-fertility, yet with the possibility to out-pollinate (Charrier and Berthaud 1985). The creation of artificial interspecific hybrids and back-cross progenies between African diploid *Coffea* species (Louarn, 1993) in Ivory Coast, between African and Malagasy species in Madagascar (Charrier, 1978) and between *C. arabica* and *Psilanthus ebracteolatus* (Couturon et al., 1998;) demonstrated the potential of possible hybridization between allopatric species, and the lack of internal reproductive barriers.

The sequencing and assembly of the genome of the diploid *Coffea canephora* (Robusta coffee) has shown that its species was not subjected to a "recent" whole-genome duplication event (Denoeud et al., 2014). Using whole-genome approaches and complete plastid reconstruction, discrepancies between maternal and nuclear phylogenies of the two main cultivated coffee species (i.e. Arabica and Robusta) were clearly identified (Guyeux et al., 2019). While *Coffea arabica* belongs to the *Coffea* clade, *C. canephora,* was seemingly more closely related to the *Psilanthus* clade. In previous works, incongruous plastid and nuclear relationships were also highlighted by Maurin et al. (2007) and Nowak et al. (2012), but clear conclusions could not be reached mainly due to the low support values obtained for the two molecular trees (nuclear and maternal). The combination of these results has an important impact on the current assumption of the evolutionary history of *C. canephora* as well as among the other wild coffee species.

Using whole plastid genome data for a set of 52 taxa (including *Coffea* s.s. and *Psilanthus*) combined with multiple nuclear SNPs we aimed to address the following research objectives: (1) to reconstruct a maternal tree based on complete plastid genomes, (2) to characterize each

maternal phylogenetic clade through detection of highly mutated genes and gene-mutation rates, (3) independently of the known phylogenetic organization, to identify the internal structure of the set of species studied by following the method of Matar et al. (2019), and 4) to construct a nuclear tree based on 28,800 nuclear SNPs. By observing discrepancies between nuclear and maternal genomes, we hoped to gain a better view of the complex evolutionary history of Coffee including *Coffea canephora* (Robusta coffee).

2. Material and Methods

2.1. Plant material

Fifty-three individuals (42 *Coffea,* 10 *Psilanthus* and *Empogona congesta* as the outgroup) were included in this current study. Their provenance, including their overall geographical location, which is available on the World Rubiaceae Checklist (http://www.kew.org/wcsp/rubiaceae), is shown in Table 1.

2.2. Illumina HiSeq sequencing

Besides the short-read sequences already available (Hamon et al., 2017; Guyeux et al., 2019), or provided by the ACGC consortium and Nestlé, France, additional sequencing was performed on DNAs from fresh or dry leaves of 27 samples (including the outgroup *Empogona congesta*). Seven taxa were sequenced by Eurofins, France, using the Illumina HiSeq2500 platform in paired-end 100-base sequencing according to the manufacturer's instructions. The 20 remaining taxa were sequenced by Nestlé Research, Switzerland, performing Kapa HyperPlus library preparation and sequencing on an Illumina 2500 platform in pared-end (125 bases). Data on the *C. canephora* DH200-94 reference genome (Denoeud et al., 2014) was also used to reconstruct the plastid genome.

2.3. Plastid genome reconstruction and annotation

De novo plastid genome reconstruction was done by mapping raw reads to the *C. arabica* chloroplast using Bowtie2 (Langmead and Salzberg, 2012), with "very-fast-local" and "al-conc" flags. Redundant reads were eliminated using the "Clumpify" function implemented in the bbmap package (https://github.com/BioInfoTools/BBMap), with the "dedupe" option activated. Finally, the obtained reads were assembled using NOVOPlasty (Version 2.7.2) (Dierckxsens, 2017), with a sequence seed corresponding to position 20000 to 20101 to the *C. arabica* chloroplast. Plastid genome annotation was performed with CPGAVAS2 (Shi et al., 2019). Plastid genomes with "Ns" or degenerated bases in the assembly were annotated with GeSeq (Tillich et al., 2017). The new annotated sequences obtained in this study are available at the IRD Dataverse (https://dataverse.ird.fr; Guyot et al., 2020).

2.4. Nuclear SNP calling

Illumina short read sequences were first evaluated for quality using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned against the *Coffea canephora* reference genome (Denoeud et al., 2014) using Bowtie2 (Langmead and Salzberg, 2012). Results were processed by SAMtools (Li et al., 2009) to call nuclear SNPs with Mpileup with a minimum 8x coverage. For every 25 species or accessions, SNPs were filtered out according to the 28,800 reference SNPs established for the *Coffea* genus (Hamon et al., 2017).

2.5. Phylogenetic analyses

In order to reconstruct the maternal phylogeny, the alignment of 52 complete plastid sequences was performed using Clustalw Ver.2.1 (Larkin et al., 2007). Next the alignment was analyzed using RaxML v8.0 (Stamatakis, 2014) with the same parameters used as in Guyeux et al. (2019): the General Time Reversible (GTR) model of nucleotide substitution under the Gamma model of rate heterogeneity, ML optimization initiated from a random

starting tree, rapid hill-climbing algorithm without the heuristic cutoff, but with autoMRE bootstopping criteria for bootstrap convergence. *Empogona congesta* was used as the outgroup.

The 28,800 nuclear SNPs were first concatenated for 49 taxa plus *Empogona congesta* and were subsequently used to construct the molecular tree, using RaxMLV8.0 with the General Time Reversible nucleotide substitution model and gamma distributed rate variation among sites. The nuclear tree included 49 taxa of *Coffea* sl (*Coffea-Psilanthus*). This taxon set is similar to that used for the reconstruction of the maternal tree except for three taxa (*C. canephora*-DH200, the reference genome for SNP mining, *C. canephora*-NC_030053.1 and *C. arabica*-EF_044213.1) for which no nuclear data was available. The SNPs used in this study are available in FASTA format (Supplementary Data 1). Trees were visualized and edited with FigTree Ver. 1.3.1 (Rambaut, 2007) and Inkscape Ver. 0.91 (https://inkscape.org/fr/).

Phylogenetic network analyses were performed with SplitsTree4 Version 4.15.1 (Huson and Bryant, 2006). The Neighbor-net model was run under Kimura 2-parameter (K2P) distances and the Ordinary Least Square Method. The nuclear SNPs and the complete chloroplast sequence data sets were analyzed separately. We used the PHI test (Bruen et al., 2006) to detect any recombination between samples (closely or distantly related), with the average of the delta scores for all quartets being selected from that set of taxa, allowing us to assess the tree likeness of phylogenetic distance data (Holland BR et al., 2002) and Q-residuals as indicators of conflicting signals, as implemented in Splits Tree. Delta scores vary for individual accessions from 0 to 1 according to the number of conflicting signals for each accession.

2.6. Ancestral chloroplast genome reconstruction

Coding sequences predicted by CPGAVAS2 (Shi et al., 2019) and GeSeq (Tillich et al., 2017) were extracted for all genes present in each of the genomes. Following the method described in Guyeux et al. (2019), the characters (nucleotide or "-" for indel) of each column were placed on the corresponding tip nodes of the plastome phylogeny. Next, the ancestral sequences were reconstructed by Phast's Prequel method (Hubisz et al., 2011) which computes the marginal probability distributions for each base in the ancestral nodes of the phylogenetic tree using the tree model and the sum-product algorithm. Reconstructed ancestors were manually inspected if the marginal probability distributions were not high enough. After computing the ancestral state at each internal node of the tree, the gene mutation rates were deduced by comparing the sequence of each genome to the reconstructed one of its last common ancestors. The following groups were used : Group 1 (eight taxa) i.e. *C. rhamnifolia*, *P. leroyi*, *P. travancorensis*, *P. horsfieldianus*, *P. benghalensis var bababudanii*, *P. benghalensis var bengalensis*, *P. brassii*, *P. wightianus*; Group 2 (21 taxa) i.e. *C. liberica*, *C.* sp3, *C. charrieriana*, *C. canephora*-BUD15, *C. kapakata*, *C. mayombensis*, *C.* sp. 'Congo', *C. dewevrei*, *C. heterocalyx*, *C. canephora*-C021, *C. canephora*-C033, *C. brevipes*, *C. congensis*-C409FL, *C. congensis*-CC53, *C. sp. 'nkolbissonii'*, *C. canephora*-FRT81, *C. canephora*-DH200, *C. canephora*_NC_030053.1, *P. mannii*, *P. melanocarpus* and *P. ebracteolatus* and Group 3 (23 taxa) with *C. macrocarpa*, *C. mauritiana*, *C. myrtifolia*-A1.1, *C. myrtifolia*-C414FL, *C. arabica*-EF044213.1, *C. arabica*-ET39, *C. eugenioides*-DA, *C. eugenioides*-BUA, *C. humilis*, *C. stenophylla*, *C. mufindiensis*, *C. sessiliflora*-C406FL, *C. sessiliflora*-PA60, *C. pseudozanguebariae*, *C. racemosa*, *C. salvatrix*, *C. boiviniana*, *C. tetragona*, *C. dolichophylla*, *C. homollei*, *C. humblotiana*, *C. perrieri*, *C. pervilleana*.

2.7. Gene mutation rates

Annotation results were then filtered to discard trna, rrna and fragments, as well as any sequences predicted to have an evaluation greater than 1e-7. In the ad hoc Python script generated for this task, reverse complements of the sequences were made as required, and the

result of this step is a description of each chloroplast in the form of a list of possibly redundant genes, along with the standardized name and orientation as well as its DNA sequence and position. Within the three clades, the number of sequence versions per gene was then counted using a simple script, thus providing information on the diversity per gene of each clade according to its mutation rate.

2.8. Clustering analyses

Clustering analyses of chloroplast genomes were carried out using the spectral method embedded in the SpClust software (Matar et al., 2019), which unlike the phylogenetic study presented above, does not only consider the SNPs, but seeks to take into account more complex relationships between divergent organisms. From the chloroplast alignments, a similarity matrix was constructed using the Needleman-Wunsch algorithm, which allows insertions and deletions to be taken into account in addition to SNPs. From this similarity matrix and the elements of its normalized Laplacian matrix, the chloroplasts were embedded into a space of small dimension, using the spectral dimension reduction technique known as Laplacian eigenmaps (Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, Mikhail Belkin and Partha Niyogi), as described in SpClust. This cloud of points was next clustered using the well-known Gaussian Mixture Model (GMM) technique, allowing clusters of more natural shapes to be recovered than when using more basic methods such as k-means. In addition, the GMM does not need to provide the number of expected clusters beforehand: it was deduced here using the Bayesian Information Criterion (BIC).

Silhouette scoring was considered for clustering performance evaluation, which is calculated using the mean intra-cluster distance and the mean nearest-cluster distance for each sample (Rousseeuw, 1987). It measures how well samples are clustered with samples that are similar to themselves. In other words, each species within clusters obtains a silhouette score from -1 to 1 that shows whether the species is well clustered or not: a high value indicates that the

species is very similar to the other members of its cluster, while a score value of 0 indicates that the species is between two clusters and could be put in one or the other; with a low score value (near -1), the assignment to the cluster is not valid and the species should not be considered as part of the cluster. In short, when all species in a cluster have high score values, the species are considered as valid members of the cluster, while in the opposite case, the clustering is questionable.

3. Results

3.1. Chloroplast genome features

Chloroplast (cp) genome reconstruction and annotation methods were used to analyze plastid genomes features. Throughout the 52 *Coffea* and *Psilanthus* accessions studied, the quadripartite structure was always present. The smallest and largest Cp genomes within *Coffea*, were for *C. canephora*-DH200 (154, 726 bp) and *C. charrieriana* (155,817 bp) respectively. Within *Psilanthus* the range was from 154,545 bp (*P. brassii)* to 155,144 bp (*P. leroyi*). Within *Psilanthus,* the length of LSC, SSC and IR varied from 84,800 (*P. brassii*) to 85,454 bp (*P. leroyi*), from 18,128 (*P. leroyi* and *P. wightianus*) to 18,888 bp (*P. benghalensis*) and from 25,505 (*P. benghalensis var bengalensis*) to 26,066 bp (*P. benghalensis var bababudanii*) respectively. Within *Coffea* the length of LSC, SSC and IR varied from 84,840 bp (*C. canephora-DH200) to* 86,022 bp (*C. charrieriana*), 18,095 bp (*C. salvatrix*) to 18,861 bp (*C. stenophylla*) and from 25,744 bp (*C. liberica*) to 25,945 bp (*C. arabica-ET39*) respectively. Also, on an intraspecific level, differences in the total genome length of the plastid were observed. For instance, in *C. canephora* a variation in length was observed from 154,726 bp (DH200) to 155,519 bp (BUD15, Table A.1.). The overall GC content was similar among all taxa (ca. 37%) but with different values recorded for the specific regions (LSC 35-36%; SSC 31-32%, and IR 43%). Gene content and order were

constant in all species. Similar to Ly et al., 2020, we identified 118 different genes including 80 unique protein-coding genes.

