

## Research Article

## Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae)

Niklas Wikström<sup>1,2\*</sup> , Birgitta Bremer<sup>1</sup>, and Catarina Rydin<sup>1,2</sup> <sup>1</sup>Bergius Foundation, The Royal Swedish Academy of Sciences and Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm SE-10691, Sweden<sup>2</sup>Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm SE-10691, Sweden

\*Author for correspondence. E-mail: niklas.wikstrom@bergianska.se

Received 30 August 2019; Accepted 2 January 2020; Article first published online 3 January 2020

**Abstract** Reconstructions of phylogenetic relationships in the flowering plant family Rubiaceae have up until now relied heavily on single- or multi-gene data, primarily from the plastid compartment. With the availability of cost- and time-efficient techniques for generating complete genome sequences, the opportunity arises to resolve some of the relationships that, up until now, have proven problematic. Here, we contribute new data from complete 58 plastid genome sequences, representing 55 of the currently 65 recognized tribes of the Rubiaceae. Also contributed are new data from the nuclear rDNA cistrons for corresponding taxa. Phylogenetic analyses are conducted on two plastid data sets, one including data from the protein coding genes only, and a second where protein coding data are combined with non-coding regions, and on a nuclear rDNA data set. Our results clearly show that simply adopting a “more characters” approach does not resolve the relationships in the Rubiaceae. More importantly, we identify conflicting phylogenetic signals in the data. Analyses of the same plastid data, treated as nucleotides or as codon-degenerated data, resolve and support conflicting topologies in the subfamily Cinchonoideae. As these analyses use the same data, we interpret the conflict to result from erroneous assumptions in the models used to reconstruct our phylogenies. Conflicting signals are also identified in the analyses of the plastid versus the nuclear rDNA data sets. These analyses use data from different genomic compartments, with different inheritance patterns, and we interpret the conflicts as representing “real” conflicts, reflecting biological processes of the past.

**Key words:** coffee family, conflicting signals, phylogenomics, plastid DNA, rDNA, Rubiaceae.

## 1 Introduction

The flowering plant family Rubiaceae, or the coffee family, comprises close to 14 000 species, and recent classifications recognize somewhere around 600 genera, 65 tribes, and three different subfamilies (Bremer et al., 1999; Bremer & Eriksson, 2009; Davis et al., 2009; Rydin et al., 2009a; Kainulainen et al., 2013; Mouly et al., 2014; Govaerts et al., 2018). Representatives of this species-rich family can be found in all regions of the world, even on the subantarctic islands, but most occur in tropical and subtropical regions across the globe (Govaerts et al., 2018). Also from ecological, morphological, and functional perspectives, the diversity of the family is considerable: life forms span from small annual and perennial herbs to large tropical trees; flowers are adapted to a broad range of different pollinators, including bats (Sazima et al., 1999); fruit types include animal dispersed fleshy fruits to dry fruits

such as nuts and capsules, the latter often with small and wind dispersed seeds (Robbrecht, 1988; Eriksson & Bremer, 1991); epiphytes, lianas, succulents, rheophytes, and aquatic life forms can be found; and members of the family inhabit a wide range of different habitats, from dry desert like conditions to wet tropical rain forests (Robbrecht, 1988).

Over the last 20–30 years, a large number of papers have been published with the primary focus of resolving phylogenetic relationships in Rubiaceae at different taxonomic levels (see Robbrecht & Manen, 2006; Bremer, 2009 for reviews of work up until 2009). This work has resulted in a reasonably well-resolved and coherent picture of the phylogenetic relationships in the family (Fig. 1). Two early diverging groups are consistently recognized and supported (e.g., Bremer et al., 1995; Andersson & Rova, 1999; Robbrecht & Manen, 2006; Bremer & Eriksson, 2009; Rydin et al., 2009a; Wikström et al., 2015). The first corresponds to the subfamily

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

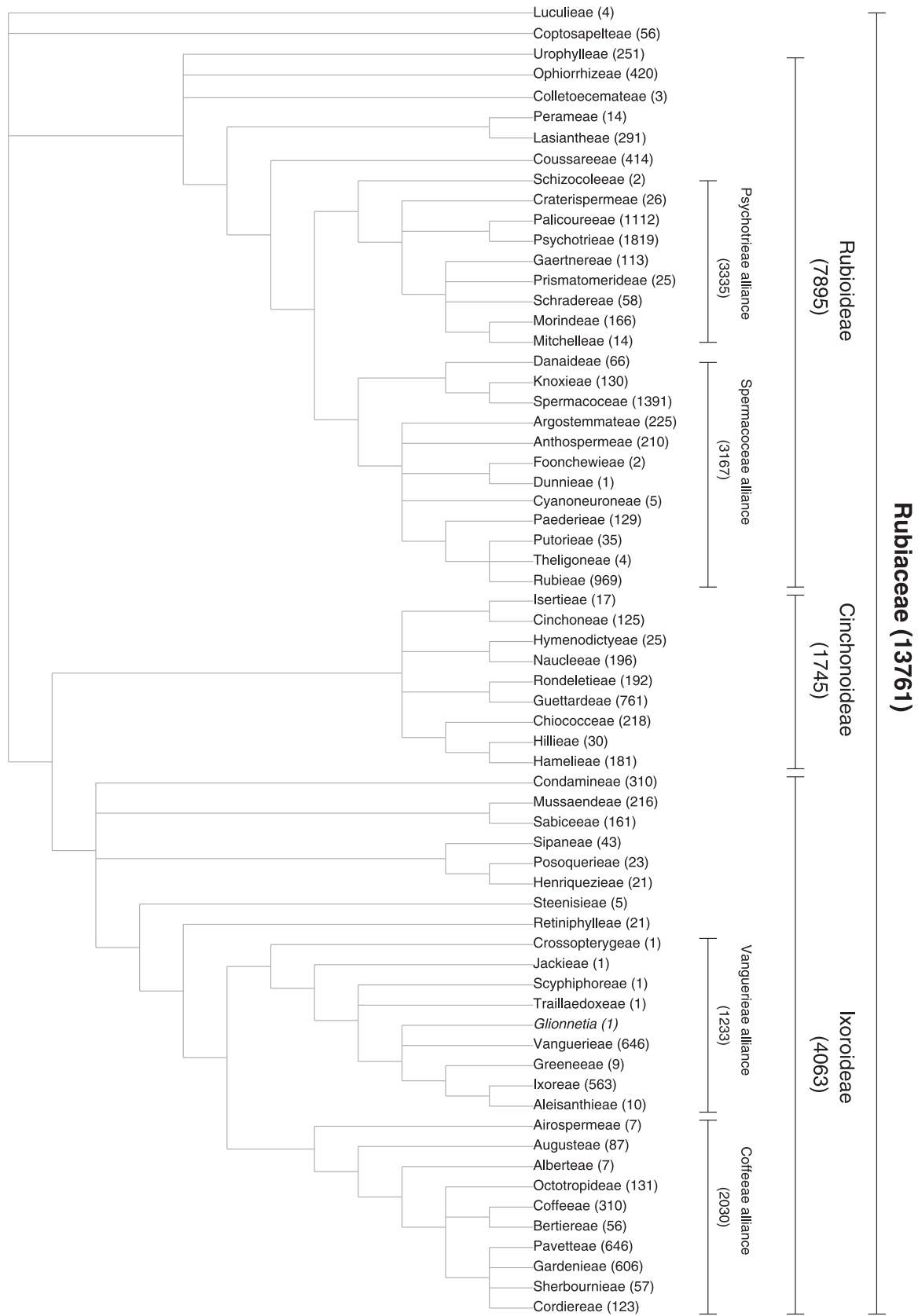


Fig. 1. Continued

Rubioideae, the largest of three subfamilies, with almost 8000 species and 27 different tribes. The second comprises the other two subfamilies: Cinchonoideae with about 1700 species and 9 (or 10) recognized tribes; and Ixoroideae with about 4000 species and 26 recognized tribes. The latter two subfamilies are sometimes united as one (Cinchonoideae s. lat. Robbrecht & Manen, 2006). More detailed analyses of the Rubioideae resolve the subfamily in two informal alliances: the Spermaceae alliance (with 10 tribes recognized by Rydin et al. (2009b) and two additional tribes described more recently by Wen & Wang (2012) and Ginter et al. (2015)); and the Psychotriaceae alliance (with 9 tribes recognized by Razafimandimbison et al., 2014, 2017), each alliance including more than 3000 species. Sister to the two alliances are Coussareeae (414 species) and a grade of tribes including Lasiantheae (291 species), Colletocemateae (3 species), Urophyllaeae (251 species), and Ophiorrhizeae with around 420 species (Bremer & Eriksson, 2009; Rydin et al., 2009a, 2009b; Wikström et al., 2015). Also, a fifth tribe, Perameae (14 species), a neotropical tribe including the single genus *Perama*, should probably be included here and has been resolved together with Lasiantheae in several analyses (Andersson & Rova, 1999; Robbrecht & Manen, 2006; Smedmark et al., 2014). In the subfamily Cinchonoideae, four groups have emerged as well supported (Antonelli et al., 2009; Manns & Bremer, 2010; Manns et al., 2012; Wikström et al., 2015): the first includes the two predominantly paleotropical tribes, Hymenodictyeae (25 species) and Naucleaeae (196 species), the second includes the “quinine-tribe” Cinchoneae (125 species) and Isertieae (17 species), the third includes Guettardeae (761 species) and Rondeletieae (192 species), and the last group includes the tribes Hamelleae (181 species), Hillieae (30 species), and Chicocceae (218 species). A fourth tribe Strumpfieae should possibly be recognized in this last group (Paudyal et al., 2014). Two informal alliances are commonly recognized also in the subfamily Ixoroideae: the Coffeaeae alliance with more than 2000 species and 10 tribes (as recognized by Kainulainen et al. (2013) and Mouly et al. (2014)); and the Vanguerieae alliance with about 1200 species and eight different tribes (Kainulainen et al., 2013), and a grade of Retiniphyllaeae (21 species), Steenisieae (5 species), Mussaendeae (216 species) together with Sabiceaeae (161 species), Condamineae (310 species), and a clade comprising Henriquezieae (21 species), Posoquerieae (23 species), and Sipaneeae with about 43 species (Kainulainen et al., 2010, 2013; Razafimandimbison et al., 2011).

There are of course uncertainties with respect to detailed relationships in many of the recognized tribes, but with respect to the overall relationships among the recognized tribes, a few problems stand out as having been particularly difficult to resolve (Fig. 1). The relationships of the two small tribes Luculieae, with a single genus and a handful of species, and Coptosapelteae, with two genera and 50–60 species, for example, are not well understood. They show inconsistent and often unsupported relationships in most analyses (Robbrecht & Manen, 2006; Bremer & Eriksson, 2009; Rydin et al., 2009a; Manns et al., 2012; Wikström et al., 2015), and they remain unclassified with respect to their subfamily placement. In the subfamily Rubioideae, relationships among early diverging lineages are problematic. Usually, the tribe Lasiantheae is grouped sister to a group comprising Coussareeae and the Psychotriaceae and Spermaceae alliances, but the tribes Colletocemateae, Ophiorrhizeae, and Urophyllaeae show inconsistent and/or poorly supported relationships in most analyses (Robbrecht & Manen, 2006; Bremer & Eriksson, 2009; Rydin et al., 2009a; Manns et al., 2012; Wikström et al., 2015). Also, some of the relationships among tribes in the Psychotriaceae and Spermaceae alliances are problematic (Rydin et al., 2009b; Razafimandimbison et al., 2017). In the Cinchonoideae, relationships among the four well supported groups that have emerged in recent analyses are not understood, and alternative resolutions have resulted in widely different ideas about the geographic origins and expansions of not only the subfamily, but of the Rubiaceae as a whole (Antonelli et al., 2009; Manns et al., 2012). In the Ixoroideae, some of the relationships in the Vanguerieae and Coffeaeae alliances (Kainulainen et al., 2013; Mouly et al., 2014), and between the early diverging Condamineae, Mussaendeae–Sabiceaeae, and Henriquezieae–Posoquerieae–Sipaneeae clades have been problematic to resolve (Kainulainen et al., 2013).

As in most plant groups (Zimmer & Wen, 2012; Davis et al., 2014; Ruhfel et al., 2014; Rothfels et al., 2015), previous work in Rubiaceae have relied almost exclusively on data from the plastid genome. There have been indications of conflicts between analyses that have strictly used data from the plastid genome, and those that have combined their plastid data with data from the nuclear genome (Robbrecht & Manen, 2006; Rydin

**Fig. 1.** Summary of overall phylogenetic relationships among recognized tribes and larger informal groups in the Rubiaceae. Collapsed nodes indicate problems that have been particularly difficult to resolve. Two small tribes, the Luculieae and the Coptosapelteae, have shown inconsistent and unsupported relationships in most analyses and remain unclassified with respect to their subfamily placement (Rydin et al., 2009a). In the subfamily Rubioideae problems concern the relationships among early diverging lineages (Rydin et al., 2009a), but also to some extent the relationships within the Psychotriaceae (Razafimandimbison et al., 2017) and Spermaceae (Rydin et al., 2009b) alliances. In the Cinchonoideae analyses have identified four well supported groups, but the same analyses show incongruent and unsupported results with respect to the relationships among these groups (Antonelli et al., 2009; Manns & Bremer, 2010; Manns et al., 2012; Wikström et al., 2015). In the Ixoroideae the delimitation of and relationships between Gardenieae and associated tribes in the Coffeaeae alliance have been problematic (Mouly et al., 2014), as have the relationships among early diverging lineages in the subfamily (Kainulainen et al., 2013). Numbers in parenthesis are approximate numbers of recognized species in each group. These were calculated by taking species counts for recognized genera from Govaerts et al. (2018) and extracting their most up to date tribal placements from the literature.

et al., 2009a, 2009b), but the reasonably well resolved and primarily plastid-based picture of overall phylogenetic relationships in the family was until recently more or less unchallenged. A recent work by Rydin et al. (2017) changed this and they demonstrated unambiguous and well supported conflicts between this plastid-based consensus and relationships obtained in their analyses of mitochondrial data. Among the supported conflicts they reported, two stood out as particularly striking: the rejection of monophyly of subfamilies Cinchonoideae and Ixoroideae with two tribes from subfamily Cinchonoideae, the “quinine-tribe” Cinchoneae and the Isertieae, nested well inside subfamily Ixoroideae; and the grouping of Jackieae from the Vanguerieae alliance with Airospermeae from the Coffeaeae alliance (Rydin et al., 2017). Neither relationship has been reported before, and the authors speculated that the conflicts possibly reflect ancient hybridization events between early members of the subfamilies Cinchonoideae and Ixoroideae and of the Vanguerieae and Coffeaeae alliances (Rydin et al., 2017).

