

Evolutionary relationships in Santalales inferred using target capture with Angiosperms353, focusing on Australasian Santalaceae *sensu lato*

Benjamin M. Anderson^{A,*} , Maja Edlund^B, Shelley A. James^A , Brendan J. Lepschi^C , Daniel L. Nickrent^D , Amir Sultan^E , Jennifer A. Tate^F  and Gitte Petersen^B 

For full list of author affiliations and declarations see end of paper

*Correspondence to:

Benjamin M. Anderson
Western Australian Herbarium, Department of Biodiversity, Conservation and Attractions, Locked Bag 104, Bentley Delivery Centre, Bentley, WA 6983, Australia.
Email: benjamin.anderson@dbca.wa.gov.au

Handling Editor:

Daniel Murphy

Received: 15 August 2024

Accepted: 30 April 2025

Published: 17 June 2025

Cite this: Anderson BM *et al.* (2025)

Evolutionary relationships in Santalales inferred using target capture with Angiosperms353, focusing on Australasian Santalaceae *sensu lato*. *Australian Systematic Botany* **38**, SB24026. doi:10.1071/SB24026

© 2025 The Author(s) (or their employer(s)).

Published by CSIRO Publishing.

This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND)

OPEN ACCESS

ABSTRACT

The angiosperm order Santalales comprises more than 2500 species, most of which are hemi- or holoparasitic on other plants, and derive water and nutrients via specialised structures that attach to host roots or stems. The parasitic lifestyle has affected the morphology and genomes of these plants, and classification of the order has been difficult, with outstanding questions about membership of and relationships between families in the order. We chose to focus on improving phylogenetic sampling in the broadly circumscribed Santalaceae *sensu lato*, with emphasis on Australasian members of Amphorogynaceae and Viscaceae as part of the Genomics for Australian Plants Initiative. We used target capture with the Angiosperms353 bait set to generate a dataset of 318 nuclear loci × 195 samples, including publicly available data from other Santalales families. Phylogenetic inferences using maximum likelihood concatenation and a summary coalescent approach were largely congruent and resolved relationships between most families, agreeing with much of the previous work on the order. Some relationships that have been difficult to resolve remained so, such as branching order among some families in Olacaceae *sensu lato* and Santalaceae *sensu lato*. Denser sampling in Amphorogynaceae and Viscaceae provided new insights into species-level relationships in genera such as *Leptomeria* and *Choretrum*, and allowed testing of recent phylogenetic work in *Korthalsella*. Our new phylogenetic hypothesis is consistent with one origin of root hemiparasitism, two origins of holoparasitism and five origins of aerial parasitism in the order. Although Angiosperms353 was successful, some phylogenetic bias in gene recovery suggests that future studies may benefit from more specific baits and deeper sequencing, especially for Viscaceae.

Keywords: Amphorogynaceae, Australia, *Choretrum*, Genomics for Australian Plants, *HybPiper*, *Korthalsella*, *Leptomeria*, mistletoe, New Zealand, parasitic plant, phylogenomics, Viscaceae.

Introduction

The angiosperm order Santalales comprises more than 180 genera and more than 2500 species (Nickrent *et al.* 2010; Nickrent 2020; Govaerts *et al.* 2024), the vast majority of which are parasitic. Parasitic plants obtain water, inorganic nutrients and sometimes photosynthate from host plants by connecting to the vascular system of roots and/or stems using specialised structures called haustoria. Parasitic plants can be categorised based on whether they completely rely on the host for carbon (holoparasites) or are able to photosynthesise themselves (hemiparasites). All these types of parasitism (holo- and hemi-, root and stem), along with non-parasites, are present in Santalales, making the group an intriguing subject for studying the evolution of parasitism. In particular, the highly specialised life history for germinating on and parasitising only aerial stems (mistletoes) has likely evolved

multiple times in the order (Vidal-Russell and Nickrent 2008), being present in Misodendraceae, Loranthaceae, Santalaceae, Amphorogynaceae and Viscaceae (Nickrent 2020).

As parasitism can have a profound effect on the morphology and genomes of these plants, classification has been challenging. The evolution of parasitism is correlated with changes to selection pressures on the plastome that are even evident in hemiparasites rather than only the holoparasites that fully rely on a host for carbon (Petersen *et al.* 2015b; Wicke *et al.* 2016). In Santalales, the evolution of the mistletoe habit coincides with increasing morphological specialisation and sometimes reductions in vegetative and floral traits (Vidal-Russell and Nickrent 2008) as well as progressive impacts to plastome gene evolution relative to root parasitic relatives (Petersen *et al.* 2015b; Chen *et al.* 2020). At the extreme, holoparasites exhibit massive morphological changes (e.g. a largely subterranean lifestyle and reductions of leaves and flowers) compared to their photosynthetic relatives and huge reductions and changes in their plastomes (e.g. Su *et al.* 2019). These morphological changes have made the taxonomic placement of holoparasites difficult (see e.g. Su *et al.* 2015), while rate changes among and even the loss of plastid loci may pose challenges to phylogenetic inference in molecular studies, which have historically relied on plastid loci to resolve family relationships (e.g. Chase *et al.* 1993; Ruhfel *et al.* 2014).

Since the 19th Century, concepts of the order have varied, particularly regarding which families are included and the circumscription of families and genera. See Kuijt (2015) for a brief review of the early taxonomic history of Santalales. As pointed out by Kuijt (2015), Schellenberg (1932) provided the first modern concept of the order, although families were included that are now excluded (e.g. Grubbiaceae) and the holoparasitic Balanophoraceae was considered an unrelated convergent group. Schellenberg (1932) also recognised a trend in morphological transformation (reduction) of ovules from Olacaceae through to Loranthaceae (including Viscaceae). In a single volume of the second edition of *Die Natürlichen Pflanzenfamilien*, treatments of most genera of Santalales were grouped together into some of the families that are currently recognised: Olacaceae (Sleumer 1935a), Opiliaceae (Sleumer 1935b), Octoknemaceae (Mildbraed 1935), Santalaceae (Pilger 1935), Misodendraceae ('Myzodendraceae'; Skottsberg 1935) and Loranthaceae (Engler and Krause 1935). In the same volume, distantly related parasites *Rafflesia* R.Br. ex Gray and *Hydnora* Thunb. were treated in Aristolochiales, and Balanophoraceae (Harms 1935) was treated in its own order, Balanophorales.

Molecular phylogenetic work has progressively clarified membership and relationships in the order, although outstanding questions remain. Most of the analyses in the group to date have relied on nuclear ribosomal DNA (18S rDNA, 18S; 26S rDNA, 26S; and *ITS*), a handful of markers from the plastid genome (*accD*, *matK*, *rbcL*, *trnL-F*) and rarely the mitochondrial genome (*matR*). Early work (Nickrent and Duff 1996; Nickrent *et al.* 1998) highlighted the genus *Schoepfia*

Schreb. as distinct from a paraphyletic Olacaceae and sister to *Misodendrum* Banks ex DC., and a paraphyletic Santalaceae with Viscaceae embedded, leading the Angiosperm Phylogeny Group (APG) classifications (The Angiosperm Phylogeny Group 1998, 2003) to recognise five families in the order: Olacaceae, Opiliaceae, Loranthaceae, Misodendraceae and Santalaceae (including Viscaceae). A flurry of studies followed APG I and II to clarify relationships in the order. Balanophoraceae (as recognised at the time) was shown to be closely related to Santalales (Nickrent *et al.* 2005), Olacaceae consisted of seven clades but was still paraphyletic (Malécot and Nickrent 2008) and Santalaceae consisted of seven supported clades but with uncertain relationships (Der and Nickrent 2008). Improved estimates of relationships in Loranthaceae led to differing hypotheses about the origin(s) of mistletoes in that family (Wilson and Calvin 2006; Vidal-Russell and Nickrent 2008) and similarly more broadly in the order (Vidal-Russell and Nickrent 2008). Given some of the phylogenetic uncertainty, the next update to APG (The Angiosperm Phylogeny Group 2009) continued to maintain broad concepts for Olacaceae and Santalaceae despite support for clades within them, and it recognised Schoepfiaceae and Balanophoraceae as families in the order.

To account for the well supported groups in Olacaceae and Santalaceae and to maintain monophyletic families, a new classification was proposed based on preceding molecular work and morphology (Nickrent *et al.* 2010). The classification split Olacaceae (excluding Schoepfiaceae) into seven families (Erythralaceae, Strombosiaceae, Coulaceae, Ximeniaceae, Aptandraceae, Olacaceae and Octoknemaceae) and Santalaceae into seven families (Comandraceae, Cervantesiaceae, Thesiaceae, Nanodeaceae, Santalaceae, Amphorogynaceae and Viscaceae). We refer to these two groups of families as Olacaceae sens. lat. and Santalaceae sens. lat. respectively. The new classification did not include Balanophoraceae, but later work (Su *et al.* 2015) found that Balanophoraceae formed two distinct clades in Santalales: (1) a clade with *Balanophora* J.R.Forst. & G.Forst. and six other genera (Balanophoraceae sens. strict.) sister to the non-Olacaceae sens. lat. Santalales and (2) a clade with *Mystropetalon* Harv., *Dactylanthus* Hook.f. and *Hachettea* Baill. (Mystropetalaceae) sister to Loranthaceae. While acknowledging the work in Santalales, APG IV (The Angiosperm Phylogeny Group 2016) did not take up the new classification, stating that relationships remained uncertain. APG IV chose to retain admittedly non-monophyletic status quo families Olacaceae sens. lat. and Santalaceae sens. lat. (though listing the new family names) and changed the order of Balanophoraceae to reflect its probable placement in Santalaceae sens. lat., citing unpublished data that purportedly supported Balanophoraceae sens. lat. as monophyletic.

Since APG IV, there have been no major changes to the classification in Santalales, although some improved sampling has continued to support the families recognised in the new classification (Nickrent *et al.* 2019). More recent studies with extensive coverage of genera in Santalales and using

five to seven nuclear, plastid and mitochondrial markers (Su *et al.* 2015; Nickrent *et al.* 2019) provided phylogenetic hypotheses for relationships among families but still faced challenges resolving portions of the backbone of the tree (e.g. Olacaceae sens. lat.) and relationships within some families (e.g. Opiliaceae, Loranthaceae, Santalaceae and Viscaceae). Outstanding questions about relationships in Santalales include: what are the branching orders within Olacaceae sens. lat. and Santalaceae sens. lat., should Balanophoraceae sens. lat. be split to recognise Mystropetalaceae, and how are major clades of Viscaceae mistletoes related? Looking more closely within families, Liu *et al.* (2018) tackled the largest family, Loranthaceae, elucidating relationships within one group of mistletoes. Broader relationships between other mistletoes and their root-parasitic relatives within Santalaceae sens. lat. remain less well studied, as recent work has mostly focused on single genera such as *Korthalsella* Tiegh. (Sultan *et al.* 2019), *Viscum* L. (Maul *et al.* 2019) and *Thesium* L. (García *et al.* 2024), relying on a small number of nuclear and plastid markers.

Alongside broader phylogenetic questions in the order, generic boundaries and relationships within two families of Santalaceae sens. lat., Amphorogynaceae and Viscaceae, have relevance to some of our ongoing research into mitochondrial evolution in *Viscum* and its relatives (see Petersen *et al.* 2020). Molecular sampling of a small number of nuclear and plastid markers in Amphorogynaceae has so far been restricted to one or two species per genus, whereas Viscaceae has had denser species-level sampling for most genera. Since these two families include genera with substantial portions of their diversity in the Australasian region (7/9 genera in Amphorogynaceae, with three root parasitic genera restricted to Australia; 3/7 genera in Viscaceae, along with representatives of widespread genera), we partnered with the Genomics for Australian Plants (GAP, see <https://www.genomicsforaustralianplants.com/>) initiative to broaden taxonomic sampling in these two families while employing a target sequence capture approach using the Angiosperms353 (A353) bait set (Johnson *et al.* 2019). Together with publicly available sequence data from sources such as the Kew Tree of Life Explorer (Baker *et al.* 2021), which have recently been included in a phylogenetic analysis across angiosperms (Zuntini *et al.* 2024), we aimed to provide new phylogenetic hypotheses for the order and within Santalaceae sens. lat., especially among genera in Amphorogynaceae and Viscaceae.

The use of target sequence capture and the A353 bait set is a promising approach to tackling these phylogenetic questions in Santalales, although there are challenges and limitations. Constructing a dataset of more than 300 nuclear loci represents a substantially larger source of phylogenetic information than has previously been used for inferring phylogenies of the order, with most previous studies relying on a handful of nuclear and organellar markers as mentioned above. Given expected conflicting phylogenetic signals among genes (Maddison 1997), increasing the number of independent loci should help to reveal and account for

this conflict when inferring relationships. The use of nuclear data may also help to avoid biases from reliance on the plastid genome, which is known to be highly affected by the evolution of parasitism (e.g. Chen *et al.* 2020), but our approach may not be particularly effective given that a large portion (~40%) of the targets for A353 are nuclear genes with putative chloroplast association (Johnson *et al.* 2019), which might also be affected by changes to chloroplast function associated with parasitism. Finally, although target capture increases sampling potential by allowing the use of herbarium specimens, success can depend on the taxonomic group and preservation of the samples (Brewer *et al.* 2019), and densely sampling at the species level across large and diverse groups can still be expensive. Nevertheless, we set out to use available and newly generated target capture data in Santalales to test phylogenetic relationships, while also evaluating the success of A353 in a group of parasites.