The variation in length for the different cp genomes as well as for the LSC, SSC and IR sub-units in comparison to the smallest cp genome (here *P. brassii*) was plotted in ascending order for the length variation (Fig. A. 1). This highlights the patterns of evolution in these regions. In general, the overall size variation of the cp genome appeared to be influenced more by the LSC than by variations in the SSC or IR. However, a noticeable increase in the length of the SSC was linked to a decrease in the length of IR, and this phenomenon was observed for several taxa (e.g., *P. benghalensis*, *C. eugenioides*-DA, *C. stenophylla,* and *C. canephora*-BUD15).

3.2 Maternal phylogenetic analysis

Phylogenetic analysis of the maternal cp genomes - including 10 *Psilanthus* (marked in blue) and 42 *Coffea* (marked in red) revealed four major clades that were well-supported (100%). These clades will be designated hereafter as MC1 to MC4 (Fig. 1). MC1 is composed of one *Coffea* (*C. rhamnifolia* from northeast coast of Africa) and one *Psilanthus* (*P. leroyi* from central northeast Africa). MC2 contains all the Asian (broadly speaking) *Psilanthus,* while MC3 includes both western and central African *Psilanthus* species as well as the majority of *Coffea* taxa from West and Central Africa, including *C. canephora.* MC4 comprises the remaining *Coffea* taxa: all Western Indian Ocean Islands species, East African *Coffea* species, two West African *Coffea* (*C. stenophylla* and *C. humilis*) and Central-East African species. *C. arabica* belongs to this clade and is closely related to *C. eugenioides,* its maternal parent. Interestingly, the six *C. canephora* accessions were clustered into tree sub-clades within the MC3 clade along with other species (*C. mayombensis* / *C. spCongo*; *C. brevipes* and C. *sp nkolbisonii / C. congensis*) (Fig. 1).

3.3 Ancestral chloroplast genome reconstruction

Ancestral plastome reconstruction was performed in order to identify the highly mutated genes per species in each clade. As it is not possible to consider a clade with only two individuals, MC1 and MC2 were merged into one group, thereby constituting three sets of taxa for further analyses (see Methods).

Table A.2 summarizes the percentage of species in which one or more genes differ from their ancestral copies. The results showed that the highly mutated genes in all the species of a given group differed from the other groups by the percentage of species involved and by the genes themselves. Therefore, some highly mutated genes were specific to one group, three genes were shared by all species (*ycf1*, *ycf2* and *mat*K), and a few by two groups. For instance, *rps3* and *ndhF* were found in all species of Group 3 while *psbD*, *psaA* and *accD* were specific to Group 2. Mutated genes *ycf3*, *rps4*, *rps2*, *rpl22*, *rpl16*, *rbcL*, *petA* and *atpF* were specific to Group 1.

Based on a total of 99 chloroplast genes studied, trees were obtained for each group (Fig. A.2), allowing the visualization of the total number of genes that differed from their ancestral form (given after the species name). Approximately one third to one half of all the genes differed from their ancestral form in each group, ranging from 39 to 55 in Group 1, 32 to 49 in Group 2 and 22 to 43 in Group 3.

Gene mutation rates per species within each group were estimated for each gene as equal to the Levenshtein distance between the translated gene sequences and its ancestor, divided by the length of their alignment (Fig. A.3). To facilitate the visual comparison, the cp genomes with genes ordered according to their genomic position have been divided into two parts (A and B), corresponding to the chloroplastic genome cut into two equal parts. Different patterns have been obtained with some genes highly mutated and others highly conserved.

3.4 Vectorial clustering analyses

In order to remain independent from the clustering derived from the molecular tree, and since the molecular phylogeny results solely from nucleotide substitutions, a separate clustering method was performed, taking into account all the genome modifications (see Methods).

Two clusters were observed. One cluster, C1, includes five of the six Asian *Psilanthus* (*P. benghalensis var bababudanii*, *P. benghalensis var bengalensis*, *P. brassii*, *P. horsfieldianus* and *P. wightianus*) and was characterized with high positive individual score values (varying from 0.413 to 0.596, except for *P. wightianus* with 0.075, the mean of the cluster being equal to 0.441). The second cluster, C2, was composed of the remaining 47 taxa. This second cluster, with a mean score value of 0.503, consisted of taxa that had positive score values ranging between 0.138 for *P. melanocarpus* and 0.616 for *C. homollei_SZ*. Only *C. charrieriana* (-.222) and *P. travancorensis* (-0.273) were characterized by negative values. Given the heterogeneity of the individual scores, the two species with negative values indicated poor clustering and were excluded from the analysis, and the sub-clustering of C2 was recomputed.

The new set of clustering analyses resulted in the formation of three sub-clusters: SC2-1, SC2-2 and SC2-3 as summarized in Fig. A.4. Sub-Cluster 2-1 (17 taxa; MC3) was composed of all *Coffea* species from West and Central Africa including *C. canephora,* except for *C. humilis* and *C. stenophylla*. In SC2-1, the mean score value was 0.583 with all accessions having highly positive score values (from 0.317 for *C. sp*3 to 0.702 for *C. sp. 'nkolbissonii'*) except for *C. liberica* (0.051).

Sub-cluster 2-2 (23 taxa; MC4) included two species from West Africa (*C. humilis* and *C. stenophylla*), and all taxa from the Western Indian Ocean Islands, East Africa and eastern Central Africa. The mean score value for SC2-2 was equal to 0.395, varying from 0.228 for *C. arabica*-EF044213.1 to 0.507 for *C. pervilleana*. Sub-cluster 2-3 (five taxa; MC1 and

MC3), which is characterized by a mean score of 0.242, ranging from 0.085 (*P. leroyi*) to 0.365 (*P. melanocarpus)*, includes the African *Psilanthus* species and *C. rhamnifolia*.

3.5 Nuclear phylogenetic relationships

The nuclear tree (Fig. 2) showed two well-supported clades. One clade included all *Psilanthus* taxa plus *Coffea rhamnifolia* whereas the other consisted only of *Coffea* taxa. Within the *Coffea* clade, the different accessions for a given species were generally very close except for *C. canephora* in which two species (*C. sp.* 'nkolbissonii' and *C. brevipes*) were included. *Coffea arabica* is positioned closely to *C. canephora,* one of its diploid progenitors (Lashermes et al., 1999). The geographic regionalization of the structure (i.e. species from one main geographic region are more closely related than species from different regions, see Fig. 2) is confirmed with the newly studied taxa placed accordingly. For instance, *C. sp*3, *C. mayombensis* and *C. sp.* 'nkolbissonii' from Central Africa fell into the West and Central Africa (WCA) clade as previously defined by genotyping with the sequencing approach and by including for example *C. canephora, C. congensis, C. liberica, C. humilis* and *C. stenophylla* (Hamon et al., 2017).

The phylogenetic networks that are generated for both plastid and nuclear data show complex relationships which occur especially in the main basal branches (Fig. 3). In order to detect conflicting signals in the data sets and to identify which taxa are involved, the average delta scores were calculated. For plastid and nuclear data, we obtained a result of 0.141 with a Q-residual score of 0.001 and 0.163 with a Q-residual score of 0.005, respectively. Regarding the nuclear data, values obtained on WCA *Coffea* and on WCA and Asian *Psilanthus* are in the same order of magnitude (Delta score = 0.189 and 0.147; Q-residual score = 0.0110 and 0.0128 respectively). These scores are very low for example for Mauritian *Coffea* (Delta score = 0.001, Q-residual score = 4.117E-6) while the highest values are present for Madagascan

and Comorian *Coffea* (0.537 and Q-residual score = 0.015). As for the nuclear and plastid data, statistically significant evidence for recombination (PHI test) is found (respectively: mean = 0.141, variance = 4.376E-7, observed = 0.136, P-value = 2.50E-10 and mean = 0.094, variance = 1.20E-5, observed = 0.062, P-value = 0.0). In the nuclear tree, as expected, due to its hybrid genetic origin, *C. arabica* appears in an intermediary position between its two progenitors *C. eugenioides* and the group of *C. canephora* accessions. The close relationships between *P. mannii* and *P. melanocarpus* and among *C. dewevrei, C. sp*3 and *C. liberica* are obvious. For the latter, this reflects the fact that *C. dewevrei* is often perceived as a variety of *C. liberica.*

4. Discussion

In this study, a significant increase in complete plastid genome sequences from *Coffea* and *Psilanthus* was obtained, complementing the Ixoroideae plastomes generated by Ly et al. 2020. The current datasets collectively will be useful for further studies in the Rubiaceae family and more generally for flowering plants. The plastid genomes obtained here as well as the approach used will offer the possibility i) to obtain an increasing number of robust molecular trees at the genus level, and ii) to detect any putative phylogenetic incoherence that may help to understand the complex evolutionary history of these genera.

The maternal phylogeny clearly identified closely-related chloroplasts between *Psilanthus* and WCA *Coffea* species. However, there are clear morphological differences in reproductive organs (short-styled and inserted stamens for *Psilanthus* versus long-styled and partially or fully exerted stamens for *Coffea,* (Leroy, 1980)), growth architecture (both sympodial and monopodial growth in *Psilanthus* but only monopodial growth for *Coffea),* biochemical compounds contents (very low caffeine or caffeine-free in *Psilanthus* and from caffeine-free to caffeine-rich in *Coffea* (Hamon et al., 2015 for review)), and on corolla and leaf anatomy (The corolla and leaves in *Psilanthus* are significantly thinner than in *Coffea* species,

Stoffelen et al., 2008.). In accordance with morphological and physiological synapomorphies for *Coffea* and *Psilanthus*, the nuclear phylogenetic tree clearly separates both genera (except for *C. rhamnifolia* which is closer to *Psilanthus* than to *Coffea*). In contrast, the maternal tree was inconsistent with the morphological evidence and nuclear data in four of the main clades observed. One clade which was sampled in West and Central Africa grouped both *Coffea* and *Psilanthus* taxa. We will further discuss the relationships among coffee species using both phylogenetic and clustering methods as well as the inconsistencies between maternal and nuclear phylogenetic trees. In addition, we will discuss the maternal phylogenetic relationships between *C. canephora* accessions and their clustering with different West African species.

4.1. Species relationships

The maternal tree adds new information to the interspecific relationships within *Coffea s.l.* that were inferred in previous studies. Maurin et al. (2007) contained a large sampling but used only a few plastid loci, whereas Guyeux et al. (2019) used complete plastid sequences but only included a limited number of samples. The present results delineated four well-supported clades two of which included both *Psilanthus* and *Coffea* species (MC1 and MC3), and two consisting solely of *Psilanthus* (MC2) or *Coffea* species (MC4).

The clades observed in the plastid phylogeny are supported i) by several distinct highly-mutated genes between clades at the species level, and also at the ancestral plastid genome-reconstruction level, and ii) by their phylogeny-independent vectorial clustering.

The nuclear phylogeny, which included data from 25 new accessions, confirms the topology reported by Hamon et al. (2017) and Guyeux et al. (2019). *Coffea* and *Psilanthus* formed two well-delineated clades except for *C. rhamnifolia* (which is part of the clade *Psilanthus*). This study also phylogenetically posits four additional African taxa (*C. sessiliflora*, *C. sp*3, *C. mayombensis* and *C. sp. 'nkolbissonii'*) previously not included in Hamon et al. (2017) nor in

Guyeux et al. (2019). All these taxa are situated in logical accordance with their geographic origins, East Africa for the Tanzanian *C. sessiliflora* and West and Central Africa for the three Cameroonian species.

4.2 Evolutionary history of Coffee species

Comparing the maternal and the nuclear trees pinpointed some interesting events, all concerning the *Coffea* species. First, nuclear and plastid data demonstrate a congruous position for East African species (EA nuclear clade and maternal MC4 clade) and for the islands' species (Western Indian Ocean Islands nuclear clade and maternal MC4 clade). Interestingly, both plastid and nuclear phylogenetic networks indicate a radiative speciation for the Madagascan species (Fig. 3). However, broader samplings need to be analyzed since this current study included only six of the total 68 Madagascan species that have been described to date.

In our current study, the northeast African *C. rhamnifolia* appears to be more closely related to *Psilanthus* and especially to the central northeast African *P. leroyi* than to the genus *Coffea* in the two molecular trees. The similar placement of *C. rhamnifolia* in both molecular trees reinforces its enigmatic taxonomic status. Indeed, the position of *C. rhamnifolia* had been uncertain for a long time and this was reflected by its placement in different genera (*Paolia jasminoides* (Chiov., 1916), *Canthium rhamnifolium* (Chiov., 1932), and *Plectronia rhamnifolia* (Bridson, 1916). Later its position in the *Coffea* genus (as *Coffea paolia* (Bridson, 1979), *Coffea rhamnifolia* (Bridson, 1983)) was confirmed by Leroy (1996), Stoffelen (1998) and Davis et al. (2006, 2011), mainly based on its flower morphology which is closer to *Coffea* than to *Psilanthus*. Based on morphological characteristics, Leroy (1996) positioned *C. rhamnifolia* in the subgenus Baracoffea, inter alia as it is a species of *Coffea* with terminal inflorescences on short monopodial shoots.