Adopting a “more characters” phylogenomic approach (Rokas et al., 2003; Delsuc et al., 2005; Davis et al., 2014) we here use complete data from the maternally inherited plastid genome with the aim to resolve relationships that have remained problematic in previous plastid-based analyses of the family. We conduct phylogenetic analyses on 59 Rubiaceae and four outgroup taxa. The effect of any compositional heterogeneity in our data, induced by synonymous substitutions, is investigated by comparing results from nucleotide-based analyses of the protein coding data with corresponding analyses of codon degenerated data that uses ambiguity codes to eliminate

the distinction between synonymous codons. To further explore the consequences of a biased employment of plastid data, we also conduct analyses on corresponding taxa using data from the biparentally inherited nuclear rDNA cistron.

## 2 Material and Methods

### 2.1 Taxon sample

Taxa were targeted with the objective of having all, or as many as possible, tribes currently recognized in the family represented in the analyses. In total, 56 of the 65 commonly recognized tribes of Rubiaceae were included. Aleisanthieae, Condamineae, Crossopterygeae, Cyanoneuroneae, Foonchewieae, Henriquezieae, Perameae, Schradereae, and Strumpfieae were the tribes not included. The genus *Glionnetia* from the Seychelles, currently unclassified at the tribal level, was also included. The phylogenetic position of *Glionnetia* has been problematic to estimate but the genus is commonly resolved in an unsupported position in the Vanguerieae alliance (Razafimandimbison et al., 2011; Kainulainen et al., 2013; Wikström et al., 2015). Two non-Rubiaceae taxa from Gentianales were included as outgroup taxa in the analyses of ribosomal rDNA data: *Mostuea* Dir. (Gelsemiaceae), and *Asclepias* L. In the analyses of plastid data; these two taxa were complemented with two additional outgroup taxa: *Rhazya* Decne. (Apocynaceae) and *Gentiana* L. (Gentianaceae). Detailed species and voucher information for all included taxa is given in Table 1.

**Table 1** List of taxon names, DNA voucher information, and accession numbers for sequences used in the analyses

| Taxon  | DNA voucher/reference                 | rDNA cistron<br>accession numbers | Plastid<br>accession number |
|--|---------------------------------------|-----------------------------------|-----------------------------|
| <i>Alberta magna</i> E.Mey.                          | J. E. Tonkin 200 (UPS)                | MK607891                          | KY348839                    |
| <i>Amphidasya ambigua</i> (Standl.) Standl.          | Clark and Watt 736 (QCNE,<br>MO, UPS) | MK607892                          | KY378703                    |
| <i>Anthospermum spathulatum</i> Spreng.              | Bremer et al. 4405 (UPS)              | MK607893                          | KY378687                    |
| <i>Argostemma yappii</i> King                        | Bremer and Bremer 1609 (S)            | MK607894                          | KY378693                    |
| <i>Asclepias syriaca</i> L.                          | Straub et al. (2013)                  | --                                | NC_022432                   |
| --   | Straub et al. (2011)                  | JF312046                          | --                          |
| <i>Augusta austrocaledonica</i> (Brogn.) J.H.Kirkbr. | Mouly et al. 237 (P)                  | MK607890                          | KY492076                    |
| <i>Bertiera longithyrza</i> Baker                    | Kårehed et al. 256 (UPS)              | MK607895                          | KY348833                    |
| <i>Boholia nematostylis</i> Merr.                    | D. Bicknell 1561A (S)                 | MK607896                          | KY348840                    |
| <i>Cephalanthus occidentalis</i> L.                  | NTBG-960457.002                       | MK607897                          | KY378678                    |
| <i>Cinchona pubescens</i> Vahl                       | Bremer 2733 (UPS)                     | MK607898                          | KY378682                    |
| <i>Coffea arabica</i> L.                             | Samson et al. (2007)                  | --                                | NC_008535                   |
| --   |                                       | NW_020849278                      | --                          |
| <i>Coltoecema dewevrei</i> (De Wild.) E.M.A.Petit    | Lisowski 47195 (K)                    | MK607899                          | KY378707                    |
| <i>Coptosapelta flavescens</i> Korth.                | Puff 950720-1/2 (WU, BKF)             | MK607900                          | KY378704                    |
| <i>Craterispermum</i> sp.                            | Eriksson et al. 999 (S)               | MK607901                          | KY378662                    |
| <i>Cubanola domingensis</i> (Britton) Aiello         | Bremer 4500 (S)                       | MK607902                          | KY378677                    |
| <i>Danais xanthorrhoea</i> (K.Schum.) Bremek.        | Bremer 3079 (UPS)                     | MK607903                          | KY378686                    |
| <i>Deppea grandiflora</i> Schltldl.                  | Bremer 2724 (UPS)                     | MK607904                          | KY378675                    |
| <i>Dunnia sinensis</i> Tutch.                        | Longmen 10 (in Ge et al., 2002)       | MK607905                          | KY378692                    |

Continued

**Table 1** Continued

| Taxon   | DNA voucher/reference             | rDNA cistron accession numbers | Plastid accession number |
|---|-----------------------------------|--------------------------------|--------------------------|
| <i>Famea multiflora</i> A.Rich. ex DC.                          | Bremer et al. 3331 (UPS)          | MK607906                       | KY378701                 |
| <i>Gaertnera phyllostachya</i> Baker                            | Kårehed et al. 272 (UPS)          | MK607907                       | KY378695                 |
| <i>Gardenia conferta</i> Guillaumin                             | Barabbé 330 (NOU)                 | MK607908                       | KY348835                 |
| <i>Gentiana straminea</i> Maxim.                                | Ni et al. (2014)                  | --                             | NC_027441                |
| <i>Glionnetia sericea</i> (Baker) Tirveng.                      | Bremer et al. 5404 (S)            | MK607909                       | KY378665                 |
| <i>Greenea oblonga</i> Craib.                                   | Larsen et al. 33451 (P)           | MK607910                       | KY378664                 |
| <i>Guettarda scabra</i> (L.) Vent.                              | Rova 2260 (GB)                    | MK607911                       | KY378680                 |
| <i>Gynochthodes officinalis</i> (F.C.How) Razafim. & B.Bremer   | Ding et al. (2015)                | --                             | NC_028009                |
| <i>Hillia triflora</i> (Oerst.) C.M.Taylor                      | Bremer 3101 (UPS)                 | MK607912                       | KY378676                 |
| <i>Hymenodictyon parvifolium</i> Oliv.                          | B. Bremer 3809 (UPS)              | MK607913                       | KY378679                 |
| <i>Iseritia laevis</i> (Triana) B.M.Boom                        | B. Bremer et al. 3364 (UPS)       | MK607914                       | KY378683                 |
| <i>Ixora hookeri</i> (Oudem.) Bremek.                           | Kainulainen et al. 182 (S)        | MK607915                       | KY378663                 |
| <i>Ixora parviflora</i> Lam.                                    | Gillis 7892 (FTG)                 | MK607916                       | KY378672                 |
| <i>Jackiopsis ornata</i> (Wall.) Ridsdale                       | K. S. Tan s.n.                    | MK607917                       | KY378669                 |
| <i>Kohautia caespitosa</i> Schnizl.                             | Bremer et al. 42566B (UPS)        | MK607918                       | KY378684                 |
| <i>Lasianthus</i> sp.   | Kainulainen et al. 17 (S)         | MK607919                       | KY378708                 |
| <i>Luculia grandifolia</i> Ghose                                | B. Bremer 2713 (S)                | MK607920                       | KY378705                 |
| <i>Mitchella repens</i> L.                                      | Atha and Gonzalez 1443a (MEXU)    | MK607921                       | KY378710                 |
| <i>Mitriostigma axillare</i> Hochst.                            | Bremer and Rydin 5014b (S)        | MK607922                       | KY348837                 |
| <i>Morinda citrifolia</i> L.                                    | Bremer 3302 (UPS)                 | MK607923                       | KY378694                 |
| <i>Mostuea brunonis</i> Didr.                                   | B. Bremer et al. 5077 (S)         | MK607924                       | KY378706                 |
| <i>Mussaenda densiflora</i> H.L.Li                              | Razafimandimbison et al. 747 (S)  | MK607925                       | KY348834                 |
| <i>Ophiorrhiza mungos</i> L.                                    | Bremer 3301 (UPS)                 | MK607926                       | KY378702                 |
| <i>Paederia foetida</i> L.                                      | Wong and Keong s.n. (KLU)         | MK607927                       | KY378691                 |
| <i>Palicourea guianensis</i> Aubl.                              | B. Bremer et al. 3332 (UPS)       | MK607928                       | KY378697                 |
| <i>Paragenipa lancifolia</i> (Bojer ex Baker) Tirveng. & Robbr. | Bremer et al. 5413 (S)            | MK607929                       | KY348838                 |
| <i>Pavetta abyssinica</i> Fresen.                               | De Block 6 (BR)                   | MK607930                       | KY378673                 |
| <i>Pentas lanceolata</i> (Forssk.) Deflers                      | Bremer 2702 (S)                   | MK607931                       | KY378685                 |
| <i>Plocama pendula</i> Aiton                                    | K. Andreasen 1 (UPS)              | MK607932                       | KY378690                 |
| <i>Posoqueria latifolia</i> (Rudge) Roem. & Schult.             | Bergius Botanic Garden SU-C-88.10 | MK607933                       | KY378662                 |
| <i>Prismatomeris fragrans</i> Geddes                            | Kainulainen et al. 39 (S)         | MK607934                       | KY378699                 |
| <i>Psychotria kirkii</i> Hiern                                  | Bremer 3102 (UPS)                 | MK607935                       | KY378696                 |
| <i>Psydrax obovata</i> (Klotzsch ex Eckl. & Zeyh.) Bridson      | Bremer 3762 (UPS)                 | MK607936                       | KY378666                 |
| <i>Retiniphyllum pilosum</i> (Spruce ex Benth.) Müll.Arg.       | Wurdack and Adderley 43270 (S)    | MK607937                       | KY378670                 |
| <i>Rhazya stricta</i> Decne.                                    | Park et al. (2014)                | --                             | KJ485849                 |
| <i>Rondeletia odorata</i> Jacq.                                 | Bremer and Andreasen 3504 (UPS)   | MK607938                       | KY378681                 |
| <i>Rubia horrida</i> (Thunb.) Puff                              | Bremer et al. 4266 (UPS)          | MK607939                       | KY378689                 |
| <i>Sabicea diversifolia</i> Pers.                               | B. Bremer et al. 5262 (S)         | MK607940                       | KY378671                 |
| <i>Schizocolea linderi</i> (Hutch. & Dalziel.) Bremek.          | Adam 20116 (UPS)                  | MK607941                       | KY378700                 |
| <i>Scyphiphora hydrophyllacea</i> C.F.Gaertn.                   | K. Bremer et al. 99 (S)           | MK607942                       | KY378668                 |
| <i>Sipanea</i> sp.  | Rova et al. 1981 (GB)             | MK607943                       | KY378674                 |
| <i>Stachyarrhena heterochroa</i> Standl.                        | Persson et al. 821 (GB)           | MK607945                       | KY348836                 |
| <i>Steenisia pleurocarpa</i> (Airy Shaw) Bakh.f.                | Puff BF 990619-1/4 (WU)           | MK607944                       | KY378709                 |
| <i>Theligonum cynocrambe</i> L.                                 | Reuterswärd and Forsslund 2 (S)   | MK607946                       | KY378688                 |
| <i>Trailliaedoxa gracilis</i> W.W.Sm. & Forrest                 | Boufford et al. 35041 (MO)        | MK607947                       | KY378667                 |

References are given for sequences taken from Genbank.

## 2.2 DNA extraction and sequencing

DNA was extracted from herbarium, live, or silica-dried material using a cetyltrimethylammonium bromide (CTAB) protocol (Doyle & Doyle, 1987; Doyle, 1991). The extracted DNA was cleaned using the QIAquick PCR cleaning kit from Qiagen (Hilden, Germany) following the protocol specified by the manufacturer.

High-throughput sequencing was done at the Science for Life Laboratory (SciLifeLab, Stockholm, Sweden) following the manufacturer's instructions for the Illumina HiSeq 2500 platform (Illumina, San Diego, California, USA). Pair-end runs with 300-bp insert size fragments and  $2 \times 150$  bp read lengths were performed. Samples were multiplexed with 14, 17, or 28 other samples in one lane in three consecutive runs. Library preparation at the SciLifeLab was done using the Illumina TruSeq DNA PCR-free library preparation kit (Illumina) for samples with high quantities of DNA, and the Thruplex DNA-seq library preparation kit from Rubicon (Rubicon Genomics, Ann Arbor, Michigan, USA) for samples with lower quantities of DNA. Demultiplexing and conversion was done using CASAVA v1.8.2 (Illumina).