Materials and methods

Sampling

Taxa were chosen based on access to existing herbarium and DNA collections and sequence data from genera across Santalales, with a particular focus on Australasian Santalaceae sens. lat. (Supplementary Table S1). Initial sampling targeted 205 members of Santalales plus two outgroup taxa (*Vitis vinifera* L. and *Berberidopsis beckleri* (F.Muell.) Veldkamp). The Santalales taxa included Olacaceae sens. lat. (7/7 families and 14/29 genera), Balanophoraceae sens. lat. (2/2 families and 4/16 genera), Misodendraceae (1/1 genus), Schoepfiaceae (1/3 genera), Loranthaceae (7/78 genera), Opiliaceae (7/11 genera) and Santalaceae sens. lat. (6/6 families: Comandraceae 1/2 genera, Cervantesiaceae 1/8 genera, Thesiaceae 2/6 genera, Nanodeaceae 2/2 genera, Santalaceae 7/11 genera, Amphorogynaceae 7/9 genera and Viscaceae 7/7 genera). Taxonomic coverage was less complete at the generic level than the two largest previous studies (Su *et al.* 2015; Nickrent *et al.* 2019), but it was denser at the species level in genera of Amphorogynaceae and Viscaceae (see Supplementary Table S2). Publicly available sequence data were obtained for 64 of the samples, mostly from the Kew Plant and Fungal Trees of Life (PAFTOL) project from the Kew Tree of Life Explorer (see <https://treeoflife.kew.org/>) including transcriptomic SRR accessions (sequencing runs submitted to the NCBI Sequence Read Archive). Other downloaded data included 16 samples from the first stage of the GAP Australian Angiosperm Tree of Life (AATOL, see <https://www.genomicsforaustralianplants.com/phylogenomics>) project. The remaining 127 samples were sequenced as part of the current study (48 samples sequenced for AATOL stage 2 and 79 samples sequenced outside Australia). For the samples sequenced as part of this study, most came from herbarium specimens held in the following institutions (codes follow *Index Herbariorum*, see [3](https://sweetgum.</p>
</div>
<div data-bbox=)

nybg.org/science/ih/): ASC, BH, CANB (including CBG), CR, GB, KATH, L, LAE, MAU, MO, MPN, MSB, PERTH, S, UPS and WELTU. Three DNA samples of *Viscum* were obtained from the African Centre for DNA Barcoding (University of Johannesburg).

DNA extraction, library preparation, sequencing and quality filtering

For samples extracted as part of AATOL, ~20–30 mg of tissue was ground in a TissueLyser II (Qiagen) with tungsten carbide beads. Genomic DNA was extracted using the DNeasy Plant mini kit (Qiagen) on a QIAcube Connect (Qiagen). For samples extracted outside Australia, DNA extraction consisted of a modified CTAB protocol (Doyle and Doyle 1987).

Libraries for AATOL samples were prepared with the NEBNext Ultra II FS Library Prep Kit (New England Biolabs, Ipswich, MA, USA) with ~350-base pair (bp) inserts. Libraries were pooled (12–16 plex) and enriched for the Angiosperms353 probes (Johnson *et al.* 2019) by hybridising at 65°C with the Arbor Biosciences MyBaits Expert Plant Angiosperms353 ver. 1 bait set with V5 chemistry (catalogue number 308108.v5). Libraries for sequencing outside Australia were prepared by Arbor Biosciences (Ann Arbor, MI, USA) using an in-house preparation, and sequence capture was done with the same bait kit.

Sequencing was done on an Illumina NovaSeq 6000 at the Australian Genome Research Facility (Melbourne, Vic., Australia) with ver. 1.5 chemistry in the 150-bp paired-end read format. Sequencing for samples outside Australia was undertaken by Arbor Biosciences on an Illumina NovaSeq 6000 in the 150-bp paired-end read format.

Raw sequencing reads were assessed with *FastQC* (ver. 0.11.7, S. Andrews, see <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), which revealed adaptor contamination and non-random sequence content in the first few base pairs of reads, consistent with enzymatic fragmentation. A custom script was run to remove optical duplicates, adaptor sequences, the first 9–10 bp of reads, and any reads less than 50 bp long after trimming, and to correct sequencing errors, all using tools from the *BBMap* package (ver. 35.85, B Bushnell, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, see <https://sourceforge.net/projects/bbmap/>). For downloaded publicly available data associated with PAFTOL in the Kew Tree of Life Explorer (Baker *et al.* 2021), the oneKP initiative (Carpenter *et al.* 2019) or other transcriptomic work, read lengths and formatting varied, so no duplicates were removed and 9, 10 or 12 bp were trimmed from the beginning of reads as appropriate.

Assembly

The *HybPiper* pipeline (ver. 2.1.5, see <https://github.com/mossmatters/HybPiper>; Johnson *et al.* 2016) was used to assemble sequences for all samples (newly sequenced and downloaded). To improve the recovery of targets, we augmented the

default A353 target file with the assistance of the ‘NewTargets’ approach (McLay *et al.* 2021) and transcriptomic data for *Viscum album* L. (Schröder *et al.* 2022). *HybPiper* was run to recover stitched exons, and exons plus introns and flanking regions (‘supercontigs’). As part of the *HybPiper* pipeline, target loci (which *HybPiper* calls ‘genes’) with multiple assembled contigs of varying depth and length may be flagged as potentially having paralogs present, with paralog sequences kept if they pass a length threshold. We assessed how many paralogs *HybPiper* kept per locus to identify and filter loci having ≥ 10 samples with recovered paralogs. We also filtered loci to remove those present in $< 50\%$ of samples. Additional details of how the pipeline and analyses were run along with custom scripts can be found in a GitHub repository (see <https://github.com/bmichanderson/Santalaceae>).

Phylogenetic analyses

Phylogenetic inferences were run for three different alignment sets: (1) exon sequences for all samples, (2) supercontig (exons + introns) sequences for samples in Amphorogynaceae and (3) supercontig sequences for samples of *Korthalsella*, *Ginalloa* Korth., *Phoradendron* Nutt. and *Dendrophthora* Eichler.

To generate the first set of alignments, stitched exons (excluding introns) were translated to protein residues for each locus and aligned with *MAFFT* (ver. 7.453, see <https://mafft.cbrc.jp/alignment/software/>; Katoh and Standley 2013) using the ‘--auto’ mode, then protein residues were converted back to corresponding nucleotide sequences with ‘pal2nal.pl’ (see <http://www.bork.embl.de/pal2nal/distribution/pal2nal.v14.tar.gz>, accessed 6 September 2023; Suyama *et al.* 2006). For the second and third alignment sets, supercontig sequences per locus were aligned with *MAFFT* using the ‘--auto’ mode. All alignments were cleaned with a custom *Python* (ver. 3, Python Software Foundation, see <https://docs.python.org/3/reference/>) script to remove positions with $> 50\%$ missing data and samples with $> 75\%$ missing data per locus.

Maximum likelihood (ML) tree inference of the alignments was run using *IQ-TREE* (ver. 2.2.2, see <http://www.iqtree.org/>; Minh *et al.* 2020b), with a partitioned analysis of all concatenated loci and inferences for each locus, using a search for the optimal model (Kalyanamoorthy *et al.* 2017) and 1000 Ultrafast Bootstrap replicates (Hoang *et al.* 2018), sampling by locus and site in the concatenation analysis. Branches in the locus trees with $< 50\%$ Ultrafast Bootstrap support (UFB) were reduced to polytomies using *Newick Utilities* (see https://github.com/tjunier/newick_utils; Junier and Zdobnov 2010). *IQ-TREE* was used to calculate gene and site concordance factors (see Lanfear and Hahn 2024) for branches in the resulting concatenation trees, using the topologies of the locus trees for gene concordance factors (Minh *et al.* 2020a) and sampling 10,000 quartets of samples per branch for site concordance factors (Mo *et al.* 2023).

A coalescent shortcut approach was used to infer species trees with *ASTRAL* (ver. 5.7.1, see <https://github.com/smirarab/ASTRAL>; Sayyari and Mirarab 2016; Zhang *et al.* 2018), using the locus trees from the ML analyses as input. A first run estimated the species tree and local posterior probabilities for each branch, along with quartet support values. A second run used the '--branch-annotate 10' option to test for polytomies (Sayyari and Mirarab 2018), providing a *P*-value per branch for whether a polytomy could be rejected. Branches with *P* > 0.05 (cannot reject a polytomy) were collapsed into polytomies with *Newick Utilities* for display.

To detect potential misassemblies or samples with alignment problems, the locus trees were run through *TreeShrink* (ver. 1.3.9, see <https://github.com/uym2/TreeShrink>; Mai and Mirarab 2018) to identify branches that were significantly longer than others (at the '0.10' threshold). Detected long branches were used to remove the corresponding sequences from those loci. For the full Santalales alignment, sequences from taxa in Balanophoraceae sens. lat. were retained regardless of whether they were detected as long branches, since the sequences are expected to be highly divergent from other Santalales (see Su *et al.* 2015). Three samples with higher numbers of flagged long branches (similar to samples of Balanophoraceae) and that showed possible issues with contamination or misidentification, alignment or quality were dropped entirely. Following removal of problematic samples and sequences, phylogenetic inferences were run again following the same approach as above.

As a test of reproducibility, we re-ran the primary analysis to generate an overall tree starting from downloading our uploaded reads. Since the European Nucleotide Archive strips Illumina tiling information from sequence headers, we were unable to remove optical duplicates during read trimming, making exact replication of the analysis impossible (see the GitHub repository). We nevertheless proceeded to assess whether our results were consistent.

To reconstruct character evolution in Santalales, we undertook stochastic character mapping (Huelsenbeck *et al.* 2003) of parasitism type (not parasitic, root parasitic and aerial parasitic) adapting the approach in Ramm *et al.* (2020). Importantly, this approach does not incorporate phylogenetic uncertainty and assumes the provided topology is correct, so results should be interpreted cautiously. We generated an ultrametric tree in *treePL* (ver. 1.0, see <https://github.com/blackrim/treePL>; Smith and O'Meara 2012) from the (re-run) ML concatenation tree, calibrating the crown age of Santalales as 125 million–204 million years ago based on two analyses in Zuntini *et al.* (2024). Parasitism types at the tips of the tree were estimated from previous work (Vidal-Russell and Nickrent 2008; Kuijt 2015; Nickrent 2020), with taxa given less informative priors when uncertain (e.g. 0.5 and 0.5 probabilities of not parasitic and root parasitic in Aptandraceae). We ran stochastic character mapping in the *R* package *phytools* (ver. 2.1-1, see <https://cran.r-project.org/package=phytools/>; Revell 2024), simulating 1000 character mappings.

The prior for the root (common ancestor of Santalales and *Vitis* + *Berberidopsis*) was fixed as not parasitic. To clarify evolution of holoparasitism, we ran a second mapping for the states: not parasitic, hemiparasitic and holoparasitic.

Results

Assembly

Following quality filtering of raw reads, three samples had < 300 000 reads and were excluded from the *HybPiper* assembly, leaving 204 to assemble. The assembled samples had on average 281 target loci with a sequence (6–352, s.d. 70) but only 195 loci with at least 50% target length (1–344, s.d. 100) (see Supplementary Table S3). The percentage of reads on-target averaged 5.6% (0.1–23.9%, s.d. 6.2%). Six samples had fewer than 98 target loci with sequences (outliers), so these samples were excluded from further analysis, leaving 198 samples. Nine loci had 10 or more taxa with recovered paralogs, so these loci were dropped. For the full dataset, 26 loci had < 50% sample coverage and were dropped, leaving 318 loci for phylogenetic analysis. For the Amphorogynaceae dataset (47 samples), 13 loci had < 50% sample coverage, leaving 331 loci for analysis, and for the *Korthalsella* dataset (37 samples), 73 loci had < 50% sample coverage, leaving 271 loci for analysis.

Recovery was variable across samples, with some indications of taxonomic and potential methodological biases. Considering families (Table 1), Balanophoraceae sens. lat. and Viscaceae had average recovery (> 50% length) below 200 loci, with Viscaceae below 130. The lower average in Viscaceae relative to the other families is more evident in the A353 samples (the bulk of the dataset), with transcriptomic data from the oneKP project and available SRR accessions performing much better, though only represented by four samples (see Table 1). The pattern is less pronounced when counting loci with a sequence, indicating target capture sequences assembled for Viscaceae were shorter than those for most other Santalales rather than only missing. This was reinforced by the consistent length recovery in the Viscaceae transcriptomic samples. Balanophoraceae sens. lat. and Viscaceae also had low percentage of reads on target (see Table 1). The average percentage of reads on target for Viscaceae samples was biased upwards by samples of *Notothixos* Oliv. (all with > 9 v. < 3.5% for the non-*Notothixos* samples). The Viscaceae transcriptomic samples had a greater sequencing effort than the A353 samples (average > 25 million v. < 9 million reads), which may partly explain the better recovery.

Phylogenetic analyses

The initial analysis of the full Santalales dataset resulted in a filtered alignment of 318 loci with a total aligned length of 185 000 bp and 18.6% missing data. The 198 samples had on average 259 loci (81–318, s.d. 60). Following evaluation

Table 1. *HybPiper* assembly of target loci for samples of Santalales.

