UNDERSTANDING ANGIOSPERM GENOME INTERACTIONS AND EVOLUTION: INSIGHTS FROM SACRED LOTUS (*NELUMBO NUCIFERA*) AND THE CARROT FAMILY (APIACEAE)

BY

RHIANNON PEERY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Plant Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Stephen R. Downie, Chair, Director of Research
Professor Ken N. Paige
Professor Sydney A. Cameron
Assistant Professor Katy Heath

ABSTRACT

Horizontal and intracellular gene transfers are driving forces in plant evolution. The transfer of DNA into a genome adds genetic diversity and successfully incorporated genes can retain their original function or develop new functions through mutation. While there are trends and hypotheses for the frequency of transfers, age of transfers, and potential mechanisms of transfer each system has its own evolutionary history. The major goal of this study was to investigate gene transfer events and organelle rare genomic changes in two plant systems – *Nelumbo* (Nelumbonaceae) and the apioid superclade of Apiaceae subfamily Apioideae.

Genome sequences from the early diverging angiosperm *Nelumbo nucifera* 'China Antique' were used to describe both intra- and interspecific patterns of variation and investigate intracellular gene transfers (IGT). A percent similarity approach was used to compare DNA from each genome and determine a possible mechanism of DNA transfer, if it occurred. The mechanisms investigated included recombination and double-strand break repair, as evidenced by repeat DNA and the presence of transposable elements. The 'China Antique' plastome retains the ancestral gene synteny of *Amborella* and has no evidence of IGT. 'China Antique' has more smaller repeats in its mitochondrial genomes than reported for other angiosperms, but does not contain any large repeats, and its nuclear genome does not have as much organelle DNA as the other angiosperms investigated, including *Arabidopsis*. The lack of large repeats within the *Nelumbo* mitochondrial genome may explain the few instances of IGT detected. The few instances of organelle IGTs into its nucleus may be the result of its history of vegetative propagation, low nucleotide substitution rate, and lack of several paleo-duplications.

Unlike *N. nucifera*, and the majority of other angiosperms, the plastomes of several members of the apioid superclade within the carrot family (Apiaceae or Umbelliferae) have instances of IGT into the plastome, in addition to other rare genomic changes (RGCs). To investigate the distribution and mechanism of IGT in species of the apioid superclade and the

variable boundary between the two single copy regions and the IR, the complete plastomes of *Anethum graveolens*, *Foeniculum vulgare*, *Carum carvi*, and *Coriandrum sativum* were sequenced. To determine the distribution of and mechanisms causing these RGCs, the extent of IGT, and changes in gene synteny, the large single copy (LSC)–inverted repeat (IR) boundary in 34 additional species was also sequenced. Analyses of these sequence data suggest that there are several mechanisms at work creating these dynamic IR changes. There is evidence of double-strand break repair in *Coriandrum*, as well as repeat mediated changes near its IR boundaries. Short dispersed repeats are also implicated as a mechanism of IR change in the 34 additional species investigated. In *Carum* (tribe Careae) there is an IR boundary expansion, in addition to two small inversions. One of these inversions is near J_{LA} and the other is between *psbM* and *trnT*. *Anethum* and *Foeniculum* plastomes contain double-strand break repair causing IGT of mtDNA into these plastomes. For the 34 additional species investigated, data support double-strand break repair as a mechanism of plastid evolution and is the likely cause of novel DNA insertions at LSC–IR boundaries. However, without a resolved phylogeny there is no context for how many gene transfer events there were or a timeline for when these events occurred.

Molecular phylogenetic studies to date have been unable to produce a well-resolved apioid superclade phylogeny. To resolve relationships among the tribes and other higher-level clades within the group, determine the phylogenetic utility of RGCs, and determine the extent and timing of plastome RGCs in the group, the plastid regions *psbM–psbD and psbA–trnH* and the nuclear gene *PHYA* were sequenced. To these sequence data four RGCs were added, as were previously available data from the nrDNA internal transcribed spacer (ITS) region. These molecular data were analyzed separately and in various combinations using maximum likelihood and Bayesian inference methods. While these data were unable to fully resolve higher-level relationships in the apioid superclade, conclusions can be made regarding the distribution and

number of RGC events that have occurred in the group. The IR boundary expansion into *rps3* occurred only once in the lineage leading to tribes Careae and Pyramidoptereae. In addition, Careae is supported as monophyletic by the presence of the inversion of *psbA* and *trnH*. The contraction of the IR to *rpl2* and the presence of putative mtDNA adjacent to J_{LA} also likely occurred only once. Alternatively, while not as parsimonious, a maximum of six events is possible if each lineage gained these RGCs independently. Other major lineages within the group are not as strongly delimited and, for these clades RGCs cannot unambiguously support monophyly. Further study of the apioid superclade is necessary to resolve relationships and make further inferences into the evolution of plastomes within the clade.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Stephen R. Downie, for welcoming me in to his lab and helping me become a better scientist and writer through his advise and editing expertise. I would like to thank my committee members Dr. Sydney A. Cameron for advice on phylogenetic analyses, Dr. Ken Paige for his positive encouragement and advice on paper composition and organization, and Dr. Katy D. Heath for always having a ready ear and great advice about everything from family to being an early-career scientist.

I would like to give a big thank you to my collaborators at other institutions. Dr. Sarah Mathews welcomed me into her lab to learn techniques associated with the amplification of *PHYA*. Dr. Linda Raubeson, my master's advisor, without whom I would never have known DNA was so interesting. Dr. Robert K. Jansen for advise, being a conference friend, and being the lead P.I. on the project that got me interested in plastomes. Lastly, I would like to thank Dr. Eric B. Knox for protocols and advise on mitochondrial DNA isolation.

A dissertation takes a village and I have had wonderful support from my community. First I would like to thank the many people in the School of Integrative Biology that helped me along the way. Foremost among these is Dr. Ray Ming for his time on my committee and help with spearheading the lotus genome project and funding other areas of my dissertation. I owe many thanks to Ray and his lab including Jennifer Han, Bob VanBuren, and Jennifer Wai for help with lab work and suggestions on how best to present genomic data. In addition, I would like to thank Dr. Ray Zielinski for his help with organelle genome isolation procedures. I need to thank Devin Quarles for the sage advice "design different primers" regarding mitochondrial DNA amplification. For their help with analyses and being a great sounding board I would like to thank the Systematics Discussion Group and HBV Consortium, especially Dr. Julie Allen, Dr. Jim Whitfield, Dr. Sydney Cameron, Dr. Tandy Warnow, Dr. Chris Dietrich, Drew Sweet, Andrew

Debevec, Dr. Bret Boyd, Patrick Gero, and Dr. Therese Catanach. Additionally I would like to thank Laura Stein, Danielle Ruffatto, Stephanie Turza, and Vince Hustad for reading early drafts and copious grant and post doctorial applications. Several undergraduate students and two visiting scholars helped with lab work related to this project: Victoria Rodriguez, Colleen Sullivan, Dr. Yue Xu, Dr. Hong-Li Chang, and Anne Curé. Finally I would like to thank the Graduate students in Ecology and Evolutionary Biology for listening to several presentations and moral support especially the ladies of wine night and the Softballogists. Last but certainly not least I thank my husband Patrick Wilson and my parents Rick and Tammy Peery. Without all their support and love I would never have aspired to achieve my education goals.

Funding for this work was provided by the Department of Plant Biology, the School of Integrative Biology, and the Botanical Society of America to Rhiannon Peery; the Downie—Katz-Downie Research Fund; the National Science Foundation Research Program Award DBI-0922545 and startup funds from Fujian Agriculture and Forestry University to Ray Ming; and National Science Foundation Program Award DEB-0120709 to Linda Raubeson.

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION 1

CHAPTER 2: REDUCED INTRACELLULAR GENE TRANSFER IN THE GENOMES OF SACRED LOTUS (*NELUMBO NUCIFERA*) 10

CHAPTER 3: THE PLASTOMES OF *ANETHUM GRAVEOLENS*, *FOENICULUM VULGARE*, *CARUM CARVI*, AND *CORIANDRUM SATIVUM* (APIACEAE): CHARACTERIZATION OF INVERTED REPEAT CHANGES 55

CHAPTER 4: THE PHYLOGENETIC UTILITY OF PLASTOME RARE GENOMIC CHANGES, PLASTID GENE REGIONS *PSBM–PSBD* AND *PSBA–TRNH*, AND NUCLEAR GENE *PHYA* IN RESOLVING RELATIONSHIPS WITHIN THE APIOID SUPERCLADE OF APIACEAE SUBFAMILY APIOIDEAE 106

CHAPTER 1: GENERAL INTRODUCTION

One of the most important processes in plant evolution is DNA transfer. Horizontal gene transfer (HGT), the transfer of DNA from one individual to another, was once thought to be rare but is now supported as a driving force in the evolution of plants (Yue et al. 2012). HGT is significant because genes acquired through transfer add to genetic diversity and can be co-opted for their original purpose or modified for new functions (Barkman et al. 2007; Noutsos et al. 2007; Kleine et al. 2009; Lloyd and Timmis 2011; Rousseau-Gueutin et al. 2011; Wang et al. 2012; Zhang et al. 2013).

A special case of HGT is intracellular gene transfer (IGT), the sharing of DNA among genomes within an individual. Immediately after endosymbiosis evolved, IGT among plant genomes began (Martin and Herrmann 1998). Initially, there was a unidirectional outflow of genes from the organelle genomes into the nuclear genome (Martin and Herrmann 1998; Martin 2003; Timmis et al. 2004; Kleine et al. 2009). After this initial purge the organelle genomes themselves followed quite different evolutionary paths regarding DNA transfer (Richardson and Palmer 2007; Smith 2011; Sanchez-Puerta 2014). The mitochondrial genome has been coined “promiscuous” (Stern and Lonsdale 1982), readily accepting DNA through IGT from both plastome and nuclear genomes and through HGT from foreign genomes (Richardson and Palmer 2007; Hao et al. 2010; Mower et al. 2010; Rice et al. 2013; Xi et al. 2013; Wang et al. 2015). Conversely, the plastome can be considered “chaste,” as IGT to the plastome is extremely rare (Rice and Palmer 2006) and HGT has never been reported. The rare cases of IGT into the plastome have so far only been reported for algal plastids (Sheveleva and Hallick 2004), one subfamily of Apocynaceae (Straub et al. 2013), and some members of the plant family Apiaceae (Goremykin et al. 2009; Iorizzo et al. 2012; Downie and Jansen 2015).

Reduced Intracellular Gene Transfer in the Genomes of Sacred Lotus (Nelumbo nucifera)

The influx of genomic DNA from next generation sequencing methods provides an excellent opportunity to study IGT across angiosperms. Currently, there are several species that have both of their organelle genomes published and, in many cases, these are from the same individuals. This permits a comparative analysis of the extent of IGT in angiosperms. The recent publication of the *Nelumbo nucifera* nuclear genome (Ming et al. 2013) provided an opportunity to investigate the frequency and type of intracellular gene transfer among all three genomes in a basal eudicot. Understanding how often and what is transferred through IGT can help understand the processes of evolution acting on plant genomes.

The Plastomes of Anethum graveolens*,* Foeniculum vulgare*,* Carum carvi*, and* Coriandrum sativum *(Apiaceae): Characterization of Inverted Repeat Changes*

The difference in gene transfer between the organelle genomes is likely related to the morphology and sequence evolution of the genomes themselves (Smith 2011). Evolution within the plastome occurs mostly through point mutations with few gene order changes, whereas the mitochondrial genome frequently undergoes changes in gene order. The DNA mutation rate among genomes also varies. The mitochondrial genome has the lowest rate, followed by the plastome, then the nuclear genome (Wolfe et al. 1987). Both chloroplast and mitochondrial genomes can each usually be assembled as a “master” circle (Fig. 1.1). However, mitochondrial genomes are far more complex (Table 1.1), with the majority investigated thus far having a multipartite organization of interconverting small and large circular genomes due to large duplications (Lonsdale 1984; Palmer and Shields 1984; Sugiyama et al. 2005). These duplications also cause the size of mitochondrial genomes to vary by hundreds of thousands of base pairs (Palmer 1990; Alverson et al. 2010). This complexity has led to a lag in mitochondrial genome publication.

Differences in plastome sizes are due primarily to small fluctuations in the amount of DNA contained within its large inverted repeat (IR). Most angiosperm plastomes have two single copy regions—a large single copy (LSC) region and a small single copy (SSC) region—that are flanked by IRs (Fig. 1.1). The boundaries where these single copy regions meet the IR can be variable, even exhibiting dramatic shifts in position. Boundaries are defined by where DNA duplication ends and single copy DNA begins. “Shifts” in the boundary imply a change from the angiosperm ancestral state resulting in more or less DNA being duplicated. Small changes in IR size (< 100 bp) are common (Goulding et al. 1996), while large expansions and contractions (> 1 kb) without IR loss are rare (Palmer et al. 1987; Raubeson and Jansen 2005; Hansen et al. 2007; Guisinger et al. 2011). Despite the general rule that the chloroplast has very stable gene adjacencies (Palmer 1985, Palmer 1991, Raubeson and Jansen 2005), these four junctions, where the IR meets the single copy regions, can be dynamic in some taxa (Palmer 1985; Palmer et al. 1987; Goulding et al. 1996; Cosner et al. 1997; Plunkett and Downie 2000; Hansen et al. 2007; Lee et al. 2007).

Apiaceae are one of the few angiosperm families to have a dynamic IR (Plunkett and Downie 1999, 2000; Downie and Jansen 2015). Within Apioideae, the largest subfamily of Apiaceae, the plastid genome has changed dramatically over time. Mapping studies of the chloroplast genome have shown that members of the apioid superclade of subfamily Apioideae have diverse IR boundaries (Plunkett and Downie 1999, 2000). These boundary differences affect the length of the IR and gene adjacencies on the J_{LA} side of the genome. Thus far no research has been done at the sequence level to determine the mechanisms of IR change in this group. In some species there is an insertion of novel DNA that has high sequence similarity to mitochondrial DNA. As such, Apiaceae provide an ideal system in which to study chloroplast genome promiscuity. My research uses Apiaceae as a model system to determine mechanisms of IR change and investigate plastome IGT within the family.

The Phylogenetic Utility of Plastome Rare Genomic Changes, Plastid Gene Regions psbM–psbD *and* psbA–trnJ*, and Nuclear Gene* PHYA *in Resolving Relationships Within the Apioid Superclade of Apiaceae Subfamily Apioideae*

The apioid superclade comprises 14 tribes and other major clades (Downie et al. 2010). Several plastid genes and non-coding DNA regions, as well as the nuclear ribosomal DNA internal transcribed spacer (ITS) region, have all been used as markers to study Apioideae phylogenetic relationships. While these studies have contributed greatly to a broad understanding of its evolutionary history, uncertainties remain with regard to the backbone relationships of the apioid superclade and other deep-level relationships within the group primarily because of a paucity of phylogenetically informative characters (reviewed in Downie et al. 2001, 2010). Currently, in the absence of a well-resolved phylogeny, it is unclear when plastome changes first occurred and what clades they support. Well-resolved phylogenies are critical when addressing hypotheses of character evolution. Therefore, a goal of this research was to place the rare genomic changes described in Chapter 3 into an evolutionary context by generating a new and robust phylogeny for the apioid superclade using two new plastid markers (*psbM–psbD* and *psbA–trnH*), rare genomic changes in the plastome (including changes in gene synteny at J_{LA}, inversions, and IGT events), and the nuclear gene phytochrome A.

References

Alverson AJ, Wei X-X, Rice DW, Stern BD, Barry K, Palmer JD (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27:1436-1448.

Barkman TJ, McNeal JR, Lim S-H, Coat G, Croom HB, Young ND, dePamphilis CW (2007) Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol Biol 7:248.

Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr Genet 31:419-429.

Downie SR, Jansen RK (2015) A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot 40:336-351.
Downie SR, Plunkett GM, Watson MG, Spalik K, Katz-Downie DS, Valiejo-Roman CM, Terentieva EI, Troitsky AV, Lee B-Y, Lahham J, El-Oqlah A (2001) Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. Edinb J Bot 58:301-330.
Downie SR, Spalik K, Katz-Downie DS, Reduron J-P (2010) Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. Plant Div Evol 128:111-136.
Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol 26: 99-110.
Goulding SE, Wolfe KH, Olmstead RG, and Morden CW (1996) Ebb and flow of the chloroplast inverted repeat. Molec General Genet 252:195-206.
Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28:583-600.
Hansen DR, Dastidara SG, Caia Z, Penaflorb C, Kuehl JV, Boore JL, Jansen RK (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). Mol Phylogenet Evol 45:547-563.
Hao W, Richardson AO, Zheng Y, Palmer JD (2010) Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proc. Natl. Acad. Sci. USA 107:21576-21581.
Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P (2012) *De novo* assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol 12:61.
Kleine Tm Maier UG, Leister D (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol 60:115-138.
Lee HL, Jansen RK, Chumley TW, Kim KJ (2007) Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple overlapping inversions. Mol Biol Evol 24:1161-1180.
Lloyd AH, Timmis JN (2011) The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. Mol Biol Evol 28:2019-2028.
Lonsdale DM (1984) A review of the structure and organization of the mitochondrial genome of higher plants. Plant Mol Biol 3:201-206.
Martin W (2003) Gene transfer from organelles to the nucleus: frequent and in big chunks. PNAS 100:8612-8614.
Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol 118:9-17.
Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li L-T, Zhang Q, Kim M-J, Schatz MC, Campbell M, Li J, Bowers JE, Tang H, Lyons E, Ferguson AA, Narzisi G, Nelson DR, Blaby-Haas CE, Gschwend AR, Jiao Y, Der JP, Zeng F, Han J, Min XJ, Hudson KA, Singh R, Grennan AK, Karpowicz SJ, Watling JR, Ito K, Robinson SA, Hudson ME, Yu Q, Mockler TC, Carroll A, Zheng Y, Sunkar R, Jia R, Chen N, Arro J, Wai CM, Wafula E, Spence A, Han Y, Xu L, Zhang J, Peery R, Haus MJ, Xiong W, Walsh JA, Wu J, Wang M-L, ZhuYJ, Paull RE, Britt AB, Du C, Downie SR, Schuler MA, Michael TP, Long SP, Ort DR, Schopf JW, Gang DR, Jiang N, Yandell M, dePamphilis CW, Merchant SS, Paterson AH, Buchanan BB, Li S,

Shen-Miller J (2013) Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). Genome Biol 14:R41.
Mower JP, Stefanović S, Hao W, Gummow JS, Jain K, Ahmed D, Palmer JD (2010) Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biol 8:150.
Noutsos C, Kleine T, Armbruster U, DalCorso G, Leister D (2007) Nuclear insertions of organelle DNA can create novel patches of functional exon sequences. Trends Genet 23:597-601.
Palmer JD (1985) Comparative organization of chloroplast genomes. Annual Review Genet 19:325-354.
Palmer JD (1990) Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet 6:115-120.
Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK, editors. The molecular biology of plastids. New York: Academic Press; 1991. p. 5-53.
Palmer JD, Nugent JM, Herbon LA (1987) Unusual structure of geranium chloroplast DNA: a triple-sized repeat, extensive gene duplications, multiple inversions, and new repeat families. PNAS USA 84:769-773.
Palmer JD, Shields CR (1984) Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature 307:437-440
Plunkett GM, Downie SR (1999) Major lineages within Apiaceae subfamily Apioideae: A comparison of chloroplast restriction site and DNA sequence data. Am J Bot 86:1014-1026.
Plunkett GM, Downie SR (2000) Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Syst Bot 25:648-667.
Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ, editor. Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants. London: CAB International.
Rousseau-Gueutin M, Ayliffe MA, Timmis JN (2011) Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. Plant Physiol 157:2181-2193.
Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Berrie K, Boore JL, Zhang Y, dePamphilis CW, Knox EB, Palmer JD (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science 342:1468-1473.
Richardson AO, Palmer JD (2007) Horizontal gene transfer in plants. J Exp Bot 58:1-9.
Sanchez-Puerta MV (2014) Involvement of plastid, mitochondrial and nuclear genomes in plant-to-plant horizontal gene transfer. Acta Soc Bot Pol 83:317-323.
Sheveleva EV, Hallick RB (2004) Recent horizontal transfer to a chloroplast genome. Nucleic Acids Res 32:803-810.
Smith DR (2011) Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol Evol 3:743-748.
Stern DB, Lonsdale DM (1982) Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. Nature 299:698-702.
Nature 299: 698–702Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A (2013) Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). Genome Biol Evol 5:1872-1885.
Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics 272:603-615.

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Reviews Genet 5:123-135.
Wang B, Climent J, Wang X-R (2015) Horizontal gene transfer from a flowering plant to the insular pine *Pinus canariensis* (Chr. Sm. Ex DC in Buch). Heredity 114:413-418.
Wang D, Rousseau-Gueutin M, Timmin JN (2012) Plastid sequences contribute to some plant mitochondrial genes. Mol Biol Evol 29:1707-1711.
Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. PNAS 84:9054-9058.
Xi, Z, Wang Y, Bradley RK, Sugumaran M, Marx CH, Rest JS, Davis CC (2013) Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genet 9:e1003265.
Yue J, Hu X, Sun H, Yang Y, Huang J (2012) Widespread impact of horizontal gene transfer on plant colonization of land. Nature Communications 3:1152.
Zhang Y, Fernandez-Aparicio M, Wafula EK, Das M, Jiao Y, Wickett NJ, Honaas LA, Ralph PE, Wojciechowski MF, Timko MP, Yoder JI, Westwood JH, dePamphilis CW (2013) Evolution of a horizontally acquired legume gene, albumin 1, in the parasitic plant *Phelipanche aegyptiaca* and related species. BMC Evol Biol 13:48.

Tables and Figures

Table 1.1 Organelle genome sizes for *Arabidopsis*, *Citrullus* (cucumber), *Daucus* (carrot), and *Zea mays* (maize). Two subspecies of maize are included to demonstrate that large differences in size can be found in mitochondrial genomes within the same species.

