Phylogeny and topological incongruence in the Rubioideae (Rubiaceae)

Olle Thureborn

Doctoral Thesis in Plant Systematics at Stockholm University, Sweden 2023

Stockholm University

Phylogeny and topological incongruence in the Rubioideae (Rubiaceae)

Olle Thureborn

Academic dissertation for the Degree of Doctor of Philosophy in Plant Systematics at Stockholm University to be publicly defended on Friday 24 March 2023 at 09.30 in Vivi Täckholmsalen (Q-salen), NPQ-huset, Svante Arrhenius väg 20.

Abstract

The work with this thesis has focused on evolutionary relationships in the Rubioideae, the most species-rich subfamily of the large and diverse coffee family (Rubiaceae). Despite considerable efforts during the last decades, uncertainty regarding several relationships in this group has remained, either as a result of unconvincing statistic support, incongruent results, or insufficient taxon sampling. Here, sequence data were obtained using both traditional and recently developed molecular methods, substantially expanding the amount of sequence data available for phylogenetic analysis within Rubioideae. Leveraging considerable amounts of data from a comprehensive sample of taxa showed to be highly effective in resolving phylogenetic relationships in the group, including its most recalcitrant parts. The results indicated that nuclear gene tree discordance at short internodes were frequently consistent with high levels of incomplete lineage sorting (ILS) owing to rapid diversification in the group. This finding may explain why some relationships have been notoriously difficult to resolve. Furthermore, while the results from coalescent simulations indicated that ILS alone can explain most of the observed cytonuclear incongruence, plastome introgression may be the more likely explanation in at least one case. Within Rubioideae the tribe Anthospermeae was studied in more detail. The phylogenetic analyses of this tribe revealed several cases and types of topological incongruence. Nevertheless, the deepest splits of the Anthospermeae phylogeny were congruent among analyses, but partly inconsistent with the traditional subtribal delimitation of the tribe. The infratribal classification of Anthospermeae was therefore partly updated. Unusual plastome features were also found within Anthospermeae, comprising large inversions and (putative) mitochondrial-to-plastome DNA transfer. With few other exceptions, plastomes across the Rubioideae tend to be highly conserved and typically conform to the canonical structure, gene content, and gene order of the majority of flowering plants.

Keywords: *Anthospermeae, cytonuclear discordance, gene trees, incomplete lineage sorting, incongruence, mitochondrial genome, nuclear genome, phylogenomics, phylogeny, plastid genome, Rubiaceae, Rubioideae, species trees.*

Stockholm 2023
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-214175

ISBN 978-91-8014-182-6
ISBN 978-91-8014-183-3

Department of Ecology, Environment and Plant Sciences

Stockholm University, 106 91 Stockholm

PHYLOGENY AND TOPOLOGICAL INCONGRUENCE IN THE RUBIOIDEAE (RUBIACEAE)

Olle Thureborn

Stockholm
University

Phylogeny and topological incongruence in the Rubioideae (Rubiaceae)

Olle Thureborn

©Olle Thureborn, Stockholm University 2023

ISBN print 978-91-8014-182-6
ISBN PDF 978-91-8014-183-3

Cover photo: The Andringitra Massif, Madagascar, with *Anthospermum* plants (probably *A. madagascariense*) in the foreground. Photo by Olle Thureborn, January 2017.

Printed in Sweden by Universitetsservice US-AB, Stockholm 2023

Abstract

The work with this thesis has focused on evolutionary relationships in the Rubioideae, the most species-rich subfamily of the large and diverse coffee family (Rubiaceae). Despite considerable efforts during the last decades, uncertainty regarding several relationships in this group has remained, either as a result of unconvincing statistic support, incongruent results, or insufficient taxon sampling. Here, sequence data were obtained using both traditional and recently developed molecular methods, substantially expanding the amount of sequence data available for phylogenetic analysis within Rubioideae. Leveraging considerable amounts of data from a comprehensive sample of taxa showed to be highly effective in resolving phylogenetic relationships in the group, including its most recalcitrant parts. The results indicated that nuclear gene tree discordance at short internodes were frequently consistent with high levels of incomplete lineage sorting (ILS) owing to rapid diversification in the group. This finding may explain why some relationships have been notoriously difficult to resolve. Furthermore, while the results from coalescent simulations indicated that ILS alone can explain most of the observed cytonuclear incongruence, plastome introgression may be the more likely explanation in at least one case. Within Rubioideae the tribe Anthospermeae was studied in more detail. The phylogenetic analyses of this tribe revealed several cases and types of topological incongruence. Nevertheless, the deepest splits of the Anthospermeae phylogeny were congruent among analyses, but partly inconsistent with the traditional subtribal delimitation of the tribe. The infratribal classification of Anthospermeae was therefore partly updated. Unusual plastome features were also found within Anthospermeae, comprising large inversions and (putative) mitochondrial-to-plastome DNA transfer. With few other exceptions, plastomes across the Rubioideae tend to be highly conserved and typically conform to the canonical structure, gene content, and gene order of the majority of flowering plants.

List of papers

This thesis is based on the following papers, which are referred to in the text by their respective roman numerals:

I: **Thureborn O**, Razafimandimbison SG, Wikström N, Khodabandeh A, Rydin C. 2019. Phylogeny of Anthospermeae of the coffee family inferred using clock and nonclock models. *International Journal of Plant Sciences* 80 (5): 386-402. https://doi.org/10.1086/703353

II: **Thureborn O**, Wikström N, Razafimandimbison SG, Rydin C. Phylogenomics and topological conflicts in the angiosperm tribe Anthospermeae (Rubiaceae). *Manuscript.*

III: **Thureborn O**, Razafimandimbison SG, Wikström N, Rydin C. 2022. Target capture data resolve recalcitrant relationships in the coffee family (Rubioideae, Rubiaceae). *Frontiers in Plant Sciences* 13: 967456. https://doi.org/10.3389/fpls.2022.967456

IV: **Thureborn O**, Wikström N, Razafimandimbison SG, Rydin C. Plastid phylogenomics and cytonuclear discordance in Rubioideae. *Manuscript.*

Author contribution:
Paper I: OT designed the work together with the co-authors. OT carried out parts of the lab work, assembled and analyzed the data. OT wrote the manuscript with comments from all co-authors.

Paper II: OT designed the work together with NW and CR. OT carried out the lab work, bioinformatics, and analyzed the data. OT wrote the manuscript with comments from all co-authors.

Paper III: OT designed the work together with the co-authors. OT carried out the lab work, bioinformatics, and analyzed the data. OT wrote the manuscript with comments from all co-authors.

Paper IV: OT designed the work together with NW and CR. OT carried out the bioinformatics and analyzed the data. OT wrote the manuscript with comments from all co-authors.

Contents

Abbreviations

CDS – Coding sequence

DNA – Deoxyribonucleic acid

GTR+G – General time reversible model with gamma distributed rates

ILS – Incomplete lineage sorting

MSC – Multispecies coalescent

NGS – Next generation sequencing

s.s. – sensu stricto (in the narrow sense)

Introduction

Systematics

The scientific field of systematic biology, or systematics, includes two major branches, (i) taxonomy, the science of discovering, describing, naming, and classifying organisms, and (ii) phylogenetics, the science of evolutionary relationships among organisms. Unlike early cornerstone classifications that predates the theory of evolution, classifications nowadays commonly involve the inference of evolutionary relationships between organisms, resulting in phylogenetic classifications, thus uniting the two major branches of systematics. The view that classifications should reflect evolutionary relationships was initiated by Hennig (1966). In such classifications names are only given to monophyletic groups (clades) - groups of organisms that comprise the most recent common ancestor and all of its descendants, which are defined by shared derived characters. The hypothesized evolutionary relationships among organisms are often depicted as a phylogenetic tree. Before the 1990ties, phylogenetic analysis was typically based on morphological data, but for example chemical data, physiological data and karyological data were also employed for the purpose of tracing the evolutionary history of a group. Phylogenetic trees are widely used across the life sciences. To first consider the impact of phylogeny, inheritance from ancestors and the ancestral vs. derived state of features, is critical when asking scientific questions in most other biological fields. Systematists too use phylogenies as a framework for investigations of other scientific questions than the actual evolutionary relationships, such as the evolution of traits, historical biogeography and biodiversity change over long time periods, i.e., how speciation and extinction contribute to variation and net diversification of organisms.

During the 1980ties and early 1990ties, systematic biology was revolutionized by methodological advances concerning phylogenetic analysis and DNA sequencing, and the use of molecular data for inferring phylogeny soon became common practice. A completely new type of data was suddenly available and could be used to make independent tests of earlier theories of evolution and contemporary classification schemes. However, this new type of data and utilization thereof are not free of issues that may mislead our understanding of evolution of life on Earth.

Molecular markers in molecular systematic studies

DNA sequence data have now been used to reconstruct phylogenetic relationships among plants for more than three decades. So far plant phylogenetic studies have primarily relied on data obtained from the plastid genome (the plastome), and data from the 18S-5.8S-26S nuclear ribosomal cistron (Small et al. 2004; Gitzendanner et al. 2018). The relatively simple mode of evolution (e.g., generally uniparental and clonal inheritance, and conserved structure and gene content) and high abundance per plant cell make the plastome relatively easily obtained and analyzed. The other organellar genome, the mitochondrial genome (mitogenome), has been largely neglected in phylogenetic studies of plants, possibly because of its relatively low substitution rate, highly dynamic genome evolution and presumably shared inheritance pattern with the plastome (Rieseberg and Soltis 1991; Small et al. 2004; Sloan et al. 2009). The large and complex biparentally inherited nuclear genome has traditionally been represented mainly by data from the nuclear ribosomal cistron. In parallel with the plastome, a high number of copies in the nuclear genome and blend of conserved and less conserved sequence regions have made the nuclear ribosomal cistron regions easily accessible and valuable for plant phylogenetic studies at several different taxonomic ranks (Baldwin and Markos 1998; Alvarez and Wendel 2003; Poczai and Hyvönen 2010). Molecular phylogenies inferred from plastome data and nuclear ribosomal data have provided major insights into the evolution of land plants (Gitzendanner et al. 2018). However, due to its mode of inheritance, the plastome (and the mitogenome) is generally regarded to evolve as a single gene (Doyle 1992, 2022; Gitzendanner et al. 2018). Also the ribosomal cistron is generally considered to evolve as a single gene due to the phenomenon of concerted evolution (Alvarez and Wendel 2003; Small et al. 2004; Poczai and Hyvönen 2010). While analysis of only a few genes can be sufficient to accurately estimate the species tree, analyses of empirical and simulated data have shown that the accuracy of the inferred phylogeny is increased with an increased number of independent genes (Pamilo and Nei 1988; Maddison 1997; Edwards et al. 2007; Ruane et al. 2015). The large and complex biparentally inherited nuclear genome has an almost endless supply of independently evolving single-copy or low-copy genes, but technical issues have for a long time constrained a broad utilization of such genes (Sang 2002; Small et al. 2004; Zimmer and Wen 2013).

The recent advent of high-throughput, next-generation sequencing (NGS) methods has again revolutionized the field of evolutionary biology, including systematics. NGS technologies yield orders of magnitude of more data at a considerably lower cost than traditional Sanger sequencing (McKain et al. 2018). NGS methods have made it relatively easy and inexpensive to retrieve the high copy number genomes (plastome and mitogenome) and the high copy number nuclear ribosomal cistron by shallow sequencing of genomic DNA (also known as genome skimming) for multiple samples in a single sequencing run (Straub et al. 2012). However, due to the large size and complexity of plant nuclear genomes it is, notwithstanding the much lower cost associated with NGS, generally not feasible to sequence and assemble whole nuclear genomes for a large sample of taxa, which is typically used for plant phylogenetic studies. Consequently, methods such as target capture or target enrichment (Weitemier et al. 2014) has been developed, which enrich genomic regions of interest prior to sequencing, thus allowing for higher sample throughput and higher read coverage of the targeted regions than whole genome sequencing. One major benefit with this method is the fact that it also functions well with degraded DNA, a common problem when extracting DNA from herbarium specimens (McKain et al. 2018; Johnson et al. 2019). Furthermore, off-target reads can often be extracted from the raw data and used to assemble large plastome datasets (Weitemier et al. 2014). Probe sets needed for "fishing out" a targeted selection of genes from the nuclear genome and designed to fit universally on angiosperms are now available in the literature, for example the 353 angiosperm probe set (Johnson et al. 2019) that was utilized in the present thesis (Paper III). However, systematists may also design new probe sets in order to target data suitable for a specific research question, although that will typically require that at least one genome or transcriptome has been produced from a related taxon and made freely available to the scientific community.