Coffea rhamnifolia, is particularly well adapted to very dry conditions (400 mm annual rainfall with a long dry season), has solitary terminal flowers often produced before the leaves, (Bridson, 1982), and an absence of caffeine in seeds (Hamon et al., 2015 for review). Together, these traits are found in *Psilanthus* particularly. Therefore, it seems that placing *Coffea rhamnifolia* in *Coffea* rather than in *Psilanthus* was only justified by the flower morphology and in particular the relative length and position of the style and stamens. The habitus, leaves and ecology of *C. rhamnifolia* and *Psilanthus leroyi* (Bridson, 1982) are very similar and indicate a close relationship between both species. Interestingly, *C. rhamnifolia* has a somewhat intermediary position between the genera *Coffea* and *Psilanthus*. Given the species phylogenetic relationships from a taxonomic point of view, if placing it into *Coffea* lacks strong evidence, we should either admit that the species should be renamed *Psilanthus rhamnifolia* or that both *Psilanthus leroyi* and *Coffea rhamnifolia* should be designated to a third genus of their own: *Paolia,* which is positioned between *Coffea* and *Psilanthus*. Another conclusion could be that *Coffea* is paraphyletic with *C. rhamnifolia* differentiated from *Psilanthus*. The current results show that there is a clear necessity for further studies on *Psilanthus leroyi* and on *Coffea rhamnifolia* in order to learn more about their vegetative morphology, leaf anatomy, phenology and genome organization and functioning. Unfortunately, both *Psilanthus leroyi* and *Coffea rhamnifolia* are absent from living collections and new collections of the species will be difficult to procure as their natural distribution coincides currently with war-torn regions.

The comparison of nuclear and plastid phylogenies has also brought forward a noticeable amount of specific discrepancies. *Coffea charrieriana* appears as enigmatic as *C. rhamnifolia* with an intermediary placement between the two nuclear clades *Psilanthus and Coffea,* while it is found nested within plastid clade MC3 where it is closely related to the West and Central African *Coffea* species. Initially classified as *Psilanthus* based on its vegetative appearance,

its flower morphology has been observed as typical for *Coffea,* which contributed to its transfer to *Coffea*. However, in contrast to the other *Coffea* species, *C. charrieriana* is characterized by very thin leaves, has a parenchymatic seed coat that lacks sclereids (Stoffelen et al., 2008), and has a very low sucrose content (the lowest in *Coffea;* 3.8% dmb; Campa et al., 2004). In addition, it is the only caffeine-free *Coffea* from West and Central Africa (Campa et al., 2005). *Coffea charrieriana* produces black fruits – a rare color in West and Central Africa. All these characteristics tend to link the species to the genus *Psilanthus*. In this study, the vectorial clustering failed to place *C. charrieriana* in a cluster. All these results (maternal and nuclear molecular trees, vectorial clustering*)* raise questions about its genetic origin, its pure or homoploid species. For Nowak et al. (2012), the inconclusive placement of *C. charrieriana* could be the result of ancient hybridization with a *Psilanthus* chloroplast capture. With this assumption, the female progenitor of *C. charrieriana* would be a *Psilanthus* species. This hypothesis might explain the difficulty of achieving favorable plant development in *ex-situ* collections. Indeed, samples collected on Mount Bakossi (Cameroon) in 1983 (Anthony et al., 1985) and placed in a forested environment in Ivory Coast could not be successfully grown. Duplication in the field at the Centre de Ressources Biologiques, Bassin-Martin, Reunion, show very slow treelet development.

Another divergence has been observed between the two West African species *C. stenophylla* and *C. humilis.* Indeed, their unexpected placement outside of the West and Central African *Coffea* lineage of the maternal tree is unexpected. Why are these two species so closely related to the *Coffea* species from both East Africa and the Western Indian Ocean islands? Surprisingly, these two species, closely related from a phylogenetic point of view, correspond to the smallest genome of West and Central Africa (*C. stenophylla*, 631 and 670 Mb, Noirot et al., 2003, Razafinarivo et al., 2012) and to the largest diploid *Coffea* genome (*C. humilis*, 861 and 890 Mb, Noirot et al., 2003, Razafinarivo et al., 2012). The two species are also very

different in morphology and habitat. *Coffea stenophylla* is a medium-sized tree with high productivity while *C. humilis* is rather a dwarf tree with a limited number of branches producing small amounts of flowers and seeds. Ripe fruits are black for *C. stenophylla* (a rare feature for *Coffea* species in West and Central Africa) and bright red for *C. humilis*. *Coffea stenophylla* is well-adapted to relatively dry forest conditions and occurs naturally on the hilltops (not on hill slopes or valleys), while *C. humilis* is well-adapted to very humid conditions which are typical beneath the canopy of evergreen rainforests where it is found along small troughs carved by stream erosion (Berthaud and Charrier, 1988). Curiously, despite these differences in morphological features, growing conditions, and genome size, both species produce artificial hybrids having pollen fertility of 27-38% (Louarn, 1993). Their presence in the MC4 clade (Fig. 1) seems in sharp contrast to other species in this group where the majority of *Coffea* species come from West and Central Africa with a significant degree of relatedness. Given that none of the current West African species share the same plastid feature, one possible explanation involves a putative plastid capture from a former extinct species which remains to be identified. However, if the past geographic distribution of the two species was greater before than at the present and extended into an area inhabited by species with the *Coffea* plastid genome, the plastid capture may have been possible. Another possibility might involve retaining an ancestral polymorphism (incomplete lineage sorting).

In our case study, we could postulate that incomplete lineage sorting possibly occurred throughout the evolution of the genus *Coffea*. Ancestral polymorphism retention could have been maintained in the ancestor of the West and Central African *Coffea* species. Afterwards, unequal random lineage sorting could have led to the current plastid genome distribution that exists in the nuclear West and Central African *Coffea* species (Fig. 4). This hypothesis implies that the retention of ancestral polymorphism would be present in the most recent common

ancestor (MRCA) of both *Psilantus* and *Coffea* lineages as in the MRCA of all the *Coffea* lineage.

Another obvious discrepancy observed from the trees is related to the position of *Coffea heterocalyx.* The species is closely related to *C. eugenioides* in the nuclear tree, and part of the East-Central-East African (E-CEA) clade of Maurin et al. (2007), Hamon et al. (2017). *C. heterocalyx* is part of the maternal MC3 clade (grouping both *Psilanthus* and *Coffea* species from the west and central region of Africa, Fig. 1). In the nuclear tree of Maurin et al. (2007), Hamon et al. (2017), *C. heterocalyx* is phylogenetically closely related to the Cameroonian *C. anthonyi*, *C. eugenioides* from Eastern Congo, Rwanda, Uganda, Tanzania and Kenya, and the Tanzanian *C. lulandoensis*, *C. kihansiensis* and *C. mufindiensis*. Based on the partial results of Maurin et al. (2007), *C. heterocalyx* seems to be the sole member of the species of the maternal MC3 clade and the SC2-1 sub-cluster of this study that occurs in the nuclear E-CEA clade. As discussed for *C. stenophylla* and *C. humilis,* the retention of ancestral polymorphisms could explain our results for *C. heterocalyx* (Fig. 4). Seen through this hypothesis, incomplete lineage sorting would be maintained in the MRCA of the E-CEA clade. Random lineage sorting would result in different plastid genomes, with *C. heterocalyx* on the one hand and with the other members of the nuclear E-CEA clade on the other (Fig. 4). Although a hybrid origin (*C. eugenioides* x *C. liberica* resulting in *C. heterocalyx*) was postulated by Maurin et al. (2007), cytogenetic observations do not support this theory. Using double fluorescence in-situ hybridization (FISH) of the 5S and 18S rDNA genes, coupled with fluorochrome banding applied to *Coffea,* differences in the number of satellite chromosomes were revealed (Hamon et al., 2009). *Coffea heterocalyx* is characterized by one SAT-chromosome, one 5S and one 18S locus; *C. liberica* has one SAT-chromosome, two 5S loci and one 18S locus, while *C. eugenioides* has two SAT-chromosomes, one 5S locus and

two 18S loci. Moreover, the acquisition of self-fertility of the nascent diploid species (*C. heterocalyx*) could not be explained.

An alternative explanation could involve a hybridization between the self-fertile *C. anthonyi* (Stoffelen et al., 2009) as the male genitor, and a species from the WCA region, member of the maternal MC3 clade as the female genitor. The male parent would be the donor of the self-fertility and the female parent the donor of the plastid genome. In the wild, interspecific hybridization of the *Coffea* species is not uncommon, yet such hybrids may not necessarily be viable and fertile. Putative hybrids within the genus *Coffea* have been reported since the end of the 19th century (Chevalier, 1942). For example, *Coffea affinis* is often considered as a natural intraspecific hybrid between *C. stenophylla* and *C. liberica* in their region of origin (Ivory Coast). Similarly, *C. bakossi* (Cheek et al., 2002) could also be a natural hybrid between *C. liberica* and *C. montekupensis* (Stoffelen et al., 1997). Both species grow in a more mountainous environment where they have a partially overlapping altitudinal range. However, even if all the hybrids reported were obtained naturally, in most cases, interspecific hybridization did occur after wild species from outside their area of origin were introduced, as with Java (*C. liberica* X *C. arabica*), in Malaysia, in Indochina, and in Trinidad (*C. liberica* X *C. stenophylla*). Gomez et al. (2010) also reported hybrids that occurred naturally after the introduction of "pure" species in New-Caledonia. Indeed, introducing new species into new habitats may induce changes in flowering, such as blossoming overlaps with the possibility of cross-pollination. In the distribution area of the parental species, detecting natural hybrids can be difficult. Survival of the F1 offspring does not guarantee success for such hybrids which need to be fertile as well as being isolated enough (reproductively from its progenitors to avoid parental gene dilution (Chase et al., 2010). Assuming that hybridization is possible, the acquisition of self-fertility could spur the process of rapid, successful speciation. Without a doubt, this new trait would greatly facilitate speciation with limited pollen fertilization from

parental species (reproductive isolation) and would ensure a rapid fixation of new allelic combinations, producing new morphological, physiological and adaptive traits. Hybridization might also lead to quicker genomic changes, such as chromosomal rearrangements and genome expansion, producing beneficial new phenotypes (Baack and Rieseberg, 2007). If its hybrid origin is confirmed, the hybridization event resulting in *C. heterocalyx* did not produced a detectable burst of transposable elements (Guyot et al., 2016).

Tree comparisons also highlighted certain inconsistencies regarding the placement of most of the Western and Central African *Coffea* species, including the widely cultivated coffee species *C. canephora.* All these coffee species belong – together with the *Psilanthus* species of the same geographic region – to the maternal MC3 clade. Putative incomplete lineage sorting for *C. stenophylla*, *C. humilis* and *C. heterocalyx*, followed by random lineage sorting could feasibly explain the position of the other WCA *Coffea* species in the maternal tree (Fig. 4). Our sampling demonstrates that this position of *C. canephora* close to the West and Central African *Psilanthus* clade, as reported by Guyeux et al. (2019), is not an effect of the accessions used by these authors. Indeed, the six accessions of the study sampled in different countries (see Table 1) are close to the WCA *Psilanthus* lineage which are both members of the maternal MC3 clade of the present study.

As expected, the recent allotetraploid *C. arabica* (Yu et al., 2011) belongs to the maternal MC4 clade as well as its female parent *C. eugenioides.* However, the hybrid species is part of the nuclear WCA clade close to its other progenitor, *C. canephora*.

Finally, it seems that several mechanisms such as incomplete lineage sorting (although this is not possible to clearly demonstrate in the present study) and hybridization which leads to homoploidy (without chromosome doubling) and alloploidy (for *C. arabica* tetraploid) have all shaped the evolutionary history of coffees. It is clear however that the evolutionary history of the genus *Coffea* is more complex than it was initially perceived. These mechanisms – by

cause or consequence – could be associated to the capability of these genomes to evolve without dramatic changes (as detected with transposable elements, Guyot et al. 2016), but with high possibilities of adaption to an immense range of habitats, (Davis et al., 2006; Charrier, 1978; http://publish.plantnet-project.org/project/wildcofdb; Guyot et al., 2020) during new niche colonization events or while undergoing climate changes.

4.3 Coffea canephora: a complex species or a complex of species?

In this study, even with a limited number of *C. canephora* accessions available, the genetic differentiation observed agrees with results based on SSRs (Gomez et al., 2009, Musoli et al., 2009), on retrotransposon diversity (Hamon et al., 2011), and on SNPs (Mérot-L'Anthoëne et al., 2019). Indeed the two samples (C021 and C033) from West Africa belonging to the Guinean Genetic Group D (Gomez et al. 2009; Mérot-L'Anthoëne et al., 2019) are sisters and are well differentiated from the Ugandan BUD15 (Genetic Group O, of Mérot-L'Anthoëne et al., 2019) and from the accession FRT81 derived from *C. canephora* material originally collected in the Kouilou region of the Republic of the Congo.

The Ugandan populations of *C. canephora* are quite distinct from the other *C. canephora* populations, as the ones from West Africa and some populations from the Central African Republic called "caféiers de la Nana (Coffee trees of the Nana)." For Berthaud and Guillaumet (1978), the "Nana" population can grow in temporarily flooded areas similar to *C. congensis* but also on dry soils, contrary to other *C. canephora* populations.

