# PLASTID PHYLOGENOMICS AND MOLECULAR EVOLUTION OF THISMIACEAE (DIOSCOREALES)

by

Natalie Garrett

B.Sc., The *University of British Columbia,* 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Plastid phylogenomics and molecular evolution of Thismiaceae (Dioscoreales)

submitted by Natalie Garrett in partial fulfillment of the requirements for

the degree of Master of Science

in Botany

**Examining Committee:**

Sean W. Graham, Professor, Botany, UBC

Supervisor

Quentin Cronk, Professor, Botany, UBC

Supervisory Committee Member

Jeannette Whitton, Associate professor, Botany, UBC

Supervisory Committee Member

Naomi Fast, Associate professor, Botany, UBC

Additional Examiner

**ABSTRACT**

PREMISE OF THE THESIS: Species in Thismiaceae can no longer photosynthesize, and instead obtain carbon from soil fungi. Here I infer Thismiaceae phylogeny using plastid genome data, and characterize the molecular evolution of this genome.

METHODS: I assembled five Thismiaceae plastid genomes from genome skimming data, adding to previously published data for phylogenomic inference. I investigated plastid genome structural changes considering locally colinear blocks (LCBs). I also characterized shifts in selection pressure in retained genes by considering changes in $\omega$, the ratio of non-synonymous to synonymous changes.

KEY RESULTS: Thismiaceae experienced two major pulses of gene loss around the early diversification of the family, with subsequent scattered gene losses in descendent lineages. In addition to massive size reduction, plastid genomes experienced occasional inversions and two losses of the inverted repeat (IR) region. Retained plastid genes remain under generally strong purifying selection ($\omega << 1$), but with significant and sporadic weakening or strengthening observed for several loci. The bifunctional *trnE-UUC* gene likely retains a role in haem biosynthesis despite repeated predicted losses of its functionality in translation. Several group IIA introns are retained despite loss of the intron maturase *matK*. The small single copy (SSC) region is reduced to a single bp in *Thismia rodwayi*.

CONCLUSIONS: I inferred that most gene losses in Thismiaceae occurred early and rapidly, following an initial loss of photosynthesis in the stem lineage. As a species-rich lineage of full mycoheterotrophs, Thismiaceae provides an excellent model system for uncovering the unique and divergent ways in which heterotrophic plastid genomes may evolve.

## LAY SUMMARY

The classification of the plant family Thismiaceae is contentious, with contradicting opinions on its evolutionary relationships within order Dioscoreales. The family comprises non-photosynthetic species that gain their nutrition from soil fungi instead of sunlight. The genetic sequences of their chloroplast (plastid) genomes have undergone substantial degradation due to gene loss. I infer a phylogenetic tree for Thismiaceae and relatives, and characterize newly sequenced plastid genomes for five species. This allowed me to investigate gene loss and retention patterns, changes in selection pressure on individual genes, and to reconstruct unusual genome structural changes in this family. My study adds to existing knowledge about the evolutionary history of this family.

## PREFACE

DNA extraction and library preparation for *Haplothismia exannulata* was performed by Marybel Soto Gomez (University of British Columbia; UBC); *Thismia huangii* was extracted and prepared by Vivienne Lam (UBC). *Thismia javanica* and *T. rodwayi* were extracted by Vincent Merckx (Naturalis Biodiversity Center) and *T. panamensis* was extracted by Juan Viruel (Royal Botanic Gardens, Kew). For *Thismia javanica*, *T. panamensis* and *T. rodwayi* library preparation was performed by Juan Viruel and contig assembly by Nate Klimpert (UBC). New plastid sequence data for *Burmannia championii*, *B. coelestis*, *B. cryptopetala*, *B. disticha*, *B. nepalensis*, and *B. oblonga* were generated by S. Jo (Korea University) and K.-J. Kim (Korea University) (unpublished). The complete annotated plastid of *Thismia tentaculata* was updated and expanded from a previous publication by Lim et al. (2016). The complete and partial annotated plastids of *Thismia alba*, *T. angustimitra*, *T. annamensis*, *T. cornuta*, *T. filiformis*, *T. gardneriana*, *T. hawkesii*, *T. hexagona*, *T. hongkongensis*, *T. kelabitiana*, *T. lanternata*, *T. mucronata*, *T. neptunis*, *T. okhaensis*, *T. puberula*, *T. thaithongiana*, and *T. viridistriata* were updated and expanded from previous publication by Yudina et al. (2021) with annotations provided pre-publication in personal communication with Yudina. Daisie Huang (UBC) provided filtering scripts used on DNA libraries. Wesley Gerelle (UBC) and Nate Klimpert advised on methods and analyses for using Compute Canada, performing dN/dS tests, and provided scripts for running PAML on ComputeCanada and for performing Benjamini-Hochberg corrections for multiple tests. I conducted all other analyses, including annotating plastid genomes, aligning sequence data, conducting phylogenetic analysis, performing locally colinear blocks analysis, and performing dN/dS tests. I also led the writing process with input from my supervisor, Sean Graham (UBC), with whom the study was also conceived and planned.

**TABLE OF CONTENTS**

**LIST OF TABLES**

**LIST OF FIGURES**

**ACKNOWLEDGEMENTS**

I would like to start by acknowledging the land on which I completed my thesis is the unceded territory of the Coast Salish Peoples, including the territories of the xʷməθkwəy̓əm (Musqueam), Skwxwú7mesh (Squamish), and Səl̓ílwətaʔ/Selilwitulh (Tsleil-Waututh) Nations.

I am grateful for the supervision and support of the Graham lab, especially for Dr. Sean Graham for taking me on as a student and continuing to support and encourage me through various health problems and data curveballs. In addition, I would like to thank Dr. Vivienne Lam, Dr. Marybel Soto Gomez, Nate Klimpert, Wesley Gerelle, and Dr. Qianshi Lin for all their technical expertise and for welcoming me into the Graduate school experience. I would not have been able to undertake this thesis without the generous data sharing of Nate Klimpert, Dr. Marybel Soto Gomez, and Dr. Vivienne Lam in particular.

I would also like to thank Paul McDade for custom coding me a few small programs and helping me troubleshoot many others.

I appreciate the support of Sophia Yudina for sending me the plastid genome consensus sequences and annotations for 18 *Thismia* species in personal communication before she published the data.

This research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca).

I am very grateful to Dr. Broadway at UBC student health service as well as Caroline Kingston at UBC's Centre for Accessibility for helping me find a path to complete my thesis when I wasn't sure there was one.

I wouldn't have been able to complete this thesis without the support of my family: Paul McDade, Pepper McDade, Alison Garrett, Bernie Garrett, and Rachel Garrett. Every time I felt overwhelmed one of them was there to help me through, and the times they have shouldered my burdens for me are innumerable.

# CHAPTER 1: INTRODUCTION

***1.1 Mycoheterotrophy in Thismiaceae***—Thismiaceae are a fully mycoheterotrophic family of flowering monocots, colloquially called 'fairy lanterns,' due to their colourful urceolate to campanulate flowers formed by fused tepals (Sochor et al., 2018). The family comprises ~90 species in up to five genera (e.g., Merckx et al., 2013; Yudina et al., 2021). Thismiaceae include two monotypic genera (*Haplothismia* and *Tiputinia*), *Oxygyne* (~6 species), *Thismia* (~80 or more species); the genus *Afrothismia* (~16 species; Shepeleva et al., 2020) should probably be excluded from the family based on phylogenetic evidence (see below). The distribution of Thismiaceae is primarily tropical, with extensions into subtropical and temperate areas (Merckx and Smets, 2014; Yudina et al., 2021). The largest genus, *Thismia*, is located primarily in tropical Southeast Asia and South America (Merckx and Smets, 2014). All Thismiaceae species are full mycoheterotrophs, as they lack the ability to photosynthesize, and instead rely on soil fungal partners to provide carbon and other nutrients (Leake, 1994; Bidartondo, 2005). Thismiaceae represent one of the largest lineages of fully mycoheterotrophic plants (Merckx, 2013). However, the age of the family is highly uncertain, with recent estimates ranging from over 100 Ma (i.e., based on the stem age of Burmanniaceae if taken to include Thismiaceae in the limited sampling in Fig. 6 of Hertweck et al., 2015), to 84 Ma (61-106 Ma) or 74.3 Ma for the stem age of Thismiaceae alone, excluding *Afrothismia* (see Merckx and Smets, 2014; Appendix 1 in Merckx et al. 2017, respectively) to only ~10 Ma (stem age of Thismiaceae, sister to Taccaceae and distinct from Burmanniaceae in Dioscoreales; see Fig. 3 in Givnish et al., 2018); note that the family was represented by a single species in Hertweck et al. (2015) and Givnish et al. (2018). These highly divergent date estimates may in part reflect the difficulty of including long-branch taxa in dating analysis (see Iles et al., 2015). The loss of photosynthesis in full mycoheterotrophs is also associated with substantial loss of genes related to photosynthesis and other plastid functions (e.g., Wicke et al., 2011; Graham et al., 2017). The compensatory mechanisms for gene losses in mycoheterotrophs are not well understood, other than that their nutrition is gained from their fungal partners; what the metabolites are that are passed from fungi to plant, and the biochemical pathways for storage of these metabolites, are all largely unresolved questions (Leake and Cameron, 2010).

***1.2 Phylogeny of Thismiaceae***—Recent angiosperm classifications (APG, 2003, 2009) lumped Thismiaceae in Burmanniaceae based on a phylogenetic study (Caddick et al., 2002) that likely included contaminant sequences (see Lam et al., 2016), and on a phylogenetic analysis (Hertweck et al., 2015; see APG, 2016) that included limited taxon sampling comprising two long-branch taxa (one species each from *Burmannia* and *Thismia*). The resulting angiosperm classifications are consistent with recognition of a single family, Burmanniaceae, with two tribes: Burmannieae and Thismieae, the latter including *Haplothismia*, *Oxygyne* and *Thismia* (Jonker, 1938; Maas et al., 1986). However, others consider Thismiaceae to be a separate family, distinct from Burmanniaceae within Dioscoreales (Chase et al., 1995; Woodward et al., 2007; Cheek et al., 2018), including authors of molecular analyses who examined one-three nuclear and/or mitochondrial genes (e.g., Merckx et al., 2006, 2009; Merckx and Smets, 2014; Shepeleva et al., 2020). Recognition of Thismiaceae as a distinct family scheme is also consistent with recent plastid and mitochondrial phylogenomic studies (Lam et al., 2018; Lin, 2020; Soto Gomez, 2020) and reconstructions based on morphological data (Soto Gomez, 2020). Species of Burmanniaceae are characterized by a persistent perianth and three stamens, whereas Thismiaceae have a circumsessile perianth and six stamens (with the exception of *Oxygyne* which has three) (Merckx et al., 2006; Cheek et al., 2018). Thismiaceae monophyly is also supported in recent studies, except that *Afrothismia* should likely be excluded from the family as it appears to represent an independent lineage of fully mycoheterotrophic plants in Dioscoreales (e.g., Merckx and Bidartondo, 2008; Merckx et al., 2009; Merckx and Smets, 2014; Shepeleva et al., 2020; Lin, 2020; Soto Gomez, 2020).

***1.3 Plastid genomes***—The plastid genomes of fully mycoheterotrophic plants are often highly modified compared to those of green plants. The typical plastid genome of photosynthetic land plants is a circular or linear-branched (Bendich, 2004; Oldenburg and Bendich, 2004) genome with a highly conserved quadripartite structure comprising a long single copy (LSC), and small single copy (SSC) separated by two inverted repeats (IRs) (Palmer, 1991; Wicke et al., 2011). The boundaries of the IRs tend to shift across species through gradual expansion and contraction (Palmer, 1991; Wicke et al., 2011). The usual counts of unique genes for photosynthetic plastid genomes are around 110–120 genes (excluding repeat copies) with a typical complement of ~78 protein-coding genes, 4 rDNA genes, and 30–32 unique tRNA genes (e.g., Wicke et al., 2011;

Rogalski et al., 2015; Xu et al., 2015). Plastid genomes in photosynthetic plants typically range between ~120–160 kb (e.g., Rogalski et al., 2015; Xu et al., 2015). In contrast, various modifications of plastid genomes in fully mycoheterotrophic plants include massive shrinkage and gene loss (e.g., Graham et al., 2017, and for a recent summary see Appendix S3 in Lam et al., 2018). Plastid genome degradation is thought to proceed in a relatively predictable manner, where genes are lost in stages in an irreversible ratchet-like pattern (e.g., Barrett and Davis, 2012; Wicke et al., 2016; Graham et al., 2017). NADH dehydrogenase gene and other genes involved in photosynthesis are predicted to be lost first, as plants transition from partial to full heterotrophy. In contrast, several non-bioenergetic genes (*accD*, *clpP*, *trnE*, *ycf1*, and *ycf2*) and the ribosomal apparatus genes needed to maintain their translation in the plastid are predicted to be retained the longest (e.g., Graham et al., 2017).

***1.4 Group IIA introns***—Mycoheterotrophic plants also tend to experience intron loss in retained plastid genes. Land-plant plastid introns are classified as group I, IIA or IIB, based on the type of conserved RNA folding patterns found at splice sites (Michel and Dujon, 1983; Christopher and Hallick, 1989; Kelchner, 2002). Typical angiosperm plastid genomes include 17–20 group II introns and one group I intron (*trnL-UAA*) (Vogel et al., 1999; McNeal et al., 2009). Group II introns can be further broken down into subclasses group IIA and IIB (Michel et al., 1989; Kelchner, 2002): single group IIA introns are found in *atpF*, *clpP* (intron 2), *rpl2*, *3'-rps12*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*, and *trnV-UAC*, and group IIB introns are found in *clpP* (intron 1), *ndhA*, *ndhB*, *rpl16*, *rpoC1*, *rps12* intron 1 (trans-spliced 1a/1b), *rps16*, *petB*, *petD*, *trnG-UCC*, *ycf3* (introns 1 and 2) (Jenkins et al., 1997; Kelchner, 2002; McNeal et al., 2009; Zoschke et al., 2010). Group IIA introns (with the possible exception of *clpP* intron 2; Zoschke et al., 2010) are thought to require a maturase (the product of the plastid gene *matK*) to be spliced (McNeal et al., 2009; Zoschke et al., 2010). However, *rpl2* and *3'-rps12* may also not require *matK* for splicing, as Barthet et al. (2020) found they were able to self-splice in vitro (at reduced levels), and several heterotrophic lineages retain these introns but lack *matK* (Graham et al., 2017). The *rps12* transcript is trans-spliced in plastid genomes, with exon 1 transcribed separately from exon 2 and exon 3; the two transcripts are being covalently ligated together in the correct reading frame (Koller et al., 1987; Hildebrand et al., 1988). Although this process has not been fully characterized, the two distantly transcribed mRNAs are thought to be trans-spliced together with

proteins EMB2654 binding to the 5'-*rps*12 transcript and PPR4 binding to the 3'-*rps*12 transcript during the trans-splicing process (Lee et al., 2019). Here I explore intron loss and retention across Thismiaceae.

***1.5 Overview of thesis goals***—I assembled five new Thismiaceae plastid genomes from two genera (*Haplothismia* and *Thismia*) with the dual goals of inferring the phylogeny of the family and investigating the molecular evolution of its plastid genomes. My explorations of molecular evolution in Thismiaceae include reconstruction of patterns of gene loss and retention, intron loss and retention, genome structural evolution, and changes in selection pressure for retained genes. Specific questions include: (1) Are Thismiaceae monophyletic at the current taxon sampling, and do they form a clade separate from Burmanniaceae in Dioscoreales? What are the overall phylogenetic relationships in the family? (2) What structural changes to the plastid genomes occurred in Thismiaceae, including gene loss, pseudogenization and intron loss, and do the losses provide structural synapomorphies that characterize individual subclades? (3) Has the mycoheterotrophic mode of life in Thismiaceae affected the strength of selection acting on individual retained plastid genes?

## CHAPTER 2: MATERIALS AND METHODS

***2.1 Taxon sampling and library preparation***—We selected taxa in Thismiaceae that expand the sampling in Yudina et al. (2021) (see Table A1 for a complete list of Thismiaceae taxa). Outgroup sampling is based on Givnish et al. (2018) but is more heavily focused on Dioscoreales and Pandanales (Soto Gomez, 2020; Soto Gomez et al., 2020). *Haplothismia exannulata* was prepared using whole-genome shotgun sequencing with a NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, USA) and sequenced on an Illumina MiSeq platform to obtain 150-bp paired-end reads (for further details, see Soto Gomez, 2020). *Thismia huangii* was prepared using whole-genome shotgun sequencing with the Bioo Nextflex DNA sequencing kit (Bioo Scientific Corp., Austin, TX) and KAPA LTP Library Preparation kit (KAPA Biosystems, Boston, MA) and sequenced on an Illumina HiSeq 2000 (Illumina, Inc., San Diego, CA) as 100-bp paired-end reads (for further details, see Lam et al., 2015). *Thismia javanica*, *T. panamensis*, and *T. rodwayi* were prepared using HybSeq, an approach that combines genome enrichment and genome skimming (Steele et al., 2012; Straub et al., 2012); the latter allows plastid genome recovery as unbaited by-product, possible because of the effectively high copy number of plastid genomes per cell (e.g., Weitemier et al., 2014; Twyford and Ness, 2017). For this we used whole-genome shotgun sequencing performed as in Soto Gomez et al. (2019) with NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, USA), which were then sequenced on an Illumina MiSeq platform to obtain 150-bp paired-end reads.

***2.2 Genome assembly and annotation***—I used CLC Genomics Workbench v.6.5.1 (CLC Bio, Aarhus, Denmark) to perform *de novo* assembly, generating contigs with at least 30X coverage and a length of at least 500 bp, which were then filtered using a custom Perl script (Daisie Huang, University of British Columbia; https://github.com/daisieh/phylogenomics/tree/master/filtering/filter_cp.pl) to separate predicted plastid genome contigs from mitochondrial and nuclear contigs. This script uses the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) against a local database of plastid genes (in this case *Tacca leontopetaloides*). To aid in the scaffolding of the filtered CLC contigs into a complete circular genome I used NOVOplasty 2.7.2 (Dierckxsens et al., 2017). I used an *H.*

*exannulata* contig as a seed for this in all cases, set the expected genome range to 10,000–200,000 bp, the K-mer to 20, with all other settings the default ones. I mapped the CLC contigs onto the NOVOplasty result to create a consensus plastid genome sequence using Sequencher 4.2.2 (Gene Codes Corporation, Ann Arbor, US). I found that the CLC and NOVOplasty assemblies only differed in a few minor instances in which repeat regions had different repeat lengths, or in calling of potentially ambiguous bases; I generally went with the NOVOplasty calls in these situations. NOVOplasty tended to yield longer contigs compared to CLC; the additional data aided greatly in finding contig overlap. To reconcile small differences, I kept the additional sequence, called versions for the handful of ambiguous bases, and kept shorter repeat regions where lengths differed. In several cases I also used Sanger sequencing to fill gaps and verify overlap between contigs, designing primers for amplification and sequencing using Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2007); see Table A2 for a list of primers. The amplification and sequencing methods follow Lam et al. (2015).

I used the online software DOGMA (Wyman et al., 2004) and GeSeq (Tillich et al., 2017) to perform initial gene recovery and create draft annotations. I used Sequencher and AliView (Larsson, 2014) to manually check and adjust Thismiaceae taxa gene and exon boundaries against autotrophic relatives *Tacca chantrieri* (KX171420), *T. leontopetaloides* (NC_036658.1), and *Dioscorea elephantipes* (NC_009601). I also performed a comparison of gene sets and locations to autotrophic relatives to identify genes that were potentially missing compared to autotrophs and to search for them. For this I used the Hmmer webserver (Potter et al., 2018) to run phmmer using the 'Ensembl Genomes Plants' sequence database, with all other settings default, and the NCBI BLAST (Altschul et al., 1990) BLASTN tool, with program selection optimized for 'somewhat similar sequences' and other settings as default. I used these approaches and manual scanning (paying particular attention to intergenic regions over 500 bp long) to uncover previously undetected genes, which we found to be relatively frequently missed for these rapidly evolving and often rearranged genomes (Lim et al., 2016; Yudina et al. 2021; e.g., recovery here of *matK* and *rps18* in *Haplothismia exannulata* and *rps3*, *rps18*, and *rrn4.5* in *Thismia rodwayi*, which were missed in Soto Gomez 2020). I found that short exons and gene fragments (such as 3'-*rps12* exon 3) are also sometimes missed by GeSeq; in addition, more divergent genes that have a high indel frequency in multi-sequence alignments (such as *accD*), often only had a fragment annotated by GeSeq, when a longer open reading frame can be

recovered. In addition, *rrn4.5* was not automatically annotated in much of Thismiaceae, and *rrn23* and *rrn16* often required substantial boundary edits, with more minor boundary edits in several other genes. The updates to previously published plastid genomes (see below) are documented more fully in Table A3. I also used tRNAscan-SE (Chan and Lowe, 2019) within GeSeq for initial tRNA locus annotations, and used the standalone version of the tRNAscan-SE program to predict the structure of all annotated Thismiaceae tRNA genes, in order to identify possible pseudogenes (Table A4). Finally, I used OGDraw (Lohse et al., 2013) to generate graphical representations of the newly sequenced full or partial plastid genomes.

***2.3 Sequence alignment and matrix creation***—I added the plastid data for the 23 Thismiaceae species—five newly assembled and 18 from Lim et al. (2016) and Yudina et al. (2021)—to a previous matrix (Soto Gomez, 2020). I generated automated alignments for each gene using MUSCLE (Edgar, 2004) run through AliView (Larsson, 2014), with default settings, and manually adjusted where needed following criteria in Graham et al. (2000), staggering difficult to align regions such as AT-rich repeat regions (e.g., Saarela & Graham, 2010). I concatenated the completed gene alignments into a single matrix using Mesquite (Maddison and Maddison, 2018). I cross-checked the matrix data for editing errors against the original data using Sequencher (none were found). The resulting matrix contains 127 species, across six Dioscoreales families (Burmanniaceae, Dioscoreaceae, Nartheciaceae, Taccaceae, Thismiaceae, and Trichopodaceae), with all 12 monocot orders (Givnish et al., 2018) represented. As more distant outgroups, I included representatives from the eudicots, magnoliids, Amborellales, Austrobaileyales and Nymphaeales (see Table A1).

***2.4 Phylogenetic inference***—I partitioned the matrices initially by gene and codon position (e.g., Lam et al., 2015), and then used PartitionFinder 2.1.1 (Lanfear et al., 2014, 2016; Stamatakis, 2014) to combine partitions with similar DNA substitution models. I used the relaxed hierarchical clustering algorithm (r-cluster), corrected Akaike Information Criterion (AICc) settings, limiting model selection to those implemented by RAxML. I then used PartitionFinder 2.1.1 to identify the optimal model for each version of the unpartitioned matrix (125- and 127-taxon versions). The invariable sites parameter "I" is accommodated by the gamma parameter

"G" (Yang, 2006) and so I set the 'Estimate proportion of invariable sites (GTRGAMMA + I)' setting to 'No' in the RAxML analyses.

I used maximum likelihood to infer phylogenetic relationships using RAxML 8.2.12 (Stamatakis, 2014) via the CIPRES portal (Miller et al., 2010), which provides access to large computational resources (NSF XSEDE) through a browser interface to run analyses. I conducted separate sets of analyses with *Apteria aphylla* and *Gymnosiphon longistylus* both either included or excluded (these are inferred to represent the longest branches in Burmanniaceae, see below).

For each analysis I ran RAxML-HPC2 for parallel fast bootstrap analysis and maximum likelihood best tree searches with 200 bootstrap replicates (40 randomized stepwise addition replicates), GTRGAMMA models, and with all other settings default. RAxML-HPC2 runs the 200 bootstrap replicates first then every fifth bootstrap tree is used as a starting tree for a subsequent maximum likelihood best tree search (Stamatakis et al., 2008). I used the fast bootstrap analysis to estimate branch support (Felsenstein, 1985), characterizing bootstrap support values of <70%, 70–94%, and 95% or greater bootstrap support, as poor, moderate, and strong support, respectively (see Soltis and Soltis, 2003).

***2.5 Genome structural evolution***—I used Mauve (Darling et al. 2004; 2015-02-06 development snapshot version) to compare gene order and rearrangements between plastid genomes. Mauve identifies regions with shared homology between sequences, referred to as locally colinear blocks (LCBs), using a sum-of-pairs based algorithm, and positions them using progressive alignment. The typical plastid genome contains two inverted repeats (IRs) separating large and small single copy (LSC and SSC, respectively). The SSC can also be found in both forward and reverse orientations in vivo, in a roughly 50:50 mixture of inversion isomers (Palmer, 1985). To avoid artificially inflating estimates of structural change, I made consensus sequences with only a single copy of the IR, and with the SSC in the same orientation in each case. I applied this process to all 23 Thismiaceae plastid genomes and for two outgroups (*Tacca chantrieri*, Taccaceae, GenBank accession KX171420; *Dioscorea elephantipes*, Dioscoreaceae, GenBank accession NC_009601). I used an arbitrary but consistent starting point to generate linearized genomes, by starting them all with the first base of the LSC at the end closest to *rpl2*, and then continuing through the LSC, the IR and ending with the last base of the SSC. To run the Mauve

analyses, I used ProgressiveMauve with a seed weight of 17, with the 'Use seed families' option enabled and all other settings set as default.

I also mapped gene losses, pseudogenization events, and inverted repeat (IR) losses (see below) onto the 125-taxon partitioned likelihood tree using Mesquite (Maddison and Maddison, 2018), focusing on the ingroup (Thismiaceae) taxa only. I used Dollo parsimony for ancestral-state reconstructions (ASRs), setting up user-defined step matrices (as in Fig. 4) with high enough penalties (biases) to reject gene (or IR) re-gain. These weighting schemes assume that complex traits are much easier to lose than to gain. Similarly, for genes with three states, the Dollo step matrix (as in Fig. 4) biases against reversions from pseudogenes to genes present, and against reversions from lost genes to pseudogenes. For incomplete genomes I coded character states as "unknown" when I had no evidence for gene presence/absence. In cases where only partial (potentially truncated) genes were found I scored the gene as present if the gene was at least 50% complete compared to the reference, and it had an open reading frame for the fragment.

***2.6 Tests of change in purifying selection***—I used the CodeML module in PAML 4.9 (Yang, 2007) on a subset of Thismiaceae chosen to represent the major clades (see below), to assess changes in selection pressure in 18 retained plastid genes in a subset of Thismiaceae species. I considered each representative Thismiaceae species in turn as a single included "foreground" branch (i.e., with other Thismiaceae excluded), with the rest of the tree counted as "background"; the tree was derived from the partitioned 125-taxon tree, pruned to only include the single focus foreground species plus autotrophic members of Dioscoreales, but excluding photosynthetic *Burmannia* species, which may be partially mycoheterotrophic (Merckx et al., 2010; Bolin et al., 2017); the taxa included in the PAML analyses are noted in Table A1. Changes in the selective regime can be detected by comparing the ratio of nonsynonymous substitutions to synonymous substitutions, ω (also known as dN/dS) in foreground (Thismiaceae) vs. background branches (autotrophic Dioscoreales) (e.g., see Lam et al., 2015). I used the "branch site" model tests in PAML 4.9 at ComputeCanada using custom Python scripts (W. Gerelle and N. Klimpert, University of British Columbia; https://github.com/wesleykg/paml_automation, https://github.com/nklimpert/mht_selection). I used a one-ratio model for the null model (where all branches are assumed to evolve under one ω-ratio) and a free-ratio model for the alternative

model (where each lineage is assumed to have its own ω-ratio). I also applied a Benjamini-Hochberg correction to adjust for multiple tests being run on the same data, considering test results significant if the branch *P*-value was less than the branch false-discovery rate (FDR).

## CHAPTER 3: RESULTS

***3.1 Phylogenetic relationships in Dioscoreales***—The tree-wide relationships in Dioscoreales recovered in the four likelihood analyses of plastid coding regions are largely congruent between partitioned and unpartitioned versions of data sets (Fig. 1; Figs. A1–A4). However, the 127- and 125-taxon analyses differ in whether two long-branch taxa in Burmanniaceae, *Apteria aphylla* and *Gymnosiphon longistylus*, are excluded (Figs. A5, A6) vs. included (Figs. A7, A8), and these analyses had substantially different placements of *Thismia* (representing the bulk of sampled Thismiaceae here, all very long-branch taxa, e.g., Fig. 1). In the case of the 127-taxon analysis *Thismia* is recovered as the sister group of *Gymnosiphon longistylus,* with a clade comprising these taxa and *Apteria aphylla* then sister to other Burmanniaceae (Fig. A3, Fig. A4). However, *Haplothismia exannulata* (the member of Thismiaceae with the shortest branch in the family, see Fig. 1) is inferred to be the sister group of Taccaceae, highly disjunct from other Thismiaceae in this analysis. These arrangements with *Thismia* close to long-branch Burmanniaceae and *Haplothismia* sister to Taccaceae have poor to strong support, respectively, in the unpartitioned vs. partitioned analyses of this taxon set.