Gene trees and species trees

An important aspect when inferring relationships among species or other taxonomic ranks is that the phylogenetic history of a gene may not be the same as that of another gene or that of the species (Pamilo and Nei 1988; Doyle 1992; Maddison 1997). Incongruence between gene trees and species trees may be explained in quite a number of ways, including methodological issues such as sampling error and systematic error, and biological causes such as incomplete lineage sorting, hybridization and gene duplication and loss (Doyle 1992; Maddison 1997).

The most common biological cause for gene tree-species tree incongruence is incomplete lineage sorting (ILS) (or deep coalescence), which happens when ancestral polymorphisms fails to reach fixation in descendant populations through successive speciation events (Pamilo and Nei 1988; Maddison 1997; Edwards 2009). An example is that of the Chimpanzee-Human-Gorilla species tree where roughly 30% of the genome is incongruent with the Human-Chimpanzee relationship, showing Gorilla-Chimpanzee or Human-Gorilla in a more or less even split (Scally et al. 2012; Dutheil et al. 2015). The probability of ILS is especially likely when the time between speciation events is short and effective population sizes are large (Pamilo and Nei 1988). Since the mitogenome and plastome are typically thought to be uniparentally inherited, thus having significantly smaller effective population sizes, divergent variants of those genomes reach fixation faster compared to nuclear genes (Rieseberg and Soltis 1991; Small et al. 2004).

Introgression, the exchange of genetic material between species via hybridization and subsequent backcrosses of hybrids to one of the parental species, is another common source of phylogenetic incongruence (Mallet et al. 2016). Compared to nuclear introgression, organellar introgression has been shown to occur more frequently, which probably reflects the differences in mode of inheritance and selective pressures between organellar and nuclear genomes (Martinsen et al. 2001). More specifically, this phenomenon, often referred to as cytoplasmic capture, plastid capture, chloroplast capture, or organellar capture, occurs if the presumably maternally inherited plastome is retained but the biparentally inherited nuclear genome is replaced, via repeated backcrossing, by the nuclear genome of the paternal parent genome (Rieseberg and Soltis 1991; Soltis and Kuzoff 1995). Organellar introgression is often suggested as an explanation when phylogenetic relationships derived from nuclear and organellar data differ.

The processes of gene duplication and loss can also cause gene tree discordance. Whole genome duplication, also known as polyploidization, and small-scale duplications are both common in flowering plants (Cui et al. 2006; Van de Peer et al. 2009; Defoort et al. 2019). While duplicated genes, referred to as paralogs, often go back to single copy status, they also can be retained, for example when they take on a new function (Lynch and Conery 2000). If the pattern of gene loss is not the same between different lineages the resulting gene tree may not match the species tree (Fitch 1970; Maddison 1997).

Technical reasons can also result in incongruence between phylogenies. One such reason is sampling error, related to the low phylogenetic information content in small datasets (Delsuc et al. 2005). An obvious remedy is to sample larger datasets, e.g., longer sequences or more genes, which may eliminate this kind of error (Rokas et al. 2003; Rasmussen and Kellis 2007). The decreased risk of sampling error is one of the main advantages with the phylogenomic approach.

Phylogenetic studies typically apply model-based analytical methods, such as maximum likelihood and Bayesian approaches. Systematic error occurs when the model used to infer the phylogeny cannot accommodate important features of the data, resulting in an incorrect phylogenetic reconstruction. Systematic error is mainly caused by heterogeneities in the data. Kapli et al. (2021) described how conventionally used models can handle three different kinds of heterogeneities (rate variation across lineages, rate variation across sites and heterogenous nucleotide (or amino acid) composition across lineages), for example by partitioning the data into homogenous subsets. However, how to adequately partition the data can be difficult to know beforehand (Kapli et al. 2021). A classic example of systematic error is the long branch attraction artifact, in which fast evolving taxa erroneously group together (Felsenstein 1978). Unlike the case of sampling error, the addition of data can make the effect of systematic error even worse, adding more support for an incorrect topology (Phillips et al. 2004). Improving the detection of multiple substitutions by increasing taxon sampling and using more realistic models are two effective strategies to combat systematic error (Heath et al. 2008; Philippe et al. 2011).

Unfortunately, the most complex models of sequence evolution require a high computational effort and are therefore often intractable for the analysis of large datasets. A more computationally effective strategy is data recoding, which recodes the data so that it is more likely to comply to the assumptions of the utilized model. For example, RY-coding converts the two purines A and G into R, and the two pyrimidines C and T into Y, thus discarding fast evolving transitions and focusing on the more slowly evolving transversions. This method has been shown to effectively reduce the effect of saturation (due to rapid evolutionary rates) and compositional bias on phylogenetic accuracy, especially when applied to large datasets (Phillips et al. 2001; Ishikawa et al. 2012). Similarly, excluding fast evolving taxa, fast evolving third codon position in protein coding sequences (CDS) or other fast evolving data partitions are other strategies often used to limit the risk of systematic error (Goremykin 2003; Karol et al. 2010; Philippe et al. 2011; Simmons 2017; Young and Gillung 2020). An advantage with RY-coding is that it is applicable to all kinds of nucleotide data, unlike for example degenerate recoding (Criscuolo and Gribaldo 2010; Zwick et al. 2012), which is designed for CDS-data only. However, these recoding strategies reduce the amount of information in the dataset (Ishikawa et al. 2012; Simmons 2017; Braun and Kimball 2021), and may thus be difficult to justify in case the data is not known to be biased. Nevertheless, in addition to analyzing the full data, employing strategies such as data recoding can be useful to allay fears that the wrong relationship(s) has been inferred.

To what extent datasets are affected by systematic error is not entirely clear, but studies have shown that while violations of assumptions are common across many datasets, conventionally used models have proven to perform well even when the model is rather severely violated (Naser-Khdour et al. 2021). In general, the problem of systematic error is likely most pronounced when dealing with ancient divergences and distantly related taxa (Kapli et al. 2021). For example, inaccurate phylogenetic relationships stemming from systematic error due to saturation and compositional bias have been suggested regarding the non-monophyly of bryophytes retrieved using plastid data (Cox et al. 2014; Sousa et al. 2019), and Lycopodiophyta rather than Monilophyta being sister to seed plants using plastid data (Ruhfel et al. 2014). It has, however, been shown that presumably saturated sites such as 3:d codon positions in protein coding genes are informative and useful also for assessing relationships in old clades, such as green plants and seed plants (e.g., Källersjö et al. 1999; Rydin et al. 2002; Simmons 2017).

Apart from different strategies to accommodate features of sequence evolution, models that can accommodate heterogeneity among gene trees have been developed. Such models have mainly focused on gene tree discordance due to ILS, as this is likely the most common cause of gene tree discordance. ILS is modeled by the multispecies coalescent (MSC) model (Pamilo and Nei 1988). The most scalable and commonly used coalescent-based approach is the two-step summary approach, which first estimates individual gene trees from individual alignments and subsequently uses these gene trees to estimate the species tree according to a model that takes gene tree discordance into account. This is different than the traditional way of concatenating all sequence data under the assumption that the phylogeny of such a supergene will be a good representation of the species tree. Such concatenation-based approaches, are effective at reducing sampling error but ignores gene tree heterogeneity. Unlike coalescent-based methods such as ASTRAL (Mirarab et al. 2014), the concatenation method can be inconsistent under the MSC and can in certain conditions infer a highly supported but incorrect phylogeny (Kubatko and Degnan 2007; Roch and Steel 2015; Mendes and Hahn 2018).

The pros and cons of these two approaches have been extensively debated in the literature (Springer and Gatesy, 2016; Edwards et al. 2016). One of the main criticisms of coalescent-based models is that they assume that all gene tree discordance is due to ILS, while it for real data often may be explained by other factors such as gene tree error due to for example low information content in short gene alignments (i.e., sampling error). Such faulty gene trees can result in errors in the resulting species tree (Molloy and Warnow 2018). Methods that jointly infer gene and species trees may be the best candidates to accommodate such errors, but due to their complexity they are impractical for most datasets (Ogilvie et al. 2016; Rannala and Yang 2017). Instead, collapsing branches with low support and increasing the information content in utilized gene alignments can improve the accuracy of species tree inference using summary methods (Zhang et al. 2018). On the other hand, increasing the length of the gene tree alignment also increases the risk of violating the MSC assumption of no recombination within loci. However, violation of that assumption has been shown to be limited in both simulations and analyses of biological datasets (Lanier and Knowles 2012; Wang and Liu 2016; Folk et al. 2017; Morales-Briones et al. 2018), the violation of free recombination between genes being more severe (Wang and Liu 2016).

The subfamily Rubioideae of the coffee family (Rubiaceae)

Rubiaceae (the coffee family) is estimated to have originated in the mid-Cretaceous and is with its over 13 000 species the fourth largest angiosperm family (Bremer and Eriksson, 2009; Davis et al., 2009). Consequently, it is no surprise that the Rubiaceae are very diverse in terms of morphology, habit and ecology, ranging from small annual herbs to tall canopy trees. Still, they can often be recognized by their opposite phyllotaxis, simple leaves, interpetiolar stipules and epigynous flowers. Regarding pollination syndrome, most species are animal-pollinated whereas wind-pollinated flowers have so far only been found in the two tribes Anthospermeae and Theligoneae (Puff, 1982), both members of the Spermacoceae alliance of the subfamily Rubioideae. Although the family has a worldwide distribution, most species are found in tropical and subtropical regions, and temperate taxa are exceptions to the rule. Rubiaceae has a high degree of species endemism. For example, in Madagascar, one of the world's main biodiversity hotspots, Rubiaceae is represented by ca 650 species in 95 genera, which makes it the second largest angiosperm family on the island (Davis et al., 2009; Janssens et al., 2016). The restricted distribution of many Rubiaceae species accompanied by (assumed) ecological sensitivity (for example the almost exclusive restriction to primary forests in the tropics) has been taken as an indication of an elevated extinction risk for members of the family (Davis et al., 2009). Among well-known and economically important species of the Rubiaceae are the sources of coffee (*Coffea*) and the antimalarian drug quinine (*Cinchona*), as well as ornamental plants (e.g., *Gardenia* and *Ixora*).

The subfamilial taxonomy of Rubiaceae has been in a state of flux since first established with the number of recognized subfamilies varying between two and eight (Schumann 1891; Verdcourt 1958; Bremekamp 1966; Robbrecht 1988; Bremer et al. 1995; Andreasen and Bremer 2000; Robbrecht and Manen 2006). Based on phylogenetic analysis of molecular data, authors have proposed a division of the coffee family into two subfamilies, Cinchonoideae and Rubioideae (Robbrecht and Manen 2006), or three subfamilies, Cinchonoideae, Ixoroideae, and Rubioideae (Bremer and Eriksson 2009). The two Asian tribes Coptosapelteae and Luculieae have been difficult to place in any of the subfamilies, frequently ending up in different and/or poorly supported positions in deep parts of the Rubiaceae tree, and they have traditionally been treated as incertae sedis.

The subfamily Rubioideae (Figs 1-2), the focus of this thesis, has been continuously recognized since its establishment, despite occasional alterations of its circumscription. The subfamily was first proposed by Bremekamp (1952) and later formalized by Verdcourt (1958). It is with over 8000 species the largest of the main lineages of Rubiaceae. The traditionally recognized morphological characteristics of Rubioideae include the presence of raphides (needle-like crystals) in tissues, valvate corolla aestivation, indumentum of septate hairs, and heterostyly (Verdcourt 1958; Robbrecht 1988; Bremer and Manen 2000; Robbrecht and Manen 2006; Bremer and Eriksson 2009). While Rubiaceae is a primarily woody family, Rubioideae is the only subfamily which also includes large herbaceous clades. In addition to presence of herbaceous habit, some groups are distributed in temperate regions, deviating from the mainly tropical and subtropical distribution pattern of the family.

Of the 29 currently recognized Rubioideae tribes, the majority are divided into two major informal groups, each containing over 3000 species: the cosmopolitan Spermacoceae alliance (13 tribes) and the pantropical Psychotrieae alliance (nine tribes). In the Psychotrieae alliance, most species are contained in the pantropical sister tribes Palicoureeae (over 1100 species) and Psychotrieae (over 1800 species), mainly reflected by their respective type genera. In the Spermacoceae alliance, two essentially herbaceous tribes make up the bulk of the species: the mainly pantropical tribe Spermacoceae (over 1300 species), and the cosmopolitan and mainly temperate tribe Rubieae (over 900 species).