Based on overall morphological differences as well as on the inferred phylogenetic relationships within *C. canephora* and within the nuclear West and Central African *Coffea* clade, it may be considered that some populations, well differentiated and genetically isolated, might evolve towards new species with several intermediary steps of (sub-)speciation.

The nested placement of *C.* sp. 'nkolbissonii' and *C. brevipes* questions the status of *C. canephora* and its link with these two species. Indeed *C. brevipes* is a small species, which is often considered as a dwarf form of *C. canephora*. Moreover, *C. brevipes* is also characterized by other morphological differences such as plant architecture, number of caliculi, number of flowers per inflorescence, length of the peduncle, shape of the leaves, etc. In fact, quite a few different and independent characters are distinctive; morphologically, both of the species *C. brevipes* and *C. canephora* can be easily distinguished from each other. The general distribution of *C. brevipes* is similar to that of *C. canephora* (except that *C. brevipes* does not occur in the Guinean domain (West of Nigeria) while *C. canephora* does occur there). Recently, *C. brevipes* was found at a higher altitude (up to 1,450 m, in Rwanda and up to 1,570 m, in Eastern Congo), an altitude similar to *C. canephora* in Uganda. However, the species have never been found sympatric. A more detailed look at their distribution and population density suggests that *C. brevipes* is better adapted to the wet and dense rainforest while *C. canephora* is a species from the less dense and less humid rainforest. It is observed that the population of *C. canephora* in dense and wet rainforest is lower and less reproductive. These observations could be an indication that *C. canephora* is surviving in dense rainforest probably as a relict species from drier periods. Further studies are needed however to shed more light on this point. The combined morphological traits and habitat preferences contribute to maintaining *C. brevipes* as a distinct species.

The case of *C. sp*. 'nkolbissonii' is not as clear, even if its habitus and morphology are quite distinct from *C. canephora. Coffea sp*. 'nkolbissonii' is a small shrub of 1 to 1,5 m tall, with thin twiglets, narrower (and smaller) leaves, light grey to white bark, and petioles that turn blackish when dried. It has quite a restricted distribution area and is only found on the hills around Yaoundé in central Cameroon. It could therefore be considered as a distinct species or as a subspecies of *C. canephora* within central Cameroon.

In both the nuclear and the chloroplast trees, *Coffea congensis* has a basal position within the *C. canephora* clade. Indeed, it is closely related to *C. canephora, C. brevipes* and *C.* sp. 'nkolbisonii', but it is always considered as a distinct species. *Coffea congensis* is adapted to a very particular habitat, namely periodically inundated islands and riverbanks of the Congo stream and some of its larger tributaries. Morphologically, it is easily distinguishable from *C. canephora* and the other species of the clade by its smaller and narrower leaves. Phenotypically, *C. congensis* resembles *C. arabica* and *C. kivuensis* more than the other species of the *canephora* clade.

Finally, while *C. congensis*, *C. brevipes* and *C. sp.* 'nkolbissonii' are considered as species with limited gene flow in the wild, *C. canephora* might be considered as a complex of species or at least one species which is currently in a state of ongoing speciation. This point of view could explain the genetic and phenotypic differentiation observed between the Western and the Central African *C. canephora* populations (Berthaud, 1986), and within the Central African populations (Mérot-L'Anthoëne et al., 2019).

In conclusion, the present study shows a new evolutionary history for coffee species revealed by comparing two molecular trees based on complete plastid genome sequences and on 28,800 nuclear SNPs, respectively. The enigmatic status of *Coffea rhamnifolia*, closely related to *Psilanthus Leroyi,* is highlighted here. The topological comparison between the maternal and nuclear trees has pointed out an inconsistent placement of most of the West and Central African *Coffea* species. These are more closely related to the Western and Central African *Psilanthus* than to the remaining *Coffea,* which could be explained both by retention of ancestral polymorphism and by hybridization without chromosomal doubling. The particular case of *C. canephora* (Robusta coffee) illustrates the complex evolutionary history of the coffee species. *Coffea canephora* could be either a complex species or in an ongoing process of sub-speciation.

Acknowledgments

We first thank Meise Botanic Garden for funding part of the sequencing. We thank the Institut de Recherche pour le Développement (IRD) for funding sequencing work and for providing a fellowship grant for S.N. Ly. Finally, we thank the Arabica Coffee Genome Consortium (ACGC; Mueller et al., 2015) for kindly providing the sequences used in this study.

Appendix and supplementary Information

Tables A.1 and A.2

Figs. A.1, A.2, A.3 and A.4

Supplementary Data 1

Chloroplast sequences are available at following the link: https://doi.org/10.23708/KWRIJJ.

Author Information

The authors declare no competing financial interests.

References

Acosta, M.C., Premoli, A.C., 2010. Evidence of chloroplast capture in South American *Nothofagus* (subgenus Nothofagus, Nothofagaceae). Molecular Phylogenetics and Evolution 54, 235-242.

Anthony, F., Couturon, E., de NAMUR, C., 1985. Les caféiers sauvages du Cameroun. Résultats d'une mission de prospection effectuée par l'ORSTOM en 1983. XIth A.S.I.C. colloquium, Lomé, Togo, 495-505.

Arnold, M.L., 1993. *Iris nelsonii* (Iridaceae): Origin and genetic composition of a homoploid hybrid species. American Journal of Botany 80, 577-583.

Baack, E.J., Rieseberg, L.H., 2007. A genomic view of introgression and hybrid speciation. Current Opinion in Genetics and Development 17-6, 513-518.

Berthaud, J., 1986. Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. ORSTOM Edition, collection Travaux et Documents n°188. 379p.

Berthaud, J., Charrier, A., 1988. Genetic Resources of *Coffea*. *In* Coffee, 4, 1-42. Eds R.J. Clarke and R. Macrae, Elsevier Applied Science London and New York.

Berthaud, J., Guillaumet, J.-L., 1978. Les caféiers sauvages en Centrafrique. Résultats d'une mission de prospection (janvier-février 1975). Café Cacao Thé 22-3.

Bouharmont, J., 1963. Somatic chromosomes of some *Coffea* species. Euphytica 12, 254-327

Brennan, A.C., Barker D., Hiscock, S.J., Abbott, R. J. 2012. Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: A study of *Senecio squalidus* (Asteraceae). Heredity 108, 87-95.

Bridson, D.M., 1979. The identity of *Paolia* (Rubiaceae). Kew Bulletin 34-2, 376.

Bridson, D.M., 1983. The identities of *Plectonia rhamnifolia* Chiov. and *P. sennii* Chiov. (Rubiaceae). Kew Bulletin 38-2, 320.

Bridson D. M., 1982. Studies in *Coffea* and *Psilanthus* (Rubiaceae subfam. Cinchonoideae) for Part 2 of Flora of Tropical East Africa: Rubiaceae. Kew Bulletin 36-4, 817-859.

Bruen, T.C., Philippe, H., Bryant, D., 2006. A Simple and Robust Statistical Test for Detecting the Presence of Recombination. Genetics 172, 2665-2681.

Campa, C., Ballester, J.F., Doulbeau, S., Dussert, S., Hamon, S., Noirot, M., Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., Dopazo, J., 2004.Trigonelline and sucrose diversity in wild *Coffea* species. Food Chemistry 88, 39-43.

Campa, C., Doulbeau, S., Dussert, S., Hamon, S., Noirot, M., 2005. Diversity in bean caffeine content among wild *Coffea* species: evidence of a discontinuous distribution. Food Chemistry 9, 633-7.

Charrier, A., 1978. La structure génétique des caféiers spontanés de la région malgache (Mascarocoffea): leur relation avec les caféiers d'origine africaine (Eucoffea). PhD Thesis Edn ORSTOM, Paris, 224 p.

Chase, M.W., Knapp. S., Cox, A.V., Clarkson, J.J., Butsko, Y., Joseph, J., Savolainen, V., Parokonny, A.S., 2003. Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Annals of Botany 92, 107-127.

Chase, M.W., Ovidiu, P., Fay, M.F., 2010. Hybridization and speciation in angiosperms: a role for pollinator shifts? BMC Biology 8, 45.

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

Chevalier, A., 1942. Les caféiers du globe. Caféiers sauvages et cultivés. Encyclopédie biologique XXII Fas. II, 1-33.

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

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

Davis, A.P., Tosh, J., Ruch, N., Fay, M.F., 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.

Davis, A.P., Chadburn, H., Moat, J., O'Sullivan, R., Hargreaves, S., Nic Lughadha, E., 2019. High extinction risk for wild coffee species and implications for coffee sector sustainability. Sci Adv. 2019 Jan 16; 5(1):eaav3473

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., Aury, J.-M., Bento, P., Bernard, M., Bocs, S., Campa, C., Cenci, A., Combes, M.-C., Crouzillat, D., Da Silva, C., Daddiego, L., De Bellis, F., Dussert, S., Garsmeur, O., Gayraud, T., Guignon, V., Jahn, K., Jamilloux, V., Joët, T., Labadie, K., Lan, T., Leclercq, J., Lepelley, M., Leroy, T., Li, L.-T., Librado, P., Lopez, L., Muñoz, A., Noel, B., Pallavicini, A., Perrotta, G., Poncet, V., Pot, D., Priyono, Rigoreau, M., Rouard, M., Rozas, J., Tranchant-Dubreuil, C., VanBuren, R., Zhang, Q., Andrade, A.C., Argout, X., Bertrand, B., de Kochko, A., Graziosi, G., Henry, R.J., Jayarama, Ming, R., Nagai, C., Rounsley, S., Sankoff, D., Giuliano, G., Albert, V.A., Wincker, P. and Lashermes, P., 2014. The Coffee Genome Provides Insight into the Convergent Evolution of Caffeine Biosynthesis. Science 345, 1181-1184.

D'Hont, A., Denoeud, F., Aury, J., Baurens, F.-C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., Da Silva, C., Jabbari, K., Céline Cardi, C., Julie Poulain, J., Souquet, M., Labadie, K., Jourda, C., Lengellé, J., Rodier-Goud, M., Alberti, A., Bernard, M., Correa, M., Saravanaraj Ayyampalayam, S., Mckain, M.R., Leebens-Mack, J., Burgess, D., Freeling, M., Mbéguié-A-Mbéguié, D., Chabannes, M., Wicker, T., Panaud, O., Barbosa, J., Hribova, E., Heslop-Harrison, P., Habas, R., Rivallan, R., Francois, P., Poiron, C., Kilian, A., Burthia, D., Jenny, C., Bakry, F., Brown, S., Guignon, V., Kema, G., Dita, M., Waalwijk, C., Joseph, S., Anne Dievart, A., Jaillon, O., Leclercq, J., Argout, X., Lyons, E., Almeida, A., Jeridi, M., Dolezel, J., Roux, N., A.-M., Weissenbach, J., Ruiz, M., Glaszmann, J.-C., Quétier, F., Yahiaoui, N., Wincker, P., 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488, 213–217.

Dierckxsens, N., Mardulyn, P., Smits, G., 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45, e18. doi:10.1093/nar/gkw955.

Duan, R., Huang, M., Yang, L., Liu, Z., 2017. Characterization of the complete chloroplast genome of *Emmenopterys henryi* (Gentianales: Rubiaceae), an endangered relict tree species endemic to China. Conservation Genetics Resources 9, 459-461.

Ferreira, T., Shuler, J., Guimaraes, R., Farah, A., 2019. Introduction to Coffee Plant and Genetics. In Coffee: production, Quality and Chemistry. Ed. Adriana Farah.

Gao, J., Wang, B., Mao, J.-F., Ingvarsson P., Zeng, Q.-Y., Wang, X.-R., 2012. Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. Molecular Ecology 21, 4811-4827.

Gomez, C., Dussert, S., Hamon, P., Hamon, S., de Kochko, A., Poncet, V., 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.

Gomez, C., Batti, A., Le Pierrès, D., Campa, C., Hamon, S., de Kochko, A., Hamon, P., Huynh, F., Despinoy, M., Poncet, V., 2010. Favourable habitats for *Coffea* inter-specific hybridization in central New Caledonia: combined genetic and spatial analyses. Journal of Applied Ecology 47, 85-95.

Guyeux, C., Charr, J.-C., Tran, H.T.M., Furtado, A., Henry, R.J., Crouzillat, D., et al. 2019. Evaluation of chloroplast genome annotation tools and application to analysis of the evolution of coffee species. PLoS ONE 14-6, e0216347.

Guyot, R., Darré, T., Dupeyron, M., de Kochko, A., Hamon, S., Couturon, E., Crouzillat, D., Rigoreau, M., Rakotomalala, J.-J., Raharimalala, N.E., Akaffou, S.D., Hamon, P., 2016.

Partial sequencing reveals the transposable element composition of *Coffea* genomes and provides evidence for distinct evolutionary stories. Molecular Genetics and Genomics 291, 1979-1990.

[Dataset] Guyot, R., Hamon, P., Couturon, E., Raharimalala, N., Rakotomalala, J.J., 2020, Photo bank for: WCSdb: A database of Wild Coffea Species. DataSuds, V1 https://doi.org/10.23708/JZA8I2.

