Universiteit Leiden
The Netherlands

Antiquities of the rainforest: evolution of mycoheterotrophic angiosperms growing on Glomeromycota

Mennes, C.B.

Citation

Mennes, C. B. (2015, October 6). *Antiquities of the rainforest: evolution of mycoheterotrophic angiosperms growing on Glomeromycota*. Retrieved from https://hdl.handle.net/1887/35731

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/35731

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The handle http://hdl.handle.net/1887/35731 holds various files of this Leiden University dissertation

Author: Mennes, Constantijn
Title: Antiquities of the rainforest : evolution of mycoheterotrophic angiosperms growing on Glomeromycota
Issue Date: 2015-10-06

Chapter 3

Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales)

Collection of *Corsia* cf. *huonensis*, Papua New Guinea. Photo: S.P. Lyon

Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales)

Constantijn B. Mennes[1], Vivienne K. Y. Lam[5], Paula J. Rudall[3], Stephanie P. Lyon[4], Sean W. Graham[5], Erik F. Smets[1,2] and Vincent S. F. T. Merckx[1]

[1]Naturalis Biodiversity Center, Leiden University, Leiden, The Netherlands,

[2]Section Ecology, Evolution and Biodiversity Conservation, KU Leuven, BE-3001 Leuven, Belgium,

[3]Royal Botanic Gardens Kew, Richmond, Surrey, UK,

[4]Department of Botany, University of Wisconsin Madison, Madison, WI 54706 USA,

[5]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4

Published in *Journal of Biogeography* 42: 1123-1136 (2015).

Materials & Methods (section 3.3) from supplemental information of this publication, available from http://onlinelibrary.wiley.com/doi/10.1111/jbi.12486/suppinfo

3.1. Abstract

Aim Many plant families have a disjunct distribution across the southern Pacific Ocean, including the mycoheterotrophic family Corsiaceae, which provides a prime example of this biogeographical pattern. A better grasp of the family's evolutionary relationships is needed to understand its historical biogeography. We therefore aimed to (1) test the uncertain monophyly of Corsiaceae, (2) define its phylogenetic position, and (3) estimate divergence times for the family, allowing us to assess whether the distribution of the family is the result of vicariance.

Location Southern South America and Australasia.

Methods We analysed various combinations of mitochondrial and nuclear data to address the monophyly, phylogenetic position and age of Corsiaceae. To test its monophyly, we used a three-locus data set including most monocot orders, and to infer its exact phylogenetic position, we used a five-locus extended data set. We corroborated these findings using an independent plastome dataset. We then used a two-locus dataset with taxa from all monocot orders, and a three-locus dataset containing only taxa of Liliales, to estimate divergence times using a fossil-calibrated uncorrelated lognormal relaxed-clock approach.

Results Corsiaceae is a monophyletic family and the sister group of Campynemataceae. This clade is the sister group of all other Liliales. The crown age of Corsiaceae is estimated to be 53 Ma (95% confidence interval 30–76 Ma).

Main conclusions Corsiaceae is an ancient family of mycoheterotrophic plants, whose crown age overlaps with the plate-tectonic split of Gondwana, consistent with a vicariance-based explanation for its current distribution.

Keywords: Australasia, BEAST, Corsiaceae, Gondwana, historical biogeography, Liliales, mycoheterotrophy, southern South America, vicariance.

3.2. Introduction

Fully mycoheterotrophic plants are non-photosynthetic and rely on their mycorrhizal symbionts for carbon uptake. They often differ from their autotrophic relatives as a result of reduced vegetative growth and modified floral structures (Leake, 1994), and show elevated rates of molecular evolution (Merckx *et al.*, 2006, 2009a) that can obscure their systematic affinities. Many clades of mycoheterotrophic plants are characterized by remarkably disjunct distribution patterns (Leake, 1994; Merckx *et al.*, 2013d). For example, the genus *Seychellaria* (Triuridaceae) grows disjunctly in Madagascar, the Seychelles and Tanzania (Maas-van de Kamer & Weustenfeld, 1998), and the otherwise South American genus *Voyria* (Gentianaceae) has a single species (*V. primuloides*) known from West and Central Africa (Merckx *et al.*, 2013c). Such distribution patterns have often been interpreted as indicative of an ancient widespread distribution for these lineages (e.g. Jonker, 1938; Leake, 1994) but, at least for *Seychellaria* and *Voyria,* dispersal mechanisms have recently been inferred (Mennes *et al.*, 2013; Merckx *et al.*, 2013c). The divergence ages of many mycoheterotrophic clades remain to be investigated.

One family of particular interest is the mycoheterotrophic Corsiaceae Becc. (Fig. 8), which is found disjunctly in South America, Australasia and China. This family currently consists of approximately 30 species in three genera (Neinhuis & Ibisch, 1998; Zhang *et al.*, 1999; Jones & Gray, 2008). The tropical genus *Corsia* Becc. (*c.* 27 species) occurs in New Guinea, the Solomon Islands and northern Australia (van Royen, 1972; Neinhuis & Ibisch, 1998; Jones & Gray, 2008). The temperate genus *Arachnitis* Phil. (one or two species) is found in southern South America, including the Falkland Islands, and reaches the northern limit of its distribution in Bolivia (Ibisch *et al.*, 1996; Neinhuis & Ibisch, 1998). The poorly known and monotypic genus *Corsiopsis* D.X. Zhang, R.M.K. Saunders & C.M. Hu has only been recorded in southern China and its description is based on a single specimen collected in 1974 (Zhang *et al.*, 1999). The disjunct distribution of Corsiaceae was suggested to be the result of a Gondwanan link between South America and Australasia via Antarctica (Zhang *et al.*, 1999).

Members of the Corsiaceae are characterized by a prominent floral synapomorphy: the outer median tepal is enlarged, forming a distinct labellum (van Royen, 1972; Neinhuis & Ibisch, 1998; Rudall & Bateman, 2002). Before molecular data were used to infer the phylogenetic relationships of Corsiaceae, the group was classified as a family or tribe closely related to Burmanniaceae, Thismiaceae or sometimes Geosiridaceae or Orchidaceae (i.e. other monocot families that contain mycoheterotrophic species) (Bentham & Hooker, 1883; Engler, 1889b; Jonker, 1938; Cronquist, 1981; Thorne, 1983; Dahlgren *et al.*, 1985; Rübsamen, 1986). In the morphological cladistic analysis of Stevenson & Loconte (1995), Corsiaceae and other families containing mycoheterotrophic plants were grouped in Orchidales, together with the bigeneric

Fig. 8. Four species of Corsiaceae included in this study: (a) *Corsia* cf. *brassii* P.Royen (photo: S.P. Lyon); (b) *Corsia* cf. *huonensis* P.Royen (photo: S.P. Lyon), (c) *Arachnitis uniflora* Phil. (photo: M. Renny); and (d) *Corsia* cf. *boridiensis* P.Royen (photo: S.P. Lyon).

chlorophyllous family Campynemataceae (*Campynema* Labill. and *Campynemanthe* Baill.). Neyland & Hennigan (2003) inferred a close relationship between *Corsia* and *Campynema* based on 26S rDNA data from 45 monocot taxa (including two species of *Arachnitis* and one of *Corsia*). The resulting clade was strongly supported, but there was only weak support for it belonging to a larger clade with other species of Liliales. They also did not recover Corsiaceae as monophyletic: *Arachnitis* was instead inferred to be sister to *Thismia* (Thismiaceae, Dioscoreales), but with poor support for this relationship. Based on several morphological differences (e.g. number of stigmas, size of seeds, presence of a rhizome), Ibisch *et al.* (1996) also raised the question of whether *Corsia* and *Arachnitis* are closely related. Later analyses that lacked *Corsia* found *Arachnitis* to be poorly supported as the sister group of Liliales based on mitochondrial *atp*A (Davis *et al.*, 2004), *atp*A and nuclear 18S rDNA (Fay *et al.*, 2006), or the mitochondrial genes *atp*A, *cob* and *nad5* (Petersen *et al.*, 2013). Kim *et al*. (2013) recovered *Arachnitis* as a poorly supported sister group of a large clade comprising commelinids, Asparagales and Liliales, based on a single plastid *rbc*L sequence (see below). This lack of support for the phylogenetic position of Corsiaceae has been explicitly noted by many authors (Fay *et al.*, 2006; Kim *et al.*, 2013; Petersen *et al.*, 2013). In summary, phylogenetic analyses based on DNA sequence data are not entirely in accordance with earlier classifications, as these phylogenetic analyses mostly suggest a close relationship between Corsiaceae and Liliales, rather than a close relationship of Corsiaceae with other mycoheterotrophic

taxa. The precise phylogenetic position and monophyly of Corsiaceae also remain ambiguous, which has prevented attempts to address questions about the intriguing disjunct distribution pattern of the family.