In contrast, the 125-taxon analysis, which excludes the two long-branch Burmanniaceae lineages (*Apteria* and *Gymnosiphon*), recovers Thismiaceae as a well-supported monophyletic group, with *Haplothismia* then sister to *Thismia*, and with the family as a whole sister to Taccaceae, also with strong support (Fig. 1; Figs. A1–A4). The 125-taxon partitioned and unpartitioned analyses recovered identical topologies within Thismiaceae, with generally similar support values for these relationships, although several branches within Thismiaceae have moderate to strong support across analyses. Well-supported branches include those defining the deepest splits in the family (Figs. 1, A1–A2). Most branches in Thismiaceae are very long compared to the rest of Dioscoreales, although *Haplothismia* is inferred to be far less rate-elevated than the other taxa in the family (see branch lengths in Fig. 1).

***3.2 Plastid genome features***—I assembled complete plastid genomes for four previously unsequenced mycoheterotrophic Thismiaceae taxa (i.e., *Haplothismia exannulata*, Fig. A5; *Thismia javanica*, Fig. A6; *T. panamensis*, Fig. A7; and *T. rodwayi*, Fig. A8), a partial mycoheterotrophic plastid genome (*T. huangii*, Fig. A9), and I also re-annotated the plastid

genomes of multiple published mycoheterotrophic plastid genomes—i.e., *T. alba*, *T. angustimitra*, *T. annamensis*, *T. cornuta*, *T. filiformis*, *T. gardneriana*, *T. hawkesii*, *T. hexagona*, *T. hongkongensis*, *T. kelabitiana*, *T. lanternata*, *T. mucronata*, *T. neptunis*, *T. okhaensis*, *T. puberula*, *T. thaithongiana*, and *T. viridistriata*—all originally assembled by Yudina et al. (2021)—in addition to *T. tentaculata,* originally assembled by Lim et al. (2016). These updates are summarized in Table A3 (and see Fig. A10 for the reannotated genome of *T. tentaculata*). I compared these plastid genomes to two published photosynthetic Taccaceae taxa; *Tacca leontopetaloides* (GenBank accession NC_036658.1) and *Tacca chantrieri* (GenBank accession KX171420). The following text focuses on the five species newly assembled here unless otherwise specified.

The five newly assembled mycoheterotrophic plastid genomes in Thismiaceae (four complete, one partial; Fig. 2) are substantially reduced in size and gene content compared to those of photosynthetic angiosperms (e.g., Hansen et al., 2007) with sizes (for complete genomes) ranging from 14,832 bp in *T. panamensis* to 22,294 bp in *H. exannulata* (compared to 14,060 bp to 22,294 bp across Thismiaceae as a whole), and the number of unique genes ranging from 12–29 (compared to 10–29 in Thismiaceae; see Table A5). Thus, the complete plastid genomes are roughly 10X smaller than photosynthetic Thismiaceae relatives in *Tacca* (i.e., 163,007–162,477 bp).

The number of unique protein-coding, tRNA, and rDNA genes present in the four newly assembled Thismiaceae complete plastid genomes range from 6–18 protein-coding genes, 2–7 tRNA genes, and four rDNA genes (compared to 6–18 protein-coding genes, 1–7 tRNA genes, and 2–4 rDNA genes across Thismiaceae as a whole; Table A5). This compares to 112 unique genes annotated in the autotrophic *Tacca leontopetaloides* plastid genome (i.e., 78 protein-coding genes, 30 tRNA genes, and four rDNA genes). Based on tRNAscan-SE predictions, some tRNA genes are identified as putative pseudogenes in individual or multiple taxa (Table A4; but see below on dual functionality of *trnE-UUC* genes). *Haplothismia exannulata* has a larger plastid genome and substantially higher gene content than all other Thismiaceae species, with 11 genes or pseudogenes found only in it and no other members of Thismiaceae (Table A5).

***3.3 Evolution of plastid genome structure in Thismiaceae***—*Haplothismia exannulata* is substantially reduced compared to autotrophic relatives and displays a considerable number of

rearrangements in gene order according to the Mauve analyses (Fig. A11). The rest of Thismiaceae is even more reduced, and further rearrangements include a series of inversions leading to frequent differences in gene order between species (Fig. 3, Fig. A12). Most taxa have a mix of genes located on both strands of the genome, but a peculiar feature in *T. panamensis* is that all of its genes (except for 3'-*rps12*) are transcribed in the same direction (the forward strand; Fig. A7).

When retained, the IR boundaries are shifted among taxa, but remain in a relatively consistent location (shaded boxes in Figs. 2, 3, A12 indicate IR boundaries, where known). Among the taxa newly assembled here, inverted repeats (IRs) are present in *T. javanica* (Fig. A6), *T. rodwayi* (Fig. A8), but absent in *H. exannulata* (Fig. A5), and *T. panamensis* (Fig. A7) (IRs are present in all other completed Thismiaceae plastid genomes; Fig. 2). In the partial plastid genome of *T. huangii* (Fig. A9), *rrn23* (the 23S rDNA gene) is present twice in separate fragments, one of which is cut off by the end of the partial sequence (see also comparisons to the *rrn*23 locus of *Tacca chantrieri* in the figure), and so it may or may not be a complete gene in the full plastid genome. It may also flag the borders of unconfirmed inverted repeats around a cryptic small single copy region. However, the incomplete genome for this taxon prevents us from confirming the presence of an IR in it. The inverted repeats across all completed Thismiaceae genomes range substantially in size, from 4,782 bp per repeat in *T. javanica* to 6,816 bp in *T. rodwayi*. For the four newly assembled complete genomes, the LSC ranges from 4,921–5,835 bp and the SSC from 1–1,825 bp (LSC from 4,439–7,799 bp and SSC from 1 bp–2,336 bp across Thismiaceae as a whole). *Thismia rodwayi* has a small single copy (SSC) region that is only a single base pair long; in addition, the two copies of *rps3* at the edge of its IRs self-overlap for 77 amino-acid residues (Figs. A8, A13). In effect, the open reading frame of this gene begins in each inverted repeat (IR), continues through the single nucleotide SSC, into the reverse strand of the other IR (Figs. A8, A13). Despite this, the gene still aligns well with the other Thismiaceae species, aside from two indels (a 21-residue and an 18-residue indel) spanning part of the reverse complement region (Fig. A14).

Several taxa have unusual gene arrangements. In *T. rodwayi*, an *rpl23* pseudogene is split into two fragments, with *trnE-UUC* apparently translocated between them (Fig. A8). Additionally, *trnfM-CAU* is located within the *rpl2* intron for *T. javanica*, *T. rodwayi* and *T. tentaculata* (Fig. A15). There are also several cases across Thismiaceae of great reduction or loss

of intergenic regions or introns between different genes, compared to photosynthetic relatives. In *Thismia tentaculata*, *rps2* and *rrn5* have an 18 bp overlap due to an apparently shifted stop codon in *rps2* (Fig. A10). In *T. javanica rps4* and *trnfM-CAU* are located on the inverted repeat boundary and have no intergenic space between them (Fig A6; note that a portion of the *trnfM-CAU* gene is also in the *rpl2* intron in both full and partial copies of the latter gene in the inverted repeat of this taxon, see above). There is also no intergenic space between *rrn23* and *rrn4.5* in *T. panamensis* (Fig. A7), or *rrn5* and *rps2* in *T. rodwayi* (note they are on opposite strands, Fig. A8). In *T. panamensis* there is only a 3 bp intergenic spacer region between *rps19* and *rpl2* (Fig. A7). In contrast, gene 'deserts' (here defined as intergenic regions at least 1 kb long) also appear to be more prevalent than in autotrophic relatives. In particular, *Thismia panamensis* has a 1,320 bp gene desert between *rps3* and *rps19*, and a 1,121 bp gene desert between *rrn16* and *rps3* (Fig. A7), and *T. tentaculata* has a 1,145 bp gene desert between *rpl2* exon 2 and *rps8* (Fig. A10).

***3.4 Reconstruction of plastid gene and IR loss***—An additional aspect of genome structural changes in Thismiaceae is gene loss. Ancestral-state reconstructions of inverted repeats, gene losses, and putative pseudogenization in Thismiaceae show mass shared losses and multiple parallel losses (Fig. 4). Eighty-two genes present in autotrophic relatives (e.g., *Tacca*) have been lost by all members of Thismiaceae (Fig. 4). An additional eleven genes present in *Haplothismia exannulata* have been lost by all *Thismia* species (Fig. 4). The genes *rps2*, *rps3*, *rps4*, *rps8*, *rps18*, *rrn4.5*, *trnE-UUC*, and *trnfM-CAU* each have multiple separate loss events within the family (2–5 steps; Fig. 4). Inverted repeats are inferred to have been lost on two separate occasions, once in the terminal lineage leading to *Haplothismia exannulata*, and once for *Thismia panamensis* (Fig. 4, see also A5, A10).

***3.5 Group IIA introns***— No Thismiaceae species appear to have retained a functional *matK* locus (*H. exannulata* has a putative pseudogene, Fig. A5; the others are all missing the gene), and none have retained the intron-containing genes *atpF*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*, or *trnV-UAC* (Table A5, A6). Only *H. exannulata* retains *clpP* within Thismiaceae, and it has lost its group IIA intron between exon 2 and exon 3 of *clpP* but retained its group IIB intron between exons 1 and 2 (Fig. A5). However, two group IIA intron-containing genes (*rpl2* and 3'-*rps12*) are retained in most Thismiaceae taxa (Table A5, A6). All complete Thismiaceae species except

*T. panamensis* have an *rpl2* intron (*T. panamensis* retains the gene but has lost the intron, Fig. A7).

Normally, *5'-rps12* intron 1b, *3'-rps12* exon 2, *3'-rps12* intron 2, and *3'-rps12* exon 3 are located consecutively in the same orientation in flowering plants, with consecutive *5'-rps12* intron 1a, and *5'-rps12* exon 1 located separately (as it is trans-spliced, e.g., in Fig. A5). However, in *T. javanica* and *T. huangii* the two exons of 3'-*rps12* (i.e., usually neighbouring exons 2 and 3 in the gene) are distantly separated and thus in non-canonical locations relative to each other (Figs. A6, and A9). There are sixteen other examples of exon 2 and exon 3 being separated in other Thismiaceae (see footnotes in Table A5). In *T. rodwayi* 3'-*rps12* exons 2 and 3 are both present consecutively but are in a non-canonical order for the orientation of the reading frames, with exon 3 being transcribed before exon 2, assuming both are required for producing a functional gene product (Fig. A8). Under the assumption that the *rps12* genes retained in these species with non-canonical exon orders are functional and that all three exons are required, additional trans-splicing would be necessary for a mature mRNA to be translated with the full reading frame.

***3.6 Selection pressure in Thismiaceae plastid genes***—In Thismiaceae there are 42 instances where the difference between a *Thismia* terminal (foreground branch) and autotrophic taxa (background branches), Δω, is negative, which is consistent with a trend of strengthening of purifying selection. In contrast, there are 36 cases with a positive trend in Δω, consistent with relaxation of purifying selection (Table 1; Table A7). *Haplothismia exannulata* and *T. tentaculata* have mainly negative Δω trends, whereas *T. hawkesii*, and *T. huangii*, and *T. panamensis* have mainly positive Δω trends. Considering individual genes, *rps4* and *rps2* have mainly negative Δω trends, and *rps3*, *rps8*, and *rps12* have mainly positive Δω trends (Table 1; Table A7). However, significant Δω values (changes in selection) were only identified in *clpP* for *Haplothismia exannulata, rps2* for *Thismia rodwayi, accD* for *T. panamensis, rpl2* for *T. panamensis,* and *rps4* for *T. panamensis* (Table 1; Table A7).

**Table 1.** Variation in Δω (difference in dN/dS, or ω, between foreground and background branches), considering individual lineages of Thismiaceae as the foreground branch (with other Thismiaceae deleted); the background includes all other taxa (autotrophic Dioscoreales). The foreground and background ω-values are all under 1.0 (see Table A7). Positive Δω values are consistent with relaxation of purifying selection, negative with strengthening. Statistically significant differences are indicated in **bold** (alpha value adjusted using the Benjamini-Hochberg correction for multiple tests).

| | *Haplothismia exannulata* | *Thismia filiformis* | *Thismia hawkesii* | *Thismia huangii* | *Thismia javanica* | *Thismia mucronata* | *Thismia panamensis* | *Thismia rodwayi* | *Thismia tentaculata* | *Thismia thaithongiana* |
|---|---|---|---|---|---|---|---|---|---|---|
| *accD* | 0.0582 | -0.0082 | -0.0222 | 0.0666 | -0.0044 | -0.0116 | **0.1709** | 0.0308 | -0.0367 | 0.0906 |
| *clpP* | **-0.1827** | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| *infA* | -0.0582 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| *matK* | -0.106 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| *rpl14* | -0.0152 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| *rpl16* | -0.0387 | -- | -- | -- | -- | -- | -0.0114 | -- | -- | -- |
| *rpl2* | 0.0057 | -0.0189 | -0.0913 | 0.0061 | -0.0436 | -0.0734 | **0.2003** | -0.0916 | -0.0376 | 0.0051 |
| *rpl20* | -0.1768 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| *rps11* | -0.0172 | -- | -- | -- | -- | -- | 0.0058 | -- | -- | -- |
| *rps12* | 0.0282 | 0.0315 | 0.0209 | -0.1018 | 0.0054 | 0.0229 | 0.0241 | -0.0036 | 0.0144 | -0.0243 |
| *rps14* | 0.035 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| *rps18* | -0.1098 | -- | 0.0424 | 0.018 | 0.0445 | -0.0006 | -- | 0.0881 | -0.006 | -0.1913 |
| *rps19* | -0.0327 | -- | -- | -- | -- | -- | 0.1006 | -- | -- | -- |

| | *Haplothismia exannulata* | *Thismia filiformis* | *Thismia hawkesii* | *Thismia huangii* | *Thismia javanica* | *Thismia mucronata* | *Thismia panamensis* | *Thismia rodwayi* | *Thismia tentaculata* | *Thismia thaithongiana* |
|---|---|---|---|---|---|---|---|---|---|---|
| *rps2* | -0.1626 | -0.1469 | -- | -- | -- | -0.091 | -- | **-0.1653** | -0.0591 | -- |
| *rps3* | -0.005 | -- | 0.0277 | -- | 0.0445 | 0.0596 | 0.0394 | 0.0387 | -- | 0.0712 |
| *rps4* | -0.0326 | -0.0485 | 0.0112 | -- | -0.1212 | -0.1854 | **0.1495** | -- | -0.0304 | -0.1631 |
| *rps7* | -0.0331 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| *rps8* | -0.0488 | 0.0723 | 0.0979 | -- | 0.1166 | 0.0041 | -- | -- | -0.0041 | 0.193 |

## CHAPTER 4: DISCUSSION

***4.1 Effect of long outgroup branches on estimates of Dioscoreales phylogeny***—Strikingly different Thismiaceae relationships are inferred when two taxa in Burmanniaceae (*Apteria* and *Gymnosiphon*) are included vs. excluded from analysis, with the bulk of the family (except the relatively slowly evolving *Haplothismia*) shifting to Burmanniaceae when *Apteria* and *Gymnosiphon* are retained (Fig. A3). In contrast, *Thismia* remains together with *Haplothismia* (representing a monophyletic Thismiaceae at the current taxon sampling) when the long-branch Burmanniaceae are excluded (Fig. 1, and cf. Figs. A1, A2 to Figs. A3, A4), and Thismiaceae is then the sister group of Taccaceae. We interpret this to be a clear case of long-branch attraction between *Thismia* and the two long-branch Burmanniaceae taxa (*Apteria* and *Gymnosiphon*) that is readily circumventable. Long-branch attraction (LBA) can lead to strikingly erroneous placements due to numerous parallel substitutions being mistaken for synapomorphies in parsimony analyses (Felsenstein, 1978; Hendy and Penny, 1989). Likelihood analyses and other model-based methods can correct for undetected multiple hits (Felsenstein, 1978; Kuhner and Felsenstein, 1994) but may also fail in more extreme long-branch cases. For example, Lam et al. (2018) successfully placed many long-branch mycoheterotrophic lineages in angiosperm phylogeny using likelihood approaches, but also demonstrated LBA for two taxa: *Epipogium* (Orchidaceae) and *Thismia* (Thismiaceae, the only representative of that family included at the time) were misleadingly pulled together. Setting aside this probable artefact, members of Thismiaceae are therefore inferred to be monophyletic at the current taxon sampling, and are phylogenetically distinct from Burmanniaceae. Because of extensive rate elevation in plastid genes, there is also potential for mis-inference of relationships within Thismiaceae. This is a general caveat for plastid-based studies of relationships in the family. Nonetheless, the removal of two long-branches in Burmanniaceae appears to minimize a strong long-branch attraction for family-level relationships in the order as a whole, and my results are consistent with recognizing Burmanniaceae and Thismiaceae as two distinct lineages in the order, conflicting with Jonker (1938), Maas et al. (1986), APG (2016), Caddick et al. (2002), but in agreement with Chase et al. (1995), Merckx et al. (2006, 2009), Woodward (2007), Merckx and Smets (2014), Cheek et al. (2018), Lam et al. (2016; 2018), and Shepeleva et al. (2020).

***4.2 Phylogenetic relationships in Dioscoreales***—The 125-taxon tree (which excludes *Apteria* and *Gymnosiphon*) has a similar topology to the nuclear/mitochondrial tree of Shepeleva et al. (2020) concerning groups labelled there as clade 1 (*T. huangii* and *T. thaithongiana* considering taxa here), clade 2 (*T. rodwayi*, considering taxa here), clade 3 (*T. kelabitiana* and *T. viridistriata* considering taxa here) and clade 4 (*T. angustimitra*, *T, mucronata*, *T. okhaensis*, and *T. puberula*, considering taxa here). These clades in their few-gene nuclear/mitochondrial tree have identical internal relationships to those found here in the 125-taxon tree, with similar relationships among them (Fig. 1). The 'Old World *Thismia'* clade found in Shepeleva et al. (2020) is also present here (Fig. 1) with 'New World' *Thismia panamensis* sister to this clade in both trees. Notably, three species (*T. hawkesii*, *T. lanternata* and *T. rodwayi*) within this clade are found in Australia/New Zealand (regions that belong to neither Old or New World geographical categories). However, a group labelled clade 5 in Shepeleva et al. (2020) (comprising *T. alba*, *T. annamensis*, *T. cornuta*, *T. filiformis*, *T. hexagona*, and *T. neptunis* considering taxa here) is also a clade in a plastid data study by Yudina et al. (2021), but is partly contradicted in the 125-taxon tree here, as *T. alba* and *T. annamensis* are instead inferred to be the sister group of a clade comprising *T. viridistriata* and *T. lanternata*, consistent with the placement of clade 3 in Shepeleva et al. (2020). In addition, *T. javanica* was recovered in a clade with *T. tentaculata*, then sister to clade 5, in Shepeleva et al. (2020), whereas *T. javanica* is recovered as sister to clade 3 and not to *T. tentaculata* here. The 125-taxon tree here (Fig. 1) has a very similar topology to the Yudina et al. (2021) plastid tree, the source of 17 of 23 Thismiaceae taxa here, with the exception of the relative placements of *T. alba* and *T. annamensis* noted above, in addition to changes in within-clade relationships in clades 3 and 5. A placement of *T. hongkongensis* as sister to *T. tentaculata* in Yudina et al. (2021) is confirmed here.

***4.3 Ancestral-state reconstructions of gene loss (and photosynthesis loss) in Thismiaceae***—Several non-terminal branches around the base of Thismiaceae are inferred to have experienced multiple gene loss or pseudogenization events (Fig. 4). Our reconstructions of these losses predict a massive set of gene losses (82 genes, with putative pseudogenization of two more, *matK*, *trnY-GUA*) along the stem lineage of Thismiaceae, representing a large set of synapomorphies for the family as a whole at current taxon sampling. A further large-scale gene loss (eleven genes; Fig. 4) is inferred in the common ancestor of all *Thismia* species, acting as a

major set of synapomorphies for this clade, effectively distinguishing *Thismia* from *Haplothismia*. Notably, two genera, *Oxygyne* and *Tiputinia*, were not sampled here and should be included in future plastid-based studies. Previous phylogenetic studies considering nuclear 18S rDNA (Yokoyama et al., 2008) alone, or in combination with mitochondrial *atp1* (Merckx and Smets, 2014), or in combination with *atp1* and nuclear ITS (nuclear ribosomal internal transcribed spacer; Shepeleva et al., 2020) placed *Oxygyne* as the sister group of the rest of Thismiaceae (Yokoyama et al., 2008; Merckx and Smets, 2014; Shepeleva et al., 2020), with *Tiputinia* sister to *Thismia panamensis*, considering the taxa sampled here (Merckx and Smets, 2014; Shepeleva et al., 2020). If *Oxygyne* is indeed sister to the rest of the family (note that *Oxygyne* in these studies is represented solely by the nuclear 18S rDNA locus), it is conceivable that it experienced different sets of gene losses, perhaps fewer, depending on where it split off from the rest of the stem lineages in Fig. 1 here.

The character reconstructions here are based on parsimony (we used Dollo-like weights to reject regain of complex traits like photosynthesis once lost, see Materials and Methods). Mycoheterotrophy has arisen more than 37 times independently in monocots, in seven families: Burmanniaceae, Corsiaceae, Iridaceae, Orchidaceae, Petrosaviaceae, Thismiaceae, and Triuridaceae (e.g., Jacquemyn and Merckx, 2019). Thismiaceae and Triuridaceae are both exclusively fully mycoheterotrophic families. The simplest interpretation of the uniform absence of photosynthetic function here—given that all species in Thismiaceae are achlorophyllous and fully mycoheterotrophic (all lack key photosynthetic genes in their plastid genomes)—is that there was a single loss of photosynthesis before the crown-clade diversification of the family. However, an alternative less-parsimonious interpretation is that photosynthesis was lost more than once in the family (as found in Burmanniaceae; Merckx et al. 2006), with subsequent extinction of nested autotrophs giving the appearance of a single loss. Convergent losses of similar sets of photosynthetic and other genes (e.g., Graham et al., 2017) would then be misinterpreted as single losses before most or all of the family diversified. Thus, while the best interpretation is that the mass gene loss at the base of the Thismiaceae clade reflects a single loss of photosynthesis in the common ancestor to Thismiaceae, the reality may be more complex. Broadly speaking, the independent loss of photosynthesis in *Afrothismia* and Thismiaceae, but not *Tacca* in the same clade (Lin, 2020), suggests this is at least a plausible possibility.

Beyond the deepest internal branches (splits) in Thismiaceae, clades supported by more than one gene loss/putative pseudogenization event include one comprising all sampled members of *Thismia* except *T. panamensis* (supported by four gene losses), and a clade comprising *T. cornuta*, *T. filiformis*, *T. gardneriana*, *T. hexagona* and *T. neptunis* (supported by one gene loss, and one putative pseudogenization event) (Fig. 4). Several terminal branches also have multiple individual gene loss or putative pseudogenization events. In some cases, the reconstructions were equivocal because of state uncertainty in individual taxa, presumably in part due to the genomes being incompletely assembled (see Table A1; note arrows in Fig. 4 indicating this uncertainty). Evidence of homoplasy in gene loss/putative pseudogenization can be seen in *rps2*, *rps3*, *rps4*, *rps8*, *rps18*, *rrn4.5*, *trnE-UUC*, and *trnfM-CAU*, which all are inferred to have experienced convergent loss or pseudogenization events (see the multiple steps noted in Fig. 4 for these characters). As a general caveat, ancestral-state reconstructions here are conditional on the tree used to infer them being correct (i.e., Fig. 1), and so the reconstructions should be interpreted cautiously, given the relatively large number of relatively poorly supported branches in the tree (Fig. 1). However, in several cases, the large number of gene loss/putative pseudogenization events on a branch adds to our confidence in those branches. An example of this is the sister-group relationship of *Haplothismia exannulata* and the remainder of the sampled taxa in the family, noted above.

***4.4 Plastid genome evolution***—Land plants have a highly conserved quadripartite structure such that almost all species have large and small single copy regions (LSC and SSC, respectively) separated by inverted repeats (IRs) (e.g. Palmer, 1985; Xu et al., 2015). Relatively minor expansions and contractions of IR boundaries (with genes entering and leaving these repeats) are quite common (Jansen and Ruhlman, 2012), and account for much of the length variation in the plastid genomes of green plants. Expansion and contraction of the IR can also lead to gene-order changes (Palmer, 1991; Jansen and Ruhlman, 2012), as seen in Chumley et al. (2006), Guisinger et al. (2011), and Fonseca and Lohmann (2017). A hypothesized function of the IR is that it contributes to plastid genome structural stability, decreasing the frequency of rearrangements when present (Palmer and Thompson, 1982; Palmer et al., 1987). However, some recent studies have argued against the correlation between IR presence and genome stability (Jansen and Ruhlman, 2012; Weng et al., 2017) with examples being found of highly rearranged plastids that

retain IRs (Knox and Palmer, 1999; Cosner et al., 2004; Lee et al., 2007), and stable plastids that lack them (e.g., Lam et al., 2015; Blazier et al., 2016). Across land plants, the IR has undergone large-scale expansion (e.g., Chumley et al., 2006; Sun et al., 2013; Joyce et al., 2018; Li et al., 2020; Darshetkar et al., 2021), large-scale contraction (e.g., Guisinger et al., 2011; Naumann et al., 2016), and has been lost completely on several occasions (see below). Novel regain of an IR has even been found in *Monotropa uniflora* (Ericaceae) (Braukmann et al., 2017), *Medicago minima* (Fabaceae) (Choi et al., 2019), and *Parasitaxus usta* (Podocarpaceae) (Lam, 2016; Qu et al., 2019).

The IR has been lost (only one copy retained) at least six times in autotrophs: in Pinaceae (e.g., Raubeson and Jansen, 1992), *Erodium* species (Geraniaceae) (Guisinger et al., 2011; Blazier et al., 2016), the inverted repeat-lacking clade (IRLC) in legumes (Fabaceae) (e.g. Sabir et al., 2014), the saguaro *Carnegia gigantea* (Cactaceae) (Sanderson et al., 2015), the putranjivoid clade (Lophopyxidaceae and Putranjivaceae) (Jin et al., 2020), and in *Passiflora capsularis* and *P. costaricensis* (Passifloraceae) (Cauz-Santos et al., 2020). Loss of one copy of the IR may be more common in heterotrophic (mycoheterotrophic or holoparasitic) lineages with degraded plastid genomes. Six independent examples include holoparasitic *Phelipanche ramosa* and *Conopholis americana* (Orobanchaceae; Wicke et al., 2013), mycoheterotrophic *Sciaphila* (Triuridaceae) (Lam et al., 2015; Petersen et al., 2018), holoparasitic *Rhopalocnemis phalloides* (Balanophoraceae; Schelkunov et al., 2018), holoparasitic *Cytinis hypocystis* (Cytinaceae; Roquet et al., 2016), all mycoheterotrophic Ericaceae except *Monotropa uniflora* (Braukmann et al., 2017), and mycoheterotrophic *Epirixanthes pallida* and *E. elongata* (Polygalaceae) (Petersen et al., 2019). In Thismiaceae, I infer two independent losses of the plastid inverted repeat (IR), once in *Haplothismia exannulata*, and once in *Thismia panamensis* (Fig. 4). In the absence of a phylogenetic tree, shared IR loss would tend to support *H. exannulata* and *T. panamensis* as a clade. However, the plastid genome tree inferred here (Fig. 1) and second large inferred mass gene loss (Fig. 4) support the arrangement with *H. exannulata* and then *T. panamensis* being successive sister groups to the rest of Thismiaceae, implying two independent IR losses in the family.

Several land plants have highly reduced small single copy regions, but none are as short as the single-base SSC observed in *T. rodwayi* here (Figs. 2; A8). Considering autotrophic taxa, the magnoliid *Asarum delavayi* (Sinn et al., 2018), eudicot *Pelargonium hortorum* (Chumley et

al., 2006) and diatom *Climaconeis scalaris* (Gastineau et al., 2021) have IRs that have expanded to 48,220 bp, 75,741 bp, and 79, 040 bp, respectively, with their SSC regions correspondingly contracted to 14 bp, 6750 bp, and 304 bp. In mycoheterotrophs *Exacum paucisquama* (Li et al., 2020) and *Geosiris australiensis* (Joyce et al., 2018) the total amount of the plastid genome sequence taken up by IRs has expanded to 17,622 bp and 36,347 bp, respectively, and the SSC has in turn contracted to 2133 bp and 515 bp. The 14 bp SSC of *A. delavayi*, the 304 bp SSC of *C. scalaris* (Gastineau et al., 2021) and the 515 bp of *G. australiensis* (Joyce et al., 2018) are all larger than the 1 bp SSC seen in *Thismia rodwayi*, which comprises only 0.005% of its total genome. *Thismia rodwayi* therefore has the most extreme case of SSC contraction found to date.