Family	Samples	Average reads ($\times 10^6$)	Average percentage on target	Average number of genes	Average number of genes at $\geq 50\%$ target length
Amphorogynaceae	49	17.4 \pm 8.7	11.6 \pm 5.4	328 \pm 20	239 \pm 85
Balanophoraceae sens. lat.	5	21.6 \pm 15.5	1.1 \pm 0.8	222 \pm 20	190 \pm 42
Loranthaceae	10	13.2 \pm 12.7	2.2 \pm 1.8	313 \pm 31	245 \pm 81
Nanodeaceae	4	14.1 \pm 6.1	18.2 \pm 6.2	348 \pm 3	320 \pm 25
Olacaceae sens. lat.	14	9.6 \pm 10.1	5.8 \pm 4.0	322 \pm 54	264 \pm 91
Opiliaceae	8	12.4 \pm 8.5	1.8 \pm 1.4	313 \pm 38	219 \pm 50
Santalaceae	16	16.4 \pm 10.8	10.5 \pm 6.0	334 \pm 24	300 \pm 59
Thesiaceae	4	16.6 \pm 12.0	4.6 \pm 2.7	332 \pm 13	307 \pm 11
Viscaceae (all)	87	9.3 \pm 5.8	2.0 \pm 3.4	233 \pm 67	129 \pm 72
Viscaceae (A353)	83	8.4 \pm 3.6	2.0 \pm 3.5	230 \pm 66	122 \pm 63
Viscaceae (transcriptomic)	4	27.6 \pm 10.6	0.7 \pm 0.3	299 \pm 48	288 \pm 49

Groups with at least four samples are shown. The metrics indicate the average number of reads (\pm s.d.), percentage of reads on target, number of loci and number with sequences at least 50% of the target length.

of locus trees with *TreeShrink*, three samples were found to have alignment issues, possibly from assembly or contamination problems or a lack of data (one sample had < 10 loci at $\geq 50\%$ target length), so these three samples were dropped from further analysis, leaving 195 samples. Following removal of sequences producing long branches in locus trees (except for those of taxa in Balanophoraceae sens. lat.), the second set of analyses resulted in a filtered alignment of 318 loci with a total aligned length of 186 000 bp and 18.5% missing data. The 195 samples had on average 259 loci (94–318, s.d. 60), with 20.7% missing data per locus (0.2–53.3%, s.d. 14.2%).

Phylogenetic relationships among our sampling of Santalales were largely well resolved (Fig. 1, Supplementary Fig. S1), and most families (under the new classification) with more than one sample received full support (100 UFB/1.00 posterior probability, pp) for monophyly. Relationships between families were also mostly well resolved, but there were some portions of the backbone with lower confidence or indications of conflict. The ML concatenation and *ASTRAL* coalescent trees (Supplementary Fig. S1) were largely congruent for relationships between and membership of family-level clades, with the exception of some of the backbone relationships in Olacaceae sens. lat. Within more densely sampled genera (e.g. *Korthalsella*, *Choretrum* R.Br. and *Leptomeria* R.Br.), however, there were indications of lower support in both the concatenation and coalescent approaches, and cases of conflict between these (Fig. 1).

The initial analysis of the Amphorogynaceae dataset resulted in a filtered alignment of 331 loci with a total aligned length of 441 000 bp and 20.4% missing data. The

47 samples had on average 303 loci (134–331, s.d. 43). Following removal of sequences producing long branches in locus trees, the second set of analyses resulted in a filtered alignment of 331 loci with a total aligned length of 445 000 bp and 20.3% missing data. The 47 samples had on average 298 loci (129–331, s.d. 47), with 22.7% missing data per locus (1.5–58.6%, s.d. 17.7%).

The ML concatenation and *ASTRAL* coalescent trees (Fig. 2) were largely congruent and provided support for the monophyly of all genera with multiple samples except for *Dendrotrophe* Miq. and *Dendromyza* Danser, which were nested together. Within both *Choretrum* and *Leptomeria* (Australian taxa), there were primary splits of species diversity into two primary clades on relatively longer branches within the genera. In *Leptomeria*, one primary clade comprised Western Australian species in section *Xeromeria* and three Western Australian species successively sister to it, whereas the other primary clade comprised a clade of three southern and eastern Australian species and a clade of four Western Australian species sister to it. In *Choretrum*, one primary clade comprised four western and southern Australian species, whereas the other primary clade comprised a clade of three eastern Australian species sister to a southern Australian species. Some conflict was evident between the concatenation and coalescent trees for relationships within the primary clades in both genera, in some cases uniting or separating multiple accessions of individual species.

The initial analysis of the *Korthalsella* dataset resulted in a filtered alignment of 271 loci with a total aligned length of 272 000 bp and 21.5% missing data. The 37 samples had on average 222 loci (80–268, s.d. 50). Following removal of

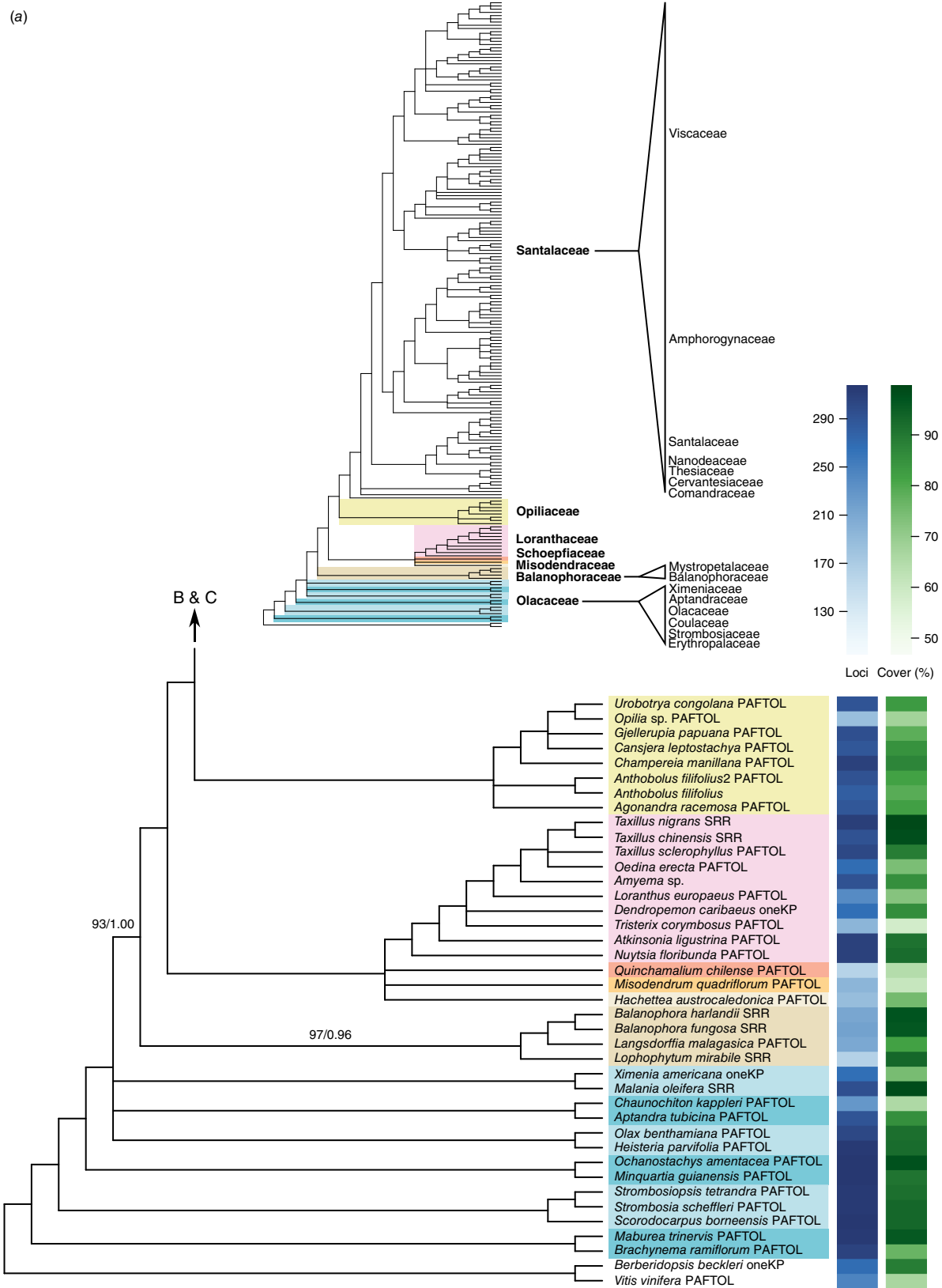


Fig. 1. (Caption on next page)

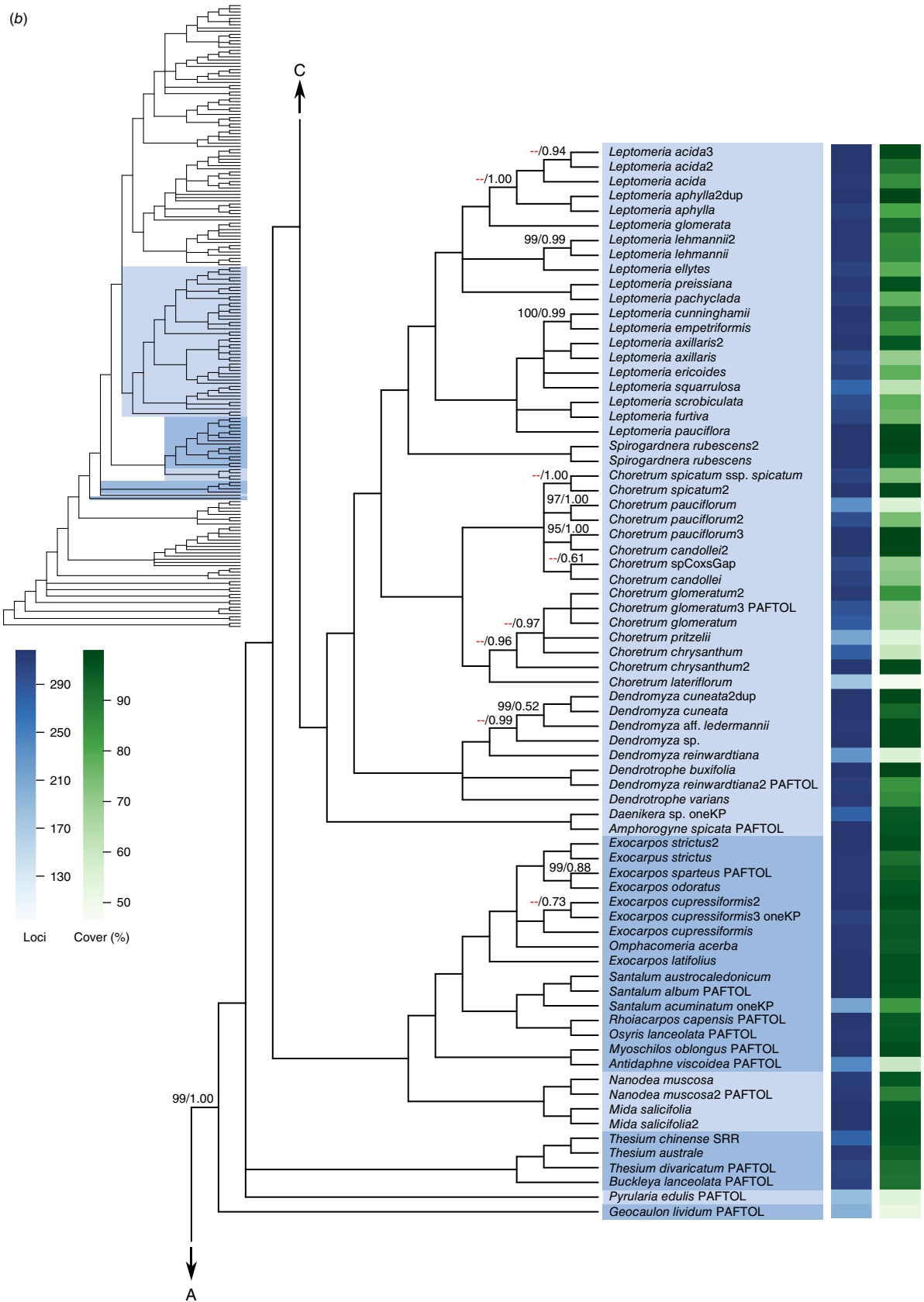


Fig. 1. (continued)