Disjunct distribution patterns across the southern Pacific Ocean are observed in many angiosperm families. For a long time, it was thought that these distributions were explained by a vicariant scenario resulting from the breakup of Gondwana (Raven & Axelrod, 1972), but recent insights suggest that vicariance cannot be the only explanation. Particularly in plants, Sanmartín & Ronquist (2004) found incongruence between Southern Hemisphere area cladograms and the generally accepted sequential breakup of Gondwana. This incongruence indicates that ancient vicariance patterns may have been obscured by more recent dispersal and/or extinction events. Subsequent biogeographical inferences based on dated phylogenies of many groups, such as *Aristotelia* (Elaeocarpaceae), *Gaultheria* (Ericaceae) and *Muehlenbeckia* (Polygonaceae), indicated that long-distance dispersal between Australasia and South America is the most likely explanation of the distribution patterns of these taxa (Crayn *et al.*, 2006; Bush *et al.*, 2009; Schuster *et al.*, 2013). Conversely, vicariance has been suggested to explain the extant distributions of a few other taxa (e.g. tribe Embothriinae, Proteaceae, Barker *et al.*, 2007; *Raukaua*, Araliaceae, Mitchell *et al.*, 2012). A notorious case is the family Nothofagaceae, for which both long-distance dispersal (e.g. Cook & Crisp, 2005) and vicariance (Heads, 2006) have been proposed independently as the most likely mechanism explaining its distribution.

The mycoheterotrophic family Corsiaceae is a prime candidate for testing ancient Gondwanan vicariance, although this is contingent on resolving its monophyly and phylogenetic affinities. We therefore aimed to (1) test the monophyly of Corsiaceae, (2) define its exact phylogenetic position, and (3) estimate divergence times, in order to investigate whether its disjunct distribution is consistent with ancient vicariance events or more recent long-distance dispersal. Because the phylogenetic placement of Corsiaceae has proved to be a difficult problem (e.g. Kim *et al.*, 2013; Petersen *et al.*, 2013), probably due to the rate variation that is typical of mycoheterotrophic plants (e.g. Merckx *et al.*, 2006, 2009a), we primarily used nuclear and mitochondrial data, which were demonstrated to be informative both for other mycoheterotrophic taxa and for Liliales as a whole (e.g. Mennes *et al.*, 2013; Petersen *et al.*, 2013). The use of plastid data has generally lagged in mycoheterotrophic plant groups, because non-photosynthetic plants tend to lack the photosynthetic genes that are commonly used for plant phylogenetic inference (e.g. Olmstead & Palmer, 1994). Recent advances in sequencing technology have, however, facilitated the recovery of whole plastid genomes from photosynthetic and non-photosynthetic monocots alike (e.g. Givnish *et al.*, 2010; Barrett & Davis, 2012; Logacheva *et al.*, 2014). We used this approach to recover plastid genes sets from *Arachnitis* and *Corsia*, allowing us to corroborate the monophyly and phylogenetic position of Corsiaceae inferred from mitochondrial

and nuclear data. We then estimated divergence times using a subset of the nuclear and mitochondrial data, based on an extensive set of monocot taxa and multiple fossil calibration points.

3.3. Materials and Methods

3.3.1. Taxon sampling

Nuclear and mitochondrial data

In total, new DNA sequence data from 47 species were obtained, supplemented by sequence data from all monocot orders from GenBank (see Appendix A.1). This study includes data from three species of *Corsia* from Papua New Guinea and one species of *Arachnitis* from Argentina (see Appendix A.1 for voucher information). No material of the rare genus *Corsiopsis* was available for inclusion. Identifications of the available *Corsia* specimens were carried out using the key of van Royen (1972). Four different nuclear and mitochondrial datasets were compiled to study different evolutionary questions.

1. To maximize monocot taxon sampling for testing the monophyly of Corsiaceae as well as its position in the monocots, a three-locus dataset (nuclear 18S rDNA and the mitochondrial *atp*A and *mat*R) was analysed, including sequence data from 33 species of Liliales (representing all nine recognized families) and from most monocot orders.

2. To maximize the analysed sequence data for Liliales, in order to infer the exact position of Corsiaceae within Liliales, a five-locus dataset (nuclear 18S rDNA and the mitochondrial *atp*A, *mat*R, *cob* and *nad*5) was analysed, including sequence data from these 33 Liliales taxa and the outgroups Pandanales, Dioscoreales and Alismatales.

3. To estimate general divergence times in monocots, a two-locus dataset (nuclear 18S rDNA and mitochondrial *atp*A) similar to the one used by Mennes *et al.* (2013) containing sequence data from all monocot orders was analysed using a fossil-calibrated uncorrelated lognormal relaxed-clock approach.

4. To obtain divergence time estimates in Liliales based on more data (i.e. more reliably, possibly with a smaller 95% confidence interval), a three-locus dataset (nuclear 18S rDNA and mitochondrial *atp*A and *mat*R) containing only sequence data from Liliales was analysed using a fossil-calibrated uncorrelated lognormal relaxed-clock approach (see Appendix A.1 for accession and voucher data). These four analyses, based on different subsets of the same nuclear and mitochondrial regions for monocots, collectively provide insights in the evolutionary biogeographical history of Corsiaceae based on two of the three plant genomes.

Plastid genome data

As independent confirmation of the position of Corsiaceae in the monocot phylogeny, we also retrieved plastid gene sets for 82 genes from 63 taxa, including newly sequenced species of Liliales (*Arachnitis uniflora, Corsia* cf. *boridiensis*, *Lilium superbum* and *Campynema lineare*) and Pandanales (*Carludovica palmata*) (Appendices A.1 & A.2). These were added to previously published plastid gene sets from Barrett *et al*. (2013) and from GenBank (Appendix A.1), and represent four rDNA genes and 78 of the 79 protein-coding genes found in most angiosperms (Raubeson & Jansen, 2005; *ycf*1 was excluded, see below). The taxon sampling represents all monocot orders and the monocot family Dasypogonaceae, and includes 11 additional angiosperms (eudicots, magnoliids and ANITA-grade taxa).

3.3.2. DNA extraction, PCR and sequencing

Nuclear and mitochondrial data

Methods for DNA extraction and PCR largely follow Mennes *et al.* (2013), as do the amplification of the nuclear 18S rDNA locus and the mitochondrial F1-ATPase alpha subunit (*atp*A) and maturase (*mat*R) loci. The mitochondrial apocytochrome B (*cob*) locus was amplified using the primer pair COB1F and COB1R (Petersen *et al.*, 2006). The PCR procedure was as follows: premelting for 4 min at 94 °C; followed by 40 cycles of denaturing for 1 min at 94 °C, annealing for 1 min at 52 °C and extension for 2 min at 72 °C; followed by a final extension step of 7 min at 72 °C. The mitochondrial nicotinamide adenine dinucleotide dehydrogenase subunit 5 (*nad*5) region was amplified using the primer pair nad5-F1 and nad5-R1 (Qiu *et al.*, 2006). The PCR procedure was as follows, as based on the *nad1b-c* procedure from Merckx *et al.* (2006): premelting for 4 min at 94 °C; followed by 30 cycles of denaturing for 30 s at 94 °C, annealing for 30 s at 52 °C and extension of 90 s at 72 °C; followed by a final extension step of 7 min at 72 °C. Sanger sequencing of the amplified DNA was carried out by the Macrogen sequencing facility (Macrogen, Amsterdam, The Netherlands).