Land-plant plastid genomes are in general highly conserved in terms of structure and gene order, and rearrangements and translocation of genes are generally very rare in green plants (Palmer, 1991; Jansen and Ruhlman, 2012). The most common mechanism for plastid genome rearrangement is inversion (e.g., Palmer, 1991; Kim et al., 2005; Lee et al., 2007; Jansen and Ruhlman, 2012). Multiple partially overlapping inversions can effectively result in translocation of genes (Chumley et al., 2006; Lee et al., 2007). We visualized the structural rearrangements using Mauve (Darling et al., 2004) and found that while multiple inversions are present, there is general consistency in locally colinear blocks (LCBs) order and IR placement among Thismiaceae species (Fig. 3, Fig. A12). As a caveat to the interpretations shown here, the circular nature of plastid genomes means that genes with slight positional shifts in the region of the artificial cut we made to linearize genomes may appear to represent a more substantial rearrangement (by appearing to jump from one end of the linearized genome to the other). I attempted to minimize this effect by careful selection of the artificial cut location. Another issue is that in divergent plastid genomes, gaps in estimates of locally colinear blocks (LCBs) may occur even in areas where shared genes are present (e.g., *trnfM-CAU* in *Haplothismia exannulata* Fig. 3). Mauve is not a perfect program for handling circular (or repetitive linear) divergent plastid genomes (as seen here), but provides a useful overview of genome rearrangements.

***4.5 Group IIA introns***—Heterotrophic lineages lacking *matK* have repeatedly been found to retain the group IIA introns from *rpl2* and *3'-rps12* (Table A6). For example, Graham et al. (2017) noted six instances of *rpl2* intron retention and two instances of *3'-rps12* intron retention across eight independent holoparasitic and mycoheterotrophic taxa that lack *matK.* Here, sixteen

Thismiaceae species retained the *rpl2* intron and all Thismiaceae species sampled with complete plastid genomes have both exons of 3'-*rps12.* However, it is unclear if the *3'-rps12* intron is retained in Thismiaceae species other than *Haplothismia exannulata* and *Thismia panamensis* (see text above, Table A5). The majority of *rpl*2 and *rps*12 genes (considering the full set of exons if fused) have open reading frames (Fig. A16 for *rps12* genes). The 5'-end motifs (i.e., 5'-GTGYG) and 3'-end motifs (i.e., 5'-AY) typical of group IIA introns (Jacquier and Michel, 1987) are largely conserved in *rpl2* (at most 1 nucleotide different; Fig. A15). For 3'-*rps12* *Thismia rodwayi*, *T. tentaculata*, and *T. huangii* there are no splice sites between exons 2 and 3 that match the standard 5'-GTGYG- and 3'-AY- Group IIA motifs (Jacquier and Michel, 1987). The matching 3'-*rps12* splice sites for other species range from a perfect match to three bases that differ from the group IIA consensus sequence (Jacquier and Michel, 1987; Fig. A17). Despite the *rps12* exon rearrangements and potential splicing issues, all five newly assembled Thismiaceae species maintain all three exons (5'-*rps12* and both 3'-*rps12* exons), and the exons (if fused) are predicted to represent a complete open reading frame (Fig. A16). As discussed above, additional trans-splicing of *rps12* would be required considering the sometimes non-canonical order of exons, if all three exons are needed for functionality. Thus, the repeated retention of *rps12* exons 1–3, when not retained in canonical order—and without retention of functional *matK*—could indicate alternative splicing mechanisms for the last two exons of the *rps*12 gene. This merits further investigation.

***4.6 trnE-UUC and haem biosynthesis***—In models of plastid genome degradation, *trnE* is one of the five core non-bioenergetic plastid genes (that also include *accD*, *clpP*, *ycf1*, and *ycf2*), and is predicted to be one of the last genes to be retained in highly modified heterotroph plastid genomes—indeed, it has been retained in members of almost every heterotrophic lineage (Barbrook et al., 2006; Graham et al., 2017). However, the endo-holoparasites *Pilostyles aethiopica* (Apodanthaceae), *P. hamiltonii* (Apodanthaceae), and *Rafflesia lagascae* (Rafflesiaceae) have plastid genomes completely lacking *trnE* genes (or in the case of *R. lagascae* potentially no plastid genome) (Molina et al., 2014; Bellot and Renner, 2015). In addition, holoparasitic *Balanophora reflexa* and *B. laxiflora* (Balanophoraceae) plastid genomes have pseudogene copies of *trnE-UUC* with respect to protein synthesis, but both possibly still function in haem biosynthesis (see below and Su et al., 2019).

All Thismiaceae species assembled here have retained some version of *trnE-UUC*, although some have divergent sequences and/or have lost their anticodon (Table A4). Using tRNAscan-SE, we predict multiple losses of *trnE-UUC* function as a translation apparatus gene in *T. panamensis*, *T. hawkesii* and *T. huangii*. However, *trnE-UUC* is known to have a secondary function in haem biosynthesis (Randau et al., 2004; Lüer et al., 2007; Layer et al., 2010). 5-aminolevulinic acid (ALA) is a precursor in the haem biosynthesis pathway and in plants the C5-pathway synthesizes ALA from tRNA-bound glutamate (Layer et al., 2010). Glutamyl-tRNA reductase (GluTR) binds to tRNA-Glu and converts it in an NADPH-dependent reaction to the intermediate glutamate-1-semialdehyde step in the ALA pathway (Layer et al., 2010). Thus, the degraded Thismiaceae *trnE* genes may still function in haem biosynthesis but not be functional in translation (Table A4). Although older studies identify a 3'-terminal CCA as being necessary for tRNA-Glu recognition, based on shared sequence between species (Willows et al., 1995) and digestion of the 3'-terminal end (Schon et al., 1986), Randau et al. (2004) found no identity elements for GluTR in the acceptor stem where CCA is located. Neither the anticodon nor acceptor stem are thought to be major identity elements for aminoacyl-tRNA synthetases (Lüer et al., 2007), and instead the tRNA substrate appears to be recognized by its overall shape (Layer et al., 2010), and so recognition may not be particularly strict (Lüer et al., 2007). Therefore, it is difficult to predict from sequence data alone whether *trnE-UUC* may retain function for haem biosynthesis, even when non-functional in primary function. However, the retention of the gene across the family, in the context of many other genes having been completely lost, is consistent with this gene retaining its functionality in haem biosynthesis in some or all of these lineages—even though it is highly modified and/or has lost the anticodon in multiple independent lineages of *Thismia*. Biochemical analysis of *trnE-UUC* and its binding to GluTR is needed to further address this prediction.

***4.7 Selection pressure in Thismiaceae plastid genes***—Despite the degradation process, retained genes in mycoheterotrophic plastid genome genes are commonly still under strong purifying pressure. For example, in a comparison of dN/dS (or ω) values, Logacheva et al. (2011) found that selection pressure was largely not relaxed in the mycoheterotrophic *Neottia nidus-avis* (Orchidaceae), with only one gene (*rpl23*) out of 23 tested showing a relaxation of purifying selection. Schelkunov et al. (2015) found that selection pressure was not relaxed in the

mycoheterotrophic *Epipogium* clade (Orchidaceae), and Lam et al. (2015) found that fifteen out of eighteen genes tested were under purifying selection comparable to green outgroups in mycoheterotrophic *Sciaphila densiflora* (Triuridaceae), with only *rps7* showing ω values consistent with neutrality, and *clpP* and *rpl14* having values consistent with weakening purifying selection. Joyce et al. (2018) found that a majority of retained plastid genes did not have a significant change in selection pressure for mycoheterotrophic *Geosiris* (Iridaceae), although *rps3* in *Geosiris aphylla* and *clpP* in *G. austaliensis* showed values consistent with relaxation of purifying selection. In mycoheterotrophic liverwort *Aneura mirabilis,* significant relaxation of purifying selection was seen in nine of 19 retained genes with strengthening of purifying selection in one (*rbcL*) (Bell et al., 2020). In mycoheterotrophic *Corallorhiza striata* (Orchidaceae), relaxed purifying selection was inferred for *ycf1* and *ycf2*, out of 41 genes tested (Barrett and Davis, 2012). Finally, Yudina et al. (2021) examined selection operating on retained genes in Thismiaceae species, considered all taxa simultaneously rather than breaking their tests down by taxon as was done here, and found some indications for relaxation of purifying selection in the inferred mycoheterotroph-specific ω values. It is worth noting that ω tests have a high sensitivity to alignment issues and should be interpreted with caution when applied to divergent and often difficult to align mycoheterotrophic species (e.g., Lam et al. 2015)—including those here, although we identified and remove difficult-to-align regions before analysis. I found that significant changes to purifying selection strength are relatively small in Thismiaceae, with only five of the 78 gene copies tested showing significant results across different genes (Table 1). For the non-significant cases, examples of strengthening and relaxation of ω values were fairly balanced (Table 1). *Thismia panamensis* had three of the five statistically significant Δω values (all indicating weakening of purifying selection) possibly indicating an elevated trend to change in selection pressure in this species, compared to the rest of Thismiaceae (Table 1).

***4.8 Conclusions***—My thesis increases sampling of Thismiaceae plastid genomes with the assembly of five new plastid genomes, including one previously unsampled genus (*Haplothismia*), although the plastid genomes of two small genera, *Oxygyne* and *Tiputinia*, remain to be sampled, as well as a broader spectrum of the ~80 recognized species of *Thismia* (Yudina et al. 2021). We infer an updated phylogeny for Thismiaceae based on plastid genome

data, and uncovered numerous modifications to their plastid genomes that are rarely seen in autotrophic plants, or even other heterotrophs. Thismiaceae form a clade supported by mass loss of genes that represent synapomorphies for the family. I also infer this clade to be distantly related to Burmanniaceae in Dioscoreales, supporting its recognition as a distinct family, and pointing to a need to update current angiosperm classification. Plastid genomes in Thismiaceae have all undergone massive reduction with scattered inversion events, and separate losses of the inverted repeat (IR) on two occasions. The reduced set of retained genes in Thismiaceae largely remain under strong purifying selection, with a significant weakening seen in only three genes of the 78 copies tested. Other examples of unusual molecular evolution in Thismiaceae include *trnE-UUC* losing function in protein synthesis while possibly retaining its secondary function in haem biosynthesis in multiple species (including loss of the anticodon in one case), non-canonical order of *rps12* exons that may imply additional trans-splicing, and a drastically reduced single base-pair small single copy (SSC) region and associated self-overlapping *rps3* gene in *Thismia rodwayi*. Further research into the function of *trnE-UUC* in Thismiaceae and splicing mechanics of *rps12* will enhance understanding of these unusual molecular evolutionary phenomena.

These numerous modifications support the utility of Thismiaceae as a model system for understanding the genomic changes that can follow the loss of photosynthesis and adaptation to a mycoheterotrophic lifestyle.

## FIGURES

**Figure 1.** Phylogenetic relationships in Thismiaceae and relatives (Dioscoreales) inferred from a DNA-based maximum likelihood (ML) analysis of 82 plastid genes (full tree in Fig. A1). Bootstrap values (ML analysis partitioned by genes and codons) beside branches; thick lines indicate 100% bootstrap support. Asterisks highlight focus species.

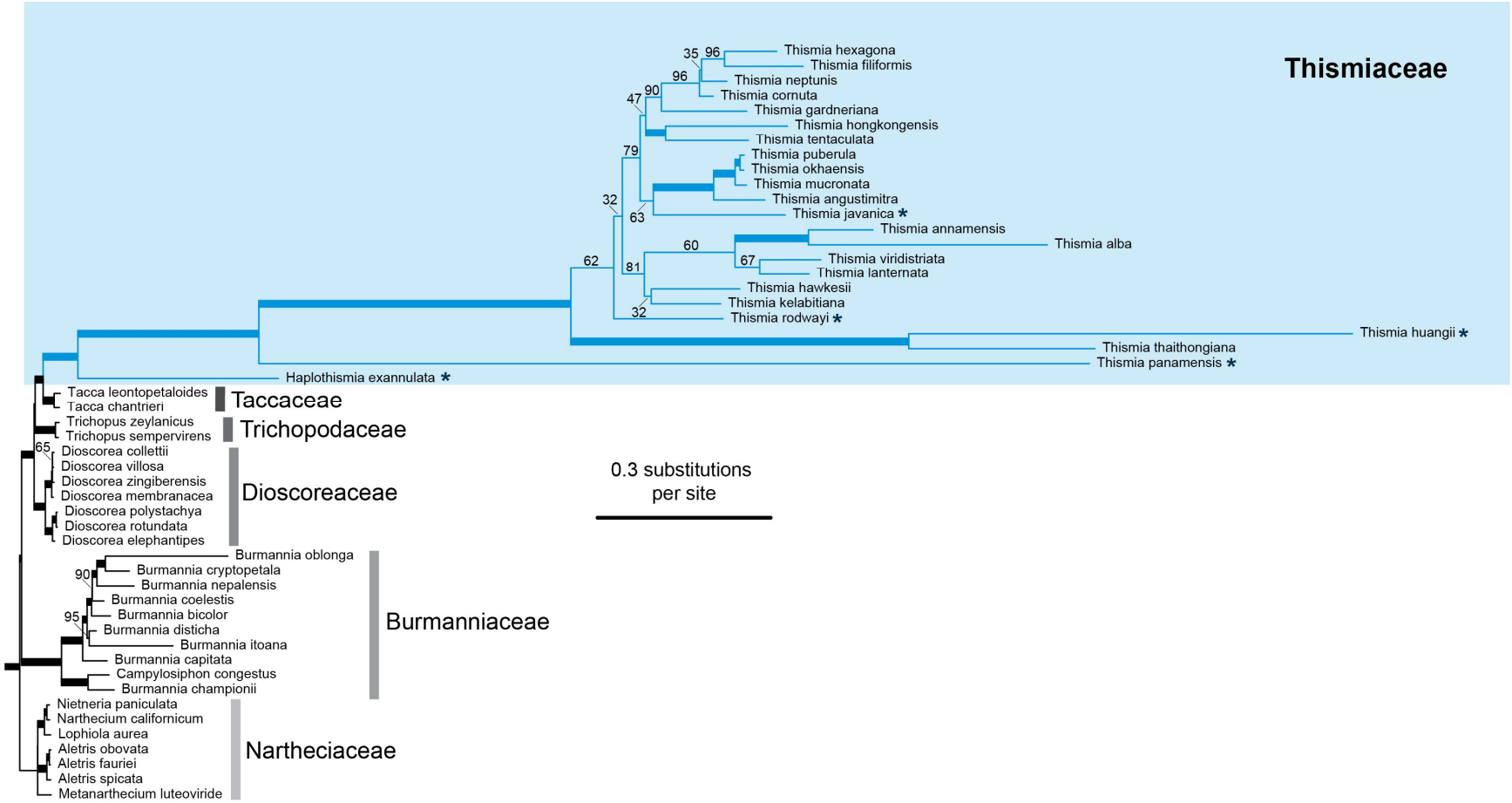


**Figure 2.** Linearized plastid genomes of newly sequenced members of Thismiaceae (*Haplothismia exannulata*, *Thismia rodwayi*, *T. javanica*, *T. panamensis* and *T. huangii*). Circularized versions of the first four genomes are shown in Figs. A5–A8. *Thismia huangii* is incomplete at each end (Fig. A9). Grey boxes indicate inverted repeat (IR) regions, present in a subset of taxa, which separate large and small single copy regions (LSC, SSC) in them (note the single base SSC in *T. rodwayi*; dashed line); the hatched box in *T. huangii* indicates a possible inverted repeat (maximum extent unclear). White boxes indicate introns, Ψ-symbols putative pseudogenes, and genes are coloured by major functional class (noted in the caption).

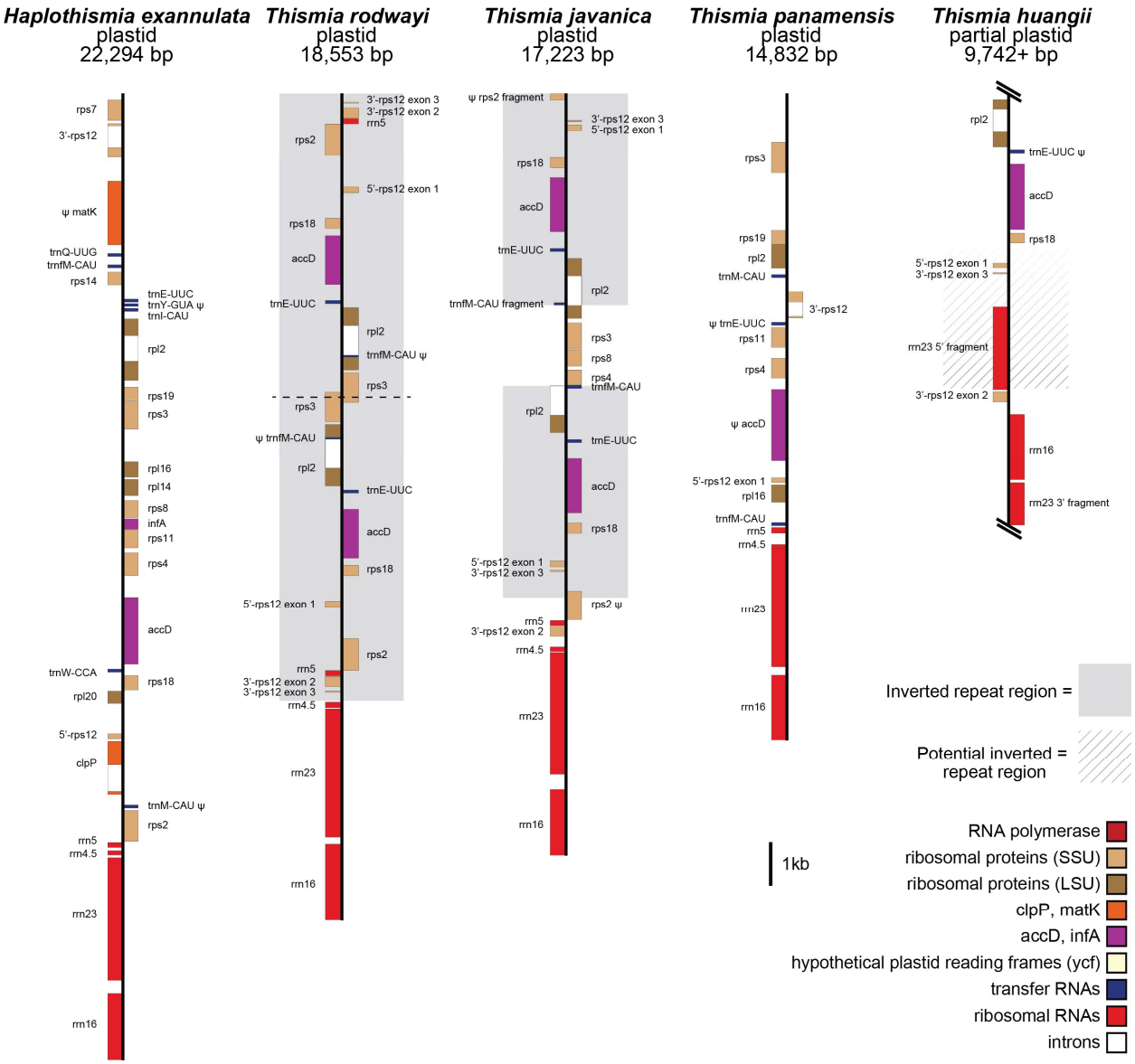

Haplothismia exannulata
plastid
22,294 bp
Thismia rodwayi
plastid
18,553 bp
Thismia javanica
plastid
17,223 bp
Thismia panamensis
plastid
14,832 bp
Thismia huangii
partial plastid
9,742+ bp
rps7
3'-rps12
ψ matK
trnQ-UUG
trnfM-CAU
rps14
trnE-UUC
trnY-GUA ψ
trnI-CAU
rpl2
rps19
rps3
rpl16
rpl14
rps8
infA
rps11
rps4
accD
trnW-CCA
rps18
rpl20
5'-rps12
clpP
trnM-CAU ψ
rps2
rrn5
rrn4.5
rrn23
rrn16
3'-rps12 exon 3
3'-rps12 exon 2
rrn5
rps2
5'-rps12 exon 1
rps18
accD
trnE-UUC
rpl2
trnfM-CAU ψ
rps3
rps3
ψ trnfM-CAU
rpl2
trnE-UUC
accD
rps18
5'-rps12 exon 1
rps2
rrn5
3'-rps12 exon 2
3'-rps12 exon 3
rrn4.5
rrn23
rrn16
ψ rps2 fragment
3'-rps12 exon 3
5'-rps12 exon 1
rps18
accD
trnE-UUC
rpl2
trnfM-CAU fragment
rps3
rps8
rps4
trnfM-CAU
rpl2
trnE-UUC
accD
rps18
5'-rps12 exon 1
3'-rps12 exon 3
rps2 ψ
rrn5
3'-rps12 exon 2
rrn4.5
rrn23
rrn16
rps3
rps19
rpl2
trnM-CAU
3'-rps12
ψ trnE-UUC
rps11
rps4
ψ accD
5'-rps12 exon 1
rpl16
trnfM-CAU
rrn5
rrn4.5
rrn23
rrn16
rpl2
trnE-UUC ψ
accD
rps18
5'-rps12 exon 1
3'-rps12 exon 3
rrn23 5' fragment
3'-rps12 exon 2
rrn16
rrn23 3' fragment
Inverted repeat region =
Potential inverted =
repeat region
1kb
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
accD, infA
hypothetical plastid reading frames (ycf)
transfer RNAs
ribosomal RNAs
introns

**Figure 3.** Mauve alignment comparing locally colinear sequences within *Haplothismia exannulata*, *Thismia panamensis*, *T. thaithongiana*, *T. huangii*, *T. rodwayi*, *T. javanica* and *T. tentaculata*. These species represent the major rearrangements within the *Thismia* species studied (refer to Fig. A12 for comparison of locally colinear sequences in other available *Thismia* species). A linear gene map of *H. exannulata* appears first for reference. A single copy of the inverted repeat regions for *Thismia thaithongiana* and *T. rodwayi* were included in this comparison and are represented by the grey boxes over portions of the strings. Coloured blocks represent 'locally colinear blocks' (LCBs) which have shared gene order between genomes. Note that parameters in Mauve for detecting LCBs do not exhaustively detect all shared sequence, which is why shorter or more divergent genes shared between species may not appear in an LCB (as in the case of *trnM-CAU*). LCBs appearing above the central line for the *Thismias* are colinear and in the same orientation as the *H. exannulata* reference sequence; those below are in the reverse complement relative to the reference. Coloured lines link orthologous LCBs shared between taxa. Large blank regions within an LCB and stretches between LCBs represent lineage-specific sequences. The double diagonal slashes indicate incomplete plastid genomes.

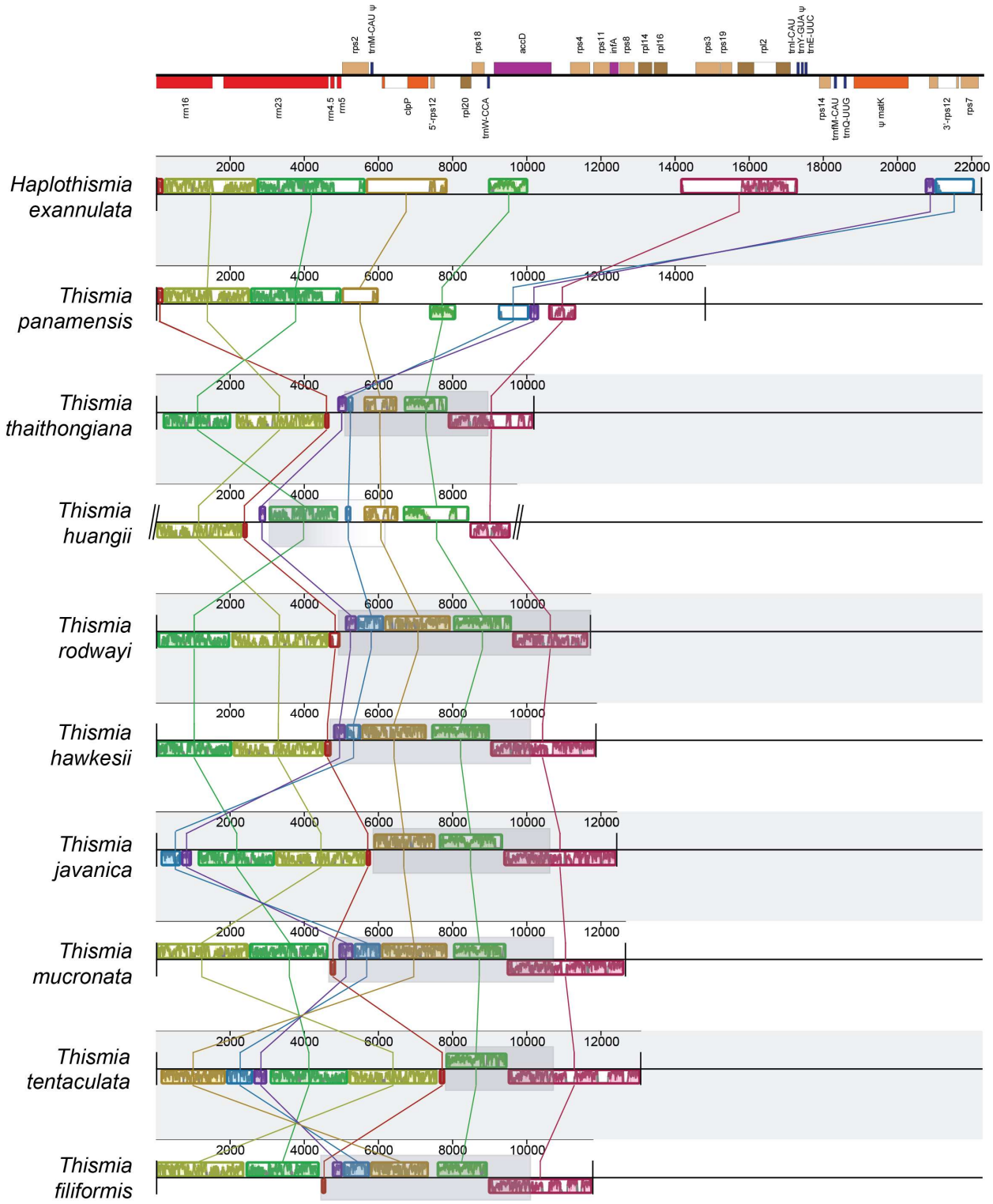

rps2
trnM-CAU ψ
rps18
accD
rps4
rps11
infA
rps8
rpl14
rpl16
rps3
rps19
rpl2
trnI-CAU
trnY-GUA ψ
trnE-UUC
rrn16
rrn23
rrn4.5
rrn5
clpP
5'-rps12
rpl20
trnW-CCA
rps14
trnfM-CAU
trnQ-UUG
ψ matK
3'-rps12
rps7
Haplothismia exannulata
Thismia panamensis
Thismia thaithongiana
Thismia huangii
Thismia rodwayi
Thismia hawkesii
Thismia javanica
Thismia mucronata
Thismia tentaculata
Thismia filiformis

**Figure 4**. Reconstruction of inverted repeat losses, gene losses and putative pseudogenization events in Thismiaceae based on the plastid-based phylogenetic tree (Fig. 1). Ancestral-state reconstructions performed with Dollo parsimony (weighting against reversion of gene loss or pseudogenization, see text); the inferred number of steps noted for each gene (or IR loss) when greater than one. The α and β characters indicate mass losses of several genes, as noted in the caption. Gene loss ticks are red and pseudogenization ticks are blue. The *trnE-UUC* locus may retain a secondary function in haem biosynthesis despite the loss of its anti-codon (putative pseudogenization, marked as "?" in the figure). Several equivocal ancestral state reconstructions are noted on the most recent branch where the state is inferred to be unambiguous; earlier losses or pseudogenization events are possible in ancestors of relatives with missing data (indicated by an arrowhead beside the gene)

