Botanical Journal of the Linnean Society, 2018, **XX**, 1–10. With 3 figures.

Phylogenomic analysis of Ranunculales resolves branching events across the order

AMANDA K. LANE[1]*, MEGAN M. AUGUSTIN[3], SARAVANARAJ AYYAMPALAYAM[2], ALASTAIR PLANT[4], STEFAN GLEISSBERG[5], VERÓNICA S. DI STILIO[6], CLAUDE W. DEPAMPHILIS[7], GANE KA-SHU WONG[8,9,10], TONI M. KUTCHAN[3] and JAMES H. LEEBENS-MACK[1,2]

[1]*Department of Genetics, University of Georgia, Athens, GA 30602, USA*
[2]*Department of Plant Biology, University of Georgia, Athens, GA 30602, USA*
[3]*Donald Danforth Plant Science Center, St. Louis, MO 63132, USA*
[4]*Centre for Plant Sciences, University of Leeds, Leeds, LS2 9JT, UK*
[5]*Athens, OH 45701, USA*
[6]*Department of Biology, University of Washington, Seattle, WA 98195, USA*
[7]*Department of Biology, Pennsylvania State University, University Park, PA 16803, USA*
[8]*Department of Biological Science, University of Alberta, Edmonton, AB, T6G 2E9, Canada*
[9]*BGI-Shenzhen, Bei shan Industrial Zone, Yantian District, Shenzhen 518083, China*
[10]*Department of Medicine, University of Alberta, Edmonton, AB, T6G 2E1, Canada*

Received 28 November 2017; accepted for publication 9 March 2018

Poppies (*Papaver*), columbines (*Aquilegia*), buttercups (*Ranunculus*) and related species, are members of the order Ranunculales, the sister lineage to all other eudicots. Using coalescent and concatenated methods to analyse alignments of 882 putatively single copy genes, we have developed a robust phylogenetic tree including 27 species representing all major lineages of Ranunculales and most tribes of Papaveraceae. This first phylogenomic analysis including examples of Ranunculales, asterids, caryophillids and rosids provides the necessary foundation for future investigations of character evolution in Ranunculales. For example, the production of benzylisoquinoline alkaloids with known and potential pharmaceutical applications is largely restricted to Ranunculales. Understanding species relationships in the order is critical for understanding the evolution of benzylisoquinoline alkaloid biosynthesis across Ranunculales, including the production of morphine and codeine in opium poppy (*Papaver somniferum*). Analysis of gene tree discordance in selected portions of the phylogenetic tree suggests that the few observed differences between trees derived from supermatrix and coalescent-based summary analyses are attributable to incomplete lineage sorting. Discordance between gene tree and species tree inferences should be taken into account in future comparative analyses of character evolution in Ranunculales.

ADDITIONAL KEYWORDS: phylogenetics – basal eudicots – ancestral relationships – phylogenomics.

INTRODUCTION

Eudicots comprise 75% of all flowering plant diversity. Among this group, Ranunculales hold a key position in the angiosperm phylogenetic tree as the sister lineage to all other extant eudicots. Evolutionary analyses of Ranunculales have contributed to elucidation of the diversification of important traits including perianth morphology, biochemical pathways and woody vs. herbaceous habits (Kim *et al.*, 2004; Liscombe *et al.*, 2005; Sharma *et al.*, 2011; Bartholmes, Hidalgo & Gleissberg, 2012; Pabón-Mora *et al.*, 2013). Ranunculales have significant pharmaceutical importance due to their unique production of benzylisoquinoline alkaloids (BIAs), including analgesics such as morphine and codeine, antibacterials such as sanguinarine and anticancer drugs such as noscapine (Liscombe *et al.*, 2005; Hagel & Facchini, 2013; Beaudoin & Facchini, 2014). In order to better understand the evolution of

*Corresponding author. E-mail: lanea13@uga.edu

BIA biosynthesis and other traits, a well-supported phylogenetic tree for Ranunculales is required. Many family-level relationships within Ranunculales have been resolved in previous studies, but there are still ambiguities remaining for the placement of some lineages and resolution of species-level relationships in families (Hoot *et al.*, 1997; Soltis *et al.*, 2000; Kim *et al.*, 2004; Anderson, Bremer & Friis, 2005; Carolan *et al.*, 2006; Worberg *et al.*, 2007; Hilu *et al.*, 2008; Wang *et al.*, 2009; Soza, Haworth & Stilio, 2013; Lehtonen, Christenhusz & Falck, 2016). For example, the placement of Eupteleaceae is controversial and little work has been done to resolve relationships among species of *Papaver* L. Whereas much of what we know about the phylogenetics of Ranunculales is based on a few genes, many of which are encoded by the plastid genome, this study utilizes transcriptome data to provide a phylogenomic reconstruction of relationships across the order.

We use a phylogenomic approach to analyse species relationships for 27 species including representatives of ranunculids, asterids, caryophillids and rosids. Putatively single-copy loci have been identified from available genome sequences and transcriptomic data. We estimated and compared concatenation and coalescent-based species tree inferences. Importantly, we also characterize nodes on the species phylogenetic tree exhibiting possible conflict between gene trees and the species tree. Studying gene tree discordance provides additional insights for understanding ancestral character states and evolution of important traits such as BIA biosynthesis.

MATERIAL AND METHODS

Transcriptome assembly

Illumina 100 bp paired-end RNA-seq reads were assembled for 20 ranunculid species and acquired from the 1000 Plant Transcriptomes (1KP) project (Matasci *et al.*, 2014; Wickett *et al.*, 2014). These 20 species represent five out of seven families of Ranunculales including six poppy species (members of Papaveraceae subfamily Papaveroideae) exhibiting a broad range of BIA profiles. Tissues were collected by collaborators (Table 1), RNA was extracted and sequencing was performed using previously described protocols (Jiao *et al.*, 2012; Johnson *et al.*, 2012). RNA Seq libraries were prepared with an insert size of *c.* 200 base pairs and at least 2 Gb were sequenced for each sample as described in Wickett *et al.* (2014).