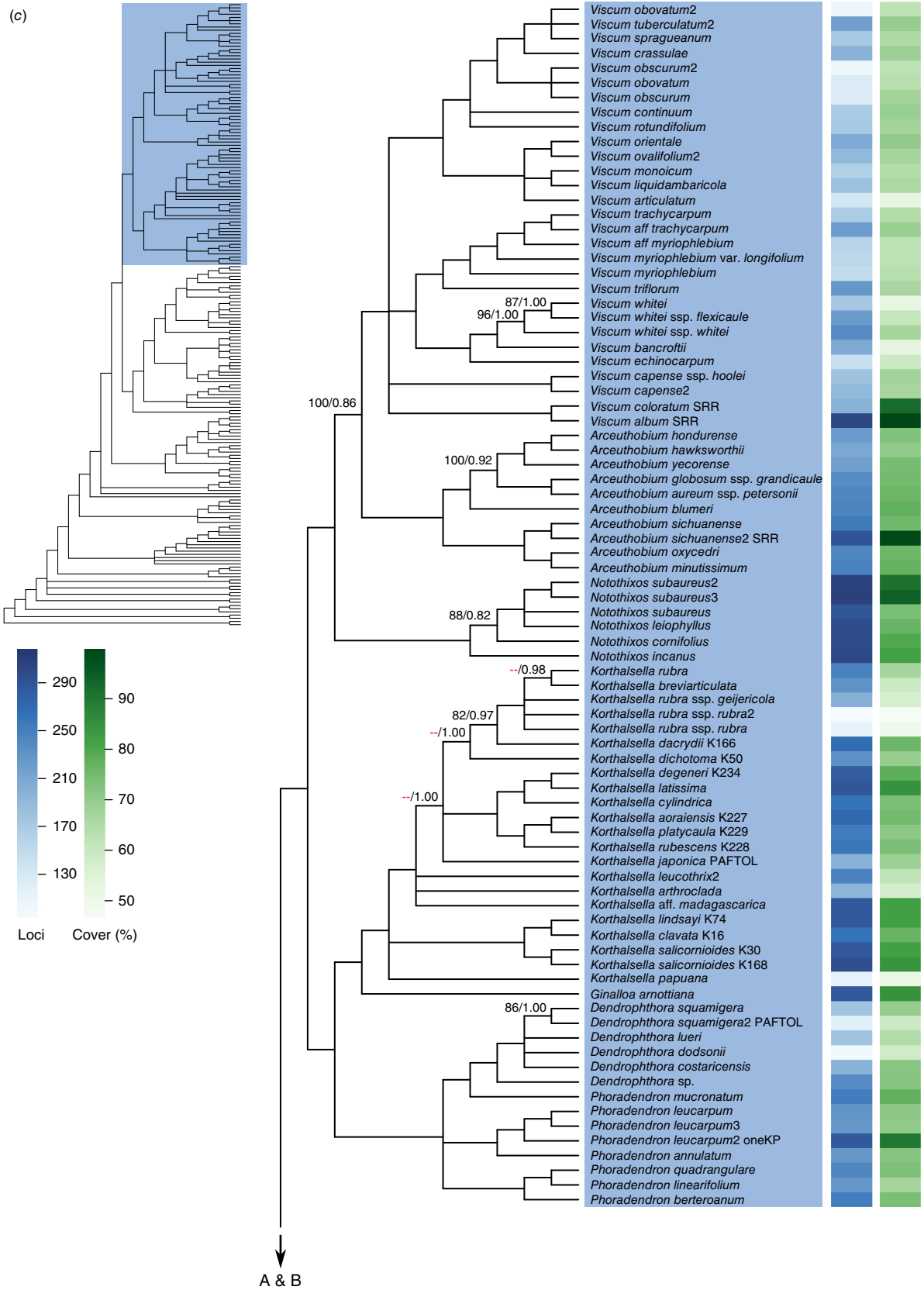


Fig. 1. (continued)

Fig. 1. (a–c) Cladogram showing relationships in Santalales inferred with ASTRAL using locus trees inferred with maximum likelihood for 318 nuclear loci. Branches where the hypothesis of a polytomy could not be rejected at the 5% significance level are collapsed. All branches have full support (100 UFB/1.00 pp, for the maximum likelihood concatenation and ASTRAL coalescent analyses respectively) except where shown. Support values designated with a red dashes (‘--’) indicate that the branch was not recovered in the concatenation analysis. Shading is by family and indicated on the overview tree for the respective portion of the tree (A, B or C). Taxon names have suffixes reflecting publicly available data sources: ‘PAFTOL’ for the Kew Plant and Fungal Trees of Life, ‘oneKP’ for the One Thousand Plant Transcriptomes Initiative and ‘SRR’ for GenBank short read data. In the upper overview tree, family names in bold represent APG IV classification, whereas non-bold family names represent an alternative classification (Nickrent 2020). The blue and green heatmaps at the right show the number of loci in the analysis and the average percentage alignment coverage per locus respectively.

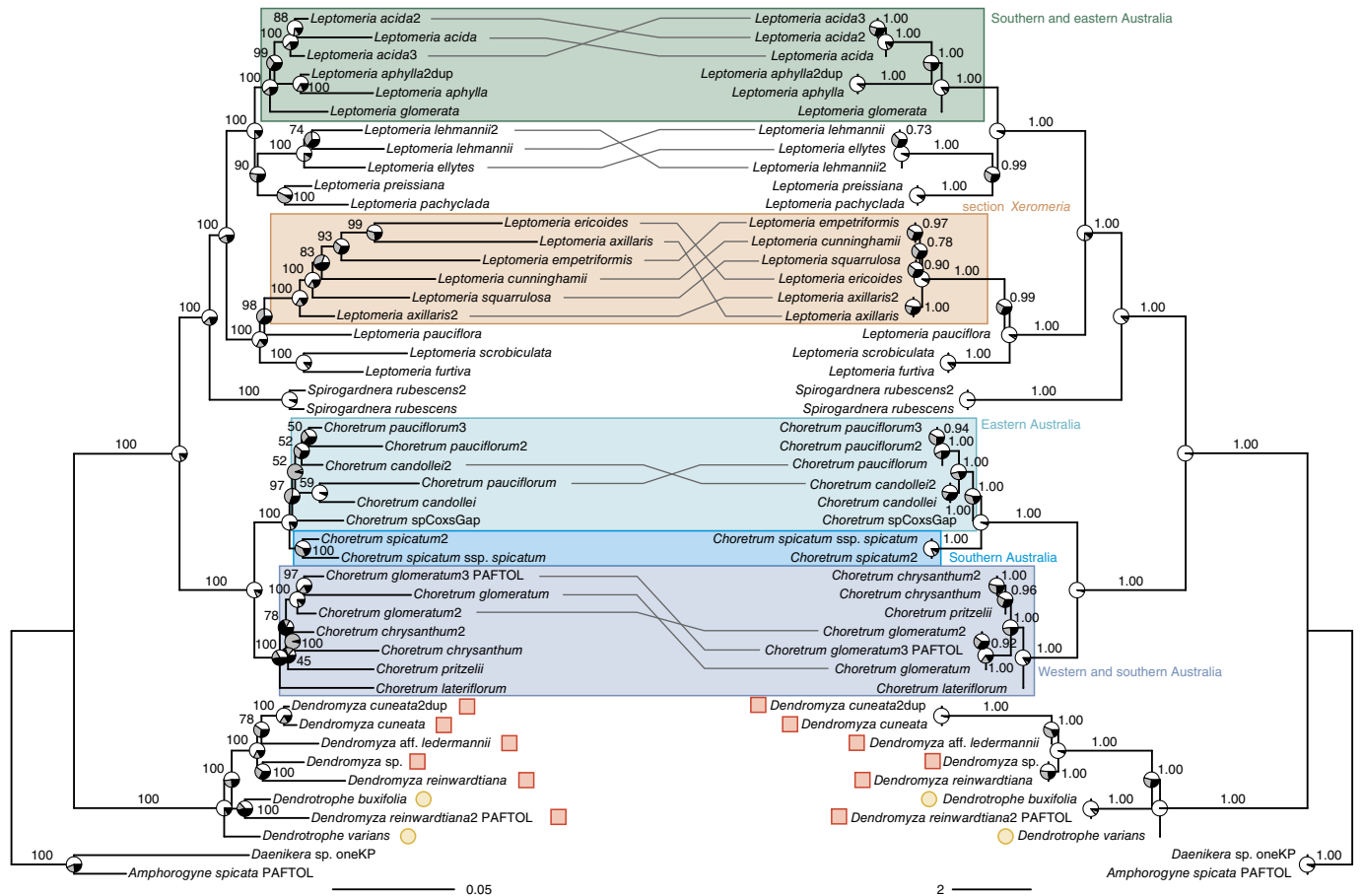


Fig. 2. Maximum likelihood concatenation (left) and ASTRAL coalescent (right) phylogenetic trees for Amphorogynaceae based on supercontigs (exons + introns) from 331 nuclear loci. For the concatenation tree, branch support values are Ultrafast Bootstrap percentages, and the scale bar shows inferred substitutions per site. Pie charts at nodes show site concordance factors for the main (white) and two alternative (grey and black) topologies. For the coalescent tree, branch support values are local posterior probabilities, and branch lengths and the scale bar are in coalescent units. Pie charts at nodes show quartet support for the main (white) and two alternative (grey and black) topologies. Grey lines connect taxa with alternative placements between the two trees. Taxon labelling is the same as in Fig. 1, with symbols for *Dendromyza* (red squares) and *Dendrotrophe* (yellow circles) emphasizing the nested relationship. Shaded and annotated boxes highlight clades mentioned in the text.

sequences producing long branches in locus trees, the second set of analyses resulted in a filtered alignment of 271 loci with a total aligned length of 273 000 bp and 21.4% missing data. The 37 samples had on average 218 loci (78–268, s.d. 51), with 24.4% missing data per locus (3.4–59.2%, s.d. 15.6%).

The ML concatenation and ASTRAL coalescent trees (Fig. 3) were largely congruent, although resolution along the backbone of *Korthalsella* was low in some places. There was limited conflict between concatenation and coalescent trees in *Korthalsella* and *Dendrophthora*. A sample of *Phoradendron mucronatum* (DC.) Krug & Urb. was recovered as sister to

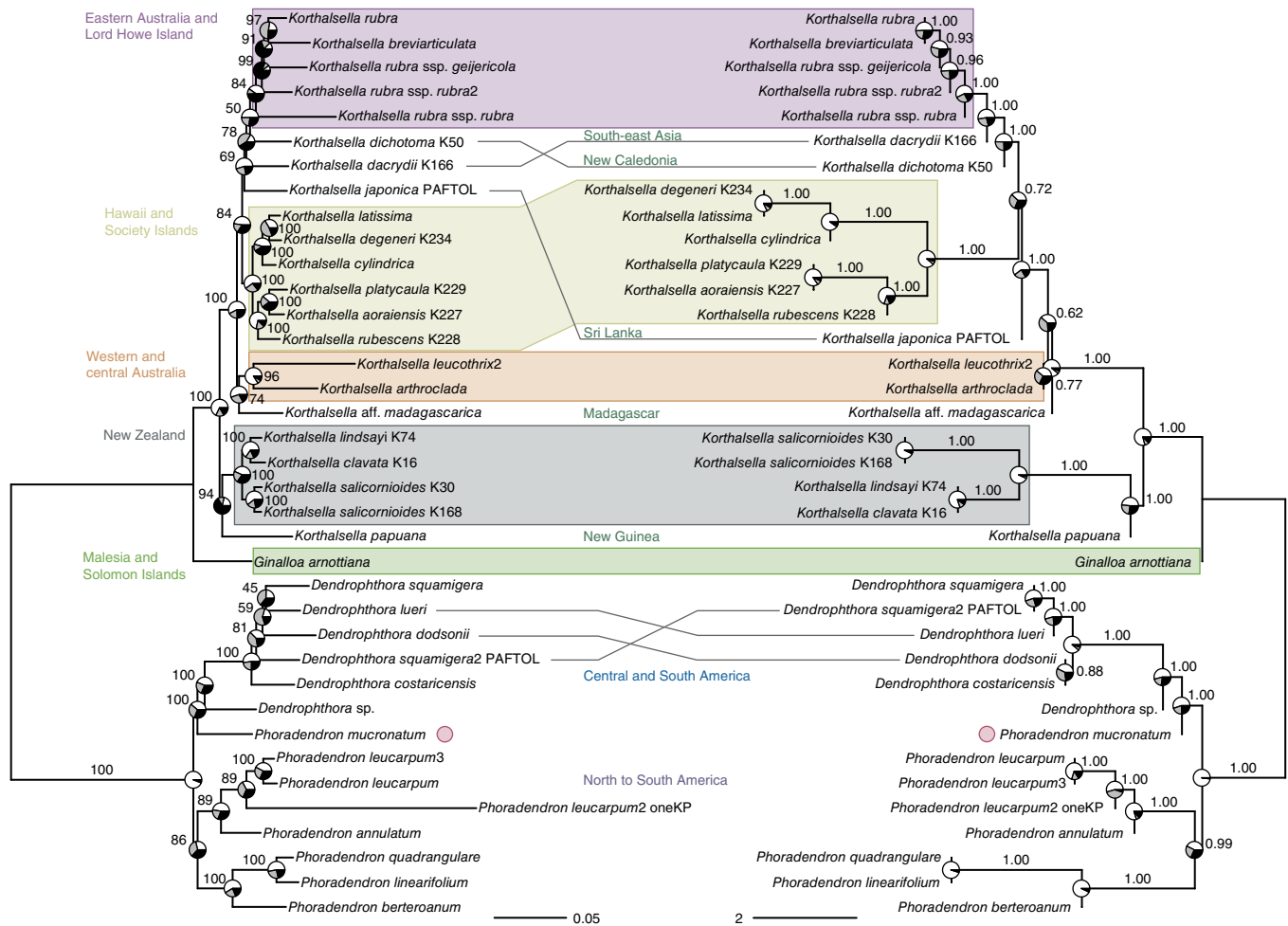


Fig. 3. Maximum likelihood concatenation (left) and ASTRAL coalescent (right) phylogenetic trees for *Korthalsella* and relatives based on supercontigs (exons + introns) from 271 nuclear loci. For the concatenation tree, branch support values are Ultrafast Bootstrap percentages, and the scale bar shows inferred substitutions per site. Pie charts at nodes show site concordance factors for the main (white) and two alternative (grey and black) topologies. For the coalescent tree, branch support values are local posterior probabilities, and branch lengths and the scale bar are in coalescent units. Pie charts at nodes show quartet support for the main (white) and two alternative (grey and black) topologies. Grey lines connect taxa with alternative placements between the two trees. Taxon labelling is the same as in Fig. 1, with a circle next to a sample sister to *Dendrophthora* that renders *Phoradendron* paraphyletic. Shading shows geographic groups mentioned in the text.

Dendrophthora, rendering our sampling of *Phoradendron* paraphyletic.

The re-run of the primary analysis using downloaded data resulted in slightly higher recovery of genes for some samples (not shown), presumably from retaining optical duplicate reads and the resulting higher read depth. The resulting phylogenetic trees (ML and ASTRAL) had no strongly supported conflicts with the primary analysis, differing only in a few poorly supported species relationships within genera.