Phylogenetic studies in the Rubioideae have mostly used plastid markers, sometimes in combination with nuclear ribosomal DNA regions (e.g., nrITS and nrETS) (e.g., Andersson and Rova, 1999; Bremer and Manen, 2000; Razafimandimbison et al. 2008; Rydin et al. 2009b; Bremer and Eriksson, 2009). Hypotheses stemming from such studies have generated a fairly well-established consensus of relationships among tribes and major groups in Rubioideae and the rest of the family. Recently, NGS approaches have facilitated the accumulation of large datasets. Two recent studies focusing on high-level relationships Rubiaceae used genome skimming to obtain large data sets from the mitogenome (Rydin et al. 2017) and plastome + nuclear ribosomal cistron data (Wikström et al. 2020), and a study focusing on high-level relationships in the order Gentianales used a target capture approach to obtain data from 353 nuclear genes and partial plastomes (Antonelli et al. 2021). Results from those tribal-level studies were overall well supported and corroborated and clarified several parts of the broadly accepted picture of intertribal relationships in Rubioideae. Despite this, several relationships within Rubioideae remained uncertain, either due to low statistic support, incongruence, or limited

Figure 1. Diversity of Rubioideae. **A**. *Schismatoclada*, Danaideae (copyright: SG Razafimandimbison* et al., Annals of Botany 2022, https://doi.org/10.1093/aob/mcac121, license CC BY v. 4.0, https://creativecommons.org/licenses/by/4.0/). **B**. *Temnopteryx*, Urophylleae (copyright: Anne-Hélène Paradis, http://legacy.tropicos.org/Image/101135412, license CC BY-NC-ND v. 3.0, https://creativecommons.org/licenses/by-nc-nd/3.0/). **C**. *Phyllis*, Anthospermeae (copyright: JT Johansson*, http://angio.bergianska.se). **D**. *Myrmecodia*, Psychotrieae (copyright: JT Johansson*, http://angio.bergianska.se). **E**. *Anthospermum*, Anthospermeae (copyright: E Larsén* and O Thureborn). **F**. *Morinda*, Morindeae (copyright: JT Johansson*, http://angio.bergianska.se). *Photo used with permission from the photographer.

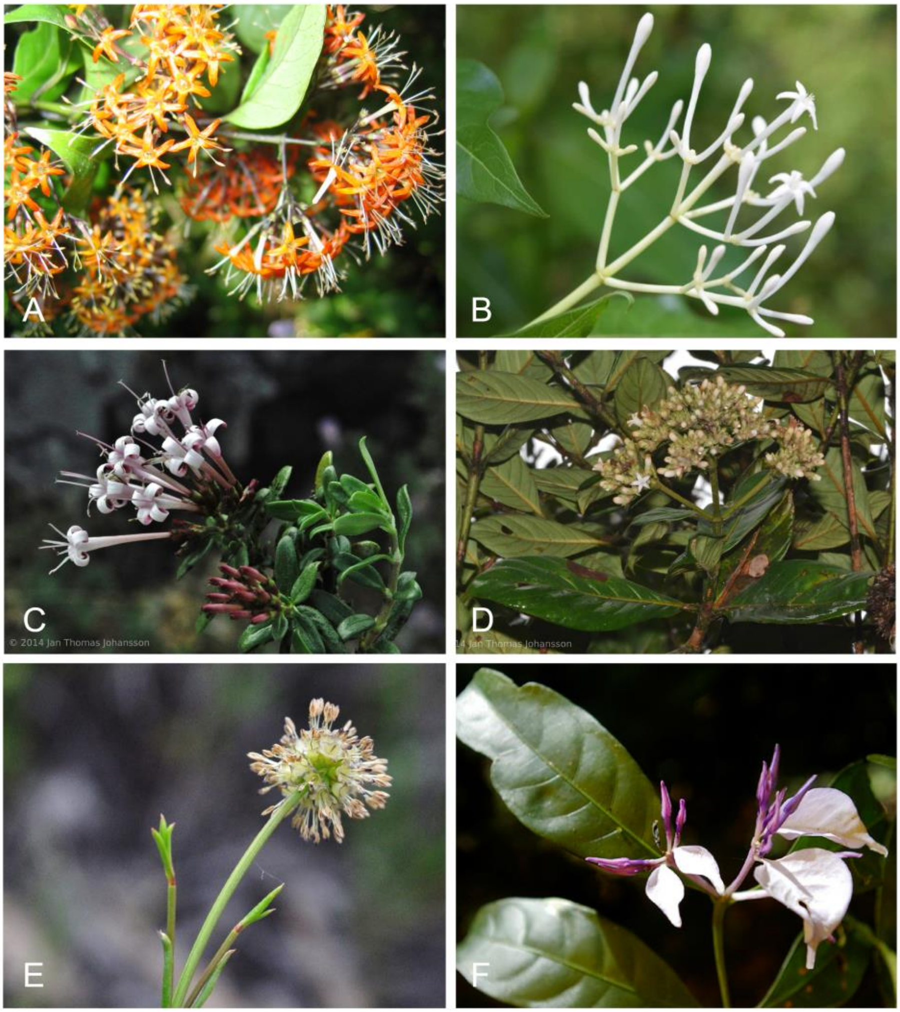

Figure 2. Diversity of Rubioideae. **A**. *Danais*, Danaideae (copyright: C Rydin*). **B**. *Chassalia*, Palicoureeae (copyright: E Larsén* and O Thureborn). **C**. *Plocama*, Putorieae (copyright: JT Johansson*, http://angio.bergianska.se). **D**. *Gaertnera*, Gaertnereae (copyright: JT Johansson*, http://angio.bergianska.se). **E**. *Opercularia*, Anthospermeae (copyright: JT Johansson*, http://angio.bergianska.se). **F**. *Faramea*, Coussareeae (DA Neill, copyright: Field Museum of Natural History, https://planttidtools.fieldmuseum.org/en/nlp/catalogue/3835245, license CC BY-NC v. 4.0, https://creativecommons.org/licenses/by-nc/4.0/). *Photo used with permission from the photographer.

taxon sampling. For example, regarding incongruences, either Ophiorrhizeae, a clade comprising Colletoecemateae and Urophylleae, or a clade comprising Colletoecemateae, Ophiorrhizeae, and Urophylleae was, depending on the genomic compartment utilized, resolved as a sister group to the remaining subfamily. Analysis of nuclear ribosomal data revealed another case of supported conflict, resolving Coussareeae (not sampled in Antonelli et al. 2021) as sister to the Spermacoceae alliance, contesting the well-documented sister-relationship between the Spermacoceae and Psychotrieae alliances in several previous studies (e.g., Bremer and Manen 2000; Razafimandimbison et al. 2008; Rydin et al. 2009b). Within the Spermacoceae and Psychotrieae alliances, relationships between tribes in the Rubieae-Theligoneae-Putorieae clade, deep splits in the Spermacoceae alliance, and the position of the tribe Gaertnereae in the Psychotrieae alliance were among the relationships that differed between analyses of data retrieved from the different genomic compartments.

These phylogenies of Rubiaceae and Gentianales have, however, typically relied on only one representative species per sampled tribe and some key taxa have not been included. Furthermore, these studies predominantly focused on CDS data; thus, non-coding data provide a relatively unexplored source of information in Rubiaceae phylogenetic studies. Moreover, while cytonuclear discordance have been detected within Rubioideae, the potential reasons for the observed conflicts have not been explicitly investigated.

Several of the tribes of Rubiaceae have been subject to a phylogenetic study using DNA sequence data offering new insights into the evolutionary relationships and monophyletic status of those tribes. One phylogenetically understudied group is the tribe Anthospermeae. This tribe of 12 genera and over 200 species is unusual within Rubiaceae as its members have wind-pollinated flowers, and some species are distributed in temperate regions. While early molecular-based studies gave valuable insight into the phylogeny the tribe (Anderson et al. 2001; Rydin et al. 2009b; Bremer and Eriksson 2009), low resolving power of used markers and sparse taxon sampling have left many phylogenetic relationships within Anthospermeae unclear. With the progressive development of both sequencing and phylogenetic inference methods there is now an opportunity to reevaluate former phylogenetic hypothesis.

Aims

The overarching aim of this thesis is to produce ample new molecular data for studying the evolutionary relationships within the subfamily Rubioideae.

In Paper I the main aims are to test the monophyletic status of the tribe Anthospermeae, its traditional subtribes, genera, and Malagasy members using markers from the plastome and the nuclear ribosomal cistron.

In Paper II the main aims are to improve the current understanding of the intergeneric relationships within Anthospermeae, and to explore phylogenetic incongruence, using data from the organellar genomes and the nuclear ribosomal cistron obtained via genome skimming.

In Paper III the main aim is to improve the current understanding of the relationships within the subfamily Rubioideae, mainly among tribes but also within tribes, using hundreds of nuclear genes.

In Paper IV the main aims are to clarify the plastome phylogeny of Rubioideae, and investigate whether ILS or hybridization/introgression best explains any detected cases of cytonuclear discordance within the group.

Materials and methods

Plant material and taxon sampling

The majority of DNA sequences used in this thesis have been newly produced within the scope of the four papers. Plant material for DNA extraction and/or sequencing was obtained from silica dried material from field collections, from herbarium specimens, and from DNA extracts already available from previous work. Publicly available sequence data retrieved from the European Nucleotide Archive and GenBank were also used.

In Paper I, molecular data were obtained for a broad sample of taxa. Taxon sampling included 87 species of Anthospermeae (Rubioideae), representing all of the 12 currently recognized genera of the tribe. Since the large genus *Coprosma* had been dealt with previously (Cantley et al. 2014, 2016), the sampling of this genus was restricted to include its main lineages including the two species *Coprosma moorei* and *Coprosma talbrockiei* that have previously been shown to be more closely related to *Durringtonia* than to the other *Coprosma* species (Cantley et al. 2016). Outgroup sampling included over 30 genera, representing 11 other recognized tribes of the Spermacoceae alliance and one representative tribe from the Psychotrieae alliance.

Anthospermeae was the focus also in Paper II. In this study the taxon sampling within Anthospermeae was decreased to comprise one representative from each of the 12 genera. Outgroup sampling included representatives from 12 non-Anthospermeae tribes, including 10 from the Spermacoceae alliance. While the taxon sampling was decreased compared to Paper I, the amount of sequence data was much increased, comprising complete (or nearly so) sequences of the plastome and nuclear ribosomal cistron, and a large fraction of the mitochondrial genome.

In Papers III and IV the taxonomic scope was greatly broadened, including 101 species of the subfamily Rubioideae. All but two (the monogeneric tribes Foonchewieae and Aitchinsonieae) of the 29 currently recognized tribes were sampled, and all but one non-monogeneric tribe were represented by more than one genus. For outgroup sampling twenty species were selected to represent the major lineages of the remaining Rubiaceae. In addition, three species from the remaining Gentianales, representing the families Gentianaceae, Loganiaceae, and Apocynaceae, were sampled.

Character sampling and laboratory work

Extraction of DNA was conducted using a modified version of the widely used cetyl trimethylammonium bromide method (Doyle and Doyle 1987). In Paper I DNA sequences of the nuclear (nrITS and nrETS) and plastid (*atpB-rbcL*, *ndhF*, *rbcL*, *rps16*, and *trnT-trnF*) markers were produced using standard protocols for PCR and Sanger sequencing. Production of sequences used in Paper II-IV was done using NGS approaches. The genome skimming approach used in Paper II included library preparation using the ThruPLEX DNA-seq library preparation kit from Rubicon (Rubicon Genomics, Ann Arbor, USA), followed by sequencing on the Illumina HiSeq2500 platform (Illumina, San Diego, CA, United States) with 2x126 bp paired-end reads .The target capture sequencing approach used in Paper III included library preparation using a modified version of the Meyer and Kircher (2010) protocol, followed by target capture of 353 nuclear genes using the Angiosperms353 probe set (Johnson et al. 2019), and sequencing on Illumina platforms (NextSeq 500 or NovaSeq 6000) with 2x151 bp paired-end reads. In Paper IV we utilized off-target data retrieved from the sequencing conducted during the work with Paper III to assemble partial and complete plastome sequences.

Sequence assembly

Obtained sequencing reads were imported into Geneious v11 (Kearse et al. 2012). The Illumina sequencing reads were preprocessed by removing adapters, removing duplicate reads, trimming low-quality bases, and discarding short reads. These steps were primarily executed using the BBduk and dedupe plugins implemented in Geneious.