De novo transcriptome assemblies were built for each species using the Trinity *de novo* Assembler (v 2.0.6) platform including the *in silico* normalization pipeline with default parameters following a procedure described by Haas and colleagues (Grabherr *et al.*, 2011; Haas *et al.*, 2013). Read mapping and abundance estimation was then performed with RSEM (v 1.2.20), and all transcript assemblies with <1% of the per-component read mappings were removed (Li & Dewey, 2011). Transdecoder was used to translate nucleotide sequences into amino acid sequences (Haas *et al.*, 2013).

Gene selection

Sequences from the resulting transcriptomes were assigned to orthogroups circumscribed in a global gene family classification for land plant genomes developed by the *Amborella* Genome Project (2013). Transcript assemblies were sorted into orthogroups using BLAST and gene families identified as single copy among the 14 eudicots included in the orthogroup circumscription were evaluated further. As described by Wickett *et al.* (2014), in cases where multiple transcripts from a species sorted into a single copy orthogroup, a scaffolding procedure was performed to collapse sequences that had 95% identity or greater. After scaffolding transcripts were sorted into the putatively single copy gene families, additional copy number assessment and filtering were performed. If three or more of the 20 species contributed more than one sequence to an orthogroup that entire orthogroup was removed from the analysis. For the remaining orthogroups, sequences from species with more than one copy were removed from the gene family. Finally, only gene families with 80% of the 20 species included were retained for phylogenetic analysis.

After filtering species, occupancy in the gene families ranged from 58.7 to 100%. The differences in these percentages can be partially explained by differences in RNA-Seq sampling and sequence depth, which varied across taxa, and taxon-specific duplication events, which would exclude sequences from consideration. Orthologues from *Amborella trichopoda* Baill., *Vitis vinifera* L., *Musa acuminata* Colla, *Solanum lycopersicum* L., *Phoenix dactylifera* L., *Populus trichocarpa* Torr. & A.Gray ex Hook. and *Aquilegia coerulea* E.James were added to the gene family datasets before multiple sequence alignment and gene tree estimation.

Phylogenetic analyses

For each orthogroup (i.e. estimated gene family cluster), peptide sequences were aligned using MAFFT (v7.215-e) and then nucleotides were mapped to the amino acids alignment using PAL2NAL (v. 14) (Katoh, 2002; Suyama, Torrents & Bork, 2006). Maximum-likelihood (ML) analyses were performed using RAxML (v. 7.3.0) with GTRGAMMA or PROTGTRGAMMA models for nucleotide and amino acid alignments,

respectively (Stamatakis, 2006). *Amborella trichopoda*, the sister lineage to the clade containing all other extant angiosperms (e.g. *Amborella* Genome Project, 2013), was used to root all gene trees. Therefore, gene families that did not include *A. trichopoda* in the alignment were dropped from further analyses. This resulted in 882 genes being used for phylogenetic analyses.

A coalescent-based analysis was conducted using RAxML bootstrap gene trees as input for ASTRAL (v. 4.7.6), which utilizes a multi-locus bootstrapping approach to assess species tree clade support while accounting for conflict within and among gene family alignments (Mirarab *et al.*, 2014). An analysis of 882 concatenated gene alignments was also performed using RAxML with GTRGAMMA and PROTGTRGAMMA models for nucleotide and amino acid alignments, respectively, with 100 bootstrap replicates. In addition, gene tree quartet frequencies were calculated for the estimated species trees (Mirarab & Warnow, 2015) and local posterior probabilities were calculated (Sayyari & Mirarab, 2016a,b).

Inferred species trees included three regions with questionable resolution as reflected by low bootstrap support or conflict among trees estimated using contrasting analyses (i.e. amino acid vs. nucleotide alignments, and concatenated vs. ASTRAL analyses). Gene tree incongruence was assessed for these select relationships using custom scripts (github.com/kheyduk/Phylogenomics, accessed 9 April 2018). For example, *Euptelea pleiosperma* Hook f. & Thompson was placed sister to the rest of Ranunculales in some gene trees or sister to Papaveraceae in others. The numbers of trees with 50 or 80% bootstrap support for alternative placements for *E. pleiosperma* were counted using the getConflict.pl script (Heyduk *et al.*, 2015; github.com/kheyduk/Phylogenomics, accessed 9 April 2018). Relationships among the four *Papaver* spp. and species of Fumarioideae were also assessed by determining the number of gene trees that supported alternative relationships among the relevant lineages.

RESULTS

Concatenated and coalescent analyses returned phylogenetic estimates with topologies that were largely congruent (Fig. 1, 2). Euptelaceae are sister to a clade including all other lineages of Ranunculales. Papaveraceae are sister to a group containing Lardizabalaceae and Ranunculaceae, which are sister to Berberidaceae. Optimizing concordance among quartets in gene trees and the species tree inference, ASTRAL has been shown to converge on the true species phylogeny in the face of incomplete lineage sorting (Mirarab *et al.*, 2014). The ASTRAL tree was supported by 87.7% of the quartets present in the gene trees estimated on peptide alignments and 92.3% of the quartets in gene trees estimated on the nucleotide alignments. This indicates a low level of incomplete lineage sorting across the majority of the species tree.