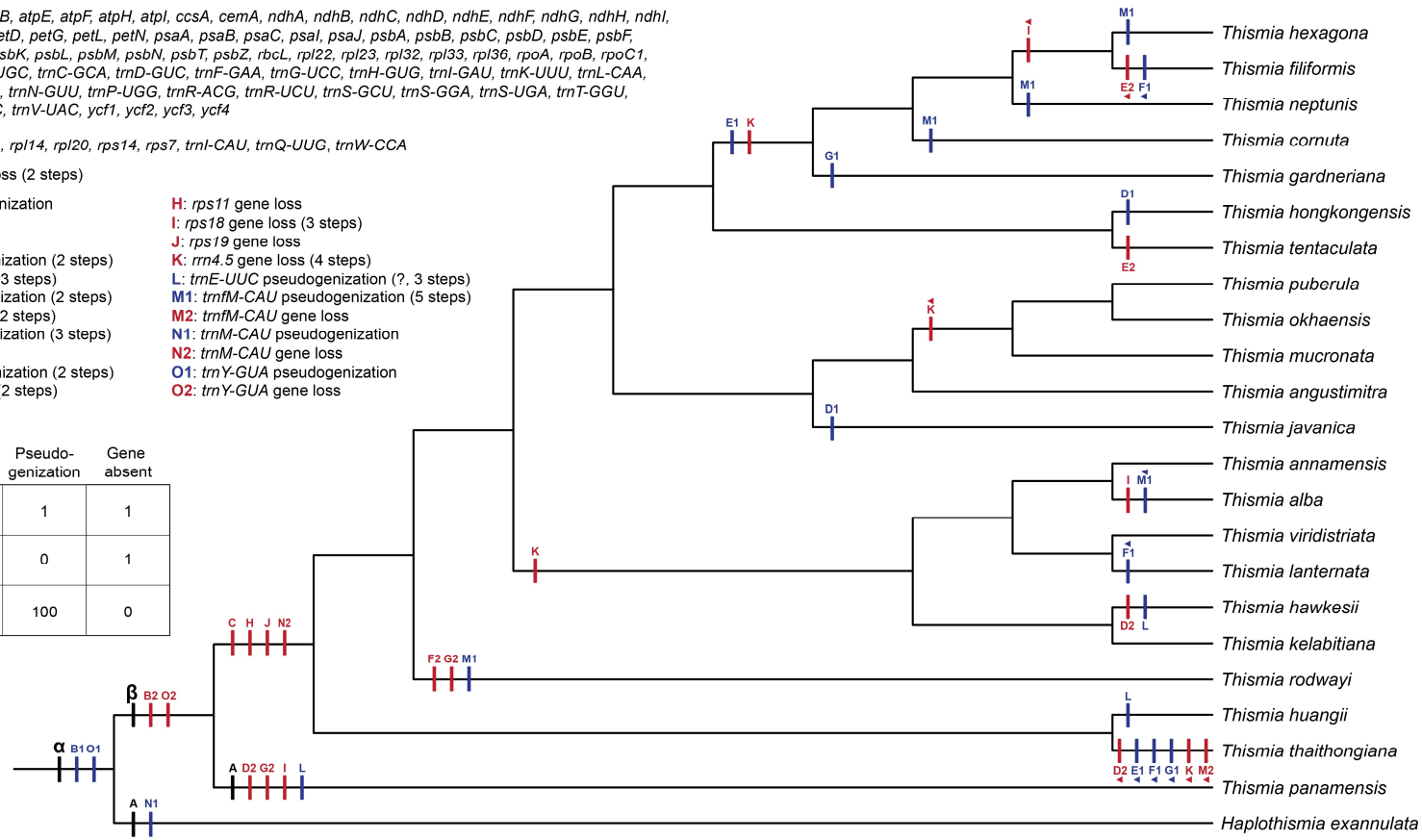
**α**: Loss of *atpA, atpB, atpE, atpF, atpH, atpI, ccsA, cemA, ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhK, petA, petB, petD, petG, petL, petN, psaA, psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, rbcL, rpl22, rpl23, rpl32, rpl33, rpl36, rpoA, rpoB, rpoC1, rpoC2, rps15, trnA-UGC, trnC-GCA, trnD-GUC, trnF-GAA, trnG-UCC, trnH-GUG, trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAA, trnL-UAG, trnN-GUU, trnP-UGG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnV-UAC, ycf1, ycf2, ycf3, ycf4*

**β**: Loss of *clpP, infA, rpl14, rpl20, rps14, rps7, trnI-CAU, trnQ-UUG, trnW-CCA*

**A**: inverted repeat loss (2 steps)

**B1**: *matK* pseudogenization
**B2**: *matK* gene loss
**C**: *rpl16* gene loss
**D1**: *rps2* pseudogenization (2 steps)
**D2**: *rps2* gene loss (3 steps)
**E1**: *rps3* pseudogenization (2 steps)
**E2**: *rps3* gene loss (2 steps)
**F1**: *rps4* pseudogenization (3 steps)
**F2**: *rps4* gene loss
**G1**: *rps8* pseudogenization (2 steps)
**G2**: *rps8* gene loss (2 steps)
**H**: *rps11* gene loss
**I**: *rps18* gene loss (3 steps)
**J**: *rps19* gene loss
**K**: *rrn4.5* gene loss (4 steps)
**L**: *trnE-UUC* pseudogenization (?, 3 steps)
**M1**: *trnfM-CAU* pseudogenization (5 steps)
**M2**: *trnfM-CAU* gene loss
**N1**: *trnM-CAU* pseudogenization
**N2**: *trnM-CAU* gene loss
**O1**: *trnY-GUA* pseudogenization
**O2**: *trnY-GUA* gene loss

| | Gene present | Pseudo-genization | Gene absent |
|---|---|---|---|
| Gene present | 0 | 1 | 1 |
| Pseudo-genization | 10 | 0 | 1 |
| Gene absent | 100 | 100 | 0 |



## LITERATURE CITED


Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

APG 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.

APG 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.

APG 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.

Barbrook, A. C., C. J. Howe, and S. Purton. 2006. Why are plastid genomes retained in non-photosynthetic organisms? *Trends in Plant Science* 11: 101–108.

Barrett, C. F., and J. I. Davis. 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany* 99: 1513–1523.

Barthet, M. M., C. L. Pierpont, and E. K. Tavernier. 2020. Unraveling the role of the enigmatic *MatK* maturase in chloroplast group IIA intron excision. *Plant Direct* 4: 1–17.

Bell, D., Q. Lin, W. K. Gerelle, S. Joya, Y. Chang, Z. N. Taylor, C. J. Rothfels, et al. 2020. Organellomic data sets confirm a cryptic consensus on (unrooted) land-plant relationships and provide new insights into bryophyte molecular evolution. *American Journal of Botany* 107: 91–115.

Bellot, S., and S. S. Renner. 2015. The plastomes of two species in the endoparasite genus Pilostyles (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution* 8: 189–201.

Bendich, A. J. 2004. Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16: 1661–1666.

Bidartondo, M. I. 2005. The evolutionary ecology of myco-heterotrophy. *New Phytologist* 167: 335–352.

Blazier, J. C., R. K. Jansen, J. P. Mower, M. Govindu, J. Zhang, M. Weng, and T. A. Ruhlman. 2016. Variable presence of the inverted repeat and plastome stability in *Erodium*. *Annals of Botany* 117: 1209–1220.

Bolin, J. F., K. U. Tennakoon, M. Bin Abdul Majid, and D. D. Cameron. 2017. Isotopic evidence

of partial mycoheterotrophy in *Burmannia coelestis* (Burmanniaceae). *Plant Species Biology* 32: 74–80.

Braukmann, T. W. A., M. B. Broe, S. Stefanović, and J. V. Freudenstein. 2017. On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytologist* 216: 254–266.

Caddick, L. R., P. J. Rudall, P. Wilkin, T. A. J. Hedderson, and M. W. Chase. 2002. Phylogenetics of Dioscoreales based on combined analysis of morphological and molecular data. *Botanical Journal of the Linnean Society* 138: 123–144.

Cauz-Santos, L. A., Z. P. da Costa, C. Callot, S. Cauet, M. I. Zucchi, H. Bergès, C. van den Berg, and M. L. C. Vieira. 2020. A repertory of rearrangements and the loss of an inverted repeat region in *Passiflora* chloroplast genomes. *Genome Biology and Evolution* 12: 1841–1857.

Chan, P. P., and T. M. Lowe. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods in Molecular Biology*, 1962: 1–14.

Chase, M. W., D. W. Stevenson, P. Wilkin, and P. J. Rudall. 1995. Monocot systematics: a combined analysis. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: systematics and evolution, 685–730. Royal Botanic Gardens, Kew.

Cheek, M., H. Tsukaya, P. J. Rudall, and K. Suetsugu. 2018. Taxonomic monograph of *Oxygyne* (Thismiaceae), rare achlorophyllous mycoheterotrophs with strongly disjunct distribution. *PeerJ* 6: e4828.

Choi, I., R. Jansen, and T. Ruhlman. 2019. Lost and found: return of the inverted repeat in the legume clade defined by its absence. *Genome Biology and Evolution* 11: 1321–1333.

Christopher, D. A., and R. B. Hallick. 1989. *Euglena gracilis* chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. *Nucleic Acids Research* 17: 7591–7608.

Chumley, T. W., J. D. Palmer, J. P. Mower, H. M. Fourcade, P. J. Calie, J. L. Boore, and R. K. Jansen. 2006. The complete chloroplast genome sequence of *Pelargonium* ×*hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution* 23: 2175–2190.

Cosner, M. E., L. A. Raubeson, and R. K. Jansen. 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evolutionary Biology* 4.

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

Darshetkar, A. M., S. Maurya, C. Lee, B. Bazarragchaa, G. Batdelger, A. Janchiv, E. J. Jeong, et al. 2021. Plastome analysis unveils inverted repeat (IR) expansion and positive selection in sea lavenders (*Limonium*, Plumbaginaceae, Limonioideae, Limonieae). *PhytoKeys* 175: 89–107.

Dierckxsens, N., P. Mardulyn, and G. Smits. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist* 125: 1–15.

Fonseca, L. H. M., and L. G. Lohmann. 2017. Plastome rearrangements in the "*Adenocalymma-neojobertia*" clade (Bignonieae, Bignoniaceae) and its phylogenetic implications. *Frontiers in Plant Science* 8: 1–13.

Gastineau, R., N. A. Davidovich, O. I. Davidovich, C. Lemieux, M. Turmel, R. J. Wróbel, and A. Witkowski. 2021. Extreme enlargement of the inverted repeat region in the plastid genomes of diatoms from the genus *Climaconeis*. *International Journal of Molecular Sciences* 22: 7155.

Givnish, T. J., A. Zuluaga, D. Spalink, M. Soto Gomez, V. K. Y. Lam, J. M. Saarela, C. Sass, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *American Journal of Botany* 105: 1888–1910.

Graham, S. W., V. K. Y. Lam, and V. S. F. T. Merckx. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist* 217: 48-55.

Graham, S. W., P. A. Reeves, A. C. E. Burns, and R. G. Olmstead. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161: S83–S96.

Guisinger, M. M., J. V. Kuehl, J. L. Boore, and R. K. Jansen. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution* 28: 583–600.

Hansen, D. R., S. G. Dastidar, Z. Cai, C. Penaflor, J. V. Kuehl, J. L. Boore, and R. K. Jansen. 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Molecular Phylogenetics and Evolution* 45: 547–563.

Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Biology* 38: 297–309.

Hertweck, K. L., M. S. Kinney, S. A. Stuart, O. Maurin, S. Mathews, M. W. Chase, M. A. Gandolfo, and J. C. Pires. 2015. Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Botanical Journal of the Linnean Society* 178: 375–393.

Hildebrand, M., R. B. Hallick, C. W. Passavant, and D. P. Bourque. 1988. Trans-splicing in chloroplasts: the *rps 12* loci of *Nicotiana tabacum*. *Proceedings of the National Academy of Sciences of the United States of America* 85: 372–376.

Iles, W. J. D., S. Y. Smith, M. A. Gandolfo, and S. W. Graham. 2015, Monocot fossils suitable for molecular dating analyses. *Botanical Journal of the Linnean Society* 178: 346–374.

Jacquemyn, H., and V. S. F. T. Merckx. 2019. Mycorrhizal symbioses and the evolution of trophic modes in plants. *Journal of Ecology* 107: 1567–1581.

Jacquier, A., and F. Michel. 1987. Multiple exon-binding sites in class II self-splicing introns. *Cell* 50: 17–29.

Jansen, R. K., and T. A. Ruhlman. 2012. Plastid genomes of seed plants. *In* R. Bock, and V. Knoop [eds.], Genomics of chloroplasts and mitochondria, 103–126. Springer, Dordrecht.

Jenkins, B. D., D. J. Kulhanek, and A. Barkan. 1997. Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* 9: 283–296.

Jin, D. M., S. Wicke, L. Gan, J. B. Yang, J. J. Jin, and T. S. Yi. 2020. The loss of the inverted repeat in the putranjivoid clade of Malpighiales. *Frontiers in Plant Science* 11: 942.

Jonker, F. P. 1938. A monograph of the Burmanniaceae. Kemink en Zoon N.V., Utrecht.

Joyce, E. M., D. M. Crayn, V. K. Y. Lam, W. K. Gerelle, S. W. Graham, and L. Nauheimer. 2018. Evolution of *Geosiris* (Iridaceae): Historical biogeography and plastid-genome evolution in a genus of non-photosynthetic tropical rainforest herbs disjunct across the Indian Ocean. *Australian Systematic Botany* 31: 504–522.

Kelchner, S. A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *American Journal of Botany* 89: 1651–1669.

Kim, K., K. Choi, and R. K. Jansen. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molecular Biology and Evolution* 22: 1783–1792.

Knox, E. B., and J. D. Palmer. 1999. The chloroplast genome arrangement of *Lobelia thuliniana* (Lobeliaceae): expansion of the inverted repeat in an ancestor of the Campanulales. *Plant Systematics and Evolution* 214: 49–64.

Koller, B., H. Fromm, E. Galun, and M. Edelman. 1987. Evidence for in vivo *trans* splicing of pre-mRNAs in tobacco chloroplasts. *Cell* 48: 111–119.

Koressaar, T., and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23: 1289–1291.

Kuhner, M. K., and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* 11: 459–468.

Lam, V. K. Y. 2016. Phylogenomics and comparative plastome analysis of mycoheterotrophic plants. Ph.D. dissertation, The University of British Columbia, Vancouver, British Columbia, Canada.

Lam, V. K. Y., H. Darby, V. S. F. T. Merckx, G. Lim, T. Yukawa, K. M. Neubig, J. R. Abbott, et al. 2018. Phylogenomic inference *in extremis*: a case study with mycoheterotroph plastomes. *American Journal of Botany* 105: 480–494.

Lam, V. K. Y., M. Soto Gomez, and S. W. Graham. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. *Genome Biology and Evolution* 7: 2220–2236.

Lanfear, R., B. Calcott, D. Kainer, C. Mayer, and A. Stamatakis. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14: 82.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder2: New methods for selecting partitioned models of evolution for molecular and morphological

phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

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

Layer, G., J. Reichelt, D. Jahn, and D. W. Heinz. 2010. Structure and function of enzymes in heme biosynthesis. *Protein Science* 19: 1137–1161.

Leake, J. 1994. The biology of myco-heterotrophic ('saprophytic') plants. *New Phytologist* 127: 171–216.

Leake, J. R., and D. D. Cameron. 2010. Physiological ecology of mycoheterotrophy. *New Phytologist* 185: 601–605.

Lee, H., R. K. Jansen, T. W. Chumley, and K. Kim. 2007. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Molecular Biology and Evolution* 24: 1161–1180.

Lee, K., S. J. Park, C. Colas des Francs-Small, M. Whitby, I. Small, and H. Kang. 2019. The coordinated action of PPR4 and EMB2654 on each intron half mediates trans-splicing of *rps12* transcripts in plant chloroplasts. *The Plant Journal* 100: 1193–1207.

Li, Z., X. Ma, Y. Wen, S. Chen, Y. Jiang, and X. Jin. 2020. Plastome of the mycoheterotrophic eudicot *Exacum paucisquama* (Gentianaceae) exhibits extensive gene loss and a highly expanded inverted repeat region. *PeerJ* 2020: 1–11.

Lim, G. S., C. F. Barrett, C. C. Pang, and J. I. Davis. 2016. Drastic reduction of plastome size in the mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. *American Journal of Botany* 103: 1129–1137.

Lin, Q. 2020. Phylogenetics and evolution of monocot mycoheterotrophs and a newly demonstrated lineage of carnivorous monocots. Ph.D. dissertation, The University of British Columbia, Vancouver, British Columbia, Canada.

Logacheva, M. D., M. I. Schelkunov, and A. A. Penin. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biology and Evolution* 3: 1296–1303.

Lohse, M., O. Drechsel, S. Kahlau, and R. Bock. 2013. OrganellarGenomeDRAW - a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 41: W575–W581.

Lüer, C., S. Schauer, S. Virus, W. D. Schubert, D. W. Heinz, J. Moser, and D. Jahn. 2007.

Glutamate recognition and hydride transfer by *Escherichia coli* glutamyl-tRNA reductase. *The FEBS Journal* 274: 4609–4614.

Maas, P. J. M., H. Maas-van de Kamer, J. Van Benthem, H. C. M. Snelders, and T. Rübsamen. 1986. Burmanniaceae. Flora Neotropica, Monograph 42. New York Botanical Garden Press, New York, New York, USA.

Maddison, W. P., and D. R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis. Computer program and documentation distributed by the author, website: http://www.mesquiteproject.org [accessed 15 Sep 2018].

McNeal, J. R., J. V. Kuehl, J. L. Boore, J. Leebens-Mack, and C. W. dePamphilis. 2009. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS ONE* 4: 1–8.

Merckx, V., F. T. Bakker, S. Huysmans, and E. Smets. 2009. Bias and conflict in phylogenetic inference of myco-heterotrophic plants: a case study in Thismiaceae. *Cladistics* 25: 64–77.

Merckx, V., and M. I. Bidartondo. 2008. Breakdown and delayed cospeciation in the arbuscular mycorrhizal mutualism. *Proceedings of the Royal Society B: Biological Sciences* 275: 1029–1035.

Merckx, V. S. F. T., J. V. Freudenstein, J. Kissling, M. J. M. Christenhusz, R. E. Stotler, B. Crandall-Stotler, N. Wickett, et al. 2013. Taxonomy and classification. *In* V. S. F. T. Merckx [eds.], Mycoheterotrophy: The Biology of Plants Living on Fungi, 19–101. Springer, New York.

Merckx, V. S. F. T., S. I. F. Gomes, M. Wapstra, C. Hunt, G. Steenbeeke, C. B. Mennes, N. Walsh, et al. 2017. The biogeographical history of the interaction between mycoheterotrophic *Thismia* (Thismiaceae) plants and mycorrhizal *Rhizophagus* (Glomeraceae) fungi. *Journal of Biogeography* 44: 1869–1879.

Merckx, V. S. F. T., and E. F. Smets. 2014. *Thismia americana*, the 101st anniversary of a botanical mystery. *International Journal of Plant Sciences* 175: 165–175.

Merckx, V., P. Schols, H. Maas-van de Kamer, P. Maas, S. Huysmans, and E. Smets. 2006. Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. *American Journal of Botany* 93: 1684–1698.

Merckx, V., M. Stöckel, A. Fleischmann, T. D. Bruns, and G. Gebauer. 2010. 15N and 13C natural abundance of two mycoheterotrophic and a putative partially mycoheterotrophic species associated with arbuscular mycorrhizal fungi. *New Phytologist* 188: 590–596.

Michel, F., and B. Dujon. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *The EMBO Journal* 2: 33–38.

Michel, F., U. Kazuhiko, and O. Haruo. 1989. Comparative and functional anatomy of group II catalytic introns - a review. *Gene* 82: 5–30.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop, GCE 2010*.

Molina, J., K. M. Hazzouri, D. Nickrent, M. Geisler, R. S. Meyer, M. M. Pentony, J. M. Flowers, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution* 31: 793–803.

Naumann, J., J. P. Der, E. K. Wafula, S. S. Jones, S. T. Wagner, L. A. Honaas, P. E. Ralph, et al. 2016. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biology and Evolution* 8: 345–363.

Oldenburg, D. J., and A. J. Bendich. 2004. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *Journal of Molecular Biology* 335: 953–970.

Palmer, J. D. 1985. Comparative organization of chloroplast genomes. *Annual Review of Genetics* 19: 325–354.

Palmer, J. D. 1991. Plastid chromosomes: structure and evolution. *In* L. Bogorad and I. K. Vasil [eds.], The Molecular Biology of Plastids, vol. 7A, Cell culture and somatic cell genetics of plants, 5–53. Academic Press, San Diego, California, USA.

Palmer, J. D., B. Osorio, J. Aldrich, and W. F. Thompson. 1987. Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. *Current Genetics* 11: 275–286.

Palmer, J. D., and W. F. Thompson. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 29: 537–550.

Petersen, G., H. Darby, V. K. Y. Lam, H. A. Pedersen, V. S. F. T. Merckx, A. Zervas, O. Seberg, and S. W. Graham. 2019. Mycoheterotrophic *Epirixanthes* (Polygalaceae) has a typical angiosperm mitogenome but unorthodox plastid genomes. *Annals of Botany* 124: 791–807.

Petersen, G., A. Zervas, H. Pedersen, and O. Seberg. 2018. Genome reports: Contracted genes and dwarfed plastome in mycoheterotrophic *Sciaphila thaidanica* (Triuridaceae, Pandanales). *Genome Biology and Evolution* 10: 976–981.

Potter, S. C., A. Luciani, S. R. Eddy, Y. Park, R. Lopez, and R. D. Finn. 2018. HMMER web server: 2018 update. *Nucleic Acids Research* 46: W200–W204.

Qu, X., S. Fan, S. Wicke, and T.-S. Yi. 2019. Plastome reduction in the only parasitic gymnosperm *Parasitaxus* is due to losses of photosynthesis but not housekeeping genes and apparently involves the secondary gain of a large inverted repeat. *Genome Biology and Evolution* 11: 2789–2796.

Randau, L., S. Schauer, A. Ambrogelly, J. C. Salazar, S. Sekine, S. Yokoyama, D. So, and D. Jahn. 2004. tRNA Recognition by Glutamyl-tRNA Reductase. *The Journal of Biological Chemistry* 279: 34931–34937.

Raubeson, L. A., and R. K. Jansen. 1992. A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochemical Systematics and Ecology* 20: 17–24.

Rogalski, M., N. Vieira, H. P. Fraga, and M. P. Guerra. 2015. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Frontiers in Plant Science* 6: 586.

Roquet, C., É. Coissac, C. Cruaud, M. Boleda, F. Boyer, A. Alberti, L. Gielly, et al. 2016. Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). *Annals of Botany* 118: 885–896.

Saarela, J. M., and S. W. Graham. 2010. Inference of phylogenetic relationships among the subfamilies of grasses (Poaceae: Poales) using meso-scale exemplar-based sampling of the plastid genome. *Botany* 88: 65–84.

Sabir, J., E. Schwarz, N. Ellison, J. Zhang, N. A. Baeshen, M. Mutwakil, R. Jansen, and T. Ruhlman. 2014. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnology Journal* 12: 743–754.

Sanderson, M. J., D. Copetti, A. Burquez, E. Bustamante, J. L. M. Charboneau, L. E. Eguiarte, S. Kumar, et al. 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. *American Journal of Botany* 102: 1115–1127.

Schelkunov, M. I., M. S. Nuraliev, and M. D. Logacheva. 2018. *Rhopalocnemis phalloides* has

one of the most reduced and mutated plastid genomes known. *PeerJ* 7: e7500.

Schelkunov, M. I., V. Y. Shtratnikova, M. S. Nuraliev, M. A. Selosse, A. A. Penin, and M. D. Logacheva. 2015. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution* 7: 1179–1191.

Schon, A., G. Krupp, S. Gough, S. Berry-Lowe, C. G. Kannangara, and D. Soll. 1986. The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. *Nature* 322: 281–284.

Shepeleva, E. A., M. I. Schelkunov, M. Hroneš, M. Sochor, M. Dančák, V. S. Merckx, I. A. Kikuchi, et al. 2020. Phylogenetics of the mycoheterotrophic genus *Thismia* (Thismiaceae: Dioscoreales) with a focus on the Old World taxa: delineation of novel natural groups and insights into the evolution of morphological traits. *Botanical Journal of the Linnean Society* 193: 287–315.

Sinn, B. T., D. D. Sedmak, L. M. Kelly, and J. V. Freudenstein. 2018. Total duplication of the small single copy region in the angiosperm plastome: rearrangement and inverted repeat instability in *Asarum*. *American Journal of Botany* 105: 71–84.

Sochor, M., M. Hroneš, and M. Dančák. 2018. New insights into variation, evolution and taxonomy of fairy lanterns (*Thismia*, Thismiaceae) with four new species from Borneo. *Plant Systematics and Evolution* 304: 699–721.

Soltis, P. S., and D. E. Soltis. 2003. Applying the bootstrap in phylogeny reconstruction. *Statistical Science* 18: 256–267.

Soto Gomez, M. 2020. Phylogenomic studies of the monocot sister orders Pandanales and Dioscoreales. Ph.D. dissertation, The University of British Columbia, Vancouver, British Columbia, Canada.

Soto Gomez, M., Q. Lin, E. da Silva Leal, T. J. Gallaher, D. Scherberich, C. B. Mennes, S. Y. Smith, and S. W. Graham. 2020. A bi-organellar phylogenomic study of Pandanales: inference of higher-order relationships and unusual rate-variation patterns. *Cladistics* 36: 481–504.

Soto Gomez, M., L. Pokorny, M. B. Kantar, F. Forest, I. J. Leitch, B. Gravendeel, P. Wilkin, et al. 2019. A customized nuclear target enrichment approach for developing a phylogenomic baseline for *Dioscorea* yams (Dioscoreaceae). *Applications in Plant Sciences* 7: e11254.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758–771.

Steele, P. R., K. L. Hertweck, D. Mayfield, M. R. McKain, J. Leebens-Mack, and J. C. Pires. 2012. Quality and quantity of data recovered from massively parallel sequencing: examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330–348.

Straub, S. C. K., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Su, H. J., T. J. Barkman, W. Hao, S. S. Jones, J. Naumann, E. Skippington, E. K. Wafula, et al. 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 of the United States of America* 116: 934–943.

Sun, Y. X., M. J. Moore, A. P. Meng, P. S. Soltis, D. E. Soltis, J. Q. Li, and H. C. Wang. 2013. Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a paleogene divergence between the two extant species. *PLoS ONE* 8: 1–14.

Tillich, M., P. Lehwark, T. Pellizzer, E. S. Ulbricht-Jones, A. Fischer, R. Bock, and S. Greiner. 2017. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45: W6–W11.

Thiers, B. Continuously updated. Index Herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. Website: http://sweetgum.nybg.org/science/ih/ [accessed 06 Sep 2021].

Twyford, A. D., and R. W. Ness. 2017. Strategies for complete plastid genome sequencing. *Molecular Ecology Resources* 17: 858–868.

Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. M. Leunissen. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35: W71–W74.

Vogel, J., T. Börner, and W. R. Hess. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Research* 27: 3866–3874.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.

Weng, M. L., T. A. Ruhlman, and R. K. Jansen. 2017. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytologist* 214: 842–851.

Wicke, S., K. F. Müller, W. Claude, D. Quandt, and S. Bellot. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences of the United States of America* 113: 9045–9050.

Wicke, S., K. F. Müller, C. W. de Pamphilis, D. Quandt, N. J. Wickett, Y. Zhang, S. S. Renner, and G. M. Schneeweiss. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25: 3711–3725.

Wicke, S., G. M. Schneeweiss, C. W. DePamphilis, K. F. Muller, and D. Quandt. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76: 273–297.

Willows, R. D., C. G. Kannangara, and B. Pontoppidan. 1995. Nucleotides of tRNA (Glu) involved in recognition by barley chloroplast glutamyl-tRNA synthetase and glutamyl-tRNA reductase. *Biochimica et Biophysica Acta.* 1263: 228–234.

Woodward, C. L., P. E. Berry, H. Maas-van De Kamer, and K. Swing. 2007. *Tiputinia foetida*, a new mycoheterotrophic genus of Thismiaceae from Amazonian Ecuador, and a likely case of deceit pollination. *Taxon* 56: 157–162.

Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

Xu, J. H., Q. Liu, W. Hu, T. Wang, Q. Xue, and J. Messing. 2015. Dynamics of chloroplast genomes in green plants. *Genomics* 106: 221–231.

Yang, Z. 2006. Computational molecular evolution. Oxford University Press, Incorporated.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.

Yokoyama, J., Y. Koizumi, M. Yokota, and H. Tsukaya. 2008. Phylogenetic position of *Oxygyne shinzatoi* (Burmanniaceae) inferred from 18S rDNA sequences. *Journal of Plant Research* 121: 27–32.

Yudina, S. V, M. I. Schelkunov, L. Nauheimer, D. Crayn, S. Chantanaorrapint, M. Hroneš, M. Sochor, et al. 2021. Comparative analysis of plastid genomes in the non-photosynthetic genus *Thismia* reveals ongoing gene set reduction. *Frontiers in Plant Science* 12: 1–13.

Zoschke, R., M. Nakamura, K. Liere, M. Sugiura, T. Börner, and C. Schmitz-Linneweber. 2010. An organellar maturase associates with multiple group II introns. *Proceedings of the National Academy of Sciences of the United States of America* 107: 3245–3250.

## APPENDICES

**Appendix A** — Supplementary table and figures

**Table A1.** A: Thismiaceae (Dioscoreales) taxa with a complete or partial plastid genome assembled (23 species). B: All taxa included in the matrices used to create RAxML maximum likelihood trees. *Apteria aphylla* and *Gymnosiphon longistylus* were excluded from a reduced version of the matrix used in most analyses (see Materials and Methods).