The Sanger sequences produced during the work with Paper I were assembled using the native Geneious assembler. Three datasets were created: one plastid data set with both ingroup (Anthospermeae) and outgroup taxa, one plastid dataset containing only Anthospermeae taxa, and one nuclear dataset containing only taxa of the Anthospermeae.

The genome-skimming data in Paper II were assembled using a combined reference guided + *de novo* approach using algorithms implemented in Geneious. Three major datasets were created, one for each genomic compartment.

The target capture data produced during work with Paper III were assembled using the HybPiper v1.3.1 pipeline (Johnson et al. 2016). Coding sequences and so called supercontig sequences, which contain both CDS and non-coding flanking sequence were extracted. During assembly HybPiper identifies putative paralogs. To test the effect of potential paralogy on phylogenetic reconstruction, datasets with and without putative paralogous genes were created. Thus, two CDS datasets (with or without putatively paralogous genes) and two supercontigs datasets (with or without putatively paralogous genes) were created, resulting in a total of four datasets.

In Paper IV the off-target reads for the corresponding samples used in Paper III were used for plastome assembly. The majority of the plastomes were assembled using a combined reference-guided + *de novo* approach. Remaining plastomes were assembled using either a reference-guided approach, or a *de novo* approach.

Alignment and data recoding

Alignment of assembled sequences was primarily done using MAFFT v7 (Katoh and Standley 2013, 2016). Typically, marker regions (such as CDSs, supercontigs, introns, and spacers) were aligned separately, but in Paper IV, whole plastomes were aligned. Manual adjustments of some of the individual alignments used in papers I and II were done in AliView (Larsson 2014) or Geneious. For the much larger datasets used in Papers III and IV, alignments were curated with automatic alignment trimming tools. In Paper III SpruceUp v2020.2.19 (Borowiec 2019) was used. In Paper IV, trimAl v1.2rev57 (Capella-Gutierrez et al. 2009) was used to create an alternative trimmed alignment. Further, to test the effects of potential substitution saturation and compositional bias, RY-coding was applied to the datasets of Papers II and IV using BMGE v1.12 (Criscuolo and Gribaldo 2010). In Papers II and IV, to detect and verify plastome rearrangements, plastome sequences were aligned with the progressiveMauve algorithm in Mauve v2.3.1 (Darling et al. 2004, 2010).

Phylogenetic analyses

While Bayesian inference was used in Paper I, maximum likelihood methods were used in Papers II, III, and IV. In Paper III the coalescent-based summary method ASTRAL was also used.

Bayesian analyses

Bayesian analyses were conducted using MrBayes v3.2.6 (Ronquist et al. 2012). The datasets were analyzed as partitioned. PartitionFinder 2 (Lanfear et al. 2016) was used to find best-fit partitioning scheme. The reversible jump

Markov chain Monte Carlo algorithm (Huelsenbeck et al. 2004) for substitution model selection implemented in MrBayes was used for each partition. In Paper I, we were interested in evaluating the potential impact of different clock models on phylogenetic results within Anthospermeae. Therefore, each dataset was analyzed using both non-clock and clock models (relaxed and strict). Bayes factor tests were used to discriminate between the competing models.

Maximum likelihood analyses

Maximum likelihood (ML) analyses were conducted using IQ-TREE (Nguyen et al. 2015; Minh et al. 2020). The datasets in Paper II were analyzed both as partitioned and unpartitioned. For the partitioned analyses we used ModelFinder (Kalyaanamoorthy et al. 2017) to search for an optimally partitioned model and for the unpartitioned analyses the GTR+G model was used. For the concatenated analyses in Paper III the data were analyzed as partitioned, treating each gene alignment as a separate partition, applying the GTR+G model to each partition. The individual single gene alignments were also analyzed using the GTR+G model. In Paper IV, we conducted unpartitioned analyses using automatic model selection (Kalyaanamoorthy et al. 2017). For all analyses branch support was evaluated using ultrafast bootstrapping (Hoang et al. 2018). In Paper II the SH-like approximate likelihood ratio test (Guindon et al. 2010) and the approximate Bayes test (Anisimova et al. 2011) were also used.

Multispecies coalescent analyses

To accommodate gene tree incongruences owing to ILS, species trees were estimated using the quartet-based summary method ASTRAL III v5.7.8 (Zhang et al. 2018) (Paper III). Since collapsing branches with very low support can help improve the accuracy of species tree inference using summary methods (Zhang et al. 2018), branches with support less than 20% were collapsed before the gene trees were fed to ASTRAL. Branch support was evaluated by local posterior probability (Sayyari and Mirarab 2016). We also used the ASTRAL polytomy test to investigate whether polytomies can be rejected at short branches by using quartet gene tree frequencies (Sayyari and Mirarab, 2018).

Evaluation of incongruence

Evaluation of incongruence between results from different datasets and analyses was primarily done by inspection of tree topologies, assuming conflict if alternative resolutions were considered supported as assessed from commonly used support thresholds such as at least 95% posterior probabilities and ultra-fast bootstrap supports (Guindon et al. 2010; Anisimova et al. 2011; Minh et al. 2013; Sayyari and Mirarab 2016). In Paper II, we used the hierarchical likelihood ratio test implemented in Concaterpillar 1.8a (Leigh et al. 2008) to test for intragenomic incongruence within the plastome, mitogenome, and nuclear ribosomal cistron. In Paper III, to examine gene tree incongruence, ASTRAL trees were annotated with quartet frequencies for alternative topologies, and the normalized quartet score, which indicates the proportion of gene tree quartets contained in the species tree and part of the ASTRAL output was used to quantify the extent of gene tree discordance. In Paper IV we tested whether the observed incongruence between nuclear- and plastid-derived trees produced during work with Papers III and IV could be explained by ILS alone, using coalescent simulation analyses (following several previous studies, Folk et al. 2017; García et al. 2017; Zhou et al. 2022). A distribution of plastome gene trees that would be expected under ILS alone was created based on simulations conducted under a coalescent model using a nuclear species tree from Paper III with branch lengths adjusted to account for organellar inheritance as a guide tree.

Main results and general discussion

The first part of this thesis work (Papers I and II) was focused on the tribe Anthospermeae. Phylogenetic analyses of selected plastid and nuclear ribosomal markers and a dense sample of taxa provided new insights into the current generic and subtribal delimitation within the tribe (Paper I). Based on the results in Paper I the previous subtribal classification was partly updated. In Paper II the taxon sampling was decreased but the sampling of characters was much increased, comprising phylogenomic data from the plastome, mitogenome and nuclear ribosomal cistron. The phylogenetic results corroborated and clarified most of the intergeneric relationships found in Paper I, including the updated subtribal classification and the deepest splits within the tribe. However, in both Papers I and II several instances of supported phylogenetic incongruences were detected. This included cytonuclear discordance (Paper I and II), organellar discordance (Paper II), and intra-plastome discordance (Paper II). While ILS and hybridization were generally considered as the most likely causes, their relative contributions could not be distinguished based on the data at hand. The phylogenomic approach taken in Paper II also resulted in the detection of large plastome inversions (a 36-kb and a 19-kb inversion in the genus *Coprosma*) and (putative) mitochondrial-to-plastid DNA transfer (in the Pacific clade), which are both considered relatively rare events.

The second part of the thesis work (Papers III and IV) took advantage of some of the recently developed methods for producing and analyzing large nuclear and plastome datasets to improve the current understanding of phylogenetic relationships within Rubioideae. It was demonstrated that using hundreds of presumably independently evolving nuclear genes and large amounts of plastome data can be hugely effective in revealing parts of the Rubioideae phylogeny that historically have proven difficult to resolve. All or nearly all intertribal relationships were highly supported in analyses based on nuclear data (Paper III) and plastome data (Paper IV). Our results clarified and corroborated previous hypotheses of relationships within the group and novel insights of relationships from both the nuclear and plastome perspectives were added. Furthermore, the hundreds of nuclear genes that were newly produced for a substantial set of species also allowed for a more comprehensive understanding of evolutionary events that have taken place in the history of the group. For example, in Paper III it was indicated that ILS contributed to most of the gene tree discordance at short internodes, suggesting rapid speciation at several parts of the Rubioideae tree. And the use of species tree and plastome tree estimates inferred in Papers III and IV allowed for detection and investigation of cytonuclear discordance. While both sources of data often resulted in concordant tree topologies, there were also some discordances, both among and within tribes. The analytical results indicate that ILS alone can explain most of the observed cytonuclear discordances, but some cases may be better explained by hybridization followed by plastid capture. Rubioideae plastomes are predominately highly conserved and mainly conform to the canonical structure, gene content, and gene order consistently found in the majority of angiosperm species.

Congruence and incongruence

In all papers included in this thesis, we compared phylogenetic results obtained using different datasets and analyses. Using such a pluralistic approach can help identifying relationships that are dependent on a specific type of data, or method, and may elucidate whether the observed differences are due to biological or technical reasons. Overall there were few indications that supported incongruence would be due to technical reasons, such as systematic error or sampling error. For example, data partitioning, data recoding, and the use of clock versus non-clock models, had little effect on topological results, yielding highly similar trees. Incongruent relationships, if any, were not supported.

Similarly, the topological results in Paper III were highly similar regardless of method (coalescence- or concatenation-based), data type (CDS or supercontigs) and inclusion/exclusion of potentially paralogous genes, and most differences were as a rule not highly supported. We found that the combined information in the CDS + noncoding regions (supercontigs) alignments and inclusion of all genes (i.e., inclusion of putatively paralogous genes) led to increased statistical support. Therefore, the preferred phylogenetic results were those from the most inclusive dataset.

There were no topological conflicts between the concatenation-based and coalescence-based analyses of that dataset, with the exception of one notable intertribal relationship. The coalescent-based analysis resolved Ophiorrhizeae sister to Urophylleae sensu stricto (s.s.) (i.e., Urophylleae excluding *Temnopteryx*), whereas the concatenation-based analysis resolved Ophiorrhizeae sister to the Seychelleeae – Colletoecemateae clade. This part of the tree includes successive rather short branches, indicating high levels of ILS, which is a characteristic pattern of the so called anomaly zone (Degnan and Rosenberg 2006). Unlike coalescent-methods, the concatenation-method can return severely misleading results in or near the anomaly zone (Kubatko and Degnan 2007;

Edwards et al. 2016; Mendes and Hahn 2018). Thus, differences in the underlying assumptions between the two approaches might account for the deviating placements of Ophiorrhizeae.

On the gene-tree level, the addition of noncoding regions increased overall gene tree congruence, suggesting that gene tree error due to sampling error contributes to more of the gene tree heterogeneity in the smaller CDS datasets than in the extended supercontig datasets. The positive effect on support and gene tree concordance when adding noncoding regions to individual nuclear gene alignments has been shown also in previous studies (Jones et al. 2019; Bagley et al. 2020; Gardner et al. 2021; Thomas et al. 2021).

Nevertheless, it is worth noting that the coalescence-based results retrieved from the paralog-filtered CDS dataset and the supercontig dataset had one supported intertribal conflict. While the results yielded from the supercontig dataset support a Danaideae+Knoxieae relationship, the results based on the CDS dataset support a Knoxieae+Spermacoceae relationship. The latter relationship is supported in all other analyses conducted in Paper III and is also well established based on analyses of organellar and nuclear ribosomal data. The quartet gene tree frequencies at the discordant branch in the CDS-based tree were indicative of ILS with rather well-matched frequencies between the two minor topologies. However, the quartet gene tree frequencies of the minor topologies at the discordant branch in the supercontig-based tree were not well matched, indicating that other factors than ILS are at play (Paper III).

It is possible that the highly variable noncoding regions included in the supercontig dataset induce gene tree estimate error because of noise in this part of the phylogeny, which would explain why the patterns of quartet frequencies differ between analysis of CDS data and supercontig data. Another possible explanation for the unbalanced quartet frequencies might be that, after a putative hybridization event, introgression is biased toward noncoding regions. Since the corresponding concatenated analysis of the paralog-filtered supercontig dataset recovered the (presumably) correct Knoxieae+Spermacoceae relationship, it could be argued that the coalescent analysis is misled by the high degree of gene tree error in this particular case.