Our analyses recovered many of the previously described familial relationships in Ranuculales, supported the circumscription of the fumarioid lineage as a subfamily in Papaveraceae and resolved previously unresolved relationships between the sampled *Papaver* spp. (Fig. 1) (Hoot *et al.*, 1997; Kim *et al.*, 2004; Wang *et al.*, 2009; Pérez-Gutiérrez *et al.*, 2012; Hoot, Wefferling, & Wulff, 2015; Sauquet *et al.*, 2015). Estimated branch lengths in the concatenated alignment analysis revealed several points on the tree with rapid rates of speciation including the early diversification of Ranunculales and *Papaver* (Fig. 2). As expected (Degnan & Rosenberg, 2009; Liu *et al.*, 2009), incongruence among quartets estimated from gene trees was highly elevated at these points in the species phylogenetic tree (Fig. 1). Euptelaceae were resolved as sister to a clade including Ranunculaceae and Papaveraceae, but the abundances of gene tree quartets supporting alternative resolutions were nearly equal (Fig. 1). We further investigated discordance among gene trees with respect to alternative relationships among Euptelaceae, Ranunculaceae and Papaveraceae (Fig. 3A). Hypotheses 1 and 2 shown in Fig. 3A have both been recovered in the phylogenetic literature (Hoot *et al.*, 1997, 2015; Kim *et al.*, 2004; Wang *et al.*, 2009; Pérez-Gutiérrez *et al.*, 2012), but hypothesis 1 clearly has stronger support in the gene trees and our species tree estimate (Fig. 1). A second point on the tree included three species for which topology varied depending on the alignment type used for reconstruction (amino acid vs. nucleotide) (Fig. 3B). Finally, three potential relationships between *Papaver* spp. were investigated due to variation in topology and bootstrap support in the different analyses (Fig. 3C). *Papaver somniferum* L. and *P. setigerum* DC. have been circumscribed as subspecies (Malik, Mary & Grover, 1979; Garnock-Jones & Scholes, 1990), and they reliably form a clade in our gene tree estimates.

The majority of gene trees support most of the relationships recovered in the ASTRAL analysis (Fig. 1), but with respect to relationships among *Papaver* spp., more gene trees supported the topology recovered in the concatenated analysis of the amino acid alignments (Fig. 3C hypothesis 1). This result may implicate an anomaly zone in which the most probable gene trees do not match the underlying species tree (Degnan & Rosenberg, 2006; Rosenberg & Tao, 2008). However, the impact of uncertainty in gene tree inference cannot be discounted since the majority of well-supported (bootstrap > 80%) gene trees were concordant with the ASTRAL trees.

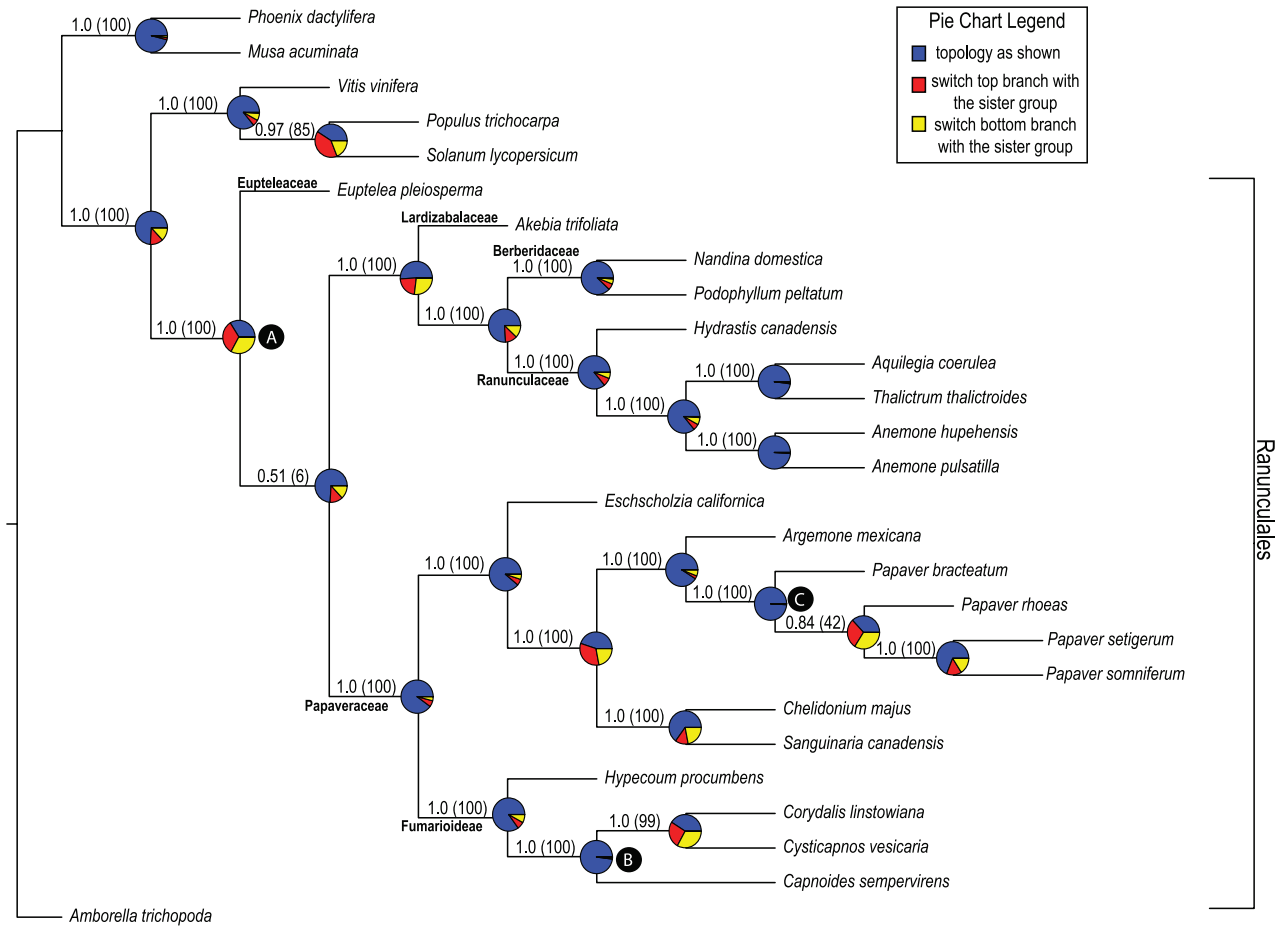

Figure 1. ASTRAL species-tree from 882 gene trees estimated from amino acid alignments. Bootstrap values are in parentheses and local posterior probabilities are adjacent. Pie charts depict the percentage of quartets from all gene trees that support one of three topologies: Q1 (blue) is the topology as shown, Q2 (red) is the lower child with the sister group and the upper child with the outgroup, Q3 (yellow) is the upper child with the sister group and the lower child with the outgroup (Mirarab & Warnow, 2015). Letters correspond to nodes discussed in the other figures. Branches are annotated with relevant family and tribe names in bold.