[Dataset] Guyot, R., Charr, J-C., Garavito, A., Guyeux, C., Crouzillat, D., Descombes, P., Ly, S., Raharimalala, E., Rakotomalala, J-J., Stoffelen, P., Janssen, S., Hamon, P. 2020. Replication Data for: Complex evolutionary history of coffees including Coffea canephora (Robusta coffee) revealed by full plastid genomes and 28,800 nuclear SNP analyses. DataSuds, V1. https://doi.org/10.23708/KWRIJJ.

Hamon, P., Siljak-Yakovlev, S., Srisuwan, S., Robin, O., Poncet, V., Hamon, S., de Kochko, A., 2009. Physical mapping of rDNA and heterochromatin in chromosomes of 16 *Coffea* species: A revised view of species differentiation. Chromosome Research 17, 291-304.

Hamon, P., Duroy, P.-O., Dubreuil-Tranchant, C., Costa, P.M.D., Duret, C., Razafinarivo, N.J., Couturon, E., Hamon, S., de Kochko, A., Poncet, V., Guyot, R., 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.

Hamon, P., Rakotomalala, J.J., Akaffou, S., Razafinarivo, N.J., Couturon, E., Guyot, R., Crouzillat, D., Hamon, S., de Kochko, A., 2015. Caffeine-free species in the genus Coffea. *In* V. Preedy [ed.], Coffee in health and disease prevention 10.1016/b978-0-12-409517-5.00005-x DOI, 39-44. Academic Press, San Diego.

Hamon, P., Grover, C.E., Davis, A.P., Rakotomalala, J.J., Raharimalala, N.E., Albert, V.A., Sreenath, H.L., Stoffelen, P., Mitchell, S.E., Couturon, E., Hamon, S., de Kochko, A., Crouzillat, D., Rigoreau, M., Sumirat, U., Akaffou, S., Guyot, R., 2017. Genotyping-by-

sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species: GBS coffee phylogeny and the evolution of caffeine content. Molecular Phylogenetics and Evolution 109, 351-361.

Holland, B.R., Huber, K.T., Dress, A., Moulton, V. Delta plots: a tool for analyzing phylogenetic distance data. Mol Biol Evol. 2002;19(12):2051-2059.

Hubisz, M.J., Pollard, K.S., Siepel, A., 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. Briefings in Bioinformatics 12-1, 41-51.

Huson, D.H., Bryant, D., 2006. Application of Phylogenetic Networks in Evolutionary Studies, Molecular Biology and Evolution 23-2, 254-267.

International Coffee Organization, 2017. Coffee statistics. http://www.ico.org/trade_statistics.asp.

Jaillon, O., Aury, J., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pè, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.-F., Weissenbach, J., Quétier, F., Wincker, P., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–467.

James, J.K., Abbott, R.J., 2005. Recent, allopatric, homoploid hybrid speciation: The origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. Evolution 59, 2533-2547.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., dePamphilis, C.W., Leebens-Mack, J.H., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the United States of America 104, 19369-19374.

Kim J-H, Lee S-I, Kim B-R, Choi I-Y, Ryser P, Kim N-S (2017) Chloroplast genomes of *Lilium lancifolium*, *L. amabile*, *L. callosum*, and *L. philadelphicum*: Molecular characterization and their use in phylogenetic analysis in the genus *Lilium* and other allied genera in the order Liliales. PLoS ONE 12(10): e0186788.

Lai, Z., Nakazato, T., Salmaso, M., Burke, J.M., Tang, S., Knapp, S. J., Rieserberg L.H., 2005. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. Genetics 171, 291-303.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 354-357.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics, 23-21, 2947-2948.

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

Leroy, J.F., 1980. Evolution et taxogénèse chez les caféiers. Hypothèse sur leur origine. Comptes Rendus de I'Académie des Sciences, Paris 291, 593-596.

Leroy, J.-F., 1996. Biogéographie: quelques grands faits relatifs à la flore angiospermienne Malgache. Biogéographie de Madagascar, 59-71.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, G.P.D.P., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, Q., Brubaker, C.L., Green, A.G., Marshall, D.R., Sharp, P.J., Singh, S.P., 2001. Evolution of the *FAD2-1* fatty acid desaturase 5′ UTR intron and the molecular systematics of *Gossypium* (Malvaceae). American Journal of Botany 88, 92-102.

Louarn, J., 1972. Introduction à l'étude génétique des Mascarocoffea: nouvelles déterminations de leurs nombres chromosomiques. Café Cacao Thé XVI-4, 312-316.

Louarn, J., 1993. Structure génétique des caféiers africains diploïdes basée sur la fertilité des hybrides interspécifiques. Association Scientifique Internationale du Café, 15th, Colloque, Montpellier 6-11 June, 243-252.

Ly, S.N., Garavito, A., De Block, P., Asselman, P., Guyeux, C., Charr, J.C., Janssens, S., Mouly, A., Hamon, P., Guyot, R,. Chloroplast genomes of Rubiaceae: Comparative genomics and molecular phylogeny in subfamily Ixoroideae. PLoS One. 2020 Apr 30;15(4):e0232295

Matar, J., Khoury, H.E., Charr, J.C., Guyeux, C., Chrétien, S., 2019. SpCLUST: Towards a fast and reliable clustering for potentially divergent biological sequences. Computers in Biology and Medicine 114,103439.

Marques, I., Jurgens, A., Aguilar, J.F., Feliner, G.N., 2016. Convergent recruitment of new pollinators is triggered by independent hybridization events in *Narcissus*. The New Phytologist 210, 731-742.

Maurin, O., Davis, A.P., Chester, M., Mvungi, E.F., Jaufeerally-Fakim, Y., Fay, M.F., 2007. Towards a Phylogeny for *Coffea* (Rubiaceae): Identifying Well-supported Lineages Based on Nuclear and Plastid DNA Sequences. Annals of Botany 100, 1565-1583.

Mérot-L'anthoëne, V., Tournebize, R., Darracq, O., Rattina, V., Lepelley, M., Bellanger, L., Tranchant-Dubreuil, C., Coulee, M., Pegard, M., Metairon, S., Fournier, C., Stoffelen, P., Janssens, S. B., Kiwuka, C., Musoli, P., Sumirat, U., Legnate, H., Kambale, J.-L., Ferreira da Costa Neto, J., Revel, C., De Kochko, A., Descombes, P., Crouzillat, D., Poncet, V., 2019. Development and evaluation of a genome-wide Coffee 8.5K SNP array and its application for high density genetic mapping and for investigating the origin of *Coffea arabica* L. Plant Biotechnology Journal https://doi.org/10.1111/pbi.13066.

Mouly, A., Razafimandimbison, S.G., Khodabandeh, A., Bremer, B., 2009. Phylogeny and classification of the species-rich pantropical showy genus *Ixora* (Rubiaceae-Ixoreae) with implications of geographical monophyletic units and hybrids. American Journal of Botany 96, 686-706.

Mueller, L., Strickler, S., Domingues, D., Pereira, L., Andrade, A., Marraccini, P., Ming, R., Wai, J., Albert, V., Giuliano, G., Fiore, A., Pietrella, M., Aprea, G., Descombes, P., Moine, D., Guyot, R., Poncet, V., Hamon, P., Hamon, S., Dubreuil-Tranchant, C., Couturon, E., de Kochko, A., Lepelley, M., Bellanger, L., Mérot-L'Anthoëne, V., Vandecasteele, C., Rigoreau, M., Crouzillat, D., Paschoal-Rossi, A., Sankoff, D., Zheng, C., Kuhn, G., Korlach, J., Chin, J., 2015. Towards a better understanding of the Coffea arabica genome structure. *In: Proceedings of the 25th international conference on coffee science*. ASIC, 42-45.

Musoli, P., Cubry, P., Aluka, P., Billot, C., Dufour, M., De Bellis, F., Pot, D., Bieysse, D., Charrier, A., Leroy, T., 2009. Genetic differentiation of wild and cultivated populations: diversity of Coffea canephora Pierre in Uganda. Genome 52, 634-646.

Noirot, M., Poncet, V., Barre, P., Hamon, P., Hamon, S., de Kochko, A., 2003. Genome Size Variations in Diploid African Coffea Species. Annals of Botany 92, 709-714.

Noirot, M., Charrier, A., Stoffelen, P., Anthony, F., 2016. Reproductive isolation, gene flow and speciation in the former *Coffea* subgenus: a review. Trees 30, 597-608.

Nowak, M.D., Davis, A.P., Yoder, A.D., 2012. Sequence Data from New Plastid and Nuclear COSII Regions Resolves Early Diverging Lineages in *Coffea* (Rubiaceae). Systematic Botany 37, 995-1005.

Osborn, T.C., 2004. The contribution of polyploidy to variation in *Brassica* species. Physiologia Plantarum 121-4, 531-536.

Pan, J., Zhang, D., Sang, T., 2007. Molecular phylogenetic evidence for the origin of a diploid hybrid of *Paeonia* (Paeoniaceae). American Journal of Botany 94, 400-408.

Rambaut, A., 2007. FigTree. <http://tree.bio.ed.ac.uk/software/figtree/>.

Razafinarivo, N.J., Rakotomalala, J.-J., Brown, S.C., Bourge, M., Hamon, S., de Kochko, A., Poncet, V., Dubreuil-Tranchant, C., Couturon, E., Guyot, R., Hamon, P., 2012. Geographical gradients in the genome size variation of wild coffee trees (*Coffea*) native to Africa and Indian Ocean islands. Tree Genetics & Genomes 8, 1345-1358.

Rieseberg, L.H., 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): Evidence from ribosomal genes. American Journal of Botany 78, 1218-1237.

Rousseeuw, P.J., 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics 20, 53-65.

Samson, N., Bausher, M.G., Lee, S.B., Jansen, R.K., Daniell, H., 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.

Servedio, M.R., Van Doorn, G.S., Kopp, M., Frame, A.M., Nosil, P., 2011. Magic traits in speciation: 'magic' but not rare? Trends in Ecology & Evolution 26, 389-397.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., Liu, C., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Research 47, 65-73.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Stoffelen, P., Cheek, M., Bridson, D., Robbrecht, E., 1997. A New Species of *Coffea* (Rubiaceae) and Notes on Mount Kupe (Cameroon). Kew Bulletin 52-4, 989-994.

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

Stoffelen, P., Noirot, M., Couturon, E., Bontems, S., De Block, P., Anthony, F., 2009. *Coffea anthonyi*, a new self-compatible Central African coffee species, closely related to an ancestor of *Coffea arabica.* Taxon 58-1, 133-140.

Stoffelen, P., Robbrecht, E., Smets, E., 1996. Coffea (Rubiaceae) in Cameroon: a new species and a nomen recognized as species. Belgian Journal of Botany 129-1, 71-76.

Stoffelen, P., Noirot, M., Couturon, E., Anthony, F., 2008. A new caffeine-free coffee from Cameroon. Botanical Journal of the Linnean Society 158, 67-72.

Sun, M., Soltis, D.E., Soltis, P.S., Zhu, X., Burleigh, J.G., Chen, Z., 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae. Molecular Phylogenetics and Evolution 83, 156-166.

Taylor, S.J., Rojas, L.D., Ho, S.W., Martin, N.H., 2013. Genomic collinearity and the genetic architecture of floral differences between the homoploid hybrid species *Iris nelsonii* and one of its progenitors, *Iris hexagona*. Heredity 110, 63-70.

The International Peach Genome Initiative. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nature Genetics 45-5, 487-94.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., et al. 2017. GeSeq{versatile and accurate annotation of organelle genomes. Nucleic acids research 45(W1):W6–W11.

Wang, Z-H., Peng, H., Kilian, N., 2013. Molecular Phylogeny of the *Lactuca* Alliance (Cichorieae Subtribe Lactucinae, Asteraceae) with Focus on Their Chinese Centre of Diversity Detects Potential Events of Reticulation and Chloroplast Capture. PLoS ONE 8-12, e82692.

Wang, X.R., Szmidt, A.E., Savolainen, O., 2001. Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan plateau. Genetics 159, 337-346.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., Quandt, D., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Molecular Biology 76, 273-297.

Wolfe, A.D., Xiang, Q.Y., Kephart, S.R., 1998. Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). Proceedings of the National Academy of Sciences, USA 95, 5112-5115.

Wu, G.A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F.R., Carbonell-Caballero, J., Alonso, R., Curk, F., Du, D., Ollitrault, P., Roose, M.L., Dopazo, J., Gmitter Jr, F.G., Rokhsar, D.S., Talon, M., 2018. Genomics of the origin and evolution of *Citrus.* Nature 554, 311-317.

Yi, T.S., Jin, G.H., Wen, J., 2015. Chloroplast capture and intra-and inter-continental biogeographic diversification in the Asian–New World disjunct plant genus *Osmorhiza* (Apiaceae). Molecular Phylogenetics and Evolution 85, 10-21.

Yakimowski, S.B., Rieseberg, L.H. 2014. The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. American Journal of Botany 101-8, 1247-1258.

Yu, Q., Guyot, R., de Kochko, A., Byers, A., Navajas-Pérez, R., Langston, B.J., Dubreuil-Tranchant, C., Paterson, A.H., Poncet, V., Nagai, C.I., Ming, R., 2011. Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*). The Plant Journal 67, 305-317.