**A**

| Species | Authority | Plastome status | Source: GenBank Accession Number or Voucher (Name(s) Collector's number, herbarium code)† |
|---|---|---|---|
| *Haplothismia exannulata* | Airy Shaw | complete | Sujanapal 30387, KFRI |
| *Thismia huangii* | P.Y.Jiang & T.H.Hsieh | partial | Hseih 3031, TAI |
| *Thismia javanica* | J.J.Sm. | complete | Ruchisansakun 770, L |
| *Thismia panamensis* | (Standl.) Jonker | complete | Aizprua 2946, RA |
| *Thismia rodwayi* | F.Muell. | complete | Merckx & Wapstra TAS4-5, L |
| *Thismia tentaculata* | K.Larsen & Aver. | complete | KX171421.1 |
| *Thismia alba* | Holttum ex Jonker | partial | MT943132.1 |
| *Thismia angustimitra* | Chantanaorr. | complete | MT943130.1 |
| *Thismia annamensis* | K.Larsen & Aver. | partial | MT943129.1 |
| *Thismia cornuta* | Hroneš, Sochor & Dančák | partial | MT943131.1 |
| *Thismia filiformis* | Chantanaorr. | complete | MT936983.1 |
| *Thismia gardneriana* | Hook.f. ex Thwaites | complete | MT943140.1 |
| *Thismia hawkesii* | W.E.Cooper | complete | MT936984.1 |
| *Thismia hexagona* | Dančák, Hroneš, Kobrlová & Sochor | partial | MT943134.1 |
| *Thismia hongkongensis* | Mar & R.M.K.Saunders | partial | MT943133.1 |
| *Thismia kelabitiana* | Dančák, Hroneš & Sochor | partial | MT936982.1 |
| *Thismia lanternata* | W.E.Cooper | complete | MT936985.1 |
| *Thismia mucronata* | Nuraliev | complete | MT943141.1 |
| *Thismia neptunis* | Becc. | partial | MT943135.1, MT943136.1 |
| *Thismia okhaensis* | Luu, Tich, Tran, G. Tran & Đinh, Quang Diệp | partial | MT943137.1, MT943139.1 |
| *Thismia puberula* | Nuraliev | complete | MT747830.1 |
| *Thismia thaithongiana* | Chantanaorr. & Suddee | complete | MT943142.1 |
| *Thismia viridistriata* | Sochor, Hroneš & Dančák | partial | MT943138.1 |

†: Herbarium abbreviations follow Thiers (continuously updated).

**B**

| Clade | Order | Family | Species | Authority | Voucher (Name(s) Collector's number, herbarium code)† | Sequence Accession |
|---|---|---|---|---|---|---|
| | Amborellales | Amborellaceae | *Amborella trichopoda* | Baill. | | NC_005086 |
| | Austrobaileyales | Schisandraceae | *Illicium parvifolium** | Merr. | | NC_009600 |
| | Nymphaeales | Nymphaeaceae | *Nuphar advena* | (Aiton) W.T. Aiton | | NC_008788 |
| Eudicots | Buxales | Buxaceae | *Buxus microphylla* | Siebold & Zucc. | | NC_009599 |
| Eudicots | Proteales | Platanaceae | *Platanus occidentalis* | L. | | NC_008335 |
| Eudicots | Ranunculales | Berberidaceae | *Nandina domestica* | Thunb. | | NC_008336 |
| Eudicots | Vitales | Vitaceae | *Vitis vinifera* | L. | | NC_007957 |
| Magnoliids | Canellales | Winteraceae | *Drimys granadensis* | L.f. | | NC_008456 |
| Magnoliids | Laurales | Calycanthaceae | *Calycanthus floridus* | L. | | NC_004993 |
| Magnoliids | Magnoliales | Magnoliaceae | *Liriodendron tulipifera* | L. | | NC_008326 |
| Magnoliids | Piperales | Piperaceae | *Piper cenocladum* | C.DC. | | NC_008457 |
| Monocots | Acorales | Acoraceae | *Acorus calamus* | L. | | NC_007407 |
| Monocots | Alismatales | Araceae | *Lemna minor* | L. | | NC_010109 |
| Monocots | Alismatales | Araceae | *Orontium aquaticum* | L. | | http://purl.org/phylo/treebase/phylows/study/TB2:S15395 |
| Monocots | Alismatales | Araceae | *Spirodela polyrrhiza* | (L.) Schleid. | | NC_015891 |
| Monocots | Alismatales | Araceae | *Wolffia australiana* | (Benth.) Hartog & Plas | | NC_015899 |
| Monocots | Alismatales | Araceae | *Wolffiella lingulata* | (Hegelm.) Hegelm. | | NC_015894 |
| Monocots | Arecales | Arecaceae | *Bismarckia nobilis* | Hildebr. & H.Wendl. | | JX088664 |
| Monocots | Arecales | Arecaceae | *Calamus caryotoides* | A.Cunn. ex Mart. | | JX088663 |
| Monocots | Arecales | Arecaceae | *Chamaedorea seifrizii* | Burret | | JX088667 |
| Monocots | Arecales | Arecaceae | *Elaeis oleifera* | (Kunth) Cortés | | EU016883-EU016962 |
| Monocots | Arecales | Arecaceae | *Phoenix dactylifera* | L. | | NC_013991 |
| Monocots | Arecales | Arecaceae | *Pseudophoenix vinifera* | (Mart.) Becc. | | JX088662 |
| Monocots | Arecales | Dasypogonaceae | *Calectasia narragara* | R.L.Barrett & K.W.Dixon | | JX088666 |
| Monocots | Arecales | Dasypogonaceae | *Dasypogon bromeliifolius* | R.Br. | | JX088665 |
| Monocots | Arecales | Dasypogonaceae | *Kingia australis* | R.Br. | | JX051651 |
| Monocots | Asparagales | Asparagaceae | *Albuca kirkii* | (Baker) Brenan | | HQ180401- HQ183670 |
| Monocots | Asparagales | Asparagaceae | *Asparagus officinalis* | L. | | HQ180403- HQ183672 |
| Monocots | Asparagales | Asparagaceae | *Chlorophytum rhizopendulum* | Bjorå & Hemp | | HQ180408-HQ183677 |
| Monocots | Asparagales | Asparagaceae | *Hesperaloe parviflora* | (Torr.) J.M.Coult. | | HQ180417-HQ183686 |
| Monocots | Asparagales | Asparagaceae | *Hosta ventricosa* | Stearn | | HQ180418-HQ183687 |
| Monocots | Asparagales | Asparagaceae | *Nolina atopocarpa* | Bartlett | | HQ180429-HQ183697 |
| Monocots | Asparagales | Asparagaceae | *Yucca schidigera* | Roezl ex Ortgies | | DQ069342-DQ069702, EU016681-EU016700 |
| Monocots | Asparagales | Iridaceae | *Iris missouriensis* | Nutt. | M.A. McPherson 000707-5a-7, ALTA | Figshare: DOI: 10.6084/m9.figshare.5480608 |
| Monocots | Asparagales | Orchidaceae | *Cypripedium formosanum* | Hayata | | NC_026772 |
| Monocots | Asparagales | Orchidaceae | *Oncidium* Gower Ramsey gx | | | NC_014056 |
| Monocots | Asparagales | Orchidaceae | *Phalaenopsis aphrodite* | Rchb.f. | | AY916449 |
| Monocots | Commelinales | Commelinaceae | *Tradescantia ohiensis* | Raf. | | HQ180441-HQ183709 |
| Monocots | Commelinales | Haemodoraceae | *Xiphidium caeruleum* | Aubl. | | JX088669 |
| Monocots | Dioscoreales | Burmanniaceae | *Apteria aphylla* | (Nutt.) Barnhart ex Small | D.M. McNair 952, USMS | Figshare: DOI: 10.6084/m9.figshare.5480608 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia bicolor* | Mart. | Maas et al. 9649, U | Figshare: DOI:10.6084/m9.figshare.1407422 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia capitata* | (Walter ex J.F.Gmel.) Mart. | Maas et al. 9606, U | Figshare: DOI: 10.6084/m9.figshare.5480608 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia championii* | Thwaites | | KT734621.1 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia coelestis* | D.Don | | KT734618.1 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia cryptopetala* | Makino | | KT734620.1 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia disticha* | L. | | NC_036661.1 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia itoana* | Makino | Kun-Ping Lo 821, PPI | Figshare: DOI: 10.6084/m9.figshare.5480608 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia nepalensis* | (Miers) Hook.f. | | KU954767.1 |
| Monocots | Dioscoreales | Burmanniaceae | *Burmannia oblonga* | Ridl. | | KT734622.1 |
| Monocots | Dioscoreales | Burmanniaceae | *Campylosiphon congestus*** | (C.H.Wright) Maas | T. Franke K/8, M | Figshare: DOI: 10.6084/m9.figshare.5480608 |
| Monocots | Dioscoreales | Burmanniaceae | *Gymnosiphon longistylus* | (Benth.) Hutch. | Merckx et al. 132, LV | Figshare: DOI: 10.6084/m9.figshare.5480608 |
| Monocots | Dioscoreales | Dioscoreaceae | *Dioscorea collettii*†† | Hook.f. | | NC_037717.1 |

| Clade | Order | Family | Species | Authority | Voucher (Name(s) Collector's number, herbarium code)† | Sequence Accession |
|---|---|---|---|---|---|---|
| Monocots | Dioscoreales | Dioscoreaceae | *Dioscorea elephantipes*†† | (L'Hér.) Engl. | | NC_009601 |
| Monocots | Dioscoreales | Dioscoreaceae | *Dioscorea membranacea*†† | Pierre ex Prain & Burkill | M.W. Chase 21050, K | |
| Monocots | Dioscoreales | Dioscoreaceae | *Dioscorea polystachya*†† | Turcz. | | NC_037716.1 |
| Monocots | Dioscoreales | Dioscoreaceae | *Dioscorea rotundata*†† | Poir. | | NC_024170.1 (under D. cayenensis subsp. Rotundata) |
| Monocots | Dioscoreales | Dioscoreaceae | *Dioscorea villosa*†† | L. | | NC_034686.1 |
| Monocots | Dioscoreales | Dioscoreaceae | *Dioscorea zingiberensis*†† | C.H.Wright | | NC_027090.1 |
| Monocots | Dioscoreales | Trichopodaceae | *Trichopus sempervirens*†† | (H.Perrier) Caddick & Wilkin | Ranaivojaona, R. 790, K | |
| Monocots | Dioscoreales | Trichopodaceae | *Trichopus zeylanicus*†† | Gaertn. | M.W. Chase 16354, K | |
| Monocots | Dioscoreales | Nartheciaceae | *Aletris faurieii*†† | H. Lév. & Vaniot | | NC_033412.1 (under A. foliata) |
| Monocots | Dioscoreales | Nartheciaceae | *Aletris obovata*†† | Nash | Iles & Norman 010, UBC | |
| Monocots | Dioscoreales | Nartheciaceae | *Aletris spicata*†† | (Thunb.) Franch. | | NC_033411.1 |
| Monocots | Dioscoreales | Nartheciaceae | *Lophiola aurea**** †† | Ker Gawl. | Whitten 95028, K | KT204866 - KT204945 |
| Monocots | Dioscoreales | Nartheciaceae | *Metanarthecium luteoviride*†† | Maxim. | | NC_029214.1 |
| Monocots | Dioscoreales | Nartheciaceae | *Narthecium californicum*†† | Baker | JM Saarela 19803, UBC | |
| Monocots | Dioscoreales | Nartheciaceae | *Nietneria paniculata*†† | Steyerm. | O. Hokche & P.J. Maas 849, U | |
| Monocots | Dioscoreales | Taccaceae | *Tacca chantrieri*†† | André | | KX171420 |
| Monocots | Dioscoreales | Taccaceae | *Tacca leontopetaloides*†† | (L.) Kuntze | | NC_036658.1 |
| Monocots | Dioscoreales | Thismiaceae | *Haplothismia exannulata*†† | Airy Shaw | Sujanapal 30387, KFRI | |
| Monocots | Dioscoreales | Thismiaceae | *Thismia alba* | Holttum ex Jonker | | MT943132.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia angustimitra* | Chantanaorr. | | MT943130.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia annamensis* | K.Larsen & Aver. | | MT943129.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia cornuta* | Hroneš, Sochor & Dančák | | MT943131.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia filiformis*†† | Chantanaorr. | | MT936983.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia gardneriana* | Hook.f. ex Thwaites | | MT943140.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia hawkesii*†† | W.E.Cooper | | MT936984.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia hexagona* | Dančák, Hroneš, Kobrlová & Sochor | | MT943134.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia hongkongensis* | Mar & R.M.K.Saunders | | MT943133.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia huangii*†† | P.Y.Jiang & T.H.Hsieh | Hseih 3031, TAI | |
| Monocots | Dioscoreales | Thismiaceae | *Thismia javanica*†† | J.J.Sm. | Ruchisansakun 770, L | |
| Monocots | Dioscoreales | Thismiaceae | *Thismia kelabitiana* | Dančák, Hroneš & Sochor | | MT936982.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia lanternata* | W.E.Cooper | | MT936985.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia mucronata*†† | Nuraliev | | MT943141.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia neptunis* | Becc. | | MT943135.1, MT943136.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia okhaensis* | Luu, Tich, Tran, G. Tran & Đinh, Quang Diệp | | MT943137.1, MT943139.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia panamensis*†† | (Standl.) Jonker | Aizprua 2946, RA | |
| Monocots | Dioscoreales | Thismiaceae | *Thismia puberula* | Nuraliev | | MT747830.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia rodwayi*†† | F.Muell. | Merckx & Wapstra TAS4-5, L | |
| Monocots | Dioscoreales | Thismiaceae | *Thismia tentaculata*†† | K.Larsen & Aver. | | KX171421.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia thaithongiana*†† | Chantanaorr. & Suddee | | MT943142.1 |
| Monocots | Dioscoreales | Thismiaceae | *Thismia viridistriata* | Sochor, Hroneš & Dančák | | MT943138.1 |
| Monocots | Liliales | Alstroemeriaceae | *Alstroemeria aurea* | Graham | | KC968976 |
| Monocots | Liliales | Campynemataceae | *Campynema lineare* | Labill. | M.F. Duretto 1842, HO | NC_026785 |
| Monocots | Liliales | Campynemataceae | *Campynemanthe viridiflora* | Baill. | | KU303580-KU303650 |
| Monocots | Liliales | Colchicaceae | *Wurmbea pygmaea* | (Endl.) Benth | | KU303506-KU303579 |
| Monocots | Liliales | Liliaceae | *Fritillaria taipaiensis* | P.Y.Li | | NC_023247 |
| Monocots | Liliales | Liliaceae | *Lilium superbum* | L. | M.W. Chase 112, NCU | NC_026787, Figshare: DOI: 10.6084/m9.figshare.5480608 |
| Monocots | Liliales | Melanthiaceae | *Veratrum oxysepalum***** | Turcz. | | NC_022715 |
| Monocots | Liliales | Petermanniaceae | *Petermannia cirrosa* | F.Muell. | | KU304256-KU304331 |
| Monocots | Liliales | Philesiaceae | *Lapageria rosea* | Ruiz & Pav. | | KU303727-KU303798 |
| Monocots | Liliales | Rhipogonaceae | *Rhipogonum album* | R.Br. | | KU304103-KU304178 |
| Monocots | Liliales | Smilacaceae | *Smilax china* | L. | | HM536959 |
| Monocots | Pandanales | Cyclanthaceae | *Carludovica palmata* | Ruiz & Pav. | M.W. Chase 14836, K | NC_026786 |
| Monocots | Pandanales | Cyclanthaceae | *Cyclanthus bipartitus* | Poit. ex A.Rich. | M.W. Chase 1237, K | KT205192 - KT205273 |
| Monocots | Pandanales | Pandanaceae | *Freycinetia banksii* | A.Cunn. | S.W. Graham 02-03-14, UBC | KT205110 - KT205191 |
| Monocots | Pandanales | Pandanaceae | *Sararanga sinuosa* | Hemsl. | Gallaher 461, BISH, HAW | KT204539 - KT204619 |

| Clade | Order | Family | Species | Authority | Voucher (Name(s) Collector's number, herbarium code)† | Sequence Accession |
|---|---|---|---|---|---|---|
| Monocots | Pandanales | Stemonaceae | *Croomia japonica* | Miq. | Rothwell & Stockey 43, ALTA | KT204620 - KT204701 |
| Monocots | Pandanales | Stemonaceae | *Pentastemona sumatrana* | Steenis | B.G. Leiden 910375, K | KT205028 - KT205109 |
| Monocots | Pandanales | Stemonaceae | *Stemona tuberosa* | Lour. | Rothwell & Stockey 46, ALTA | KT204702 - KT204783 |
| Monocots | Pandanales | Stemonaceae | *Stichoneuron caudatum* | Ridl. | Rothwell & Stockey 45, ALTA | KT204946 - KT205027 |
| Monocots | Pandanales | Velloziaceae | *Xerophyta retinervis* | Baker | B.G. Reeves 14, NBG | KT204784 - KT204865 |
| Monocots | Petrosaviales | Petrosaviaceae | *Japonolirion osense* | Nakai | | JQ068951-JQ069028 |
| Monocots | Poales | Bromeliaceae | *Brocchinia micrantha* | (Baker) Mez | | HQ183672- HQ183674 |
| Monocots | Poales | Bromeliaceae | *Neoregelia carolinae* | (Beer) L.B.Sm. | | HQ180428-HQ183696 |
| Monocots | Poales | Bromeliaceae | *Puya laxa* | L.B.Sm. | | HQ180433-HQ183702 |
| Monocots | Poales | Poaceae | *Hordeum vulgare* | L. | | NC_008590 |
| Monocots | Poales | Poaceae | *Oryza sativa* | L. | | NC_001320 |
| Monocots | Poales | Poaceae | *Saccharum officinarum* | L. | | NC_006084 |
| Monocots | Poales | Poaceae | *Sorghum bicolor* | L. (Moench) | | NC_008602 |
| Monocots | Poales | Poaceae | *Triticum aestivum* | L. | M.W. Chase 3000, K | NC_002762 |
| Monocots | Poales | Poaceae | *Zea mays* | L. | | NC_001666 |
| Monocots | Poales | Typhaceae | *Typha latifolia* | L. | | NC_013823 |
| Monocots | Zingiberales | Heliconiaceae | *Heliconia collinsiana* | Griggs | | JX088660 |
| Monocots | Zingiberales | Musaceae | *Musa acuminata* | Colla | | EU016983-EU017063 |
| Monocots | Zingiberales | Zingiberaceae | *Alpinia zerumbet* | (Pers.) B.L.Burtt & R.M.Sm. | | JX088668 |
| Monocots | Zingiberales | Zingiberaceae | *Zingiber spectabile* | Griff. | | JX088661 |

†: Herbarium abbreviations follow Thiers (continuously updated).
††: Species included in PAML ω analysis
*: *Illicium parvifolium* was originally labelled *Illicium oligandrum* in Lam et al. (2018) paper
**: *Campylosiphon congestus* was originally labelled *Burmannia congesta* in Lam et al. (2018) paper
***: *Lophiola aurea* was originally labelled *Lophiola americana* in Lam et al. (2018) paper
****: *Veratrum oxysepalum* was originally labelled *Veratrum patulum* in Lam et al. (2018) paper

**Table A2.** List of primer sequences used to create consensus plastid genome sequence by filling gaps and verifying contig overlap.

| *Taxon* | *Primer name* | *Primer sequence (5' to 3')* | *Primer pair* |
|---|---|---|---|
| *Haplothismia exannulata* | HAEX_2L | GGTATCAAGCCAAAGTGCCC | HAEX_2R |
| | HAEX_2R | AATCGCTAGTAATCGCCGGT | HAEX_2L |
| | HAEX_NL | TCGCGCTCTGTAGGATTTGA | HAEX_NR |
| | HAEX_NR | TGACCCTCCGATCCAAATGT | HAEX_NL |
| *Thismia panamensis* | THPA_AL | TGTAAGGCAGGATCCAAAATTTG | THPA_AR |
| | THPA_AR | TCGATAAGCCTCTTTAAACGGG | THPA_AL |

**Table A3. A:** Annotation edits and additions for *Thismia tentaculata* published by Lim et al. (2016). **B:** Annotation edits and additions for *Thismia* species published by Yudina et al. (2021) as compared to their annotations.

**A**

**Newly Annotated**

| |
|---|
| *3'-rps12* exon 2 |
| *3'-rps12* exon 3 |
| *rrn4.5* |

**Gene boundaries altered slightly (under 21 bp)**

| |
|---|
| *3'-rps12* exon 1 |
| *rrn5* |
| *rrn16* |

**Gene boundaries altered significantly (at least 21 bp)**

| |
|---|
| *rrn23* |
| *rpl2* exon 1 |
| *rpl2* exon 2 |
| *rps4* |

**B**

**Newly Annotated**

| | *3'-rps12* exon 3 | *trnfM-CAU* | *rrn16* | *trnE-UUC* | *rps2* | *5'-rps12* exon 1 | *accD* | *3'-rps12* exon 2 |
|---|---|---|---|---|---|---|---|---|
| *Thismia angustimitra* | recovered | 2nd copy in IR | | 2nd copy in IR | | | | |
| *Thismia alba* | recovered | | fragment recovered | | | | | |
| *Thismia annamensis* | recovered | | | | | | | |
| *Thismia cornuta* | recovered | | fragment recovered | | | | | |
| *Thismia filiformis* | recovered | | | | | | | |
| *Thismia gardneriana* | recovered | 2nd copy in IR | | | | | | |
| *Thismia hawkesii* | recovered | 2nd copy in IR | | | | | | |
| *Thismia hexagona* | recovered | | | | | | | |
| *Thismia hongkongensis* | recovered | | | | | | | |
| *Thismia kelabitiana* | recovered | | | | | | | |
| *Thismia lanternata* | recovered | 2nd copy in IR | | | | | | |
| *Thismia mucronata* | recovered | | | | | | | |
| *Thismia neptunis* | | | | | 2nd copy recovered | 2nd copy recovered | 2nd copy recovered | |
| *Thismia okhaensis* | recovered | | | | | | | |
| *Thismia puberula* | recovered | | | | | | | |
| *Thismia thaithongiana* | recovered | | | | | | | 2nd copy in IR |
| *Thismia viridistriata* | recovered | | | | | | | |

**Gene boundaries altered slightly (under 21 bp)**

| | *5'-rps12* exon 1 | *rrn5* | *rrn16* | *trnfM-CAU* | *rpl2* exon 2 | *trnE-UUC* | *rrn23* | *3'-rps12* exon 2 | *accD* | *rps8* | *rps3* | *rps18* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Thismia angustimitra* | Yes | Yes | | | | | | | | | | |
| *Thismia alba* | | Yes | | Yes | Yes | Yes | | Yes | | Yes | | |
| *Thismia annamensis* | Yes | Yes | Yes | | Yes | | | | | | | |
| *Thismia cornuta* | Yes | Yes | | Yes | | | | | | | | |
| *Thismia filiformis* | Yes | Yes | Yes | Yes | | | | | | | | |
| *Thismia gardneriana* | Yes | Yes | | | | | | | | | | |
| *Thismia hawkesii* | Yes | Yes | Yes | | Yes | | Yes | | | | | |
| *Thismia hexagona* | Yes | | | Yes | Yes | Yes | | | | | | |
| *Thismia hongkongensis* | Yes | Yes | Yes | | | | | | | | | |
| *Thismia kelabitiana* | Yes | Yes | | | | | Yes | Yes | | | Yes | |
| *Thismia lanternata* | Yes | Yes | Yes | | | | | | | | | |
| *Thismia mucronata* | Yes | Yes | | Yes | Yes | | | | Yes | | | Yes |
| *Thismia neptunis* | Yes | | | Yes | | | | | Yes | | | |
| *Thismia okhaensis* | Yes | | Yes | | | | | | | | | |
| *Thismia puberula* | Yes | Yes | | | | | | | | | | |
| *Thismia thaithongiana* | | | Yes | | | | | | | | | |
| *Thismia viridistriata* | Yes | Yes | | | | Yes | | | | | | |

**Gene boundaries altered significantly (at least 21 bp)**

| | *rpl2* exon 1 | *3'-rps12* exon 2 | *rrn23* | *rrn16* | *rpl2* exon 2 | *rps4* | *rps2* | *trnE-UUC* | *rps18* | *rrn5* | *5'-rps12* exon 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Thismia angustimitra* | Yes | Yes | Yes | Yes | | | | | | | |
| *Thismia alba* | Yes | Yes | Yes | Yes | | | | | | | |
| *Thismia annamensis* | Yes | Yes | Yes | | | | | | | | |
| *Thismia cornuta* | Yes | Yes | Yes | | | | | | | | |
| *Thismia filiformis* | Yes | Yes | Yes | Yes | Yes | Yes | | | | | |
| *Thismia gardneriana* | Yes | Yes | Yes | Yes | | | | Yes | Yes | | |
| *Thismia hawkesii* | Yes | Yes | | | | Yes | | | | | |
| *Thismia hexagona* | Yes | Yes | | Yes | Yes | | | | | Yes | |
| *Thismia hongkongensis* | Yes | Yes | Yes | | | | Yes | | | | |
| *Thismia kelabitiana* | Yes | | | | | | | | | | |
| *Thismia lanternata* | Yes | Yes | Yes | | | Yes | | | | | |
| *Thismia mucronata* | Yes | | Yes | Yes | Yes | | Yes | | | | |
| *Thismia neptunis* | Yes | | | | Yes | | | | | | |
| *Thismia okhaensis* | Yes | | Yes | | | Yes | | | | | |
| *Thismia puberula* | Yes | Yes | Yes | Yes | Yes | | Yes | | | | |
| *Thismia thaithongiana* | Yes | Yes | Yes | | Yes | | | | | | Yes |
| *Thismia viridistriata* | Yes | Yes | Yes | | | | | | | | |

**Other Changes**

| | |
|---|---|
| *Thismia okhaensis* | *rpl2* exon 2 removed (false hit?) |
| *Thismia alba* | *rrn16* fragment potentially indicates IR presence |

**Table A4.** Putative pseudogene status of tRNA genes across representative Thismiaceae mycoheterotrophs and autotrophic relatives as determined by tRNAscan-SE (ignoring the possible role of *trnE-UUC* in haem biosynthesis; see text). Orange-coloured rows indicate predicted pseudogenes.