## 2.3 Plastid sequence assembly

To isolate plastid sequences from the original reads, a BLAT (BLAST-like alignment tool v36; Kent, 2002) search of forward and reverse reads against an initial database of whole plastid genomes of eight Lamiid taxa was performed. The initial database included the plastid genomes of *Coffea arabica* (Rubiaceae; NC\_008535), *Gynochthodes officinalis* (Rubiaceae; NC\_028009), *Nicotiana undulata* (Solanaceae; NC\_016068), *Catharanthus roseus* (Apocynaceae; NC\_021423), *Asclepias nivea* (Apocynaceae, NC\_022431), *Rhazya stricta* (Apocynaceae; KJ485849), *Boea hygrometrica* (Gesneriaceae; NC\_016468), and *Scrophularia takasimensis* (Scrophulariaceae; MK590983). Forward and reverse reads were both saved if either showed at least 70% similarity to any of the Lamiid genomes. After the BLAT search, reads were extracted from the original fastq data files using pullseq v1.0.1 ([github.com/bctomas/pullseq](https://github.com/bctomas/pullseq)) into new forward and reverse fastq data files representing a “chloroplast” subset. De novo assembly of the “chloroplast” subset of reads was performed for each taxon using ABySS v1.5.2 and eight different k-mer lengths (55, 61, 67, 73, 85, 91, 97, 103). All generated contigs were pooled and mapped onto a reference genome using bwa v0.7.5a-r405. Initially, the chloroplast genome of *Coffea arabica* (NC\_008535) was used as reference. This resulted in complete or near complete draft genomes. All original reads were subsequently mapped onto the draft genomes using bwa v0.7.5a-r405 allowing unfinished gaps to be filled and sequencing depths to be evaluated. Assemblies were reviewed and edited using gap5 from the Staden Package v2.0.ob10 (Staden, 1996; Staden et al., 2000). After the completion of each chloroplast genome, it was added to the database of chloroplast genomes used in the BLAT search and also considered as a potential reference genome in the final mapping stage in subsequent assemblies. Protein coding (CDS), tRNA, and rRNA genes were annotated using Sequin v15.10 (available by anonymous FTP at <https://www.ncbi.nlm.nih.gov/Sequin/>; National Center for Biotechnology

Information, Bethesda, Maryland, USA) by transferring the annotation of *Coffea arabica* (NC\_008535).

### 2.3.1 Assembly of nuclear rDNA cistrons

Nuclear rDNA cistrons were assembled in a corresponding way as the plastids (see above). The initial database for the BLAT search comprised the rDNA cistron sequence of *Asclepias syriaca* (JF312046), which was also used as the initial reference sequence for mapping the de novo generated contigs. Annotation of the sequences was done using Sequin v15.10 and boundaries for the 18S, 5.8S, and 28S rDNA genes were inferred from alignment with the *Coffea arabica* sequence NW\_020849278 REGION: complement (11780..17784).

## 2.4 Phylogenetic analyses

### 2.4.1 Plastid data

**2.4.1.1 Alignment and pretrimming** Protein coding (CDS), intron, tRNA, rDNA, and intergenic spacer (IGS) regions were extracted from the annotated GenBank files using an in-house python script built on biopython v1.63 (Cock et al., 2009). Script is available from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.q573n5tf4>). Stop codons and regions of uncertain homology were identified by eye and removed using a python script developed and made available at the Dryad Digital Repository (<https://doi.org/10.5061/dryad.q8nov>) by Fučíková et al. (2016). Extracted DNA regions were aligned individually using Muscle v3.8.31 (Edgar, 2004), with protein-coding gene regions aligned as amino acid sequences and converted back to nucleotides using Seaview 4 (Gouy et al., 2010).

**2.4.1.2 Nucleotide based analyses** Individual protein-coding genes were concatenated into a combined CDS data set. Also, non-CDS data were assembled comprising the intron, tRNA, and rDNA regions, and phylogenetic analyses were subsequently conducted on two alternative data sets: one including only the CDS data, and a second with CDS and non-CDS data combined. Intergenic spacer regions were as a rule too variable to produce reliable alignments and were therefore not included in any of the analyses. Bayesian analyses were conducted in MrBayes v3.2.6 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). The CDS data were partitioned by codon position and the non-CDS data into two partitions, one comprising the tRNA and rDNA regions and a second including the intron data. The GTR+I+ $\Gamma$ 4 substitution model was used for all partitions based on the corrected Akaike information criterion (AICc) as calculated using the program MrAIC c1.4.6 (Nylander, 2004) and PHYML v3.0 (Guindon et al., 2010). Substitution models were unlinked across all partitions. MrBayes was run for 10 000 000 generations sampling trees and parameter estimates every 1000 generations. Two independent runs, each with four chains and heating parameters set to default values, were conducted for all analyses. Bayesian posterior probability values were calculated after discarding the first 25% of the trees and parameters as burnin, and this was well beyond the burnin phase of the chains based on the PSRF convergence diagnostic (Gelman & Rubin, 1992) reported by MrBayes.

Maximum Likelihood (ML) analyses were conducted in RAXML v8.2.9 (Stamatakis, 2014) with CDS and non-CDS data partitions and nucleotide substitution models set up as in the Bayesian analyses. Branch support was estimated using the

non-parametric bootstrap (Felsenstein, 1985) and 1000 ML pseudoreplicates.

**2.4.1.3 Degenerate codon analyses** To investigate the effect of possible compositional heterogeneity in our data induced by synonymous substitutions, we replaced the codons of all nucleotide sequences in the CDS data set with degenerate codons and analyzed this degenerate codon data set separately. Degenerate codons use nucleotide ambiguity codes to eliminate the distinction between synonymous codons (Criscuolo & Gribaldo, 2010; Regier et al., 2010; Criscuolo & Gribaldo, 2011; Regier & Zwick, 2011; Zwick et al., 2012; Rota-Stabelli et al., 2013; Cox et al., 2014). Most synonymous substitutions differ by a single substitution at the third codon position. However, codon variants of leucine (TTT/CTN) and arginine (CGN/AGR) may differ also in the first codon position, and those of serine (TCN/AGY) may differ in all three codon positions. All recoding was done using a modified version of the python translation script from Fučíková et al. (2016) available from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.q573n5tf4>). Both Bayesian and ML analyses of the degenerate codon data set were done in the same way as the nucleotide based analyses of CDS data.

#### 2.4.2 Nuclear rDNA data

Nuclear 18S, 5.8S, and 28S rDNA gene regions, and ITS1 and ITS2 rDNA spacer regions were extracted from the annotated GenBank files and aligned individually using MAFFT v7.271 (Katoh et al., 2002; Katoh & Standley, 2013). Parts of the ITS1 and ITS2 spacer regions were highly variable across our taxon sample and we used the transitive consistency score (TCS; Notredame et al., 2000; Chang et al., 2014) to assess alignment accuracy and to filter out ambiguously aligned columns from the ITS1 and ITS2 alignments. Alignment positions with a TCS score of 3 or less were excluded in the phylogenetic analyses. Also the two outgroup taxa were excluded from the alignments of the ITS spacer regions as they were highly divergent compared to our ingroup taxa. The five extracted rDNA regions were subsequently concatenated into a combined data set. Analyses were unpartitioned and the GTR+I+dΓ4 substitution model was used based on the corrected Akaike information criterion (AICc) as calculated using the program MrAIC c1.4.6 (Nylander, 2004) and PHYML v3.0 (Guindon et al., 2010). Bayesian and ML runs were set up in the same way as the plastid nucleotide based analyses and conducted on two alternative data sets: one where the ITS1 and ITS2 spacer regions were excluded from the analyses, and a second where they were included, but where positions with a TCS score of 3 or less were excluded.

## 3 Results

### 3.1 Assembled plastids

Plastid sequences were successfully assembled for 58 accessions, one from Gelsemiaceae and 57 from Rubiaceae. Assembled sequences comprise the large single copy region (LSC), the inverted repeat region (IR), and the small single copy region (SSC) and ranges in length from 125 128 bp in *Theligonum cynocrambe* to 130 778 in *Faramea multiflora*. The total number

of sequenced fragments varied across samples from  $1.5 \times 10^6$  fragments in *Cephalanthus occidentalis* to  $16.7 \times 10^6$  fragments in *Prismatomeris fragrans* with an average number of  $8.9 \times 10^6$  fragments. Assembled sequences were submitted to the DDBJ/EMBL/GenBank databases and a comprehensive list of accession numbers for all deposited sequences, together with taxon names and DNA voucher information is given in Table 1.

### 3.2 Assembled rDNA cistrons

Nuclear rDNA cistrons were assembled for the corresponding accessions. Assembled sequences comprise incomplete external transcribed spacer (ETS), 18S ribosomal RNA (18S rRNA) gene region, internal transcribed spacer 1 (ITS1), 5.8S ribosomal RNA (5.8S rRNA) gene region, internal transcribed spacer 2 (ITS2), 28S ribosomal RNA (28S rRNA) gene region, and incomplete non-transcribed spacer (NTS). Assembled sequences ranges in length from 5835 bp in *Rubia horrida* to 5980 bp in *Lasianthus* sp. See Table 1 for a comprehensive list of accession numbers for all deposited sequences.

### 3.3 Phylogenetic analyses

#### 3.3.1 Plastid data

**3.3.1.1 Nucleotide based analyses** The plastid CDS data set include a total of 77 protein-coding gene regions and 69 828 aligned nucleotide characters for 59 ingroup Rubiaceae (57 newly generated here) and four outgroup (one newly generated here) non-Rubiaceae Gentianales taxa. The combined CDS + non-CDS data set for corresponding taxa include a total of 95 494 nucleotide characters: 69 828 CDS and 25 666 non-CDS (18 899 intron; 2225 tRNA; 4542 rDNA). The complete partitioned and annotated data sets are available from Dryad Digital Repository (S1 and S2 Files; <https://doi.org/10.5061/dryad.q573n5tf4>). Results from the phylogenetic analyses are reported as phylograms from the Bayesian analyses in Fig. 2 (analysis of CDS data) and Fig. S1 (analysis of combined CDS and non-CDS data). Branch support is indicated in all trees by Bayesian posterior probabilities (BPP) and ML bootstrap frequencies (BS).

In the CDS analyses relationships are generally resolved and, with a few exceptions, well supported (Fig. 2). In subfamily Rubioideae the tribes Ophiorrhizeae, Urophyllaeae, and Colletocemateae are in unsupported positions as early diverging lineages. Lasiantheae followed by Coussareeae are successive sisters to a clade comprising the Psychotrieae and Spermaceae alliances. In the Psychotrieae alliance Schizocoleae are sister to remaining tribes followed by Craterispermeae, Psychotrieae + Palicoureeae, Prismatomerideae + Gaertnereae, and Morindeae + Mitchelleae. In the Spermaceae alliance Danaideae are grouped with Spermaceae plus Knoxieae and together they are sister to remaining tribes. Anthospermeae and Dunnieae are successive sisters to an unsupported group (BPP = 0.75; BS < 50%) with Argostemmatae sister to a clade comprising Paederieae, Putorieae, Theligoneae, and Rubieae. Within the latter clade, Rubieae are grouped with Theligoneae, with Putorieae and Paederieae as successive sisters. Coptosapelteae (BPP = 0.79; BS = 75%) and Luculieae (BPP = 0.91; BS = 81%) are resolved in positions at the base of subfamily Rubioideae.

In subfamily Cinchonoideae, Naucleaeae together with Hymenodictyeae are resolved sister to a clade (BPP = 1.00; BS = 77%) comprising remaining tribes in the subfamily. Chiococceae are sister to Hillieae plus Hamelieae, and they

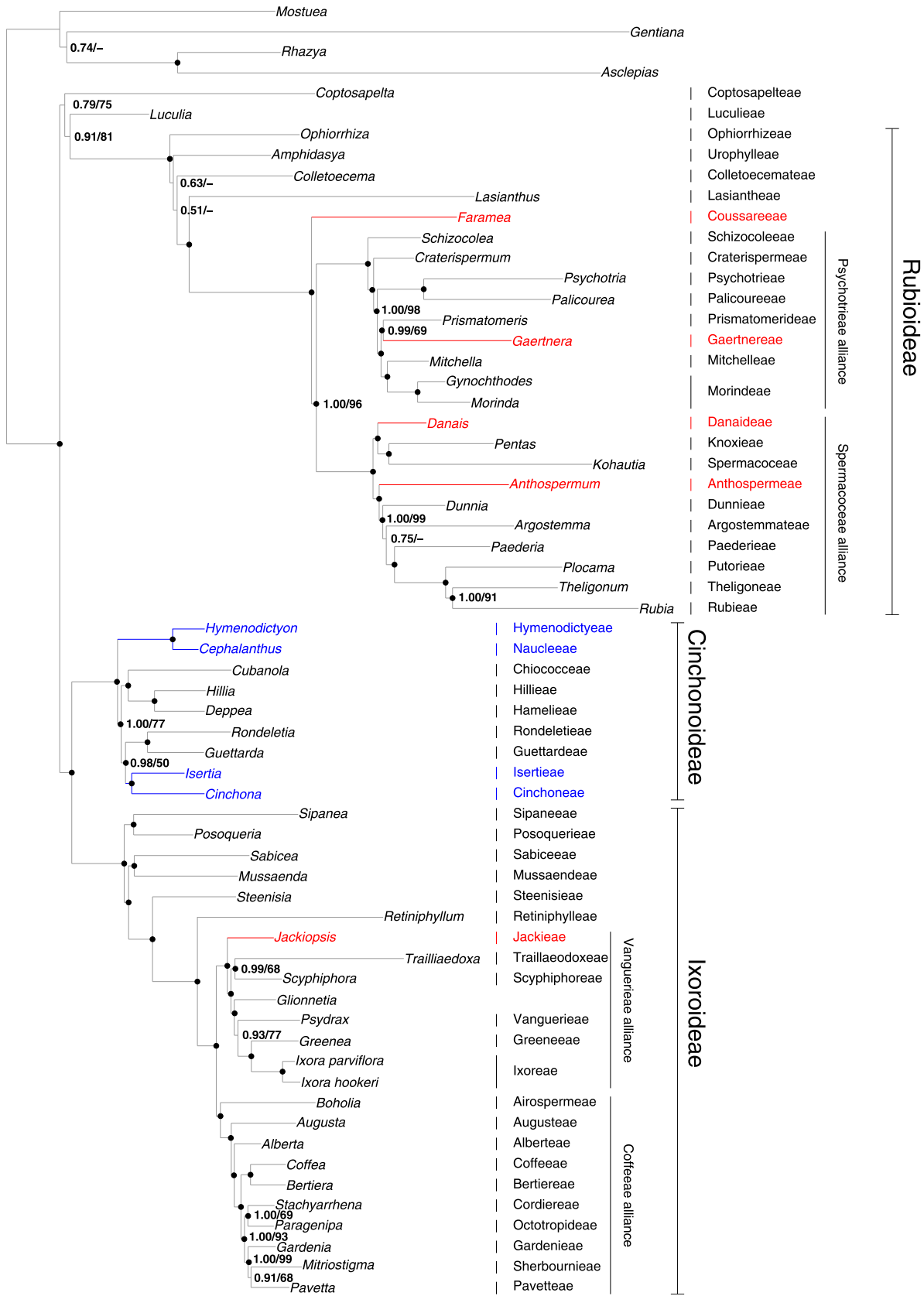


Fig. 2. Continued



are sister to a clade (BPP = 0.98; BS = 50%) where Rondeletieae plus Guettardeae are sister to Cinchoneae plus Isertieae.