Zhou, Y., Duvaux, L., Ren, G., Zhang, L., Savolainen, O., Liu, J., 2017. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. Heredity 118, 211-220.

CAPTIONS

Fig. 1. Maternal phylogeny.

Maternal phylogeny on 52 *Coffea* and *Psilanthus* accessions plus *Empogona congesta* as outgroup. The tree was built using RaxML and GTR as model of substitution. Only the support values equal or superior to 85% are considered significant. Apart from *C. canephora* accessions marked in green, *Psilanthus* taxa are marked in blue while the *Coffea* taxa are in red. The geographical regions of origin are as follows, NEA: North East Africa; WCA: West and Central Africa; MUS: Mauritius; MDG-COM: Madagascar & Comoros; EA: East Africa; WA: West Africa; E-CEA: East Central East Africa. MC1, MC2, MC3 and MC4 refer to clades determined in this study (colored boxes).

Fig. 2. Nuclear phylogeny.

Coffee species relationships of 49 accessions plus *Empogona congesta* as an outgroup based on 28,800 SNPs according to the methodology developed in Hamon et al. (2017). The tree was built using RaxML and GTR as the model of substitution. The *Psilanthus* clade and taxa close to *Psilanthus* in the phylogenetic plastid-based tree (MC1, MC2, MC3; Fig. 2) are marked in blue while the *Coffea* clade and taxa with *Coffea* plastid (MC4) are indicated in red. *C. canephora* accessions are marked in green. The geographic region of origin is as follows, NEA: North-Eastern Africa, WCA: West & Central Africa; E-CEA: East Central-East Africa; EA: East Africa; MUS: Mauritius; MDG-COM: Madagascar and Comoros. MC1, MC2, MC3 and MC4 clades are indicated by colored boxes.

Fig. 3. Phylogenetic networks.

Phylogenetic networks from complete plastid sequences (top), and from 28,800 nuclear SNPs (bottom) showing complex species relationships for all clades except the Madagascan-

Comorian *Coffea* (MDG-COM). Species with *Coffea*- or *Psilanthus*-like plastid genomes are marked in red and blue respectively.

Fig. 4. Scenario for Retention of ancestral polymorphism.

Scenario of the distribution of plastid genomes according to a hypothesis of retention of ancestral polymorphism followed by unequal random sorting, especially for the *Coffea* species originating from West and Central Africa. Blue circles indicate MC1, MC2 and MC3 plastid genomes, while the red circles are for the MC4 plastid genomes. MC1, MC2, MC3 and MC4 clades are indicated with the same colors as for Figures 1 and 2.

Fig. A.1. Pattern of length variation of the plastid genomes and their different regions. Relative to the smallest plastid genome (here, *P. brassii*) the variation of the total length in bp and length for the different regions (LSC, SSC and IR) is plotted for each species by total length in ascending order.

Fig. A.2. Number of mutated genes per species (given next to the species name) when compared to the ancestors.

Phylogenetic trees showing the number of mutated genes per species and per group. Branch values are bootstraps according to the maximum likelihood of phylogenetic reconstruction. An arbitrary suffix has been added to provide unique names to ancestral nodes (for instance 100 x and 100 5) during the ancestral reconstruction process. The number of mutated genes relative to the ancestor per species is listed next to the species name for Group 1 including all Asian *Psilanthus,* the Central Northeast African *P. leroyi* and the Coastal Northeast African *C. rhamnifolia*; Group 2 is composed of West and Central African *Psilanthus* and the majority of West and Central African *Coffea*; Group 3 is formed exclusively of *Coffea* species (mainly island species and East African species).

Fig. A.3. Gene mutation rates per species among each group.

The gene mutation rate between a species and its ancestor for a given gene is equal to the Levenshtein distance between the translated gene sequences of the species and the ancestor divided by the length of their alignment. These rates were estimated independently for the three groups, Gene rate mutation is given on the y-axis. On the x-axis, the Cp genomes with the genes ordered by their position are divided into two parts (A and B) with Groups 1 to 3 from top to bottom. Each species within a group is represented by a color listed in the legend of the figure with Group 1: All-Asian *Psilanthus, P. leroyi* and *C. rhamnifolia*), Group 2:

West and Central African species including both *Psilanthus* and *Coffea* with the exception of the western African *C. humilis* and *C. stenophylla,* and Group 3: the remaining *Coffea* species.

Fig. A.4. Vectorial clustering.

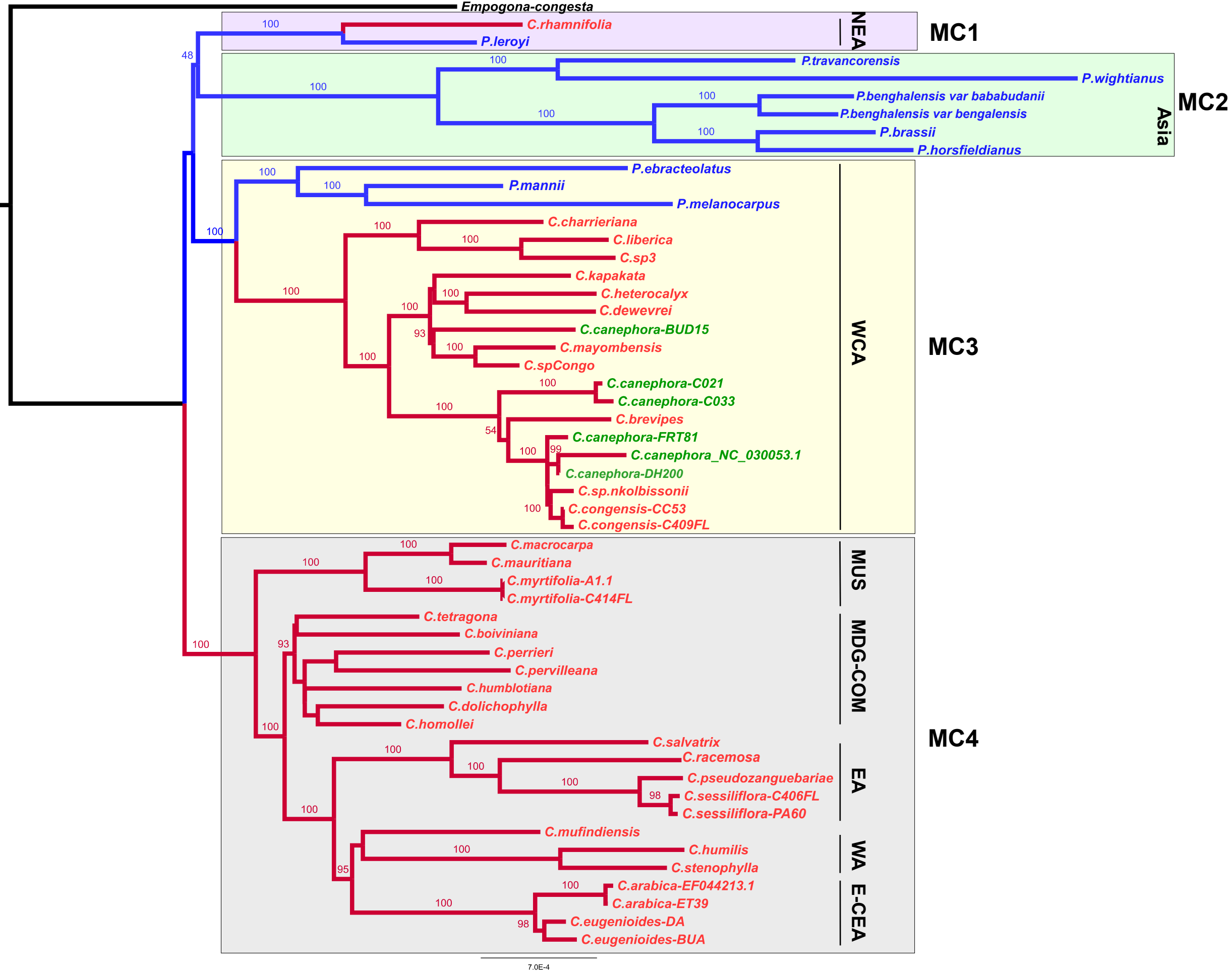
Clusters and sub-clusters resulting from the vectorial clustering analyses. WCA: West & Central Africa; EA: East Africa; E-CA: East-Central Africa. MC groups are indicated.

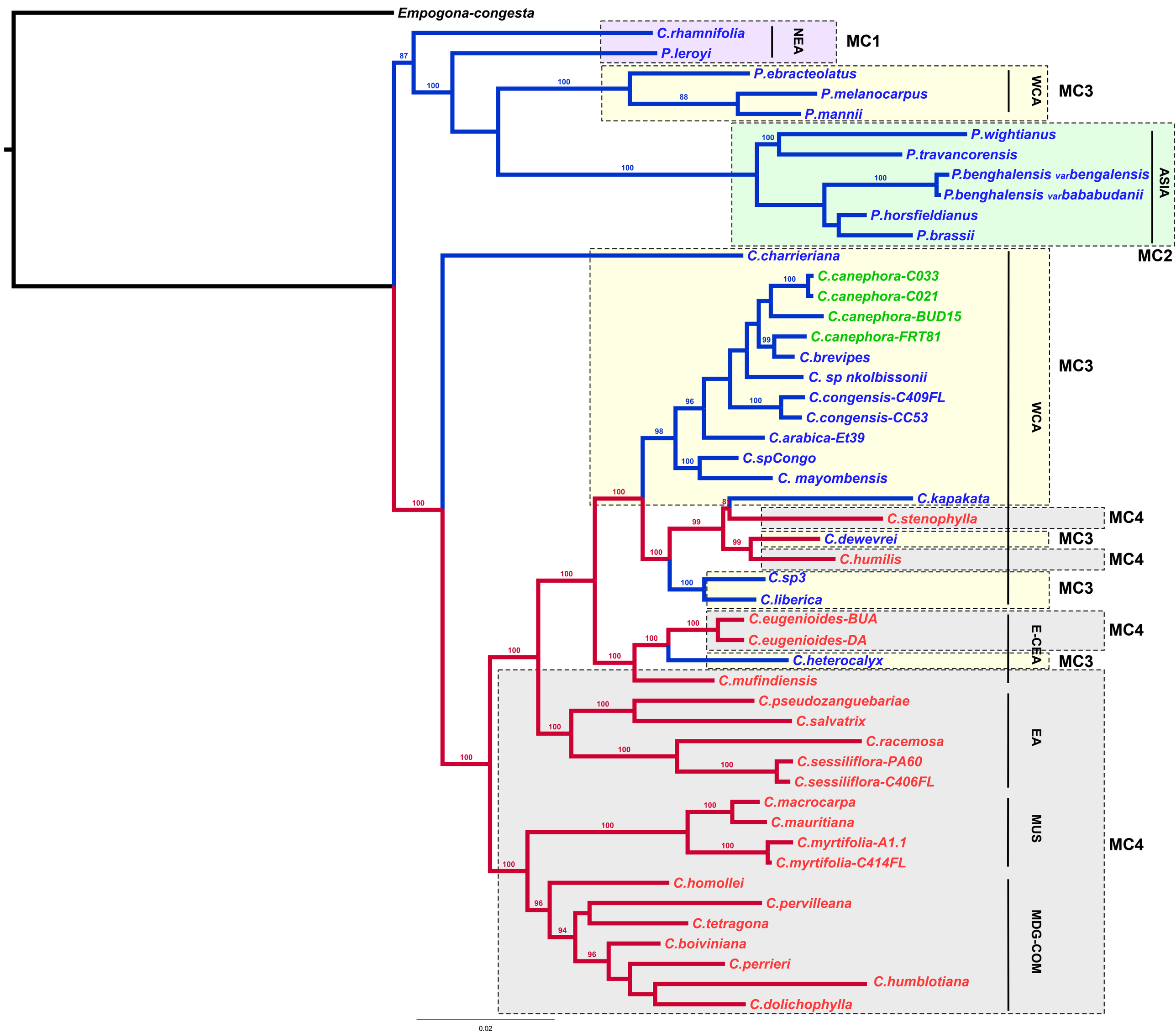
Supplementary data 1.