Taxon	Mitochondrial genome size (kb) and GenBank accession number	Chloroplast genome size (kb) and GenBank accession number
Arabidopsis thaliana	36.7 (NC_001284)	15.4 (NC_000932)
Citrullus lanatus	168.5 (NC_014043)	15.5 (NC_007144)
Daucus carota subsp. *sativus*	28.1 (NC_017855)	15.6 (NC_008325)
Zea mays subsp. *mays*	57.0 (NC_007982)	14.0 (NC_001666)
Zea mays subsp. *parviglumis*	68.0 (NC_008332)	No data

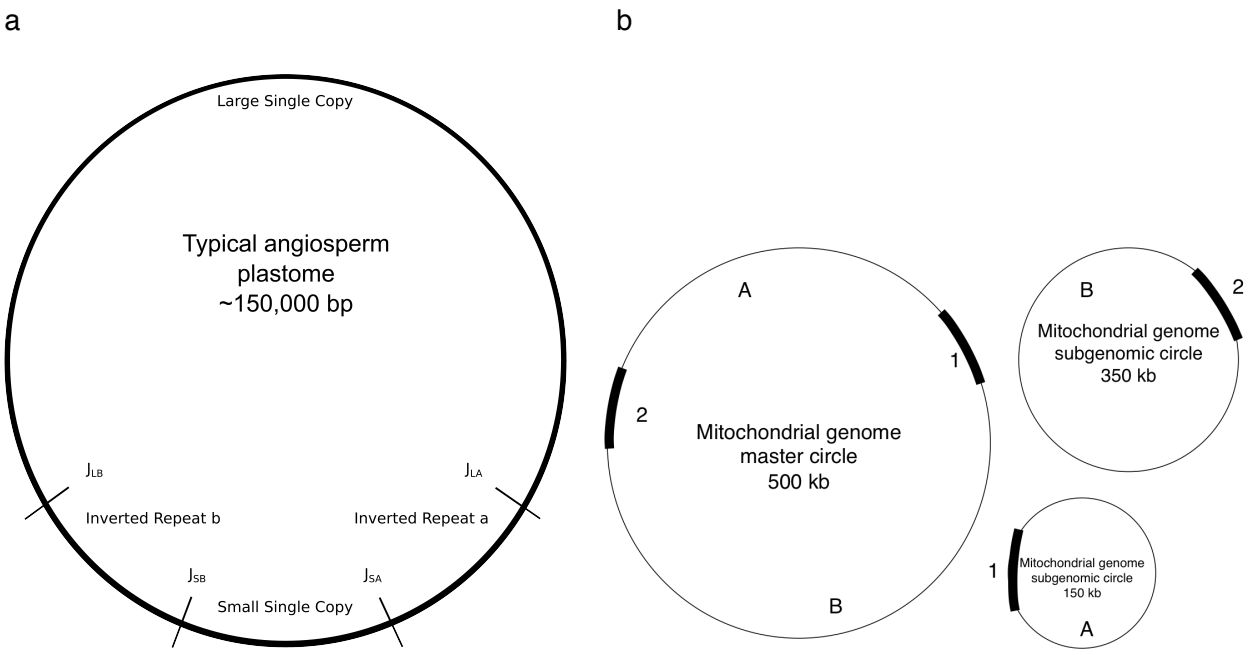


Fig. 1.1 Comparison of angiosperm plastid and mitochondrial genomes. (a) Typical genome configuration of an angiosperm plastome. The boundaries where single copy regions meet the IR can be variable (J_{LA} = junction at the large single copy and inverted repeat a; J_{LB} = junction at the large single copy and inverted repeat b; J_{SA} = junction at the small single copy and inverted repeat a; J_{SB} = junction at the small single copy and inverted repeat b. (b) Basic structural organization of an angiosperm mitochondrial genome. Subgenomic circles are possible through recombination at repeats 1 and 2 (shown as bold regions), breaking apart gene regions A and B into separate molecules.

CHAPTER 2: REDUCED INTRACELLULAR GENE TRANSFER IN THE GENOMES OF SACRED LOTUS (*NELUMBO NUCIFERA*)

Abstract

Intracellular gene transfer from the organelles into the nuclear genome and from the plastome and nuclear genome into the mitochondrial genome is an ongoing and dynamic process; however, the amount, location, and timing of these transfers are different in all species examined thus far. The basal eudicot *Nelumbo nucifera* 'China Antique' genome was sequenced and its organelle genomes were captured bioinformatically and assembled and annotated. Herein, I describe these organelle genomes, compare both intra- and interspecific patterns of variation with other taxa, and investigate intracellular gene transfers. The 'China Antique' plastome does not vary from the ancestral angiosperm plastome in its structural organization and gene arrangement, the draft mitochondrial genome has more smaller repeats than reported for other angiosperm mitochondrial genomes, but does not contain any large repeats, and the nuclear genome is depauperate in organelle DNA. The lack of large repeats within its mitochondrial genome may explain the few instances of plastid DNA introgression. The even distribution of nuclear genes may also be preventing successful integration and retention of organelle DNA. The nuclear genome of 'China Antique' has undergone only one paleo-duplication and shows a reduction in its overall mutation rate. These factors along with seed longevity and vegetative propagation could be the cause of reduced levels of intracellular gene transfers in *Nelumbo*.

Introduction

Nelumbo nucifera Gaertn., sacred lotus, is one of two species of aquatic plants in the family Nelumbonaceae. It is an economically and culturally important species native to Asia and Australia and is classified in the eudicot order Proteales, sister group to the core eudicots (APG III 2009). *Nelumbo lutea* (Willd.) Pers. is the other member of the family and is native to North America and the Caribbean.

Nelumbo nucifera 'China Antique' is the most basal angiosperm eudicot to have its entire genome sequenced (Ming et al. 2013). Its nuclear genome has a slow rate of evolution and lacks the paleo-triplication found in core eudicots (Jiao et al. 2012; Ming et al. 2013). The species is commonly cultivated, with several hundred cultivars described (Xue et al. 2012). It has exceptionally long-lived seeds, with seedlings that are initially very fragile but quickly becoming hardy (Shen-Miller 2002a; Shen-Miller et al. 2002b). The plants are mostly vegetatively propagated (Guo 2009). In addition to extraordinary seed longevity (Shen-Miller et al. 2002b), *N. nucifera* is known for having extremely hydrophobic leaves (Ensikat et al. 2011).

Next generation sequencing (NGS) technologies have allowed for faster acquisition and processing of sequence data than ever before, and algorithms have advanced to handle repeats and assembly without detailed mapping from BAC libraries. These technological advances have led to numerous plant genomes being sequenced; however, less than 10 of these have complete sequence for plastid and mitochondrial organelle genomes as well. Having sequence data from three genomes within and individual organism permits their comparative analysis, including studies of intracellular gene transfer (IGT). In angiosperms, intracellular transfer of DNA is frequent, but the directionality of the transfer is biased (Leister 2005). Nuclear and mitochondrial genomes often accept foreign DNA, whereas plastomes generally do not (Rice and Palmer 2006). The transfer of plastid and mitochondrial DNA into the nuclear genome began immediately after the origins of symbiosis and is an ongoing process (Martin and

Hermann 1998; Gould et al. 2009), such that all plant nuclear genomes have varying levels and ages of organelle DNA content (Blanchard and Schmidt 1995; Martin 2003; Timmis et al. 2004). This DNA is termed NORG, nuclear organellar DNA, and can be categorized according to which organelle genome donated the DNA, either nuclear mitochondrial DNA (NUMT) or nuclear plastid DNA (NUPT). The largest intact NUMT is a 620 kb fragment in *Arabidopsis thaliana* (Stupar et al. 2001) and the largest NUPTs are 33 kb and 131 kb fragments in rice (Guo et al. 2008). Such insertions are not distributed evenly across nuclear chromosomes. DNA integration more often occurs in large chunks, each several thousand nucleotides in size or in concatenated smaller fragments, rather than small transcripts being integrated individually (Yuan et al. 2002; Hazkani-Covo et al. 2010). The site of integration is often near centromeres (Matsuo et al. 2005), or on single chromosomes such as chromosome 2 in *Arabidopsis* (Meinke et al. 1998; Lin et al. 1999; A.G.I. 2000) and chromosome 3 in *Sorghum* (Paterson et al. 2009). The overall amount of introgression reported is strongly correlated with nuclear genome size (Hazkani-Covo et al. 2010).

Herein, I report on the *N. nucifera* plastid and mitochondrial genomes, as part of the *N. nucifera* 'China Antique' genome sequencing project (Ming et al. 2013). Specifically, I characterize the amount of NORGs present, investigate the amount of plastid DNA within the mitochondrial genome (MTPT), and compare its mitochondrial genome to those of other eudicots. I also compare its plastome to previously published plastomes of *N. nucifera* and *N. lutea* to investigate rates and types of mutations occurring among them. To investigate other instances of IGT in angiosperms, I examine seven published genomes (GenBank database accessed July 20, 2014) for which annotated mitochondrial and plastid genomes are also available. These angiosperms include three grasses (*Oryza sativa*, *Sorghum bicolor*, and *Zea mays*) and four rosids (*Arabidopsis thaliana*, *Carica papaya, Glycine max*, and *Vitis vinifera*).

Materials and Methods

DNA Isolation and Sequencing

Etiolated leaf tissues were used for nuclei preparation as per Ming et al. (2013). Whole-genome shotgun sequencing was done at the University of Illinois Roy J. Carver Biotechnology Center (www.biotech.uiuc.edu/htdna). As described in Ming et al. (2013), several rounds of Illumina Solexa sequencing generated the majority of the raw data. Sequencing followed standard protocols used with the Illumina HiSeq 2000 sequencing system. Four paired-end Illumina libraries were created with inserts of 180 bp, 500 bp, 3.8 kb, and 8 kb. A paired-end 20 kb insert library was also generated and used for nuclear scaffolding with the Roche/454 circularization protocol. 454 sequencing was done using the 454 FLX+ system. Organelle specific reads were separated from nuclear reads bioinformatically.

Plastome

All available *Nelumbo* genomic data were included in the assembly of the plastome using reference guided assembly in the CLC GENOMICS WORKBENCH 4 (http://www.clcbio.com/). The unpublished plastome from *Nelumbo nucifera* (GenBank accession NC_015610) was used as the reference. The sequence depth of the aligned reads averaged >78,000 along the entire genome. Inverted repeat boundaries were confirmed by PCR amplification across boundaries followed by sequencing. The two LSC/IR boundary amplicons were aligned in CLUSTAL OMEGA (http://www.ebi.ac.uk/Tools/msa/clustalo/) and the point of mismatch was deemed the IR boundary (Raubeson et al. 2007). The same process of identification was used for determining the SSC/IR boundary. No additional PCR was necessary to improve the quality of the DNA base calls or to join contigs. Annotation was done using DOGMA (Wyman et al. 2004). The circular gene map was produced using CIRCOS V. 0.56 (Krzywinski et al. 2009). Gene synteny was

determined using *Nicotiana tabacum* as the reference (Shinozaki et al. 1986). Alignments of the newly generated 'China Antique' plastome and the four available *Nelumbo* plastomes (Xue et al. 2012) were done using MESQUITE v. 2.75 (build 566) and the plug-in OPAL (Wheeler and Kececioglu 2007). Detailed differences among plastomes were identified in SEQUENCHER v. 5.0 (http://genecodes.com/). Identification of repeat DNA was done using SSR-Extractor (Dolan unpublished) for identification of simple sequence repeats (SSRs), with a minimum size of 10 for mono- and dinucleotide repeats and 15 for trinucleotide repeats. Short dispersed repeats (SDRs) were identified in VMATCH v. 2.2.2 (http://www.vmatch.de/) using a Hamming distance of three and a minimum repeat size ≥ 30 bp.

Mitochondrial Genome

Putative mitochondrial contigs were created using the GS DE NOVO ASSEMBLER v. 2.6 (Roche, USA). Any contig with a 20-fold higher than average coverage was investigated and verified using BLAST v. +2.2.28 alignment to conserved mitochondrial genic sequences. Contigs that had high sequence similarities to mitochondrial DNA were then used as a backbone for assembling Illumina paired-end reads. Illumina reads were assembled to 454 contigs using the CLC GENOMICS WORKBENCH. PCR primers were designed for the ends of these contigs and long-range PCR amplifications, followed by sequencing when appropriate, were done to try and complete the *Nelumbo* mitochondrial genome. Long range PCR was performed using Biolines RANGER DNA polymerase following the manufacturer's protocol (http://www.bioline.com/). Annotation of the draft mitochondrial genome was done with the assistance of MITOFY (Alverson et al. 2010), which automates the search for known mitochondrial proteins and tRNAs using BLAST and TRNASCAN-SE. DOGMA was used to identify plastid genes or pseudogenes within the mitochondrial genome. When verifying exon boundaries using SEQUIN v. 12.3, the TAIR database (http://www.arabidopsis.org/) and the annotated *Carica papaya* (EU431224) and

Nicotiana tabacum (Sugiyama et al. 2005) genomes were used. The gene map was produced using CIRCOS (Krzywinski et al. 2009). Identification of repeat DNA was as described for the plastome.

The 'China Antique' mitochondrial genome was compared to mitochondrial genomes of *Beta* (Kubo et al. 2000), *Arabidopsis* (Unseld et al. 1997), *Carica* (EU431224), *Glycine* (Chang et al. 2013), *Oryza* (Notsu et al. 2002), *Sorghum* (Saski et al. 2007; Paterson et al. 2009), *Vitis* (Jansen et al. 2006, Jaillon et al. 2007, Goremykin et al. 2009), and *Zea* (Maier et al. 1995) for gene content, presence of shared DNA and plastid DNA, and proportion of repeat DNA. To compare the amount of DNA shared among genomes, BLASTN searches were performed with a length cutoff of 60 and a percent identity of 70. This was done to reduce the amount of repeat DNA matching by chance but still capture tRNA-length genes. BLASTN was also used to identify MTPT DNA with cutoffs for length and percent identity of 60 bp and 70%, respectively.

DNA Introgression

Although care was taken with the assembly of the *Nelumbo* nuclear genome (Ming et al. 2013), there were many contigs in that final assembly and thus I am conservative in determination of introgression because of the possibility of organellar DNA contamination. Nuclear contigs were screened for NORGs using both high (penalty -3, reward 1, gapopen 5, gapextend 2) and low (penalty -4, reward 5, gapopen 8, gapextend 6) stringency searches using BLAST, with a word size of 11 and a percent identity cutoff of 90 (Rice et al. unpublished). These searches capture only recent incorporation of organellar DNA, but the certainty of correctly identifying an actual introgression is higher. Only matches larger than 100 bp were considered, helping to eliminate false instances of introgression. If a nuclear region matched several organelle regions, only the best match (based on e-value and length) was reported. The same searches were performed using the mitochondrial genome as the subject of the plastome

query. The overall percentage of intracellular gene transfer was calculated for each type of transfer, enabling a direct comparison of how much of the genome is composed of foreign DNA.

Comparison of NORGs Among Published Genomes

All genomes published (as of July 20, 2014) that had both organelle genomes sequenced were used for a comparative study of organellar DNA introgression. These include *Arabidopsis* (Unseld et al. 1997; Lin et al. 1999; Mayer et al. 1999; Sato et al. 1999; Salanoubat et al. 2000; Tabata et al. 2000; Theologis et al. 2000), *Carica* (Ming et al. 2008; Rice et al. unpublished), *Glycine* (Saski et al. 2005; Schmutz et al. 2010; Chang et al. 2013), *Oryza* (Hiratsuka et al. 1989; Notsu et al. 2002; Tanaka et al. 2008), *Sorghum* (Saski et al. 2007; Paterson et al. 2009), *Vitis* (Jansen et al. 2006; Jaillon et al. 2007; Goremykin et al. 2009), and *Zea* (Maier et al. 1995; Clifton et al. 2004; Schnable et al. 2009). BLASTN searches between the plastid and nuclear genomes were done using only one copy of the IR. Searches of the mitochondrial genomes were done without removal of duplications, which may lead to a slight overestimate of the amount of nuclear introgression. Analyses were done using both low and high stringency parameters as outlined above. As with the *Nelumbo* comparisons, the amount of plastid DNA in the mitochondrial genome (and visa versa) was also calculated using a percent identity cutoff of 90 and both low and high stringency searches.

Results

'China Antique' Plastome

Coverage of the *N. nucifera* 'China Antique' plastome was very deep, with a maximum depth of 78,699 reads and an average of 70,000, and assembly of these data resulted in a single plastid contig. The 'China Antique' plastome (GenBank accession NC_025339) is 163,330 bp in size (Fig. 2.1). The large single copy (LSC), small single copy (SSC), and inverted

repeat (IR) regions each span 91,910 bp, 19,358 bp, and 26,031 bp, respectively (Table 2.1). The plastome codes for 115 genes, of which 34 are RNA and 17 are completely contained within the IR. The genome is wholly collinear with the *Nicotiana tabacum* plastome, with all IR-single copy junctions occurring in the same relative positions (Shinozaki et al. 1986). The only inconsistent feature of the 'China Antique' plastome in comparison to *Nicotiana* is that the gene *rpl2* has undergone a mutation at the accepted start codon location. Instead of the codon ATG in that position, the codon is ACG. The closest alternative start codon is 12 bp upstream and this is an ATA codon.

The plastome of 'China Antique' has 30 short dispersed repeats (SDRs), with the SSC and IR regions lacking short inverted repeats (Fig. S2.1). The longest SDRs in the LSC, SSC, and IR regions are 74 bp, 33 bp, and 85 bp, respectively. There are 46 simple sequence repeats (SSRs), the majority of which are A/T mononucleotide repeats (Table S2.1).

Comparison of Nelumbo *Plastomes*

The *N. nucifera* 'China Antique' plastome was compared to four complete plastomes of *N. nucifera* and *N. lutea* (Xue et al. 2012). All *Nelumbo* plastomes are collinear and share the same alternate start codon identified for the 'China Antique' *rpl2* gene. Plastomes from the three *N. nucifera* accessions range in size from 163,307 bp to 163,639 bp, representing a 332 bp size difference, while plastomes from the two *N. lutea* accessions range in size from 163,206 bp to 163,510 bp, representing a 304 bp size difference (Table 2.1). Plastome size range differences do not reflect species designations and there is a 433 bp disparity in size between plastomes of *N. nucifera* and *N. lutea*. The smallest and second largest plastomes belong to *N. lutea,* while the two middle-sized and largest plastomes belong to *N. nucifera*. These size differences not only vary among accessions, but also among genome compartments. 'China Antique'

(NC_025339) has the largest LSC region, the two other *N. nucifera* accessions have the largest IRs, and *N. nucifera* accession JQ336993 has the largest SSC region (Table 2.1).

An alignment of the five plastome sequences was examined for point mutations, indels, and repeat motif length differences. While the locations of repeat regions are shared among all accessions, variation in repeat size occurs within and between species (Fig. 2.2). 'China Antique' has the most SDRs among all accessions (30) and the three *N. nucifera* accessions have the most (17-22) and largest (85 bp) direct repeats. The *N. lutea* accessions have the largest number of inverted repeats (11-12).

Two of the three *N. nucifera* accessions (NC_025339 and JQ336993) have identical numbers of SSRs (46), although their composition is slightly different. One additional SSR was detected in *N. nucifera* accession NC_015610. Both *N. lutea* accessions have 54 SSRs, but with fewer C+G motifs than in *N. nucifera*. The *N. lutea* accessions also have an additional trinucleotide SSR not detected in *N. nucifera*.

Most of the sequence variation detected among the five *Nelumbo* accessions occurs within the first 20 kb of the LSC region. These differences are due to several small, tandem repeats, averaging about 6 bp in length, and varying lengths of the SSRs. Additionally, for the *N. nucifera* accessions, there is a 167 bp insertion in the *psbA* and *trnK* intergenic spacer and a 176 bp insertion between *ndhC* and *trnV*. The *trnT-trnE* intergenic spacer is also variable in length, ranging between 827 bp in *N. nucifera* and 1018 bp in *N. lutea*. Furthermore, intraspecific variation is apparent, with length differences of 17 bp and 8 bp occurring in accessions of *N. lutea* and *N. nucifera*, respectively. The most striking variable region within the *trnT-trnE* spacer is a 127 - 293 bp tandem, imperfect, A+T repeat. In *N. lutea* this repeat is 264 – 293 bp in size, much larger than the 127 – 134 bp repeat occurring in *N. nucifera*. In addition, within this large repeat, *N. nucifera* has an inverted repeat of 54 bp in accessions NC_025339

and NC_015610, but not in JQ336993, and an inverted repeat of 64 bp in *N. lutea* accession JQ336992, but not in NC_015605.

Plastomes from two of the three *N. nucifera* accessions have identical IR lengths, while the 'China Antique' IR has 34 fewer nucleotides (Table 2.1). This difference is due to one bp difference in a mononucleotide repeat, a 15 bp indel, a 6 bp difference in the LSC/IR boundary, and a 15 bp contraction of the 'China Antique' SSC/IR boundary. IRs of the *N. lutea* accessions differ in length by only two nucleotides. This variation is due to a 6 bp difference in the LSC/IR boundary and a 4 bp difference in the SSC/IR boundary. Comparisons between the IRs of *N. nucifera* and *N. lutea* result in a total of 91 differences, including 14 point mutations, that are largely attributable to small changes in IR boundary positions (≤ 15 bp) and variations in mononucleotide repeat length.

Among the three plastome compartments, the size of the SSC region differs most across all accessions (up to 290 bp in *N. nucifera* and 269 bp in *N. lutea*; Table 2.1). Within *N. lutea*, SSC size differences are explained by three indels, two repeats, and positions of the SSC/IR junctions. In *N. lutea* JQ336992 there is a 282 bp insertion in the *ndhA* intron, whereas in *N. lutea* NC_015605 this insertion is only 22 bp in size. Surprisingly, *N. lutea* accession JQ336992 shares the same large insertion within the *ndhA* intron as does *N. nucifera* accession JQ336993. This insertion is not found in the other *Nelumbo* plastomes. An additional difference between the two species includes the amount of *ycf1* retained within the SSC region (25 bp). The remaining length variations are accounted for by differences in repeat DNA.

The majority of point mutations occur within the LSC region (Table S2.2). There are fewer mutations from C to G and from G to C than any other point mutation. Mutations from T to A, A to T, and T to G are the most prevalent. This trend of minimal C/G mutations is consistent within all plastome compartments. Within the two single copy regions the percentage of types of mutations is similar, with the exception of T to A mutations that occur twice as often in the LSC

region, even when the large size of this region is accounted for. Within the IR, A to C mutations are the most frequent, followed by T to C mutations.

'China Antique' Mitochondrial Genome

The initial draft of the *N. nucifera* mitochondrial genome consisted of 21 contigs totaling approximately 450 kb. The final draft genome has 12 contigs and 454,603 bp (Fig. 2.3; GenBank accessions AQOG01058426-AQOG01058443). The contigs are oriented and ordered as they are hypothesized to be joined based on evidence from PCR and scaffolding. The exception is contig 12, shown as separate from the remaining contigs, which has no supported connectivity with the rest of the genome. Within these 12 contigs, 43 protein coding genes were identified, including 2 rRNAs, and 22 tRNAs (Table S2.3). There are also 14 mitochondrial gene fragments (called pseudogenes herein) that are likely partial or degenerate duplications. The 'China Antique' mitochondrial genome has all of the expected protein coding genes, with the exception of *nad6*.

Within the draft genome there are numerous plastid-derived pseudogenes (10 protein coding and one tRNA) termed MTPTs (Smith 2011; Wang et al. 2012; Sloan and Wu 2014). Transfer RNA genes that are plastid-derived (labeled "-cp" in Table S2.3) are counted as mitochondrial genes due to their incorporation and probable use by the mitochondrial genome (Dietrich et al. 1996; Adams et al. 2002). Other MTPTs within the mitochondrial contigs, such as rRNA genes and protein-coding pseudogenes, are counted as plastid-derived pseudogenes, as they are unlikely to be transcribed or translated due to their fragmented nature.

Coding sequence was not evenly dispersed among contigs. Contig 5 has no complete genes, while contigs 3, 6, 9, 11, and 12 have at least 6 coding regions and contig 1 has 20 (Fig. S2.2). Contig 1 also has the most protein-coding genes and tRNAs, contig 4 has the most rRNAs, and contig 3 has the most pseudogenes.

Within the draft mitochondrial genome, 95 SSRs were detected (Table 2.2). Mononucleotide repeats are dominated by A+T motifs, both in abundance and in length. There is more diversity in the number and base pair composition of dinucleotide repeats than of mononucleotide repeats. Dinucleotide repeats with an AG or GA motif are the most prevalent, while the longest repeat had 9 AT/TA duplications. There are only 5 instances of trinucleotide repeats. 'China Antique' has over 3,000 small SDRs (between 30 and 50 bp) and hundreds of larger SDRs. However, there are no repeats larger than 1 kb within the draft genome. Direct and inverted repeats are equally represented in all size classes.