Relationships in the subfamily Ixoroideae are generally supported by both Bayesian and ML analyses, exceptions being the relationships of *Glionnetia* sister to Vanguerieae and the Greeneae–Ixioereae clade (BPP = 0.93; BS = 77%), the sister group relationship between Traillaeodoxeae and Scyphiphoreae (BPP = 0.99; BS = 68%), and some relationships in the Gardenieae complex. Sipaneae with Posoquerieae are sister to remaining Ixoroideae followed by Sabiceae with Mussaendeae, Steenisieae, Retiniphyllae, and the Coffeae and Vanguerieae alliances. In the Vanguerieae alliance Jackieae are sister to remaining tribes followed by Traillaeodoxeae with Scyphiphoreae (BPP = 0.99; BS = 68%), *Glionnetia*, and an unsupported node (BPP = 0.93; BS = 77%) comprising the Vanguerieae and a Greeneae + Ixoreae clade. In the Coffeae alliance Airospermeae are sister to remaining tribes followed by Augusteae, Alberteae, Coffeae with Bertiereae, Cordiereae with Octotropideae (BPP = 1.00; BS = 69%), and Gardenieae sister to the unsupported Pavetteae + Sherbourneae clade (BPP = 0.91; BS = 68%).

In the analyses of combined CDS and non-CDS data the results obtained in the CDS analyses are as a rule corroborated with increased support (Fig. S1). Exceptions are the relationships of Urophyllae, Coltoecemateae, Theligoneae, and Octotropideae. Urophyllae and Coltoecemateae are resolved as sistergroups (BPP = 0.74; BS = 62%), the support for grouping Theligoneae and Rubieae is substantially decreased (BPP = 0.92; BS < 50%), and Octotropideae are no longer supported as sister to Cordiereae, but placed in an unsupported position as sister to a group with Gardenieae, Sherbourneae, and Pavetteae (BPP = 0.91; BS < 50%).

**3.3.1.2 Degenerate codon analyses** The complete partitioned and annotated codon degenerated data set is available in Nexus file format from Dryad Digital Repository (S3 File; <https://doi.org/10.5061/dryad.q573n5tf4>) Results from the phylogenetic analyses are reported as a phylogram (from the Bayesian analyses) in Fig. 3, with branch support indicated for each node by Bayesian posterior probabilities (BPP) and ML bootstrap frequencies. In the subfamilies Rubioideae and Ixoroideae the analyses show results highly similar to the nucleotide analyses. Both place Ophiorrhizeae as sister to remaining tribes in subfamily Rubioideae, but the support for this relationship (BPP = 0.99; BS = 76%) increases considerably here in the degenerate codon analyses (Fig. 3). Also the sister group relationship between Argostemmateae and the clade comprising Paederieae, Putorieae, Theligoneae, and Rubieae becomes supported (BPP = 1.00; BS = 85%), whereas the support for grouping Prismatomerideae with Gaertnereae decreases (BPP = 0.83; BS = 59%) in

the degenerate codon analyses. In subfamily Ixoroideae the support for grouping Cordiereae with Octotropideae is lost (BPP = 0.90; BS < 50%), and Pavetteae goes from an unsupported position as sister to Sherbourneae to a position as sister to Gardenieae (BPP = 0.95; BS = 69%).

In the subfamily Cinchonoideae the situation is different. Here the results clearly conflict with those from the nucleotide based analyses (Figs. 2, S1) with Isertieae together with Cinchoneae forming a sistergroup relationship to remaining tribes (BPP = 0.99; BS = 59%), and with Naucleae together with Hymenodictyeae now in a more derived but unsupported position (BPP = 0.89; BS = 50%) as sister to a group comprising Rondeletieae and Guettardeae.

### 3.3.2 Nuclear rDNA data

The nuclear rDNA data set included a total of 6045 aligned nucleotide characters for 58 ingroup Rubiaceae (57 newly generated here) and two outgroup (one newly generated here) non-Rubiaceae Gentianales taxa. Included were data from the 18S rDNA gene (1814 bp), 5.8S rDNA gene (156 bp), 28S rDNA gene (3427 bp), ITS1 rDNA spacer (331 bp; 207 excluded), and the ITS2 rDNA spacer (317 bp; 153 excluded). The complete partitioned and annotated rDNA data set is available in Nexus file format from Dryad Digital Repository (S4 File; <https://doi.org/10.5061/dryad.q573n5tf4>).

Results from the phylogenetic analyses of the rDNA gene data set, only including the 18S, 5.8S, and 28S rDNA gene regions, are reported as a phylogram (from the Bayesian analyses) in Fig. 4. In the subfamily Rubioideae relationships are generally resolved and well supported. Coltoecemateae, Ophiorrhizeae, and Urophyllae form a monophyletic sister group (BPP = 1.00; BS = 89%) to remaining tribes (BPP = 0.98; BS = 93%) followed by Lasiantheae. Coussareae are resolved sister (BPP = 0.98; BS = 76%) to the Spermaceae alliance (BPP = 1.00; BS = 82%), Anthospermeae are sister to Knoxiaceae plus Spermaceae (BPP = 1.00; BS = 70%), and together they form a monophyletic sister group to a clade comprising remaining tribes (BPP = 0.95; BS < 50%) in the Spermaceae alliance. Among the remaining tribes Dunnieae followed by Danaideae, Argostemmateae, and Paederieae are resolved as successive sister groups to a clade where Rubieae are sister to Theligoneae plus Putorieae (BPP = 0.93; BS = 92%), but the Paederieae (BPP = 0.93; BS = 73%) and Danaideae (BPP = 0.89; BS < 50%) positions are poorly supported. In the Psychotriaceae alliance (BPP = 0.99; BS = 83%) Morindeae together with Mitchellaeae (BPP = 0.98; BS = 76%) are resolved sister to Prismatomerideae (BPP = 1.00; BS = 93%), Psychotriaceae together with Palicoureeae are resolved sister to Gaertnereae, and together the six tribes are resolved as monophyletic with

**Fig. 2.** Phylogram resulting from the MrBayes nucleotide based analysis of the plastid CDS data set. Nodes indicated by a black dot are well supported and have a Bayesian posterior probability (BPP) of 0.95 or more (Alfaro et al., 2003; Erixon et al., 2003). Maximum Likelihood non-parametric bootstrap support (BS) is also indicated for each node in the tree. Support values are 1.00 (BPP) and 100% (BS) unless indicated in the figure. Subfamilies Rubioideae, Cinchonoideae, and Ixoroideae, and the Psychotriaceae, Spermaceae (Rubioideae), and Vanguerieae, Coffeae (Ixoroideae) alliances as traditionally recognized are indicated on the tree. Taxa in blue show conflicting relationships in different analyses of data from the same (plastid) genomic compartment (see Fig. 3). Taxa in red show conflicting relationships in analyses of data from different genomic compartments (see Figs. 4, S1).

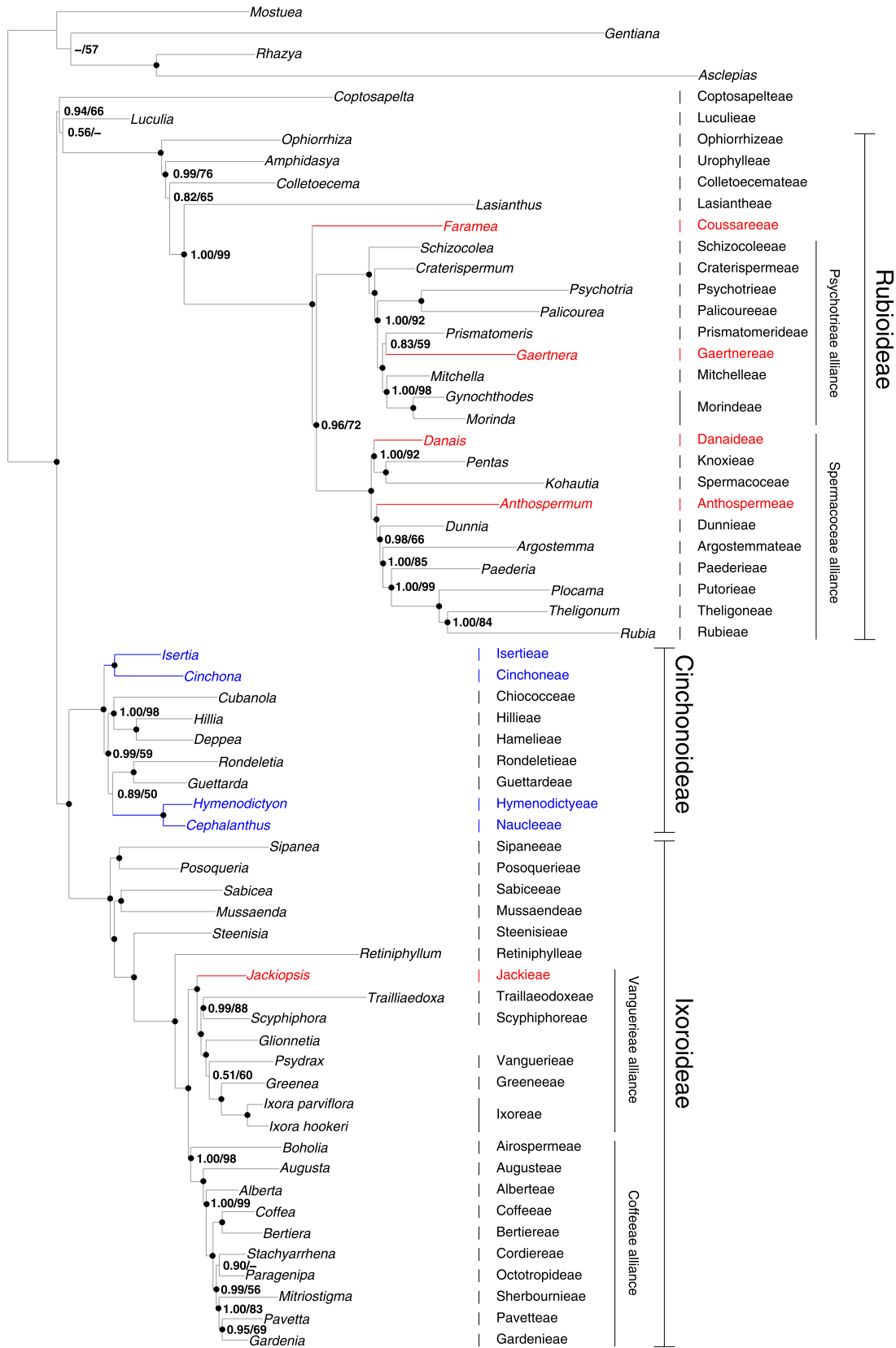


Fig. 3. Continued

Craterispermeae (unsupported with BPP = 0.57; BS < 50%) and Schizocoleae as successive sister groups.

In subfamilies Cinchonoideae and Ixoroideae we see a highly deviating pattern. Here relationships are mainly unresolved or poorly supported, and neither the two subfamilies, nor the two informal Ixoroideae alliances are retrieved by the analyses (Fig. 4). Hillieae are sister to Hamelieae (BPP = 1.00; BS = 78%), and Naucleae are sister to Hymenodictyeae (BPP = 1.00; BS = 98%), but remaining relationships of the Cinchonoideae tribes are unresolved or poorly supported. Among Ixoroideae, the tribe Ixoreae is sister to Greeneae (BPP = 1.00; BS = 81%), with *Glionnetia* as their sister group (BPP = 1.00; BS = 65%). Octotropideae are sister to Sherbournieae (BPP = 1.00; BS = 81%) and together with Gardenieae and Pavetteae they form a monophyletic group (BPP = 1.00; BS = 80%). Alberteae, Jackieae, and Airospermeae are also supported as a monophyletic group (BPP = 0.99; BS = 69%). Coptosapelteae are grouped with Luculieae and together in an unresolved position at the base of the family, but the grouping of Coptosapelteae and Luculieae is not supported by the ML bootstrap (BS < 50%).

Results from the analyses of rDNA data including also the variable ITS1 and ITS2 spacer regions are reported as supporting information (Fig. S2). Compared to the analyses not including the ITS regions there are only minor differences, and the general pattern with most relationships in subfamily Rubioideae resolved and well supported, and with almost no relationships resolved and well supported in subfamilies Cinchonoideae and Ixoroideae remain. Support and resolution are sometimes lower than those of the analysis without nrITS, in particular within the subfamily Rubioideae. The opposite can, however, also be observed for some nodes; there is for example, increased support for grouping Coussareae with the Spermaceae alliance, the Psychotrieae alliance, and the sister relationships between Pavetteae and Gardenieae, and between Isertieae and Cinchoneae.