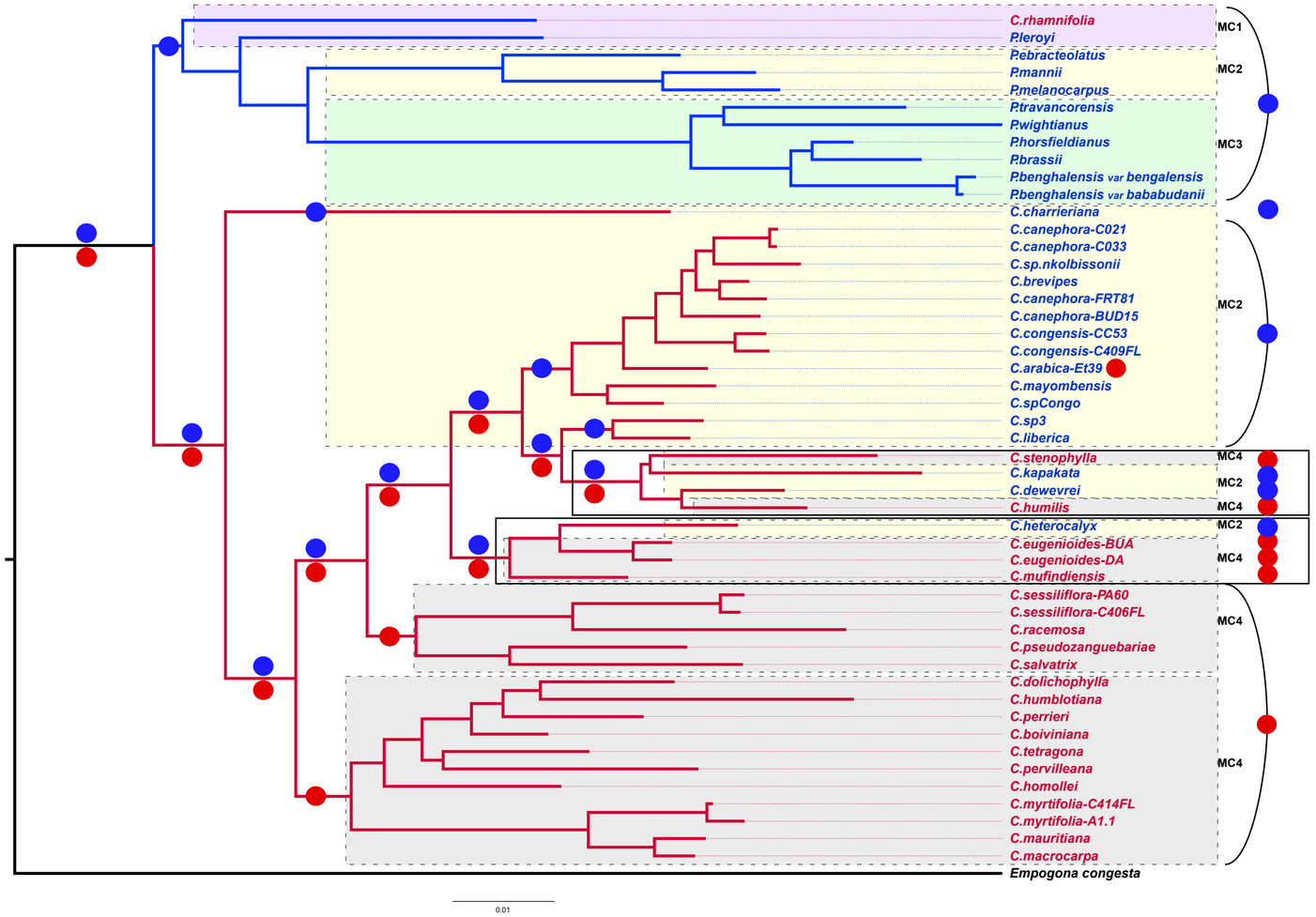
Fasta sequences of nuclear SNPs.

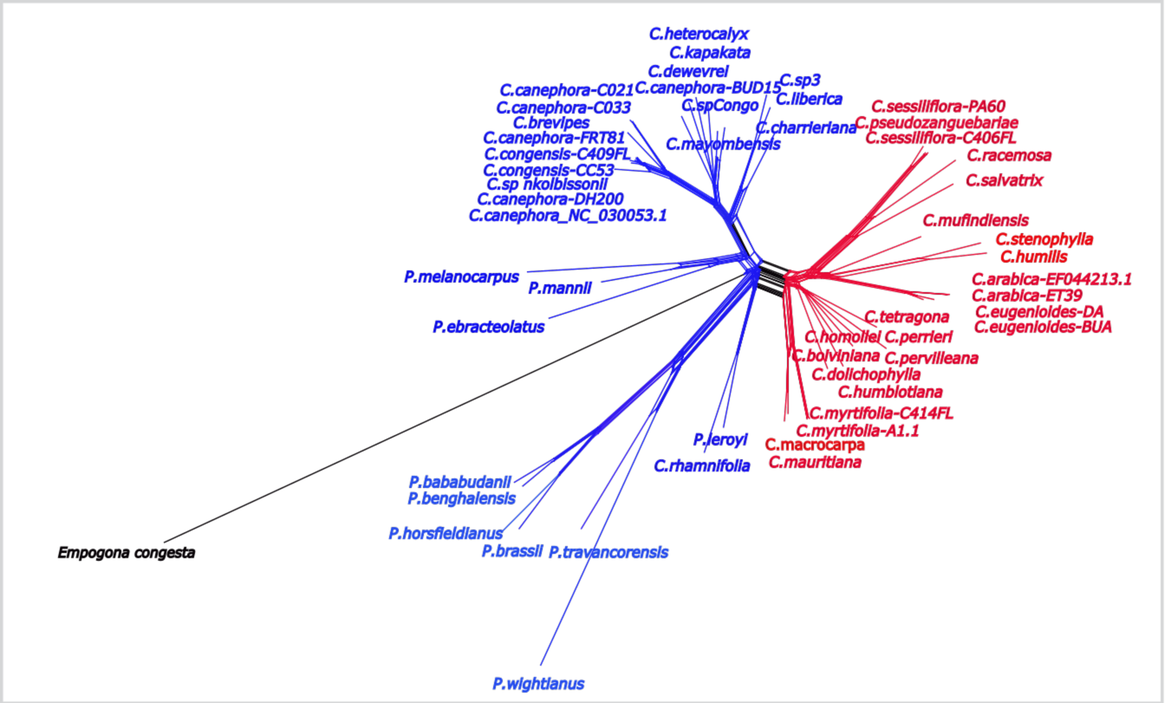
Table 1. Species considered in this study

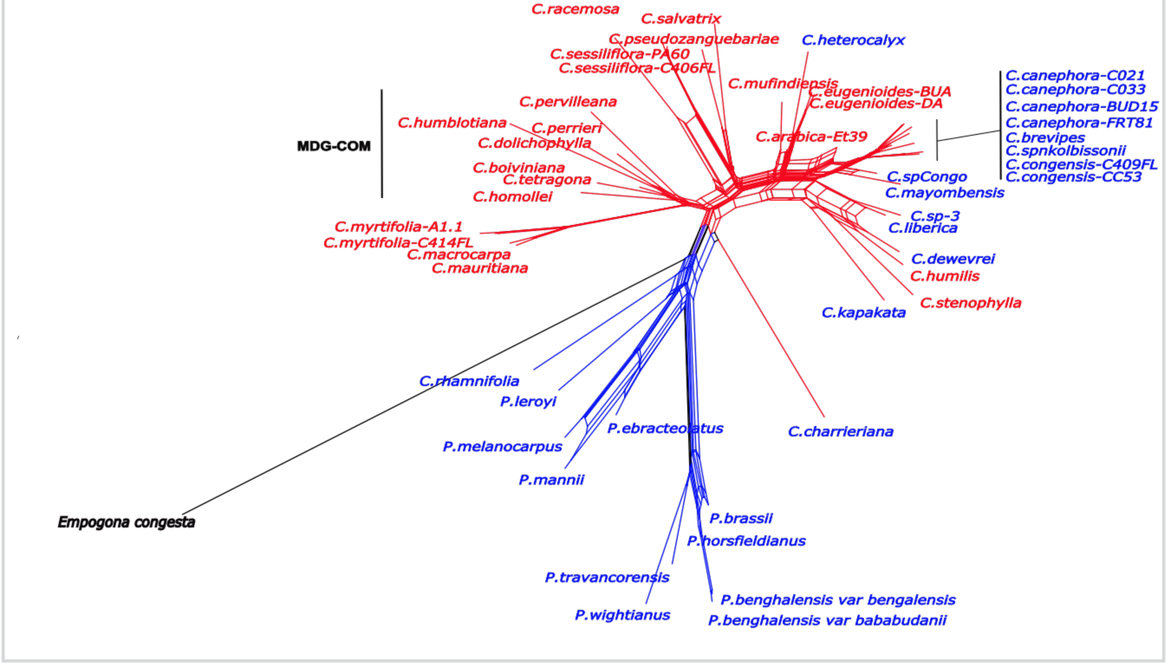
Herbarium codes follow Holmgren et al. (1990). Germplasm source: BR (Meise Botanic Garden, Belgium), BRC (Biological Resources Center, Reunion), CBI (Coffee Board of India), CNS (Australian Tropical Herbarium), K (Royal Botanic Gardens, Kew, UK), KCRS (Kianjavato Coffee Research Station, Madagascar), P (Natural History Museum, Paris, France).


Empogona-congesta
C.rhamnifolia
P.leroyi
NEA
MC1
P.travancorensis
P.wightianus
P.benghalensis var bababudanii
P.benghalensis var bengalensis
P.brassii
P.horsfieldianus
Asia
MC2
P.ebracteolatus
P.mannii
P.melanocarpus
C.charrieriana
C.liberica
C.sp3
C.kapakata
C.heterocalyx
C.dewevrei
C.canephora-BUD15
C.mayombensis
C.spCongo
C.canephora-C021
C.canephora-C033
C.brevipes
C.canephora-FRT81
C.canephora_NC_030053.1
C.canephora-DH200
C.sp.nkolbissonii
C.congensis-CC53
C.congensis-C409FL
WCA
MC3
C.macrocarpa
C.mauritiana
C.myrtifolia-A1.1
C.myrtifolia-C414FL
MUS
C.tetragona
C.boiviniana
C.perrieri
C.pervilleana
C.humblotiana
C.dolichophylla
C.homollei
MDG-COM
C.salvatrix
C.racemosa
C.pseudozanguebariae
C.sessiliflora-C406FL
C.sessiliflora-PA60
EA
C.mufindiensis
C.humilis
C.stenophylla
WA
C.arabica-EF044213.1
C.arabica-ET39
C.eugenioides-DA
C.eugenioides-BUA
E-CEA
MC4
7.0E-4

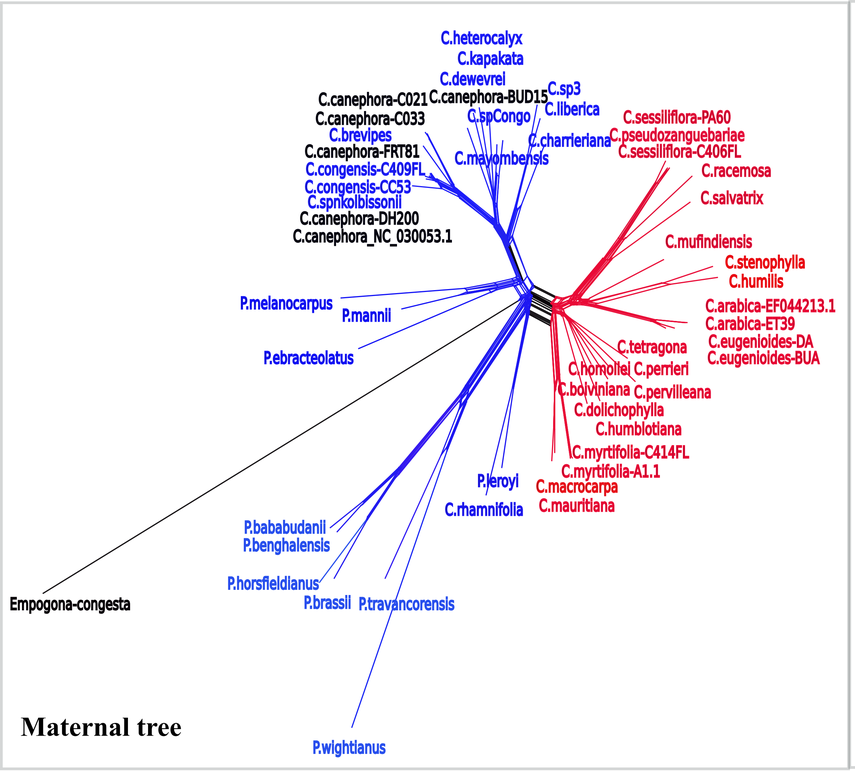

Empogona-congesta
C.rhamnifolia
P.leroyi
NEA
MC1
P.ebracteolatus
P.melanocarpus
P.mannii
WCA
MC3
P.wightianus
P.travancorensis
P.benghalensis varbengalensis
P.benghalensis varbabadudanii
P.horsfieldianus
P.brassii
ASIA
MC2
C.charrieriana
C.canephora-C033
C.canephora-C021
C.canephora-BUD15
C.canephora-FRT81
C.brevipes
C. sp nkolbissonii
C.congensis-C409FL
C.congensis-CC53
C.arabica-Et39
C.spCongo
C. mayombensis
C.kapakata
MC3
C.stenophylla
MC4
C.dewevrei
MC3
C.humilis
MC4
C.sp3
C.liberica
MC3
C.eugenioides-BUA
C.eugenioides-DA
E-CEA
MC4
C.heterocalyx
MC3
C.mufindiensis
C.pseudozanguebariae
C.salvatrix
C.racemosa
C.sessiliflora-PA60
C.sessiliflora-C406FL
EA
C.macrocarpa
C.mauritiana
C.myrtifolia-A1.1
C.myrtifolia-C414FL
MUS
C.homollei
C.pervilleana
C.tetragona
C.boiviniana
C.perrieri
C.humblotiana
C.dolichophylla
MDG-COM
MC4
0.02