DISCUSSION

In the first phylogenomic study of the Ranunculales, we investigate the species tree and the underlying gene tree incongruences in order to understand the evolutionary framework of this important clade in angiosperm phylogeny. Here we illustrate the importance of understanding the nature of the input data including the incongruence among gene trees and among species tree estimates based on different methods of analysis. We identified lineages with significant gene tree incongruence that may be impossible to resolve even with large data sets.

The relationship between *Euptelea* Siebold & Zucc. and the rest of Ranunculales has been difficult to predict due to low bootstrap support in previous analyses (Wang *et al.*, 2009), but understanding the phylogenetic position of *Euptelea* is important for understanding character state evolution in Ranunculales (Kim *et al.*, 2004). For example, the placement of *Euptelea* influences reconstruction of woody versus herbaceous habit in Ranunculales (Kim *et al.*, 2004; Ren *et al.*, 2007). The placement also has implications for understanding the evolution of BIA biosynthesis because there is no evidence for BIA biosynthesis in *Euptelea*, whereas other lineages of Ranunculales are known to produce BIAs (Liscombe *et al.*, 2005).

In order to investigate the source of conflicting inferences for placement of *Euptelea* in previous

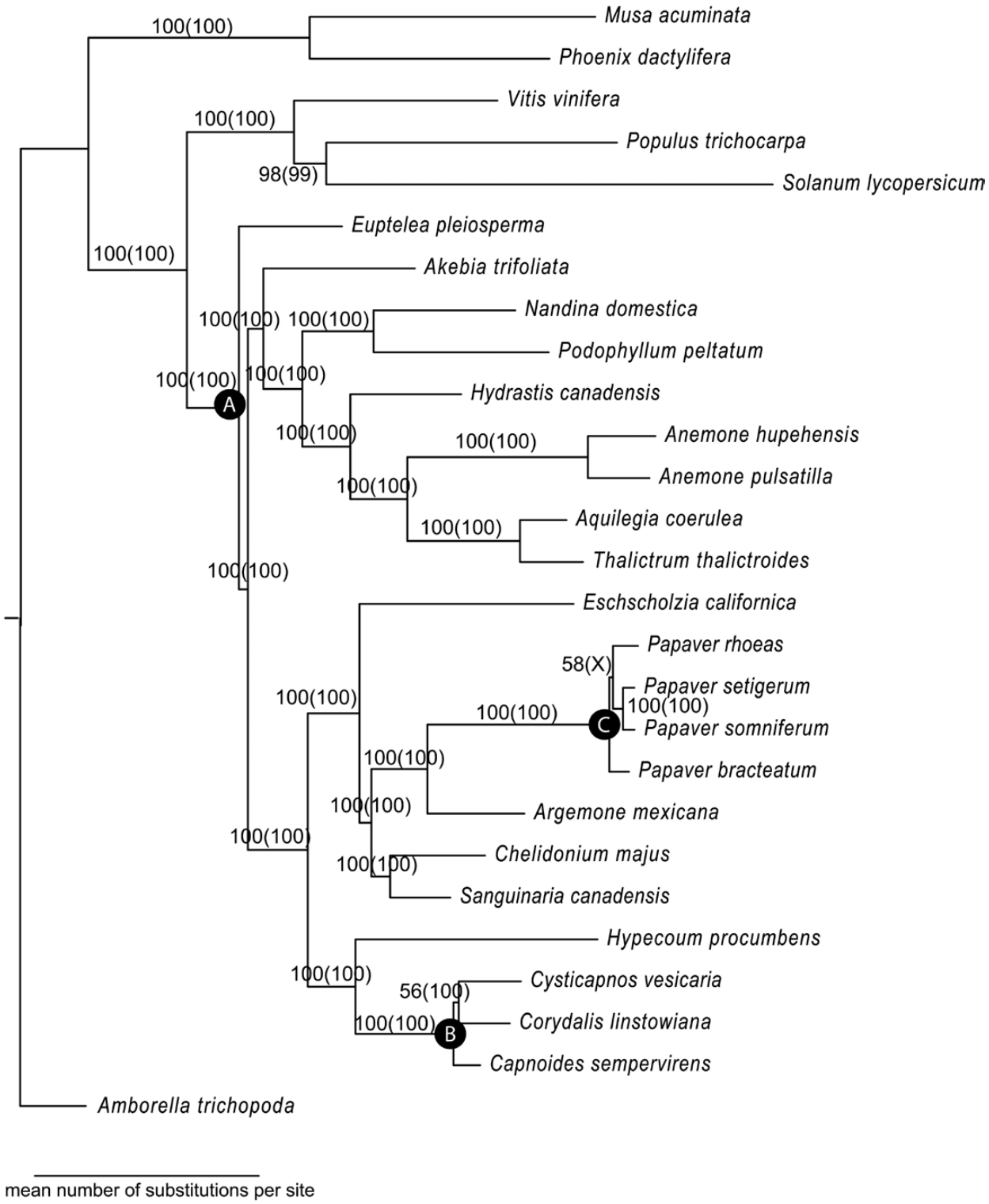

Figure 2. Concatenated species tree. This is a species-level tree was constructed using RAxML from a concatenated sequence from alignments of all genes included in the ASTRAL analysis. Bootstrap values for nucleotide and amino acid reconstructions are included with the nucleotide topology shown and the corresponding bootstrap values listed first. In parentheses are the bootstrap values from the amino acid-based analysis. X indicates partitions not included in the amino acid topology.