Most of the instances of supported discordance across the Rubioideae phylogeny detected during the work with this thesis were between results retrieved from analyses of different genomes, mainly between organellar and nuclear data. Given factors such as highly similar (or even identical) taxon sampling and robustness to alternative analyses of the different genomic compartments, we have argued that the topological conflicts are likely due to biological, rather than technical causes. Cytonuclear discordance is commonly attributed to organellar introgression and ILS, but relatively few studies have attempted to distinguish between these two phenomena. The results from the coalescent simulation analysis conducted in Paper IV indicated that ILS alone may explain the majority of reported cytonuclear conflicts, although to various degrees. Several of those examples and other cases of discordance are described in more detail in the following section.

Relationships

The deepest splits

The deepest divergences in Rubioideae have often been resolved differently in previous phylogenetic studies, frequently including alternative placements of the tropical African tribe Colletoecemateae, the Australasian tribe Ophiorrhizeae and the pantropical tribe Urophylleae. The relationships among those tribes have differed between analyses of different genomes, but also between different studies using plastid data. Recently, the monospecific seychellean tribe Seychelleeae was found to be the sister-taxon to the species-poor monogeneric tribe Colletoecemateae. In Papers III and IV (respectively based on same sample nuclear data and plastome data), a clade comprising the sampled members of the four aforementioned tribes referred to as the SCOUT clade is consistently recovered as the sister group to the remaining subfamily Rubioideae. This result agrees with previous results based on nuclear data, but not with previous results based on organellar data.

Compared to previous studies based on plastid data, the increased congruence between the plastid-based trees in Paper IV with the nuclear-based species tree (Paper III) may lie in the increased sampling of characters and taxa (Paper IV). In comparison to earlier phylogenomic studies, the taxon sampling employed in Paper IV was much increased, making it more comparable to taxon-rich Sanger datasets used in the literature. Furthermore, unlike earlier phylogenomic studies, which mostly focused on information from coding regions, the data analyzed in Paper IV included all coding and noncoding regions. While noncoding regions have not been as commonly used as the overall more conserved coding regions in phylogenetic studies involving relatively deep relationships, their utility for resolving relationships at both deep and shallow levels have been shown in several previous studies (Bremer et al. 2002; Chen et al. 2017; Fu et al. 2017; Givnish et al. 2018; Literman and Schwartz 2021).

Regarding the internal phylogeny of the SCOUT clade, the coalescent-based and concatenation-based trees in Paper III showed strong support for all relationships. Both trees resolved the African monospecific genus *Temnopteryx* (Fig. 1B) of the Urophylleae sensu Smedmark et al. (2008) as the sister to the remaining clade, and both trees resolved the Seychelleeae – Colletoecemateae clade although the relationships among the Seychelleeae – Colletoecemateae clade, Ophiorrhizeae, and Urophylleae s.s. differed between the two trees, as described above.

Furthermore, while *Temnopteryx* is strongly supported as sister to the remaining SCOUT clade based on nuclear data, the analyses of plastid data (Paper IV) resolve *Temnopteryx* as the highly supported sister to the (remaining) tribe Urophylleae, thus corroborating results in several previous studies based on (predominately) plastid data (Smedmark et al. 2008, 2010; Smedmark and Bremer 2011; Yang et al. 2016). The coalescent simulations conducted in Paper IV indicated that this case of cytonuclear discordance could be explained by ILS alone rather than hybridization and subsequent plastome capture.

The next divergence along the Rubioideae backbone involves the split between the Lasiantheae-Perameae clade and a clade comprising the Psychotrieae and Spermacoceae alliances and the tribe Coussareeae. This result is congruent with those of most previous molecular studies. However, the position of the morphologically heterogeneous Neotropical tribe Coussareeae differs between results based on nuclear data and plastome data. Nuclear data (Paper III) resolve Coussareeae as sister to the Spermacoceae alliance, whereas plastome data (Paper IV) resolve the Coussareeae as sister to a clade comprising both the Spermacoceae alliance and the Psychotrieae alliance. This case of cytonuclear discordance is consistent with comparisons of previous studies based on nuclear ribosomal cistron data versus organellar data (Wikström et al. 2020). The pattern of gene tree discordance at the short Coussareeae + Spermacoceae alliance branch is consistent with high levels of ILS (Paper III). The results from the coalescent simulation analysis (Paper IV), showed that the empirical result based on plastome data (a Psychotrieae alliance + Spermacoceae alliance relationship) is expected even when accounting for organellar inheritance (roughly 29% of the simulated trees contained this relationship), clearly indicating that this case of cytonuclear discordance can be explained by ILS alone. In sum, these results suggest that the divergence events that separated the three lineages the Coussareeae and the large Spermacoceae and Psychotrieae alliances occurred within a short time frame, leading to extensive ILS.

Psychotrieae alliance

The nuclear and plastome datasets used in Papers III and IV included representatives of all nine currently recognized tribes of the Psychotrieae alliance and demonstrated, unlike previous studies based on Sanger-data (e.g., Razafimandimbison et al. 2008), strong support for all intertribal relationships. The results based on nuclear data agree with earlier phylogenomic estimates based on nuclear data (Wikström et al. 2020; Antonelli et al. 2021), and also those retrieved from mitochondrial data (Rydin et al. 2017), although Schradereae and Mitchelleae were not included in Antonelli et al. (2021), and Schradereae was not included in Wikström et al. (2020) and Rydin et al. (2017). The results based on plastome data largely agree with those of earlier plastid analyses, and detected differences may be explained by differences in sampling (characters and taxa). Within the Psychotrieae alliance, the results retrieved from nuclear data (Paper III) resolve Schizocoleeae and Craterispermeae as subsequent sisters to the remaining clade, which in turn is divided into two major clades (1) a Gaertnereae + Psychotrieae-Palicoureeae clade and (2) a clade where Schradereae and Prismatomerideae are subsequent sisters to the Morindeae – Mitchelleae clade. The results retrieved from plastid data (Paper IV) largely agree but place Gaertnereae (Fig. 2D) as sister to the Morindeae – Mitchelleae clade. Overall, our results indicate that rapid divergences and relatively high levels of ILS have been common in the early history of the Psychotrieae alliance (Paper III). Interestingly, the cytonuclear discord involving the placement of Gaertnereae may, however, be best explained by an ancient hybridization event leading to plastome capture (Paper IV).

Spermacoceae alliance

Relationships in the Spermacoceae alliance have been difficult to resolve, with relationships often being poorly supported or incongruent. In contrast, in Papers III and IV all intertribal relationships were completely congruent and highly supported. The results from Papers III and IV support a Danaideae-Knoxieae-Spermacoceae clade as sister taxon to the remaining members of the alliance which is congruent with results based on phylogenomic plastid and mitochondrial data (Rydin et al. 2017; Wikström et al. 2020), and plastid (sometimes together with nrITS) Sanger datasets (Rydin et al. 2009a; Wikström et al. 2015). However, several other plastid (sometimes together with nrITS) Sanger datasets have instead supported Danaideae as sister to the remaining Spermacoceae alliance (Bremer and Manen 2000; Bremer and Eriksson 2009; Rydin et al. 2009b; Yang et al. 2016).

Previous analyses of nuclear data have resulted in some rather unexpected results. Based on nuclear ribosomal data Anthospermeae was sister to the Knoxieae-Spermacoceae clade with Danaideae found in a clade containing the other sampled members of the alliance (Wikström et al. 2020), and based on Angiosperms353 data, Argostemmateae followed by the tribe Spermacoceae were supported as consecutive sisters to the remaining members of the alliance (Antonelli et al. 2021). The sparse taxon sampling in the latter study, for example the Danaideae-Knoxieae-Spermacoceae clade was only represented by one species of Spermacoceae, may explain the different placements of these tribes compared to the results presented in this thesis. The phylogenetic trees presented here (both based on nuclear, Paper III, and plastome, Paper IV, data) support Anthospermeae, the Dunnieae-Cyanoneuroneae clade and Argostemmateae as successive sisters to a Paederieae-Rubieae complex (here Putorieae-Rubieae-Theligoneae) clade, a result concordant with the plastid-based results in Wikström et al. (2020). This phylogenetic position of the Anthospermeae has often been well supported based on plastid (sometimes combined with nrITS) Sanger datasets (Rydin et al. 2009b, 2009a; Wikström et al. 2015; Yang et al. 2016), and the Paederieae-Rubieae complex relationship has been supported in several studies that have used plastid and nuclear data, either alone or in combination (Rydin et al. 2009b, 2009a; Wikström et al. 2015, 2020; Yang et al. 2016; Antonelli et al. 2021). Other relationships among these groups have been either poorly supported or incongruent among analyses, including findings from previous studies based on nuclear data (Wikström et al. 2020; Antonelli et al. 2021), as described above, and analysis of mitochondrial data, which supported an Anthospermeae + Dunnieae clade as sister to a clade in which Paederieae was sister to a Argostemmateae + Rubieae complex clade (Rydin et al. 2017).

The results in Papers III and IV support a clade comprising the recently described small Bornean and Sulawesian tribe Cyanoneuroneae (Ginter et al. 2015) and the small Chinese and relatively recently described tribe Dunnieae (Rydin et al. 2009b). A third tribe, the monogeneric Chinese tribe Foonchewieae (Wen and Wang 2012), may also belong in this clade. Foonchewieae formed a clade with Cyanoneuroneae and Dunnieae in Ginter et al. (2015) based on plastid + nuclear (nrETS and nrITS) data, and plastid data have supported a Foonchewieae + Dunnieae clade (Wen and Wang 2012; Wikström et al. 2015; Yang et al. 2016). In Paper I representatives from all three tribes were included and they were, based on plastid data, found in a clade comprising also Argostemmateae. Most other studies have typically only sampled one of the three tribes, including the phylogenomic studies by Rydin et al. (2017) and Wikström et al. (2020), which included Dunnieae, and Antonelli et al. (2021), which included Cyanoneuroneae. A surprising result in Antonelli et al. (2021) was the placement of Cyanoneuroneae in the Psychotrieae alliance based on plastid data. That result is likely an artifact associated with limited sampling (Paper IV).

Within the Rubieae complex our results support a Theligoneae + Putorieae clade, together sister to Rubieae, corroborating previous results based on nuclear data (Wikström et al. 2020; Antonelli et al. 2021) and plastid data or a combination of plastid and nrITS data (Yang et al. 2016; Antonelli et al. 2021; Rincón-Barrado et al. 2021). However, several other plastid and plastid + nrITS analyses, have instead supported a Theligoneae + Rubieae clade (e.g., Backlund et al., 2007; Bremer and Eriksson, 2009; Rydin et al., 2009b; Deng et al., 2017; Ehrendorfer et al., 2018; Wikström et al., 2020), congruent with analyses of mitochondrial data (Rydin et al., 2017). Although apparent morphological support for the Theligoneae + Rubieae clade is lacking, there are some shared morphological features between some Putorieae members and some taxa within Rubieae (Natali et al. 1995; Ehrendorfer et al. 2018). The recently described monospecific tribe Aitchinsonieae also belong in the Rubieae complex (Bordbar et al. 2021). The tribe was described to accommodate the single species *Plocama rosea* (Hemsl. ex Aitch.) M.Backlund & Thulin (= *Aitchisonia rosea* Hemsl. ex Aitch.), which grouped with Rubieae based on plastid *trnL-F* data. The remaining members of *Plocama*/Putorieae and Theligoneae were subsequent sister groups to this clade (Bordbar et al. 2021).

The tribe Anthospermeae – an atypical Rubiaceae lineage

Papers I and II focused on the tribe Anthospermeae. Paper I included a broad sample of taxa, including representatives from all twelve currently recognized genera of the tribe. The results in Paper I consistently supported the monophyly of the tribe with the South African genus *Carpacoce* sister to remaining species, which were divided into an African clade comprising the genera *Anthospermum* (Fig. 1E), *Galopina*, *Nenax* and *Phyllis* (Fig. 1C), and a Pacific clade comprising the genera *Coprosma*, *Durringtonia*, *Leptostigma*, *Nertera*, *Normandia*, *Pomax* and *Opercularia* (Fig. 2E). The phylogenetic relationships were, however, not compatible with the traditional subtribal classification of Anthospermeae, (Anthosperminae, Coprosminae, and Operculariinae; Puff 1982), which was based primarily on floral and fruit characters. Paper I revealed that the two genera *Anthospermum* and *Nenax* are not monophyletic since *Nenax* species were found intermixed with *Anthospermum* species in the *Anthospermum-Nenax* clade. A revised subtribal classification of the tribe was proposed, including the description of the new monogeneric tribe Carpacocinae, and an emended classification of Anthosperminae (*Anthospermum*, *Galopina*, *Nenax* and *Phyllis*) (Paper I). Other consistently supported relationships included an *Anthospermum-Nenax* clade, a *Durringtonia-Coprosma* (i.e., *Durringtonia* + *C. moorei* and *C. talbrockiei*), and a *Nertera-Coprosma* s.s. (i.e., all species of *Coprosma* except *Coprosma moorei* and *Coprosma talbrockiei*) clade. The subsequent phylogenomic studies (i.e., Papers II, III, IV) have consistently yielded results congruent with these sets of relationships, although *C. moorei* and *C. talbrockiei* were not included in those studies, and *Nenax* was not included in Papers III and IV.