Comparative Mitochondrial Genomics

Of the complete mitochondrial genomes available on GenBank, the closest relative to *Nelumbo* is *Beta* of the family Amaranthaceae (Kubo et al. 2000). The *Beta* mitochondrial genome has 29 protein coding genes, all of which are shared with *Nelumbo* with the exception of *tatC*, which is found only in *Beta*. *Nelumbo* has 12 additional protein coding genes predicted. There are 20 tRNA genes in common between the two genomes. However, there are several tRNA genes predicted for *Nelumbo* that are not predicted for *Beta*. *Beta* has only one tRNA that is not predicted in *Nelumbo*. The rRNA genes are conserved. Broadening the comparison reveals that 'China Antique' has more duplications of mitochondrial protein-coding genes in the form of gene fragments than the other mitochondrial genomes considered herein (Table S2.3). However, when plastid pseudogenes are considered, *Vitis* has the most (69).

'China Antique' has 4 to 16 times more repeats 30-50 bp in length than any of the other seven mitochondrial genomes (Table 2.3). *Vitis* and 'China Antique' have the most similar pattern in repeat size, with all dispersed repeats less than 1 kb. *Carica* and *Glycine* have fewest small repeats (less than 1 kb), but also have several of the largest repeats (1 – 20 kb). The

monocot genomes have the largest repeats, with *Oryza* having a single direct repeat of over 40 kb in size. The largest inverted repeat occurs in *Zea* (16,870 bp).

Examining the percentage of mitochondrial DNA shared in pairwise comparisons, 'China Antique' shares more DNA with *Vitis* (30.84%) than it does with its closest relative *Beta* (20.53%; Table S2.4). 'China Antique' has the least similarity to *Zea* (19.09%) and *Arabidopsis* (19.12%). Taxonomy is not a good predictor of how much DNA will be shared among taxa. *Sorghum* and *Oryza* share with *Zea* 56.20% and 47.28% of their DNA, respectively, and *Zea* shares 46.23% of its DNA with *Sorghum*; however, the *Oryza* and *Sorghum* genomes have only 17.06 - 17.86% of their DNA in common, depending on the directionality of the comparison.

Plastid-Derived Mitochondrial DNA

Within 'China Antique' there are MTPTs in six of the 12 mitochondrial contigs (Fig. 2.3; Fig. S2.2). The majority of these are rRNA and photosystem pseudogenes. *Carica* and *Zea* each have over 12 kb of contiguous MTPTs within their mitochondrial genomes. *Sorghum*, *Oryza*, *Nelumbo*, and *Vitis* all have MTPTs ranging from 1 – 6 kb. *Arabidopsis* and *Glycine* have smaller fragments of plastid DNA, all under 1 kb. These fragments, when summed and divided by the total size of the genome, are reported as percentages of introgression in Table 2.4. Amongst 'China Antique', *Arabidopsis*, and *Glycine,* the total percentage of MTPT within the mitochondrion is comparable, at 1 to 1.6% (Table 2.4). The *Vitis* mitochondrial genome has the highest percentage of MTPT (8.14%), followed by the monocot species at 4.32 – 7.07%.

Organelle DNA Introgression Into the Nuclear Genome

The total percentage of NUPTs within 'China Antique' was low (Table 2.4). There are only 143 instances of transfer in this direction, totaling 35,836 bp (93.7% of which is non-coding DNA). This amount of introgression is less than that detected in the other species. Identified

fragments range in size from 101 to 1128 bp (Fig. 2.4). Among the eudicots there are no NUPTs larger than 8 kb (*Glycine)*. *Carica*, *Glycine*, and *Vitis* all have comparable percentages of NUPTs (0.052 – 0.088%). Characterization of NUPTs differs among eudicots examined: *Glycine* has fewer, larger NUPTs; *Carica* and *Vitis* have more, shorter fragments. Within the monocot genomes, *Zea* has more, larger fragments (≥30,000 bp) than either *Oryza* or *Sorghum,* although *Oryza* has the most NUPTs overall (0.267% of the nuclear genome).

In 'China Antique', there are 126 instances of NUMTs for a total of 29,163 bp (Table 2.4). There are fewer NUMTs in 'China Antique' than in any of the other species (Fig. 2.4). Within 'China Antique' the majority of NUMTs match non-coding DNA (78.7%), with their sizes ranging from 100 to 1172 bp. There are only 23 NUMTs that match mtDNA coding sequence and these range in size from 105 to 729 bp. The largest fragments of organelle DNA within the nuclear genome tend to be mitochondrial in origin and the ratio is especially biased in the eudicot genomes analyzed. Within these genomes, *Arabidopsis* has the most NUMTs with 0.411% of the nuclear genome made up of mitochondrial DNA. *Arabidopsis* also has the largest NUMTs. The monocot genomes have a broader size range of NUMTs than the eudicots. As with NUPTs, *Oryza* has the most NUMTs, totaling 0.252% of the nuclear genome. While *Oryza* has the most total base pairs of NUMT DNA, *Zea* has the largest fragments and is the only monocot genome to have NUMT fragments ≥ 30 kb.

Discussion

'China Antique' Plastid Genome and Intraspecific Comparisons

Nelumbo nucifera 'China Antique' has a typical land plant plastome, with no structural mutations or gene adjacency changes from plastomes having an organization considered ancestral within angiosperms, such as *Amborella trichopoda* (Goremykin et al. 2003) and *Nicotiana tabacum* (Shinozaki et al. 1986). The only inconsistency is the alternative start codon

hypothesized for the gene *rpl2*, where the codon ATG is replaced by ACG. This same point mutation occurs within all other accessions of *Nelumbo*, as well as in many other land plants (such as *Amborella*, some magnoliids, Chloranthaceae, Ceratophyllaceae, some monocots, and some core eudicots), therefore the presence of an ACG start codon in *Nelumbo* is not remarkable.

Differences among the five *Nelumbo* plastomes are due primarily to point mutations, several large indels, and repeat motif length differences. Repeat DNA, specifically mononucleotide repeats adjacent to IR boundaries, is likely the cause of the observed, small boundary shifts. Xue et al. (2012) investigated SSR diversity in *Nelumbo* and reported 38 SSR loci, eight fewer than are present for 'China Antique' and 16 fewer than in *N. lutea*. The methods used by Xue et al. (2012) and ourselves to detect SSRs require motifs to repeat at least five times; in our study, however, these anlyses required a minimum length of 10 bp for mononucleotide repeats, whereas they only required six. The SSR Hunter v. 1.3 (Li and Wan 2005) program used by Xue et al. (2012) appears to be underestimating the total number of SSRs. Unsurprisingly for an A+T rich plastome, the majority of SSRs are A's or T's. 'China Antique' has 30 SDRs and this number is comparable to what has been reported for the plastomes of *Arabidopsis* (Sato et al. 1999), *Vitis* (Jansen et al. 2006), *Sorghum* (Saski et al. 2007), *Oryza* (Hiratsuka et al. 1989), and *Zea* (Maier et al. 1995).

Other than the search for microsatellite loci in the plastomes of four populations of *Nelumbo* (Xue et al. 2012), this is the first study to report on intraspecific plastome variation within the Proteales. Indeed, such studies of plastome intraspecific variation in other major lineages of flowering plants are generally few. Cultivars of *Solanum lycopersicum*, *Jacobaea vulgaris*, and *Oryza sativa* var. *indica*, as examples, have a much lower plastome genetic diversity than what is reported for *Nelumbo lutea* or *Nelumbo nucifera* (Tang et al. 2004; Kahlau et al. 2006; Doorduin et al. 2011). The only other study of intraspecific comparisons to find

similar levels of SNPs and sequence length differences is that of *Colocasia esculenta* (Ahmed et al. 2012). However, these length differences are due mostly to where *rps19* straddles the IR. In *Nelumbo,* the IR has a more conserved length, with variation in size related to a large insertion within the *ndhA* intron plus other smaller indels.

At the intrageneric level, a comparison of chloroplast genomes of *Camellia* species shows the same trend of low diversity among individuals (Yang et al. 2013). In contrast, the differences in length and SNPs between the two *Nelumbo* species are similar to what was found among 12 *Gossypium* (Xu et al. 2012) and seven *Camellia* species (Yang et al. 2013). Additional intrageneric studies are necessary to determine if the levels of divergence seen between *Nelumbo nucifera* and *Nelumbo lutea* are high or low in relation to what has been reported in other genera.

In *N. lutea* JQ336992 and *N. nucifera* JQ336993 there is a 282 bp insertion in the *ndhA* intron that is not present in the other *Nelumbo* plastomes. It is surprising that these two accessions share this insertion while the other three do not. The original publication of *Nelumbo* plastid microsatellites does not detail the variety or cultivar names of the *Nelumbo* accessions examined (Xue et al. 2012). If such information was known, then paternal relationships of the accessions could provide hypotheses as to why JQ336992 and JQ336993 share this insertion while the other accessions do not. There is strong potential for interspecific hybridization during cultivation and without further information on source material, paternal and maternal contributions to the genome cannot be explored.

'China Antique' Mitochondrial Genome and Comparative Mitochondrial Genomics

The draft mitochondrial genome of 'China Antique' has all of the expected genes for an angiosperm with the exception of *nad6*. This is noteworthy considering *nad6* is present in all other genomes examined. Therefore, it is likely that a small portion mitochondrial genome

containing this region is missing. Comparing the location of *nad6* in other genomes is not helpful in knowing what portion or if any of the 'China Antique' genic sequence is potentially missing, because the position and gene adjacencies of *nad6* are different in all other genomes examined.

Each mitochondrial genome sequenced to date has a unique order of genes and genic content; however, there are some gene clusters that are conserved. For example, the gene clusters *nad5-nad4L-ORF25* and *nad2-rps12*, predicted from early mitochondrial genome studies (Unseld et al. 1997), are broken up in 'China Antique'. In *Carica*, Ming et al. (2008) reported that *rrn5* and *rrn18* are linked, as is the clustering of *atp4-nad4L* and *cob-rps14-rpl5*. Within 'China Antique', these same gene clusters are retained. In addition, the gene order *rpl16-rps3-rps19-rpl2* in 'China Antique' is collinear with that occurring in *Nicotiana*, *Arabidopsis*, *Zea*, and *Vitis*, as is the position of *nad3* adjacent to *rps12*.

Within the draft mitochondrial genome, coding regions tend to cluster and are not evenly distributed among or within the contigs. This clustering of coding DNA may help with retaining genic material since the mitochondrial genome is constantly rearranging, accepting, and losing DNA. If coding DNA is clustered there is less chance of rearrangements breaking up operons or otherwise disrupting essential processes required by the plant.

The draft 'China Antique' mitochondrial genome has the smallest SDRs of the eight genomes compared. Monocot genomes tend to have larger and more SDRs than eudicot genomes, such as the 120 kb repeat in maize (Allen et al. 2007), while eudicot genomes have fewer large repeats. The differences in SDR number and size between the results I report herein and those from each of the original publications of the genomes I compared are due to the different parameters and algorithms used to determine amounts of shared and repeat DNA.

Following the pattern established for genic and repeat DNA content, each mitochondrial genome has varying amounts of MTPTs. The process of intracellular transfer of nuclear or plastid DNA into the mitochondrion is useful since the successfully integrated genes have a

chance to develop new functions (Wang et al. 2012). However, this process is not essential to mitochondrial genome function, since only 1% of the *Arabidopsis* genome is attributed to MTPTs and another 4% of it is identified as being nuclear in origin (Unseld et al. 1997). The majority of MTPTs within 'China Antique' were identified as tRNAs and photosystem genes. The overall number of transfers was relatively low, especially in comparison to *Glycine* and *Vitis*. In the analysis of *Glycine* MTPTs, Chang et al. (2013) detected 7.1 kb of plastid DNA, while in *Vitis* almost 50% of the plastome is duplicated within its mitochondrial genome (Goremykin et al. 2009). The amount of MTPT does not correlate to the number or sizes of repeat DNA currently present within the mitochondrial genomes. With the exception of the hypothesis relating to acquiring new gene function, little is known about why and how MTPTs occur. Additional empirical studies of the mechanism and frequency of DNA introgression into the mitochondrial genome, such as the studies of double-stranded break repair in *Arabidopsis* (Davila et al. 2011) and yeast (Ricchetti et al. 1999), are needed to further understand the processes involved in, and consequences of, MTPTs.

As more mitochondrial genomes become available for analysis, it is clear that there are no rules for predicting how similar mitochondrial genomes may be, for even closely related species can have very different genomes (Kubo and Newton 2008; Darracq et al. 2011). Taxonomic relationship is a poor indicator of predicted size of a genome and the amount of shared DNA (Palmer et al. 2000; Alverson et al. 2010). Even with this caveat of mitochondrial genome non-comformity, given the completeness of the coding DNA found within the draft mitochondrial genome presented herein, it is likely that the majority of the genome is present.

Comparison of Organelle DNA Introgression – NORGs

Introgression of organellar DNA into the nuclear genome is not scattered across chromosomes but concentrated on only a few (Yoshida et al. 2013). As an example, in

Arabidopsis, chromosome 2 has large amounts of NUMTs (Lin et al. 1999). However, the large plastid insertion of 620 kb reported by Stupar et al. (2001) in *Arabidopsis* was not detected using our search methods – I found no NUPTs over 10 kb. This is likely due to the degenerative nature of the insertion and our search parameters, which found several smaller NUPTs rather than few larger ones. Within *Glycine*, NORGs were detected on all but one chromosome; however, introgressions were concentrated near centromeres and on chromosome 17 (Chang et al. 2013). *Zea* also has biases in location of NORGs, with NUMTs concentrated on chromosome 1 (25 of 43 fragments), but in discontinuous order from that occurring on its mitochondrial genome (Notsue et al. 2002).

Among monocot genomes, only three grasses have annotated mitochondrial and plastid genomes available. Thus, it is unclear if the abundance of NORGs in *Oryza* is unique among monocots, or if similar large numbers might be found elsewhere. Among eudicots, *Arabidopsis* has the most NUMTs, while *Vitis* has many NUPTs. The amount of NORGs within 'China Antique' is much lower than that of other taxa, with less NUPTs and NUMTs. This paucity of NORGs begs the question – what is so different about *N. nucifera*? The composition of its nuclear genome is within expected norms, with all standard eukaryotic genes present and possession of a typical number of repeat elements (Ming et al. 2013). 'China Antique' is unique among eudicots in its gene distribution, rates of nucleotide substitution, and having only one paleo-duplication (Ming et al. 2013). With less of the genome available in 'China Antique,' it may be more difficult for integration to be retained in further generations (Wang and Timmis 2013). In addition, the 'China Antique' nuclear genome has a 30% reduction in genome-wide mutation rate in comparison to *Vitis*, and this may reduce the likelihood of successful organelle DNA integration into its nuclear genome (Ming et al. 2013).

Double-stranded break repair is reported to be the most frequent mechanism causing NORGs (Hazkani-Covo et al. 2010). With the high density of coding sequence of the nuclear

genome and its low mutation rate perhaps there are fewer non-fatal double-stranded breaks and therefore fewer NORGs in *Nelumbo*. The mode of propagation of *Nelumbo* may also be a factor leading to the lack of NORGs, for the absence of sexual reproduction will result in fewer instances of integration. The sequencing of additional basal eudicots outside the core eudicot group, especially from within the ANITA grade and magnoliids, will help illuminate if the density of coding sequence or propagation methods affect the accumulation of NUMTs and NUPTs.

References

Adams KL, Daley DO, Whelan J, Palmer JD (2002) Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. Plant Cell 14:931–943.

A.G.I. [Arabidopsis Genome Initiative] (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-814.

Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD, Lockhart PJ (2012) Mutational dynamics of aroid chloroplast genomes. Genome Biol Evol 4:1316–1323.

Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177:1173-1192.

Alverson AJ, Wyman SK, Boore JL (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27:1436-1448.

A.P.G. [Angiosperm Phylogeny Group] III (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linnean Soc 161:105-121.

Blanehard JL, Schmidt GW (1995) Pervasive migration of organellar DNA to the nucleus in plants. J Mol Evol 41:397-406.

Chang S, Wang Y, Lu J, Gai J, Li J, Chu P, Guan R, Zhao T (2013) The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels PLoS One 8:e56502.

Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, Meyer L, Wilson RK, Newton KJ (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol 136:3486-503.

Darracq A, Varre JS, Marechal-Drouard L, Courseaux A, Castric V, Saumitou-Laprade P, Oztas S, Lenoble P, Vacherie B, Barbe V, Touzet P (2011) Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. Genome Biol Evol 3:723-736.

Davila JI, Arrieta-Montiel AP, Wambolt Y, Cao J, Hagmann J, Shedge V, Xu Y-Z, Weigel D, Mackenzie SA (2011) Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. BMC Biol 9:64.

Dietrich A, Small I, Cosset A, Weil J, Marechal-Drouard L (1996) Editing and import: strategies for providing plant mitochondria with a complete set of functional transfer RNAs. Biochimie 78:518–529.

Dolan A. Simple Sequence Repeat (SSR) Extractor Utility. http://www.aridolan.com/ssr/ssr.aspx. Accessed 23 July, 2014.

Doorduin L, Gravendeel B, Lammers Y, Ariyurek Y, Chin-A-Woeng T, Vrieling K (20110 The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA Res 18:93-105.

Ensikat HJ, Ditsche-Kuru P, Neinhuis C, Barthlott W (2011) Superhydrophobicity in perfection: the outstanding properties of the lotus leaf. Beilstein J Nanotechnol 2:152–161.

Goremykin VV, Hirsch-Ernst KI, Wolfl S, Hellwig, FH (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. Mol Biol Evol 20:1499-1505.

Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol 26: 99-110.

Gould SB, Waller RF, McFadden GI (2008) Plastid evolution. Annu Rev Plant Biol 59:491–517.

Guo HB. 2009. Cultivation of lotus (*Nelumbo nucifera* Gaertn. ssp. *nucifera*) and its utilization in China. Genet Resour Crop Ev 56:323–330.

Guo X, Ruan S, Hu W, Cai D, Fan L (2008) Chloroplast DNA insertions into the nuclear genome of rice: the genes, sites, and ages of insertions involved. Funct Integr Genomics 8:101-108.

Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular poltergeists: mitochondrial DNA copies (NUMTs) in sequenced nuclear genomes. PLoS Genet 6:e1000834.

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185-194.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P, French-Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467.

Jansen RK, Kaittanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, Daniell H (2006) Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol 6:32.

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

Kahlau S, Aspinall S, Gray JC, Bock R (2006) Comparison of solanaceous plastid genomes. J Mol Evol 63:194–207.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639-1645.

Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. Mitochondrion 8:5-14.

Kubo T, Nishizawa S, Sugawara A, Itchoda N., Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res 28:2571-2576.

Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organellar DNA. TRENDS Genet 21:655-663.

Li Q, Wan JM (2005) SSR Hunter: development of a local searching software for SSR sites. Hereditas 27:808-810.

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldblyum TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Shen M, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Nierman WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter JC (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761-768.

Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614-628.

Martin W (2003) Gene transfer from organelles to the nucleus: frequent and in big chunks. P Natl Acad Sci USA 100:8612–8614.

Martin W, Hermann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol 118:9-17.

Matsuo M, Ito Y, Yamauchi R, Obokata J (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. Plant Cell 17:665-675.

Mayer K, Schüller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Düsterhöft A, Stiekema W, Entian KD, Terryn N, Harris B, Ansorge W, Brandt P, Grivell L, Rieger M, Weichselgartner M, de Simone V, Obermaier B, Mache R, Müller M, Kreis M, Delseny M, Puigdomenech P, Watson M, Schmidtheini T, Reichert B, Portatelle D, Perez-Alonso M, Boutry M, Bancroft I, Vos P, Hoheisel J, Zimmermann W, Wedler H, Ridley P, Langham SA, McCullagh B, Bilham L, Robben J, Van der Schueren J, Grymonprez B, Chuang YJ, Vandenbussche F, Braeken M, Weltjens I, Voet M, Bastiaens I, Aert R, Defoor E, Weitzenegger T, Bothe G, Ramsperger U, Hilbert H, Braun M, Holzer E, Brandt A, Peters S, van Staveren M, Dirske W, Mooijman P, Klein Lankhorst R, Rose M, Hauf J, Kötter P, Berneiser S, Hempel S, Feldpausch M, Lamberth S, Van den Daele H, De Keyser A, Buysshaert C, Gielen J, Villarroel R, De Clercq R, Van Montagu M, Rogers J, Cronin A, Quail M, Bray-Allen S, Clark L, Doggett J, Hall S, Kay M, Lennard N, McLay K, Mayes R, Pettett A, Rajandream MA, Lyne M, Benes V, Rechmann S, Borkova D, Blöcker H, Scharfe M, Grimm M, Löhnert TH, Dose S, de Haan M, Maarse A, Schäfer M, Müller-Auer S, Gabel C, Fuchs M, Fartmann B, Granderath K, Dauner D, Herzl A, Neumann S, Argiriou A, Vitale D, Liguori R, Piravandi E, Massenet O, Quigley F, Clabauld G, Mündlein A, Felber R, Schnabl S, Hiller R, Schmidt W, Lecharny A, Aubourg S, Chefdor F, Cooke R, Berger C, Montfort A, Casacuberta E, Gibbons T, Weber N, Vandenbol M, Bargues M, Terol J, Torres A, Perez-Perez A, Purnelle B, Bent E, Johnson S, Tacon D, Jesse T, Heijnen L, Schwarz S, Scholler P, Heber S, Francs P, Bielke C, Frishman D, Haase D, Lemcke K, Mewes HW, Stocker S, Zaccaria P, Bevan M, Wilson RK, de la Bastide M, Habermann K, Parnell L, Dedhia N, Gnoj L, Schutz K, Huang E, Spiegel L, Sehkon M,

Murray J, Sheet P, Cordes M, Abu-Threideh J, Stoneking T, Kalicki J, Graves T, Harmon G, Edwards J, Latreille P, Courtney L, Cloud J, Abbott A, Scott K, Johnson D, Minx P, Bentley D, Fulton B, Miller N, Greco T, Kemp K, Kramer J, Fulton L, Mardis E, Dante M, Pepin K, Hillier L, Nelson J, Spieth J, Ryan E, Andrews S, Geisel C, Layman D, Du H, Ali J, Berghoff A, Jones K, Drone K, Cotton M, Joshu C, Antonoiu B, Zidanic M, Strong C, Sun H, Lamar B, Yordan C, Ma P, Zhong J, Preston R, Vil D, Shekher M, Matero A, Shah R, Swaby IK, O'Shaughnessy A, Rodriguez M, Hoffmann J, Till S, Granat S, Shohdy N, Hasegawa A, Hameed A, Lodhi M, Johnson A, Chen E, Marra M, Martienssen R, McCombie WR (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature 402:769-777.

Meinke, DW, Cherry JM, Dean C, Rounsley SD, Koornneef M (1998) *Arabidopsis thaliana*: a model plant for genome analysis. Science 282:662-665.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Pérez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452:991-996.

Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li L-T, Zhang Q, Kim M-J, Schatz MC, Campbell M, Li J, Bowers JE, Tang H, Lyons E, Ferguson AA, Narzisi G, Nelson DR, Blaby-Haas CE, Gschwend AR, Jiao Y, Der JP, Zeng F, Han J, Min XJ, Hudson KA, Singh R, Grennan AK, Karpowicz SJ, Watling JR, Ito K, Robinson SA, Hudson ME, Yu Q, Mockler TC, Carroll A, Zheng Y, Sunkar R, Jia R, Chen N, Arro J, Wai CM, Wafula E, Spence A, Han Y, Xu L, Zhang J, Peery R, Haus MJ, Xiong W, Walsh JA, Wu J, Wang M-L, ZhuYJ, Paull RE, Britt AB, Du C, Downie SR, Schuler MA, Michael TP, Long SP, Ort DR, Schopf JW, Gang DR, Jiang N, Yandell M, dePamphilis CW, Merchant SS, Paterson AH, Buchanan BB, Li S, Shen-Miller J (2013) Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). Genome Biol 14:R41.