## 4 Discussion

It is clear that simply adopting a “more characters” phylogenomic approach does not resolve the phylogenetic relationships of Rubiaceae. The plastid based analyses presented here use 69 828 bp from 77 protein-coding genes, and an additional 25 666 bp of nonprotein coding data, yet some of the relationships that have been difficult to resolve in previous analyses remain problematic and appear to be unaffected by the massive addition of new data contributed in the present analyses. The tribes Luculieae and

Coptosapelteae, for example, are still placed in poorly supported positions at the base of the family, although the analyses of combined CDS and non-CDS data support a sister group relationship between Luculieae and subfamily Rubioideae (BPP = 0.98; BS = 89%). Also the relationships of the early diverging tribes Urophyllae and Colletocemateae in subfamily Rubioideae remain difficult to resolve (Figs. 2, 3, S1).

It is also clear that there are conflicting phylogenetic signals in the data analyzed. Some of those conflicts come from the plastid data alone where different analyses of the same protein coding data, treated either as nucleotides (Fig. 2) or as codon degenerated data (Fig. 3), resolve and support conflicting topologies, in particular in the subfamily Cinchonoideae. As these analyses use the same data the conflicting signals cannot be “real” in a biological sense, but are likely resulting from erroneous assumptions in the models used to reconstruct our phylogenies (Philippe et al., 2011; Cooper, 2014).

Conflicting signals of a different kind are seen in the analyses of plastid data on the one hand, and of nuclear rDNA data on the other. The analyses of plastid data behave as one would expect, with results that are more or less congruent with the coherent picture of phylogenetic relationships in the family built during the last 20–30 years. The analyses of nuclear rDNA data, however, yield less congruent results. In the subfamily Rubioideae the analyses both resolve and support a large number of relationships (Figs. 4, S2), many also congruent with those indicated by the analyses of plastid data, but other relationships are not only well supported, but also clearly in conflict with the relationships supported by the analyses of plastid data. In the subfamilies Cinchonoideae and Ixoroideae we see a very different general pattern in the results of the phylogenetic analyses. Here the analyses of nuclear rDNA data almost completely fail to resolve and support any phylogenetic relationships, yet also here there are clear indications of conflicting signals in the analyses of plastid versus nuclear data. Plants typically have a maternal inheritance of the plastids (Corriveau & Coleman, 1988; Birky, 1995, 2001), whereas nuclear rDNA have a biparental inheritance, although it may appear uniparental as a result of concerted evolution and differential elimination of one of the parental rDNA (Volkov et al., 2007). Hence, these conflicting signals possibly represent “real” conflicts, in the sense that the different relationships actually reflect alternative evolutionary histories of the two data types.

### 4.1 Conflicting signals in data of the same genomic compartment within Cinchonoideae

In the subfamily Cinchonoideae our analyses of the same protein-coding gene sequence data treated either as nucleotides (Fig. 2), or as codon degenerated data, where nucleotide ambiguity codes are used to eliminate the

**Fig. 3.** Phylogram resulting from the MrBayes degenerate codon analysis of the CDS data set using nucleotide ambiguity codes to eliminate the distinction between synonymous codons. Nodes indicated by a black dot are well supported and have a Bayesian posterior probability (BPP) of 0.95 or more (Alfaro et al., 2003; Erixon et al., 2003). Maximum Likelihood non-parametric bootstrap support (BS) is also indicated for each node in the tree. Support values are 1.00 (BPP) and 100% (BS) unless indicated in the figure. Subfamilies Rubioideae, Cinchonoideae, and Ixoroideae, and the Psychotrieae, Spermaceae (Rubioideae), and Vanguerieae, Coffeae (Ixoroideae) alliances as traditionally recognized are indicated on the tree. Taxa in blue show conflicting relationships in different analyses of data from the same (plastid) genomic compartment (see Fig. 2). Taxa in red show conflicting relationships in analyses of data from different genomic compartments (see Figs. 4, S1).

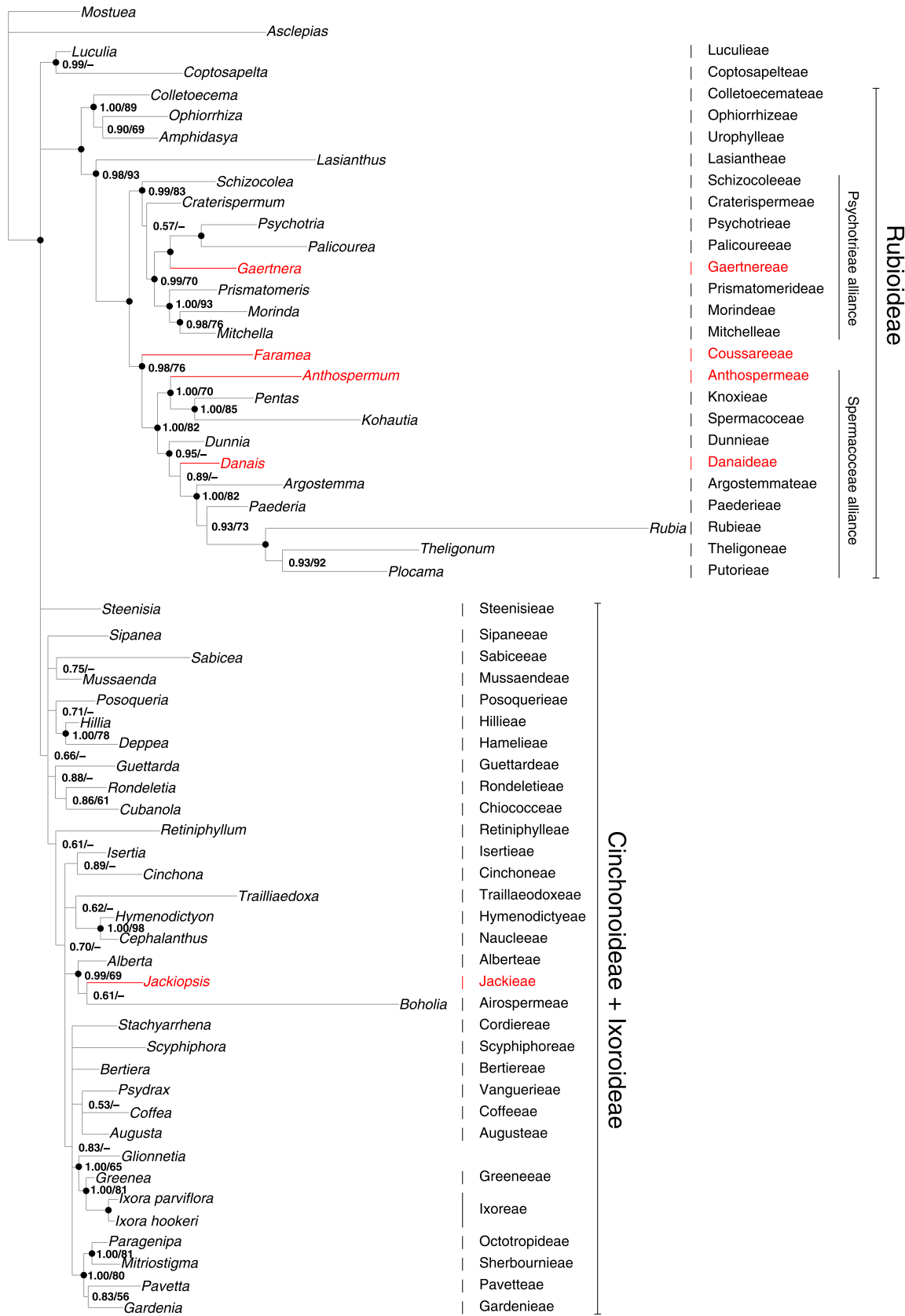


Fig. 4. Continued

distinction between synonymous codons (Fig. 3), provide unequivocal support for conflicting relationships. The subfamily has a primary distribution in the Neotropics but two of the tribes, the Naucleaeae and the Hymenodictyeae, have predominantly paleotropical distributions (Razafiman-dimbison & Bremer, 2001; Antonelli et al., 2009; Manns & Bremer, 2010). Our nucleotide based analyses of the CDS data set (Fig. 2) support that the paleotropical tribes Hymenodictyeae and Naucleaeae are sisters to remaining tribes, whereas the codon degenerated version of the same data indicate corresponding support for a conflicting topology with the neotropical tribes Cinchoneae and Isertieae in that same position (Fig. 3). As both types of data are derived from the same set of observations, they must have evolved under the same species tree and the supported conflict must result from erroneous assumptions in the models used to reconstruct our phylogenies. Both topologies have, however, been obtained and reported in previous studies (Andersson & Antonelli, 2005; Robbrecht & Manen, 2006; Antonelli et al., 2009; Manns & Bremer, 2010; Manns et al., 2012; Wikström et al., 2015; Yang et al., 2016), and the alternative positions for the paleotropical tribes have contributed to widely different and conflicting ideas about the geographic origins, not only of the subfamily Cinchonoideae, but of Rubiaceae as a whole. Antonelli et al. (2009) resolved the paleotropical Hymenodictyeae and Naucleaeae as sisters to remaining tribes with the neotropical tribes Isertieae and Cinchoneae nested well inside a clade of mainly Central American and Antillean tribes. They considered the pattern to corroborate a boreotropical origin of the family, and suggested Rubiaceae as a whole to have originated in the Paleotropics, and to have dispersed into North America via the North Atlantic Land Bridge (NALB) in the Late Paleocene (Antonelli et al., 2009). The neotropical tribes Isertieae and Cinchoneae were suggested to have dispersed into South America more recently, sometime in the Early Eocene (Antonelli et al., 2009). Rydin et al. (2009a) reported an alternative topology with an initial split in subfamily Cinchonoideae separating the neotropical tribes Rondeletieae and Guettardeae, but most recent studies have reported a topology with the neotropical Cinchoneae and Isertieae as sister to remaining tribes (Manns & Bremer, 2010; Manns et al., 2012; Wikström et al., 2015). The alternative positions for the paleotropical tribes have, however, remained poorly supported in most of these analyses, but following ancestral-area and divergence-time reconstructions Manns et al. (2012) rejected entirely the hypothesis of Rubiaceae having used the NALB to reach the Neotropics in the Late Paleocene. They suggested a reversed scenario with the ancestor of both subfamilies Cinchonoideae and Ixoroideae present in South America already during the Late Cretaceous, and with the ancestor of Hymenodictyeae

and Naucleaeae having dispersed out of America, from the Neotropics to the Paleotropics, using the NALB in the Late Paleocene or Eocene (Manns et al., 2012). We consider this out of America scenario as best supported by the analyses presented here with paleotropical tribes Hymenodictyeae and Naucleaeae nested well inside the subfamily Cinchonoideae (Fig. 3).

The conflicting positions of the paleotropical Hymenodictyeae and Naucleaeae in the nucleotide based analyses, as sisters to remaining tribes (Fig. 2), appear to be caused by a composition bias resulting from synonymous substitutions. Synonymous substitutions primarily result from changes in the third codon position, whereas substitutions in first and second codon positions are mostly nonsynonymous, except for codon variants of Leucine, Arginine, and Serine. Before the era of phylogenomics, analyses of one or a few genes often found “greater phylogenetic signal” in rapidly evolving third codon positions relative to their corresponding first and second codon positions (Källersjö et al., 1998; Björklund, 1999; Sennblad & Bremer, 2000; Simmons et al., 2002, 2006). However, recent analyses of both plants (Zhong et al., 2011; Cox et al., 2014; Li et al., 2014) and animals (van den Bussche et al., 1998; Tarrío et al., 2001; Philippe et al., 2011; Regier & Zwick, 2011; Breinholt & Kawahara, 2013; Morgan et al., 2013) clearly indicate that composition and dependent codon-usage biases, resulting from synonymous substitutions, are potential causes behind conflicting and erroneous phylogenetic signals.

#### 4.2 Conflicting signals in data from different genomic compartments

Our analyses of plastid data on the one hand, and nuclear rDNA data on the other reveal significant conflicts in their phylogenetic signals. Relationships, well supported by the analyses of plastid data, are clearly in conflict with relationships equally well supported in the analyses of nuclear rDNA data, and these conflicting signals we interpret to represent “real” conflicts, in the sense that the different relationships actually reflect alternative evolutionary histories of the two data types. In the subfamily Rubioideae there are several such conflicts: the close relationship between the tribe Coussareeae and tribes traditionally united in the Spermaceae alliance; the sister group relationship between the tribe Anthospermeae and the Knoxiaceae–Spermaceae clade; the relationships of Danaideae to the Argostemmatae–Putorieae clade, not close to the Knoxiaceae–Spermaceae clade; and the sister group relationship between Gaertnereae and the Psychorieae–Palicourieae clade. These relationships are supported in the analyses of nuclear rDNA data (Figs. 4, S2), but have to our knowledge never been supported by any previous analyses. Two of the conflicts concern tribes in the Spermaceae alliance, and resolving the relationships in this group have been problematic

**Fig. 4.** Phylogram resulting from the MrBayes analyses of nuclear rDNA gene data. Nodes indicated by a black dot are well supported and have a Bayesian posterior probability (BPP) of 0.95 or more (Alfaro et al., 2003; Erixon et al., 2003). Maximum Likelihood non-parametric bootstrap support (BS) is also indicated for each node in the tree. Support values are 1.00 (BPP) and 100% (BS) unless indicated in the figure and nodes with BPP <0.50 are collapsed. Subfamilies Rubioideae, and Cinchonoideae plus Ixoroideae as traditionally recognized are indicated on the tree. Taxa in red show conflicting relationships in analyses of data from different genomic compartments (see Figs. 2, 3).

with relationships either unresolved or poorly supported in previous analyses (Robbrecht & Manen, 2006; Rydin et al., 2009b; Wikström et al., 2015). From a data and taxon sample point of view Rydin et al. (2009b) is the most comprehensive analysis of the group, and they considered conflicting signals in different loci to cause some of the problems they encountered in their analyses. They used five different plastid loci in their analyses, but they also included data from the nuclear ITS rDNA spacer regions. The limited amount of non-plastid characters did, however, prevent them from further evaluating conflicting signals between plastid and nuclear data. Our results clearly indicate that there are such conflicting signals, not only in the Spermaceae alliance but also in the Psychotrieae alliance, and instead of combining data from the plastid genome, that has an almost strict maternal inheritance pattern (Corriveau & Coleman, 1988; Birky, 1995, 2001), with biparentally inherited data from the nucleus, we need to start teasing apart the different signals and investigate the underlying causes behind their conflicts.