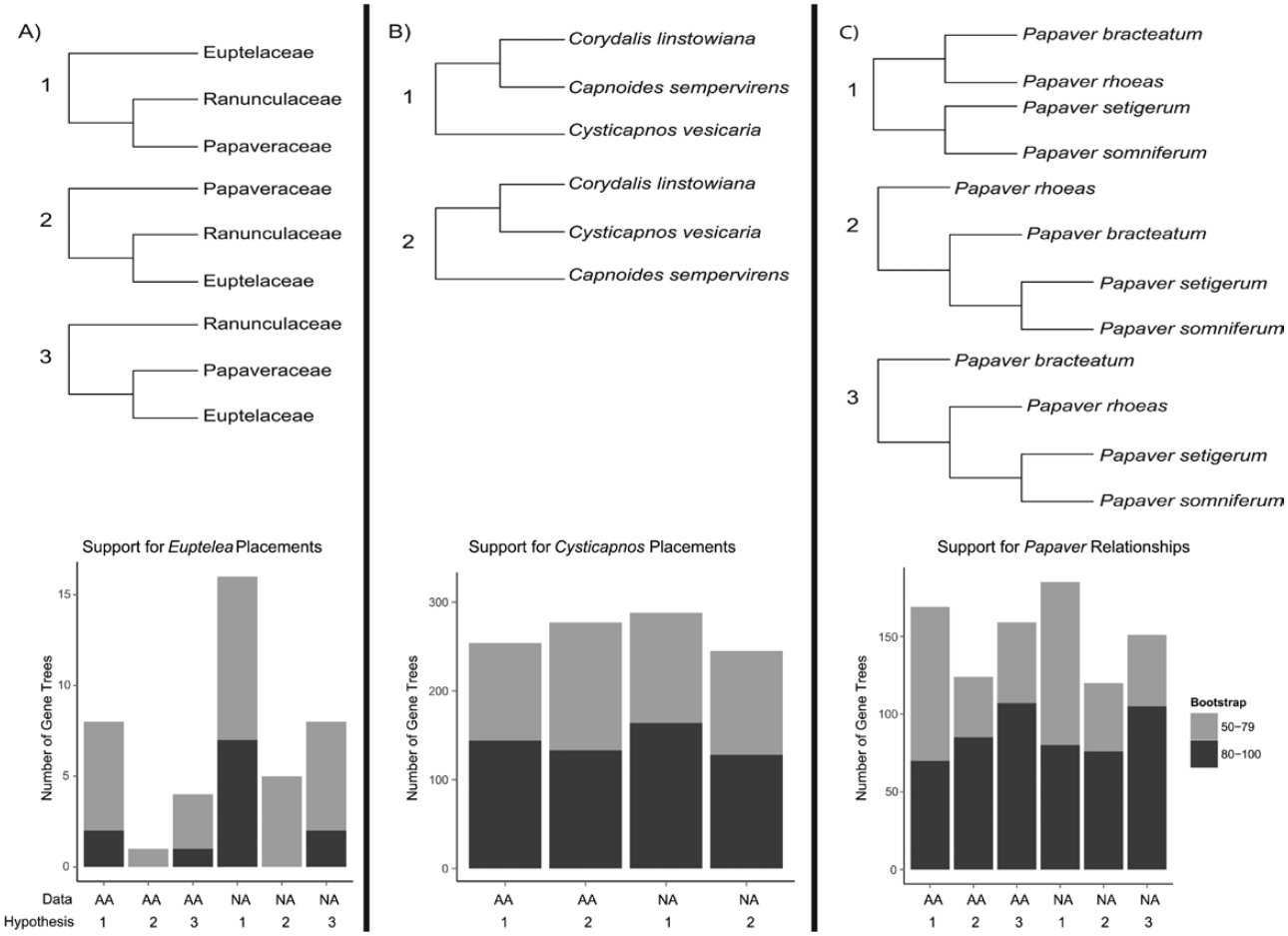

Figure 3. (Top) Hypotheses for unresolved species relationships. The letters of these hypotheses correspond to the nodes shown in other figures. (Bottom) Bar graphs showing number of gene trees that support varying hypotheses. The figure legend applies to parts A–C. The hypotheses numbers reference the hypotheses introduced in Supplementary Figure 1 and the data type represents if nucleotide (NA) or amino acid (AA) alignments were used to make the trees. Gene trees had to have a bootstrap support for each hypothesis either from 50–79 or >80% as indicated by shading. (A) Input gene trees were required to have Ranunculaceae and Papaveraceae to each be monophyletic with a bootstrap value of 50 or more, but did allow for missing taxa. *Euptelea pleiosperma* had to be present in all trees. (B) Input trees were required to have all three sampled members of Fumarioideae (*Corydalis linstowiana*, *Cysticapnos vesicaria* and *Capnoides sempervirens*). (C) Input gene trees were required to have *Papaver somniferum* and/or *Papaver setigerum* present. If both species were present, they were assumed to be sister.

studies (Soltis *et al.*, 2000; Kim *et al.*, 2004; Anderson *et al.*, 2005; Worberg *et al.*, 2007; Hilu *et al.*, 2008; Wang *et al.*, 2009), we assessed variation in phylogenetic signal among our gene trees (Fig. 3, see also Supporting Information, Fig. S1). Among the 795 gene trees that included *Euptelea*, the greatest number of trees supported *Euptelea* as sister to remaining Ranuculales, in agreement with our phylogenetic analysis (Fig. 3A hypothesis 1). Additionally, the internode branch for *Euptelea* is short, one condition for incomplete lineage sorting, which could explain why some gene trees are incongruent with the species tree (Maddison & Knowles, 2006).