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics 268:434-445.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. P Natl Acad Sci USA 97:6960–6966.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551-556.

Raubeson LA, Peery R, Chumley TW, Dzuibek C, Fourcade HM, Boore J, Jansen RK (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics 8:174.

Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs doublestrand breaks in yeast chromosomes. Nature 402:96-100.

Rice DW, Palmer JD (2006) An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. BMC Biol 4:31.

Rice DW, Saw JH, Yu Q, Feng Y, Wang W, Hou S, Dionne-Laporte A, Ly B, Lewis K, Hall CM, Senin P, Ming R, Wang L, Alam M, Plamer JD (Unpublished) Analysis of the plastid and mitochondrial genomes of papaya, organelle-like DNA in the nucleus, and substoichiometric mitochondrial configurations.

Salanoubat M, Lemcke K, Rieger M, Ansorge W, Unseld M, Fartmann B, Valle G, Blöcker H, Perez-Alonso M, Obermaier B, Delseny M, Boutry M, Grivell LA, Mache R, Puigdomènech P, De Simone V, Choisne N, Artiguenave F, Robert C, Brottier P, Wincker P, Cattolico L, Weissenbach J, Saurin W, Quétier F, Schäfer M, Müller-Auer S, Gabel C, Fuchs M, Benes V, Wurmbach E, Drzonek H, Erfle H, Jordan N, Bangert S, Wiedelmann R, Kranz H, Voss H, Holland R, Brandt P, Nyakatura G, Vezzi A, D'Angelo M, Pallavicini A, Toppo S, Simionati B, Conrad A, Hornischer K, Kauer G, Löhnert TH, Nordsiek G, Reichelt J, Scharfe M, Schön O, Bargues M, Terol J, Climent J, Navarro P, Collado C, Perez-Perez A, Ottenwälder B, Duchemin D, Cooke R, Laudie M, Berger-Llauro C, Purnelle B, Masuy D, de Haan M, Maarse AC, Alcaraz JP, Cottet A, Casacuberta E, Monfort A, Argiriou A, flores M, Liguori R, Vitale D, Mannhaupt G, Haase D, Schoof H, Rudd S, Zaccaria P, Mewes HW, Mayer KF, Kaul S, Town CD, Koo HL, Tallon LJ, Jenkins J, Rooney T, Rizzo M, Walts A, Utterback T, Fujii CY, Shea TP, Creasy TH, Haas B, Maiti R, Wu D, Peterson J, Van Aken S, Pai G, Militscher J, Sellers P, Gill JE, Feldblyum TV, Preuss D, Lin X, Nierman WC, Salzberg SL, White O, Venter JC, Fraser CM, Kaneko T, Nakamura Y, Sato S, Kato T, Asamizu E, Sasamoto S, Kimura T, Idesawa K, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakayama S, Nakazaki N, Shinpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S; European Union Chromosome 3 Arabidopsis Sequencing Consortium; Institute for Genomic Research; Kazusa DNA Research Institute (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. Nature 408:820-822.

Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG, Jansen RK (2005) Complete chloroplast genome sequence of *Gycine max* and comparative analyses with other legume genomes. Plant Mol Biol 59:309-322.

Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, Tomkins J, Rognli OA, Daniell H, Clarke JL (2007) Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. Theor Appl Genet 115:571-590.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res 6:283-290.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2013) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115.

Shen-Miller J (2002a) Sacred lotus, the longliving fruits of China Antique. Seed Sci Res. 12:131-143.

Shen-Miller J, Schopf JW, Harbottle G, Cao RJ, Ouyang S, Zhou KS, Southon JR, Liu GH (2002b) Long-living lotus: germination and soil γ-irradiation of centuries-old fruits, and cultivation, growth, and phenotypic abnormalities of offspring. Am J Bot 89:236–247.

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043-2049

Sloan DB, Wu Z (2014) History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. Genome Biol Evol 6:3210-3221.

Smith DR (2011) Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol Evol 3:743-748.

Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang J (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of 138 potential sequencing errors caused by large-unit repeats. P Natl Acad Sci USA 98:5099–5103

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics 272:603-615.

Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, Miyajima N, Sasamoto S, Kimura T, Hosouchi T, Kawashima K, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakayama S, Nakazaki N, Naruo K, Okumura S, Shinpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Sato S, de la Bastide M, Huang E, Spiegel L, Gnoj L, O'Shaughnessy A, Preston R, Habermann K, Murray J, Johnson D, Rohlfing T, Nelson J, Stoneking T, Pepin K, Spieth J, Sekhon M, Armstrong J, Becker M, Belter E, Cordum H,

Cordes M, Courtney L, Courtney W, Dante M, Du H, Edwards J, Fryman J, Haakensen B, Lamar E, Latreille P, Leonard S, Meyer R, Mulvaney E, Ozersky P, Riley A, Strowmatt C, Wagner-McPherson C, Wollam A, Yoakum M, Bell M, Dedhia N, Parnell L, Shah R, Rodriguez M, See LH, Vil D, Baker J, Kirchoff K, Toth K, King L, Bahret A, Miller B, Marra M, Martienssen R, McCombie WR, Wilson RK, Murphy G, Bancroft I, Volckaert G, Wambutt R, Düsterhöft A, Stiekema W, Pohl T, Entian KD, Terryn N, Hartley N, Bent E, Johnson S, Langham SA, McCullagh B, Robben J, Grymonprez B, Zimmermann W, Ramsperger U, Wedler H, Balke K, Wedler E, Peters S, van Staveren M, Dirkse W, Mooijman P, Lankhorst RK, Weitzenegger T, Bothe G, Rose M, Hauf J, Berneiser S, Hempel S, Feldpausch M, Lamberth S, Villarroel R, Gielen J, Ardiles W, Bents O, Lemcke K, Kolesov G, Mayer K, Rudd S, Schoof H, Schueller C, Zaccaria P, Mewes HW, Bevan M, Fransz P; Kazusa DNA Research Institute; Cold Spring Harbor and Washington University in St Louis Sequencing Consortium; European Union Arabidopsis Genome Sequencing Consortium (2000) Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. Nature 408:823-826.

Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, Numa H, Sakai H, Wu J, Itoh T, Sasaki T, Aono R, Fujii Y, Habara T, Harada E, Kanno M, Kawahara Y, Kawashima H, Kubooka H, Matsuya A, Nakaoka H, Saichi N, Sanbonmatsu R, Sato Y, Shinso Y, Suzuki M, Takeda J, Tanino M, Todokoro F, Yamaguchi K, Yamamoto N, Yamasaki C, Imanishi T, Okido T, Tada M, Ikeo K, Tateno Y, Gojobori T, Lin YC, Wei FJ, Hsing YI, Zhao Q, Han B, Kramer MR, McCombie RW, Lonsdale D, O'Donovan CC, Whitfield EJ, Apweiler R, Koyanagi KO, Khurana JP, Raghuvanshi S, Singh NK, Tyagi AK, Haberer G, Fujisawa M, Hosokawa S, Ito Y, Ikawa H, Shibata M, Yamamoto M, Bruskiewich RM, Hoen DR, Bureau TE, Namiki N, Ohyanagi H, Sakai Y, Nobushima S, Sakata K, Barrero RA, Sato Y, Souvorov A, Smith-White B, Tatusova T, An S, An G, OOta S, Fuks G, Fuks G, Messing J, Christie KR, Lieberherr D, Kim H, Zuccolo A, Wing RA, Nobuta K, Green PJ, Lu C, Meyers BC, Chaparro C, Piegu B, Panaud O, Echeverria M (2008) The rice annotation project database (RAP-DB): 2008 update. Nucleic Acids Res 36:D1028-33.

Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, Tong W, Wang J, Wang J, Yu J, Yang H, Zhu L (2004) A comparison of rice chloroplast genomes. Plant Physiol 135:412-420.

Theologis A, Ecker JR, Palm CJ, Federspiel NA, Kaul S, White O, Alonso J, Altafi H, Araujo R, Bowman CL, Brooks SY, Buehler E, Chan A, Chao Q, Chen H, Cheuk RF, Chin CW, Chung MK, Conn L, Conway AB, Conway AR, Creasy TH, Dewar K, Dunn P, Etgu P, Feldblyum TV, Feng J, Fong B, Fujii CY, Gill JE, Goldsmith AD, Haas B, Hansen NF, Hughes B, Huizar L, Hunter JL, Jenkins J, Johnson-Hopson C, Khan S, Khaykin E, Kim CJ, Koo HL, Kremenetskaia I, Kurtz DB, Kwan A, Lam B, Langin-Hooper S, Lee A, Lee JM, Lenz CA, Li JH, Li Y, Lin X, Liu SX, Liu ZA, Luros JS, Maiti R, Marziali A, Militscher J, Miranda M, Nguyen M, Nierman WC, Osborne BI, Pai G, Peterson J, Pham PK, Rizzo M, Rooney T, Rowley D, Sakano H, Salzberg SL, Schwartz JR, Shinn P, Southwick AM, Sun H, Tallon LJ, Tambunga G, Toriumi MJ, Town CD, Utterback T, Van Aken S, Vaysberg M, Vysotskaia VS, Walker M, Wu D, Yu G, Fraser CM, Venter JC, Davis RW (2000) Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. Nature 408:816-820.

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123-135.

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet 15:57-61.

Wang D, Rousseau-Gueutin M, and Timmis JN (2012) Plastid sequences contribute to some plant mitochondrial genes. Mol Biol Evol 29:1707–1711.

Wang D, Timmis JN (2013) Cytoplasmic organelle DNA preferentially inserts into open chromatin. Genome Biol Evol 5:1060–1064.
Wheeler TJ, Kececioglu JD (2007) Multiple alignments by aligning alignments. Bioinformatics 23:i559-i568.
Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252-3255.
Xu Q, Xiong G, Li P, He F, Huang Y, Wang K, Li Z, Hua J (2012) Analysis of complete nucleotide sequences of 12 *Gossypium* chloroplast genomes: origin and evolution of allotetraploids. PLoS ONE 7:e37128.
Xue J, Wang S, Zhou S-L (2012) Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). Am J Bot 99:e240–e244.
Yang Y-B, Yang S-X, Li H-T, Yang J, Li D-Z (2013) Comparative chloroplast genomes of *Camellia* species. PLoS ONE 8:e73053.
Yoshida T, Furihata HY, Kawabe A (2013) Patterns of genomic integration of nuclear chloroplast DNA gragments in plant species. DNA Res 21:127-140.
Yuan Q, Hill J, Hsiao J, Moffat K, Ouyang S, Cheng Z, Jiang J, Buell C (2002) Genome sequencing of a 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. Mol Genet Genomics 267:713-720.

Tables and Figures

Table 2.1 Comparison of genome compartment lengths (bp) in *Nelumbo* plastome accessions.

	N. nucifera			*N. lutea*	
	NC_025339	NC_015610	JQ336993	NC_015605	JQ336992
Total length	163,330	163,307	163,639	163,206	163,510
LSC	91,910	91,847	91,889	91,759	91,798
IR	26,031	26,065	26,065	26,054	26,052
SSC	19,358	19,330	19,620	19,339	19,608

Table 2.2 Number, length, and type of simple sequence repeats in the *N. nucifera* 'China Antique' mitochondrial genome. If length of a repeat motif is inapplicable, the cell was left empty.

	Simple sequence repeat type											
	Mononucleotide		Dinucleotide					Trinucleotide				
Length (bp)	A/T	C/G	AT/TA	AC/CA	AG/GA	GT/TG	TC/CT	AGG	ATA	ATT	CTC	TAT
10	22	4	4	1	9	2	1					
11	4	8										
12	6	4	4	1	0	0	2					
13	6	2										
14	4	0	0	0	1	0	1					
15	0	0						1	1	1	1	1
16	0	0	0	0	0	0	1					
17	2	0										
18	0	0	1	0	0	0	0	0	0	0	0	0
Total	44	18	9	2	10	2	5	1	1	1	1	1

Table 2.3 Comparison of short dispersed direct (D) and inverted (I) repeats among eight angiosperm mitochondrial genomes.

	'China Antique'		*Arabidopsis*		*Carica*		*Glycine*		*Vitis*		*Oryza*		*Sorghum*		*Zea*	
Length (bp)	D	I	D	I	D	I	D	I	D	I	D	I	D	I	D	I
30-50	3375	3287	249	157	173	195	108	85	846	808	157	168	170	114	301	200
51-70	365	332	29	30	18	21	13	15	42	61	58	74	31	31	29	21
71-90	158	150	12	15	10	3	11	12	41	22	5	19	10	3	6	8
91-110	70	54	5	10	1	2	8	9	12	18	2	6	7	6	2	4
111-200	53	52	6	15	2	10	20	26	11	12	6	7	11	1	8	5
201-300	12	4	1	4	0	2	4	3	7	3	1	2	2	2	2	1
301-999	5	2	6	4	0	1	1	4	1	4	7	5	1	0	6	0
1000-5000	0	0	1	0	0	2	0	7	0	0	8	0	1	2	0	0
5001-10,000	0	0	0	1	0	3	0	2	0	0	1	0	0	0	1	0
10,001-20,000	0	0	0	0	1	0	0	0	0	0	2	0	1	0	2	1
20,001-40,000	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
>40,000	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Table 2.4 Percentage (number of bp), hypothesized directionality, and type of intracellular gene transfer. Percentage of genome is calculated by dividing the total amount of NORG or plastid DNA by the total number of bp in a genome.

	NUPT	NUMT	MTPT
'China Antique'	0.005 (35,836)	0.004 (29,163)	1.60 (7287)
Arabidopsis	0.025 (29,441)	0.411 (490,157)	1.35 (4958)
Carica	0.087 (236,657)	0.116 (315,447)	4.68 (22,324)
Glycine	0.052 (492,127)	0.048 (461,158)	1.00 (4041)
Oryza	0.276 (1,055,767)	0.252 (962,986)	7.07 (34,673)
Sorghum	0.048 (335,216)	0.038 (264,436)	6.08 (28,506)
Vitis	0.088 (414,379)	0.145 (679,983)	8.14 (62,953)
Zea	0.065 (1,340,545)	0.109 (2,242,570)	4.32 (24,565)

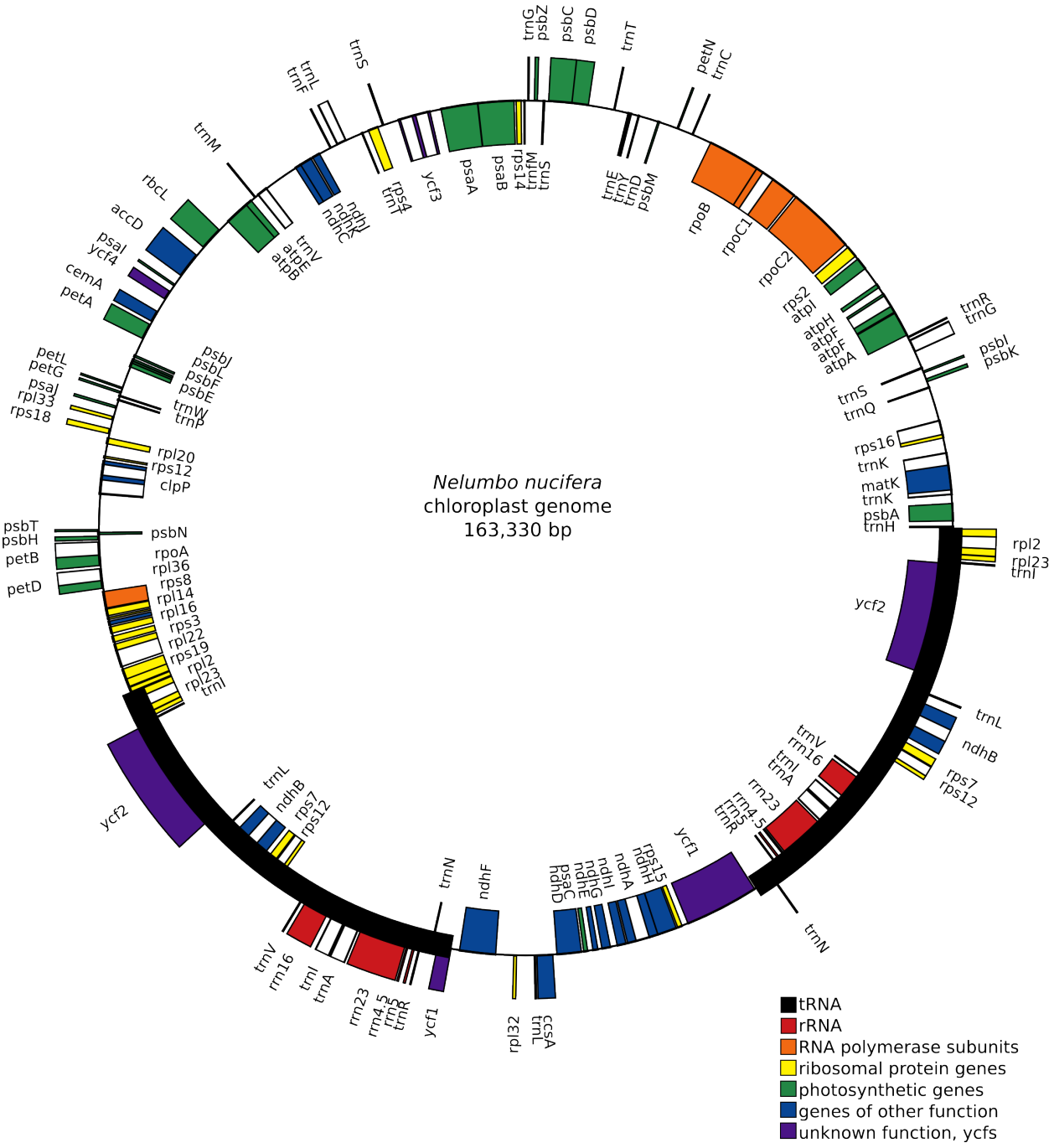


Fig. 2.1 *N. nucifera* 'China Antique' plastome gene map. Blocks show location, adjacencies, gene type (provided in the legend), strandedness (genes on the inside of the circle are transcribed clockwise and outside the circle are counter-clockwise), and the presence of introns.

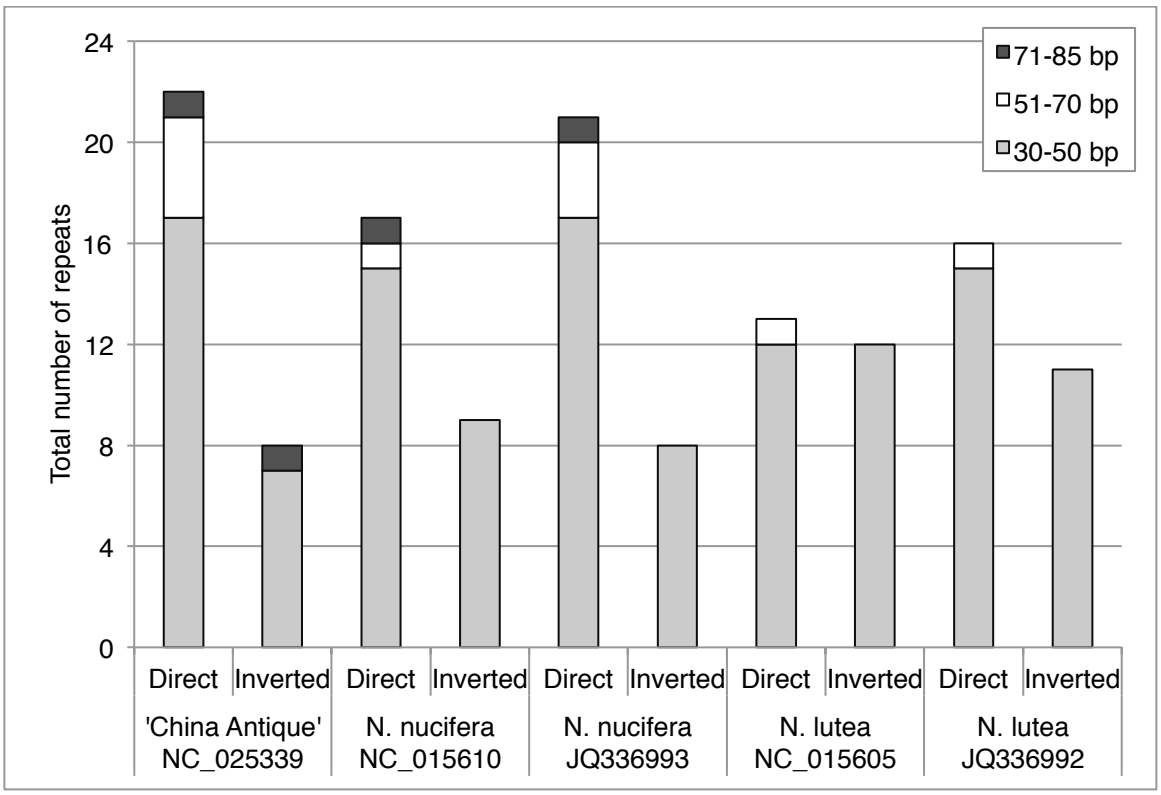


Fig. 2.2 Comparison of total number and lengths of direct and inverted repeats in the plastomes of ‘China Antique’ and four previously published *Nelumbo* accessions (Xue et al. 2012). Within ‘China Antique’ one additional repeat is shared when the genome is compared as a whole versus when each compartment is compared individually.