Plastid genome data

DNA was extracted using a modified CTAB extraction protocol (Rai *et al.*, 2003). Libraries for next-generation sequencing were made using Bioo NEXTflex DNA sequencing kits (Bioo Scientific, Austin, TX, USA) or KAPA LTP Library Preparation kits (Kapa Biosystems, Boston, MA, USA). Paired-end sequencing was performed on an Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA); samples were multiplexed on individual lanes (Cronn *et al.,* 2008). Sequence data were processed using CASAVA 1.8.2 (Illumina), and *de novo* assembly was performed using CLC Genomics Workbench 6.5.1 (CLC Bio, Aarhus, Denmark) with the default parameters (length fraction 0.85; similarity fraction 0.85). We compared *de*

novo contigs to a local database of plastid genes using BLAST, to filter for plastid genes: occasional mitochondrial inserts of plastid genes were identified using BLAST searches on contiguous non-plastid sequences, and were removed. We designed primers using PRIMER3 (Koressaar & Remm, 2007; Untergrasser *et al.,* 2012) to link contigs, by using them to amplify and sequence across gaps or regions with low overlap (< 20 bp). Amplification was performed using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA), following the manufacturer's instructions. Sanger sequencing was carried out on an Applied Biosystems 3730S 48-capillary DNA Analyzer platform (Applied Biosystems, Foster City, CA, USA), using BigDye Terminator v3.1 sequencing chemistry (Applied Biosystems) following the manufacturer's instructions. *De novo* contigs and Sanger sequences were manually assembled and checked carefully using SEQUENCHER 4.2.2 (Gene Codes Corporation, Ann Arbor, MI, US).

We used DOGMA (Wyman *et al.*, 2004) to annotate the plastid gene sets. Annotations were checked using CONVERT GENBANK TO FASTA (http://rocaplab.ocean.washington.edu/tools/genbank_to_fasta) and by comparison to published GenBank plastomes. In total, 82 plastid genes were extracted from most photosynthetic taxa (78 protein-coding genes and four rDNA genes; see Appendix A.2), and a subset of plastid genes were retrieved for *Arachnitis* (the four rDNA genes and 16 protein-coding genes retrieved as open reading frames) and *Corsia* (the four rDNA genes and 23 protein-coding genes as open reading frames; note that *ycf*1 was obtained for this taxon, but omitted from the alignment due to alignment uncertainty). Full circular plastid genomes for these two taxa were obtained and will be published elsewhere (here, the focus was on plastid gene sets for phylogenetic inference). We combined individual alignments into a concatenated matrix (unretrieved genes for *Arachnitis* and *Corsia* were represented by blanks).

3.3.3. Phylogenetic analyses

Nuclear and mitochondrial data

Two nuclear and mitochondrial datasets were analysed to infer the phylogeny of Corsiaceae. Assembly of DNA sequences and manual editing were performed using GENEIOUS PRO 6.1.7 (Biomatters, Auckland, New Zealand; available at http://www.geneious.com/). Sequence alignment was conducted using the MUSCLE (Edgar, 2004) plug-in in GENEIOUS PRO, maintaining the default settings. The optimality criteria maximum likelihood (ML) and Bayesian inference (BI) were used for inferring phylogenies. The best substitution model to fit the data under the Akaike information criterion (AIC) was found using JMODELTEST 2.1.4 (Guindon & Gascuel, 2003; Darriba *et al.*, 2012) for individual genes. ML analyses were performed using GARLI 2.0 (Zwickl, 2006). Branch support values in the ML results were calculated using the bootstrap support method (BS; Felsenstein, 1985), with 500 pseudoreplicates.

Majority-rule consensus trees were built using the Tree Builder plug-in in Geneious Pro. BI analyses were performed using MrBayes 3.2.2 (Ronquist *et al.,* 2012). The Markov chain Monte Carlo (MCMC; Geyer, 1991) analyses were carried out with two runs of four chains for at least 4,000,000 generations and only terminated after the standard deviation of split frequencies was lower than 0.01. One tree was sampled every 1000 generations. A burn-in of 25% was set. Results were evaluated using Tracer 1.6 (Rambaut *et al.*, 2013) by checking the effective sample size values (ESS) for model parameters (200 or higher was considered good, 100 or higher was considered sufficient) and the distribution and stability of the obtained likelihood scores (a normal distribution of the values was considered sufficient). Clade support was assessed using the following criteria. In the ML analyses, only clades with over 85% bootstrap support were considered strongly supported; clades with between 75% and 85% bootstrap support were considered moderately supported; clades with lower bootstrap support values were not considered to be reliably supported. In the BI analyses, clades with 95% or higher posterior probability (PP) values were considered to be significantly supported by the data. All the presented phylogenetic trees resulted from BI analyses and were visualized using FigTree 1.4 (Rambaut, 2012).

The first analysed dataset contained sequence data from most monocot orders, and included the loci 18S rDNA, *atp*A and *mat*R. This dataset was analysed to evaluate the monophyly of Corsiaceae and its position in the monocots, due to the emergence of *Arachnitis* in different positions in the monocots in several previous studies (Neyland & Hennigan, 2003; Kim *et al.*, 2013). An alignment containing all included Corsiaceae taxa was analysed, as well as alignments excluding *Arachnitis* and *Corsia*, respectively, evaluating the individual and independent placement of each genus. For each alignment, the best substitution model was found for each locus. For 18S rDNA, GTR+G was selected as the best model (for all three alignments). For *atp*A, the model TVM+I+G was suggested as the best fit. Because this model was not included in any of the phylogenetic inference packages used, the best available model (i.e. the model with the lowest AIC value) was used. Therefore, GTR+I+G was used for all three alignments. For *mat*R, GTR+G was selected for all three alignments. Each dataset was partitioned into three subsets, corresponding to the three amplified loci, and each locus was analysed with its best model found using the AIC.

The second dataset focused on the order Liliales and contained sequence data from 18S rDNA, *atp*A, *mat*R, *cob* and *nad*5. This dataset was analysed to infer the exact phylogenetic position of Corsiaceae. The model TIM2+G was suggested for the 18S rDNA dataset, and GTR+G was used. Likewise, for *atp*A, TIM2+I+G was suggested as best fit, and GTR+I+G was used. For *mat*R, TPM1uf+G was suggested, and GTR+G was used. For *cob*, TIM2+I+G was suggested, and GTR+I+G was used. For *nad*5, TIM1+I+G was suggested, and GTR+I+G was used. The dataset was

partitioned into five subsets, corresponding to the five amplified loci, and each locus was analysed with the corresponding best model found using AIC. Furthermore, each dataset was analysed separately to identify possible conflicts between the loci.