Recent work by Rydin et al. (2017) documented conflicting phylogenetic signals very much similar to those seen here, but in their case the conflicts were between plastid data on the one hand, and mitochondrial data on the other. Their analyses of mitochondrial data rejected monophyly of two out of three subfamilies of Rubiaceae and placed the quinine-tribe Cinchoneae and the Isertieae from subfamily Cinchonoideae well inside subfamily Ixoroideae. In the subfamily Ixoroideae they also found strong support for grouping the tribe Jackieae, traditionally placed in the Vanguerieae alliance, with Airospermeae from the Coffeaeae alliance, and they considered ancient hybridization events between early occurring members of the subfamilies Cinchonoideae and Ixoroideae and of the Vanguerieae and Coffeaeae alliances as possible reasons behind the conflicting signals (Rydin et al., 2017). Although the analyses of nuclear rDNA data presented here fail to resolve almost any relationships in the subfamilies Cinchonoideae and Ixoroideae, they do support one of the conflicting relationships seen in the mitochondrial analyses, the close relationship between the tribes Jackieae and Airospermeae. Together they form a well supported group with the tribe Alberteae (Figs. 4, S2). *Boholia* (Airospermeae) deviate from most other taxa by having a very long branch in the phylograms, and it is possible that the close relationship to Jackieae is caused by long branch attraction artifacts (Felsenstein, 1978). However, a close relationship between Airospermeae and Alberteae is not in conflict with the analyses using plastid data (Figs. 2, 3), the conflict concerns their close relationship to Jackieae, and neither *Alberta* (Alberteae), nor *Jackiopsis* (Jackieae) are associated with long branches in the rDNA phylograms (Figs. 4, S2).

#### 4.3 Usefulness of nuclear DNA

The almost complete failure to resolve the relationships in the subfamilies Cinchonoideae and Ixoroideae using nuclear rDNA data is surprising (Figs. 4, S2). One possible explanation could be that the 18S, 5.8S, and 26S nuclear rDNA gene regions analyzed are not variable enough for resolving the relationships in these groups. A simple evaluation of individual phylograms from the posterior distribution of trees of our Bayesian analyses show that the rDNA gene

regions, without doubt, are less variable in these groups than in subfamily Rubioideae. However, adding data from the highly variable ITS spacer regions to our analyses have no significant effect on the resolution in the Cinchonoideae and Ixoroideae (Fig. S2), which indicate that lack of information is inadequate as explanation for the poorly resolved results based on nuclear data. If anything, the ITS regions appear variable to the extent that they cannot be reliably aligned across the family. Two previous studies (Rydin et al., 2009a; Manns & Bremer, 2010) in fact report conflicting topologies likely caused by their respective alignment of the ITS regions. While data from the 18S, 5.8S, and 26S rDNA gene regions thus may be too conserved to be useful for resolving the relationships in the Cinchonoideae and Ixoroideae, the ITS spacer regions appear too variable.

To expand phylogenetic investigations to also allow for detection of potential anastomosing evolutionary patterns caused by for example hybridization will require other approaches than basing studies of yet larger sets of data. Although it has long been known that inheritance of the organellar genomes is not always strictly maternal (see e.g., Rydin et al., 2017 and references therein), the true inheritance patterns for individual taxa of interest have often been left unknown or unconsidered in studies based on organellar DNA. Further, data from biparentally inherited DNA will often be needed in order to dig deeper into “difficult” phylogenetic questions, for example involving taxa of hybrid and/or allopolyploid origin, and nuclear data is therefore an important source of information. As in many phylogenetic studies, we selected the ribosomal DNA for our nuclear data set, DNA regions that are well documented and conveniently present in many copies in the genome, thus easy to sequence. The ribosomal DNA is, however, not useful for addressing questions involving reticulate evolutionary patterns because the many existing copies undergo a process of homogenization, or concerted evolution (Elder & Turner, 1995), that with time will mask the original genetic information in taxa of hybrid origin. It will therefore be necessary to utilize other parts of the nuclear DNA in order to further explore the nature and causes of the conflicting patterns we observe.

#### 4.4 More on Rubiaceae relationships

4.4.1 The unplaced tribes Luculieae and Coptosapelteae  
Two South East Asian tribes, the Luculieae and the Coptosapelteae, have been notoriously difficult to associate with any other taxa (Robbrecht & Manen, 2006; Bremer, 2009; Rydin et al., 2009a). Both tribes have commonly been placed in an unresolved “basal” position outside of the three recognized subfamilies (Fig. 1), sometimes together as sister groups (Bremer, 2009; Rydin et al., 2009a; Manns et al., 2012), but more commonly not (Bremer et al., 1999; Robbrecht & Manen, 2006; Bremer & Eriksson, 2009; Manns et al., 2012; Wikström et al., 2015; Yang et al., 2016). Robbrecht & Manen (2006) considered their analyses of nuclear versus plastid data to show conflicting results. Their nuclear data placed the tribes together, inside Rubiaceae and as sister to a clade comprising subfamilies Cinchonoideae and Ixoroideae. In contrast, their analyses of plastid data placed them separately, outside of Rubiaceae and sisters to the entire family, or separately with Coptosapelteae outside and sister to the entire family and

Luculieae as sister to subfamily Rubioideae (Robbrecht & Manen, 2006). Rydin et al. (2009a) also included nuclear data in their analyses of deep divergences in the family, in their case data from the ITS rDNA spacer regions, and they reported similar patterns. Their combined analyses of plastid and nuclear ITS data resolved Luculieae and Coptosapelteae together, and in an unresolved “basal” position in the family, whereas their analyses excluding the nuclear ITS data resolved the two tribes separately (Rydin et al., 2009a). Again, their limited amount of non-plastid data prevented them to further evaluate this potential conflict. One clear indication as to the relationships of at least one of these tribes came recently from an analyses of mitochondrial data by Rydin et al. (2017). The relationships of Luculieae remained poorly supported in their analyses, but Coptosapelteae was unequivocally supported as sister to a clade comprising the two subfamilies Cinchonoideae and Ixoroideae. There were, however, several supported conflicts in their analyses of mitochondrial data compared to corresponding analyses of plastid data, specifically with respect to tribal relationships in the subfamilies Cinchonoideae and Ixoroideae. As argued by Rydin et al. (2017), these conflicts potentially result from one, or several, ancient hybridization events between early occurring members of the two subfamilies and, if correct, could make strict comparisons with the plastid-based results presented here invalid.

Despite the substantial increase in the amount of data analyzed, uncertainties surrounding the relationships of these two tribes remain also in the analyses presented here. Both the nucleotide based (Fig. 2), and the degenerated codon (Fig. 3) analyses of plastid CDS data place the tribes separately, and in poorly supported “basal” positions together with the subfamily Rubioideae. Nucleotide analyses of combined CDS and non-CDS data (Fig. S1) also have poor support for placing Coptosapelteae, but resolve Luculieae as sister to subfamily Rubioideae with good support (BPP = 0.98; BS = 89%). Although consistent with most previous analyses of plastid data (Bremer et al., 1999; Robbrecht & Manen, 2006; Bremer & Eriksson, 2009; Manns et al., 2012; Wikström et al., 2015), the continued lack of support for their relationships in most analyses is problematic. Also the nuclear rDNA analyses produce results consistent with those from previous analyses including nuclear data, with Luculieae and Coptosapelteae placed together in a monophyletic group, and in an unresolved “basal” position in the family (Robbrecht & Manen, 2006; Rydin et al., 2009a). The analyses that only include the conserved and unambiguously aligned rDNA gene regions provide unambiguous support for grouping the two tribes in the Bayesian analyses (Fig. 4), but the grouping is not supported by the ML bootstrap, and the Bayesian support is significantly reduced when also the variable and less easily aligned ITS spacer regions are included in the analyses (Fig. S2). This last result is somewhat surprising as the ITS regions were the only nuclear regions included in the analyses by Rydin et al. (2009a), and the ones supporting their grouping of Luculieae and Coptosapelteae.

An alternative to adopting a “more characters” phylogenomic approach for improving our phylogenetic inferences, is to expand on the taxon sample (Heath et al., 2008), as done by Rydin et al. (2009a) in their analyses of deep divergences in the

family. Luculieae comprise a single genus with no more than four species, and adding further species to our analyses would likely be of limited value. Coptosapelteae, on the other hand, comprise 2 genera and altogether 56 recognized species, and *Acranthera* Arnott, the genus not represented in our analyses, was only recently included in the tribe (Rydin et al., 2009a). Morphologically, the genus also deviates significantly from *Coptosapelta*, and expanding the taxon sample to also include *Acranthera* should be done in future attempts at resolving the relationships of this problematic group.

#### 4.4.2 Subfamily Rubioideae

The early diverging tribes in the subfamily Rubioideae, the Ophiorrhizeae, Urophyllaeae, and Colletocemateae, also remain problematic. Earlier analyses of plastid data have sometimes placed and supported Colletocemateae as sister to remaining tribes in the subfamily (Robbrecht & Manen, 2006; Rydin et al., 2008, 2009a), but there is no support for this early split in the subfamily when using much larger sets of genomic data, neither based on plastid and nuclear data (the present study), nor based on mitochondrial data (Rydin et al., 2017). Relaxed-clock analyses of plastid data have instead resolved the Colletocemateae as sister to the Lasiantheae (Manns et al., 2012; Wikström et al., 2015), but also this relationship is rejected here (Figs. 2–4). Here, the degenerate codon analyses of plastid CDS data, as well as the nucleotide based analyses of combined CDS and non-CDS data (Fig. S1), support Ophiorrhizeae as sister to remaining tribes of the subfamily, but the positions of Urophyllaeae and Colletocemateae remain unsupported. In contrast to our plastid-based analyses, our analyses of nuclear rDNA data resolve and support the three tribes as a monophyletic group, sister to remaining tribes in the subfamily (Figs. 4, S2). However, as no support is obtained for the relationships of the tribes Urophyllaeae and Colletocemateae in the analyses of plastid data, the different relationships indicated for the three tribes cannot be unambiguously attributed to conflicting phylogenetic signals in the two data types. Rydin et al. (2017), in their analyses of mitochondrial data, found strong support for a third alternative with the Colletocemateae as sister to the Urophyllaeae, and with this Colletocemateae–Urophyllaeae clade sister to remaining tribes of the subfamily. The grouping of the three tribes in a monophyletic group, as supported by our analyses of nuclear rDNA data (Figs. 4, S2), clearly conflicts with this mitochondrial based result reported by Rydin et al. (2017).

In the Spermaceae alliance the Danaideae were for some time considered problematic. Some analyses placed the Danaideae sister to remaining tribes of the Spermaceae alliance (Bremer & Manen, 2000; Bremer, 2009; Bremer & Eriksson, 2009; Rydin et al., 2009b; Yang et al., 2016), or sister to a clade comprising the Anthospermeae, Argostemmataeae, Paederieae, Rubieae, and the Theligoneae (Robbrecht & Manen, 2006), but more recent analyses of plastid data and mitochondrial data respectively, have more or less firmly placed the tribe as sister to a Knoxiaceae–Spermaceae clade (Rydin et al., 2009a; Krüger et al., 2012; Krüger, 2014; Wikström et al., 2015; Rydin et al., 2017). Here however, the placement of Danaideae is once again challenged since our analysis of nuclear data support Danaideae as member of a clade also comprising Dunnieae, Argostemmataeae, Paederieae, Rubieae, Theligoneae

and Putorieae. Among these remaining tribes of the alliance the Rubieae, Theligoneae, and the Putorieae have consistently formed a monophyletic group (Robbrecht & Manen, 2006; Backlund et al., 2007; Rydin et al., 2009b; Wikström et al., 2015; Yang et al., 2016; Rydin et al., 2017), and together they have commonly been placed in a sister group relationships to the Paederieae (Robbrecht & Manen, 2006; Backlund et al., 2007; Rydin et al., 2009a, 2009b; Yang et al., 2016). All these relationships in the Spermaceae alliance are corroborated and well supported also in the present analyses of plastid data (Figs. 2, 3, S1). The relationships of the Anthospermeae, Argostemmatae, and Dunnieae have been more problematic. Rydin et al. (2009b) specifically addressed tribal relationships in the alliance and included a comprehensive sample of taxa from all tribes in their analyses. They resolved the Anthospermeae, Argostemmatae, and Dunnieae as successive sister groups to the Rubieae–Theligoneae–Putorieae–Paederieae clade (Rydin et al., 2009b). Their position of the Anthospermeae as the first diverging group was well supported, and this position has been supported also in other analyses (Rydin et al., 2009a; Wikström et al., 2015). The relative positions of the Argostemmatae and Dunnieae in their analyses were, however, unsupported (Rydin et al., 2009b). Our analyses of plastid data support the Anthospermeae, Dunnieae, and Argostemmatae as successive sister groups to the Rubieae–Theligoneae–Putorieae–Paederieae clade (Figs. 2, 3, S1), and these relationships are entirely consistent with those supported by Rydin et al. (2009b).

Among conflicting tribal relationships in the Spermaceae alliance supported in our analyses of nuclear rDNA data, the position of Anthospermeae as sister to the Knoxiaceae–Spermaceae clade is perhaps the most apparent and striking example. The tribe Anthospermeae comprise 12 genera with around 200 species (Puff, 1982; Thureborn et al., 2019), and although previous analyses of a single or a few genes from the plastid genome have been unable to find support for their precise relationships to other tribes in the alliance, a sister group relationship to the Knoxiaceae–Spermaceae clade have never before been reported. A second example is the position of the Coussareeae, a morphologically heterogeneous group of Neotropical plants with 10 different genera and a possibly 400 species (Andersson & Rova, 1999; Bremer & Manen, 2000; Löfstrand et al., 2019). Previous analyses of plastid data (Bremer & Manen, 2000; Rydin et al., 2008, 2009a, 2009b; Wikström et al., 2015), the analyses of plastid data presented here (Figs. 2, 3, S1), as well as analyses of mitochondrial data (Rydin et al., 2017) all support a position of the Coussareeae as sister to a clade comprising the Spermaceae and the Psychotrieae alliances. In contrast however, the analyses of nuclear rDNA data presented here support a position of the Coussareeae nested inside the Spermaceae alliance (Figs. 4, S2).