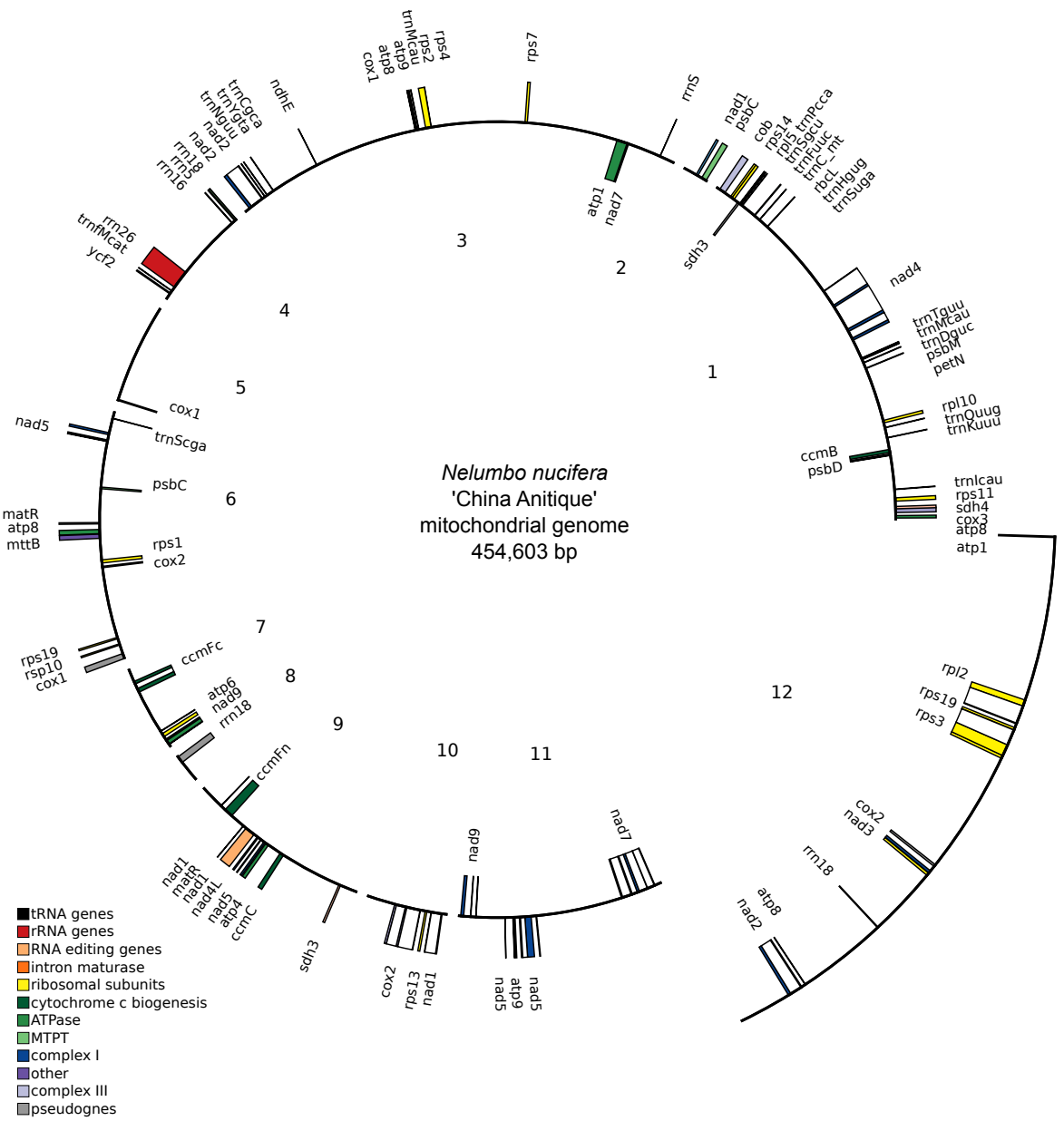
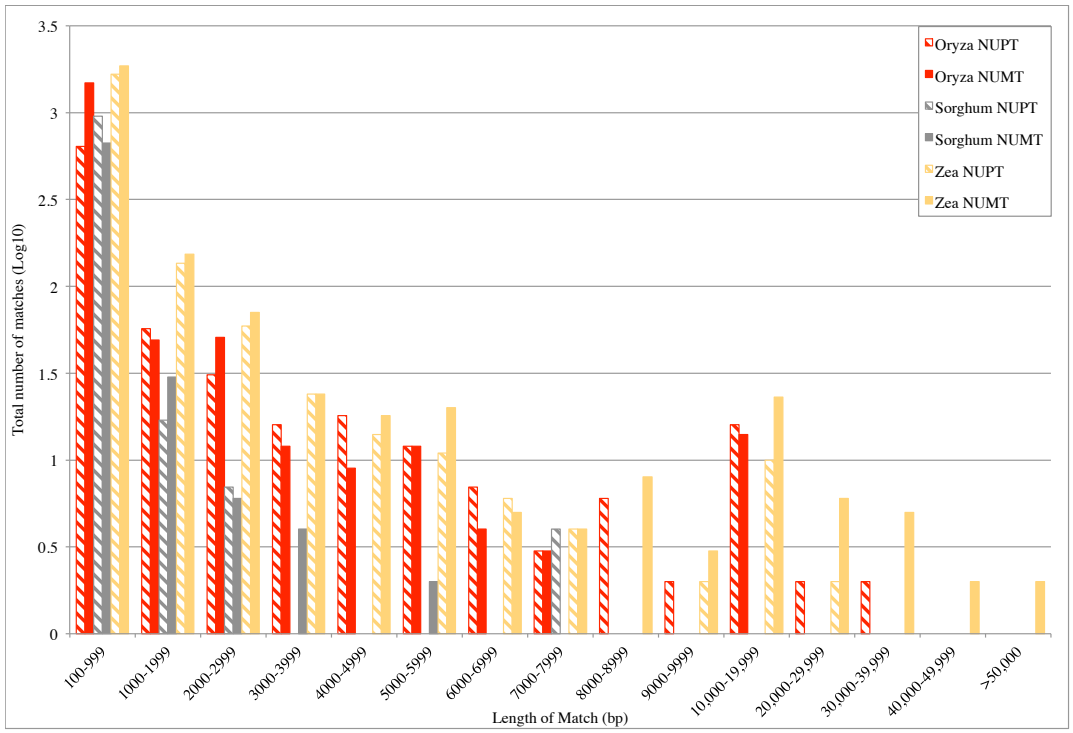


Fig. 2.3 *N. nucifera* 'China Antique' draft mitochondrial genome. Contigs are oriented in the way that they are hypothesized to assemble based on evidence from PCR and scaffolding. Contig 12 is shown as separate because there is no supported assembly for this contig. Blocks show gene locations, adjacencies, type of gene, strandedness (genes on the inside of the circle are transcribed clockwise and outside the circle are counter clockwise), and the presence of introns.

a



b

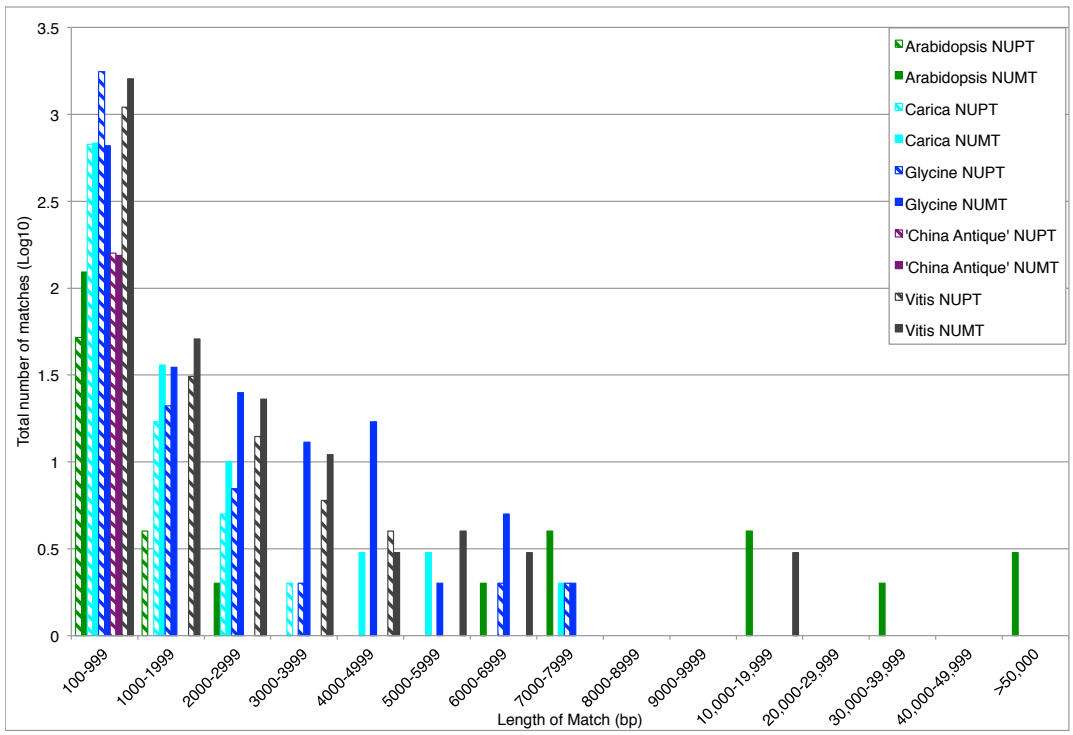


Fig. 2.4 Total number and sizes of nuclear organellar DNA (NORGs) in: (**a**) monocot genomes, and (**b**) eudicot genomes. The total number of NORGs is shown as Log_{10} values due to the large variation in total number of NORGs found.

Supplementary Tables and Figures

Table S2.1 Comparison of simple sequence repeats among *Nelumbo* plastomes. If length of a repeat motif is inapplicable, the cell was left empty.

N. nucifera 'China Antique' accession NC_025339

	Simple sequence repeat type					
	Mononucleotide		Dinucleotide			Trinucleotide
Length (bp)	A/T	C/G	AT/TA	GT/TG	TC	TTA
10	15	2	6	2	2	
11	3	0				
12	7	1	2	0	0	
13	3	0				
14	1	0	0	0	0	
15	1	0				1
Total	30	3	8	2	2	1

N. nucifera accession JQ336993

	Simple sequence repeat type					
	Mononucleotide		Dinucleotide			Trinucleotide
Length (bp)	A/T	C/G	AT/TA	GT/TG	TC	TTA
10	14	1	6	2	2	
11	4	1				
12	7	1	2	0	0	
13	4	0				
14	0	0	0	0	0	
15	1	0				1
Total	30	3	8	2	2	1

N. nucifera accession NC_015610

	Simple sequence repeat type					
	Mononucleotide		Dinucleotide			Trinucleotide
Length (bp)	A/T	C/G	AT/TA	GT/TG	TC	TTA
10	18	2	6	2	2	
11	3	0				
12	5	1	2	0	0	
13	2	0				
14	2	0	0	0	0	
15	1	0				1
Total	31	3	8	2	2	1

N. lutea accession JQ336992

	Simple sequence repeat type						
	Mononucleotide		Dinucleotide			Trinucleotide	
Length (bp)	A/T	C/G	AT/TA	GT/TG	TC	TTA	ATA
10	23	0	4	2	2		
11	6	0					
12	3	0	3	0	0		
13	4	0					
14	2	1	0	0	0		
15	1	0				1	1
16	0	0	0	0	0		
17	0	0					

Table S.2.1 (cont.)

18	0	0	0	0	0	0	0
19	0	0					
20	0	0	0	0	0		
21	1	0				0	0
Total	40	1	7	2	2	1	1

N. lutea accession NC_015605

	Simple sequence repeat type						
	Mononucleotide		Dinucleotide			Trinucleotide	
Length (bp)	A/T	C/G	AT/TA	GT/TG	TC	TTA	ATA
10	25	0	4	2	2		
11	4	1					
12	4	0	3	0	0		
13	3	0					
14	3	0	0	0	0		
15	0	0				1	1
16	0	0	0	0	0		
17	1	0					
Total	40	1	7	2	2	1	1

Table S2.2 Comparison of nucleotide changes in plastome compartments among the five *Nelumbo* accessions. The diagonals show conserved nucleotides among the plastomes. Numbers not on the diagonal show the amount, type, and direction of change of point mutations with reference to 'China Antique' (NC_025339).

	Large single copy				Inverted repeat				Small single copy			
	States In Compared Taxa				States In Compared Taxa				States In Compared Taxa			
NC_025339	A	C	G	T	A	C	G	T	A	C	G	T
A	142,902	88	61	152	37,064	10	0	0	32,805	22	28	36
C	72	84,866	9	76	3	27,241	0	7	19	16,365	4	12
G	67	10	80,965	63	4	0	28,976	0	20	0	14,954	22
T	184	62	102	148,394	2	8	1	36,845	22	17	21	32,723

Table S2.3 Comparison of gene content in the draft *Nelumbo* mitochondrial genome and seven other angiosperm genomes used in introgression comparisons. Presence of a gene is indicated by the '+' symbol, absence of a gene is indicated by a '-' symbol, and pseudogenes (duplications with internal stops or fragments) are indicated with 'Ψ'. Superscript numbers denote the number of exons for each gene. Subscript numbers denote the number of duplications, if applicable. Due to the draft status of *N. nucifera*, if a gene was not present its absence was not inferred and the cell was left empty.

Gene	'China Antique'	*Arabidopsis* NC_001284	*Carica* NC_012116	*Glycine* NC_020455	*Oryza* NC_011033	*Sorghum* NC_008360	*Vitis* NC_012119	*Zea* NC_007982
atp1	$+^{2}$, Ψ	+	+, Ψ	$+_{3}$	+	$+_{2}$	$+_{2}$	$+_{2}$
atp4	+	+	+	+	+	+	+	+
atp6	+	$+_{2}$	+	$+_{2}$	+	+	+	+
atp8	$+_{2}$, Ψ_{2}	Ψ	+	+	+	+	+	+
atp9	$+_{2}$, Ψ	+	+, Ψ	+	+	+	$+_{2}$	+
ccmB	+	+	+	+	+	+	+	+
ccmC	+	+	+	+	+	$+_{2}$	+	+
ccmFc	$+^{2}$	$+^{2}$	$+^{2}$	$+^{2}$	$+^{2}$	$+^{2}$	$+^{2}$, Ψ	$+^{2}$
ccmFn	+	$+^{2}$	$+_{2}$	+	+	+	+	+
cob	+	+	+	+	+	$+_{2}$	+	+
cox1	$+^{2}$, Ψ_{3}	+	+, Ψ	+	+	+	+	+
cox2	$+^{3}$, Ψ_{2}	$+^{2}$	$+^{2,}$ Ψ	+	+	$+^{2}$	+	$+^{2}$
cox3	+	+	+	+	+	+	+	+
matR	+, Ψ	+	+, Ψ	+	Ψ	+	+	+
mttB (orfX, tatC, ymf16)	+	Ψ	+	+	+	+	Ψ	+
nad1	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	+	$+^{5}$	$+^{5}$
nad2	$+^{4}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$
nad3	+	+	+	+	+	+	+	+
nad4	$+^{4}$	$+^{4}$	$+^{4}$	$+^{4}$	$+^{4}$	$+^{4}$	$+^{4}$	$+^{4}$
nad4L	+	+	+	$+_{2}$	+	+	+	+
nad5	$+^{5,}$ Ψ	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$
nad6		+	+	+	+	+	$+_{2}$	+
nad7	$+^{5}$, Ψ	$+^{5}$	$+^{5}$	$+^{5}$, Ψ_{2}	$+^{5}$	$+^{5}$	$+^{5}$	$+^{5}$
nad9	+, Ψ	+	+	+	+	+	+	+
rpl2	$+^{2}$	$+^{2}$	$+^{2}$	-	$+^{2}$	Ψ	Ψ	+
rpl5	+	+	+	+	+	-	+	-
rpl10	+	-	+	-	-	-	Ψ	-
rpl16	+	+	+	+	Ψ	+	+	+
rps1	+	-	+	+	+	+	+	+
rps2	+	-	$+^{2}$	-	+	+	Ψ	+
rps3	$+^{2}$	$+^{2}$	-	$+^{2}$	$+^{2}$	$+^{2}$	$+^{2}$, Ψ_{2}	$+^{2}{}_{2}$
rps4	+	+	+	+	+	+	+	+
rps7	+	+	+	-	+	$+_{2}$	+	+
rps10	+	-	$+^{2}$	$+^{2}$	-	-	$+^{2}$	-
rps11	+	-	-	-	Ψ	-	-	-
rps12	+	+	+	+	+	+	+	+
rps13	+	-	+	-	+	+	+	+
rps14	+	Ψ	+	+	Ψ	-	+	-

Table S2.3 (cont.)

rps19	+$_2$	Ψ	+	-	+	-	+$_2$	-
sdh3	+, Ψ	-	+	-	-	-	+, Ψ	-
sdh4	+	Ψ	Ψ	-	Ψ	Ψ	+	Ψ
Total number of protein-coding genes	*43*	*30*	*38*	*36*	*33*	*36*	*40*	*35*
Ala-cp		-	-	-	-	Ψ	-	Ψ$^2_2$
Arg		-	-	-	-	-	-	Ψ
Arg-cp		-	-	-	+	-	+	Ψ$_2$
Asn	+	-	-	+	-	-	-	-
Asn-cp		+	+	+	+	+	+$_2$	+$_2$
Asp	+	-	-	+	+	+	+	+, Ψ
Asp-cp	+	+	+	+	Ψ	-	+	+
Cys-bacterial		-	-	+	-	-	-	-
Cys-cp	+	-	-	-	+	+	-	+
Cys-mt	+	+	+	+	+	+	+	+
Gln	+	+	+	+	-	+	+	-
Gln-cp		+	-	+	+	-	+	+
Glu	+	+	+	+	+$_2$	+	+	+$_2$
Gly	+	+	+	+	-	-	+	-
Gly-cp		-	-	-	-	-	Ψ	-
His-cp	+	+	+, Ψ	+	+	+	+	+
Ile	+	+	+	+	+	+	+	+$_2$
Ile-cp	+	+	+, Ψ	-	Ψ	Ψ	Ψ	+$^2_2$
Leu	+	-	-	-	+	Ψ	-	+, Ψ
Leu-cp		-	+	-	Ψ	Ψ	+	+$_2$
Lys	+	+$_2$	+	+	+	+$_2$	+	+
Lys-cp		-	-	-	-	-	Ψ	+
Met	+	-	+, Ψ	-	-	-	-	+$_2$
Met-cp	Ψ	+$_2$	+	+	+$_2$	+	+	+
fMet	+$_2$	+	+	+$_4$	+	+	+$_2$	+
Phe	+	+$_2$	+	+		-	+	-
Phe-cp		Ψ$_2$	Ψ	-	+	+	-	+
Pro		-	+	+	+$_2$	+	+	+
Pro-cp		+	+$_2$	+	+$_2$	+	+$_2$	+, Ψ
Ser	+$_3$	+$_2$	+$_2$	+	+	+$_4$	+$_2$	+
Ser-cp		+$_3$	+	-	+	+	-	-
Thr-cp	+	Ψ$_2$	Ψ	Ψ	-	Ψ	+	Ψ
Trp-cp		+	+, Ψ	+	+$_2$, Ψ	+	+$_2$	-
Tyr	+	+$_3$	+	+	+	+	+$_2$	+
Tyr-cp		-	-	-	-	-	+	-
Val-cp		-	-	-	Ψ	Ψ	-	+
rrn5		+	+	+	+	+	+	+
rrn18 (rrnS)	+^{4}	+	+, Ψ$_2$	+	+	+	+	+
rrn26 (rrnL)	+^{5}	+	+, Ψ$_2$	+	+	+	+, Ψ	+
Total number of tRNAs	*22*	*27*	*23*	*24*	*25*	*22*	*29*	*29*
Total number of rRNAs	*2*	*3*	*3*	*3*	*3*	*3*	*3*	*3*
Total number of pseudogenes	*17*	*9*	*16*	*3*	*10*	*8*	*12*	*10*
accD-cp		-	-	-	-	-	Ψ	-

Table S2.3 (cont.)

atpA-cp		-	-	-	-	-	Ψ	-
atpB-cp		-	-	-	Ψ	Ψ	Ψ	-
atpE-cp		-	-	-	Ψ	Ψ	Ψ	-
atpF-cp		-	-	-	-	-	Ψ	-
atpH-cp		-	-	-	Ψ	Ψ	-	-
atpI-cp		-	-	Ψ	-	-	Ψ	-
ccsA-cp		-	-	-	-	-	Ψ	-
clpP-cp		-	-	-	-	-	Ψ	-
cemA-cp		-	-	-	-	-	Ψ	-
infA-cp		-	-	-	-	Ψ	Ψ	-
matK-cp		-	-	-	-	-	Ψ	-
ndhA-cp		Ψ	Ψ	Ψ	-	-	Ψ	-
ndhB-cp		-	Ψ_2	-	-	-	Ψ	Ψ^2
ndhC-cp		-	-	-	-	Ψ	-	-
ndhD-cp	Ψ	-	-	-	-	-	Ψ_2	-
ndhE-cp	Ψ	-	-	-	-	-	Ψ_2	Ψ_2
ndhF-cp		-	-	-	-	-	Ψ	-
ndhH-cp		-	-	-	-	-	Ψ	-
ndhI-cp		-	-	-	-	Ψ	-	-
ndhJ-cp		-	-	-	Ψ	-	-	-
ndhK-cp		-	-	-	Ψ	Ψ_2	-	Ψ
petA-cp		-	-	-	-	-	Ψ	-
petB-cp		-	Ψ	-	-	Ψ	Ψ_2	-
petD-cp		-	-	-	-	Ψ	Ψ_2	-
petG-cp		-	Ψ	-	-	-	Ψ_2	-
petL-cp		-	-	-	-	-	Ψ_2	-
petN-cp	Ψ	-	-	-	-	-	-	-
psaA-cp		Ψ	-	-	-	-	Ψ	-
psaB-cp		-	-	-	-	-	Ψ	Ψ_2
psaC-cp		-	-	-	-	-	Ψ	-
psaJ-cp		-	-	-	-	-	Ψ	-
psbA-cp		Ψ	-	-	-	-	Ψ	Ψ
psbB-cp		-	-	-	-	Ψ	Ψ_2	-
psbC-cp	Ψ_3	-	-	-	-	-	Ψ	-
psbD-cp	Ψ	-	-	-	-	-	Ψ	-
psbE-cp		-	-	-	-	-	Ψ	-
psbF-cp		-	-	-	-	-	Ψ	-
psbH-cp		-	-	Ψ_5	-	-	Ψ	-
psbJ-cp		-	-	-	-	-	Ψ	-
psbL-cp		-	-	-	-	-	Ψ	-
psbM-cp	Ψ	-	-	-	-	-	Ψ	-
psbN-cp		-	-	-	-	-	Ψ	-
psbT-cp		-	-	-	-	-	Ψ	Ψ
rbcL-cp	Ψ	Ψ	-	Ψ	Ψ	Ψ	Ψ	Ψ
rpoA-cp		-	-	-	-	Ψ	Ψ	-
rpoB-cp		Ψ	-	-	Ψ_2	Ψ	-	-
rpoC1-cp		-	Ψ	-	Ψ	Ψ_2	-	Ψ
rpoC2-cp		-	-	-	Ψ	Ψ_2	Ψ	-
rpl2-cp		-	-	-	Ψ	Ψ	Ψ_2	Ψ
rpl14-cp		-	-	-	Ψ	Ψ_2	Ψ	-
rpl16-cp		-	-	-	-	Ψ	Ψ	-
rpl22-cp		-	Ψ_2	-	-	Ψ_2	Ψ	Ψ

Table S2.3 (cont.)

rpl23-cp		-	Ψ	-	Ψ	Ψ	-	Ψ
rpl32-cp		-	-	-	-	-	Ψ	-
rpl33-cp		-	-	-	-	-	Ψ	-
rpl36-cp		-	-	-	-	Ψ	Ψ	-
rps3-cp		-	Ψ	-	-	-	Ψ	-
rps4-cp			Ψ	-	Ψ	-	-	-
rps7-cp		-	Ψ	-	-	-	-	Ψ
rps8-cp		-	-	-	Ψ	Ψ	Ψ	-
rps11-cp		-	-	-	-	Ψ	Ψ_2	-
rps12-cp		Ψ	Ψ	Ψ	Ψ	Ψ	Ψ_2	Ψ_3
rps15-cp		-	-	-	-	-	Ψ	-
rps16-cp		-	-	-	-	-	Ψ	-
rps18-cp		-	-	-	-	-	Ψ	-
rps19-cp		-	Ψ	-	Ψ	Ψ	Ψ	Ψ
ycf1-cp		Ψ	-	-	-	-	Ψ	-
ycf2-cp	Ψ	-	Ψ	Ψ	-	Ψ_2	-	Ψ_3
ycf3-cp		-	-	-	-	-	Ψ	-
ycf4-cp		-	-	-	-	-	Ψ	-
4.5S-cp		-	-	-	-	-	-	Ψ
5S-cp	Ψ	-	-	-	-	-	-	Ψ
16S-cp	Ψ	Ψ_2	Ψ_2	Ψ	Ψ	Ψ_2	Ψ_3	Ψ_3
23S-cp		Ψ_2	-	Ψ	-	Ψ_2	-	Ψ_4
Total number of cp derived gene fragments	10	7	15	10	17	32	69	21

Table S2.4 Percentage of mitochondrial DNA shared in pairwise comparisons[a]. Higher values (bold) indicate more DNA in common, while lower values (red) indicate less shared DNA between the two mitochondrial genomes being compared. Percentages are derived from the number of shared base pairs divided by the total number of base pairs within the subject genome.