Plastid data

We inferred a plastid-based phylogeny using maximum likelihood. Genes were aligned individually using the default settings in MUSCLE 3.8.31 (Edgar, 2004), followed by manual editing in MESQUITE (Maddison & Maddison, 2007) using the procedure described by Graham *et al.* (2000). The ends of each sequence were adjusted in order to retain open reading frames in protein-coding regions. Gaps were coded as missing regions. We partitioned our plastid data by gene and codon position (first, second and third codon positions for 78 protein-coding genes; 5′-*rps*12 and 3′-*rps*12 were treated as separate genes, and the four rDNA genes were treated as additional data partitions) for a total of 241 initial data partitions. We used the 'hcluster' algorithm in PARTITIONFINDER 1.1.1. (Lanfear *et al.,* 2012) to attempt to group together data partitions with similar models of DNA substitution, according to the corrected Akaike information criterion (AIC_c; Hurvich & Tsai, 1989). The resulting scheme retained all 241 subsets, with either the GTR+G or the GTR+G+I model of evolution. The total length of the aligned plastid dataset was 101,698 bp. We conducted 20 ML replicate searches using the GTR+G model for all data partitions, using a graphical user interface for RAxML 7.4.2 (Stamatakis, 2006; Silvestro & Michalak, 2012). We evaluated branch support for the ML analysis using 500 bootstrap replicates (Felsenstein, 1985).

3.3.4. Estimation of divergence times

Divergence times of taxa in Corsiaceae were estimated by performing two unlinked uncorrelated lognormal relaxed clock analyses using BEAST 2.1.3 (Bouckaert *et al.*, 2014). First, a dataset consisting of 18S rDNA and *atp*A DNA sequence data for all monocot orders was analysed (see Appendix A.1 for taxon information). The dataset was partitioned into two subsets, corresponding to the 18S rDNA and *atp*A loci. Each locus was analysed with the GTR+I+G substitution model, which was selected as best fit model under the AIC in JMODELTEST 2.1.4 (Guindon & Gascuel, 2003; Darriba *et al.*, 2012). Second, a dataset consisting of 18S rDNA, *atp*A and *mat*R DNA sequence data for Liliales only was analysed. This dataset was partitioned into three subsets, corresponding to the 18S rDNA, *atp*A and *mat*R loci. For 18S rDNA, TIM2+I+G was suggested and GTR+I+G was used. For *atp*A, TPM2uf+I+G was suggested and GTR+I+G was used. For *mat*R, TPM1uf+G was suggested and GTR+G was used. All these models were selected as best fit under AIC in JMODELTEST. The Yule process of speciation (Gernhard, 2008) was set as tree prior in both analyses.

In the first analysis, calibration-point priors were modelled as lognormal distributions with the age of dated angiosperm fossils from literature as offset values. The mean

and standard deviation of these distributions were set to 5.0 and 1.25, respectively, assuming an optimum slightly older than the fossils and a 95% confidence interval spanning roughly 10 million years. Based on a recent review (Iles *et al*., 2015), we selected the following six fossil calibration points.

1. On the basis of fossil seeds of *Spirematospermum chandlerae*, the stem node of Zingiberales was constrained to a minimum age of 72.1 Ma (Friis, 1988; Sohl & Owens, 1991).

2. On the basis of fossil leaves from *Sabalites carolinensis*, the crown node of Arecaceae was constrained to a minimum age of 83.6 Ma (Berry, 1914; Gohn *et al.*, 1992).

3. On the basis of fossil epidermis and phytoliths from *Changii indicum*, the crown node of Poaceae was constrained to a minimum age of 66 Ma (Courtillot & Renne, 2003; Prasad *et al.*, 2011).

4. On the basis of fossil pollen from *Liliacidites*, the stem node of Asparagales was constrained to a minimum age of 93 Ma (Ramirez *et al.*, 2007; Gustafsson *et al.*, 2010).

5. On the basis of fossil leaves from *Ripogonum tasmanicum*, the stem node of Ripogonaceae was constrained to a minimum age of 51 Ma (Carpenter *et al.*, 2007; Conran *et al.*, 2009).

6. On the basis of fossil leaves of *Luzuriaga* the stem node of this genus was constrained to a minimum age of 23.2 Ma (Lindqvist & Lee, 2009; Conran *et al.*, 2014). This node corresponds with the crown node of Alstroemeriaceae for our taxon sampling.

The root age of the tree was modelled as a normal distribution (mean 0.0, standard deviation 1.0), constrained to an offset age of 134 Ma, based on the crown age of the monocots as found by Bremer (2000). The substitution rate was set to 1.0 and priors for each substitution were modelled as gamma distributions with default parameter values. The gamma shape prior of the substitution model, and the priors of the mean and standard deviation of the unlinked uncorrelated lognormal clock models were set as exponential distributions. The default settings were kept, except for the mean of the clock models, which was set to 10.0. All other priors were kept to their default settings. Clades in the phylogeny were constrained according to the latest insights in monocot phylogeny (Chase *et al.*, 2006; Soltis *et al.*, 2011; Barrett *et al.*, 2013) and recent Dioscoreales (Merckx *et al.*, 2009a) and Pandanales (Mennes *et al.*, 2013) phylogenies, as well as the multi-locus Liliales phylogeny from this study. Clades were constrained to order level, but clades within Dioscoreales, Pandanales and Liliales were constrained to family or intrafamilial level. The chain

length of the Markov chain Monte Carlo (MCMC; Geyer, 1991) algorithm was 200 million generations, sampling one tree every 10,000 generations. A maximum clade credibility tree was built using TreeAnnotator 1.7.5 (Drummond *et al.*, 2012). The resulting trees were evaluated using Tracer 1.6 (Rambaut *et al.*, 2013) by checking the ESS values for model parameters (see phylogenetic analyses section) and the stability and distribution of the obtained likelihood values (an approximation of the prior distribution of the values was considered sufficient). Trees were visualized using FigTree 1.4 (Rambaut, 2012). In the second analysis, only the fossil calibration points from Liliales [i.e. (5) and (6)] were included, using the following parameter settings. The substitution rate was set to 1.0, priors for each substitution were modelled as gamma distributions (default parameter values), the gamma shape prior of the substitution model and the priors of the mean and standard deviation of the clock models were set as exponential distributions. The mean of the clock models was set to 10.0, and the remaining settings were kept to their default settings. Additionally, the posterior probability distribution obtained for the crown age of Liliales resulting from the first divergence times estimation was set as a secondary calibration point. This prior was modelled as normal distribution (mean 85; sigma 9.7). All remaining prior settings were identical to the first analysis and kept to their default settings.

3.4. Results

3.4.1. Phylogenetic analyses

The BI phylogenetic inferences for the complete three-locus dataset (5741 bp) recovered Corsiaceae as monophyletic with strong support (100% posterior probability, PP; 100% bootstrap support, BS). This clade was sister to Campynemataceae (Liliales), also with strong support (100% PP, 100% BS; Fig. 9). *Arachnitis* and *Corsia* were recovered in this position independently in analyses of the equivalent data set that excluded either *Arachnitis* (see Fig. S6 in Appendix B; based on 5822 bp) or *Corsia* (see Fig. S7 in Appendix B; based on 5881 bp). The phylogenies resulting from the BI analyses are not in conflict with those resulting from the ML analyses (not shown), although removal of either *Arachnitis* or *Corsia* resulted in a slight topological change within Triuridaceae for both BI and ML analyses (concerning the relative position of *Triuris* spp. and *Sciaphila ledermannii*); this was not a strong conflict, in that only the BI analysis depicted a relationship with strong support (100% PP; Fig. 9, Appendix B Figs S6 & S7). In the analysis that excluded *Arachnitis*, the standard deviation of split frequencies was 0.015 (i.e. higher than 0.01) after 1,000,000 additional generations.