Not all areas of conflict in the species tree showed clear gene tree support for one topology over another. The relationships between *Corydalis linstowiana* Fedde, *Capnoides sempervirens* (L.) Borkh and *Cysticapnos vesicaria* (L.) Fedde were examined because they were the only relationships that changed between the amino acid and nucleotide-based species trees in the coalescent analysis (Fig. 1, see also Supporting Information, Fig. S1). Both analyses yielded strong bootstrap support for their respective branching orders and the gene trees show some support for both topologies (Fig. 3C).

Conflicting signal among gene trees was also observed in the resolution of relationships among sampled

Table 1. Voucher table for RNA-Seq and genomic samples. Relevant information for identifying sample sources for both the RNA-Seq samples (all Ranunculales) and the acquired genomic data (Ranunculales and various angiosperm outgroups) is listed. Samples used from 1KP include the 1KP identifier. Genomic data are indicated with GENOME and are further identified in Wickett *et al.* (2014)

Identifier	Family	Species	Tissue type	Voucher data
WFBF	Berberidaceae	*Podophyllum peltatum*	leaves and root	*Deyholos 2013-11*
YHFG	Berberidaceae	*Nandina domestica*	young leaves and flower buds	*Soltis and Miles 2972*
QTJY	Eupteleaceae	*Euptelea pleiosperma*	leaves	*Chase 33135*
CCID	Lardizabalaceae	*Akebia trifoliata*	leaves	*Soltis and Miles 2751*
NMGG	Papaveraceae	*Hypecoum procumbens*		*GDAC 43575*
AUGV	Papaveraceae	*Capnoides sempervirens*		*NBGB 19871844*
UDHA	Papaveraceae	*Cysticapnos vesicaria*		*Goldblatt and Porter 12396 (MO)*
XHKT	Papaveraceae	*Sanguinaria canadensis*	young shoot and flower buds	*JLM2013-65 (GA)*
YDEH, GOQJ, SMZF, CCHG, BFMT	Papaveraceae	*Argemone mexicana*	leaf, stem, root, flower bud, developing fruit	*T. Kutchan 6173585 (MO)*
UNPT, NJKC, RKGT, TUHA, ERXG	Papaveraceae	*Eschscholzia californica*	leaf, stem, root, flower bud, developing fruit	*T. Kutchan 6173583 (MO)*
RRID, TMWO, ZSNV	Papaveraceae	*Papaver bracteatum*	leaf, bulb, root, stem	*M. Augustin MO 5452578*
ACYX, GMAM, QZBA, IORZ, MVTZ	Papaveraceae	*Papaver rhoeas*	leaf, stem, root, flower bud, developing fruit	*T. Kutchan 6173586 (MO)*
FNXH, MLPX, JSVC, STDO, EPRK	Papaveraceae	*Papaver setigerum*	leaf, stem, root, flower bud, developing fruit	*T. Kutchan 6173582 (MO)*
BMRX, SUFP, MIKW, FPYZ, KKCW	Papaveraceae	*Papaver somniferum*	leaf, stem, root, flower bud, developing fruit	*T. Kutchan 6173586 (MO)*
ZGQD	Papaveraceae	*Corydalis linstowiana*	leaves	*Chase 34442 K*
XMVD	Papaveraceae	*Chelidonium majus*	leaf	*M.K. Deyholos 2016–100*
VGHH	Ranunculaceae	*Hydrastis canadensis*	young shoots and flowers	*ABG19970771*
ZUHO	Ranunculaceae	*Anemone hupehensis*	leaf	*DWS NYBG 99/89A*
GBVZ	Ranunculaceae	*Thalictrum thalictroides*	floral buds	*V. Di Stilio 123 (WTU)*
UPOG	Ranunculaceae	*Anemone pulsatilla*	leaf	*M.K. Deyholos 2016–101*
GENOME	Amborellaceae	*Amborella trichopoda*		
GENOME	Vitaceae	*Vitis vinifera*		
GENOME	Salicaceae	*Populus trichocarpa*		
GENOME	Ranunculaceae	*Aquilegia coerulea*		
GENOME	Solanaceae	*Solanum lycopersicum*		
GENOME	Arecaceae	*Phoenix dactylifera*		
GENOME	Musaceae	*Musa acuminata*		

Papaver spp. (Fig. 3C). The trees estimated from the concatenated analyses differ from those inferred in the ASTRAL analyses (compare Figs 1 and 2). Previous studies have failed to resolve relationships among *P. bracteatum* Lindl., *P. somniferum* and *Papaver rhoeas* L. (Hoot *et al.*, 1997; Carolan *et al.*, 2006). Relationships between *Papaver* spp. are important for understanding the evolution of morphinan alkaloids because *P. somniferum* and its potential subspecies *P. setigerum* are the only plant species to produce some of these alkaloids (Malik *et al.*, 1979; Garnock-Jones & Scholes, 1990; Liscombe *et al.*, 2005). Additionally, *P. bracteatum* has been shown to produce morphinan precursors, such as thebaine, whereas *P. rhoeas* has not been shown to produce these compounds (e.g. Sharghi & Lalezari, 1967). Resolving the relationships among these three species is important for understanding the evolution of BIA biosynthesis. We determined the number of gene trees that supported each of three viable hypotheses with *P. setigerum* and *P. somniferum* forming a well-supported clade (Fig. 3C). The distribution of gene tree resolutions suggests that it is least likely for *P. bracteatum* to be sister to *P. somniferum*. However, the relationship observed in the majority of estimated gene trees matches the topology recovered by the concatenated amino acid analysis (169 total for amino acid and 185 total for nucleotide; Fig. 3C) with a balanced tree placing *P. bracteatum* and *P. rhoeas* in one clade sister to the *P. setigerum* + *P. somniferum* clade (Fig. 3C tree 1). This result is consistent with the expectation that with short intervals between speciation events, balanced gene trees may be more common than an unbalanced species tree, a situation described as the anomaly zone (Degnan & Rosenberg, 2006; Rosenberg & Tao, 2008).