Taxon	'China Antique'	*Arabidopsis*	*Beta*	*Carica*	*Glycine*	*Oryza*	*Sorghum*	*Vitis*	*Zea*
'China Antique'		23.69	25.30	26.39	24.77	22.77	19.28	18.13	15.24
Arabidopsis	19.12		22.36	20.72	20.95	18.90	16.29	12.52	12.94
Beta	20.53	22.47		23.42	23.65	20.44	16.82	13.84	14.11
Carica	27.69	**26.93**	**30.28**		**28.62**	21.82	19.31	**20.04**	16.15
Glycine	21.93	22.99	25.81	24.16		20.76	17.95	14.68	14.69
Oryza	24.57	25.27	27.19	22.45	25.29		17.86	14.97	40.71
Sorghum	19.87	20.80	21.38	18.97	20.89	17.06		12.64	**46.23**
Vitis	**30.84**	**26.39**	29.01	**32.49**	**28.20**	23.60	20.85		15.63
Zea	19.09	20.09	21.79	19.29	20.79	**47.28**	**56.20**	11.51	

[a] Percent identity between mitochondrial genomes calculated without masking repetitive DNA.

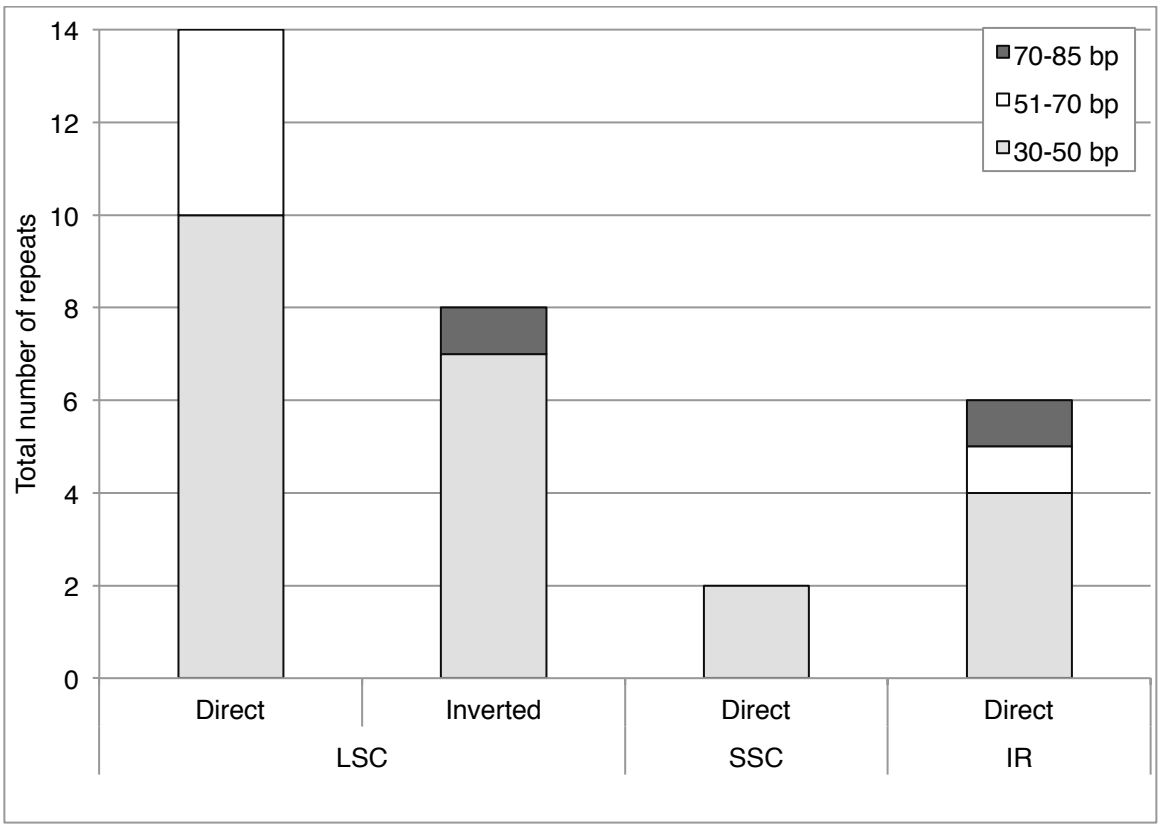


Fig. S2.1 Number and length of short dispersed repeats (SDRs) within each of the three plastome compartments of *N. nucifera* 'China Antique'. The LSC has the most SDRs and the only instances of inverted repeats.

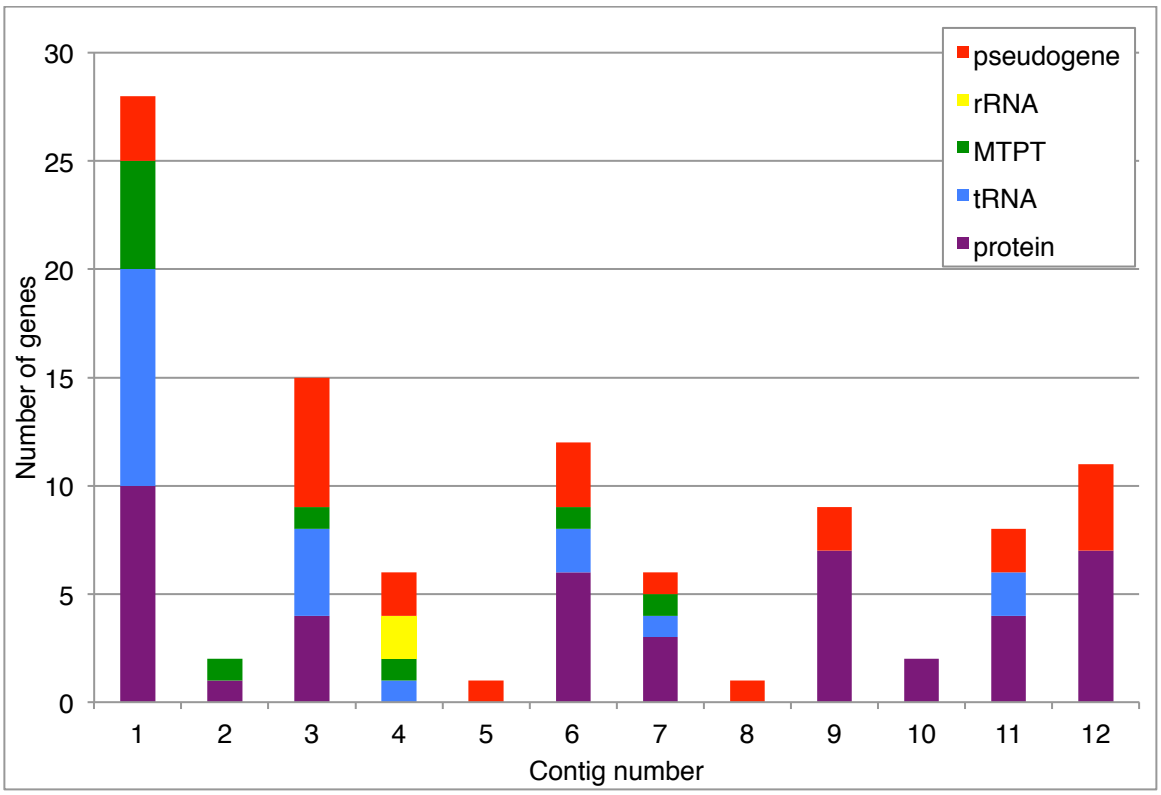


Fig. S2.2 Number and type of genes and pseudogenes predicted within *N. nucifera* 'China Antique' mitochondrial contigs.

CHAPTER 3: THE PLASTOMES OF *ANETHUM GRAVEOLENS*, *FOENICULUM VULGARE*, *CARUM CARVI*, AND *CORIANDRUM SATIVUM* (APIACEAE): CHARACTERIZATION OF INVERTED REPEAT CHANGES

Abstract

Land plant plastomes can be divided into three regions, two of which are single copy and the third a large inverted duplication known as the inverted repeat (IR). The boundary between the two single copy regions and the IR can vary by small amounts in closely related species and in some groups the variation in gene content within the IR is large. However, these larger fluctuations in gene content are rare and are only seen in a subset of eudicot families, such as the apioid superclade of Apiaceae subfamily Apioideae. The apioid superclade comprises 12 tribes and other major clades and exhibits much variation in IR size. These sizes range from an expansion of a few thousand nucleotides to a contraction of over 16 kb. The mechanism(s) and timing of changes in IR size are unknown. Through sequencing of complete plastomes from *Anethum graveolens*, *Foeniculum vulgare*, *Carum carvi*, and *Coriandrum sativum*, and through sequencing the large single copy (LSC)–IR boundary in 34 additional species, I show that there are several mechanisms at work creating the dynamic IR changes seen. In *Coriandrum* (tribe Coriandreae), the IR was likely shortened as a result of double-strand break repair, supporting a mechanism previously suggested. In addition, *Coriandrum* has many repeats that may have contributed to additional changes near its IR boundaries. Short dispersed repeats are also implicated as a mechanism of IR change in the 34 additional species investigated. In *Carum* (tribe Careae) there is an IR boundary expansion, in addition to two small inversions. One of these inversions is near J_{LA} and the other is between *psbM* and *trnT*. *Anethum* and *Foeniculum* (both tribe Apieae) do not have extreme IR boundary changes, elevated levels of repeat DNA, or inversions. Instead, these two plastomes contain unique DNA in the LSC region adjacent to J_{LA} having high sequence similarity to mitochondrial non-coding DNA. A transfer of cpDNA from the

S10 operon into the mitochondrial genome may have donated a template for homologous recombination near J_{LA}, leading to an insertion of non-coding mtDNA within these plastomes. This insertion may have also caused the small IR contraction seen in all examined members of tribe Apieae. These results shed new light on IR boundary changes and describe a potential new instance of angiosperm intracellular gene transfer from the mitochondrial genome to the plastome. For the 34 additional species investigated our data support double-strand break repair as a mechanism of plastid evolution and is the likely cause of novel DNA insertions at J_{LA}.

Introduction

The majority of angiosperm plastid genomes (plastomes) are highly conserved in structure and gene content. These plastomes share the same basic organization, with a large inverted repeat (IR) separating the remainder of the molecule into large single copy (LSC) and small single copy (SSC) regions. Belying this structural conservatism, the boundaries between the LSC and IR regions may be quite dynamic, resulting in gene adjacency changes. In a typical angiosperm plastome, the LSC–IR boundaries occur within or near *rps19* of the S10 operon. This boundary has been termed J_{LB} (Sugiura et al. 1986). At the other end of the LSC region, the interrupted *rps19* gene located at the terminus of the IR is adjacent to genes *trnH* and *psbA*; this boundary has been termed J_{LA}. Small changes in LSC–IR boundary positions of less than 100 bp are frequent during angiosperm evolution (Goulding et al. 1996) whereas extreme contractions without a complete loss of the IR are rare (Palmer et al. 1987; Hansen et al. 2007; Guisinger et al. 2011). The plastomes of Apiaceae subfamily Apioideae are unusual among angiosperms in that they exhibit increased variation in the position of J_{LB} (Palmer 1985; Plunkett and Downie 1999, 2000).

Mechanisms proposed to explain IR structural changes include gene conversion and double-strand break repair (DSBR; Goulding et al. 1996; Odom et al. 2008) and recombination facilitated by repetitive DNA (Palmer 1985; Palmer et al. 1987; Aii et al. 1997; Lee et al. 2007; Cai et al. 2008). Recombination across repeats is the most reported cause of plastome structural changes (Ogihara et al. 1988; Goulding et al. 1996; Hansen et al. 2007; Lee et al. 2007; Catalano et al. 2009; Guo et al. 2014). Gathering data to test these hypotheses can be problematic because evidence suggesting any one mechanism of IR structural change can be masked by additional mutations.

Within plastomes, gene adjacencies can change through mechanisms other than expansion and contraction of the IR. Some plastid genes have been relocated to the

mitochondrial and/or nuclear genomes, and such intracellular gene transfers include *tufA*, *rbcS*, *accD*, *rpl22*, and the *ndh* gene family (Palmer 1991; Martin et al. 1998; Millen et al. 2001; Cummings et al. 2003; Richardson and Plamer 2007). The loss of genes from the plastome is an ongoing process (Martin and Herrmann 1998), with recent transfers resulting in pseudogenes (Kleine et al. 2009). Intracellular transfers of DNA from the plastid into the mitochondrion or nucleus are well documented; however, until recently, the chloroplast was believed to be exempt from acquiring foreign DNA (Rice and Palmer 2006). While the transfer of DNA from the nuclear genome into the plastome of land plants has not been reported, there is a growing body of evidence that mitochondrial DNA (mtDNA) has made its way into the Apiaceae plastome (Goremykin et al. 2009; Iorizzo et al. 2012; Downie and Jansen 2015). In this chapter, I further study the possible mtDNA transfer into the Apiaceae plastome.

Coriandrum sativum (coriander; Apiaceae subfamily Apioideae) has a greatly reduced IR (Palmer 1985), yet the mechanism explaining this contraction is unclear. Plunkett and Downie (2000) used restriction site mapping to investigate the extent of IR change in Apiaceae and allied families by assessing variation in the position of J_{LB}. Of the 113 species they surveyed, nine different J_{LB} boundaries were detected. Such boundary shifts, without further rearrangements elsewhere in the plastome, are highly unusual among angiosperms. In addition to the typical J_{LB} boundary within or near *rps19*, as occurring in *Nicotiana tabacum* (tobacco) and other species having the ancestral angiosperm plastome structural organization (Raubeson et al. 2007), they identified one expansion and seven different contractions, ranging in size from 1 to 16 kb. *Coriandrum* was deemed to have the most contracted IR; however, the overall size of its plastome (~150 kb) was only slightly smaller than that of a typical species, a result of a ~5.7 kb insertion of unknown composition near the terminus of the IR (Plunkett and Downie 2000). All boundary shifts were restricted to the apioid superclade of Apiaceae subfamily

Apioideae, a large group comprising 12 tribes and other major clades of dubious relationship (Plunkett and Downie 1999, 2000).

The goals of this study are to further characterize the J_{LA} boundary and investigate hypotheses of IR change in the apioid superclade of Apiaceae subfamily Apioideae. I focus on J_{LA} because through IR expansion and contraction, there is no gene adjacency change at J_{LB}. To address these goals I have determined the complete plastome sequences of four species: *Foeniculum vulgare* and *Anethum graveolens* (fennel and dill; tribe Apieae); *Coriandrum sativum* (tribe Coriandreae); and *Carum carvi* (caraway; tribe Careae). Through previous restriction site mapping studies, the two species of Apieae are resolved as sister taxa and have a 1.6 kb contraction of J_{LB} relative to its position in tobacco (Plunkett and Downie 1999, 2000). *Carum* and *Coriandrum* represent the extremes of IR change known in Apiaceae, with an expansion of about 1 kb in *Carum* and a contraction of about 16 kb in *Coriandrum*. To further characterize J_{LA} in other members of the apioid superclade, investigate the insertion of putative mtDNA into the Apiaceae plastome, and bolster support for any evidence of mechanism leading to IR boundary changes, I report on sequencing through J_{LA} in 34 additional species.

Methods

Plastid DNA Isolation and Sequencing

Isolation and sequencing of the *Coriandrum*, *Foeniculum*, *Carum*, and *Anethum* plastomes followed the procedures described in Jansen et al. (2005) and summarized by Chumley et al. (2006). Leaf tissue was obtained from seedlings propagated from seeds, and plastid isolations consisted of several individual plants. For *Coriandrum* and *Carum,* total genomic DNA was isolated from these same seedlings; for *Anethum* and *Foeniculum,* total genomic DNA was isolated from plants obtained from a local grocery store (Table 3.1). Total genomic DNA from *Coriandrum*, *Anethum*, and *Foeniculum* was isolated using the CTAB

method (Doyle and Doyle 1987) modified by adding 2% polyvinylpyrrolidone (MW 40,000) and re-suspending in Tris-EDTA buffer. The extractions were cleaned using the Wizard® DNA Clean-up System (Promega, Madison, WI) following their protocol. *Carum* total genomic DNA was isolated using Invitrogen's PureLink Plant Total DNA Purification kit, with no protocol modifications.

Draft genome sequences of *Coriandrum*, *Anethum*, and *Foeniculum* were produced at the Joint Genome Institute (http://www.jgi.doe.gov/sequencing/protocols/prots_production.html). The draft genome of *Carum* was generated using Roche 454 sequencing at the University of Illinois W.M. Keck Center using standard protocols. For all plastomes, PCR was used to improve quality scores (any base pair <Q40 or equivalent) using either total genomic or RCA (rolling circle amplification) product followed by sequencing at the University of Washington or the University of Illinois. Plastomes acquired through shotgun sequencing at JGI were assembled using CONSED (Gordon et al. 1998) and Sequencher v. 4.9 (Gene Codes Corporation). All 454 reads were assembled using gsAssembler (Roche). Genome finishing and IR boundary identification followed the methods outlined in Raubeson et al. (2007).

Characterization of J_{LA} in the Apioid Superclade

Plunkett and Downie (2000) identified nine different J_{LB} boundary positions in the 113 species they surveyed (A-I; Fig. 3.1). These boundary locations represent one expansion (B) and seven different contractions (C-I), including the typical position within *rps19* (A). To further characterize J_{LA} in the apioid superclade, I examined 15 species used in the Plunkett and Downie (2000) survey plus 19 additional species (Table 3.1). Collectively, these 34 species represent at least one species each from 10 of the 12 tribes and other major clades comprising the apioid superclade and all previously recognized J_{LB} boundary positions, with the exceptions of C and G. Genomic DNA for these species was isolated as described in the original

publications or by using the DNeasy Plant Mini kit (Qiagen Inc., Valencia, CA). Primers were designed to amplify and sequence through the LSC–IR boundary at J_{LA} in each species (Table S3.1). The locations of *trnH* and the S10 operon facilitated LSC–IR boundary identification. When IR boundaries were not readily identified, J_{LB} was sequenced for that taxon. Amplicons from J_{LA} and J_{LB} were aligned in CLUSTAL OMEGA (http://www.ebi.ac.uk/Tools/msa/clustalo/) and the point of mismatch was deemed the IR boundary (Raubeson et al. 2007).

To determine the origin of the novel plastid DNA fragments adjacent to J_{LA}, these sequences were queried against NCBI's nucleotide DNA database using BLAST. All BLAST searches resulted in multiple hits to angiosperm mitochondrial DNA (mtDNA) sequences, with the best alignment scores showing sequence similarity to the *cob–atp4* and *nad4L–atp4* intergenic spacer regions. Primers anchored within each pair of mitochondrial genes (Table S3.1; Kubo et al. 2000) were used to PCR amplify and sequence the intervening region in 14 of the 34 examined members of the apioid superclade, with the goal of identifying the novel plastid DNA fragments within the mitochondrial genome. An additional primer was designed for a conserved region within the novel plastid DNA fragments and used with an *atp4* primer to confirm adjacency of these regions within the mitochondrial genome. Genome walking within the cob–*atp4* intergenic spacer region and away from *atp4* was also attempted using the APA Genome Walking kit (Bio S&T Inc., Montreal, Canada).

Mechanisms of IR Change

To determine if repetitive DNA, such as simple sequence repeats (SSRs) and short dispersed repeats (SDRs), was affecting IR boundary shifts, *rps19–rpl2* sequences from *Anethum*, *Carum*, *Coriandrum*, and *Foeniculum* were aligned. *Daucus carota* was also included in these comparisons, since its plastome has ancestral IR boundaries, no gene rearrangements, and was the closest relative to the apioid superclade published at the time of analysis (Ruhlman

et al. 2006). The alignment was scanned by eye to locate repeats. In addition, plastome sequences from eight other angiosperms (Table S3.2), representing species having ancestral IR boundaries, LSC–IR boundaries different from ancestral, or lacking an IR, were analyzed using SSR Extractor (Dolan unpublished) and compared to the four plastomes sequenced herein. SSRs were only counted if they were at least 15 bp long and motifs ranged in size from 1 to 5 bp. The location of these repeats was also reported to assist in determining if they were a potential mechanism of IR change.

Vmatch (http://www.vmatch.de/) was used to locate SDRs and SSR Extractor was used to locate SSRs in all plastomes. SDRs were identified with a minimum length of 30 bp and a Hamming distance of 3. SSRs were located as previously described, except minimum repeat size was 10, 12, or 15 bp for each repeat motif length. The total amount of SDRs in each of the newly sequenced plastomes was compared to published reports for other Apiales and eudicot plastomes, the latter with and without major IR structural changes (Table 3.2). Duplication of DNA, such as tRNA genes, may provide evidence of double-strand break repair (Haberle et al. 2008); thus, the newly sequenced plastomes were also scanned for larger duplications at or near their IR boundaries.

Results

Plastomes

A comparison of the major structural features of the four Apiaceae plastomes and the previously published *Daucus* plastome is presented in Table 3.3. Plastome sizes differed by 8,930 bp, between *Carum* and *Coriandrum*. *Coriandrum* had the largest LSC region and the smallest IR. *Foeniculum* had the smallest SSC region. The number of single copy genes was the same across all plastomes. Differences in gene content were due to the number of genes contained within the expanded or contracted IR.

The total number of SSRs varied from 60 (*Carum*) to 66 (*Coriandrum*) with *Daucus* having 59 (Table 3.4). The majority of SSRs are mono- or dinucleotide repeat motifs. The total number of SDRs ranged from 19 (*Carum*) to 417 (*Coriandrum*) with the latter being exceptional in its large number. Furthermore, *Coriandrum* has more repeats than any other plastome across all motif size classes. The longest direct repeat, 254 bp, also belongs to *Coriandrum*. When repeat motifs (both SDRs and SSRs) are plotted according to their locations within genes, introns, and intergenic spacers as the plastome is read from J_{LA} to J_{SA} it is evident that repeat DNA is dispersed evenly across the genome with most partitions having one motif (Fig. 3.2).

Much of the repetitive DNA occurring in *Coriandrum* is located near its LSC–IR boundaries, specifically between IR genes *trnH* and *psbA*. In the other three plastomes, this spacer occurs in the LSC region adjacent to J_{LA} and does not show an increase in repeat DNA content. The intergenic spacer regions near the *psbM* and *trnT* inversion break points in *Carum*, located in the intergenic sequence between *psbM* and *trnE* and between *trnD* and *trnT*, have twice as much repetitive DNA as any of the other three plastomes (Fig. 3.2).

The plastome of *Coriandrum* has a contracted IR encompassing only 12 genes (Fig. 3.3). These genes include *ycf1* through *trnV* (which contains the four ribosomal RNA genes) plus *trnH* and *psbA*. The reduction of the IR to the rRNA genes was designated as boundary position type I (Plunkett and Downie 2000). With the inclusion of *trnH* and *psbA* in the IR, I have designated this updated boundary type as I' (Fig. 3.1). *Coriandrum* also has a partial duplication of *trnV* within the LSC region adjacent to J_{LB}. All other gene adjacencies within the LSC and SSC regions are collinear with those of the *Daucus* plastome.