The phylogeny resulting from the BI analysis of the five-locus dataset (Fig. 10; 6965 bp) is not in conflict with that resulting from the ML analysis (not shown). The trees resulting from separate analyses of individual genes (see Figs S8–S12 in

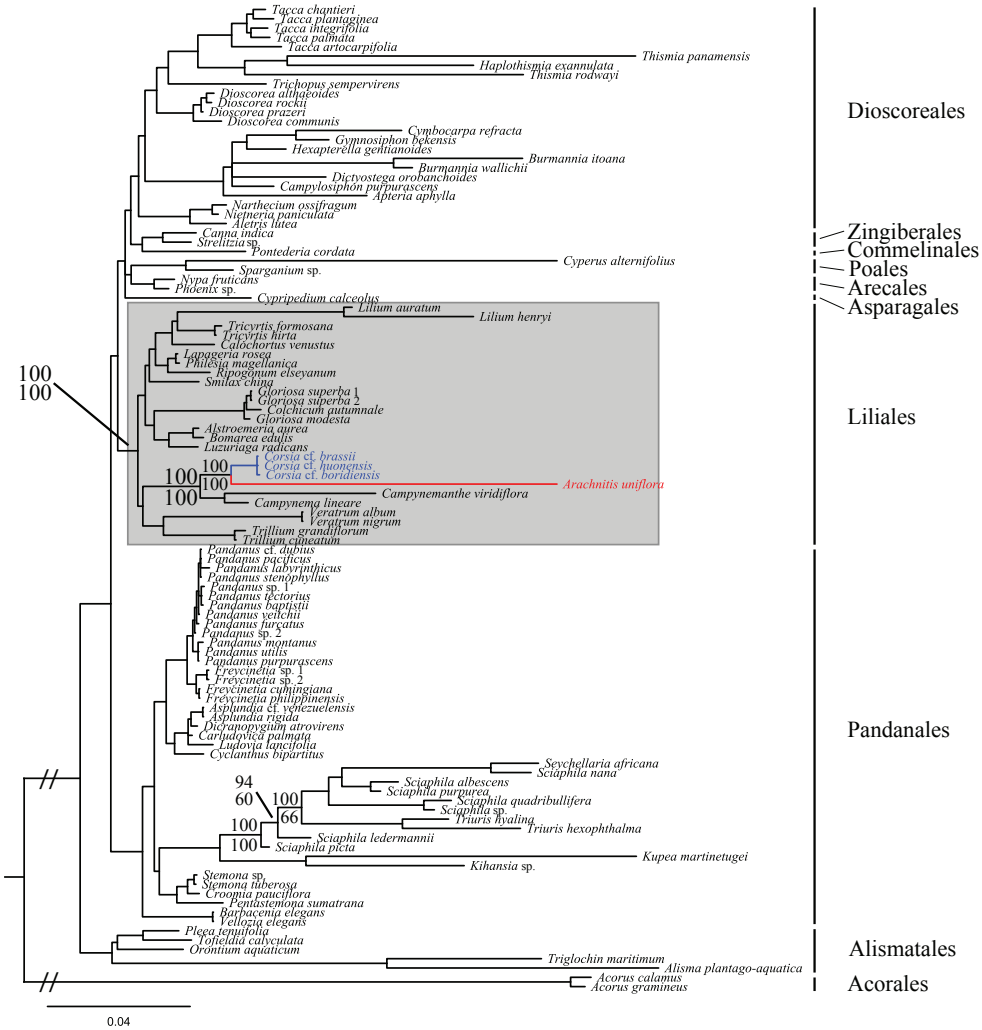

Fig. 9. Phylogeny of Liliales and related monocot orders based on a Bayesian-inference (BI) analyses of a three-locus molecular dataset (nuclear 18S rDNA and mitochondrial *atp*A and *mat*R), including all available *Corsia* (blue) and *Arachnitis* (red) species from this study. The grey shaded area shows the Liliales taxa. Values above branches are posterior probabilities (expressed as percentages) resulting from the BI analysis; those below branches are bootstrap support percentages resulting from the maximum-likelihood (ML) analysis. Support values are only given for Corsiaceae and the weakly conflicting Triuridaceae taxa (see text). Dashes ("//") indicate branches that are longer than indicated in the figure. The scale bar indicates the number of substitutions per site. (See Figs S6 & S7 in Appendix B for phylogenies based on the same analyses, but excluding *Arachnitis* and *Corsia*, respectively.)

Appendix B) do not show highly supported conflict. In the separate *nad5* analysis, the standard deviation of split frequencies was 0.023 after 1,000,000 additional generations. Additionally, in one of the runs, a single effective sample size (ESS) did not reach a sufficient value (i.e. above 100), namely the ESS of tree length (sum

of all branch lengths, TL)=69. With the current taxon sampling, the monophyly of Liliales is strongly supported (100% PP; 100% BS), and all the families in Liliales were recovered as monophyletic. Within Corsiaceae, *Corsia* was represented by three species that formed a strongly supported monophyletic group (100% PP; 100% BS). A group comprising Corsiaceae and Campynemataceae was also recovered as monophyletic (100% PP; 100% BS). Limited support (78% PP; not recovered in ML analysis) was found for a clade consisting of all Liliales families excluding Corsiaceae and Campynemataceae, sister to the latter two families.

The plastid dataset contained full gene data (protein-coding genes and plastid rDNA genes) for representatives of all monocot orders, supplemented by additional angiosperms. A total of 83 plastid genes were retrieved for most species (*ycf*1 was retrieved but not included in the analysis due to alignment difficulties). Corsiaceae was represented by *Arachnitis uniflora* and *Corsia* cf. *boridiensis*. For *A. uniflora*, 20 non-photosynthetic genes were retrieved, and for *C.* cf. *boridiensis*, 27 non-photosynthetic genes were retrieved (see Appendix A.2). No photosynthetic genes were retrieved as open reading frames in the plastid genomes for Corsiaceae, although several pseudogenes were recovered. A number of tRNA sequences were

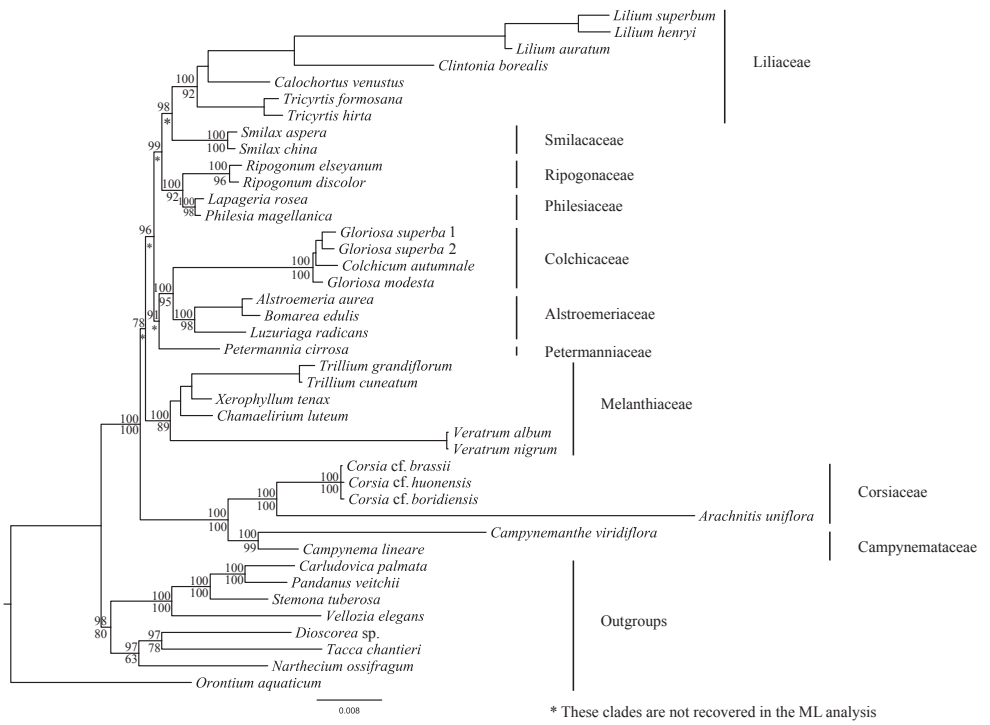

Fig. 10. Phylogeny of Liliales based on a Bayesian-inference (BI) analysis of a five-locus molecular dataset (nuclear 18S rDNA and mitochondrial *atp*A, *mat*R, *cob* and *nad*5). Values above branches are posterior probabilities (expressed as percentages) resulting from the BI analysis; those below branches are bootstrap support percentages resulting from the maximum-likelihood (ML) analysis. The scale bar indicates the number of substitutions per site.