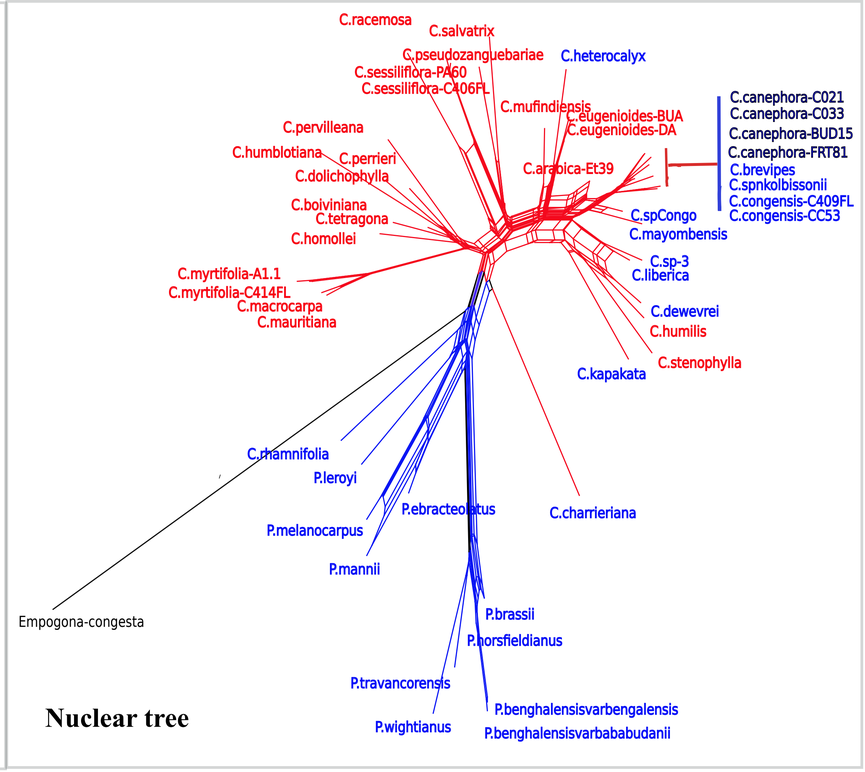

C.rhamnifolia
P.leroyi
P.ebracteolatus
P.mannii
P.melanocarpus
P.travancorensis
P.wightianus
P.horsfieldianus
P.brassii
P.benghalensis var bengalensis
P.benghalensis var bababudanii
C.charrieriana
C.canephora-C021
C.canephora-C033
C.sp.nkolbissonii
C.brevipes
C.canephora-FRT81
C.canephora-BUD15
C.congensis-CC53
C.congensis-C409FL
C.arabica-Et39
C.mayombensis
C.spCongo
C.sp3
C.liberica
C.stenophylla
C.kapakata
C.dewevrei
C.humilis
C.heterocalyx
C.eugenioides-BUA
C.eugenioides-DA
C.mufindiensis
C.sessiliflora-PA60
C.sessiliflora-C406FL
C.racemosa
C.pseudozanguebariae
C.salvatrix
C.dolichophylla
C.humblotiana
C.perrieri
C.boiviniana
C.tetragona
C.pervilleana
C.homollei
C.myrtifolia-C414FL
C.myrtifolia-A1.1
C.mauritiana
C.macrocarpa
Empogona congesta
MC1
MC2
MC3
MC2
MC4
MC2
MC4
MC2
MC4
MC4
MC4
0.01


C.heterocalyx
C.kapakata
C.dewevrei
C.canephora-C021
C.canephora-BUD15
C.sp3
C.liberica
C.canephora-C033
C.spCongo
C.sessiliflora-PA60
C.brevipes
C.charrieriana
C.pseudozanguebariae
C.canephora-FRT81
C.mayombensis
C.sessiliflora-C406FL
C.congensis-C409FL
C.racemosa
C.congensis-CC53
C.salvatrix
C.sp nkolbissonii
C.canephora-DH200
C.mufindiensis
C.canephora_NC_030053.1
C.stenophylla
C.humilis
P.melanocarpus
C.arabica-EF044213.1
P.mannii
C.arabica-ET39
C.eugenioides-DA
P.ebracteolatus
C.tetragona
C.eugenioides-BUA
C.homollei
C.perrieri
C.boiviniana
C.pervilleana
C.dolichophylla
C.humblotiana
C.myrtifolia-C414FL
C.myrtifolia-A1.1
P.leroyi
C.macrocarpa
C.rhamnifolia
C.mauritiana
P.babadudanii
P.benghalensis
P.horsfieldianus
Empogona congesta
P.brassii
P.travancorensis
P.wightianus


C.racemosa
C.salvatrix
C.pseudozanguebariae
C.heterocalyx
C.sessiliflora-PA60
C.sessiliflora-C406FL
C.mufindiensis
C.canephora-C021
C.eugenioides-BUA
C.canephora-C033
C.eugenioides-DA
C.canephora-BUD15
C.pervilleana
C.canephora-FRT81
C.humblotiana
C.perrieri
C.arabica-Et39
C.brevipes
MDG-COM
C.dolichophylla
C.spnkolbissonii
C.congensis-C409FL
C.boiviniana
C.spCongo
C.congensis-CC53
C.tetragona
C.mayombensis
C.homollei
C.sp-3
C.myrtifolia-A1.1
C.liberica
C.myrtifolia-C414FL
C.dewevrei
C.macrocarpa
C.humilis
C.mauritiana
C.stenophylla
C.kapakata
C.rhamnifolia
P.leroyi
C.charrieriana
P.ebracteolatus
P.melanocarpus
P.mannii
Empogona congesta
P.brassii
P.horsfieldianus
P.travancorensis
P.benghalensis var bengalensis
P.wightianus
P.benghalensis var bababudanii

Species name	Plant code	voucher herbarium	Country of origin	Source of leave material or DNA sequence
Coffea				
C. arabica L.	NA	NA	NA	EF044213.1 / NC_008535
C. arabica L.	Et39	-	Ethiopia	Mueller et al. 2015
C. boiviniana (Baill.) Drake	A.980	-	Madagascar	KCRS
C. brevipes Hiern.	C417FL	-	Cameroon	BRC
C. canephora Pierre ex A.Froehner	DH200-94	-	Democratic Republique of Congo	Denoeud et al. 2014
C. canephora Pierre ex A.Froehner	-	-	-	NC_030053.1
C. canephora Pierre ex A.Froehner	BUD15	-	Uganda	Mueller et al., 2015
C. canephora Pierre ex A.Froehner	C021	-	Ivory Coast	Mueller et al., 2015
C. canephora Pierre ex A.Froehner	C033	-	Ivory Coast	Mueller et al. 2015
C. canephora Pierre ex A.Froehner	FRT81 (cultivated)	-	Brazil	R&D Nestlé
C. charrieriana Stoff.	OA22	-	Cameroon	BRC
C. congensis A.Froehner	C409FL	-	NA	BRC
C. congensis A.Froehner	CC53	-	Republique of Congo	BRC
C. dewevrei De Wild. & T.Durand	EB51	-	Centrafrique Republique	BRC
C. dolichophylla J.-F.Leroy	A.206 (P)	-	Madagascar	KCRS
C. eugenioides S.Moore	DA (P)	-	Kenya	BRC
C. eugenioides S.Moore	BUA	-	Uganda	Mueller et al. 2015
C. heterocalyx Stoff.	C413FL	-	Cameroon	BRC
C. homollei J.-F.Leroy	SZ	-	Madagascar	KCRS
C. humblotiana Baill.	BM19/20 (K, MO, TAN)	-	Comoros	BRC
C. humilis A.Chev.	G57 (K)	-	Ivory Coast	BRC
C. kapakata (A.Chev.) Bridson	KAP	-	Angola	BRC
C. liberica W.Bull. ex Hiern	EA61	-	Ivory Coast	BRC
C. macrocarpa A.Rich.	PET (P, K)	-	Mauritius	BRC
C. mauritiana Lam	BM17-25	-	Mauritius	BRC
C. mayombensis A.Chev.	-	Dessein & Sonké 1447	Cameroon	BR
C. mufindiensis Hutch. ex Bridson	-	2917	Tanzania	K
C. myrtifolia (A.Rich. ex DC.) J.-F.Leroy	A1.1	-	Mauritius	BRC
C. myrtifolia (A.Rich. ex DC.) J.-F.Leroy	C414FL	-	Mauritius	BRC
C. sp. 'nkolbisonii'	-	Lachenaud 594	Cameroon	BR
C. perrieri Drake ex Jum. & H.Perrier	A.12	-	Madagascar	KCRS
C. pervilleana Drake	A.957	-	Madagascar	KCRS
C. pseudozanguebariae Bridson	H53 (K)	-	Kenya	BRC
C. racemosa Lour.	IB62 (K)	-	Mozambique	BRC
C. rhamnifolia (Chiov.) Bridson	-	P04003534	Somalia	P
C. salvatrix Swynn. & Philipson	C408FL	-		BRC
C. sessiliflora Bridson	C406FL	-	Tanzania	BRC
C. sessiliflora Bridson	PA60	-	Tanzania	BRC
C. sp 3	-	Dessein, Lachenaud, Janssens, Issembe & Nzabi 2583	Cameroon	BR
C. sp. 'Congo'	C416FL	-	Congo	BRC
C. stenophylla G.Don.	FB55 (K)	-	Ivory Coast	BRC
C. tetragona Jum. & H.Perrier	A.252 (K, MO, TAN)	-	Madagascar	KCRS
Psilanthus				
P. benghalensis var. bababudanii (Sivar., Biju & P.Mathew) A.P.Davis	PBT1 (CCRI)	-	India	CBI
P. benghalensis (Heyne ex J.A.Schult.) J.-F. Leroy	PBT5 (CCRI)	-	India	CBI
P. brassii (J.-F.Leroy) A.P.Davis	-	D. Crayn 1196 (CNS)	Australia	CNS
P. ebracteolatus Hiern	PSI11 (K, P)	-	Ivory Coast	BRC
P. horsfieldianus (Miq.) J.-F.Leroy	HOR (K)	-	Indonesia	ICRI
P. leroyi Bridson	-	6008	Sudan	K
P. mannii Hook.f.	-	2003 1365-45	Cameroon	BR
P. melanocarpus (Welw. ex Hiern) J.-F.Leroy	-	P00128349	Angola	P
P. travancorensis (Wight & Arn.) J.-F.Leroy	-	PBT2-S51	India	CBI
P. wightianus (Wall. ex Wight & Arn.) J.-F.Leroy	-	PBT3-S50	India	CBI
Coffeeae tribe (outgroup)				
Empogona congesta (Oliv.) Cheek	-	Dessein et al. 1103	Zambia	BR


C.heterocalyx
C.kapakata
C.dewevrei
C.canephora-C021
C.canephora-BUD15
C.sp3
C.liberica
C.canephora-C033
C.spCongo
C.sessiliflora-PA60
C.brevipes
C.charrieriana
C.pseudozanguebariae
C.canephora-FRT81
C.mayombensis
C.sessiliflora-C406FL
C.congensis-C409FL
C.racemosa
C.congensis-CC53
C.salvatrix
C.spnkolbissonii
C.canephora-DH200
C.mufindiensis
C.canephora_NC_030053.1
C.stenophylla
C.humilis
P.melanocarpus
C.arabica-EF044213.1
P.mannii
C.arabica-ET39
C.eugenioides-DA
P.ebracteolatus
C.tetragona
C.eugenioides-BUA
C.homollei
C.perrieri
C.boiviniana
C.pervilleana
C.dolichophylla
C.humblotiana
C.myrtifolia-C414FL
C.myrtifolia-A1.1
P.leroyi
C.macrocarpa
C.rhamnifolia
C.mauritiana
P.babadudanii
P.benghalensis
P.horsfieldianus
Empogona-congesta
P.brassii
P.travancorensis
P.wightianus
Maternal tree


C.racemosa
C.salvatrix
C.pseudozanguebariae
C.heterocalyx
C.sessiliflora-PA60
C.sessiliflora-C406FL
C.mufindiensis
C.canephora-C021
C.eugenioides-BUA
C.canephora-C033
C.eugenioides-DA
C.canephora-BUD15
C.pervilleana
C.canephora-FRT81
C.humblotiana
C.perrieri
C.arabica-Et39
C.brevipes
C.dolichophylla
C.spnkolbissonii
C.boiviniana
C.congensis-C409FL
C.tetragona
C.spCongo
C.congensis-CC53
C.homollei
C.mayombensis
C.sp-3
C.myrtifolia-A1.1
C.liberica
C.myrtifolia-C414FL
C.macrocarpa
C.dewevrei
C.mauritiana
C.humilis
C.stenophylla
C.kapakata
C.rhamnifolia
P.leroyi
C.charrieriana
P.ebracteolatus
P.melanocarpus
P.mannii
Empogona-congesta
P.brassii
P.horsfieldianus
P.travancorensis
P.benghalensisvarbengalensis
P.wightianus
P.benghalensisvarbababudanii
Nuclear tree