Nonetheless, the observed discordance among inferred gene trees must be considered in comparative analysis of trait evolution. The histories of genes underlying traits of interest may not be consistent with the species tree. For ancestral character state reconstructions and comparative analyses of trait evolution the consequences of incomplete sorting and ancestral genes and all alternative gene topologies should be considered. For example, we find that *P. rhoeas* is sister to the *P. setigerum* + *P. somniferum* clade in the ASTRAL trees. However, *P. rhoeas* is the only sampled *Papaver* sp. that does not produce thebaine. Genes underlying production of thebaine may have been lost in the *P. rhoeas* lineage or null alleles for one or more of these genes may have existed in the ancestral *Papaver* population and subsequently became fixed in the *P. rhoeas* lineage. This second scenario is clearly possible given the high frequency of inferred gene trees with *P. bracteatum* as sister to the *P. setigerum* + *P. somniferum* clade.

Clear understanding of gene tree discordance and processes that contribute to discordance are necessary in order to guide inferences about species relationships and character evolution. For example, a whole genome duplication event has been inferred at the base of the Ranunculales and has been shown to contribute significantly to the evolution of certain gene families (Cui *et al.* 2006; Pabón-Mora *et al.* 2013). In addition to incomplete lineage sorting and interspecific gene flow, polyploidization and subsequent loss of paralogous genes in diverging lineage could contribute to gene tree discordance even in studies such as this that focus on putatively single copy genes. Comprehensive analyses of large numbers of gene alignments can help produce the data required to resolve species relationships in the face of ILS and other sources of gene tree discordance at various depths of a phylogenetic tree (e.g. Wickett *et al.*, 2014; Smith *et al.*, 2015).

The inferred phylogeny for Ranunculales (Fig. 1, see also Supporting Information, Fig. S1) supports previous work on the order, but also resolves nodes that were difficult to resolve with fewer genes. We were able to resolve a polytomy for four *Papaver* spp., which is useful for understanding BIA biosynthesis. A phylogenomic approach was able to elucidate species relationships at the same time as accounting for gene tree incongruence due to incomplete lineage sorting.

ACKNOWLEDGEMENTS

This work was supported by funding from the U.S. National Science Foundation IOS 0922742 (to Cwd and JL-M) and IOS 1121669 (to VSD), the U.S. National Institutes of Health 1R01DA025197 (to TMK, CWd and JL-M) and support for The 1000 Plants (1KP) initiative (GK-SW) from the Alberta Ministry of Innovation and Advanced Education, the Alberta Innovates Technology Futures, the Innovates Centres of Research Excellence, Musea Ventures and BGI-Shenzhen.

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES

***Amborella* Genome Project. 2013.** The *Amborella* genome and the evolution of flowering plants. *Science* **342:** 1241089–1241089.

Anderson CL, Bremer K, Friis EM. 2005. Dating phylogenetically basal eudicots using *rbcL* sequences and multiple fossil reference points. *American Journal of Botany* **92:** 1737–1748.

Bartholmes C, Hidalgo O, Gleissberg S. 2012. Evolution of the YABBY gene family with emphasis on the basal eudicot *Eschscholzia californica* (Papaveraceae). *Plant Biology* **14:** 11–23.

Beaudoin GAW, Facchini PJ. 2014. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **240:** 19–32.

Carolan JC, Hook ILI, Chase MW, Kadereit JW, Hodkinson TR. 2006. Phylogenetics of *Papaver* and related genera based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. *Annals of Botany* **98:** 141–155.

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* **16:** 738–749.

Degnan JH, Rosenberg NA. 2006. Discordance of species trees with their most likely gene trees. *PLoS Genetics* **2:** e68.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* **24:** 332–340.

Garnock-Jones PJ, Scholes P. 1990. Alkaloid content of *Papaver somniferum* subsp. *setigerum* from New Zealand. *New Zealand Journal of Botany* **28:** 367–369.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29:** 644–652.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8:** 1494–512.

Hagel JM, Facchini PJ. 2013. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant and Cell Physiology* **54:** 647–672.

Heyduk K, Trapnell DW, Barrett CF, Leebens-Mack J. 2015. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* **117:** 106–120.

Hilu KW, Black C, Diouf D, Burleigh JG. 2008. Phylogenetic signal in *matK* vs. *trnK*: a case study in early diverging eudicots (angiosperms). *Molecular Phylogenetics and Evolution* **48:** 1120–1130.

Hoot SB, Kadereit JW, Blattner FR, Schwarzbach AE, Crane PR. 1997. Data congruence and phylogeny of the Papaveraceae *s.l.* based on four data sets: *atpB* and *rbcL* sequences, *trnK* restriction sites, and morphological characters. *Systematic Botany* **22:** 575–590.

Hoot SB, Wefferling KM, Wulff JA. 2015. Phylogeny and character evolution of Papaveraceae *s. l.* (Ranunculales). *Systematic Botany* **40:** 474–488.

Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, Wu X, Zhang Y, Wang J, Zhang Y, Carpenter EJ, Deyholos MK, Kutchan TM, Chanderbali AS, Soltis PS, Stevenson DW, McCombie R, Pires JC, Wong GKS, Soltis DE, dePamphilis CW. 2012. A genome triplication associated with early diversification of the core eudicots. *Genome Biology* **13:** R3.