Phylogenetic discordance is extensive within Anthospermeae. Across all papers included in the present thesis (I-IV) several instances of supported phylogenetic incongruences were detected, both within and among genera. This included cytonuclear discordance, organellar discordance, and intra-plastome discordance. When comparing with other relatively well sampled tribes (at the generic level) the cytonuclear discordance within the tribe was rivalled only by Argostemmateae (Papers III-IV). Coalescent simulation analyses conducted in Paper IV indicated that the observed intergeneric cytonuclear discordance within the Anthospermeae (i.e., topological discordance between results retrieved from plastome data and nuclear data) could be attributed to ILS alone in all cases, albeit to various degrees.

While the here observed discrepancies between mitochondrial and plastome data in Paper II was unexpected given that both these genomes are generally expected to be maternally inherited, such organellar conflicts have been reported previously and both ILS and hybridization have been suggested as possible explanations (Folk et al. 2017; Rydin et al., 2017).

Intra-plastome discordance, i.e., incongruence between results retrieved from different subsets of the plastome data, was more unexpected since the plastome is typically regarded to be uniparentally (usually maternally) inherited and evolve as single unit and is commonly treated as a single locus in phylogenetic analyses. However, several recent studies have reported high levels of gene tree discordance among plastome loci, and have suggested that the traditional treatment of the plastome may not always be optimal and within-plastome discordance should be more thoroughly explored (Gonçalves et al. 2019; Walker et al. 2019; Zhang et al. 2020). The hierarchical ratio test conducted in Paper II indicated two supported conflicting signals within the Anthospermeae plastome data. The cause for these phylogenomic discords is not clear but whatever the cause, it shows that plastome phylogenies can be highly sensitive to alternative choice of markers.

In addition to the high degree of phylogenetic incongruence within Anthospermeae, the tribe also exhibits some unusual plastome features. Particularly noteworthy are two large inversions (36-kb and 19-kb) found in *Coprosma*, and a (putative) transfer of mitochondrial DNA into the plastome of the members of the Pacific clade (Paper II). Intriguingly, it was found that conflicting phylogenetic signal in the Pacific clade is, at least partly, associated with the exclusion/inclusion of this lineage-specific insertion of putative mitochondrial origin (Paper II and IV). Sequencing of additional species from those groups reinforced that these features are possible synapomorphies for the respective groups (Paper IV). Both large inversions and mitochondrial-to-plastid DNA transfer are considered to be relatively rare events (Smith 2014; Gandini and Sanchez-Puerta 2017; Charboneau et al. 2021). For example, it was not long ago that the plastome was thought to be more or less immune to the uptake of foreign DNA (Smith 2011).

Future studies

In a sense this thesis has followed the same trajectory that the broad systematic community has over the last decade or so, transitioning from a relatively data-sparse environment into a data-rich environment, driven by the progressive development of both molecular and computational methods. The benefits of employing these considerably larger datasets have been demonstrated to be highly valuable in this thesis' work. For example, it was shown that using hundreds of presumably independent nuclear genes and large amounts of plastome data can be highly efficient in resolving longstanding questions of the Rubioideae phylogeny. Utilizing these big datasets has also allowed for a more comprehensive understanding of the root of these problems, which include evolutionary events such as ILS and hybridization during the history of the group. The phylogenomic approach has also provided new insights into plastome evolution in the group.

Approaches like those used here would most probably be useful when addressing similarly difficult phylogenetic questions in the remaining coffee family. Recent studies based on nuclear genome-scale data have shown that all remaining species of the coffee family constitute the sister group to subfamily Rubioideae (Antonelli et al. 2021; Paper III). However, relationships within this Ixoroideae-Cinchonoideae-Luculieae-Coptosapelteae clade are only partly understood, for example with conflicting topologies indicated in previous work (Rydin et al. 2017; Wikström et al. 2020). Further studies of this clade using similar approaches as those used here, would most probably provide additional clarification.

Furthermore, there are many remaining questions concerning the evolutionary history of Rubioideae. The coalescent simulation approach in Paper IV was useful for characterizing instances of cytonuclear discordance. However, to deepen the understanding of the role of ILS and hybridization in the evolutionary history of Rubioideae, their relative contributions should be investigated more specifically for individual clades affected by cytonuclear discordance. Extended sampling of taxa and applying additional approaches such as species networks and other statistical tests used on nuclear datasets (Yu et al. 2012; Edelman et al. 2019) would possibly provide more insight into this matter. For example, it would be interesting to investigate if the invoked putative introgression scenario regarding the placement of Gaertnereae (Paper IV) would be supported using such approaches.

Once the phylogeny of a group has been settled, a framework for many biological studies has been established. Further, taxonomic work important for assessment and communication of biodiversity should be based on phylogenetic results. Taxonomic revision will be needed as a consequence of some of the results contributed in the present thesis. One example is provided by the work with Paper I, which was based on roughly half of the species of the two closely related genera *Anthospermum* and *Nenax* (Anthospermeae). The results of the study indicated that these two genera should be merged into one genus. This was, however, not done because many interspecific relationships were not resolved, and several incongruences were found between the nuclear ribosomal and plastome phylogenies. A denser taxon sampling and phylogenomic nuclear and organellar sequence data could help shed more light on the evolutionary history of this group and the monophyletic status of the two genera.

Svensk sammanfattning

Systematisk biologi, eller systematik, kan definieras som studier av biologisk mångfald, evolution och släktskapsrelationer mellan organismer. Mer specifikt så inkluderar systematik två huvudgrenar, (i) taxonomi, vetenskapen om att upptäcka, beskriva, namnge och klassificera organismer och (ii) fylogeni, vetenskapen om evolutionära relationer mellan organismer. De flesta moderna klassifikationssystem är grundade på analys av evolutionära släktskapsförhållanden, vilket förenar de två stora grenarna av systematiken.

Numera använder systematiker ofta information från DNA-sekvenser för att förstå evolutionära händelser och undersöka hur olika organismer är släkt med varandra. DNA ärvs från generation till generation, men kan också förändras genom mutationer. Genom att jämföra likheter och olikheter i DNA-sekvenserna kan man få fram ett utvecklingsträd (fylogeni). Ett sådant utvecklingsträd, eller släktträd, representerar en hypotes om hur olika organismer är släkt.

Innan tekniken för DNA-analyser blev allmänt tillgänglig för forskarsamhället för några decennier sedan baserades släktskapsanalyser vanligen på morfologi, det vill säga utseendemässiga egenskaper. Ibland stämmer sådana utvecklingsträd överens med hypoteser baserade på DNA, och ibland inte. Analyser av olika delar av växters DNA kan också resultera i olika släktskapshypoteser. Hos växter är det till exempel ganska vanligt att fylogenier baserade på data från det (oftast) maternellt nedärvda kloroplast-genomet och fylogenier baserade på det biparentalt nedärvda nukleära genomet skiljer sig åt.

Skillnader mellan hypoteser om släktskap baserade på olika genetiska dataset kan ha många orsaker. Till exempel kan hybridisering ha gjort att DNA hos en individ innehåller en mosaik av data från två föräldraarter. Det kan också bero på något som kallas ofullständig linjesortering (av gener), ett fenomen som har sin grund i stokastiska populationsgenetiska processer. Förenklat kan man säga att det kan finnas tidsmässiga skillnader mellan arters evolution och geners evolution; att gener inte alltid diversifierar och sorteras upp i olika evolutionära linjer samtidigt som artbildning sker.

Motsägelsefulla resultat mellan olika släktskapsträd kan också bero på att en för enkel evolutionär modell har använts i analyserna, eller att en otillräcklig mängd data har använts. I och med den snabba utvecklingen av sekvenseringsmetoder och utveckling av nya analysmetoder finns det nu goda möjligheter att omvärdera tidigare fylogenetiska hypoteser.

Den här avhandlingen fokuserar på en grupp växter som ingår i familjen Rubiaceae (måreväxter, krappväxter, eller kaffeväxter på svenska), en familj med stor artrikedom och morfologisk mångfald. De flesta arter i denna familj lever i tropikerna men arter finns spridda över i stort sett hela jordklotet. Till familjen kaffeväxter hör förutom kaffe även kinaträd, måror och många prydnadsväxter såsom *Gardenia*. Växterna i kaffefamiljen är ofta relativt lätta att känna igen med sina motsatta, enkla och hela blad, interpetiolära stipler och översittande hylle.

Majoriteten av arterna i kaffefamiljen ingår i underfamiljen Rubioideae, vilken också är fokus för denna avhandling. Medan kaffefamiljen består främst av vedartade växter så är Rubioideae den enda av underfamiljerna som även inkluderar stora grupper av örtartade växter. Några av grupperna i Rubioideae finns även i tempererade områden vilket annars är ovanligt inom kaffefamiljen. Vindpollinerade blommor är också ett ovanligt drag i kaffefamiljen men finns hos ett fåtal grupper i Rubioideae.

Rubioideae är indelad i 29 olika tribus (grupp av en till flera närbesläktade släkten). Majoriteten av dessa tribus ingår i någon av de två stora informella grupperna Psychotrieae-alliansen (nio tribus) och Spermacoceae-alliansen (13 tribus), vardera med över 3000 arter. Psychotrieae-alliansen har pantropisk utbredning och består övervägande av vedartade arter. Spermacoceae-alliansen förekommer mer globalt och består övervägande av örtartade arter.

Det övergripande syftet med detta avhandlingsarbete har varit att ta fram en riklig mängd nya data i form av DNA-sekvenser för att undersöka evolutionära relationer inom gruppen; både hur underfamiljens olika tribus är släkt med varandra, och även släktskap inom vissa tribus har undersökts. Trots många fylogenetiska studier under de senaste decennierna har osäkerheten om flera evolutionära relationer i denna grupp kvarstått, bland annat på grund av att olika studier visat konflikterande resultat och/eller att resultaten haft lågt statistiskt stöd.

I den här avhandlingen användes både traditionella molekylära metoder och den nya generationens metoder för att erhålla sekvensdata, vilket väsentligt utökade mängden sekvensdata tillgängligt för fylogenetisk analys i Rubioideae.

Att utnyttja stora mängder data från ett omfattande urval av arter visade sig vara mycket effektivt för att lösa fylogenetiska relationer i Rubioideae, inklusive de historiskt mest motsträviga frågorna.

Nästan alla (eller alla) relationer mellan olika tribus fick högt statistiskt stöd, både i analyser baserade på nukleära data och på kloroplast data. Dessutom medförde analys av hundratals nukleära gener och stor mängd plastid data för en relativt stor uppsättning arter också en mer omfattande förståelse av evolutionära händelser som har ägt rum i gruppens historia. Resultaten indikerade att perioder av snabb artbildning varit ganska vanligt i gruppens historia, vilket troligtvis också är en starkt bidragande orsak till att vissa relationer inom gruppen tidigare har varit svåra att lösa.

Medan separata analyser av plastid data och nukleära data ofta resulterade i överensstämmande fylogenetiska hypoteser fanns det också olikheter. Vi fann få eller inga tydliga indikationer på att konflikterande resultat baserade på olika genom har orsakats av metodologiska svagheter eller andra mer analystekniska brister. Istället verkar det som om verkliga biologiska händelser i växternas evolutionära historia kan förklara varför vissa släktskapsrelationer skiljer sig åt beroende på vilka data som analysen baserats på. Resultaten tydde på att ofullständig linjesortering kan förklara de flesta inkongruenserna, men hybridisering verkar vara mer sannolikt i åtminstone ett fall.

En tribus, Anthospermeae, studerades i mer detalj. De fylogenetiska analyserna påvisade flera fall och flera typer av inkongruens. Dock var de djupaste förgreningarna i gruppens släktträd kongruenta mellan de olika analyserna. Resultaten var däremot inte helt överensstämmande med den traditionella indelningen i undertribus, baserad på morfologi. Därför gjordes en uppdatering av klassifikationen av gruppen i samband med studien.