The plastomes of *Anethum* and *Foeniculum*, both members of tribe Apieae, are 99.32% similar with 422 single nucleotide polymorphisms and 453 indels across 128,726 aligned positions. Each has about a 1500 bp contraction of their IR (Figs. S1 and S2; Fig. 3.1 boundary type D). In addition, there is an insertion of novel, non-coding DNA between J_{LA} and the 3' end

of *trnH*. *Foeniculum* has 392 bp of non-coding sequence between J_{LA} and *trnH*, while *Anethum* has 244 bp of the same non-coding sequence in this region (Fig. S3.3). The remainder of their plastomes are collinear with *Daucus*.

The *Carum* IR has expanded to include all of *rps3* (Fig. S3.4), an expansion that has also resulted in the duplication of *rpl22* (Fig. 3.1 boundary type B). Additionally, two major rearrangements were detected that did not involve the IR: a 571 bp inversion between *psbM* and *trnT*, resulting in the inversion of *trnD-trnY-trnE*; and a 2178 bp inversion from J_{LA} to the 3'*trnK* exon, resulting in the inversion of *trnH* and *psbA*.

In all four plastomes, the SSC–IR boundaries occur within *ycf1*. The amount of *ycf1* contained within the IR varies by 217 bp, with *Daucus* having 1676 bp of duplicated sequence and *Coriandrum* having 1893 bp of duplicated sequence. Both *Anethum* and *Foeniculum* have 1885 bp of *ycf1* duplicated and identical SSC–IR boundary endpoints. Comparisons of the 50 bp of sequence on either side of J_{SB} support the possibility that the presence of two small duplications (18 and 43 bp) in *Carum*, *Anethum*/*Foeniculum*, and *Coriandrum* could facilitate boundary shifts through recombination. *Anethum*/*Foeniculum* have 18 out of the 100 bp flanking J_{SB} in common with *Carum* and 43 bp in common with *Coriandrum*. *Carum* and *Coriandrum* also have 18 bp in common and these identical sequences are in the same location as those in *Anethum*/*Foeniculum*. None of these genomes share any sequence similarity with the 100 bp of sequence flanking the *Daucus* J_{SB}.

Novel DNA Characterization

Between J_{LA} and 3' *trnH*, *Anethum* and *Foeniculum* contain novel, non-coding sequences. These sequences, at 244 and 392 bp in size, are identical over the 244 bp they share (Fig. S3.3). These novel fragments do not match any published plastid DNA sequence.

Instead, they show a short but significant match to non-coding mtDNA in the intergenic spacers between *cob–atp4* and *nad4L–atp4* (Table 3.5).

Primers anchored in mitochondrial gene pairs *cob–atp4* and *nad4L–atp4*, as well as in the novel plastid fragment and *atp4*, were used to try and locate the novel plastid DNA fragment within the mitochondrial genomes of *Anethum* and *Foeniculum* (Table S3.1). Amplifications using primer pairs "mt.cob3f*"* and "mt.orf25.3r," "Kubo1*"* and "Kubo6," and "Kubo1*"* and "Kubo5mod*"* did not produce products. A primer designed within the novel fragment ("fragShortR") was used with "Kubo6" or "mt.orf25.3r", both within *atp4*, and resulted in an amplicon of about 400 bp in *Anethum* and no product in *Foeniculum*. Genome walking from the intergenic sequence between *atp4* and the sequence that matches the novel plastid fragment away from *atp4* in *Anethum* and *Foeniculum* did not produce any new data that were not already available.

Characterization of J_{LA} in 34 additional species of the apioid superclade reveals that for those species having an IR boundary in *rpl2* (Fig. 3.1 boundary type D) there was an insertion of novel DNA in the LSC region bounded by J_{LA} and *trnH*, ranging in size from 40 to 447 bp (Table 3.6). This novel DNA occurs in all examined species with boundary type D, with the exception of *Oedibasis platycarpa* (Fig. S3.5; Table 3.6). While *Oedibasis platycarpa* does have novel DNA within the J_{LA}–*trnH* intergenic spacer it does not match the sequence found in all other species having a type D boundary.

Novel DNA in the J_{LA}–*trnH* region was also detected in *Crithmum maritimum* and *Trachyspermum ammi* (tribe Pyramidoptereae, boundary type B) and in *Aethusa cynapium* and *Enantiophylla heydeana* (tribe Selineae, boundary types E or F; Table 3.6). The *Crithmum* and *Enantiophylla* novel sequence share 106 bp with 78% similarity. Within the 1528 bp fragment in *Aethusa*, there are 83 bp with 92% similarity to non-coding sequences occurring between 5' *rps12* exon and *clpP*. The remaining novel fragments and the hundreds of remaining bp in

Crithmum, *Enantiophylla,* and *Aethusa*, show no similarity to each other or to any other sequences in GenBank (as of 17 April 2013).

The primer "fragShortR" was used with primers "mt.orf25.3r" or "Kubo6" to amplify mtDNA in 13 of these 34 additional species. Sequence data obtained from this region ranged in size from 178 to 525 bp and contains between 16 and 27 bp of sequence that matches the plastome sequence adjacent to the "fragShortR" primer location; the remaining sequences are fragments of non-coding DNA adjacent to *atp4,* as reported in *Daucus* (Iorizza et al. 2012). For these 13 species, I confirmed that a small fragment of sequence matching the novel non-coding DNA in type D plastomes also occurs in their mtDNA. Genome walking in *Anethum*, *Foeniculum*, *Ridolfia segetum,* and *Pastinaca sativa* from the intergenic sequence between *atp4* and the sequence that matches the novel plastid fragment away from *atp4* did not yield any information beyond what was already available in GenBank, with the exception of the *Ridolfia segetum* sequence. In *Ridolfia, cytB* is adjacent to *atp4* and the remaining sequence does not match anything in GenBank.

Inverted Repeat Changes

A survey of J_{LA} in 34 species of the apioid superclade confirms several LSC–IR boundary shifts. The ancestral J_{LA} as typified by *Daucus* (Fig. 3.1 boundary type A) has its LSC–IR boundary within *rps19*. No other species examined herein has its boundary in the same relative position.

The type B boundary location, with an expansion of the IR into *rps3*, is characteristic of tribe Careae and two of four members of tribe Pyramidoptereae (Table 3.6). The inversion of *trnH* and *psbA* in *Carum* also occurs in *Aegokeras caespitosa* and *Falcaria vulgaris*, both members of tribe Careae. However, not all species with an IR expansion to *rps3* have an inversion of *trnH* and *psbA*. *Crithmum* and *Trachyspermum* have IR expansions to *rps3* but

neither has the inversion. Instead, these two species have novel insertions at J_{LA}, of 1463 and 62 bp, respectively.

All species sampled from tribes Apieae and Pimpinelleae, the *Cachrys*, *Conium*, and *Opopanax* clades, and two of four species of tribe Pyramidoptereae have IR boundaries within *rpl2* that is characteristic of the type D IR boundary location (Fig. 3.1). These species have an insertion of novel DNA between J_{LA} and *trnH*. With the exception of *Oedibasis platycarpa*, whose 1034 bp insertion has no sequence similarity to any other taxon examined, sequence alignments of the other taxa indicate that they all share the same fragment (Fig. S3.5).

Species from tribe Selineae have boundary types E and F, characterized by a contraction of the IR into either the *ycf2–trnL* intergenic spacer region or *ycf2* (Table 3.6). The amount of *ycf2* duplicated varies by 293 bp. *Aethusa* and *Enantiophylla* both have additional changes beyond the contraction of the IR, as previously described. The IR boundary type H in *Tordylium aegyptiacum* var. *palaestinum* (tribe Tordylieae) does not have any additional changes beyond the contraction of the IR.

The remaining boundary types, I and I' (Fig. 3.1), occur in the two members of tribe Coriandreae—*Bifora radians* and *Coriandrum.* These two species share little sequence similarity in the genes adjacent to their LSC–IR boundaries. This is due to the inclusion of *trnH* and *psbA* in the IR of *Coriandrum* and repetitive DNA that does not occur in *Bifora*. Within *Coriandrum* the IR is located in the 5' end of *psbA* with only the first 10 bp of the gene being single copy. In *Bifora* the IR has contracted to the intergenic region between *rrn16* and *trnV* making *trnV*, *trnH*, and *psbA* single copy.

Mechanisms of LSC–IR Boundary Change

The partial duplication of *trnV* in *Coriandrum* was the only evidence supporting DSBR as a mechanism of IR change in the plastomes examined herein. The *Coriandrum* plastome had

more repeats than the other three plastomes combined, and more than any other Apiales plastome sequenced to date (Table 3.2). This was due in large part to three repeat motifs in the *trnH–trnV* region, one of 24 bp repeated 7 times, one of 18 bp repeated 21 times, and a motif of 21 bp repeated 4 times. The 18 bp repeat was tandem in most cases with the other motifs breaking up those tandem duplications. Evidence of DSBR was also found near J_{LA} in *Aethusa* in the form of an 83 bp fragment that is a duplication of non-coding DNA between 5' *rps12* and *clpP*. There was no evidence of DSBR among any other species sequenced.

There was only one difference in repeat content among *Anethum*, *Foeniculum*, *Carum*, *Coriandrum*, and *Daucus* plastomes between the ancestral *rps19* boundary location A and the modified boundaries D and I. A thymine mononucleotide repeat ranging in size from nine to 17 bp was present, the longest found in *Daucus*. There was no matching repeat at the LSC–IR boundary in *Anethum*, *Foeniculum*, *Carum*, or *Coriandrum* to facilitate recombination (Table S3.2).

The *Anethum/Foeniculum*, *Carum*, *and Coriandrum* plastomes had 2 to 7 SSRs, with a minimum length of 15 bp, throughout their entire plastomes (Tables 3.2, 3.4, and 3.7). In *Anethum*, *Foeniculum*, and *Carum* SSRs of 15 bp or more are not found at present or ancestral (i.e., near *rps19*) boundary locations. In *Coriandrum* there are more complex repeat motifs identified (5 bp) and these do occur near the present but not the ancestral IR boundary (Table S3.2). *Coriandrum* is the only plastome to have SDRs near the LSC–IR boundary. Sequences near J_{LA} in 13 of the 34 species analyzed contain several direct and inverted SDRs of 20 bp (*Oedibasis platycarpa*) to 300 bp (*Ammi majus*; Table 3.7).

Discussion

Plastomes

The angiosperm plastome is static in structure, with only a few groups exhibiting frequent, dynamic changes. In addition to members of Apiaceae subfamily Apioideae, the following taxa are recognized as regularly having major structural changes involving the IR: Berberidaceae (Kim and Jansen 1994; Ma et al. 2013), Campanulaceae (Cosner et al. 1997; Knox 2014), Fabaceae subfamily Papilionoideae (Palmer et al. 1987; Lavin et al. 1990; Cai et al. 2008; Jansen et al. 2008), and Geraniaceae (Price et al. 1990; Chumley et al. 2006; Guisinger et al. 2011; Weng et al. 2013). Mapping studies of the Apiaceae plastome (Plunkett and Downie 2000) and sequence data presented herein have shown that members of the apioid superclade have diverse IR boundaries. These boundary differences affect the length of the IR and gene adjacencies on the J_{LA} side of the plastome.

In comparison to other Apiales plastomes with and without IR changes *Foeniculum*, *Anethum*, and *Carum* have similar amounts of repetitive elements. When the four Apioideae plastomes are compared to other eudicot species with and without IR changes *Coriandrum* is the only species with similar amounts of SDRs present. These SDRs are located between *trnV* and *trnH* (near J_{LA}) within the IR and are potential sites of recombination.

Recombination across repeat DNA has resulted in many different major structural rearrangements of the plastome including LSC–IR boundary changes (Ogihara et al. 1998; Hupfer et al. 2000; Guo et al. 2007; Lee et al. 2007; Greiner et al. 2008; Martin et al. 2014). In general, in those plastomes with IR boundaries that vary from the ancestral type, there is an increased rate of rearrangement (Cosner et al. 2004; Chumley et al. 2006; Weng et al. 2013). Among the plastomes sequenced herein, *Carum* and *Coriandrum* have different IR boundaries and additional gene order changes from those typical among eudicots. *Carum* has an inversion of the genes between *psbM* and *trnT* that is likely repeat mediated and an inversion of *trnH* and *psbA* that does not have any repeat DNA associated with it. This first inversion occurs in other angiosperm plastomes (Sloan et al. 2012b; Sloan et al. 2013).

Origin of Novel Plastid DNA at J_{LA}

In typical angiosperm plastomes there are 2–9 bp of non-coding DNA between J_{LA} and 3' *trnH* (Raubeson and Jansen 2005). In many members of the apioid superclade there is a larger, novel DNA insertion in this same region. This novel plastid sequence is similar to angiosperm non-coding mtDNA. However, even though this similarity to mtDNA is high, the lengths of the matches are small. The majority of significant hits were near genes *atp4* and *nad4L*, although I was unable to determine the origin of the insertion using PCR or genome walking approaches.

There have been other reported instances of intracellular gene transfer (IGT) within Apiaceae. Goremykin et al. (2009) showed that the intergenic spacer between 3' *rps12* and *trnV* plastid genes in *Daucus* had high sequence similarity to published mtDNA coding sequence. Evidence of transfer, however, was based solely on sequence similarity. Subsequently, Iorizzo et al. (2012) confirmed the presence of this mtDNA fragment within both the *Daucus* plastome and mitochondrial genome and suggested that the transfer was the result of a retrotransposon event.

Other angiosperms possess plastid DNA fragments having sequence similarities to mtDNA. Within *Pelargonium* (Geraniaceae), for example, there is a possible insertion of mtDNA within the *trnA* intron, with this insertion having sequence similarity to the mitochondrial *ACRS* and *pvs-trnA* genes (Chumley et al. 2006). Chumley et al. (2006) reported further that these mtDNA sequences within the *trnA* intron are conserved across many angiosperms. No mechanism was inferred for how these genes were incorporated into the plastome, however, since *ACRS* is in the mitochondrial *tRNA-Ala* intron (Ohtani et al. 2002) recombination is probable.

More recently, Straub et al. (2013) determined that there is an insertion of 2,427 bp into the *rps2–rpoC2* intergenic spacer of the plastome in several species of Apocynaceae tribe

Asclepiadeae. This instance of intracellular gene transfer is a mitochondrial copy of the plastid *rpl2* gene – a mitochondrial paralog of the plastid *rpl2* is transferred back to the plastome. The authors hypothesized that recombination of mtDNA and plastid DNA via traditional DSBR or synthesis-dependent strand annealing (an alternative mechanism of break repair) led to this introgression.

DSBR is a proposed mechanism of plastid DNA introgression into the mitochondrial and nuclear genomes (Leister 2005; Klein et al. 2009) and is the most likely explanation for the incorporation of mtDNA into the Apiaceae plastome. The mitochondrial genome of *Daucus* has paralogs of plastid genes *rpl2* and *trnH* that have maintained gene adjacency. This presents a plausible scenario for incorporation of mtDNA into the plastome through strand hybridization at J_{LA}. The high frequency of mitochondrial genome rearrangements (Palmer and Herbon 1988; Shirzadegan et al. 1989; Sloan et al. 2012a; Gualberto et al. 2013; Noyszewski et al. 2014) would explain why the novel DNA of the plastome is no longer located between the mtDNA paralogous genes *rpl2* and *trnH* and why I was unable to find significant matches near *atp4* or *nad4L* in the examined mitochondrial genomes. Occasionally, the DNA incorporated through DSBR is "filler DNA" that does not match any other region of the genome (Ricchetti et al. 1999; Windels et al. 2003; Cai et al. 2008). The novel DNA adjacent to J_{LA} in members of the apioid superclade, having no sequence similarity whatsoever to any other sequence currently in GenBank, may have been integrated into the plastome though DSBR as "filler DNA."

IR Boundary Changes and Their Mechanisms

The mechanisms proposed to explain IR expansions, such as DSBR and the presence of short dispersed repeats (Palmer 1985; Palmer et al. 1987; Ogihara et al. 1988; Aii et al. 1997; Haberle et al. 2008; Odom et al. 2008), have facilitated IR changes in several species of the

apioid superclade. The short, direct, and inverted repeats occurring around J_{LA} in 13 of the 34 additional species sequenced provide evidence for repeat-mediated IR boundary changes.

The mechanisms relating to IR change are complex and no single mechanism can explain all the variation present. As an example, the *Coriandrum* plastome has at least two different causes that explain its IR boundary changes. First, the contraction of the IR to *trnV* adjacent to the rRNA genes; this contraction is shared with *Bifora radians*, also of tribe Coriandreae. Second, a subsequent expansion of the IR to include *trnH* and *psbA* that may have been repeat mediated. *Bifora's* plastome has a larger contraction of the IR and is the only species examined herein to have *trnV* occurring within the LSC region.

There is no evidence that SSRs are mediating LSC–IR boundary changes (Table S3.2). Species without IRs (i.e., *Erodium*) do not have any SSRs that met our minimum criteria while species with IR boundary changes (i.e., *Pelargonium*) do not have SSRs at ancestral or present IR boundary locations. However, there is evidence that more complex repeats like SDRs may be a common mechanism of LSC–IR boundary change in species of the apioid superclade (Table 3.7). Guisinger et al. (2011) reported that size of a repeat motif correlates with frequency of inversions, with larger repeats rearranging more frequently. This implies that *Coriandrum* and *Ammi majus* should have more genomic rearrangements than *Carum* and *Spermolepis*. This trend, however, was not observed in the four Apioideae plastomes sequenced herein, where *Carum* had more inversions than *Coriandrum*.

Insertion of tRNAs is often cited as evidence of DSBR (Haberle et al. 2008) and DSBR is the most likely mechanism for the partial duplication of *trnV* in *Coriandrum*. Evidence of DSBR is also found in *Aethusa* in the form of an 83 bp fragment located near *trnH* that is the duplication of non-coding DNA between 5' *rps12* and *clpP*. This duplication is not a likely cause of IR boundary change since it occurs within a larger fragment of novel DNA.

Conclusions

IR changes within members of the apioid superclade of Apiaceae subfamily Apioideae are complex, with multiple mechanisms generating changes at their single copy – IR boundaries. DSBR and SDRs are the most likely mechanisms of IR change. To better understand how novel DNA has been integrated into these plastomes, targeted sequencing of additional plastomes and mitochondrial genomes will be useful to show how frequently DSBR is occurring in the group and what the souce of this novel DNA might be.

References

Aii J, Kishima Y, Mikami T, Adachi T. 1997. Expansion of the IR in the chloroplast genomes of buckwheat species is due to incorporation of an SSC sequence that could be mediated by an inversion. Curr Genet 31:276-279.

Ajani Y, Ajani A, Cordes JM, Watson MF, Downie SR (2008) Phylogenetic analysis of nrDNA ITS sequences reveals relationships within five groups of Iranian Apiaceae subfamily Apioideae. Taxon 57:383–401.

Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK (2008) Extensive reorganization of the plastid genome of *Trifolium* subterraneum (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol 67:696-704.

Catalano SA, Saidman BO, Vilardi JC (2009) Evolution of small inversions in chloroplast genome: a case study from a recurrent inversion in angiosperms. Cladistics 25: 93-104.

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175-2190.

Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr Genet 31:419-429.

Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol Biol 4:27.

Cummings MP, Nugent JM, Olmstead RG, Palmer JD (2003) Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms. Curr Genet 43:131-138.

Dolan A. Simple Sequence Repeat (SSR) Extractor Utility. http://www.aridolan.com/ssr/ssr.aspx. Accessed 23 July, 2014.

Downie SR, Jansen RK (2015) A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot 40:336-351.

Downie SR, Katz-Downie DS (1996) A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal transcribed spacer sequences. Am J Bot 83:234–251.

Downie SR, Katz-Downie DS, Watson MF (2000) A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpoC1* intron sequences: towards a suprageneric classification of subfamily Apioideae. Am J Bot 87:273–292.

Downie SR, Ramanath S, Katz-Downie DS, Llanas E (1998) Molecular systematics of Apiaceae subfamily Apioideae: phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer and plastid rpoC1 sequences. Am J Bot 85:563–591.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15.

Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol 26:99-110.

Goulding SE, Wolfe KH, Olmstead RG, and Morden CW (1996) Ebb and flow of the chloroplast inverted repeat. Molec General Genet 252:195-206.

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8:195-202.

Greiner S, Wang X, Rauwold U, Silber MV, Mayer K, Meurer J, Haberer G, Herrmann RG (2008) The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. 2008. Nucl Acids Res 36:2366-2378.

Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A (2013) The plant mitochondrial genome: dynamics and maintenance. Biochimie 100:107-120.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28:583-600.

Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, Schwarzbach AE, Mower JP (2014) Predominant and substoichiometric isomers of the plastid genome coexist within juniperus plants and have shifted multiple times during cupressophyte evolution. Genome Biol Evol 6:580-590.

Guo X, Catillo-Ramírez S, González V, Bustos P, Fernádez-Vázquez JL, Santamaría RI, Arellano J, Cevallos MA, Dávila G (2007) Rapid evolutionary change of common bean (*Phaseolus vulgaris* L) plastome, and the genomic diversification of legume chloroplasts. BMC Genomics 8:228.

Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol 66:350-361.

Hansen DR, Dastidara SG, Caia Z, Penaflorb C, Kuehl JV, Boore JL, Jansen RK (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). Mol Phylogenet Evol 45:547-563.

Hupfer H, Swaitek M, Hornung S, Herrmann RG, Maier RM, Chiu W-L, Sears B (2000) Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenothera* plastomes. Mol Gen Genet 263:581-585.

Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P (2012). De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into anangiosperm plastid genome. BMC Plant Biol 12:61.

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HW, Kuehl JV, McNeal JR, Leebens-Mack J, Cui L (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Method Enzymol 395:348-384.

Jansen RK, Wojciechowski MF, Sanniyasi E, Lee S-B, Daniell H (2008) Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). Mol Phylogenet Evol 48:1204-1217.

Katz-Downie DS, Valiejo-Roman CM, Terentieva EI, Troitsky AV, Pimenov MG, Lee B, Downie SR (1999) Towards a molecular phylogeny of Apiaceae subfamily Apioideae: additional information from nuclear ribosomal DNA ITS sequences. Plant Syst Evol 216:167–195.

Kim YD, Jansen RK (2009) Characterization and phylogenetic distribution of a chloroplast DNA rearrangement in the Berberidaceae. Plant Syst Evol 193:107-14.

Kleine T, Maier UG, Leister D (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol. 60:115–138.

Knox E (2014) The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. PNAS 111:11097–11102.

Kubo T, Yamamoto MP, Mikami T (2000) The *nad4L–orf25* gene cluster is conserved and expressed in sugar beet mitochondria. Theor Appl Genet 100:214-220.

Lavin M, Doyle JJ, Palmer JD (1990) Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evol 44:390-402.

Lee HL, Jansen RK, Chumley TW, Kim KJ (2007) Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple overlapping inversions. Mol Biol Evol 24:1161-1180.

Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organellar DNA. TRENDS Genet 21:655-663.

Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X (2013) The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene 528:120-131.

Martin W, Hermann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol 118:9-17.

Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393:162–165.

Martin GE, Rousseau-Gueutin M, Cordonnier S, Lima O, Michon-Coudouel S, Naquin D, de Carvalho JF, Aïnouche M, Salmon A, Aïnouche A (2014) The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. Annals of Bot 113:1197-1210.

Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggied L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermiin LS, Wolfe KH (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13:645-658.

Noyszewski AK, Ghavami F, Alnermer LM, Soltani A, G YQ, Huo N, Meinhardt S, Kianian PMA, Kianian SF (2014) Accelerated evolution of the mitochondrial genome in an alloplasmic line of durum wheat. BMC Genomics 15:67.