In the Psychotrieae alliance our analyses of plastid data more or less corroborate patterns reported in previous analyses of the group. One exception is Prismaticerideae, here resolved sister to the Gaertnereae. This sister relationship is, however, only supported in our nucleotide based analysis of plastid CDS and combined CDS and non-CDS data (Figs. 2, S1), not in the degenerate codon analyses of CDS data, where silent substitutions are ignored (Fig. 3), and

it has not been seen in previous analyses of the group (Robbrecht & Manen, 2006; Razafimandimbison et al., 2008, 2017; Wikström et al., 2015). Based on our analysis of nuclear data, previously reported relationships in the Psychotrieae alliance are also corroborated with one exception, the position of the tribe Gaertnereae, which is supported as sister to the Psychotrieae–Palicoureeae clade in the analysis of rDNA data (Figs. 4, S2), but more closely related to Mitchelleae, Morindeae, and Prismaticerideae based on plastid data (Figs. 2, 3, S1). Mitochondrial data support the sister group relationship of Gaertnereae to a Psychotrieae–Palicoureeae clade (Rydin et al., 2017), and yield entirely congruent results with nuclear ribosomal data (the present study) in the Psychotrieae alliance.

Notwithstanding these conflicting signals in plastid and nuclear data, in subfamily Rubioideae the analyses of nuclear rDNA data also show considerable congruence to the results of the plastid-based analyses, and thus to the coherent picture of phylogenetic relationships in the family developed during the last 20–30 years. The Rubieae, Theligoneae, and Putorieae, for example, are resolved as monophyletic, which is consistent with most previous analyses (see Bremer, 2009; Rydin et al., 2009b). This Rubieae–Theligoneae–Putorieae clade are grouped with the Paederieae, Argostemmatae, and Dunnieae (Rydin et al., 2009b; Wikström et al., 2015), although the nuclear rDNA analyses also include the Danaideae in this group. The Spermaceae alliance are supported as monophyletic (Andersson & Rova, 1999; Bremer & Manen, 2000; Robbrecht & Manen, 2006; Rydin et al., 2009b; Wikström et al., 2015), even though the nuclear rDNA data also place the Coussareeae in this group. The Knoxiaceae and Spermaceae are supported as sister groups (Bremer, 1996; Robbrecht & Manen, 2006; Kårehed & Bremer, 2007; Rydin et al., 2009b; Wikström et al., 2015). The Psychotrieae alliance are supported as monophyletic (Bremer & Manen, 2000; Robbrecht & Manen, 2006; Rydin et al., 2008; Razafimandimbison et al., 2008, 2017), and the sister group relationships of the Morindeae and Mitchelleae, and the Psychotrieae and Palicoureeae are both supported (Razafimandimbison et al., 2008; Rydin et al., 2008; Wikström et al., 2015).

#### 4.4.3 Subfamilies Cinchonoideae and Ixoroideae

The “more characters” approach adopted here also provides some additional resolution to relationships that have been problematic in subfamilies Cinchonoideae and Ixoroideae, at least with respect to the relationships of the maternally inherited plastid. Disregarding the conflicting signal in subfamily Cinchonoideae, caused by synonymous substitutions (see discussion above), relationships among the four previously identified and well supported groups in the Cinchonoideae are with one exception supported in our analyses of plastid data. Still problematic is the position of the Hymenodictyeae–Naucleaeae clade, here resolved in a poorly supported position as sister to the Rondeletieae–Guettardeae clade (Fig. 3). This position as sister to the Rondeletieae–Guettardeae clade is, however, unambiguously supported by analyses of mitochondrial data (Rydin et al., 2017).

In the subfamily Ixoroideae our plastid data analyses mainly corroborate relationships obtained in recent



analyses of the group where the alternative “more taxa” approach was used (Kainulainen et al., 2013, 2017; Mouly et al., 2014). Relationships among the early diverging Condamineae, Mussaendeae–Sabiceeae, and Henriquezieae–Posoquerieae–Sipaneeae clades have been difficult to support (Kainulainen et al., 2013; Wikström et al., 2015), but the main problem has been the position of the tribe Condamineae, which we were unable to include in the present analyses. In the Vanguerieae alliance the small tribes Scyphiphoreae and Traillaedoxeae, and the genus *Glionnetia* from the Seychelles, currently unclassified at tribal level, have been problematic (Razafimandimbison et al., 2011; Kainulainen et al., 2013; Wikström et al., 2015). *Glionnetia* have been supported in a group also including the tribes Vanguerieae, Greeneae, Ixoreae, and Aleisanthiae (Kainulainen et al., 2013; Wikström et al., 2015), but the relationships between *Glionnetia*, the Vanguerieae, and a Greeneae–Ixoreae–Aleisanthiae clade have been unsupported. The tribes Scyphiphoreae and Traillaedoxeae have commonly been placed in an unresolved sister group positions to this group of four tribes and *Glionnetia* (Razafimandimbison et al., 2011; Kainulainen et al., 2013; Wikström et al., 2015). Our analyses of plastid data provide some support for the relationships of *Glionnetia*. In the nucleotide based and degenerate codon analyses of CDS data *Glionnetia* remain resolved in a corresponding polytomy as in previous analyses, together with the Vanguerieae and a Greeneae–Ixoreae clade, but a sister group relationship to these tribes obtains support in our nucleotide-based analyses of combined CDS and non-CDS data (Fig. S1). In contrast, there is good support for the relationships of the Scyphiphoreae and Traillaedoxeae, which are here supported as sister groups to each other, and together supported as sister to the *Glionnetia*–Vanguerieae–Greeneae–Ixoreae clade (Fig. 3). Also in the Coffeae alliance we see only minor differences in the results from our analyses of plastid data compared to analyses that have adopted a “more taxa” approach (Kainulainen et al., 2013, 2017; Mouly et al., 2014). They all supported monophyly of a group comprising four tribes, the Gardenieae, Pavetteae, Sherbournieae, and Cordiereae, and with Octotropideae sister to this group. Here, the tribe Octotropideae is instead resolved sister to the Cordiereae and they together are sister to a Gardenieae–Pavetteae–Sherbournieae clade (Figs. 2,3). Octotropideae as sister to Cordiereae is, however, not supported by the nucleotide analyses of combined CDS and non-CDS data (Fig. S1).

## 5 Conclusions

It is important to note that most phylogenetic results in the large coffee family are consistently retrieved, regardless of kind of data and analytical approach. However, our study here together with our analyses of mitochondrial data (Rydin et al., 2017) reveals that there are (not so few) clades that are well supported by data from one or two of the genomic compartments, but not from all three, and some of these conflicts may be “real”, reflecting biological processes of the past. A striking example concerns the deepest splits in the

family and the previously unresolved position of the small tribe Coptosapelteae, which we show is sister to subfamily Rubioideae plus Luculieae based on plastid data (the present study), sister to Luculieae based on nuclear data (the present study), and sister to Cinchonoideae plus Ixoroideae based on mitochondrial data (Rydin et al., 2017). We also show that phylogenetic questions that have been considered answered are in fact not. For example, the neotropical tribe Coussareeae have been consistently resolved as sister to a large clade comprising the Psychotriaceae and the Spermaceae alliances (see e.g., Löfstrand et al., 2019), a result seemingly supported by data from all three genomic compartments including the mitochondrion (Rydin et al., 2017). Here we show, analyzing the entire cistron, that the nuclear ribosomal DNA with strong support places the Coussareeae in the Spermaceae alliance, thus strongly conflicting with our results based on data from the plastome and the mitochondrion. Adding genomic data for additional taxa may be relevant in order to address some still unresolved parts of the Rubiaceae phylogeny, but for the future it will most of all be important to take the possibility of reticulate evolutionary patterns into consideration. A straightforward way will be to utilize biparentally inherited low-copy genes from the nuclear genome, an approach that should be adopted in future analyses trying to resolve these relationships.

## Acknowledgements

The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. The project was funded by grants from the Royal Swedish Academy of Sciences, Stockholm University, and the Swedish Research Council (A0529602 to B.B. and 2017-03985 to C.R.).

## References

- Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* 20: 255–266.
- Andersson L, Antonelli A. 2005. Phylogeny of the tribe Cinchoneae (Rubiaceae), its position in Cinchonoideae, and description of a new genus, *Ciliosemina*. *Taxon* 54: 17–28.
- Andersson L, Rova JHE. 1999. The *rps16* intron and the phylogeny of the Rubioideae (Rubiaceae). *Plant Systematics and Evolution* 214: 161–186.
- Antonelli A, Nylander JAA, Persson C, Sanmartin I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences USA* 106: 9749–9754.
- Backlund M, Bremer B, Thulin M. 2007. Paraphyly of Paederieae, recognition of Putorieae and expansion of *Plocama* (Rubiaceae–Rubioideae). *Taxon* 56: 315–328.
- Birky CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. *Proceedings of the National Academy of Sciences USA* 92: 11331–11338.

- Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148.
- Björklund M. 1999. Are third positions really that bad? A test using vertebrate cytochrome b. *Cladistics* 15: 191–197.
- Breinholt JW, Kawahara AY. 2013. Phylotranscriptomics: Saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biology and Evolution* 5: 2082–2092.
- Bremer B. 1996. Phylogenetic studies within Rubiaceae and relationships to other families based on molecular data. *Opera Botanica Belgica* 7: 33–50.
- Bremer B. 2009. A review of molecular phylogenetic studies of Rubiaceae. *Annals of the Missouri Botanical Garden* 96: 4–26.
- Bremer B, Andreasen K, Olsson D. 1995. Subfamilial and tribal relationships in the Rubiaceae based on *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 82: 383–397.
- Bremer B, Eriksson T. 2009. Timetree of Rubiaceae: Phylogeny and dating the family, subfamilies and tribes. *International Journal of Plant Sciences* 170: 766–793.
- Bremer B, Jansen RK, Oxelman B, Backlund M, Lantz H, Kim K-J. 1999. More characters or more taxa for a robust phylogeny – Case study from the coffee family (Rubiaceae). *Systematic Biology* 48: 413–435.
- Bremer B, Manen JF. 2000. Phylogeny and classification of the subfamily Rubioideae (Rubiaceae). *Plant Systematics and Evolution* 225: 43–72.
- Chang J-M, Tommaso P, Notredame C. 2014. TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Molecular Biology and Evolution* 31: 1625–1637.
- Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25: 1422–1423.
- Cooper ED. 2014. Overly simplistic substitution models obscure green plant phylogeny. *Trends in Plant Science* 19: 576–582.
- Corriveau JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* 75: 1443–1458.
- Cox CJ, Li B, Foster PG, Embley TM, Civič P. 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Systematic Biology* 63: 272–279.
- Crisuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology* 10: 210.
- Crisuolo A, Gribaldo S. 2011. Large-scale phylogenomic analyses indicate a deep origin of primary plastids within Cyanobacteria. *Molecular Biology and Evolution* 28: 3019–3032.
- Davis AP, Govaerts R, Bridson DM, Ruhsam M, Moat J, Brummit NA. 2009. A global assessment of distribution, diversity, endemism, and taxonomic effort in the Rubiaceae. *Annals of the Missouri Botanical Garden* 96: 68–78.
- Davis CC, Xi Z-X, Mathews S. 2014. Plastid phylogenomics and green plant phylogeny: Almost full circle but not quite there. *BMC Biology* 12: 11.
- Delsuc F, Brinkmann H, Hervé P. 2005. Phylogenomics and the reconstruction of the tree of life. *Nature Review Genetics* 6: 361–375.
- Ding P, Zhang R, Li Q, Qu M. 2015. Direct submission, 21 May 2015.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Doyle JJ. 1991. DNA protocols for plants. In: Hewitt G, Johnson AWB, Young JPW eds. *Molecular techniques in taxonomy*. Berlin: Springer. 283–293.
- Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.
- Elder JF, Turner BJ. 1995. Concerted evolution of repetitive DNA-sequences in eukaryotes. *Quarterly Review of Biology* 70: 297–320.
- Eriksson O, Bremer B. 1991. Fruit characteristics, life forms, and species richness in the plant family Rubiaceae. *American Naturalist* 138: 751–761.
- Erixon P, Sennblad B, Britton T, Oxelman B. 2003. Reliability of bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology* 52: 665–673.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.
- Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.
- Fučíková K, Lewis PO, Lewis LA. 2016. Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution. *Molecular Phylogenetics and Evolution* 98: 176–183.
- Ge X-J, Chiang Y-C, Chou C-H, Chiang T-Y. 2002. Nested clade analysis of *Dunnia sinensis* (Rubiaceae), a monotypic genus from China based on organelle DNA sequences. *Conservation Genetics* 3: 351–362.
- Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7: 457–472.
- Ginter A, Razafimandimbison SG, Bremer B. 2015. Phylogenetic affinities of *Myrioneuron* and *Cyanoneuron*, generic limits of the tribe Argostemmatae and description of a new Asian tribe, *Cyanoneuroneae* (Rubiaceae). *Taxon* 64: 286–298.
- Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27: 221–224.
- Govaerts R, Andersson L, Robbrecht E, Bridson D, Davis A, Schanzer I, Sonke B. 2018. World checklist of Rubiaceae. Facilitated by the Royal Botanic Gardens, Kew [online]. Available from [wcsp.science.kew.org](http://wcsp.science.kew.org) [accessed 7 February 2018].
- Guindon S, Dufayard JF, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307–310.
- Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.
- Kainulainen K, Persson C, Eriksson T, Bremer B. 2010. Molecular systematics and morphological character evolution of the Condamineae (Rubiaceae). *American Journal of Botany* 97: 1961–1981.