Stochastic mapping of parasitism type (see Supplementary Fig. S2) suggested one origin of root parasitism in Santalales and five origins of aerial parasitism in the families Misodendraceae, Loranthaceae, Santalaceae sens. str., Amphorogynaceae and Viscaceae. The second mapping

suggested two origins of holoparasitism in Balanophoraceae sens. str. and *Hachettea*.

Discussion

Our analysis of phylogenetic relationships in the Santalales based on more than 300 nuclear loci provides a test of previous hypotheses for the order and new insights into its evolution, particularly within Santalaceae sens. lat. Making use of publicly available data that was part of a recent angiosperm-wide phylogenetic analysis (Zuntini *et al.* 2024), we identify concordance and conflict with existing classifications of the order and note remaining knowledge gaps. Our results contribute to a framework that will be

useful for understanding mistletoe and mitogenome evolution in Santalaceae sens. lat. and support ongoing taxonomic revision of Australian Amphorogynaceae.

Phylogenetic relationships among and classification of Santalales families

Family relationships outside of Santalaceae sens. lat. were not the focus of this study, but publicly available data with limited sampling allowed for some inferences. Sampling in this study comprised 61 genera (of ~183), with the primary lack of sampling in Loranthaceae (only 8 of ~78 genera). Findings from other studies using fewer loci but with denser taxon sampling (Su *et al.* 2015; Nickrent *et al.* 2019) were nonetheless largely reiterated here.

Relationships among Olacaceae sens. lat. have been difficult to resolve (Malécot and Nickrent 2008), and recent studies (Su *et al.* 2015; Nickrent *et al.* 2019) have recovered clades but with variable support for how they are related. Our sampling of publicly available data comprised 13 (of 29) genera from 6 (of 7) families (Octoknemaceae was not successful), and we recovered clades for the same families as proposed by Nickrent *et al.* (2010), with one exception: our downloaded sample of *Heisteria parvifolia* Sm. was recovered as sister to *Olax benthamiana* Miq. (Olacaceae) rather than with *Maburea* Maas and *Brachynema* Benth. (Erythralaceae) where it should group (Su *et al.* 2015; Nickrent *et al.* 2010, 2019; Edlund *et al.* 2024). This anomalous relationship was unsurprisingly also recovered in the full tree in Zuntini *et al.* (2024) that used the same samples. The identification of that sample should be reviewed and additional sampling of *Heisteria* Jacq. included in future studies of the group. For our stochastic mapping, we accounted for this potential misidentification by coding *Heisteria* as equally likely to be not parasitic (the expected condition) or root parasitic (expected for a relative of *Olax* L.). Although our sampling is limited, relationships among Olacaceae, Ximeniaceae and Aptandraceae were recovered differently in concatenation and coalescent trees (see Supplementary Fig. S1) and also differed in Zuntini *et al.* (2024), where Ximeniaceae was recovered as sister to the non-Olacaceae Santalales with 0.95 posterior probability (39% gene tree support). Our ASTRAL results could not reject a polytomy for that relationship (i.e. the branch lengths or coalescent times were too short for the analysis to conclusively resolve the relationship and it should be represented as a polytomy), whereas our concatenation results provided low support (79% UFB) that the three families instead formed a clade sister to the remaining non-Olacaceae Santalales, similarly to Nickrent *et al.* (2019) except for *Octoknema* Pierre (which had little data and a poor assembly in this study and was dropped). Given the lack of phylogenetic resolution, the results of our stochastic mapping that suggest a single origin of parasitism should be taken with care (also because of the uncertainty around the parasitic status of Aptandraceae), as the approach does not allow for polytomies or take into account topological uncertainty.

Despite the uncertainty in some of these relationships (also unresolved in Li *et al.* (2021)), there was no support for the monophyly of Olacaceae sens. lat., and we advocate the recognition of the clades as families.

As in Su *et al.* (2015), we recovered a monophyletic (97 UFB/0.96 pp) Balanophoraceae sens. strict. as sister (93 UFB/1.00 pp) to the remaining non-Olacaceae sens. lat. Santalales (Fig. 1, Supplementary Fig. S1), despite evidence of long branches. The other holoparasite, *Hachettea*, was the sole representative of Mystropetalaceae, which Su *et al.* (2015) recovered (with two other genera) as sister to Loranthaceae (80 ML bootstrap/0.97 Bayesian posterior probability), but that we recovered as part of a clade containing Loranthaceae, Schoepfiaceae (one of three genera sampled) and Misodendraceae. Although there was strong support that our samples from these four families belong to a single clade, there was considerable gene tree incongruence and potential conflict for relationships among them, with relatively short branches recovered in both trees (see Supplementary Fig. S1). Broader sampling of Schoepfiaceae (including *Arjona* Cav. and *Schoepfia*) and Mystropetalaceae (including *Dactylanthus* and *Mystropetalon*) may help clarify their relationships. In contrast to our results, Zuntini *et al.* (2024) recovered the same sample of *Hachettea* as sister to the Loranthaceae (0.84 pp/40% gene tree support), although their samples were similarly recovered in the same clade with samples of Schoepfiaceae and Misodendraceae. An additional sample of *Mystropetalon* in Zuntini *et al.* (2024) was placed without support between Balanophoraceae and the remaining non-Olacaceae Santalales, but sequence data for that sample had not been released and only 72 loci were reported as recovered, which may have made it difficult to confidently place the sample. We did not recover Mystropetalaceae in a clade with Balanophoraceae sens. strict., so we advocate the recognition of two separate families as first proposed by Su *et al.* (2015).

Our sampling of Loranthaceae was limited (Fig. 1a), but the recovered topology is consistent with previous work on the group (Liu *et al.* 2018; Nickrent *et al.* 2019), with the Western Australian root parasite *Nuytsia* R.Br. ex G.Don sister to the remainder of the family (mostly mistletoes). We also recovered root parasite *Atkinsonia* F.Muell. as successive sister to the remainder of our Loranthaceae samples, a result that had some support, along with *Gaiadendron* G.Don (not sampled here), in previous work (Liu *et al.* 2018; Nickrent *et al.* 2019). There has been limited contention about these relationships (Grímsson *et al.* 2017; Liu *et al.* 2018), and although Zuntini *et al.* (2024) also recovered the sister relationship of *Nuytsia* to the remaining Loranthaceae, they recovered *Atkinsonia* + *Gaiadendron* nested within tribe Elytrantheae + *Tupeia* Cham. & Schltdl., sister to tribe Psittacanthae. Denser sampling of Loranthaceae may help to resolve the placement of tribe Gaiadendreae.

We recovered Opiliaceae as sister (99 UFB/1.00 pp) to Santalaceae sens. lat., in agreement with previous work

(Su *et al.* 2015; Nickrent *et al.* 2019). With similar sampling, Zuntini *et al.* (2024) reported low support (0.26 pp/38% gene tree support) for this sister relationship due to the placement of *Geocaulon lividum* (Richardson) Fernald, and our gene concordance factors also indicate potential conflict for that branch in the concatenation tree (see Supplementary Fig. S1). Relationships within Opiliaceae were partly supported and largely congruent with previous work, with the exception of the sister relationship between our samples of *Urobotrya* Stapf and *Opilia* Roxb. Previous work (Su *et al.* 2015; Le *et al.* 2018; Nickrent *et al.* 2019) indicated a closer relationship between *Urobotrya* and *Cansjera* Juss. Given the long branches subtending our sister relationship between *Urobotrya* and *Opilia*, the identification of the *Opilia* sample may need to be revisited. There was also some moderately supported conflict between our concatenation and coalescent trees as to whether *Agonandra* Miers ex Benth. was sister to *Anthobolus* R.Br. (coalescent) or sister to the remaining Opiliaceae (concatenation). Our ASTRAL analysis could not reject a polytomy for this relationship, which has been unresolved in previous work. The ASTRAL analysis of Zuntini *et al.* (2024) also recovered a sister relationship between *Agonandra* and *Anthobolus* with similar support to our ASTRAL analysis. The placement of *Anthobolus* in Opiliaceae here rather than in Santalaceae traditionally (e.g. Kuijt 2015) was first suggested by Der and Nickrent (2008) and supported by subsequent studies (Su *et al.* 2015; Le *et al.* 2018; Nickrent *et al.* 2019).

Branching order among families in Santalaceae sens. lat. has been challenging to fully resolve (Der and Nickrent 2008; Su *et al.* 2015; Nickrent *et al.* 2019), and our results suggest some improvements while also showing uncertainty. Although previous work (Su *et al.* 2015, plastid data; Nickrent *et al.* 2019) has hinted at a sister relationship between Comandraceae (*Geocaulon* Fernald here; *Comandra* Nutt. not sampled) and Thesiaceae, we recover our *Geocaulon* sample as sister to the remaining Santalaceae sens. lat., although there are indications of conflict (mentioned above). Analysis of nuclear and mitogenome loci in Su *et al.* (2015) suggested the same relationship with low support. We recover a sister relationship between Cervantesiaceae (*Pyrularia* Michx. here; seven genera unsampled) and Thesiaceae with low support (97 UFB/0.66 pp), which had also been hinted at in previous work (Der and Nickrent 2008; Su *et al.* 2015), and our ASTRAL analysis could not reject a polytomy for this relationship. Zuntini *et al.* (2024) instead recovered *Pyrularia* sister to the remaining Santalaceae sens. lat. (excluding *Geocaulon*) with low support (0.49 pp/38% gene tree support) and Thesiaceae successively sister to the remainder. The main strongly supported difference between our results and previous work including Zuntini *et al.* (2024) is a sister relationship between Nanodeaceae and Santalaceae (100 UFB/1.00 pp), which are sister to Amphorogynaceae + Viscaceae. Previous work has suggested (Santalaceae (Nanodeaceae (Amphorogynaceae + Viscaceae))) (Nickrent *et al.* 2019) or (Nanodeaceae

(Santalaceae (Amphorogynaceae + Viscaceae))) (Der and Nickrent 2008; Zuntini *et al.* 2024). Zuntini *et al.* (2024) recovered a sister relationship between Santalaceae and Amphorogynaceae + Viscaceae with strong support (0.99 pp/42% gene tree support). Our sampling of Nanodeaceae is slightly larger (two genera v. one genus) but our concordance factor is lower for the sister relationship between Nanodeaceae and Santalaceae and suggests some conflict (see Supplementary Fig. S1). Despite this, our support for the relationship is still strong and has high quartet support in the ASTRAL analysis.

Our results suggest some differences in relationships within Santalaceae sens. strict., with *Myoschilos* Ruiz & Pav. + *Antidaphne* Poepp. & Endl. sister to the remainder of the family, a relationship also found for the same samples in Zuntini *et al.* (2024). Previous work (Der and Nickrent 2008; Su *et al.* 2015; Nickrent *et al.* 2019) instead recovered *Exocarpos* Labill. + *Omphacomeria* A.DC. as sister to the remainder of the family. The former Eremolepidaceae mistletoes (*Antidaphne* here; *Eubrachion* Hook.f. and *Lepidoceras* Hook.f. not sampled) are on relatively longer branches in Santalaceae sens. strict., possibly accounting for some of the variability in their placement. Our concatenation tree (Supplementary Fig. S1) suggests some conflict in the backbone of Santalaceae sens. strict. but generally good support for the recovered relationships. Our sampling of *Exocarpos* also indicates that *Omphacomeria* is nested within it, in agreement with recent work that chose to combine *Omphacomeria* with *Exocarpos* (Pillon *et al.* 2023). Pillon *et al.* (2023) recovered *Exocarpos* sister to the remaining Santalaceae sens. strict. samples with limited support (0.93 pp), in agreement with previous work. Although there appears to be limited support for the *Exocarpos*-sister hypothesis in previous work, our much larger dataset suggests that *Exocarpos* is more closely related to *Santalum* L. and its relatives than to *Antidaphne*. Denser sampling of the Santalaceae sens. strict. mistletoes may help to clarify these relationships.

The remaining families in Santalaceae sens. lat., Amphorogynaceae and Viscaceae, have long been supported as monophyletic (Der and Nickrent 2008; Su *et al.* 2015; Nickrent *et al.* 2019) and continue to be so here and in Zuntini *et al.* (2024). Consistent with previous work, we recovered a clade of the two New Caledonian taxa *Amphorogyne* Stauffer & Hürl. + *Daenikera* Hürl. & Stauffer sister to the remaining Amphorogynaceae and a clade of the south-east Asian mistletoes (*Dendromyza* + *Dendrotrophe*) sister to the Australian root parasites (*Chorettrum* and *Leptomera* + *Spirogardnera* Stauffer). Within Viscaceae, there has been uncertainty in the relationships among clades comprising *Notothixos*, *Arceuthobium* M.Bieb., *Viscum*, *Korthalsella* + *Ginalloa* and *Phoradendron* + *Dendrophthora* (Su *et al.* 2015; Nickrent *et al.* 2019). Our results recovered a sister relationship between *Arceuthobium* and *Viscum* (100 UFB/0.86 pp) with indications of conflict, but this relationship had previously received some support (Nickrent *et al.* 2019) and was also recovered in Zuntini *et al.* (2024) (1.00 pp/48% gene

tree support). We recovered *Notothixos* as sister to *Arceuthobium* + *Viscum*, again with some indications of conflict and short branches (Supplementary Fig. S1), and this relationship was also recovered in Zuntini *et al.* (2024). Our results strongly supported (100 UFB/1.00 pp and high gene concordance) a sister relationship between *Korthalsella* + *Ginallia* and *Phoradendron* + *Dendrophthora*, as has been recovered before (Su *et al.* 2015; Nickrent *et al.* 2019).