I Anthospermeae hittades också relativt ovanliga evolutionära förändringar av kloroplastens genom i form av stora inversioner och (förmodad) horisontell överföring av genetiskt material från mitokondriens genom till kloroplastens genom. Överlag är annars kloroplastens genom hos växterna i Rubioideae konserverade med en organisation, ett geninnehåll och en genordning som stämmer väl överens med den som finns hos majoriteten av andra blomväxter.

References

Alvarez I., Wendel J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417–434.

Anderson C.L., Rova J.H.E., Andersson L. 2001. Molecular phylogeny of the tribe Anthospermeae (Rubiaceae): systematic and biogeographic implications. Aust. Syst. Bot. 14:231–244.

Andersson, L., Rova, J.H.E. 1999. The *rps16* intron and the phylogeny of the Rubioideae (Rubiaceae). Plant Syst. Evol. 214:161–186

Andreasen K., Bremer B. 2000. Combined phylogenetic analysis in the Rubiaceae-Ixoroideae: morphology, nuclear and chloroplast DNA data. Am. J. Bot. 87:1731–1748.

Anisimova M., Gil M., Dufayard J.-F., Dessimoz C., Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60:685–699.

Antonelli A., Clarkson J.J., Kainulainen K., Maurin O., Brewer G.E., Davis A.P., Epitawalage N., Goyder D.J., Livshultz T., Persson C., Pokorny L., Straub S.C.K., Struwe L., Zuntini A.R., Forest F., Baker W.J. 2021. Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. Am. J. Bot. 108:1143–1165.

Backlund M, B Bremer, M Thulin 2007 Paraphyly of Paederieae, recognition of Putorieae and expansion of *Plocama* (Rubiaceae-Rubioideae). Taxon 56:315–328.

Bagley J.C., Uribe-Convers S., Carlsen M.M., Muchhala N. 2020. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Burmeistera* bellflowers as a case study. Mol. Phylogenet. Evol. 152:106769.

Baldwin B.G., Markos S. 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S–26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). Mol. Phylogenet. Evol. 10:449–463.

Bordbar F., Mirtadzadini M., Razafimandimbison S.G. 2021. Phylogenetic re-assessment of the delimitation of *Plocama* and its species relationships and limits (Rubiaceae, Putorieae): resurrection of the monospecific genus *Aitchisonia* and a description of trib. nov. Aitchisonieae. Plant Syst. Evol. 308:7.

Borowiec M. 2019. Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. JOSS. 4:1635.

Braun E.L., Kimball R.T. 2021. Data types and the phylogeny of Neoaves. Birds. 2:1–22.

Bremekamp C.E.B. 1952. The African species of *Oldenlandia* L. sensu Hiern et K. Schumann. Verh. Kon. Ned. Akad. Wetensch., Afd. Natuurk., Sect. 2. 48 (2): 1-297.

Bremekamp C.E.B. 1966. Remarks on the position, the delimitation and the subdivision of the Rubiaceae. Acta Bot. Neerl. 15:1–33.

Bremer B., Andreasen K., Olsson D. 1995. Subfamilial and tribal relationships in the Rubiaceae based on *rbcL* sequence data. Ann. Mo. Bot. Gard. 82:383.

Bremer B., Bremer K., Heidari N., Erixon P., Olmstead R.G., Anderberg A.A., Källersjö M., Barkhordarian E. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Mol. Phylogenet. Evol. 24:274–301.

Bremer B., Eriksson T. 2009. Time tree of Rubiaceae: phylogeny and dating the family, subfamilies, and tribes. Int. J. Plant Sci. 170:766–793.

Bremer B., Manen J.-F. 2000. Phylogeny and classification of the subfamily Rubioideae (Rubiaceae). Plant Syst. Evol. 225:43–72.

Cantley J.T., Markey A.S., Swenson N.G., Keeley S.C. 2016. Biogeography and evolutionary diversification in one of the most widely distributed and species rich genera of the Pacific. AoB Plants. 8:plw043.

Cantley J.T., Swenson N.G., Markey A., Keeley S.C. 2014. Biogeographic insights on Pacific *Coprosma* (Rubiaceae) indicate two colonizations to the Hawaiian Islands: Biogeographic insights on Hawaiian *Coprosma*. Bot. J. Linn. Soc. 174:412–424.

Capella-Gutierrez S., Silla-Martinez J.M., Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.

Charboneau J.L.M., Cronn R.C., Liston A., Wojciechowski M.F., Sanderson M.J. 2021. Plastome structural evolution and homoplastic inversions in Neo-*Astragalus* (Fabaceae). Genome Biol. Evol. 13:evab215.

Chen M.-Y., Liang D., Zhang P. 2017. Phylogenomic resolution of the phylogeny of laurasiatherian mammals: exploring phylogenetic signals within coding and noncoding sequences. Genome Biol. Evol. 9:1998–2012.

Cox C.J., Li B., Foster P.G., Embley T.M., Civáň P. 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. 63:272–279.

Criscuolo 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 Evol. Biol. 10:210.

Cui L., Wall P.K., Leebens-Mack J.H., Lindsay B.G., Soltis D.E., Doyle J.J., Soltis P.S., Carlson J.E., Arumuganathan K., Barakat A., Albert V.A., Ma H., dePamphilis C.W. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16:738–749.

Darling A.C.E., Mau B., Blattner F.R., Perna N.T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394–1403.

Darling A.E., Mau B., Perna N.T. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 5:e11147.

Davis, A. P., Govaerts, R., Bridson, D. M., Ruhsam, M., Moat, J., and Brummitt, N. A. (2009). A global assessment of distribution, diversity, endemism, and taxonomic effort in the Rubiaceae. Ann. Mo. Bot. Gard. 96:68–78.

Defoort J., Van de Peer Y., Carretero-Paulet L. 2019. The Evolution of gene duplicates in angiosperms and the impact of protein–protein interactions and the mechanism of duplication. Genome Biol. Evol. 11:2292–2305.

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68.

Delsuc F., Brinkmann H., Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6:361–375.

Deng, T., Zhang, J.-W., Meng, Y., Volis, S., Sun, H., and Nie, Z.-L. (2017). Role of the Qinghai-Tibetan Plateau uplift in the Northern Hemisphere disjunction: evidence from two herbaceous genera of Rubiaceae. Sci. Rep. 7:13411.

Doyle J.J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17:144–163.

Doyle J.J. 2022. Defining coalescent genes: theory meets practice in organelle phylogenomics. Syst. Biol. 71:476–489.

Doyle J.J., Doyle J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.

Dutheil J.Y., Munch K., Nam K., Mailund T., Schierup M.H. 2015. Strong selective sweeps on the X chromosome in the Human-Chimpanzee ancestor explain its low divergence. PLoS Genet. 11:e1005451.

Edelman N.B., Frandsen P.B., Miyagi M., Clavijo B., Davey J., Dikow R.B., García-Accinelli G., Van Belleghem S.M., Patterson N., Neafsey D.E., Challis R., Kumar S., Moreira G.R.P., Salazar C., Chouteau M., Counterman B.A., Papa R., Blaxter M., Reed R.D., Dasmahapatra K.K., Kronforst M., Joron M., Jiggins C.D., McMillan W.O., Di Palma F., Blumberg A.J., Wakeley J., Jaffe D., Mallet J. 2019. Genomic architecture and introgression shape a butterfly radiation. Science. 366:594–599.

Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution. 63:1–19.

Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. PNAS 104:5936–5941.

Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.

Ehrendorfer F., Barfuss M.H.J., Manen J.-F., Schneeweiss G.M. 2018. Phylogeny, character evolution and spatiotemporal diversification of the species-rich and world-wide distributed tribe Rubieae (Rubiaceae). PLOS ONE. 13:e0207615.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Biol. 27:401–410.

Fitch W.M. 1970. Distinguishing homologous from analogous proteins. Syst. Zool. 19:99–113.

Folk R.A., Mandel J.R., Freudenstein J.V. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. Syst. Biol. 66:320–337.

Fu C.-N., Li H.-T., Milne R., Zhang T., Ma P.-F., Yang J., Li D.-Z., Gao L.-M. 2017. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. BMC Genomics 18:956.

Gandini C.L., Sanchez-Puerta M.V. 2017. Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. Sci. Rep. 7:43402.

García N., Folk R.A., Meerow A.W., Chamala S., Gitzendanner M.A., Oliveira R.S. de, Soltis D.E., Soltis P.S. 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). Mol. Phylogenet. Evol. 111:231–247.

Gardner E.M., Johnson M.G., Pereira J.T., Puad A.S.A., Arifiani D., Sahromi, Wickett N.J., Zerega N.J.C. 2021. Paralogs and off-target sequences improve phylogenetic resolution in a densely sampled study of the breadfruit genus (*Artocarpus*, Moraceae). Syst. Biol. 70:558–575.

Ginter A., Razafimandimbison S.G., Bremer B. 2015. Phylogenetic affinities of *Myrioneuron* and *Cyanoneuron*, generic limits of the tribe Argostemmateae and description of a new Asian tribe, Cyanoneuroneae (Rubiaceae). Taxon. 64:286–298.

Gitzendanner M.A., Soltis P.S., Wong G.K.-S., Ruhfel B.R., Soltis D.E. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am. J. Bot. 105:291–301.

Givnish T.J., Zuluaga A., Spalink D., Soto Gomez M., Lam V.K.Y., Saarela J.M., Sass C., Iles W.J.D., de Sousa D.J.L., Leebens-Mack J., Chris Pires J., Zomlefer W.B., Gandolfo M.A., Davis J.I., Stevenson D.W., dePamphilis C., Specht C.D., Graham S.W., Barrett C.F., Ané C. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. Am. J. Bot. 105:1888–1910.

Gonçalves D.J.P., Simpson B.B., Ortiz E.M., Shimizu G.H., Jansen R.K. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. Mol. Phylogenet. Evol. 138:219–232.

Goremykin V.V. 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. Mol. Biol. Evol. 20:1499–1505.

Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

Heath T.A., Hedtke S.M., Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46:239–257.

Hennig, W. 1966. Phylogenetic Systematics. The University of Illinois Press, Urbana.

Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35:518–522.

Huelsenbeck J.P., Larget B., Alfaro M.E. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Mol. Biol. Evol. 21:1123–1133.

Ishikawa S.A., Inagaki Y., Hashimoto T. 2012. RY-coding and non-homogeneous models can ameliorate the maximum-likelihood inferences from nucleotide sequence data with parallel compositional heterogeneity. Evol. Bioinform. Online 8:EBO.S9017.

Janssens S.B., Groeninckx I., De Block P.J., Verstraete B., Smets E.F., Dessein S. 2016. Dispersing towards Madagascar: biogeography and evolution of the Madagascan endemics of the Spermacoceae tribe (Rubiaceae). Mol. Phylogenet. Evol. 95:58–66.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4:1600016.

Johnson M.G., Pokorny L., Dodsworth S., Botigué L.R., Cowan R.S., Devault A., Eiserhardt W.L., Epitawalage N., Forest F., Kim J.T., Leebens-Mack J.H., Leitch I.J., Maurin O., Soltis D.E., Soltis P.S., Wong G.K., Baker W.J., Wickett N.J.

2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Syst. Biol. 68:594–606.

Jones K.E., Fér T., Schmickl R.E., Dikow R.B., Funk V.A., Herrando-Moraira S., Johnston P.R., Kilian N., Siniscalchi C.M., Susanna A., Slovák M., Thapa R., Watson L.E., Mandel J.R. 2019. An empirical assessment of a single family-wide hybrid capture locus set at multiple evolutionary timescales in Asteraceae. Appl. Plant Sci. 7:e11295.

Källersjö M., Albert V.A., Farris J.S. 1999. Homoplasy increases phylogenetic structure. Cladistics. 15:91–93.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.

Kapli P., Flouri T., Telford M.J. 2021. Systematic errors in phylogenetic trees. Current Biol. 31:R59–R64.

Karol K.G., Arumuganathan K., Boore J.L., Duffy A.M., Everett K.D., Hall J.D., Hansen S., Kuehl J.V., Mandoli D.F., Mishler B.D., Olmstead R.G., Renzaglia K.S., Wolf P.G. 2010. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol. Biol. 10:321.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Katoh K., Standley D.M. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. Bioinformatics. 32:1933–1942.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.

Lanier H.C., Knowles L.L. 2012. Is recombination a problem for species-tree analyses? Syst. Biol. 61:691–701.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 30:3276–3278.

Leigh J.W., Susko E., Baumgartner M., Roger A.J. 2008. Testing congruence in phylogenomic analysis. Syst. Biol. 57:104–115.