| Species | Gene | Predicted Pseudogene[1] | Predicted Clover-folding | Sequence |
|---|---|---|---|---|
| *Tacca leontopetaloides* | *trnE-UUC* | No | Yes | GCCCCTATCGTCTAGTGGTTCAGGACATCTCTCTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Haplothismia exannulata* | *trnE-UUC* | No | Yes | GCCCCTATTGTCTAGTGGTTTAGGACATCTCCCTTTCGAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia huangii* | *trnE-UUC* | Yes | No | GCCCCTATAGTCTAGTGGTTTAGGACACCTTATATATTAATTAAAGAAGGTAACGTGGATTCAACTTCCTCTAGGGGT |
| *Thismia javanica* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia panamensis* | *trnE-UUC* | Yes | Yes | GCCCCTATCGTCTAATGGTTAAGGACATCTCTTTTTCGAGGAGATAACGTGGATTCGATTTCTACTAGGGGTT |
| *Thismia rodwayi* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCAACTTCCCCTAGGGGTT |
| *Thismia tentaculata* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGGGGCAGCGGGGATTCGACTTCTCCTAGGGGTA |
| *Thismia angustimitra* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGGGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia alba* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCGACGGGGATTCGACTTCCCTTAGGGGTT |
| *Thismia annamensis* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia cornuta* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia filiformis* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTT |
| *Thismia gardneriana* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia hawkesii* | *trnE-UUC* | Yes | Yes | GCCCTTATCGTCTAGTGTGGGTCAGGACATCTCTTTTTCAAGGAGGCAACGGGGATTCAACTTCCCCTAGGGGTA |
| *Thismia hexagona* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTT |
| *Thismia hongkongensis* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTGGGGGTT |
| *Thismia kelabitiana* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGAAAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia lanternata* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia mucronata* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia neptunis* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia okhaensis* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia puberula* | *trnE-UUC* | No | Yes | GCCCTTATCGTCTAGTGGTTCAGGACATCTCCTTTTCAAGGAGGCAGCGGGGATTCGACTTCCCCTAGGGGTA |
| *Thismia thaithongiana* | *trnE-UUC* | No | Yes | GCCCCTATAGTCTAGTGGCTCAGGACACCTCTTTTTCAAAGAGGCAACGGGGATTCGACTTCCCCTAGGGGTT |
| *Thismia viridistriata* | *trnE-UUC* | No | Yes | GCCCTTATAGTCTAGTGTGGTTCAGGACACCTCCTTTTCAAGGAGGCAACGGGGATTCAACTTCCCCTAGGGGTT |
| *Tacca leontopetaloides* | *trnfM-CAU* | No | Yes | CGCGGAGTAGAGCAGTTTGGTAGCTCGCAAGGCTCATAACCTTGAGGTCACGGGTTCAAATCCTGTCTCCGCAA |
| *Haplothismia exannulata* | *trnfM-CAU* | No | Yes | CGCGGAGTAGAGCAGTCTGGTAGCTCGCAAGGCTCATAACCTTGAGGCCGCGGGTTCGAATCCCGTCTCCGCAA |
| *Thismia javanica* | *trnfM-CAU* | No | Yes | TTGCGAAAACAGGATTTGAACCTGTGATTTCAAGGTTATGAGCCTTGCGAGCTTCCGTAAAACTACTCTATTTCGCA |
| *Thismia panamensis* | *trnfM-CAU* | No | Yes | TGCAGAGTGGAGCAATCTGGTAGCTCGTGAGTTTCATAGTCTTGAGGTTATGGGTTCAAATCCCGTCTCTGCAT |
| *Thismia rodwayi* | *trnfM-CAU* | Yes | No | TGCGAAATAGAGCAGTTTTACGGAAGCTCACAAGGCTCATAACCT |
| *Thismia tentaculata* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTTACGGAAACTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTTCGCAA |
| *Thismia angustimitra* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGTAATTTATTGAAGCTCACAAGGCTCATAACCTTGAAATTATAGGTTCAAATCCTGTTTTCGCAA |
| *Thismia alba* | *trnfM-CAU* | Yes | Yes | TGCGAAATAGAGCAGTTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTCAATAA |
| *Thismia cornuta* | *trnfM-CAU* | Yes | Yes | TGCGAAATAGAGCAGTTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTT |
| *Thismia filiformis* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTTCGCTT |
| *Thismia gardneriana* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTTCGTAA |
| *Thismia hawkesii* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAGTCACAGGTTCAAATCCTGTTTTCGCAA |
| *Thismia hexagona* | *trnfM-CAU* | Yes | Yes | TGCGAAATAGAGCAGTTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTCAATAA |
| *Thismia hongkongensis* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGTAGTGGCAACTCGCAAGGCTCATAACCTTGAGATCACAGGTTCAAGTCCTGTTTTCGCAT |
| *Thismia kelabitiana* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAATCATAGGTTCAAATCCTATTTTCGCAA |
| *Thismia lanternata* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTTATGGAAGCTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTTCGCAA |
| *Thismia mucronata* | *trnfM-CAU* | No | Yes | CGCGAAATAGAGCAGTTTCGGAAGCTTACAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTTCGCAA |
| *Thismia neptunis* | *trnfM-CAU* | Yes | Yes | TGCGAAATAGAGCAGTTTTACGGAAGCTCGCAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTCAATAA |
| *Thismia okhaensis* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTCGGAAGCTTACAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTTCGCAA |
| *Thismia puberula* | *trnfM-CAU* | No | Yes | TGCGAAATAGAGCAGTTTCGGAAGCTTACAAGGCTCATAACCTTGAAATCACAGGTTCAAATCCTGTTTTCGCAA |
| *Tacca leontopetaloides* | *trnI-CAU* | No | Yes | GCATCCATGGCTGAATGGTTAAAGCGCCCAACTCATAATTGGCAAATTCGTAGGTTCAATTCCTGCTGGATGCA |
| *Haplothismia exannulata* | *trnI-CAU* | No | Yes | GCATCCATGGCTGAATGGTTAAAGCGACCAACTCATAATTGGTAAATTCGTAGGTTCGATTCCTTCTGGATGCA |
| *Tacca leontopetaloides* | *trnM-CAU* | No | Yes | ACCTACTTAACTCAGTGGTTAGAGTATTGCTTTCATACGGCAGGAGTCATTGGTTCAAATCCAATAGTAGGTA |
| *Haplothismia exannulata* | *trnM-CAU* | Yes | Yes | CCCTGCTAAACTCAGTGGTTAGAGTATTGTTTTCATACGACAAGAGACATTGGTTCGAATCCAATAGTAGGTT |
| *Thismia panamensis* | *trnM-CAU* | No | Yes | GCATCTATGACTGAATGGATAAAGTGCTCAACTCATGATTGAGTTTTTGCGGGTTCAATTCCTGCTAGATGCA |
| *Tacca leontopetaloides* | *trnQ-UUG* | No | Yes | TGGGGTGTGGCCAAGTGGTAAGGCAGCGGGTTTTGGTCCCGTTACTCGGAGGTTCGAATCCTTCCGTCCCAG |
| *Haplothismia exannulata* | *trnQ-UUG* | No | Yes | TGGGCCGTGGCCGAGTGGTAAGGCAGCAGGTTTTGGTCCCGCTATTCGTAGGTTCGAATCCTTCCGTCCCAG |
| *Tacca leontopetaloides* | *trnW-CCA* | No | Yes | GCGCTCTTAGTTCAGTTCGGTAGAACACGGGTCTCCAAAACCCGATGTCGTAGGTTCAAATCCTACAGAGCGTG |
| *Haplothismia exannulata* | *trnW-CCA* | No | Yes | GCGTTCTTAGTTCAGTTCGGTAGAACGATGGTCTCCAAAACCCGACGTCGTAGGTTCAAATCCTACAGAGCGCG |
| *Tacca leontopetaloides* | *trnY-GUA* | No | Yes | GGGTCGATGCCCGAGTGGTTAATGGGGACGGACTGTAAATTCGTTGGCGATATGTCTACGCTGGTTCAAATCCAGCTCGGCCCA |
| *Haplothismia exannulata* | *trnY-GUA* | Yes | No | GTAAAAGGGGACCGACTGTAAATTCGTCGGTGATATGTCTACGCTGTTTCGAACCCAGCTCGGCCCA |

1. Potential pseudogenes based on structure and anticodon (see text); note that the *trnE-UUC* instances may be functional (with secondary functions outside translation)

**Table A5.** "Heat-map" of gene loss and retention across representative Thismiaceae mycoheterotrophs and autotrophic relatives. For newly assembled Thismiaceae bright blue blocks indicate retained genes and light blue blocks with the Ψ-symbol indicate a putative pseudogene, for *T. tentaculata* from Lim *et al*. (2016), green blocks indicate retained genes, and for Thismiaceae from Yudina *et al*. (2021), orange blocks indicate retained genes and pastel orange blocks with the Ψ-symbol indicate a putative pseudogene. Grey blocks with question marks indicate unfound genes from incomplete plastid genomes where the gene is predicted to be in the missing sequence area, and white blocks indicate missing genes. An asterisk ('*') indicates an intron is present.

| | | accD | clpP | infA | matK | rpl14 | rpl16 | rpl2 | rpl20 | rps11 | 5'-rps12[5] | 3'-rps12[5] | rps14 | rps18 | rps19 | rps2 | rps3 | rps4 | rps7 | rps8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 22,294 bp | *Haplothismia exannulata* | | * [3] | | Ψ | | | * | | | | * | | | | | | | | |
| 17,223 bp | *Thismia javanica* | | | | | | | * | | | [6] | | | | | Ψ | | | | |
| 18,553 bp | *Thismia rodwayi* | | | | | | | * | | | [7] | | | | | | | | | |
| 14,832 bp | *Thismia panamensis* | | | | | | [4] | | | | * | | | | | | | | | |
| 9742 bp | *Thismia huangii*[2] | | | | ?[2] | | | * | | | [6] | | ?[2] | | | | ?[2] | | ?[2] | |
| 16,031 bp | *Thismia tentaculata*[1] | | | | | | | * | | | [6] | | | | | | | | | |
| 17,480 bp | *Thismia filiformis*[12] | | | | | | | * | | | separated | | | | | | | Ψ | | |
| 17,645 bp | *Thismia gardneriana*[12] | | | | | | | * | | | separated | | | | | | Ψ | | | Ψ |
| 17,312 bp | *Thismia hawkesii*[12] | | | | | | | * | | | separated | | | | | | | | | |
| 18,428 bp | *Thismia lanternata*[12] | | | | | | | * | | | separated | | | | | | | Ψ | | |
| 18,718 bp | *Thismia mucronata*[12] | | | | | | | * | | | separated | | | | | | | | | |
| 18,768 bp | *Thismia puberula*[12] | | | | | | | * | | | separated | | | | | | | | | |
| 14,060 bp | *Thismia thaithongiana*[12] | | | | | | | * | | | separated | | | | | | Ψ | Ψ | | Ψ |
| 10,127 bp | *Thismia alba*[12, 13] | | | | | | | * | | | separated | | | | | | ? | | | partial |
| 18,204 bp | *Thismia angustimitra*[12, 13] | | | | | | | * | | | separated | | | | | | | | | |
| 10,641 bp | *Thismia annamensis*[12, 13] | | | | | | | exon 2 not found | | | separated | | | | | | ? | ? | | ? |
| 6578 bp | *Thismia cornuta*[12, 13] | | | | | | | exon 2 not found | | | separated | | | | | | ? | ? | | ? |
| 5904 bp | *Thismia hexagona*[12, 13] | | | | | | | * | | | separated | | | | | | ? | ? | | ? |
| 14,791 bp | *Thismia hongkongensis*[12, 13] | | | | | | | exon 2 not found | | | separated | | | | | Ψ | | | | |
| 8966 bp | *Thismia kelabitiana*[12, 13] | | | | | | | exon 2 not found | | | separated | | | | | | partial | | | ? |
| 7687 bp | *Thismia neptunis*[12, 13] | | | | | | | * | | | ? | | | | | partial | ? | ? | | ? |
| 8120 bp | *Thismia okhaensis*[12, 13] | | | | | | | exon 2 not found | | | exon 2 not found | | | | | | ? | partial | | ? |
| 8324 bp | *Thismia viridistriata*[12, 13] | | | | | | | exon 2 not found | | | separated | | | | | | ? | ? | | ? |

| | | rrn16 | rrn23 | rrn4.5 | rrn5 | trnE-UUC[10] | trnfM-CAU[10] | trnI-CAU | trnM-CAU[10] | trnQ-UUG | trnW-CCA | trnY-GUA[10] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 22,294 bp | *Haplothismia exannulata* | | | | | | | | Ψ | | | Ψ |
| 17,223 bp | *Thismia javanica* | | | | | | | | | | | |
| 18,553 bp | *Thismia rodwayi* | | | | | | Ψ | | | | | |
| 14,832 bp | *Thismia panamensis* | | [8] | [8] | | ?[11] | | | | | | |
| 9742 bp | *Thismia huangii*[2] | | [9] | | | ?[11] | | ?[2] | | ?[2] | | ?[2] |
| 16,031 bp | *Thismia tentaculata*[1] | | | | | | | | | | | |
| 17,480 bp | *Thismia filiformis*[12] | | | | | | | | | | | |
| 17,645 bp | *Thismia gardneriana*[12] | | | | | | | | | | | |
| 17,312 bp | *Thismia hawkesii*[12] | | | | | ?[11] | | | | | | |
| 18,428 bp | *Thismia lanternata*[12] | | | | | | | | | | | |
| 18,718 bp | *Thismia mucronata*[12] | | | | | | | | | | | |
| 18,768 bp | *Thismia puberula*[12] | | | | | | | | | | | |
| 14,060 bp | *Thismia thaithongiana*[12] | | | | | | | | | | | |
| 10,127 bp | *Thismia alba*[12, 13] | | | | | | Ψ | | | | | |
| 18,204 bp | *Thismia angustimitra*[12, 13] | | | | | | | | | | | |
| 10,641 bp | *Thismia annamensis*[12, 13] | | | | | | ? | | | | | |
| 6578 bp | *Thismia cornuta*[12, 13] | partial | partial | | | | Ψ | | | | | |
| 5904 bp | *Thismia hexagona*[12, 13] | partial | ? | | | | Ψ | | | | | |
| 14,791 bp | *Thismia hongkongensis*[12, 13] | | | | | | | | | | | |
| 8966 bp | *Thismia kelabitiana*[12, 13] | ? | partial | | | | | | | | | |
| 7687 bp | *Thismia neptunis*[12, 13] | ? | ? | | ? | | Ψ | | | | | |
| 8120 bp | *Thismia okhaensis*[12, 13] | | partial | | ? | | | | | | | |
| 8324 bp | *Thismia viridistriata*[12, 13] | ? | partial | | | | ? | | | | | |

1. *T. tentaculata* data modified from (Lim et al., 2016)
2. *T. huangii* is a partial plastome. The grey filled boxes indicate genes expected to be found outside the partial sequence area based on gene order comparison with the other Thismiaceae species
3. clpP exon 2 and 3 are merged in *H. exannulata*
4. rpl2 exon 1 and 2 are merged in *T. panamensis*
5. 5'-rps12 and 3'-rps12 are trans-spliced
6. The two exons of 3'-rps12 are separated and 5'-rps12 exon 1 and 3'-rps12 exon 3 are adjacent in *T. tentaculata*, *T. javanica*, and *T. huangii*
7. 3'-rps12 exon 2 and 3 are both present consectutively but are in the wrong order for the orientation of the reading frames in *T. rodwayi*
8. rrn23 and rrn4.5 may be merged in *T. panamensis* (there is no intergenic sequence between them)
9. *T. huangii* is only a partial plastome sequence and rrn23 is present in one fragment and one incomplete gene (cut off so it may or may not be a complete gene in the full plastome)
10. Putative pseudogenic tRNAs were identified using tRNAscan-SE to predict tRNA folding and determine if primary and secondary structure were within typical bounds (Lowe et al., 2016)
11. *trnE-UUC* functions both as a tRNA and as a precursor in haem biosynthesis. In *T. panamensis*, *T. huangii*, and *T. hawkesii* tRNAscan-SE identifies *trnE-UUC* as a pseudogene with regard to tRNA function, but it is unclear whether it may still be functional in haem biosynthesis
12. *T. filiformis*, *T. gardneriana*, *T. hawkesii*, *T. lanternata*, *T. mucronata*, *T. puberula*, *T. thaithoniana*, *T. alba*, *T. angustimitra*, *T. annamensis*, *T. cornuta*, *T. hexagona*, *T. hongkongensis*, *T. kelabitiana*, *T. neptunis*, *T. okhaensis*, and *T. viridistriata* are modified from (Yudina et al., 2021)
13. *T. alba*, *T. angustimitra*, *T. annamensis*, *T. cornuta*, *T. hexagona*, *T. hongkongensis*, *T. kelabitiana*, *T. neptunis*, *T. okhaensis*, and *T. viridistriata* are partial plastomes. The grey filled boxes indicate genes expected to be found outside the partial sequence area based on gene order comparison with the other Thismiaceae species

**Table A6.** Retention of group IIA introns in heterotrophic species lacking *matK*. White blocks with an 'x' indicate the gene as a whole has been lost (or likely pseudogenized); those without an 'x' indicate the intron has been lost in a retained gene. Bright blue blocks indicate intact genes with retained introns (despite *matK* loss); the light blue block with an asterisk indicates uncertainty due to non-canonical exon order (see text).

| | | | Group IIA introns | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **Taxon** | *clpP* -intron 2[3] | trnA-UGC | *trnI-GAU* | *trnK-UUU* | *trnV-UAC* | *atpF* | *rpl2* | *3'-rps12* |
| **Parasitic plants** | Apodanthaceae | *Pilostyles aetiopica* [1] | x | x | x | x | x | x | x | x |
| | | *Pilostyles hamiltonii* [1] | x | x | x | x | x | x | x | |
| | Convolvulaceae | *Cuscuta gronovii* [1] | | x | x | x | x | | | |
| | | *Cuscuta obtusiflora* [1] | | x | x | x | x | | | |
| | Cynomoriaceae | *Cynomorium coccineum* [1] | | x | x | x | x | x | | |
| | Cytinaceae | *Cytinus hypocistis* [1] | | x | x | x | x | x | | |
| | Hydnoraceae | *Hydnora visseri* [1] | x | x | x | x | x | x | | |
| **Mycoheterotrophs** | Corsiaceae | *Arachnitis uniflora* [1] | | x | x | x | x | x | | |
| | Orchidaceae | *Epipogium aphyllum* [1] | | x | x | x | x | x | | x |
| | | *Epipogium roseum* [1] | | x | x | x | x | x | | x |
| | | *Rhizanthella gardneri* [1] | | x | x | x | x | x | | x |
| | Thismiaceae | *Haplothismia exannulata* | | x | x | x | x | x | | |
| | | *Thismia panamensis* | x | x | x | x | x | x | | |
| | | 11 other *Thismia* species[2] | x | x | x | x | x | x | | * |

1: Data from summary in Graham et al (2017), Table 1.

2: The eleven *Thismia* are *T. rodwayi* , *T. javanica* , *T. huangii* , *T. tentaculata* , *T. filiformis* , *T. gardneriana* , *T. hawkesii* , *T. lanternata* , *T. mucronata* , *T. puberula* , and *T. thaithongiana*

3: *clpP* intron 2 is a group IIA intron that does not require *matK* for splicing (Zoschke et al., 2010)

*: Exon 2 and exon 3 of *3'-rps12* are not in canonical order, and thus it is unclear if *3'-rps12* intron 2 is present

**Table A7.** Variation in ω ratio between lineages. LRT is the Likelihood Ratio Test; FDR is the False Discovery Rate (calculated using the Benjamini-Hochberg method). Statistically significant findings in **bold**. Findings are significant if the *P*-value is less than the FDR.

| *Species* | *Gene* | *Background ω* | *Foreground ω* | *Branch LRT* | *Branch P-value* | *Branch FDR* |
|---|---|---|---|---|---|---|
| *Haplothismia exannulata* | *accD* | 0.2979 | 0.2397 | 1.5846 | 0.2081 | 0.0111 |
| | ***clpP*** | **0.1268** | **0.3095** | **10.1104** | **0.0015** | **0.0028** |
| | *infA* | 0.0597 | 0.1179 | 1.5754 | 0.2094 | 0.0139 |
| | *matK* | 0.4098 | 0.5158 | 1.9204 | 0.1658 | 0.0083 |
| | *rpl14* | 0.1029 | 0.1181 | 0.1189 | 0.7302 | 0.0389 |
| | *rpl16* | 0.1556 | 0.1943 | 0.4434 | 0.5055 | 0.0222 |
| | *rpl2* | 0.2514 | 0.2457 | 0.0056 | 0.9403 | 0.0500 |
| | *rpl20* | 0.3322 | 0.509 | 1.1821 | 0.2769 | 0.0167 |
| | *rps11* | 0.117 | 0.1342 | 0.1730 | 0.6775 | 0.0333 |
| | *rps12* | 0.1427 | 0.1145 | 0.2429 | 0.6221 | 0.0306 |
| | *rps14* | 0.2487 | 0.2137 | 0.1433 | 0.7050 | 0.0361 |
| | *rps18* | 0.2264 | 0.3362 | 1.1155 | 0.2909 | 0.0194 |
| | *rps19* | 0.2466 | 0.2793 | 0.0936 | 0.7596 | 0.0417 |
| | *rps2* | 0.1669 | 0.3295 | 7.2352 | 0.0071 | 0.0056 |
| | *rps3* | 0.1577 | 0.1627 | 0.0121 | 0.9123 | 0.0472 |
| | *rps4* | 0.2004 | 0.233 | 0.2735 | 0.6010 | 0.0278 |
| | *rps7* | 0.1924 | 0.2255 | 0.0864 | 0.7688 | 0.0444 |
| | *rps8* | 0.2424 | 0.2912 | 0.2931 | 0.5882 | 0.0250 |
| | | | | | | |
| *Thismia filiformis* | *accD* | 0.3026 | 0.3108 | 0.0187 | 0.8912 | 0.0500 |
| | *rpl2* | 0.2556 | 0.2745 | 0.0577 | 0.8102 | 0.0417 |
| | *rps12* | 0.1344 | 0.1029 | 0.4294 | 0.5123 | 0.0250 |
| | *rps2* | 0.1582 | 0.3051 | 5.6447 | 0.0175 | 0.0083 |
| | *rps4* | 0.1945 | 0.243 | 0.1880 | 0.6646 | 0.0333 |
| | *rps8* | 0.2401 | 0.1678 | 0.4522 | 0.5013 | 0.0167 |
| | | | | | | |
| *Thismia hawkesii* | *accD* | 0.3014 | 0.3236 | 0.1345 | 0.7138 | 0.0429 |
| | *rpl2* | 0.255 | 0.3463 | 1.1254 | 0.2888 | 0.0143 |
| | *rps12* | 0.1328 | 0.1119 | 0.1801 | 0.6713 | 0.0286 |
| | *rps18* | 0.224 | 0.1816 | 0.2291 | 0.6322 | 0.0214 |
| | *rps3* | 0.1667 | 0.139 | 0.1355 | 0.7128 | 0.0357 |
| | *rps4* | 0.1945 | 0.1833 | 0.0206 | 0.8858 | 0.0500 |
| | *rps8* | 0.2404 | 0.1425 | 1.7741 | 0.1829 | 0.0071 |
| | | | | | | |
| *Thismia huangii* | *accD* | 0.295 | 0.2284 | 1.4504 | 0.2285 | 0.0250 |
| | *rpl2* | 0.2601 | 0.254 | 0.0051 | 0.9432 | 0.0500 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | *rps12* | 0.1305 | 0.2323 | 2.0123 | 0.1560 | 0.0125 |
| | *rps18* | 0.226 | 0.208 | 0.0147 | 0.9036 | 0.0375 |
| | Gene | Background ω | Foreground ω | Branch LRT | Branch p-value | Branch FDR |
| *Thismia javanica* | *accD* | 0.2997 | 0.3041 | 0.0057 | 0.9397 | 0.0500 |
| | *rpl2* | 0.2579 | 0.3015 | 0.2888 | 0.5910 | 0.0286 |
| | *rps12* | 0.1336 | 0.1282 | 0.0102 | 0.9196 | 0.0429 |
| | *rps18* | 0.2277 | 0.1832 | 0.2625 | 0.6084 | 0.0357 |
| | *rps3* | 0.1663 | 0.1218 | 0.3298 | 0.5658 | 0.0214 |
| | *rps4* | 0.1941 | 0.3153 | 1.1517 | 0.2832 | 0.0143 |
| | *rps8* | 0.2406 | 0.124 | 2.1231 | 0.1451 | 0.0071 |
| | | | | | | |
| *Thismia mucronata* | *accD* | 0.2996 | 0.3112 | 0.0381 | 0.8453 | 0.0375 |
| | *rpl2* | 0.2536 | 0.327 | 0.7885 | 0.3746 | 0.0250 |
| | *rps12* | 0.133 | 0.1101 | 0.2174 | 0.6410 | 0.0313 |
| | *rps18* | 0.2277 | 0.2283 | 0.0000 | 0.9950 | 0.0500 |
| | *rps2* | 0.1581 | 0.2491 | 2.7373 | 0.0980 | 0.0063 |
| | *rps3* | 0.1672 | 0.1076 | 0.8159 | 0.3664 | 0.0188 |
| | *rps4* | 0.1913 | 0.3767 | 2.7229 | 0.0989 | 0.0125 |
| | *rps8* | 0.2402 | 0.2361 | 0.0023 | 0.9617 | 0.0438 |
| | | | | | | |
| *Thismia panamensis* | ***accD*** | **0.2866** | **0.1157** | **7.6685** | **0.0056** | **0.0188** |
| | *rpl16* | 0.1457 | 0.1571 | 0.0487 | 0.8253 | 0.0438 |
| | ***rpl2*** | **0.2572** | **0.0569** | **10.3865** | **0.0013** | **0.0125** |
| | *rps11* | 0.1234 | 0.1176 | 0.0083 | 0.9274 | 0.0500 |
| | *rps12* | 0.1326 | 0.1085 | 0.2418 | 0.6229 | 0.0375 |
| | *rps19* | 0.2346 | 0.134 | 2.0738 | 0.1499 | 0.0250 |
| | *rps3* | 0.1656 | 0.1262 | 0.4671 | 0.4943 | 0.0313 |
| | ***rps4*** | **0.1924** | **0.0429** | **12.7793** | **0.0004** | **0.0063** |
| | | | | | | |
| *Thismia rodwayi* | *accD* | 0.3027 | 0.2719 | 0.3144 | 0.5750 | 0.0333 |
| | *rpl2* | 0.2599 | 0.3515 | 1.0809 | 0.2985 | 0.0250 |
| | *rps12* | 0.1322 | 0.1358 | 0.0044 | 0.9471 | 0.0500 |
| | *rps18* | 0.226 | 0.1379 | 1.1079 | 0.2925 | 0.0167 |
| | ***rps2*** | **0.1594** | **0.3247** | **6.9752** | **0.0083** | **0.0083** |
| | *rps3* | 0.166 | 0.1273 | 0.1490 | 0.6995 | 0.0417 |
| | | | | | | |
| *Thismia tentaculata* | *accD* | 0.2996 | 0.3363 | 0.3459 | 0.5564 | 0.0143 |
| | *rpl2* | 0.2569 | 0.2945 | 0.2246 | 0.6355 | 0.0214 |
| | *rps12* | 0.1342 | 0.1198 | 0.0791 | 0.7785 | 0.0357 |
| | *rps18* | 0.2275 | 0.2335 | 0.0037 | 0.9513 | 0.0429 |
| | *rps2* | 0.1578 | 0.2169 | 1.2861 | 0.2568 | 0.0071 |
| | *rps4* | 0.1937 | 0.2241 | 0.1232 | 0.7256 | 0.0286 |
| | *rps8* | 0.2398 | 0.2439 | 0.0022 | 0.9625 | 0.0500 |

| | Gene | Background ω | Foreground ω | Branch LRT | Branch p-value | Branch FDR |
|---|---|---|---|---|---|---|
| *Thismia thaithongiana* | *accD* | 0.2971 | 0.2065 | 3.4255 | 0.0642 | 0.0143 |
| | *rpl2* | 0.2573 | 0.2522 | 0.0037 | 0.9512 | 0.0500 |
| | *rps12* | 0.1319 | 0.1562 | 0.1732 | 0.6773 | 0.0429 |
| | *rps18* | 0.2214 | 0.4127 | 1.5384 | 0.2148 | 0.0214 |
| *Thismia thaithongiana* | *rps3* | 0.1661 | 0.0949 | 0.6707 | 0.4128 | 0.0357 |
| | *rps4* | 0.1937 | 0.3568 | 1.2151 | 0.2703 | 0.0286 |
| | *rps8* | 0.2411 | 0.0481 | 5.7987 | 0.0160 | 0.0071 |

.

**Figure A1.** Phylogenetic relationships in Thismiaceae and relatives inferred from a partitioned DNA-based likelihood analysis of 82 plastid genes, using a gene-by-codon partitioning scheme (see text for analysis details). Bootstrap values are indicated beside branches. The scale bar indicates estimated substitutions per site. The two long-branch Burmanniaceae lineages (*Apteria* and *Gymnosiphon*) are excluded (125-taxon set).

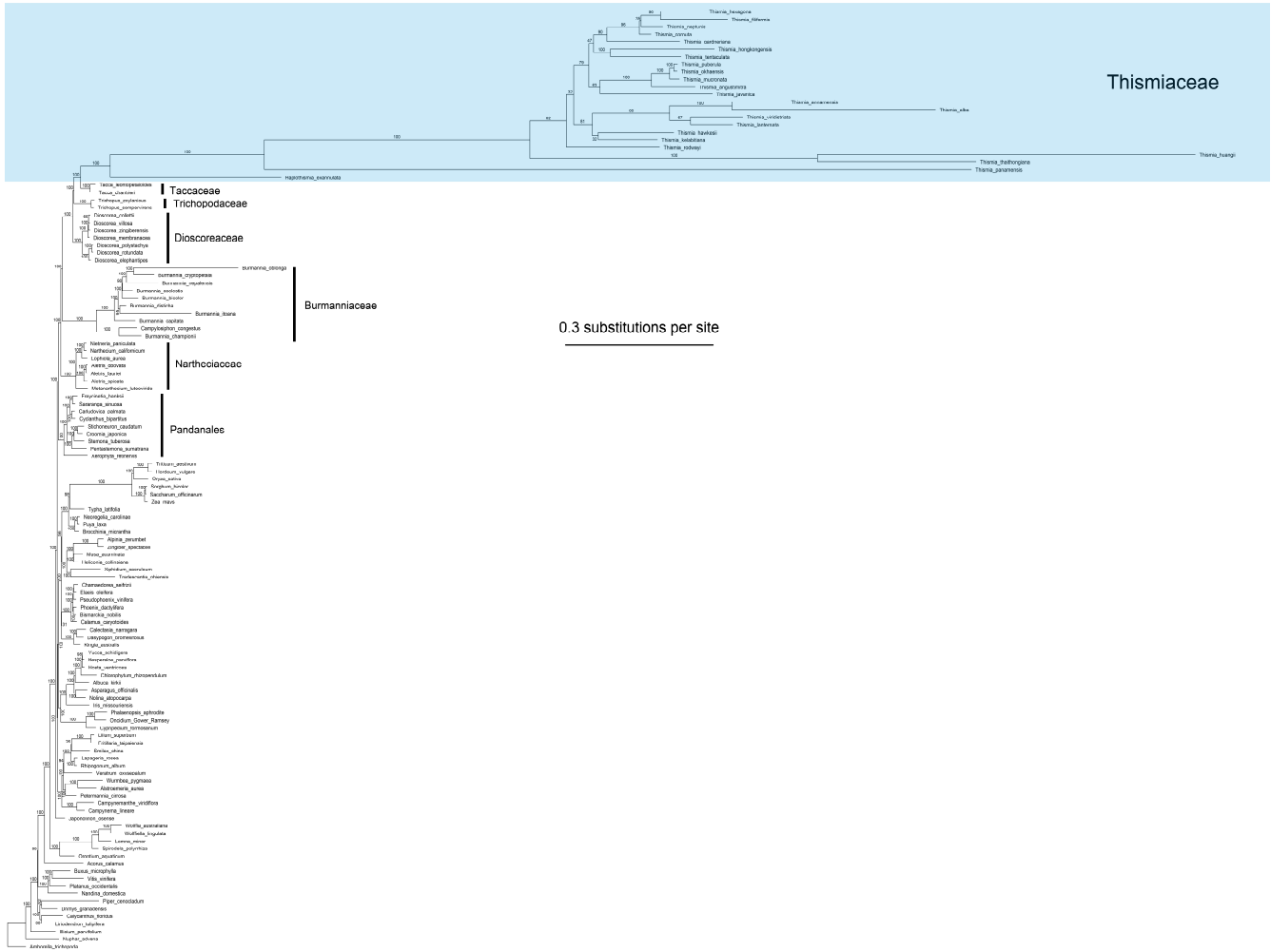


**Figure A2.** Phylogenetic relationships in Thismiaceae and relatives inferred from an unpartitioned DNA-based likelihood analysis of 82 plastid genes (see text for analysis details). Bootstrap values are indicated beside branches. The scale bar indicates estimated substitutions per site. The two long-branch Burmanniaceae lineages (*Apteria* and *Gymnosiphon*) are excluded (125-taxon set).