Focused sampling in Amphorogynaceae and Viscaceae

We undertook denser sampling of predominantly Australasian Amphorogynaceae and Viscaceae, particularly within the genera *Choretrum*, *Leptomeria* and *Korthalsella*, as one of the motivating reasons for this study was to explore diversity and relationships in these groups as part of the GAP initiative. We created new alignments using both coding and non-coding portions of the targets for taxa in these densely sampled groups to increase the number of informative sites while avoiding alignment ambiguity by using closely related species. The larger alignments led to the same overall relationships between genera observed in the main tree and some improvements in resolution within them, although substantial uncertainty and conflicting phylogenetic signal remained (Fig. 2 and 3).

Within Amphorogynaceae (Fig. 2), we recovered a clade containing intermixed samples of *Dendromyza* (4 of ~30 species sampled) and *Dendrotrophe* (2 of ~5 species sampled). The lack of reciprocal monophyly between these two genera was also found in a study using ITS (Devkota *et al.* 2015). Further taxonomic work on the two genera appears to be needed and Kuijt (2015) considered there to be no clear morphological separation between the two, provisionally treating them as the same taxon. Additional sampling of both may help to clarify whether monophyletic groups can be delimited or if they should be synonymised as our results suggest. Complicating this, the two samples of *Dendromyza reinwardtiana* (Blume) Danser did not group together, suggesting that the identifications may need to be reviewed, although they also had substantially different gene recovery with *HybPiper* (241 v. 85 genes at 50% target length; see Supplementary Table S3).

Both *Leptomeria* and *Choretrum* were strongly supported as monophyletic, suggesting that generic boundaries between these Australian root parasites are good, and *Spirogardnera* is divergent and sister to *Leptomeria*, not nested in it. Current infrageneric classification in *Leptomeria* recognises two sections: section *Leptomeria* (12 species) with plants having scale-like leaves that fall off, and section *Xeromeria* (Endl.) Miq. (5 species) with plants having well-developed and typically persistent leaves (Lepschi 1999). We recovered section *Xeromeria* as monophyletic with full support (100 UFB/1.00 pp), but it was nested in a clade with three species of section *Leptomeria*, rendering that section paraphyletic. Most species

of *Leptomeria* are restricted to the south-west of Western Australia, and the close relatives of section *Xeromeria* are also geographically in the same region. The three species of *Leptomeria* outside of Western Australia were recovered in a strongly supported clade (see Fig. 2) sister to another group of south-west species, which may point to a single dispersal out of the centre of diversity for the genus, but this pattern should be interpreted with care given potential confounding effects from differential extinction between areas (Sanmartín and Meseguer 2016). The sister genus to *Leptomeria* is *Spirogardnera*, which is restricted to Western Australia, consistent with the inference that ancestors of the two occurred in Western Australia. By contrast, *Choretrum* showed a more nuanced biogeographic pattern. We recovered two well supported clades consistent with morphological characters. One clade comprised species from eastern Australia, sister to *C. spicatum*, which extends westwards into South Australia. Species in this clade have an adaxial tuft of hairs on their petals, which are longitudinally striate when dry, and obscurely stellate stigmas. The other clade included species restricted to Western Australia (*Choretrum pritzelii* Diels and *C. lateriflorum* R.Br.), and two species distributed in Western Australia and extending eastwards into the southern regions of eastern Australia (*C. glomeratum* R.Br. and *C. chrysanthum* F.Muell.). Species in this clade have glabrous petals, which are wrinkled when dry, and distinctly stellate stigmas. Although the geographic pattern is not inconsistent with a Western Australian origin for all Australian Amphorogynaceae root parasites, that hypothesis would require multiple eastward dispersals in *Choretrum*. Further studies with denser sampling and biogeographic models are needed to confirm an area of origin for the group.

Within our tree of Viscaceae mistletoes allied to *Korthalsella* (Fig. 3), samples of *Dendrophthora* (5 of >100 species) formed a clade, but a sample of *Phoradendron mucronatum* (circle in Fig. 3) was recovered sister to them (100 UFB/1.00 pp), rendering our sampling of *Phoradendron* (6 of >250 species) paraphyletic. Branches along the backbone of the clade were short, with indications of conflicting signal, and the remaining *Phoradendron* samples formed a clade with low support (86 UFB/0.99 pp). The two American genera are known to be challenging to separate (Kuijt 2015) and have overlapping distributions, so further sampling of the substantial diversity is needed to assess generic boundaries.

Within *Korthalsella* (Fig. 3), we recovered multiple clades reflecting geographic distribution, which was in general agreement with overall findings from a recent study with broader taxon sampling (Sultan *et al.* 2019). We sampled 18 species (of ~30) compared to Sultan *et al.* (2019) who sampled 22 plus subspecific taxa and multiple accessions for some widespread species (e.g. *Korthalsella japonica* (Thunb.) Engl.). In agreement with Sultan *et al.* (2019), we recovered a strongly supported clade of the New Zealand species, with moderate support (94 UFB/1.00 pp) for a sister relationship with *K. papuana* Danser from

New Guinea, which Sultan *et al.* (2019) instead recovered as sister (with *K. geminata* (Korth.) Engl.) to the remainder of the genus (with the New Zealand clade successively sister). The branching order at the backbone had limited support in our tree, but our results are consistent with these two groups being largely distinct from the remainder of the diversity in the genus and underscore that New Zealand *Korthalsella* are distantly related to geographically closer New Caledonian and eastern Australian species. Morphologically, there are similarities in inflorescence structure between *K. papuana* and two New Zealand species, *K. lindsayi* (Oliv. ex Hook.f.) Engl. and *K. clavata* Cheeseman, concordant with a close relationship between these taxa, but apparent differences with the other New Zealand species, *K. salicornioides* (A.Cunn.) Tiegh., which has been thought to have an ‘undifferentiated’ inflorescence. Molvray *et al.* (1999) noted that this was a misinterpretation based on the differing vegetative branch shape in *K. salicornioides* relative to *K. lindsayi*, when all these closely related species actually share ‘differentiated’ inflorescences (only on distal nodes *v.* ‘undifferentiated’ inflorescences on every node as in most of the rest of the genus). The sister genus *Ginalloa* also has a more differentiated inflorescence, supporting the ‘differentiated’ condition as symplesiomorphic, and *Ginalloa* occurs in Malesia and the Solomon Islands, consistent with a Malesian or Australasian origin for *Ginalloa* + *Korthalsella* (Molvray *et al.* 1999). Our results agreed with this, with the New Guinea *K. papuana* + New Zealand species sister to the remainder of the genus.

The backbone of the remainder of *Korthalsella* had relatively low support and short branch lengths, similar to the tree in Sultan *et al.* (2019), and both sets of trees support some distinct clades arising from the uncertain branching order. In contrast to Sultan *et al.* (2019), we recovered some support (96 UFB/0.77 pp) for a sister relationship between the western and central Australian species (*K. arthroclada* Cranfield and *K. leucothrix* Barlow) as a separate lineage arising from the radiation around the Indian Ocean (74 UFB as sister to the sample from Madagascar in the concatenation tree; part of the backbone polytomy in the *ASTRAL* tree). The remainder of our sampling included a single sample of the widespread *K. japonica*, and samples from South-east Asia and across the Pacific. There is some support (84 UFB/1.00 pp) for this geographically widespread clade, in which east Australian *Korthalsella* samples were nested (exclusive of the western and central Australian species), and that also includes a strongly supported (100 UFB/1.00 pp) clade comprising two lineages for species in Hawaii and the south Pacific, as found in Sultan *et al.* (2019). Our results separating western and central Australian species with terete internodes from eastern Australian species with flattened internodes, as also found by Sultan *et al.* (2019), suggests that the two groups are not closely related and may help to explain differences in where they occur (more arid *v.* temperate or tropical).

For the remainder of Viscaceae (Fig. 1c), our sampling of *Arceuthobium* (nine of ~26 species) and *Viscum* (21 of ~100

species) was limited but allowed some comparisons with previous work. We recovered two divergent clades within *Arceuthobium* corresponding to the two subgenera: New World *Vaginata* Hawksw. & Wiens and Old World *Arceuthobium* (Nickrent *et al.* 2004). Species that were found to be closely related and considered to be potentially synonymous in Nickrent *et al.* (2004) formed highly similar sister pairs in our results (e.g. *A. hondurensis* Hawksw. & Wiens and *A. hawksworthii* Wiens & Shaw). Within *Viscum*, our results showed a lack of resolution and conflict along the backbone, with uncertainty in the branching order of four well-supported clades (Fig. 1, Supplementary Fig. S1). These same four clades were represented in a recent study by Maul *et al.* (2019), although with broader taxon sampling (52 species *v.* 21 here; 16 species shared). The first clade is represented here by *Viscum coloratum* (Kom.) Nakai and *V. album*, which Maul *et al.* (2019) recovered (the corresponding clades A and B) as sister to the remainder of the genus. We had two samples of *V. capense* L.f. representing a second clade that Maul *et al.* (2019) found with *V. schaeferi* Engl. & K.Krause (clade F). The third clade includes mostly species from Madagascar along with two of the Australian species, *V. whitei* Blakely and *V. bancroftii* Blakely, and corresponds to clades C, D and E in Maul *et al.* (2019). The fourth clade contains African and Asian species, including the other two species with Australian occurrences, *V. ovalifolium* Wall. ex DC. and *V. articulatum* Burm.f., which corresponds to clades H, I and J in Maul *et al.* (2019). In our results as in Maul *et al.* (2019), *V. ovalifolium* and *V. articulatum* are more closely related to Asian species than to the other Australian species *V. whitei* and *V. bancroftii*, suggesting multiple arrivals of *Viscum* in Australia.

Prospects for studying mistletoe evolution

Our phylogenetic results reinforce previous work on the origins of aerial parasitism in Santalales (mistletoes). There are five families with species that parasitise stems: Misodendraceae (all species), Loranthaceae (most species), Santalaceae sens. strict. (some species), Amphorogynaceae (some species) and Viscaceae (all species). Given these five non-monophyletic instances of aerial parasitism, Vidal-Russell and Nickrent (2008) hypothesised five origins of aerial parasitism. Our results are consistent with this hypothesis but may also hint at more origins. Misodendraceae (1 species) is not recovered sister to other mistletoes, supporting one origin. We recovered root parasites in Loranthaceae successively sister to the mistletoes, supporting a single origin in that family. We recovered (with some conflict) a closer relationship between *Exocarpos* (root and debatedly stem parasites) and *Santalum* (root parasites) than with *Antidaphne* (mistletoes), suggesting two potential origins of aerial parasitism in that family. Although most *Exocarpos* species are root parasites, three species, *E. aphyllus* R.Br., *E. cupressiformis* Labill. and *E. pullei* Pilg., have been reported to

parasitise stems (Coleman 1934; Lam 1945; Kuijt 1969; Baird 2014), although Kuijt (2015) expressed doubt and the need for more observational work. Although we coded *E. cupressiformis* as uncertain in our stochastic mapping (Supplementary Fig. S2), there was little indication for more than one origin in Santalaceae. Within Amphorogynaceae, there are indications of varying degrees of stem and root parasitism in *Daenikera*, *Dendrotrophe* and *Dendromyza* (and the unsampled *Dufrenoya* Chatin and *Phacellaria* Benth.) that may suggest at least two origins in the family. Given our sampling, the stochastic mapping nevertheless only supported one origin there. Alternatively, and perhaps unlikely given the specialised nature of mistletoes, aerial parasitism may have been ancestral in Amphorogynaceae and lost at least twice, once in *Amphorogyne* and once in the ancestor of the Australian root parasites (*Choretrum*, *Leptomeria* and *Spirogardnera*), although our lack of sampling of *Dufrenoya* and *Phacellaria* limit what can be postulated. Another alternative is that ancestors of the group may have been able to parasitise both roots and stems (Vidal-Russell and Nickrent 2008), although whether this condition exists has been questioned with regard to seedling establishment (Kuijt 2015). Given that all of Viscaceae are mistletoes, a common aerial condition or ability to parasitise both roots and stems in the ancestor of Amphorogynaceae and Viscaceae may be more parsimonious, but this is not supported with our current sampling and stochastic mapping. As stressed by Kuijt (2015), more observations and study of modes of parasitism, seedling establishment and occurrences of both root and stem parasitism in the same genus (e.g. *Daenikera* and *Dendrotrophe*) are needed.

Mistletoes in Viscaceae have become a focus for mitogenome research, in part due to the discovery that *Viscum album* has lost mitochondrial respiratory complex I (Maclean *et al.* 2018; Senkler *et al.* 2018), something that had not been documented in multicellular eukaryotes. Other work showing apparent loss of key mitochondrial genes in other species of *Viscum* (Petersen *et al.* 2015a; Skippington *et al.* 2015) and *Phoradendron* (Zervas *et al.* 2019) suggests that these fundamental mitogenomic changes (Petersen *et al.* 2020, 2022) may have occurred more deeply in the Viscaceae phylogeny. Our results resolved *Phoradendron* in a clade sister to *Korthalsella* + *Ginalloa*, separate from the clade containing *Viscum*. If that topology is correct, this would suggest that shared mitochondrial changes occurred in the ancestors of the entire family. Ongoing work in Amphorogynaceae (unpublished) suggests that similar mitochondrial changes have occurred in that family that would indicate that these changes may have begun in the ancestors of both families. Further sequencing of mitogenomes in these parasites should help to clarify patterns of mitogenome evolution in these two closely related families.