Literman R., Schwartz R. 2021. Genome-scale profiling reveals noncoding loci carry higher proportions of concordant data. Mol. Biol. Evol. 38:2306–2318.

Lynch M., Conery J.S. 2000. The evolutionary fate and consequences of duplicate genes. Science. 290:1151–1155.

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

Mallet J., Besansky N., Hahn M.W. 2016. How reticulated are species? BioEssays. 38:140–149.

Martinsen G.D., Whitham T.G., Turek R.J., Keim P. 2001. Hybrid populations selectively filter gene introgression between species. Evolution. 55:1325–1335.

McKain M.R., Johnson M.G., Uribe-Convers S., Eaton D., Yang Y. 2018. Practical considerations for plant phylogenomics. Appl. Plant Sci. 6:e1038.

Mendes F.K., Hahn M.W. 2018. Why concatenation fails near the anomaly zone. Syst. Biol. 67:158–169.

Meyer, M., Kircher, M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010:db.rot5448.

Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30:1188–1195.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534.

Mirarab S., Reaz R., Bayzid Md.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 30:i541–i548.

Molloy E.K., Warnow T. 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst. Biol. 67:285–303.

Morales-Briones D.F., Liston A., Tank D.C. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). New Phytol. 218:1668–1684.

Naser-Khdour S., Minh B.Q., Lanfear R. 2021. The influence of model violation on phylogenetic inference: a simulation study. bioRxiv 2021.09.22.461455.

Natali A., Manen J.-F., Ehrendorfer F. 1995. Phylogeny of the Rubiaceae-Rubioideae, in particular the tribe Rubieae: evidence from a non-coding chloroplast DNA sequence. Ann. Mo. Bot. Gard. 82:428–439.

Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Ogilvie H.A., Heled J., Xie D., Drummond A.J. 2016. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Syst. Biol. 65:381–396.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9:e1000602.

Phillips M.J., Delsuc F., Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21:1455–1458.

Phillips M.J., Lin Y.-H., Harrison G., Penny D. 2001. Mitochondrial genomes of a bandicoot and a brushtail possum confirm the monophyly of australidelphian marsupials. Proc. R. Soc. Lond. B. 268:1533–1538.

Poczai P., Hyvönen J. 2010. Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. Mol. Biol. Rep. 37:1897–1912.

Puff C. 1982. The delimitation of the tribe Anthospermeae and its affinities to the Paederieae (Rubiaceae). Bot. J. Linn. Soc. 84:355–377.

Rannala B., Yang Z. 2017. Efficient bayesian species tree inference under the multispecies coalescent. Syst. Biol. 66:823–842.

Rasmussen M.D., Kellis M. 2007. Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. Genome Res. 17:1932–1942.

Razafimandimbison S.G., 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. Mol. Phylogenet. Evol. 48:207–223.

Rieseberg L.H., Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants 5:65–84.

Rincón-Barrado M., Olsson S., Villaverde T., Moncalvillo B., Pokorny L., Forrest A., Riina R., Sanmartín I. 2021. Ecological and geological processes impacting speciation modes drive the formation of wide-range disjunctions within tribe Putorieae (Rubiaceae). J. Syst. Evol. 59:915–934.

Robbrecht E. 1988. Tropical woody Rubiaceae. Characteristic features and progressions. Contribution to a new subfamilial classification. Opera Bot. Belg. 1:272.

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. Syst. Geogr. Plants 76:85–146.

Roch S., Steel M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.

Rokas A., Williams B.L., King N., Carroll S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 425:798–804.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Ruane S., Raxworthy C.J., Lemmon A.R., Lemmon E.M., Burbrink F.T. 2015. Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. BMC Evol. Biol. 15:1–14.

Ruhfel B.R., Gitzendanner M.A., Soltis P.S., Soltis D.E., Burleigh J. 2014. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 14:23.

Rydin C., Kainulainen K., Razafimandimbison S.G., Smedmark J.E.E., Bremer B. 2009a. Deep divergences in the coffee family and the systematic position of *Acranthera*. Plant Syst. Evol. 278:101–123.

Rydin C., Källersjö M., Friis E.M. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems, and the monophyly of conifers. Int. J. Plant Sci. 163:197–214.

Rydin C., Razafimandimbison S.G., Khodabandeh A., Bremer B. 2009b. Evolutionary relationships in the Spermacoceae 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 views of Rubiaceae phylogeny. Am. J. Bot. 104:1522–1532.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37:121–147.
Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33:1654–1668.
Sayyari, E., and Mirarab, S. (2018). Testing for polytomies in phylogenetic species trees using quartet frequencies. Genes 9:132.
Scally A., Dutheil J.Y., Hillier L.W., Jordan G.E., Goodhead I., Herrero J., Hobolth A., Lappalainen T., Mailund T., Marques-Bonet T., McCarthy S., Montgomery S.H., Schwalie P.C., Tang Y.A., Ward M.C., Xue Y., Yngvadottir B., Alkan C., Andersen L.N., Ayub Q., Ball E.V., Beal K., Bradley B.J., Chen Y., Clee C.M., Fitzgerald S., Graves T.A., Gu Y., Heath P., Heger A., Karakoc E., Kolb-Kokocinski A., Laird G.K., Lunter G., Meader S., Mort M., Mullikin J.C., Munch K., O'Connor T.D., Phillips A.D., Prado-Martinez J., Rogers A.S., Sajjadian S., Schmidt D., Shaw K., Simpson J.T., Stenson P.D., Turner D.J., Vigilant L., Vilella A.J., Whitener W., Zhu B., Cooper D.N., de Jong P., Dermitzakis E.T., Eichler E.E., Flicek P., Goldman N., Mundy N.I., Ning Z., Odom D.T., Ponting C.P., Quail M.A., Ryder O.A., Searle S.M., Warren W.C., Wilson R.K., Schierup M.H., Rogers J., Tyler-Smith C., Durbin R. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483:169–175.
Schumann, K., 1891. Rubiaceae. Pp. 1-156 in A. Engler and K. Prantl (eds.), Die natürlichen Pflanzenfamilien, vol. 4 (4). Wilhelm Engelmann, Leipzig.
Simmons M.P. 2017. Relative benefits of amino-acid, codon, degeneracy, DNA, and purine-pyrimidine character coding for phylogenetic analyses of exons: Relative benefits of character-coding methods. J. Syst. Evol. 55:85–109.
Sloan D.B., Oxelman B., Rautenberg A., Taylor D.R. 2009. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. BMC Evol. Biol. 9:260.
Small R.L., Cronn R.C., Wendel J.F. 2004. Use of nuclear genes for phylogeny reconstruction in plants. Aust. Syst. Bot. 17:145.
Smedmark J.E.E., Bremer B. 2011. Molecular systematics and incongruent gene trees of Urophylleae (Rubiaceae). Taxon. 60:1397–1406.
Smedmark J.E.E., Eriksson T., Bremer B. 2010. Divergence time uncertainty and historical biogeography reconstruction – an example from Urophylleae (Rubiaceae). J. Biogeogr. 37:2260–2274.
Smedmark J.E.E., Rydin C., Razafimandimbison S.G., Khan S.A., Liede-Schumann S., Bremer B. 2008. A phylogeny of Urophylleae (Rubiaceae) based on *rps16* intron data. Taxon 57:24–32.
Smith D.R. 2011. Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol. Evol. 3:743–748.
Smith D.R. 2014. Mitochondrion-to-plastid DNA transfer: it happens. New Phytol. 202:736–738.
Soltis D.E., Kuzoff R.K. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evolution 49:727–742.
Sousa F., Foster P.G., Donoghue P.C.J., Schneider H., Cox C.J. 2019. Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). New Phytol. 222:565–575.
Springer M.S., Gatesy J. 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94:1–33.

Straub S.C.K., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99:349–364.
Thomas A.E., Igea J., Meudt H.M., Albach D.C., Lee W.G., Tanentzap A.J. 2021. Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand *Veronica*. Am. J. Bot. 108:1289–1306.
Van de Peer Y., Fawcett J.A., Proost S., Sterck L., Vandepoele K. 2009. The flowering world: a tale of duplications. Trends Plant Sci. 14:680–688.
Verdcourt B. 1958. Remarks on the classification of the Rubiaceae. Bull. Jard. Bot. État Bruxelles 28:209–290.
Walker J.F., Walker-Hale N., Vargas O.M., Larson D.A., Stull G.W. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. PeerJ. 7:e7747.
Wang Z., Liu K.J. 2016. A performance study of the impact of recombination on species tree analysis. BMC Genomics 17:785.
Weitemier K., Straub S.C.K., Cronn R.C., Fishbein M., Schmickl R., McDonnell A., Liston A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. 2:1400042.
Wen H.-Z., Wang R.-J. 2012. *Foonchewia guangdongensis* gen. et sp. nov. (Rubioideae: Rubiaceae) and its systematic position inferred from chloroplast sequences and morphology. J. Syst. Evol. 50:467–476.
Wikström N., Bremer B., Rydin C. 2020. Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae). J. Syst. Evol. 58:440–460.
Wikström N., Kainulainen K., Razafimandimbison S.G., Smedmark J.E.E., 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., Yang T., Kuang D.-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: Phylogeny of Gentianales. J. Syst. Evol. 54:400–415.
Young A.D., Gillung J.P. 2020. Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. Syst. Entomol. 45:225–247.
Yu Y., Degnan J.H., Nakhleh L. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. PLoS Genet. 8:e1002660.
Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.
Zhang R., Wang Y.-H., Jin J.-J., Stull G.W., Bruneau A., Cardoso D., De Queiroz L.P., Moore M.J., Zhang S.-D., Chen S.-Y., Wang J., Li D.-Z., Yi T.-S. 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. Syst. Biol. 69:613–622.
Zhou B.-F., Yuan S., Crowl A.A., Liang Y.-Y., Shi Y., Chen X.-Y., An Q.-Q., Kang M., Manos P.S., Wang B. 2022. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. Nat. Commun. 13:1320.
Zimmer E.A., Wen J. 2013. Reprint of: Using nuclear gene data for plant phylogenetics: Progress and prospects. Mol. Phylogenet. Evol. 66:539–550.

Zwick A., Regier J.C., Zwickl D.J. 2012. Resolving discrepancy between nucleotides and amino acids in deep-level arthropod phylogenomics: differentiating serine codons in 21-amino-acid models. PLoS ONE. 7:e47450.

Acknowledgments

I want to begin with thanking my main supervisor Catarina Rydin for giving me the unique opportunity of pursuing a PhD. Thanks for all the knowledge sharing, support, advice, encouragement, and our countless discussions. Your relentless drive and passion for science have been a great source of inspiration, and I am grateful that I have had you as my supervisor over the years. Many thanks to my two co-supervisors Niklas Wikström and Sylvain Razafimandimbison. Niklas, thank you for all the constructive feedback, sharing of knowledge and discussions during this time. Sylvain, thank you for sharing your taxonomy, Rubiaceae, and field expertise, and thanks for showing Madagascar - one of the main highlights during my PhD time. Thanks also to Per Ola Karis for help with loans and taxonomical matters.

Thanks to all the other systematists, past and present, including Frida Karis Stångberg, Kent Kainulainen, Kajsa Axelius, Aelys Humphreys, Gitte Petersen, Natasha Barbolini, Benjamin Anderson, Marian Schubert and Stefan Löfstrand, for advice and discussions over the years. Special thanks to the people involved in the floristic and OME courses, including greenhouse, BIG, and Bergian Garden staff members, and fellow teaching assistants - it has been great fun and educational to work together with all of you on these courses.

Thanks to Anbar Khodabandeh for educating me in laboratory work. Thanks also to Bodil Cronholm and Martin Irestedt for being so helpful and welcoming when I came to the lab at the Swedish Museum of Natural History.

Thank you Eva Larsén for being a fantastic travel companion, the conversations, and for listening to the occasional rant. Thanks to the rest of my fellow PhD students, including Emma Wahlberg, Julia Ferm, Nannie Persson, Ruben Blokzijl, Maja Edlund, Lovisa Thilén, Jan-Niklas Nuppenau, Laura Schat, Lotta Berntzon, Chen Hou, and Stina Bolinder - without you these years would have been far less memorable and enjoyable.

Thanks to my family and friends for being family and friends, with all that that implies. To Petter and Rasmus for consistently offering welcome and important distractions from PhD work. To my kids, Jon and the baby in the belly, and my other half Lena, thank you for making everything better and magical.

To anyone I may have forgotten, my apologies and thanks.