Johnson MTJ, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, DePamphilis CW, Edger PP, Goh F, Graham S, Greiner S, Hibberd JM, Jordon-Thaden I, Kutchan TM, Leebens-Mack J, Melkonian M, Miles N, Myburg H, Patterson J, Pires JC, Ralph P, Rolf M, Sage RW, Soltis D, Soltis P, Stevenson D, Stewart CN Jr, Surek B, Thomsen CJM, Villarreal JC, Wu X, Zhang Y, Deyholos MK, Wong GKS. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS One* **7:** e50226.

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30:** 3059–3066.

Kim S, Soltis DE, Soltis PS, Zanis MJ, Suh Y. 2004. Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody? *Molecular Phylogenetics and Evolution* **31:** 16–30.

Lehtonen S, Christenhusz MJM, Falck D. 2016. Sensitive phylogenetics of *Clematis* and its position in Ranunculaceae. *Botanical Journal of the Linnean Society* **182:** 825–867.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323.

Liscombe DK, MacLeod BP, Loukanina N, Nandi OI, Facchini PJ. 2005. Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry* **66:** 1374–1393.

Liu L, Yu L, Pearl DK, Edwards SV. 2009. Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* **58:** 468–477.

Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55:** 21–30.

Malik CP, Mary TN, Grover IS. 1979. Cytogenetic studies in *Papaver* V. Cytogenetic studies on *P. somniferum* × *P. setigerum* hybrids and amphiploids. *Cytologia* **44:** 59–69.

Matasci N, Hung LH, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, Burleigh JG, Gitzendanner MA, Wafula E, Der JP, dePamphilis CW, Roure B, Philippe H, Ruhfel BR, Miles NW, Graham SW, Mathews S, Surek B, Melkonian M, Soltis DE, Soltis PS, Rothfels C, Pokorny L, Shaw JA, DeGironimo L, Stevenson DW, Villarreal JC, Chen T, Kutchan TM, Rolf M, Baucom RS, Deyholos MK, Samudrala R, Tian Z, Wu X, Sun X, Zhang Y, Wang J, Leebens-Mack J, Wong GKS. 2014. Data access for the 1000 Plants (1KP) project. *GigaScience* **3:** 17.

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

Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics (Oxford, England)* **31:** i44–i52.

Pabón-Mora N, Hidalgo O, Gleissberg S, Litt A. 2013. Assessing duplication and loss of APETALA1/FRUITFULL homologs in Ranunculales. *Frontiers in Plant Science* **4:** 358.

Pérez-Gutiérrez MA, Romero-García AT, Salinas MJ, Blanca G, Carmen Fernández M, Suárez-Santiago VN. 2012. Phylogeny of the tribe Fumarieae (Papaveraceae *s.l.*) based on chloroplast and nuclear DNA sequences: evolutionary and biogeographic implications. *American Journal of Botany* **99:** 517–528.

Ren Y, Li HF, Zhao L, Endress PK. 2007. Floral morphogenesis in *Euptelea* (Eupteleaceae, Ranunculales). *Annals of Botany* **100:** 185–193.

Rosenberg NA, Tao R. 2008. Discordance of species trees with their most likely gene trees: the case of five taxa. *Systematic Biology* **57:** 131–140.

Sayyari E, Mirarab S. 2016a. Anchoring quartet-based phylogenetic distances and applications to species tree reconstruction. *BMC Genomics* **17:** 783.

Sayyari E, Mirarab S. 2016b. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* **33:** 1654–1668.

Sauquet H, Carrive L, Poullain N, Sannier J, Damerval C, Nadot S. 2015. Zygomorphy evolved from disymmetry in Fumarioideae (Papaveraceae, Ranunculales): new evidence from an expanded molecular phylogenetic framework. *Annals of Botany* **115:** 895–914.

Sharghi N, Lalezari I. 1967. *Papaver bracteatum* Lindl., a highly rich source of thebaine. *Nature* **213:** 1244–1244.

Sharma B, Guo C, Kong H, Kramer EM. 2011. Petal-specific subfunctionalization of an APETALA3 paralog in the Ranunculales and its implications for petal evolution. *New Phytologist* **191:** 870–883.

Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* **15:** 150.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* **133:** 381–461.

Soza VL, Haworth KL, Di Stilio VS. 2013. Timing and consequences of recurrent polyploidy in meadow-rues (*Thalictrum*, Ranunculaceae). *Molecular Biology and Evolution* **30:** 1940–1954.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

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.

Wang W, Lu AM, Ren Y, Endress ME, Chen ZD. 2009. Phylogeny and classification of Ranunculales: evidence from four molecular loci and morphological data. *Perspectives in Plant Ecology, Evolution and Systematics* **11:** 81–110.

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, Ruhfel BR, Wafula E, Der JP, Graham SW, Mathews S, Melkonian M, Soltis DE, Soltis PS, Miles NW, Rothfels CJ, Pokorny L, Shaw AJ, DeGironimo L, Stevenson DW, Surek B, Villarreal JC, Roure B, Philippe H, dePamphilis CW, Chen T, Deyholos MK, Baucom RS, Kutch TM, Augustin MM, Wang J, Zhang Y, Tian Z, Yan Z, Wu X, Sun X, Wong GKS, Leebens-Mack J. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences* **111:** E4859–E4868.

Worberg A, Quandt D, Barniske AM, Löhne C, Hilu KW, Borsch T. 2007. Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. *Organisms Diversity & Evolution* **7:** 55–77.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. ASTRAL species tree from 882 gene trees estimated from nucleotide alignments. Bootstrap values are in parentheses and local posterior probabilities are adjacent. Pie charts depict the percentage of quartets from all gene trees that support one of three topologies: Q1 (blue) is the topology as shown, Q2 (red) is the lower child with the sister group and the upper child with the outgroup and Q3 (yellow) is the upper child with the sister group and the lower child with the outgroup (Mirarab & Warnow, 2015). Letters correspond to nodes discussed in the other figures. Branches are annotated with relevant family and tribe names in bold.