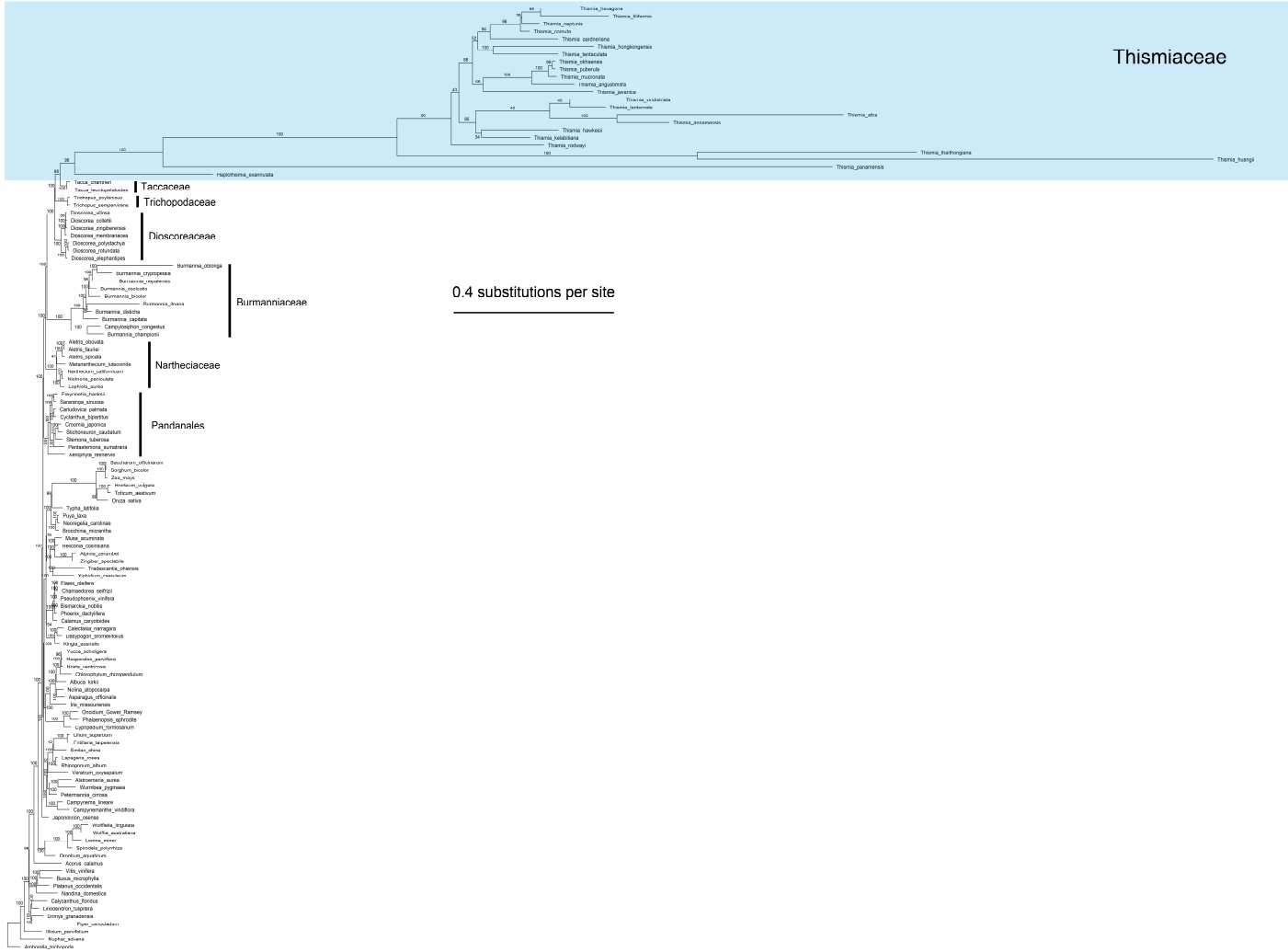


**Figure A3.** Phylogenetic relationships in Thismiaceae and relatives inferred from a partitioned DNA-based likelihood analysis of 82 plastid genes, using a gene-by-codon partitioning scheme (see text for analysis details). Bootstrap values are indicated beside branches. The scale bar indicates estimated substitutions per site. The two long-branch Burmanniaceae lineages (*Apteria and Gymnosiphon*) are included (127-taxon set).

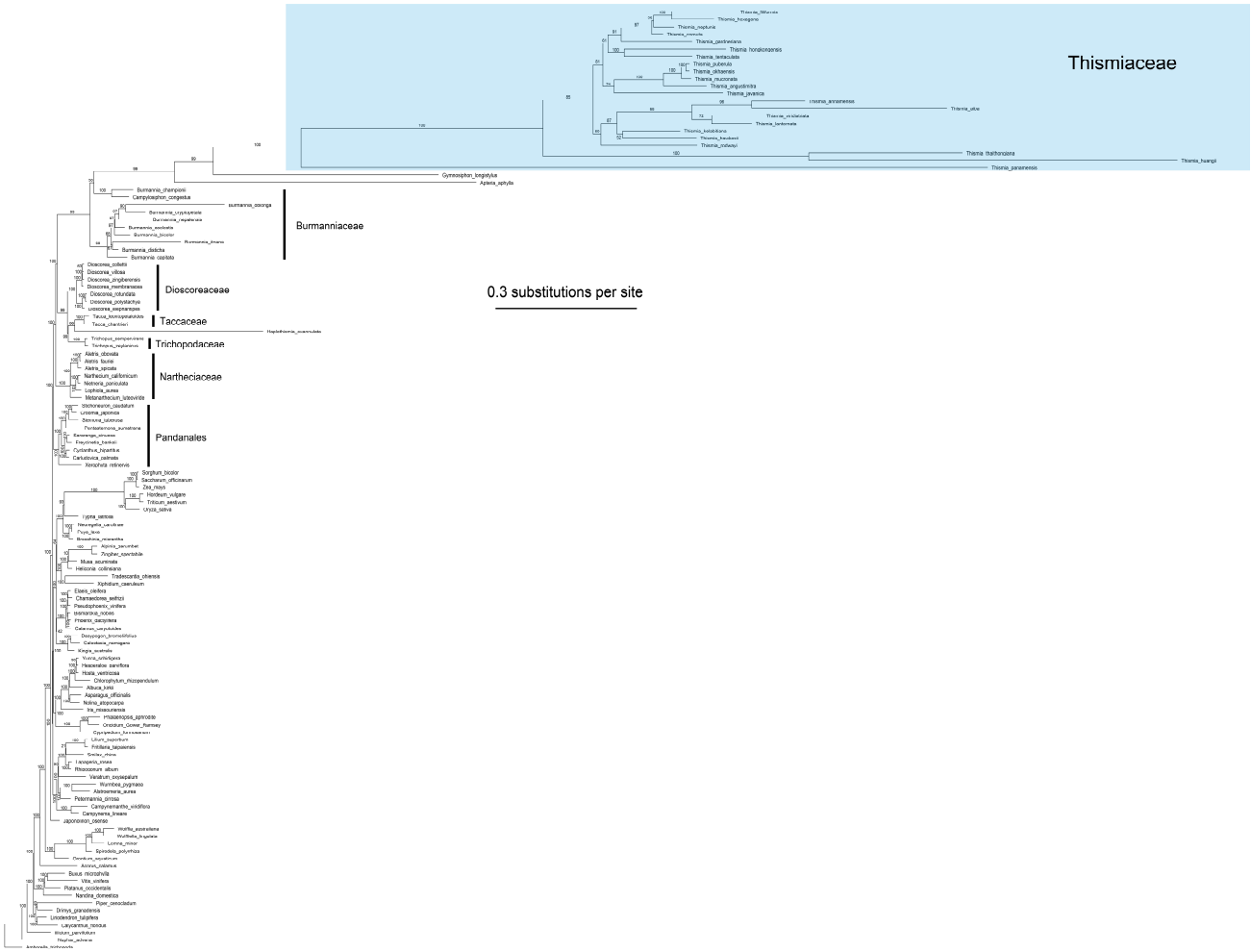


**Figure A4.** Phylogenetic relationships in Thismiaceae and relatives inferred from an unpartitioned DNA-based likelihood analysis of 82 plastid genes (see text for analysis details). Bootstrap values are indicated beside branches. The scale bar indicates estimated substitutions per site. The two long-branch Burmanniaceae lineages (*Apteria* and *Gymnosiphon*) are included (127-taxon set).

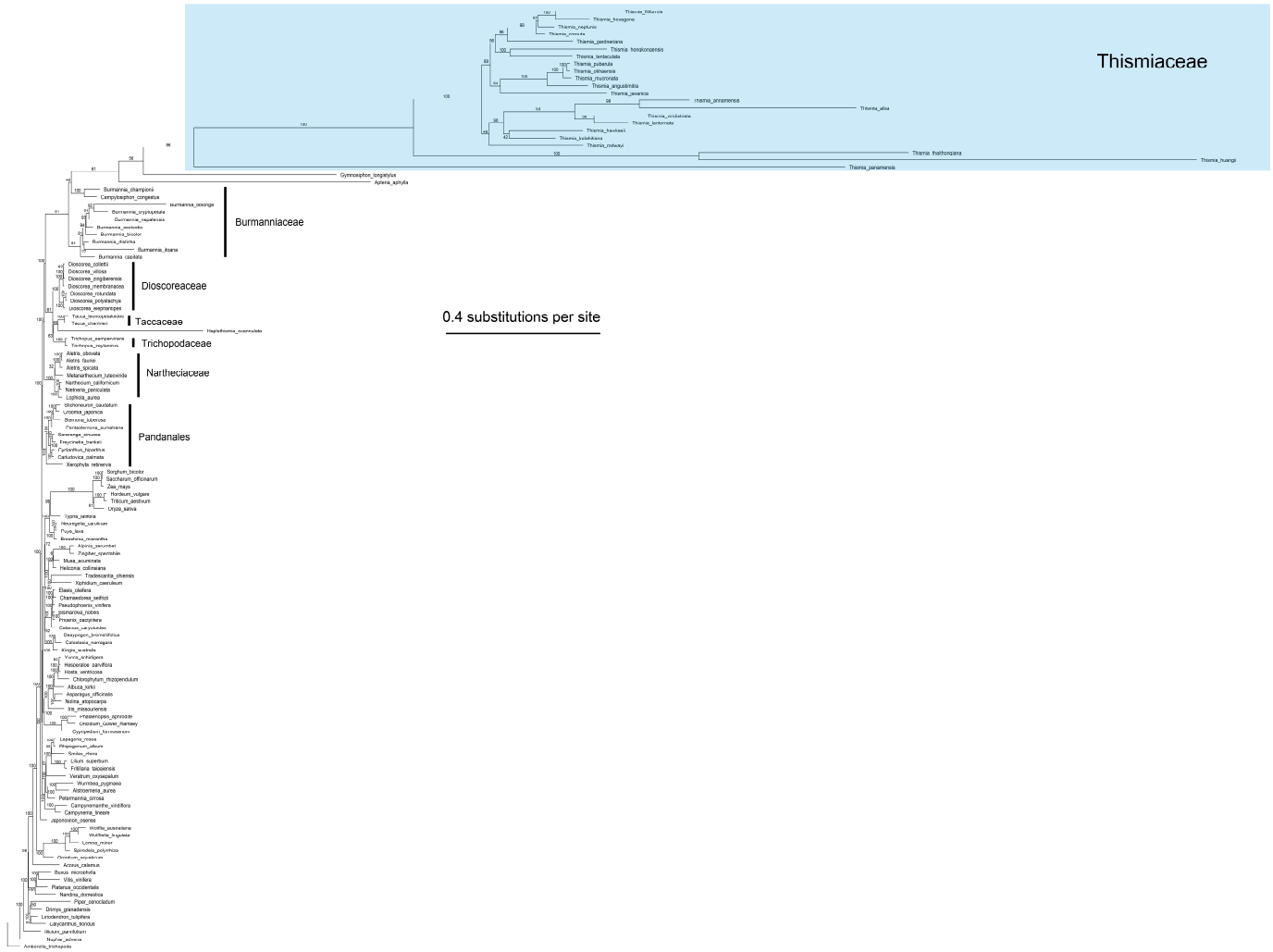


**Figure A5.** Plastid genome of *Haplothismia exannulata* presented as a circle. This plastid genome lacks an inverted repeat (IR). Grey arrows indicate the direction of transcription. White boxes indicate introns (dotted line in *clpP* indicates a missing intron), the Ψ-symbols putative pseudogenes (*matK, trnM-CAU, trnY-GUA*), and genes are coloured by major functional classes (noted in the caption). Grey bars in the centre represent average GC content across a sliding window.

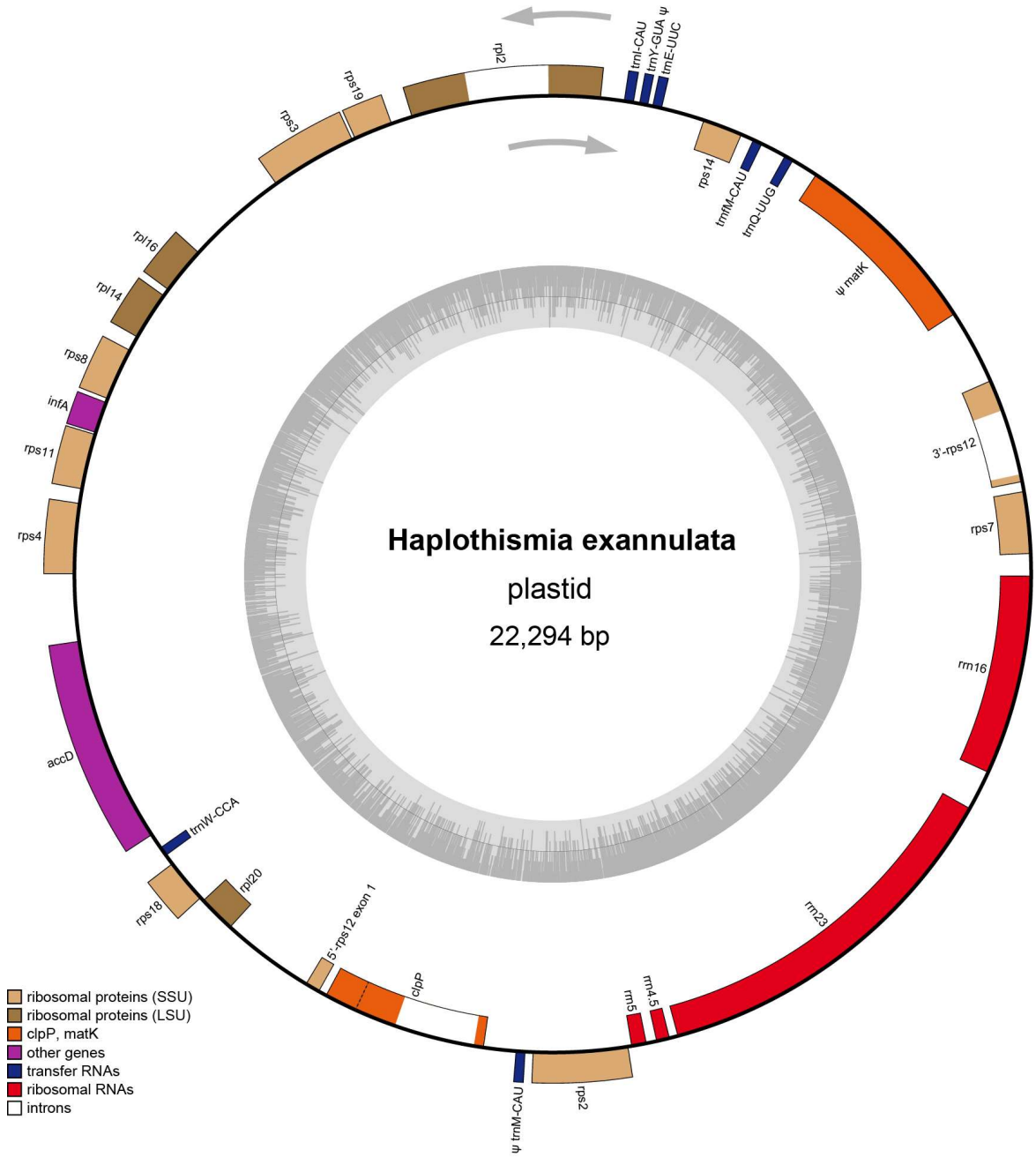


**Figure A6.** Plastid genome of *Thismia javanica* presented as a circle. Grey arrows indicate the direction of transcription. White box indicates an intron (note that this *rpl2* intron has *trnfM-CAU* inserted in it), dashed line and question mark ('?') a possible extended intron in 3'-*rps12* (i.e., *rps12* exons 2 and 3, separated by several genes), the Ψ-symbol a putative pseudogene (*rps2)* and genes are coloured by major functional classes (noted in the caption). LSC = large single-copy region, SSC = small single-copy region, IR = inverted repeat regions. Grey bars in the centre represent average GC content across a sliding window.

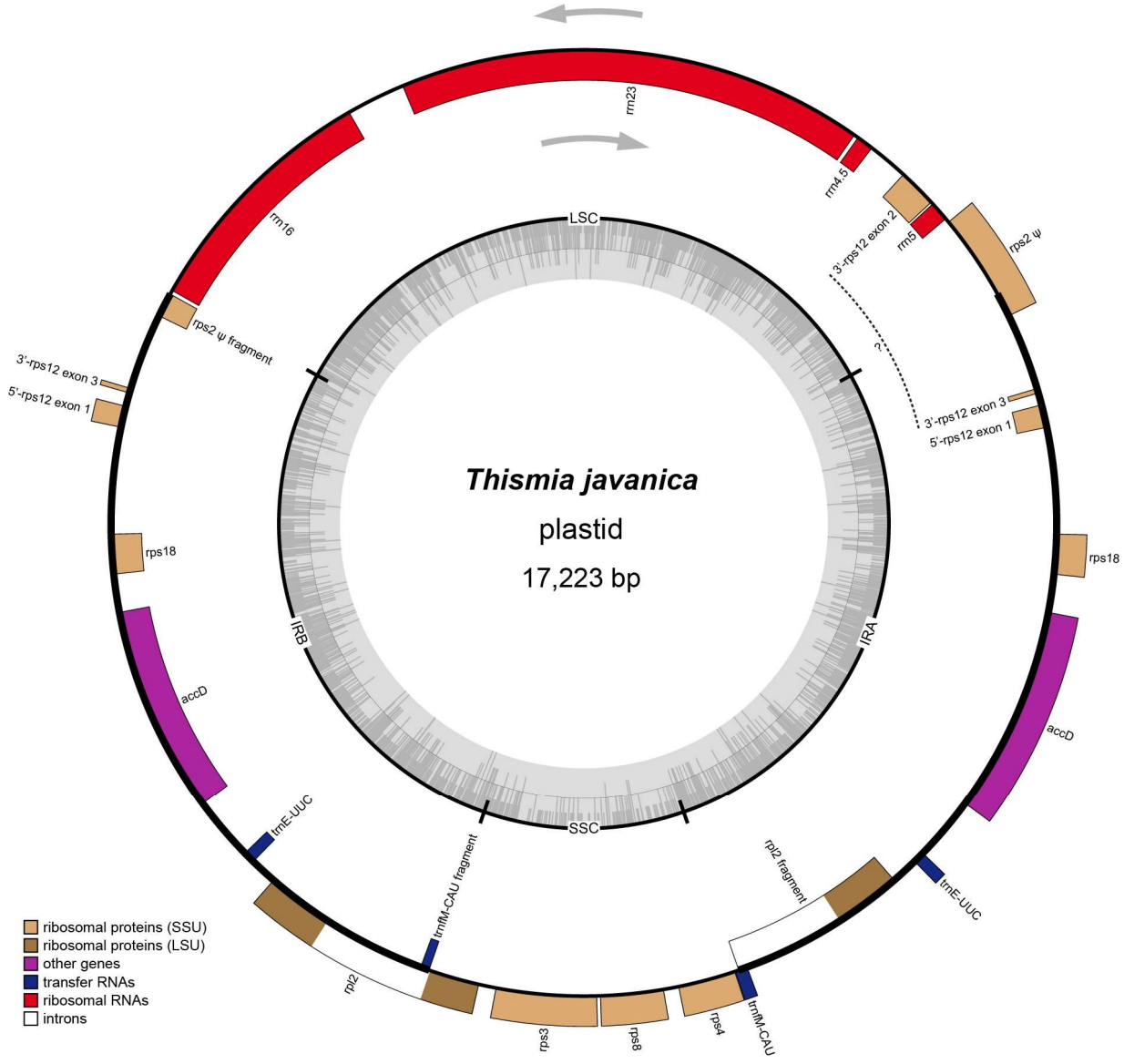


**Figure A7.** Plastid genome of *Thismia panamensis* presented as a circle. This plastid genome lacks an inverted repeat (IR). Grey arrows indicate the direction of transcription. White box indicates an intron (dotted line in *rpl2* indicates a missing intron), the Ψ-symbols putative pseudogenes (*accD, trnE-UUC*), and genes are coloured by major functional classes (noted in the caption). Grey bars in the centre represent average GC content across a sliding window.

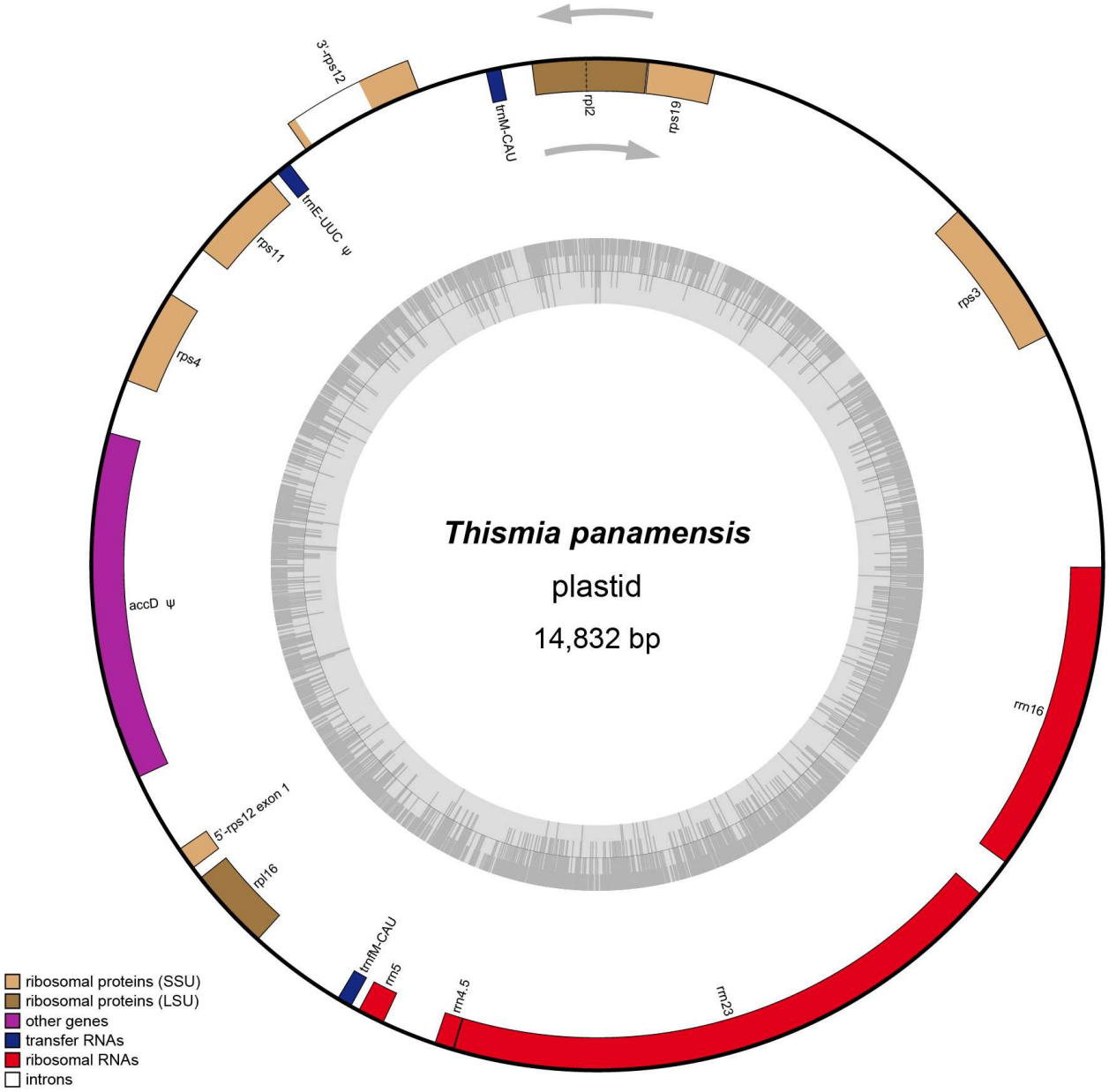


**Figure A8.** Plastid genome of *Thismia rodwayi* presented as a circle. Grey arrows indicate the direction of transcription. White box indicates an intron (note that this *rpl2* intron has *trnfM-CAU* inserted in it), the Ψ-symbols putative pseudogenes (rps23 [note *trnE-UUC* inserted between two fragments of this gene], *trnfM-CAU)* and genes are coloured by major functional classes (noted in the caption). LSC = large single-copy region, SSC = small single-copy region, IR = inverted repeat regions; the dashed line indicates the position of the single nucleotide SSC. Only one copy of *rps3* has a complete coding region (see Fig. A14). Grey bars in the centre represent average GC content across a sliding window.

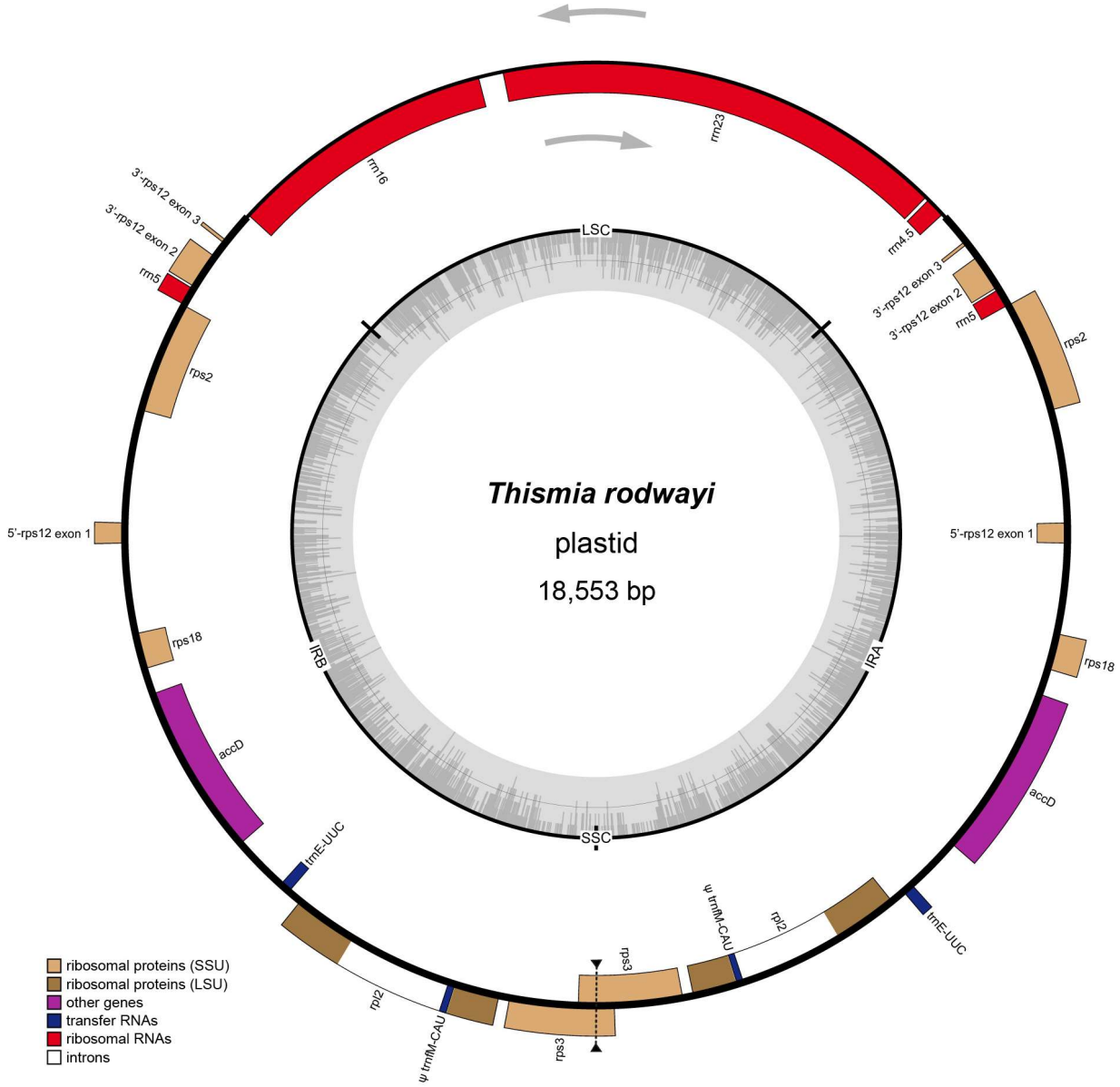


**Figure A9.** Incomplete plastid genome of *Thismia huangii*. Grey arrows indicate the direction of transcription. Black and green lines (including the dashed portion in the former, with a question mark) indicate potential inverted repeat (IR) region and small single copy regions, respectively. It is unclear whether *T. huangii* has an IR, this may be consistent with the two incomplete *rrn23* regions sequenced in this taxon (*rrn23* copy in *Tacca chantrieri* included to show the extent of the gene in a taxon with a full IR; this map implies that if *Thismia huangii* has an IR, at least one IR has a complete copy of the *rrn23* gene).