Future phylogenetic studies of mistletoes in Santalaceae sens. lat. may be well served using target capture approaches but may be improved by using a more targeted bait set. Our results largely resolved phylogenetic relationships among genera and were consistent with previous work, indicating that

A353 was successful, but gene recovery was variable and could have affected resolution in portions of the tree. We noticed a phylogenetic bias in recovery, with holoparasites and mistletoes in Viscaceae performing particularly poorly (Fig. 1, Table 1). Given that species of Viscaceae showed good recovery with data that was not from target capture (thus not a bioinformatic problem), the problem likely relates to low DNA quality, sequencing depth or poor specificity in the baits given sequence divergence. As *Viscum* has one of the largest genomes among flowering plants (Zonneveld 2010; Novák *et al.* 2020), the low percentage of reads on target may mean that greater sequencing effort is needed to effectively recover targets. Future studies might also explore custom bait sets designed for Viscaceae or the divergent holoparasites to improve recovery and completeness of alignments for phylogenetic analyses. Although we have presented a phylogeny with denser sampling within Santalaceae sens. lat., there are still outstanding questions about family relationships in Santalales (e.g. branching order in Olacaceae sens. lat.) that would benefit from more complete sampling at the generic level and may be improved with more specific bait sets and deeper sequencing.

Supplementary material

Supplementary material is available online.

References

- Anderson B (2025) bmchanderson/Santalaceae: Publication. *Zenodo* 2025, v1.1.0. [Dataset, published 23 May 2025] doi:10.5281/zenodo.15496487
- Baird IRC (2014) A novel observation of putative aerial hemiparasitism in *Exocarpus aphyllus* (Santalaceae). *Queensland Naturalist* 52, 48–52.
- Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, Botigué LR, Brewer G, Carruthers T, Clarkson JJ, Cook J, Cowan RS, Dodsworth S, Epitawalage N, Françoise E, Gallego B, Johnson MG, Kim JT, Leempoel K, Maurin O, McGinnie C, Pokorny L, Roy S, Stone M, Toledo E, Wickett NJ, Zuntini AR, Eiserhardt WL, Kersey PJ, Leitch IJ, Forest F (2021) A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71, 301–319. doi:10.1093/sysbio/syab035
- Brewer GE, Clarkson JJ, Maurin O, Zuntini AR, Barber V, Bellot S, Biggs N, Cowan RS, Davies NMJ, Dodsworth S, Edwards SL, Eiserhardt WL, Epitawalage N, Frisby S, Grall A, Kersey PJ, Pokorny L, Leitch IJ, Forest F, Baker WJ (2019) Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* 10, 1102. doi:10.3389/fpls.2019.01102
- Carpenter EJ, Matasi N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA, Li L, Du W, Ullrich K, Wickett NJ, Barkmann TJ, Barker MS, Leebens-Mack JH, Wong GK-S (2019) Access to RNA-sequencing data from 1,173 plant species: the 1000 Plant transcriptomes initiative (1KP). *Gigascience* 8, giz126. doi:10.1093/gigascience/giz126
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q-Y, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. *Annals of the Missouri Botanical Garden* 80, 528–580. doi:10.2307/2399846

- Chen X, Fang D, Wu C, Liu B, Liu Y, Sahu SK, Song B, Yang S, Yang T, Wei J, Wang X, Zhang W, Xu Q, Wang H, Yuan L, Liao X, Chen L, Chen Z, Yuan F, Chang Y, Lu L, Yang H, Wang J, Xu X, Liu X, Wicke S, Liu H (2020) Comparative plastome analysis of root- and stem-feeding parasites of Santalales untangle the footprints of feeding mode and lifestyle transitions. *Genome Biology and Evolution* 12, 3663–3676. doi:10.1093/gbe/evz271
- Coleman E (1934) Notes on *Exocarpos*. *Victorian Naturalist* 51, 132–139.
- Der JP, Nickrent DL (2008) A molecular phylogeny of Santalaceae (Santalales). *Systematic Botany* 33, 107–116. doi:10.1600/036364408783887438
- Devkota MP, Macklin J, Nickrent DL (2015) The status of the mistletoe genus *Dufrenoya* Chatin (Amphorogynaceae) with a specific focus on Nepal. *Flora* 215, 75–83. doi:10.1016/j.flora.2015.07.009
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11–15.
- Edlund M, Anderson BM, Su H-J, Robison T, Caraballo-Ortiz MA, Der JP, Nickrent DL, Petersen G (2024) Plastome evolution in Santalales involves relaxed selection prior to loss of *ndh* genes and major boundary shifts of the inverted repeat. *Annals of Botany* 135, 515–530. doi:10.1093/aob/mcae145
- Engler A, Krause K (1935) Lorantheaceae. In 'Die Natürlichen Pflanzenfamilien', 2nd edn, vol. 16b. (Eds A Engler, K Prantl) pp. 98–203. (Engelmann: Leipzig, German Republic) [In German]
- García MA, Mucina L, Nickrent DL (2024) A tough nutlet to crack: resolving the phylogeny of *Thesium* (Thesiaceae), the largest genus in Santalales. *Taxon* 73, 190–236. doi:10.1002/tax.13123
- Govaerts R, Alrich P, Andersson L, Andrella GC, Andrews S, et al. (2025) The World Checklist of Vascular Plants (WCVP). (Royal Botanic Gardens: Kew, UK) doi:10.15468/6h8ucr
- Grimsson F, Grimm GW, Zetter R (2017) Evolution of pollen morphology in Lorantheaceae. *Grana* 57, 16–116. doi:10.1080/00173134.2016.1261939
- Harms H (1935) Balanophoraceae. In 'Die Natürlichen Pflanzenfamilien', 2nd edn, vol. 16b. (Eds A Engler, K Prantl) pp. 296–339. (Engelmann: Leipzig, German Republic) [In German]
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) *UFBoot2*: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35, 518–522. doi:10.1093/molbev/msx281
- Huelsenbeck JP, Nielsen R, Bollback JP (2003) Stochastic mapping of morphological characters. *Systematic Biology* 52, 131–158. doi:10.1080/10635150390192780
- Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJ, Wickert NJ (2016) *HybPiper*: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4, 1600016. doi:10.3732/apps.1600016
- Johnson MG, Pokorny L, Dodsworth S, Botigué LR, Cowan RS, Devault A, Eisehardt WL, Epiatawalage N, Forest F, Kim JT, Leebens-Mack JH, Leitch IJ, Maurin O, Soltis DE, Soltis PS, Wong GK, Baker WJ, Wickert NJ (2019) A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68, 594–606. doi:10.1093/sysbio/syy086
- Junier T, Zdobnov EM (2010) The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26, 1669–1670. doi:10.1093/bioinformatics/btq243
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) *ModelFinder*: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14, 587–589. doi:10.1038/nmeth.4285
- Katoh K, Standley DM (2013) *MAFFT* multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30, 772–780. doi:10.1093/molbev/mst010
- Kuijt J (1969) 'The Biology of Parasitic Flowering Plants.' (University of California Press: Berkeley, CA, USA)
- Kuijt J (2015) Santalales. In 'The Families and Genera of Vascular Plants. Flowering Plants. Eudicots: Santalales, Balanophorales'. (Ed K Kubitzki) pp. 1–189. (Springer International Publishing: Cham, Switzerland) doi:10.1007/978-3-319-09296-6
- Lam HJ (1945) Fragmenta papuana (observations of a naturalist in Netherlands New Guinea). *Sargentina* 5, 1–196. doi:10.5962/p.315952
- Lanfear R, Hahn MW (2024) The meaning and measure of concordance factors in phylogenomics. *Molecular Biology and Evolution* 41, msae214. doi:10.1093/molbev/msae214
- Le C, Liu B, Barrett RL, Lu L, Wen J, Chen Z (2018) Phylogeny and a new tribal classification of Opiliaceae (Santalales) based on molecular and morphological evidence. *Journal of Systematics and Evolution* 56, 56–66. doi:10.1111/jse.12295
- Lepschi BJ (1999) Taxonomic revision of *Leptomeria* (Santalaceae). *Australian Systematic Botany* 12, 55–100. doi:10.1071/SB97004
- Li H-T, Luo Y, Gan L, Ma P-F, Gao L-M, Yang J-B, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, Jin J-J, Zeng C-X, Wang H, Yu W-B, Zhang R, Van Der Bank M, Olmstead RG, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Yi T-S, Li D-Z (2021) Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biology* 19, 232. doi:10.1186/s12915-021-01166-2
- Liu B, Le CT, Barrett RL, Nickrent DL, Chen Z, Lu L, Vidal-Russell R (2018) Historical biogeography of Lorantheaceae (Santalales): diversification agrees with emergence of tropical forests and radiation of songbirds. *Molecular Phylogenetics and Evolution* 124, 199–212. doi:10.1016/j.ympev.2018.03.010
- Maclean AE, Hertle AP, Ligas J, Bock R, Balk J, Meyer EH (2018) Absence of complex I is associated with diminished respiratory chain function in European mistletoe. *Current Biology* 28, 1614–1619. doi:10.1016/j.cub.2018.03.036
- Maddison WP (1997) Gene trees in species trees. *Systematic Biology* 46, 523–536. doi:10.1093/sysbio/46.3.523
- Mai U, Mirarab S (2018) *TreeShrink*: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* 19, 272. doi:10.1186/s12864-018-4620-2
- Malécot V, Nickrent DL (2008) Molecular phylogenetic relationships of Olacaceae and related Santalales. *Systematic Botany* 33, 97–106. doi:10.1600/036364408783887384
- Maul K, Krug M, Nickrent DL, Müller KF, Quandt D, Wicke S (2019) Morphology, geographic distribution, and host preferences are poor predictors of phylogenetic relatedness in the mistletoe genus *Viscum* L. *Molecular Phylogenetics and Evolution* 131, 106–115. doi:10.1016/j.ympev.2018.10.041
- McLay TGB, Birch JL, Gunn BF, Ning W, Tate JA, Nauheimer L, Joyce EM, Simpson L, Schmidt-Lebuhn AN, Baker WJ, Forest F, Jackson CJ (2021) New targets acquired: improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* 9, e11420. doi:10.1002/aps.3.11420
- Mildbraed J (1935) Octoknemaceae. In 'Die Natürlichen Pflanzenfamilien', 2nd edn, vol. 16b. (Eds A Engler, K Prantl) pp. 42–45. (Engelmann: Leipzig, German Republic) [In German]
- Minh BQ, Hahn MW, Lanfear R (2020a) New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 37, 2727–2733. doi:10.1093/molbev/msaa106
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020b) *IQ-TREE 2*: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37, 1530–1534. doi:10.1093/molbev/msaa015
- Mo YK, Lanfear R, Hahn MW, Minh BQ (2023) Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics* 39, btac741. doi:10.1093/bioinformatics/btac741
- Molvray M, Kores PJ, Chase MW (1999) Phylogenetic relationships within *Korthalsella* (Viscaceae) based on nuclear ITS and plastid *trnL-F* sequence data. *American Journal of Botany* 86, 249–260. doi:10.2307/2656940
- Nickrent DL (2020) Parasitic angiosperms: How often and how many? *Taxon* 69, 5–27. doi:10.1002/tax.12195
- Nickrent DL, Duff RJ (1996) Molecular studies of parasitic plants using ribosomal RNA. In 'Advances in Parasitic Plant Research'. (Eds MT Moreno, JI Cubero, D Berner, D Joel, LJ Musselman, C Parker) pp. 28–52. (Junta de Andalucía, Dirección General de Investigación Agraria Córdoba: Córdoba, Spain)
- Nickrent DL, Duff RJ, Colwell AE, Wolfe AD, Young ND, Steiner KE, dePamphilis CW (1998) Molecular phylogenetic and evolutionary studies of parasitic plants. In 'Molecular Systematics of Plants II: DNA Sequencing'. (Eds DE Soltis, PS Soltis, JJ Doyle) pp. 211–241. (Kluwer: Boston, MA, USA) doi:10.1007/978-1-4615-5419-6_8
- Nickrent DL, García MA, Martín MP, Mathiasen RL (2004) A phylogeny of all species of *Arceuthobium* (Viscaceae) using nuclear and