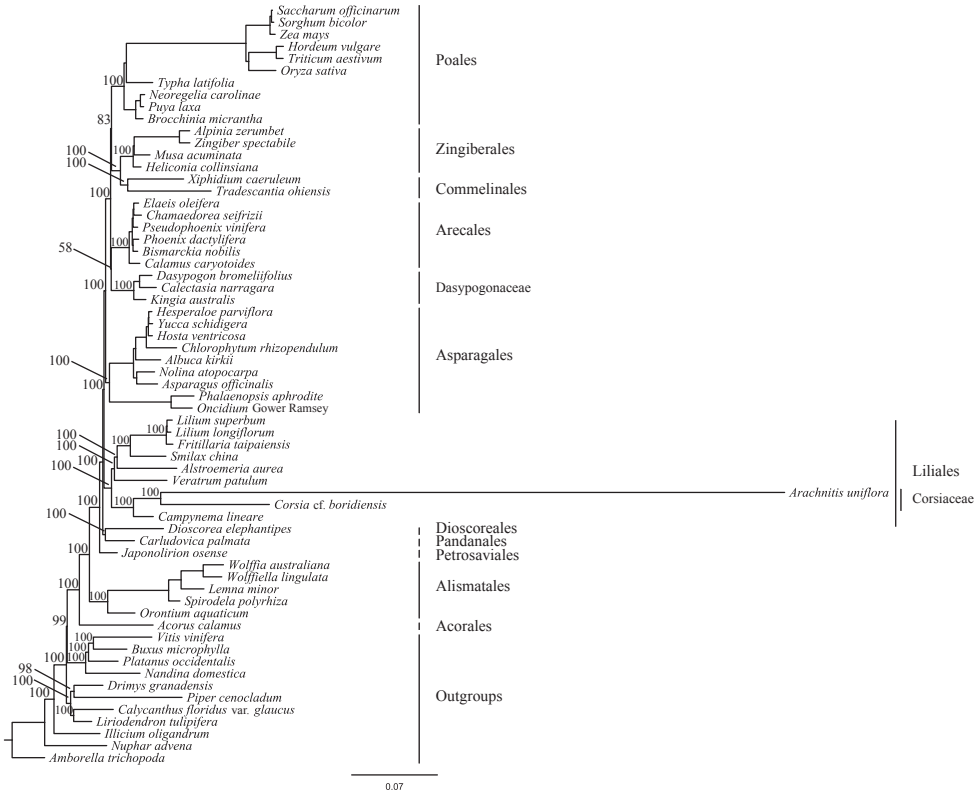

Fig. 11. Phylogeny of monocots based on a maximum-likelihood (ML) analysis of an 82-gene molecular plastid genome dataset. Values above branches are bootstrap support percentages. The scale bar indicates the number of substitutions per site.

also retrieved, but were not included in the analysis. With the current taxon sampling (which lacks any representatives of Colchicaceae, Ripogonaceae, Petermanniaceae or Philesiaceae), the relationships in the order Liliales agree with the results from the five-locus analysis (Fig. 11). Corsiaceae and Campynemataceae are sister to all other Liliales and, notably, the clade containing all other sampled members of Liliales is strongly supported (100% BS). The relationships among the rest of the monocots also agree with those in recent studies (e.g. Soltis *et al.*, 2011; Barrett *et al.*, 2013).

3.4.2. Estimation of divergence times

The estimated ages for Liliales clades are roughly the same in both analyses, except for Corsiaceae, which had notably older age estimates in the three-locus analysis (Table 3). For Corsiaceae, the three-locus analysis resulted in a mean stem age of 70 Ma (95% confidence interval, CI, 47–93 Ma; Fig. 12). The mean crown age of Corsiaceae was estimated at 53 Ma (95% CI, 30–76 Ma; Fig. 12). The two-locus analysis resulted in a mean stem age of 54 Ma (95% CI, 30–78 Ma; see Fig. S13 in Appendix B), and a mean crown age of 36 Ma (95% CI, 14–58 Ma; Fig. S13).

Table 3. Estimated divergence times of Liliales.

		Age estimation (mean and 95% confidence interval in Ma)	
Family	Node	Two-locus monocot-wide analysis	Three-locus Liliales analysis
Alstroemeriaceae	Stem	44 (28-61)	45 (27-63)
	Crown	26 (23.2-34)	28 (23.2-39)
Campynemataceae	Stem	54 (30-78)	70 (47-93)
	Crown	38 (12-62)	41 (16-70)
Colchicaceae	Stem	44 (28-61)	45 (27-63)
	Crown	10 (2-21)	9 (3-17)
Corsiaceae	Stem	54 (30-78)	70 (47-93)
	Crown	36 (14-58)	53 (30-76)
Liliaceae	Stem	59 (44-73)	56 (42-69)
	Crown	53 (38-68)	46 (31-60)
Liliales	Stem	110 (97-122)	na
	Crown	85 (67-104)	90 (75-106)
Melanthiaceae	Stem	74 (59-90)	77 (63-93)
	Crown	54 (32-77)	59 (36-80)
Petermanniaceae	Stem	56 (37-74)	na
	Crown	na	na
Philesiaceae	Stem	53 (51-58)	53 (51-56)
	Crown	6 (0.3-16)	6 (0.4-14)
Ripogonaceae	Stem	53 (51-58)	53 (51-56)
	Crown	na	na
Smilacaceae	Stem	59 (44-73)	56 (42-69)
	Crown	na	na

The estimates obtained in the two-locus monocot-wide analysis for the crown and stem-node ages of monocot orders are generally in accordance with the findings of earlier studies (Table 4). Within Liliales, the results of Chacón *et al.* (2012) resemble our results in showing a crown age of Campynemataceae of 36.5 Ma (95% CI, 13–61 Ma), compared to our estimate of 41 Ma (95% CI, 16–70 Ma). Differences between the two studies are probably the result of differences in the use of a fossil calibration point in *Luzuriaga*, reflecting our taxon sampling – we conservatively constrained the crown node of Alstroemeriaceae using this fossil, whereas Chacón *et al*. (2012) were able to constrain a subclade in this family using the same fossil, because of their denser taxon sampling (see Iles *et al*., 2015, for justification).