- Kainulainen K, Razafimandimbison SG, Bremer B. 2013. Phylogenetic relationships and new tribal delimitations in subfamily Ixoroideae (Rubiaceae). *Botanical Journal of the Linnean Society* 173: 387–406.
- Kainulainen K, Razafimandimbison SG, Wikström N, Bremer B. 2017. Island hopping, long-distance dispersal and species radiation in the Western Indian Ocean: Historical biogeography of the Coffeae alliance (Rubiaceae). *Journal of Biogeography* 44: 1966–1979.
- Källersjö M, Farris JS, Chase MW, Bremer B, Day MF, Humphries CJ, Petersen G, Seberg O, Bremer K. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbcl* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Systematics and Evolution* 213: 259–287.
- Kårehed J, Bremer B. 2007. The systematics of Knoxiaceae (Rubiaceae) – Molecular data and their taxonomic consequences. *Taxon* 56: 1051–1076.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.
- Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kent JW. 2002. BLAT—The BLAST-like alignment tool. *Genome Research* 12: 656–664.
- Krüger Å. 2014. *Phylogenetic position and biogeography of the Madagascar-centered tribe Danaideae (Rubiaceae: Rubioideae)*. Ph.D. Dissertation. Stockholm: Stockholm University.
- Krüger Å, Razafimandimbison SG, Bremer B. 2012. Molecular phylogeny of the tribe Danaideae (Rubiaceae: Rubioideae): Another example of out-of-Madagascar dispersal. *Taxon* 61: 629–636.
- Li B, Lopes JS, Foster PG, Embley TM, Cox CJ. 2014. Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. *Molecular Biology and Evolution* 31: 1697–1709.
- Löfstrand SD, Razafimandimbison SG, Rydin C. 2019. Phylogeny of Coussareeae (Rubioideae, Rubiaceae). *Plant Systematics and Evolution* 305: 293–304.
- Manns U, Bremer B. 2010. Towards a better understanding of intertribal relationships and stable tribal delimitations within Cinchonoideae s.s. (Rubiaceae). *Molecular Phylogenetics and Evolution* 56: 21–39.
- Manns U, Wikström N, Taylor CM, Bremer B. 2012. Historical biogeography of the predominantly neotropical subfamily Cinchonoideae (Rubiaceae): Into or out of America? *International Journal of Plant Sciences* 173: 261–289.
- Morgan CC, Foster PG, Webb AE, Pisani D, McInerney JO, O'Connell MJ. 2013. Heterogenous models place the root of the placental mammal phylogeny. *Molecular Biology and Evolution* 30: 2145–2156.
- Mouly A, Kainulainen K, Person C, Davis AP, Wong KM, Razafimandimbison SG, Bremer B. 2014. Phylogenetic structure and clade circumscriptions in the Gardenieae complex (Rubiaceae). *Taxon* 63: 801–818.
- Ni L-H, Zhao Z-L, Xu H-X, Chen S-L. 2014. Direct submission, 03 April 2014.
- Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302: 205–217.
- Nylander JAA. 2004. MrAIC.pl. Uppsala: Uppsala University. Available from [github.com/nylander/MrAIC](https://github.com/nylander/MrAIC). [Accessed June 30, 2014].
- Park S, Ruhlman TA, Sabir JS, Mutwakil MH, Baeshen MN, Sabir MJ, Baeshen NA, Jansen RK. 2014. Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *BMC Genomics* 15: 405.
- Paudyal SK, Delprete PG, Motley TJ. 2014. Using molecular, morphological, and palynological evidence to transfer *Strumpfia maritima* to the monotypic tribe Strumpfiaceae (Cinchonoideae, Rubiaceae), and a re-delimitation of the tribe Chiococceae. *Systematic Botany* 39: 1197–1203.
- Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology* 9: e1000602.
- Puff C. 1982. The delimitation of the tribe Anthospermeae and its affinities to the Paederieae. *Botanical Journal of the Linnean Society* 84: 355–377.
- Razafimandimbison SG, Bremer B. 2001. Tribal delimitation of Naucleaeae (Rubiaceae): Inference from molecular and morphological data. *Systematics and Geography of Plants* 71: 515–538.
- Razafimandimbison SG, Kainulainen K, Wikström N, Bremer B. 2017. Historical biogeography and phylogeny of the pantropical Psychotrieae alliance (Rubiaceae), with particular emphasis on the Western Indian Ocean region. *American Journal of Botany* 104: 1407–1423.
- Razafimandimbison SG, Kainulainen K, Wong KM, Beaver K, Bremer B. 2011. Molecular support for a basal grade of morphologically distinct, monotypic genera in the species-rich Vanguerieae alliance (Rubiaceae, Ixoroideae): Its systematic and conservation implications. *Taxon* 60: 941–952.
- Razafimandimbison SG, Rydin C, Bremer B. 2008. Evolution and trends in the Psychotrieae alliance (Rubiaceae) – A rarely reported evolutionary change of many-seeded carpels from one-seeded carpels. *Molecular Phylogenetics and Evolution* 48: 207–223.
- Razafimandimbison SG, Taylor CM, Wikström N, Pailler T, Khodabandeh A, Bremer B. 2014. Phylogeny and generic limits in the sister tribes Psychotrieae and Palicoureae (Rubiaceae): Evolution of schizocarps in Psychotria and origins of bacterial leaf modules of the Malagasy species. *American Journal of Botany* 101: 1102–1126.
- Regier JC, Schultz JW, Zwick A, Hussey A, Ball B, Wetzler R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463: 1079–1083.
- Regier JC, Zwick A. 2011. Sources of signal in 62 protein-coding nuclear genes for higher-level phylogenetics of Arthropods. *PLoS One* 8: e23408.
- Robbrecht E. 1988. Tropical woody Rubiaceae. *Opera Botanica Belgica* 1: 1–271.
- Robbrecht E, Manen J-F. 2006. The major evolutionary lineages of the coffee family (Rubiaceae, angiosperms). Combined analysis (nDNA and cpDNA) to infer the position of *Coptosapelta* and *Luculia*, and supertree construction based on *rbcl*, *rps16*, *trnL-trnF* and *atpB-rbcl* data. A new classification in two subfamilies, Cinchonoideae and Rubioideae. *Systematics and Geography of Plants* 76: 85–145.
- Rokas A, Williams BL, King N, Carrol SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

- Rota-Stabelli O, Lartillot N, Phillippe H, Pisani D. 2013. Serine codon-usage bias in deep phylogenomics: Pancrustacean relationships as a case study. *Systematic Biology* 62: 121–133.
- Rothfels CJ, Li F-W, Sigel EM, Huiet L, Larsson A, Burge DO, Ruhsam M, Deyholos M, Soltis DE, Stewart CN, Shaw SW, Pokorny L, Chen T, dePamphilis C, DeGironimo L, Chen L, Wei X, Sun X, Korall P, Stevenson DW, Graham SW, Wong GK, Pryer KM. 2015. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. *American Journal of Botany* 102: 1089–1107.
- Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh GJ. 2014. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.
- Rydin C, Kainulainen K, Razafimandimbison SG, Smedmark JEE, Bremer B. 2008. Rare and enigmatic genera (*Dunnia*, *Schizocolea*, *Colletocema*), sisters to species-rich clades: Phylogeny and aspects of conservation biology in the coffee family. *Molecular Phylogenetics and Evolution* 48: 74–83.
- Rydin C, Kainulainen K, Razafimandimbison SG, Smedmark JEE, Bremer B. 2009a. Deep divergences in the coffee family and the systematic position of *Acranthera*. *Plant Systematics and Evolution* 278: 101–123.
- Rydin C, Razafimandimbison SG, Khodabandeh A, Bremer B. 2009b. Evolutionary relationships in the Spermaceae alliance (Rubiaceae) using information from six molecular loci: Insights into systematic affinities of *Neohymenopogon* and *Mouretia*. *Taxon* 58: 793–810.
- Rydin C, Wikström N, Bremer B. 2017. Conflicting results from mitochondrial genomic data challenge current view of Rubiaceae phylogeny. *American Journal of Botany* 104: 123–140.
- Samson N, Bausher MG, Lee S-B, Jansen RK, 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.
- Sazima M, Buzato S, Sazima I. 1999. Bat-pollinated flower assemblages and bat visitors at two Atlantic forest sites in Brazil. *Annals of Botany* 83: 705–712.
- Sennblad B, Bremer B. 2000. Is there a justification for differential a priori weighting in coding sequences? A case study from *rbCL* and *Apocynaceae* s.l. *Systematic Biology* 49: 101–113.
- Simmons MP, Ochoterena H, Freudenstein JV. 2002. Amino acid vs. nucleotide characters: Challenging preconceived notions. *Molecular Phylogenetics and Evolution* 24: 78–90.
- Simmons MP, Zhang L-B, Webb CT, Reeves A. 2006. How can third codon positions outperform first and second codon positions in phylogenetic inference? An empirical example from the seed plants. *Systematic Biology* 55: 245–258.
- Smedmark JEE, Razafimandimbison SG, Wikström N, Bremer B. 2014. Inferring geographic range evolution of a pantropical tribe in the coffee family (Lasiantheae, Rubiaceae) in the face of topological uncertainty. *Molecular Phylogenetics and Evolution* 70: 182–194.
- Staden R. 1996. The Staden sequence analysis package. *Molecular Biotechnology* 5: 233–241.
- Staden R, Beal KF, Bonfield JK. 2000. The Staden Package, 1998. In: Misener S, Krawets SA eds. *Bioinformatics methods and protocols, volume 132 of methods in molecular biology*. New Jersey, USA: Humana Press Inc. 115–130.
- Stamatakis A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. 2013. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biology and Evolution* 5: 1872–1885.
- Straub SC, Fishbein M, Livshultz T, Foster Z, Parks M, Weitemier K, Cronn RC, Liston A. 2011. Building a model: Developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12: 211.
- Tarrío R, Rodríguez-Trelles F, Ayala FJ. 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of Drosophilidae. *Molecular Biology and Evolution* 18: 1464–1473.
- Thureborn O, Razafimandimbison SG, Wikström N, Khodabandeh A, Rydin C. 2019. Phylogeny of Anthospermeae of the coffee family inferred using clock and non-clock models. *International Journal of Plant Sciences* 180: 386–402.
- Van Den Bussche RA, Baker RJ, Huelsenbeck JP, Hillis DM. 1998. Base compositional bias and phylogenetic analyses: A test of the "Flying DNA" hypothesis. *Molecular Phylogenetics and Evolution* 13: 408–416.
- Volkov RA, Komarova NY, Hemleben V. 2007. Ribosomal DNA in plant hybrids: Inheritance, rearrangement, expression. *Systematics and Biodiversity* 5: 261–276.
- Wen HZ, Wang RJ. 2012. *Foonchewia guangdongensis* gen. et sp. nov. (Rubiaceae: Rubiaceae) and its systematic position inferred from chloroplast sequences and morphology. *Journal of Systematics and Evolution* 50: 467–476.
- Wikström N, Kainulainen K, Razafimandimbison SG, Smedmark JEE, Bremer B. 2015. A revised time tree of the Asterids: Establishing a temporal framework for evolutionary studies of the coffee family (Rubiaceae). *PLoS One* 10: e0126690.
- Yang L-L, Li H-L, Wei L, Kuang S-Y, Li M-H, Liao Y-Y, Chen Z-D, Wu H, Zhang S-Z. 2016. A supermatrix approach provides a comprehensive genus-level phylogeny for Gentianales. *Journal of Systematics and Evolution* 54: 400–415.
- Zhong B, Deusch O, Goremykin VV, Penny D, Biggs PJ, Atherton RA, Nikiforova SV, Lockhart PJ. 2011. Systematic error in seed plant phylogeny. *Genome Biology and Evolution* 3: 1340–1348.
- Zimmer EA, Wen J. 2012. Using nuclear gene data for plant phylogenetics: Progress and prospects. *Molecular Phylogenetics and Evolution* 65: 774–785.
- Zwick A, Regier JC, Zwickl DJ. 2012. Resolving discrepancy between nucleotides and amino acids in deep-level arthropod phylogenomics: Differentiating serine codons in 21-amino-acid models. *PLoS One* 7: e47.

## Supplementary Material

The following supplementary material is available online for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12566/supinfo>:

**Fig. S1.** Phylogram resulting from nucleotide based bayesian analysis of the combined plastid CDS and non-CDS data sets. Nodes indicated by a black dot are well supported and have a bayesian posterior probability (BPP) of 0.95 or more (Alfaro et al., 2003; Erixon et al., 2003). Maximum Likelihood non-parametric bootstrap support (BS) is also indicated for each node in the tree. Support values are 1.00 (BPP) and 100% (BS) unless indicated in the figure. Subfamilies Rubioideae, Cinchonoideae, and Ixoroideae, and the Psychotriaceae, Spermaceae (Rubiaceae), and Vanguerieae, Coffeae (Ixoroideae) alliances as traditionally recognized are indicated

on the tree. Taxa in blue show conflicting relationships in different analyses of data from the same (plastid) genomic compartment (see Fig. 3). Taxa in red show conflicting relationships in analyses of data from different genomic compartments (see Figs. 4, S2).

**Fig. S2.** Phylogram resulting from the MrBayes analyses of nuclear rDNA data including also the variable ITS1 and ITS2 spacer regions. Ambiguous ITS1 and ITS2 alignment positions were filtered out prior to the analyses using transitive consistency scores (TCS; Notredame et al., 2000; Chang et al., 2014). Alignment positions with a TCS score of 3 or less were considered ambiguous and 360 of the total

648 ITS spacer characters were excluded from the analyses. Nodes indicated by a black dot are well supported and have a bayesian posterior probability (BPP) of 0.95 or more (Alfaro et al., 2003; Erixon et al., 2003). Maximum Likelihood non-parametric bootstrap support (BS) is also indicated for each node in the tree. Support values are 1.00 (BPP) and 100% (BS) unless indicated in the figure and nodes with BPP < 0.50 are collapsed. Subfamilies Rubioideae, and Cinchonoideae plus Ixoroideae as traditionally recognized are indicated on the tree. Taxa in red show conflicting relationships in analyses of data from different genomic compartments (see Figs. 2, 3).