- chloroplast DNA sequences. *American Journal of Botany* 91, 125–138. doi:10.3732/ajb.91.1.125
- Nickrent DL, Der JP, Anderson FE (2005) Discovery of the photosynthetic relatives of the ‘Maltese mushroom’ *Cynomorium*. *BMC Evolutionary Biology* 5, 38. doi:10.1186/1471-2148-5-38
- Nickrent DL, Malécot V, Vidal-Russell R, Der JP (2010) A revised classification of Santalales. *Taxon* 59, 538–558. doi:10.1002/tax.592019
- Nickrent DL, Anderson F, Kuijt J (2019) Inflorescence evolution in Santalales: integrating morphological characters and molecular phylogenetics. *American Journal of Botany* 106, 402–414. doi:10.1002/ajb2.1250
- Novák P, Guignard MS, Neumann P, Kelly LJ, Mlinarec J, Koblížková A, Dodsworth S, Kovařík A, Pellicer J, Wang W, Macas J, Leitch IJ, Leitch AR (2020) Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nature Plants* 6, 1325–1329. doi:10.1038/s41477-020-00785-x
- Petersen G, Cuenca A, Møller IM, Seberg O (2015a) Massive gene loss in mistletoe (*Viscum*, Viscaceae) mitochondria. *Scientific Reports* 5, 17588. doi:10.1038/srep17588
- Petersen G, Cuenca A, Seberg O (2015b) Plastome evolution in hemiparasitic mistletoes. *Genome Biology and Evolution* 7, 2520–2532. doi:10.1093/gbe/evv165
- Petersen G, Anderson B, Braun H-P, Meyer EH, Møller IM (2020) Mitochondria in parasitic plants. *Mitochondrion* 52, 173–182. doi:10.1016/j.mito.2020.03.008
- Petersen G, Shyama Prasad Rao R, Anderson B, Zervas A, Seberg O, Rasmussen AG, Møller IM (2022) Genes from oxidative phosphorylation complexes II–V and two dual-function subunits of complex I are transcribed in *Viscum album* despite absence of the entire mitochondrial holo-complex I. *Mitochondrion* 62, 1–12. doi:10.1016/j.mito.2021.10.006
- Pilger R (1935) Santalaceae. In ‘Die Natürlichen Pflanzenfamilien’. (Eds A Engler, K Prantl) pp. 52–91. (Engelmann: Leipzig, German Republic) [In German]
- Pillon Y, Gotty K, Lepschi BJ (2023) A revised generic circumscription of *Exocarpos* (Santalaceae), including the transfer of *Omphacomeria* to *Exocarpos*. *Muelleria* 42, 9–14.
- Ramm T, Roycroft EJ, Müller J (2020) Convergent evolution of tail spines in squamate reptiles driven by microhabitat use. *Biology Letters* 16, 20190848. doi:10.1098/rsbl.2019.0848
- Revell LJ (2024) *phytools* 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12, e16505. doi:10.7717/peerj.16505
- Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG (2014) From algae to angiosperms – inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14, 23. doi:10.1186/1471-2148-14-23
- Sanmartín I, Meseguer AS (2016) Extinction in phylogenetics and biogeography: from timetrees to patterns of biotic assemblage. *Frontiers in Genetics* 7, 35. doi:10.3389/fgene.2016.00035
- Sayyari E, Mirarab S (2016) Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* 33, 1654–1668. doi:10.1093/molbev/msw079
- Sayyari E, Mirarab S (2018) Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* 9, 132. doi:10.3390/genes9030132
- Schellenberg G (1932) Über Systembildung und über die Reihe der Santalales. *Berichte der Deutschen Botanischen Gesellschaft* 50a, 136–145. doi:10.1111/j.1438-8677.1932.tb03824.x
- Schröder L, Hohnjec N, Senkler M, Senkler J, Küster H, Braun H (2022) The gene space of European mistletoe (*Viscum album*). *The Plant Journal* 109, 278–294. doi:10.1111/tpj.15558
- Senkler J, Rugen N, Eubel H, Hegermann J, Braun H-P (2018) Absence of complex I implicates rearrangement of the respiratory chain in European mistletoe. *Current Biology* 28, 1606–1613. doi:10.1016/j.cub.2018.03.050
- Skippington E, Barkman TJ, Rice DW, Palmer JD (2015) Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proceedings of the National Academy of Sciences* 112, E3515–E3524. doi:10.1073/pnas.1504491112
- Skottsberg C (1935) Myzodendraceae. In ‘Die Natürlichen Pflanzenfamilien’, 2nd edn, vol. 16b. (Eds A Engler, K Prantl) pp. 92–97. (Engelmann: Leipzig, German Republic) [In German]
- Sleumer H (1935a) Olacaceae. In ‘Die Natürlichen Pflanzenfamilien’, 2nd edn, vol. 16b. (Eds A Engler, K Prantl) pp. 5–32. (Engelmann: Leipzig, German Republic) [In German]
- Sleumer H (1935b) Opiliaceae. In ‘Die Natürlichen Pflanzenfamilien’, 2nd edn, vol. 16b. (Eds A Engler, K Prantl) pp. 33–41. (Engelmann: Leipzig, German Republic) [In German]
- Smith SA, O’Meara BC (2012) *treePL*: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690. doi:10.1093/bioinformatics/bts492
- Su H-J, Hu J-M, Anderson FE, Der JP, Nickrent DL (2015) Phylogenetic relationships of Santalales with insights into the origins of holoparasitic Balanophoraceae. *Taxon* 64, 491–506. doi:10.12705/643.2
- Su H-J, Barkman TJ, Hao W, Jones SS, Naumann J, Skippington E, Wafala EK, Hu J-M, Palmer JD, dePamphilis CW (2019) Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. *Proceedings of the National Academy of Sciences* 116, 934–943. doi:10.1073/pnas.1816822116
- Sultan A, Robertson AW, Callmander MW, Phillipson PB, Meyer J, Tate JA (2019) Widespread morphological parallelism in *Korthalsella* (Santalaceae, tribe Viscaceae): a molecular phylogenetic perspective. *Taxon* 68, 1204–1218. doi:10.1002/tax.12152
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34, W609–W612. doi:10.1093/nar/gkl315
- The Angiosperm Phylogeny Group (1998) An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85, 531–553. doi:10.2307/2992015
- The Angiosperm Phylogeny Group (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141, 399–436. doi:10.1046/j.1095-8339.2003.t01-1-00158.x
- The Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161, 105–121. doi:10.1111/j.1095-8339.2009.00996.x
- The Angiosperm Phylogeny Group (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181, 1–20. doi:10.1111/boj.12385
- Vidal-Russell R, Nickrent DL (2008) The first mistletoes: origins of aerial parasitism in Santalales. *Molecular Phylogenetics and Evolution* 47, 523–537. doi:10.1016/j.ympev.2008.01.016
- Wicke S, Müller KF, dePamphilis CW, Quandt D, Bellot S, Schneeweiss GM (2016) Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences* 113, 9045–9050. doi:10.1073/pnas.1607576113
- Wilson CA, Calvin CL (2006) An origin of aerial branch parasitism in the mistletoe family, Loranthaceae. *American Journal of Botany* 93, 787–796. doi:10.3732/ajb.93.5.787
- Zervas A, Petersen G, Seberg O (2019) Mitochondrial genome evolution in parasitic plants. *BMC Evolutionary Biology* 19, 87. doi:10.1186/s12862-019-1401-8
- 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. doi:10.1186/s12859-018-2129-y
- Zonneveld BJM (2010) New record holders for maximum genome size in eudicots and monocots. *Journal of Botany* 2010, 527357. doi:10.1155/2010/527357
- Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Francisco E, Gallego-Paramo B, McGinnie C, Negrão R, Roy SR, Simpson L, Toledo Romero E, Barber VMA, Botigüé L, Clarkson JJ, Cowan RS, Dodsworth S, Johnson MG, Kim JT, Pokorny L, Wickett NJ, Antar GM, DeBolt L, Gutierrez K, Hendriks KP, Hoewener A, Hu A-Q, Joyce EM, Kikuchi IABS, Larridon I, Larson DA, de Lirio EJ, Liu J-X, Malakasi P, Przelomska NAS, Shah T, Viruel J, Allnutt TR, Ameka GK, Andrew RL, Appelhans MS, Arista M, Ariza MJ, Arroyo J, Arthan W, Bachelier JB, Bailey CD, Barnes HF, Barrett MD, Barrett RL, Bayer RJ, Bayly MJ, Biffin E, Biggs N, Birch JL, Bogarín D, Borosova R, Bowles AMC, Boyce PC, Bramley GLC, Briggs M, Broadhurst L, Brown GK, Bruhl JJ, Bruneau A, Buerki S, Burns E, Byrne M, Cable S, Calladine A, Callmander MW, Cano Á, Cantrill DJ, Cardinal-McTeague WM, Carlsen MM, Carruthers AJA, de Castro

Mateo A, Chase MW, Chatrou LW, Cheek M, Chen S, Christenhusz MJM, Christin P-A, Clements MA, Coffey SC, Conran JG, Cornejo X, Couvreur TLP, Cowie ID, Csiba L, Darbyshire I, Davidse G, Davies NMJ, Davis AP, van Dijk K, Downie SR, Duretto MF, Duvall MR, Edwards SL, Egli U, Erkens RHJ, Escudero M, de la Estrella M, Fabriani F, Fay MF, Ferreira PDL, Ficinski SZ, Fowler RM, Frisby S, Fu L, Fulcher T, Galbany-Casals M, Gardner EM, German DA, Giaretta A, Gibernau M, Gillespie LJ, González CC, Goyder DJ, Graham SW, Grall A, Green L, Gunn BF, Gutiérrez DG, Hackel J, Haevermans T, Haigh A, Hall JC, Hall T, Harrison MJ, Hatt SA, Hidalgo O, Hodkinson TR, Holmes GD, Hopkins HCF, Jackson CJ, James SA, Jobson RW, Kadereit G, Kahandawala IM, Kainulainen K, Kato M, Kellogg EA, King GJ, Klejevska B, Klitgaard BB, Klopper RR, Knapp S, Koch MA, Leebens-Mack JH, Lens F, Leon CJ, Léveillé-Bourret E, Lewis GP, Li D-Z, Li L, Liede-Schumann S, Livshultz T, Lorence D, Lu M, Lu-Irving P, Lubert J, Lucas EJ, Luján M, Lum M, Macfarlane TD, Magdalena C, Mansano VF, Masters LE, Mayo SJ, McColl K, McDonnell AJ, McDougall AE, McLay TGB, McPherson H, Meneses RI, Merckx VSFT, Michelangeli FA, Mitchell JD, Monro AK, Moore

MJ, Mueller TL, Mummenhoff K, Munzinger J, Muriel P, Murphy DJ, Nargar K, Nauheimer L, Nge FJ, Nyffeler R, Orejuela A, Ortiz EM, Palazzesi L, Peixoto AL, Pell SK, Pellicer J, Penneys DS, Perez-Escobar OA, Persson C, Pignal M, Pillon Y, Pirani JR, Plunkett GM, Powell RF, Prance GT, Puglisi C, Qin M, Rabeler RK, Rees PEJ, Renner M, Roalson EH, Rodda M, Rogers ZS, Rokni S, Rutishauser R, De Salas MF, Schaefer H, Schley RJ, Schmidt-Lebuhn A, Shapcott A, Al-Shehbaz I, Shepherd KA, Simmons MP, Simões AO, Simões ARG, Siros M, Smidt EC, Smith JF, Snow N, DE Soltis, Soltis PS, Soreng RJ, Sothers CA, Starr JR, Stevens PF, Straub SCK, Struwe L, Taylor JM, Telford IRH, Thornhill AH, Tooth I, Trias-Blasi A, Udovicic F, Utteridge TMA, Del Valle JC, Verboom GA, Vonow HP, Vorontsova MS, de Vos JM, Al-Wattar N, Waycott M, Welker CAD, White AJ, Wieringa JJ, Williamson LT, Wilson TC, Wong SY, Woods LA, Woods R, Worboys S, Xanthos M, Yang Y, Zhang Y-X, Zhou M-Y, Zmarzty S, Zuloaga FO, Antonelli A, Bellot S, Crayn DM, Grace OM, Kersey PJ, Leitch IJ, Sauquet H, Smith SA, Eiserhardt WL, Forest F, Baker WJ (2024) Phylogenomics and the rise of the angiosperms. *Nature* 629, 843–850. doi:10.1038/s41586-024-07324-0

Data availability. Raw reads for all newly generated sequences are available under projects PRJEB49212 (GAP stage 1), PRJEB78980 (GAP stage 2) and PRJEB79126 (sequences generated outside of Australia) at the European Nucleotide Archive (see <https://www.ebi.ac.uk/ena/browser/home>). Scripts and steps for the analyses are available at a GitHub repository (see <https://github.com/bmichanderson/Santalaceae>), archived at Zenodo (Anderson 2025). Data assembly and analyses were run on the National Computational Infrastructure (NCI) Gadi supercomputer.

Conflicts of interest. Brendan Lepschi and Jennifer Tate are Associate Editors for *Australian Systematic Botany* but were not involved in the peer review or decision-making process for this paper. *Australian Systematic Botany* encourages its editors to publish in the journal and they are kept totally separate from the decision-making processes for their manuscripts. The authors have no further conflicts of interest to declare.

Declaration of funding. The GAP Initiative is supported by funding from Bioplatforms Australia (enabled by NCRIS), the Ian Potter Foundation, Royal Botanic Gardens Foundation (Victoria), Royal Botanic Gardens Victoria, the Royal Botanic Gardens and Domain Trust, the Council of Heads of Australasian Herbaria, CSIRO, Centre for Australian National Biodiversity Research and the Department of Biodiversity, Conservation and Attractions, Western Australia. Additional funding was provided by the Royal Swedish Academy of Sciences and the Helge Ax:son Johnson's foundation. The Australian BioCommons Leadership Share (ABLES) program provided computing and data resources, and this program is co-funded by Bioplatforms Australia (enabled by NCRIS), the NCI and Pawsey Supercomputing Research Centre. The funders did not participate in the preparation of the manuscript. The decision to publish this work followed approval by the Department of Biodiversity, Conservation and Attractions, Western Australia.

Acknowledgements. We acknowledge the contribution of the Genomics for Australian Plants Framework Initiative consortium (see <https://www.genomicsforaustralianplants.com/consortium>) in the generation of the data used in this publication. We acknowledge the provision of computing and data resources provided by the ABLeS program.

Author affiliations

^AWestern Australian Herbarium, Department of Biodiversity, Conservation and Attractions, Locked Bag 104, Bentley Delivery Centre, Bentley, WA 6983, Australia.

^BDepartment of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91 Stockholm, Sweden.

^CAustralian National Herbarium, Centre for Australian National Biodiversity Research, Canberra, ACT 2601, Australia.

^DPlant Biology Section, School of Integrative Plant Science, College of Agriculture and Life Science, Cornell University, Ithaca, NY 14853, USA.

^ENational Herbarium of Pakistan (Stewart Collection), Plant Genetic Resources Institute, Pakistan Agricultural Research Council, National Agricultural Research Centre, Islamabad, Pakistan.

^FSchool of Agriculture and Environment, Massey University, Palmerston North, 4442, New Zealand.