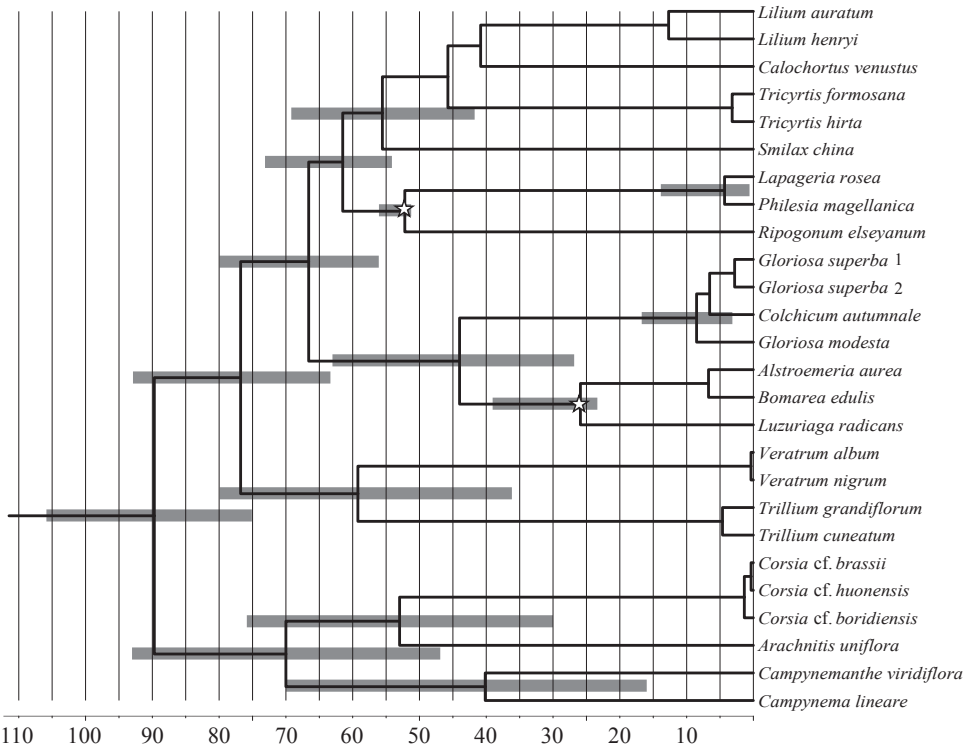

Fig. 12. Maximum clade credibility tree resulting from the divergence time estimation of Liliales inferred using the uncorrelated lognormal relaxed clock method, based on a three-locus molecular dataset (nuclear 18S rDNA and mitochondrial *atp*A and *mat*R), two fossil calibration points, one secondary calibration point and assuming a Yule process of speciation (see text). The scale bar and the numbers in the figure both represent time (million years ago; Ma), and bars around nodes represent 95% confidence intervals. Nodes with 95% confidence interval bars were constrained. Stars indicate fossil calibration points (see text).

3.5. Discussion

3.5.1. Phylogenetic affinities and monophyly of Corsiaceae

The monophyletic relationship revealed between *Corsia* and *Arachnitis* (Figs 9–11) supports the current circumscription of the family (van Royen, 1972; Cronquist, 1981; Dahlgren *et al.*, 1985; Neinhuis & Ibisch, 1998; Zhang *et al.*, 1999; Jones & Gray, 2008). Although the members of Corsiaceae form a well-supported clade, the genus *Arachnitis* is always placed on a rather long branch (Figs 9–11). Thismiaceae taxa are also placed on long branches, suggesting that the weakly recovered affinities between *Arachnitis* and *Thismia* in an earlier analysis (Neyland & Hennigan, 2003) were the result of long-branch attraction (Felsenstein, 1978), possibly exacerbated by the relatively limited taxon sampling in that study. In our analyses, *Arachnitis* is sister to Campynemataceae rather than to *Thismia* when *Corsia* is excluded (and *Corsia* emerges as sister to Campynemataceae after the removal of *Arachnitis*) (Figs

Table 4. Estimated divergence times of monocots.

		95% confidence interval (Ma)			
Order	Node	This study	Mennes et al. 2013	Merckx et al. (2008)	Janssen & Bremer (2004) (Ma)
Acorales	Stem	132-136	134-138	134	na
	Crown	4-46	4-52	7-44	na
Alismatales	Stem	116-135	116-137	123-133	131
	Crown	89-132	90-134	97-133	128
Petrosaviales	Stem	108-132	108-133	121-132	126
	Crown	9-91	8-96	87-102	123
Dioscoreales	Stem	96-123	96-123	119-130	124
	Crown	85-116	85-116	113-126	123
Pandanales	Stem	96-123	96-123	119-130	124
	Crown	70-110	69-110	116-130	114
Liliales	Stem	97-122	97-121	109-131	124
	Crown	67-104	43-105	78-131	117
Asparagales	Stem	94-118	94-116	98-126	122
	Crown	63-113	93-96	101-127	119
Arecales	Stem	87-109	90-93	94-122	120
	Crown	84-90	14-71	15-98	110
Commelinales	Stem	72-82	83-85	83-114	114
	Crown	47-82	47-84	50-104	110
Zingiberales	Stem	72-82	83-85	91-116	114
	Crown	26-66	30-75	52-96	88
Poales	Stem	88-111	84-105	89-120	117
	Crown	82-104	78-98	88-116	113

S6 & S7). Moreover, both the analyses resulting from the nuclear and mitochondrial datasets (Figs 9 & 10) and the plastid dataset (Fig. 11) independently showed strong support for a sister-group relationship between *Arachnitis* and *Corsia*. These results collectively indicate that the placement of *Arachnitis* inferred in this study is not an artefact of long-branch attraction; the family is monophyletic under the current taxon sampling. The inclusion of more species of Corsiaceae, particularly *Corsiopsis*, would further test the monophyly of the family.

The sister-group relationship of Corsiaceae and Campynemataceae inferred from all data sources investigated here (Figs 9–11) identifies Campynemataceae as the closest chlorophyllous relatives of the fully mycoheterotrophic Corsiaceae. Corsiaceae and Campynemataceae then form a clade that is sister to the rest of Liliales, well

supported by the plastid genome data. These findings moreover are in accordance with the previous molecular analyses of Davis *et al.* (2004), Fay *et al.* (2006) and Petersen *et al.* (2013), based on analyses that included a single taxon of each family, although we recovered higher support values for the clade containing Corsiaceae and Campynemataceae. Erdtman (1952) had earlier suggested the affinity of Corsiaceae with Campynemataceae, based on similarities in pollen surface patterning. Based on plastid *rbc*L sequence data, Kim *et al.* (2013) had found a possible sister-group relationship for Corsiaceae with a clade comprising commelinids, Asparagales and Liliales. However, BLAST results of this sequence (obtained from the authors) showed high sequence similarity (95.8%) with *Aglaia* (Meliaceae, core eudicots), suggesting that the sequence does not belong to *Arachnitis*. Furthermore, *rbc*L is not present in the draft circular plastid genome of *A. uniflora.*

All the studied datasets indicate that Liliales forms a well-supported monophyletic order (Figs 9–11) containing the same monophyletic families as in other recent phylogenetic studies of Liliales (Kim *et al.*, 2013; Petersen *et al.*, 2013). The clade containing all sampled members of Liliales except Corsiaceae and Campynemataceae is more strongly supported by the plastid data than by the mitochondrial and nuclear data. The order is supported by some strong morphological synapomorphies, especially the presence of perigonal (tepal-borne) nectaries, compared with septal (ovary-borne) nectaries in many other monocots (Rudall *et al.*, 2000; Smets *et al.*, 2000). The absence of septal nectaries is often linked with congenital carpel fusion (Rudall, 2002; Remizowa *et al.*, 2010). The relationships between other families within Liliales are roughly in accordance with previous molecular analyses (Kim *et al.*, 2013; Petersen *et al.*, 2013) (Fig. 10). The relationships between most other monocot lineages (Fig. 11) are also in accordance with earlier studies (e.g. Soltis *et al.*, 2011; Barrett *et al.*, 2013).

There are few obvious morphological synapomorphies shared by the families Campynemataceae and Corsiaceae, partly because Corsiaceae have reduced vegetative characters. Erdtman (1952) suggested similarities in pollen sexine sculpturing between Corsiaceae and Campynemataceae (as 'Campynematoideae'). However, more extensive comparative studies of pollen in Liliales (C.A. Furness and P.J. Rudall, RBG Kew, personal communication) show that both families have reticulate or microreticulate pollen sculpturing of a type that is relatively common in Liliales and could represent a plesiomorphic condition for the order as a whole. Pollen of Corsiaceae and *Campynema* is monosulcate (porate in *Corsia*), but inaperturate in *Campynemanthe*. Perhaps the strongest similarity between the two families is the shared presence of an inferior ovary, a feature that is otherwise rare in Liliales, although also present in Alstroemeriaceae (Rudall, 2002; Rudall & Eastman, 2002). The inferior ovary is unilocular in *Arachnitis* and probably also in *Corsiopsis*, but trilocular in *Corsia* and Campynemataceae. The style is almost entirely fused in *Corsia* (albeit trilobed at the tip), but has separate styluli in *Arachnitis* and

Campynemataceae, whereas the stigmas are sessile and fused in *Corsiopsis* (Zhang *et al.*, 1999). The primary synapomorphy of Corsiaceae is the presence of a labellum that is (uniquely among monocots) formed from the outer median tepal (Rudall & Bateman, 2002). The labellum is reflexed forwards in *Arachnitis* and *Corsiopsis*, but more upright in some species of *Corsia,* in which it has a prominent callus that could be nectariferous, like the perigonal nectaries of other Liliales, or (perhaps more likely) could represent osmophores. No labellum is present in Campynemataceae, although the tepals also possess thickened regions.