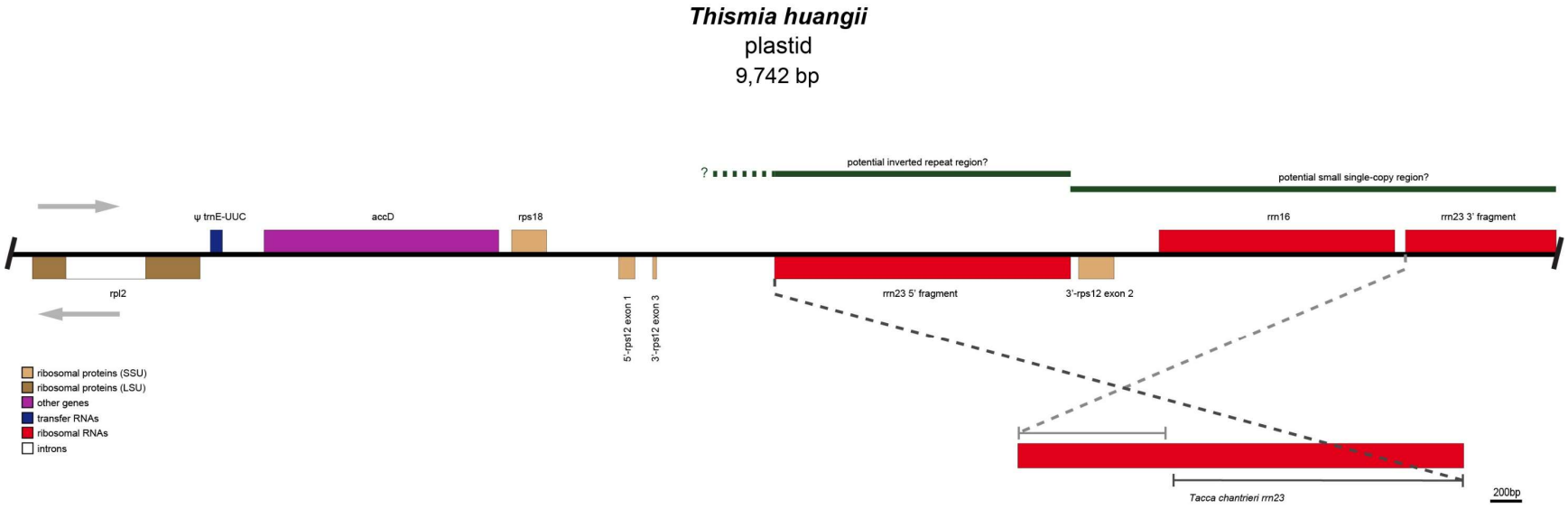


**Figure A10.** Plastid genome of *Thismia tentaculata* presented as a circle (this updates Lim et al., 2016; see Table A3). Grey arrows indicate the direction of transcription. White box indicates an intron (note that this *rpl2* intron has *trnfM-CAU* inserted in it), and genes are coloured by major functional classes (noted in the caption). LSC = large single-copy region, SSC = small single-copy region, IR = inverted repeat regions. Grey bars in the centre represent average GC content across a sliding window.

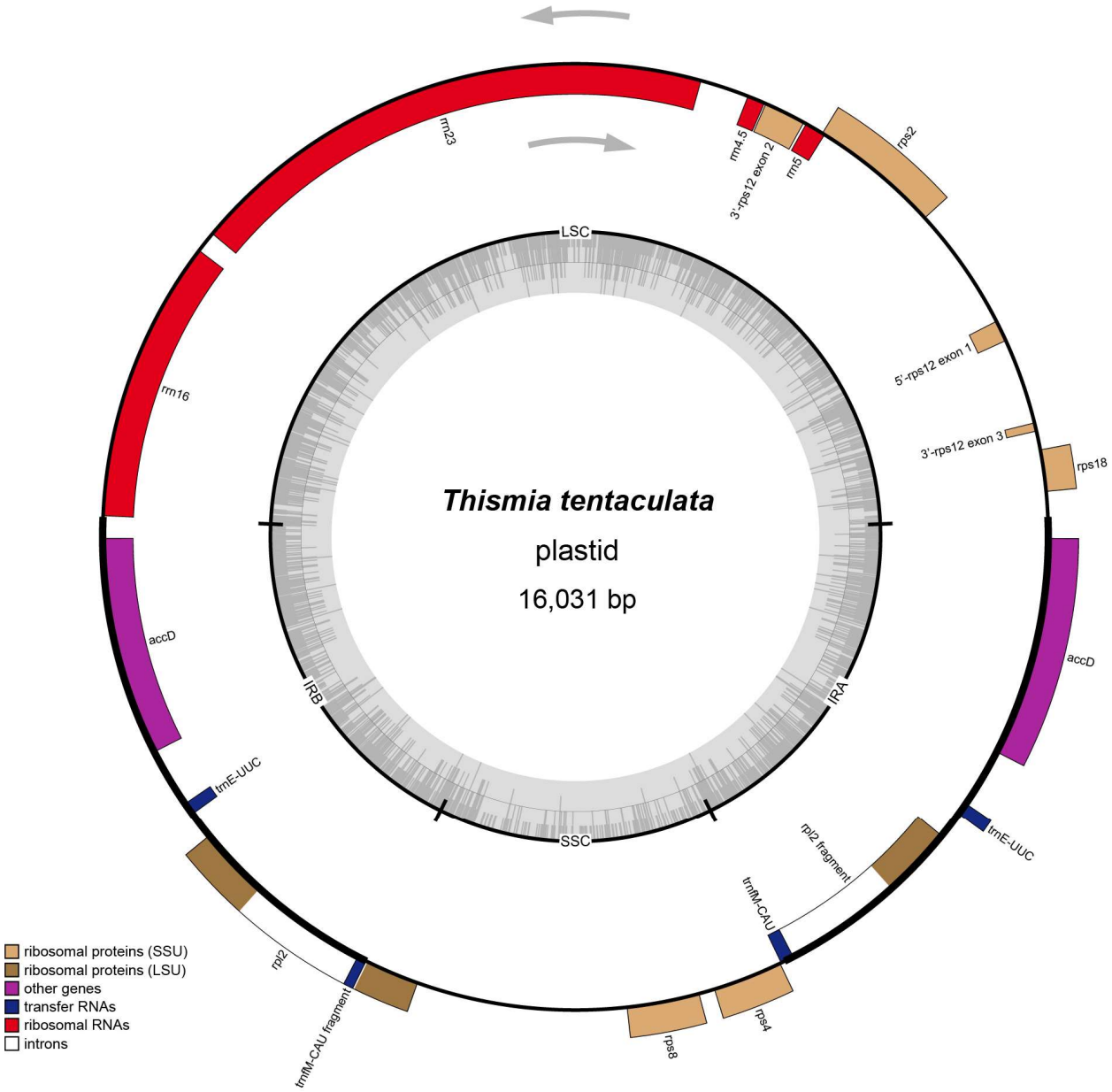


**Figure A11.** Mauve-based alignment comparing mycoheterotrophic *Haplothismia exannulata* to two autotrophic relatives (*Dioscorea elephantipes*, Dioscoreaceae; *Tacca chantrieri*, Taccaceae), with one copy of the inverted repeat removed (remaining copy highlighted with grey box). Coloured 'locally colinear blocks' (LCBs) have shared gene order between plastid genomes; (gene order for *Dioscorea* shown at top of figure for reference); coloured lines link LCBs shared between taxa. LCBs appearing above the central line for *H. exannulata* and *T. chantrieri* are colinear and in the same orientation as the *D. elephantipes* reference sequence; those below are in the reverse complement relative to the reference.

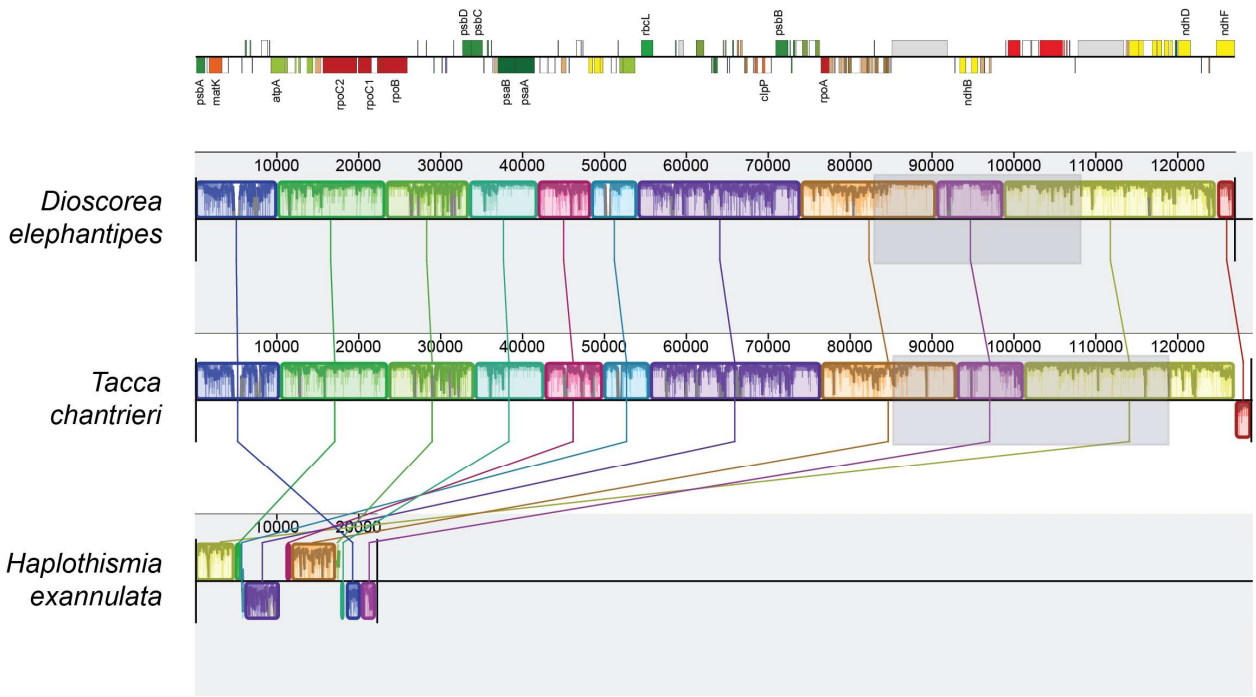


**Figure A12.** Mauve alignment comparing locally colinear sequences within available plastid genomes of genus *Thismia* (*Thismia panamensis*, *T. thaithongiana*, *T. huangii*, *T. rodwayi*, *T. kelabitiana*, *T. hawkesii*, *T. lanternata*, *T. viridistriata*, *T. alba*, *T. annamensis*, *T. javanica*, *T. angustimitra*, *T. mucronata*, *T. okhaensis*, *T. puberula*, *T. tentaculata*, *T. hongkongensis*, *T. gardneriana*, *T cornuta*, *T. neptunis*, *T. filiformis*, and *T. hexagona*). A linear gene map of *T. panamensis* appears first for reference. A single copy of the inverted repeat regions for *T. hawkesii*, *T. lanternata*, *T. javanica*, *T. angustimitra*, *T. mucronata*, *T. puberula*, *T. tentaculata*, *T. gardneriana*, and *T. filiformis* were included in this comparison and are represented by the grey boxes over portions of the strings. Coloured blocks represent 'locally colinear blocks' (LCBs) which have shared gene order between genomes. LCBs appearing above the central line for the *Thismia*s are colinear and in the same orientation as the *T. panamensis* reference sequence; those below are in the reverse complement relative to the reference. Coloured lines link orthologous LCBs shared between taxa. Large blank regions within an LCB and stretches between LCBs represent lineage-specific sequences. The double diagonal slashes indicate incomplete plastid genomes.

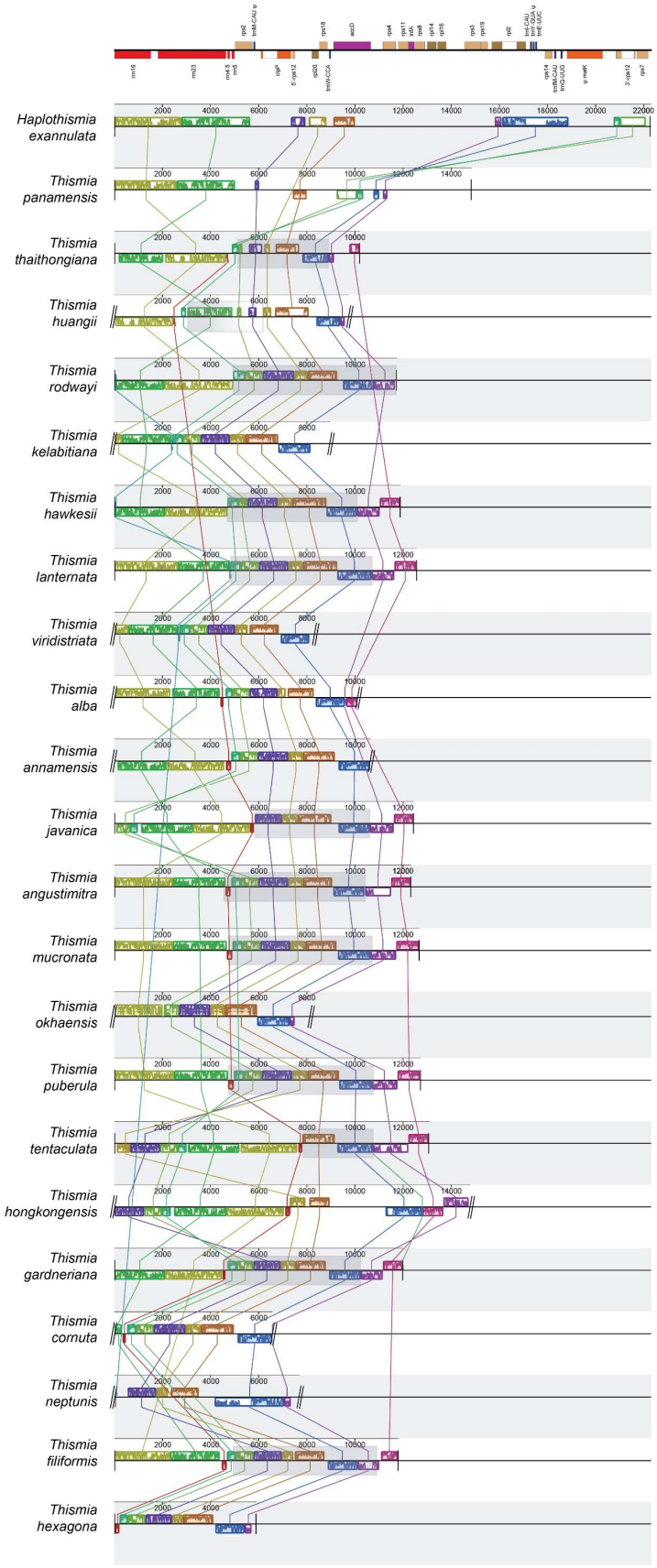

rpl2
trnK-CAU
rps18
accD
rps4
rps11
infA
rps8
rpoA
rpl16
rps3
rps19
rpl2
trnH-GUG
trnC-GCA
trnE-UUC
rrn16
rrn23
rrn4.5
rrn5
ctrP
5'-rps12
rpl20
trnV-CA
rps14
trnfM-CAU
trnQ-UUG
ψ rrn6K
3'-rps12
rps7
Haplothismia exannulata
Thismia panamensis
Thismia thaithongiana
Thismia huangii
Thismia rodwayi
Thismia kelabitiana
Thismia hawkesii
Thismia lanternata
Thismia viridistriata
Thismia alba
Thismia annamensis
Thismia javanica
Thismia angustimitra
Thismia mucronata
Thismia okhaensis
Thismia puberula
Thismia tentaculata
Thismia hongkongensis
Thismia gardneriana
Thismia cornuta
Thismia neptunis
Thismia filiformis
Thismia hexagona

**Figure A13.** Codon alignment of *rps3* in *Thismia rodwayi* and inverted repeat regions showing the self-overlapping nature of this gene in *T. rodwayi*. The magenta tag emphasizes the location of the single nucleotide small single-copy region (SSC).

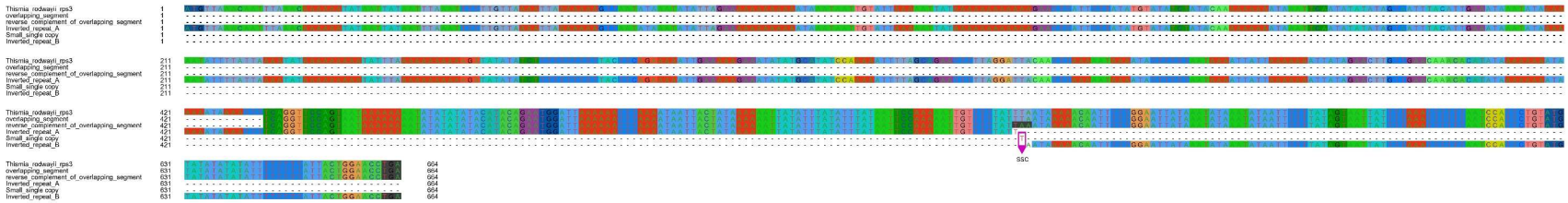


**Figure A14.** Alignments of *rps3* in members of Thismiaceae that retain the gene. The nucleotide and amino acid regions labelled with the magenta tags contain the single nucleotide small single-copy region (SSC) for *Thismia rodwayi*. **A**. Codon alignment of *rps3*. **B**. Amino acid alignment of *rps3*. The periods ('.') in grey indicate an unknown amino acid and the asterisks ('*') in dark grey indicate a stop codon.

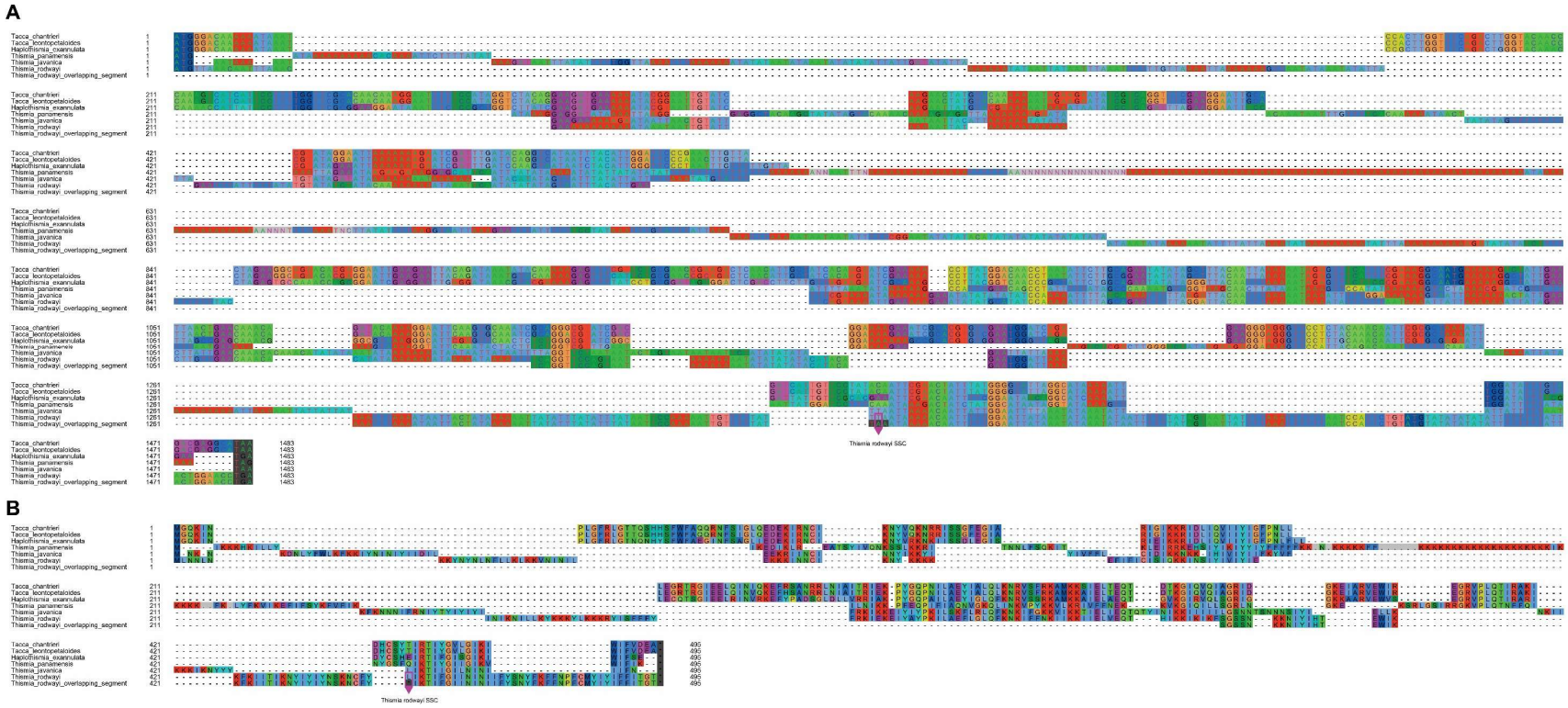


**Figure A15.** Nucleotide alignment of *rpl2* introns showing the location of *trnfM-CAU* within the intron. Conserved splicing sites for group IIA introns are included above the top nucleotide string (Y indicates the nucleotide may be either cytosine or thymine). Magenta highlighted boxes indicate the location of the most likely group IIA splice site for each sequence. The grey bar underneath indicates the location of *trnfM-CAU* within the intron. In *Thismia javanica* and *T. tentaculata*, the *rpl2* gene is complete in one inverted repeat (IR) copy and incomplete in the other (see Figs. A6 and A10 respectively; complete copies shown here).

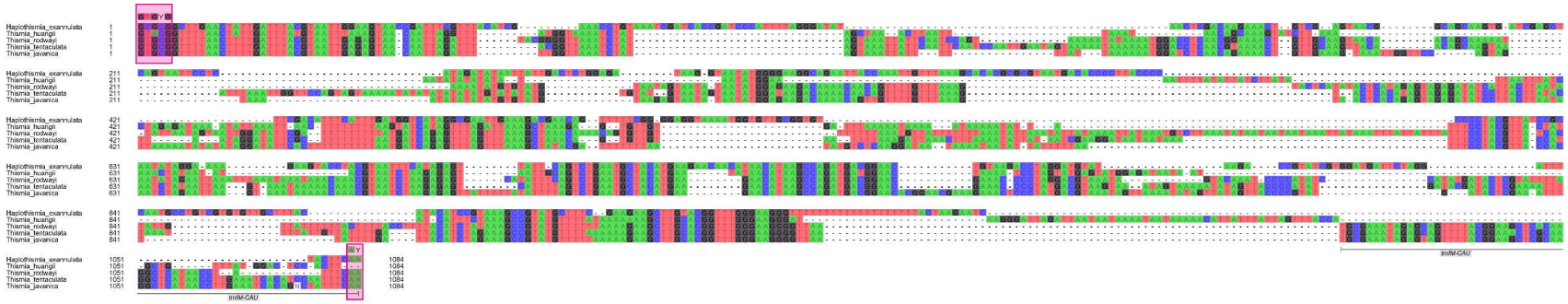


**Figure A16.** Alignments of *rps12*. Grey bars along the top string of nucleotides mark the location of exon and intron boundaries. As the red arrow indicates, there is no intron between exon 1 and exon 2 because *rps12* is a trans-spliced gene. Exons may not be spliced in this manner in vivo*,* splicing boundaries presented here are based on comparison to green Thismiaceae taxa only. **A**. Nucleotide alignment of *rps12*. Putative introns for *Thismia huangii, T. javanica*, *T. rodwayi*, *T. tentaculata,* and are not included as they are much longer due to exon rearrangement and in the cases of *T. huangii*, *T. rodwayi*, and *T. tentaculata* do not align well. See Figs. A9, A6, A8, and A10 for *Thismia huangii, T. javanica*, *T. rodwayi*, and *T. tentaculata rps12* exon rearrangements, respectively. See Fig. A17 for alignment of *T. javanica*'s putative intron. **B**. Amino acid alignment of *rps12*.

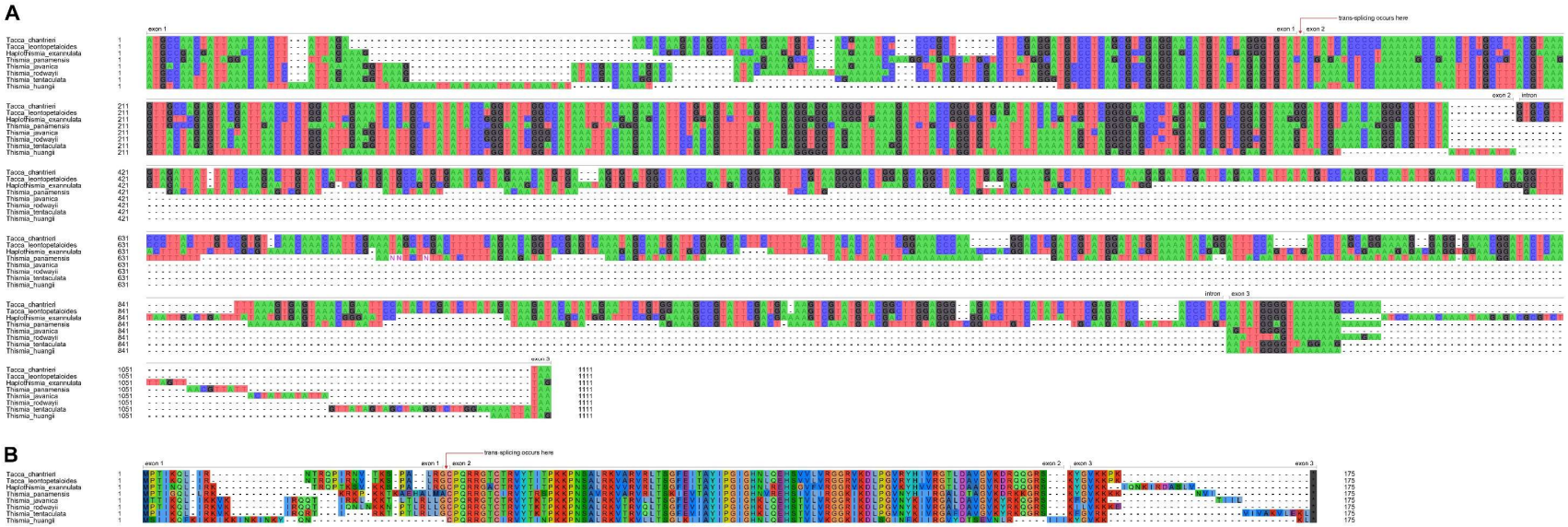


**Figure A17.** Nucleotide alignment of *rps12* introns from members of Thismiaceae that retain the intron with the putative *Thismia javanica* intron. The grey bars underneath the *T. javanica* string indicate locations of other *T. javanica* genes within the putative intron. Conserved splicing sites for group IIA introns are included above the top nucleotide string (Y indicates the nucleotide may be either cytosine or thymine). Magenta highlighted boxes indicate the location of the most likely group IIA splice site for each sequence. See Fig. A6 for the exon rearrangement that caused the long putative intron in *T. javanica*.