3.5.2. Biogeographical history of Corsiaceae

The estimated divergence times are roughly in accordance with previous studies, as shown by overlapping 95% confidence intervals (Table 4; Vinnersten & Bremer, 2001; Janssen & Bremer, 2004; Merckx *et al.*, 2008; Chacón *et al.*, 2012; Mennes *et al.*, 2013). Differences from earlier studies are likely to be the result of different fossil calibration points or a different tree topology used, coupled with the use of less comprehensive gene sampling [i.e. the use of a single *rbc*L sequence by Vinnersten & Bremer (2001) and Janssen & Bremer (2004)]. The two divergence-time estimates in this study indicate similar ages for families in Liliales (Table 3). Notable differences are found in the estimated stem and crown ages of Corsiaceae, however. This source of uncertainty may be caused by the branch leading to *Arachnitis* being much longer than that leading to *Corsia*, indicating rather strong heterogeneity of substitution rates, which we suspect might have had an effect on the estimated divergence times. Moreover, the low support values for the sister clade of Corsiaceae and Campynemataceae (the rest of Liliales) resulting from analyses of nuclear and mitochondrial data, might further complicate the analysis. Inspection of ESS values indicates that the two-locus monocot dataset only reached moderately high values (between 200 and 300) for Corsiaceae, Campynemataceae and the clade consisting of both families, after 200 million generations. The three-locus Liliales dataset reached ESS values above 3000 for the same clades, and we therefore conclude that this dataset may provide more reliable divergence-time estimates for these nodes.

The inferred sister-group relationship between the genera *Arachnitis* and *Corsia* establishes Corsiaceae as a genuine example of a disjunct South American–Australasian plant lineage. Based on the three-locus Liliales dataset, the estimated age for the crown node of Corsiaceae (53 Ma; 95% CI, 30–76 Ma; Fig. 12) overlaps tightly with age estimates of the tectonic splits between Australasia and South America/Antarctica (35.5-52 Ma) and South America and Antarctica (36 Ma) (Scotese *et al.*, 1988; Veevers *et al.*, 1991; Woodburne & Case, 1996). This implies that the current disjunct distribution of Corsiaceae is best explained by Gondwanan vicariance, although the lower boundary of the 95% CI (30 Ma) leaves the possibility that the distribution of the family is the result of a more recent dispersal event. Moreover, recent dispersal events combined with extinction (possibly in Australia)

cannot be completely ruled out.

Nevertheless, our results are consistent with a scenario in which Corsiaceae provides a rare example of vicariance between South America and Australasia in Southern Hemisphere biogeography. The mycoheterotrophic lifestyle of the species of Corsiaceae might have prevented subsequent dispersal events, as these species can only grow in symbiosis with mycorrhizal fungi. *Arachnitis uniflora* is known to be highly specialized on a clade of arbuscular mycorrhizal fungi (Bidartondo *et al.*, 2002), which may have limited its dispersal capabilities in the past. Both Corsiaceae and the Tasmanian/New Caledonian Campynemataceae are distributed in the southern Pacific region. The inferred mean stem age of Corsiaceae, which is also the date of its split from Campynemataceae (70 Ma; 95% CI, 47–93 Ma; Fig. 12), is on average older than the age estimates of the tectonic splits between Australasia and South America/Antarctica (35.5-52 Ma), and between South America and Antarctica (36 Ma) (Scotese *et al.*, 1988; Veevers *et al.*, 1991; Woodburne & Case, 1996). This suggests that the two families diverged in Gondwana, although the lower boundary of the 95% CI (47 Ma) leaves the possibility open that the split between the families is more recent than the mentioned tectonic splits. The divergence between *Campynema* (Tasmania) and *Campynemanthe* (New Caledonia), however, is estimated to be more recent (16–70 Ma; Fig. 12) than the plate-tectonic split between Australia and New Caledonia/New Zealand (80 Ma) (Raven & Axelrod, 1972; Scotese *et al*., 1988). This implies that the current distribution of Campynemataceae is explained by a long-distance dispersal event, as found for many New Caledonian taxa (e.g. Sapotaceae; Swenson *et al.*, 2014). The age estimates based on the two-locus monocot-wide dataset are younger for these families, particularly for Corsiaceae (Table 3), which makes long-distance dispersal events slightly more likely. The mean value for the estimated crown age of Corsiaceae (36 Ma), however, is still in accordance with the described scenarios. We assume that including the Chinese *Corsiopsis* would not alter the findings of this study, although it might involve an additional scenario (e.g. a long-distance dispersal event) within Corsiaceae. Moreover, the placement of *Corsiopsis* might alter the monophyly of Corsiaceae as currently circumscribed. There is, however, only a single collection for *Corsiopsis* (Zhang *et al.*, 1999), and it is not clear whether the genus is extant.

3.5.3. Plastome evolution in Corsiaceae

Our plastome dataset indicates gene loss and putative pseudogenization of genes involved in photosynthesis [i.e. genes involved in ATP synthase, NADH dehydrogenase, electron transport (cytochrome), and photosystems I and II; Wicke *et al.*, 2011] compared with chlorophyllous Liliales species (Appendices A.1 & A.2). We can assume that pseudogenization and the initial plastid gene-loss events related to the loss of photosynthesis took place between the inferred crown and stem ages of Corsiaceae, which implies an apomorphy age for mycoheterotrophy in Corsiaceae of

between 30 and 93 Ma. *Arachnitis* seems to have a more degraded plastid genome, as several further genes (the group II intron maturase *mat*K and open reading frames encoding the plastid genes of uncertain function *ycf*1 and *ycf*2) appear to be lost from it but are retained in *Corsia*. Because the split between these genera was estimated as 53 Ma (95% CI, 30–76 Ma), these gene losses in *Arachnitis* are likely to have occurred in the last 76 Myr. The *mat*K locus is hypothesized to be one of the last genes lost after the origin of mycoheterotrophy (Barrett & Davis, 2012). The largest class of genes to be retained include genetic apparatus genes (ribosomal protein genes, ribosomal RNA regions and tRNA genes), similar to the findings in the mycoheterotrophic genus *Petrosavia* (Petrosaviaceae), although some genes directly involved in photosynthesis were retained in that genus (Logacheva *et al.*, 2014) but are not present in Corsiaceae. The retained non-photosynthetic plastid genes show relatively longer branches in the plastid phylogeny (Fig. 11) than those based on the nuclear and mitochondrial data (Fig. 10), and not all genes in this functional class have been retained in the plastid genome, showing that losses are not restricted to photosynthesis-related genes. These results will be addressed in a follow-up study.

3.6. Acknowledgements

The authors thank Elke Bellefroid, Martin Bidartondo, Barbara Briggs, Tom Givnish, Ray Neyland, Gerda van Uffelen, Ed de Vogel and Rogier van Vugt for their help and assistance in obtaining plant material. Furthermore, we thank William Iles for assisting with the selection of fossil calibration points, Marcel Eurlings for technical support and Mauricio Renny for providing photographs. V.M. is financially supported by NWO (Veni fellowship 863.11.018); support from an NSERC Discovery grant to S.W.G., and U.B.C. and NSERC postgraduate fellowships to V.L. are also gratefully acknowledged.