Odom OW, Baek K-H, Dani RN, Herrin DL (2008) *Chlamydomonas* chloroplasts can use short dispersed repeats and multiple pathways to repair a double-strand break in the genome. Plant J 53:842-853.

Ogihara Y, Terachi T, Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. PNAS USA 85:8573-8577.
Ogihara Y, Terachit T, Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc. Natl. Acad. Sci. USA 85:8573-8577.
Ohtani K, Yamamoto H, Akimitsu K (2002) Sensitivity to *Alternaria alternata* toxin in citrus because of altered mitochondrial RNA processing. PNAS 99:2439-2444.
Palmer JD (1985) Comparative organization of chloroplast fenomes. Annual Review Genet 19:325-354.
Palmer JD, Nugent JM, Herbon LA (1987) Unusual structure of geranium chloroplast DNA: a triple-sized repeat, extensive gene duplications, multiple inversions, and new repeat families. PNAS USA 84:769-773.
Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol 28:87-97.
Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK, editors. The molecular biology of plastids. New York: Academic Press; 1991. p. 5-53.
Plunkett GM, Downie SR (1999) Major lineages within Apiaceae subfamily Apioideae: A comparison of chloroplast restriction site and DNA sequence data. Am J Bot 86:1014-1026.
Plunkett GM, Downie SR (2000) Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Syst Bot 25:648-667.
Price RA, Calie PJ, Downie SR, Logsdon JM, Palmer JD (1990) Chloroplast DNA variation in the Geraniaceae—a preliminary report. In: Vorster P, editor. The International Geraniaceae Symposium. Republic of South Africa: The University of Stellenbosch p. 237-244.
Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ, editor. Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants. London: CAB International.
Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics 8:174.
Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs doublestrand breaks in yeast chromosomes. Nature 402:96-100.
Rice DW, Palmer JD (2006) An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. BMC Biology 4:31.
Richardson AO, Palmer JD (2007) Horizontal gene transfer in plants. J Exp Bot 58:1-9.
Ruhlman T, Lee S, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H (2006) Complete plastid genome sequence of *Daucus carota*: Implications for biotechnology and phylogeny of angiosperms. BMC Genomics 7: 222.
Shirzadegan M, Christey M, Earle ED, Palmer JD (1989) Rearrangement, amplification, and assortment of mitochondrial DNA molecules in cultured cells of *Brassica campestris*. Theor Appl Genet 77:17-25.
Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLOS Biol 10:e1001241.
Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene.* Genome Biol Evol 4:294-306.

Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR (2013) A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). Molec Phylogen Evol 72:82-89.

Spalik K, Reduron J-P, Downie SR (2004) The phylogenetic position of *Peucedanum* sensu lato and allied genera and their placement in tribe Selineae (Apiaceae, subfamily Apioideae). Plant Syst Evol 243:189-210.

Spalik K, Downie SR (2007) Intercontinental disjunctions in *Cryptotaenia* (Apiaceae, Oenantheae): an appraisal using molecular data. J Biogeog 34:2039-2054.

Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A (2013) Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). Genome Biol Evol 5:1872-1885.

Sugiura M, Shinozaki K, Zaita N, Kusuda M, Kumano M (1986) Clone bank of the tobacco (*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments: mapping of eleven ribosomal protein genoes. Plant Sci 44:211-216.

Weng M-L, Blazier JC, Govindu M, Jansen RK (2013) Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol 31:645-659.

Windels P, De Buck S, Van Bockstaele E, De Loose M, Depicker A (2003) T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. Plant Physiol 133:2061-2068.

Winter PJD, Magee AR, Phephu N, Tilney PM, Downie SR, and van Wyk B-E (2008) A new generic classification for African peucedanoid species (Apiaceae). Taxon 57:347-364.

Tables and Figures

Table 3.1 Accessions of Apiaceae subfamily Apioideae examined for J_{LA} changes. The four species whose entire plastomes have been sequenced herein are indicated by asterisks.

Species	Source
Aegokeras caespitosa (Sibth. & Sm.) Raf.	Plunkett and Downie 2000
Aethusa cynapium L.	Plunkett and Downie 2000
Ammi majus L.	Downie et al. 1998
Ammoselinum butleri (Engelm. ex S. Watson) J.M. Coult. & Rose	USA, Mississippi, Leflore Co., West of Greenwood, *Cryson 13404* (MO)
Anethum graveolens L.*	Cultivated at Central Washington University from seeds purchased from Burpee® (ELRG).
Anethum graveolens L.	Total genomic DNA from plant material obtained from a local market (ELRG).
Apiastrum angustifolium Nutt. ex Torr. & A. Gray	USA, California, Riverside Co., Vail Lake Area; *Boyd et al. 3848* (MO 4000398)
Apium graveolens L.	Downie et al. 1998
Apium prostratum Vent.	Spalik et al. 2010
Azilia eryngioides (Pau) Hedge & Lamond	Ajani et al. 2008
Bifora radians Bieb.	Downie et al. 1998
Cachrys libanotis L.	Ajani et al. 2008
Carum carvi L.*	Cultivated at UIUC from seeds purchased from Burpee®; *Downie 3219* (ILL)
Conium maculatum L.	Downie et al. 1998
Coriandrum sativum L.*	Cultivated at Central Washington University from seeds purchased from Burpee® (ELRG).
Coriandrum satrivum L.	Total genomic DNA from plant material obtained from a local market (ELRG).
Crithmum maritimum L.	Plunkett and Downie 2000
Deverra burchellii (DC.) Eckl. & Zeyh.	Winter et al. 2008
Deverra triradiata Hochst. ex Boiss.	Downie et al. 2000
Diplotaenia cachrydifolia Boiss.	Ajani et al. 2008
Enantiophylla heydeana J.M. Coult. & Rose	Downie and Katz-Downie 1996
Falcaria vulgaris Bernh.	Downie et al. 1998
Ferulago nodosa (L.) Boiss.	Italy, Sicily, Melilli, Monti Iblei; leaf material provided by S. Brullo, Departimento di Botanica, Università di Catania, Catania, Italy
Foeniculum vulgare Mill.*	Cultivated at Central Washington University from seeds purchased from Burpee® (ELRG).
Foeniculum vulgare Mill.*	Total genomic DNA from plant material obtained from a local market (ELRG).
Haussknechtia elymaitica Boiss.	Ajani et al. 2008
Naufraga balearica Constance & Cannon	Downie et al. 2000

Table 3.1 (cont.)

Oedibasis platycarpa (Lipsky) Koso-Pol.	Katz-Downie et al. 1999
Opopanax persicus Boiss.	Ajani et al. 2008
Petroselinum crispum (P. Mill.) A.W. Hill	Downie et al. 1998
Pimpinella major (L.) Huds.	Plunkett & Downie 2000
Pimpinella peregrina L.	Downie et al. 1998
Prangos goniocarpa (Boiss.) Zohary	Ajani et al. 2008
Ridolfia segetum (L.) Moris	Downie et al. 1998
Selinum carvifolia (L.) L.	Spalik et al. 2004
Seseli webbii Coss.	Spalik et al. 2004
Sison segetum L.	France, Val-de-Marne, Créteil, au Mont-Mesly. *Reduron 19770711-01* (ILL)
Spermolepis inermis (Nutt. ex DC.) Mathias & Constance	USA, Illinois, Carroll Co., Savanna Army Depot., Green Island, 30 June 1993, *Phillippe et al. 22290* (ILLS)
Stoibrax dichotomum (L.) Raf.	Spalik & Downie 2007
Tordylium aegyptiacum (L.) Lam. var. *palaestinum* (Zoh.) Zoh.	Downie et al. 1998
Trachyspermum ammi (L.) Sprague ex Turrill	Downie et al. 1998

Table 3.2 Comparison of short dispersed repeats among plastomes with and without LSC–IR boundary changes.

			Repeat length (bp)			
Family	Species	GenBank accession	30-49	50-69	70-99	≥100
Apiaceae	*Anethum graveolens*[a]	KR011055	19	1	0	0
	Anthriscus cerefolium	NC_015113	22	2	0	0
	Carum carvi[a]	KR048286	17	2	0	0
	Coriandrum sativum[a]	KR002656	278	86	39	14
	Daucus carota	NC_008325	22	1	1	0
	Foeniculum vulgare[a]	KR011054	17	1	2	3
Araliaceae	*Aralia undulata*	NC_022810	23	5	1	2
	Eleutherococcus senticosus	NC_016430	22	4	1	0
	Kalopanax septemlobus	NC_022814	19	4	1	0
	Metapanax delavayi	NC_022812	19	4	1	0
	Panax ginseng	NC_006290	17	3	0	2
	Panax ginseng 'Damaya'	KC686331	17	3	0	2
	Panax ginseng 'Ermaya'	KC686332	17	3	0	2
	Panax ginseng 'Gaolishen'	KC686333	17	3	0	2
	Schefflera delavayi	NC_022813	25	3	1	0
	Brassaiopsis hainla	NC_022811	21	3	1	0
Asteraceae	*Helianthus annuus*	NC_007977	98	1	0	0
	Parthenium argentatum	NC_013553	64	2	2	1
Camanulaceae	*Trachelium caeruleum*[a]	NC_010442	242	61	22	25
Fabaceae	*Pisum sativum*[b]	NC_014057	46	4	3	1
	Trifolium subterraneum[b]	NC_011828	216	102	63	112
Geraniaceae	*Erodium carvifolium*[b]	NC_015083	41	6	8	2
	Geranium palmatum[b]	NC_014573	230	80	53	35
	Monsonia speciosa[a]	NC_014582	59	27	14	10
	Pelargonium x *hortorum*[a]	NC_008454	120	30	12	20
Ranunculaceae	*Megaleranthis saniculifolia*	NC_012615	13	0	0	0
	Ranunculus macranthus[a]	NC_008796	7	0	0	0
Schisandraceae	*Illicium oligandrum*[a]	NC_009600	8	0	0	0
Solanaceae	*Nicotiana tabacum*	NC_001879	12	1	0	0

[a] IR is different than ancestral IR type (i.e., LSC–IR junctions are not in or near *rps19*)
[b] Plastome does not have an IR

Table 3.3 Comparison of features of the four apioid superclade plastomes sequenced herein and *Daucus* (Ruhlman et al. 2006). Gene counts within parentheses include both copies of the IR.

Feature	*Coriandrum*	*Anethum*	*Foeniculum*	*Carum*	*Daucus*
Total length (bp)	146,519	153,356	153,628	155,449	155,911
LSC length (bp)	99,231	86,506	86,659	83,672	84,242
SSC length (bp)	17,486	17,518	17,471	17,549	17,567
IR length (bp)	14,901	24,666	24,749	27,114	27,051
No. of protein coding genes	79 (81)	79 (85)	79 (85)	79 (88)	79 (86)
No. of tRNA genes	30 (36)	30 (37)	30 (37)	30 (37)	30 (37)
No. of rRNA genes	4 (8)	4 (8)	4 (8)	4 (8)	4 (8)

Table 3.4 Total amount of SDRs and SSRs in the four apioid superclade plastomes sequenced herein and *Daucus* (Ruhlman et al. 2006).

	SDR repeat motif length (bp)					SSR repeat motif length (bp)					
Species	30-49	50-69	70-99	≥100	Total	1	2	3	4	5	Total
Anethum graveolens	19	1	0	0	20	39	15	3	6	1	64
Carum carvi	17	2	0	0	19	33	17	2	6	2	60
Coriandrum sativum	278	86	39	14	417	39	17	2	5	3	66
Daucus carota	22	1	1	0	24	32	12	6	7	2	59
Foeniculum vulgare	21	1	2	3	27	40	15	2	6	2	65

Table 3.5 Results of top 10 BLAST searches of the GenBank nucleotide database (as of 9 March 2015) querying the J_{LA} insertion sequence in *Foeniculum vulgare*. Only hits with lengths longer than 60 bp and a percent similarity of at least 90 are reported (note that all whole mitochondrial genomes have secondary matches of shorter lengths as well).

Accession	Species	Location	Length of match (bp)	Percent similarity
JQ248574	*Daucus carota* subsp. *sativus*	*cob–atp4*	121	92
AY007821	*Daucus carota*	*cob–atp4*	121	92
AY007816	*Daucus carota*	*cob–atp4*	121	92
HM367685	*Vigna radiata*	*nad4L–atp4*	67	93
HM367685	*Vigna radiata*	*cob–trnW–cp*	67	93
KF815390	*Helianthus annuus*	*nad4L–atp4*	67	91
AP012599	*Vigna angularis*	*nad4L–atp4*	67	91
JN87255	*Lotus japonicus*	*nad4L–atp4*	67	90
HQ874649	*Ricinus communis*	*nad4L–atp4*	64	91
JX065074	*Gossypium hirsutum*	*nad4L–trnS*	71	85

Table 3.6 Characterization of J_{LA} changes in species of Apiaceae subfamily Apioideae. Accession information is provided in Table 3.1.

Tribe	Species	LSC–IR boundary location	Boundary type	No. of bp between J_{LA} and *trnH*
Apieae	*Ammi majus*	*rpl2*	D	390
	Apium graveolens	*rpl2*	D	214
	Apium prostratum	*rpl2*	D	212
	Anethum graveolens	*rpl2*	D	244
	Deverra burchellii	*rpl2*	D	363
	Deverra triradiata	*rpl2*	D	364
	Foeniculum vulgare	*rpl2*	D	392
	Naufraga balearica	*rpl2*	D	206
	Petroselinum crispum	*rpl2*	D	203
	Ridolfia segetum	*rpl2*	D	377
	Seseli webbii	*rpl2*	D	127
	Stoibrax dichotomum	*rpl2*	D	195
Cachrys clade	*Azilia eryngioides*	*rpl2*	D	248
	Cachrys libanotis	*rpl2*	D	447
	Diplotaenia cachrydifolia	*rpl2*	D	443
	Ferulago nodosa	*rpl2*	D	220
	Prangos goniocarpa	*rpl2*	D	443
Careae	*Aegokeras caespitosa*	*rps3*[1]	B	0
	Carum carvi	*rps3*[1]	B	0
	Falcaria vulgaris	*rps3*[1]	B	0
Conium clade	*Conium maculatum*	*rpl2*	D	224
Coriandreae	*Bifora radians*	*16S-trnV*[2] IGS	I	0

Table 3.6 (cont.)

	Coriandrum sativum	*psbA*[3]	I’	0
Opopanax clade	*Opopanax persicus*	*rpl2*	D	242
Pimpinelleae	*Haussknechtia elymaitica*	*rpl2*	D	40
	Pimpinella major	*rpl2*	D	161
	Pimpinella peregrina	*rpl2*	D	61
Pyramidoptereae	*Crithmum maritimum*	*rps3*	B	1463
	Oedibasis platycarpa	*rpl2*	D	1034
	Sison segetum	*rpl2*	D	127
	Trachyspermum ammi	*rps3*	B	62
Selineae	*Aethusa cynapium*	*ycf2*	E	1528
	Ammoselinum butleri	*ycf2*	E	0
	Apiastrum angustifolium	*ycf2*	E	8
	Enantiophylla heydeana	*ycf2–trnL* IGS	F	657
	Selinum carvifolia	*ycf2*	E	0
	Spermolepis inermis	*ycf2*	E	0
Tordylieae	*Tordylium aegyptiacum* var. *palaestinum*	*ndhB* intron	H	12

[1] inversion of *trnH–psbA.*
[2] *trnV* and some intergenic sequence are within the LSC; 18 bp between IRa and *trnV.*
[3] only 10 bp of *psbA* are within the LSC.

Table 3.7 SDRs present within each J_{LA} boundary sequence identified using BLASTN.

Species	Repeat location	Repeat type	Length of repeat
Aethusa cynapium	*trnH–psbA*	Inverted	60
	ndhB–trnL	Inverted	58
	Within novel DNA	Inverted	61
Ammi majus	*ycf2–trnI*	Inverted	300
Ammoselinum butleri	*trnH–psbA*	Inverted	72
Apiastrum angustifolium	*trnH–psbA*	Inverted	60
Apium graveolens	Within novel DNA	Inverted	58
Apium prostratum	Within novel DNA	Inverted	76
Conium maculatum	*trnH–psbA*	Inverted	100
Crithmum maritimum	*rps3–trnH*	Direct	105
	rps3–trnH	Direct	105
	rps3–trnH	Direct	84
	rps3–trnH	Direct	56
Enantiophylla heydeana	*trnH–psbA*	Inverted	60
	trnH–psbA	Inverted	57
	trnH–psbA	Direct	31
	Within novel DNA	Direct	31
	ndhB–trnL	Inverted	58
Oedibasis platycarpa	*trnH–psbA*	Inverted	76
	rpl2	Inverted	34
	rpl2	Inverted	20
Petroselinum crispum	*trnH–psbA*	Inverted	58
Spermolepis inermis	*trnH–psbA*	Inverted	50
Tordylium aegyptiacum var. *palaestinum*	*trnH–psbA*	Inverted	56

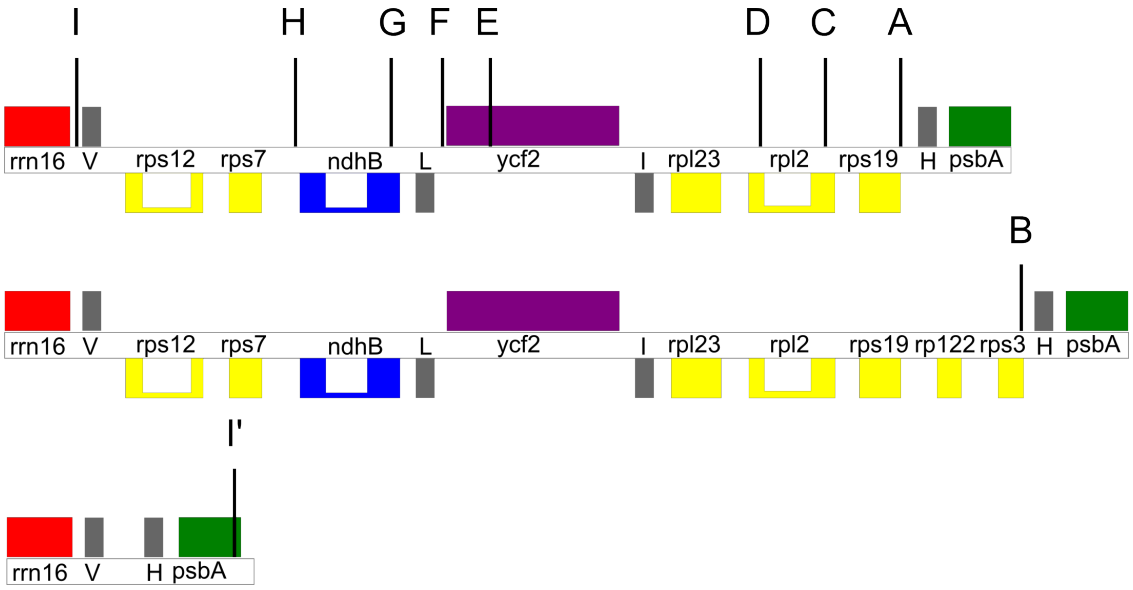


Fig. 3.1 Schematic showing the genes (boxes) and gene adjacencies possible at J_{LA} in species of Apiaceae subfamily Apioideae. Genes transcribed counterclockwise are on the top and genes transcribed clockwise are on the bottom. The horizontal lines show which gene region is adjacent to LSC gene *trnH* in species with that particular IR boundary shift with letter designations and lines to show IR boundary labels first described by Plunkett and Downie (2000). Their boundary types are approximate and cover a range near boundaries indicated in figure and not exact locations due to size of probes used in mapping. Boundary location I' in the bottom panel is a modification of their system for *Coriandrum* to show an IR contraction to the rRNA genes with the inclusion of *trnH* and *psbA*.

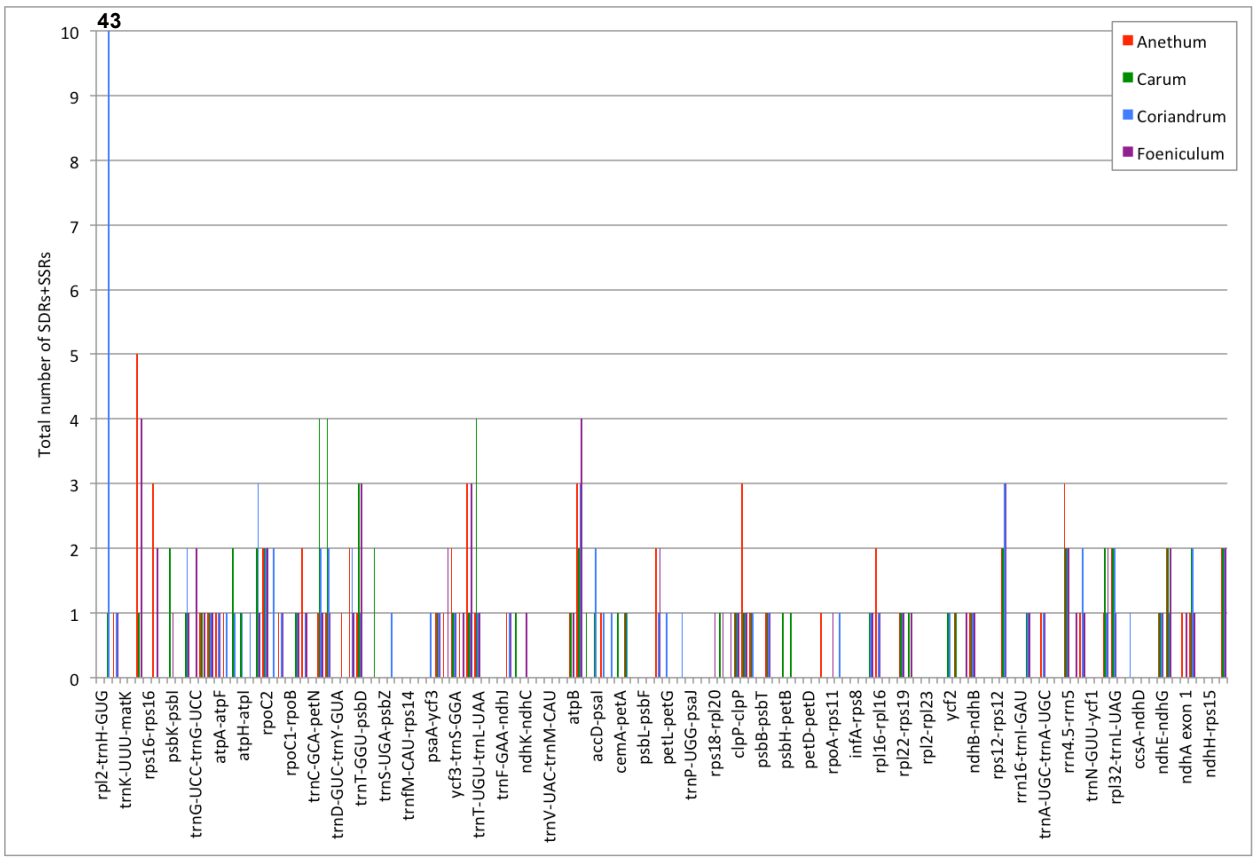


Fig. 3.2 Comparison of total number of repetitive elements, SDRs and SSRs, in each region of a linearized plastid genome from J_{LA} through J_{SA}. The position of genes and intergenic regions that are involved in rearrangements have been moved from their original location to match the orientation found in *Anethum graveolens.* Every gene, intergenic region, and intron that had an SDR or SSR is included (Table S3.3), however, only 48 labels on the X axis are included for readability. Although the Y axis terminates at 10, *Coriandrum* has one region with 43 repeat elements (noted above the bar).

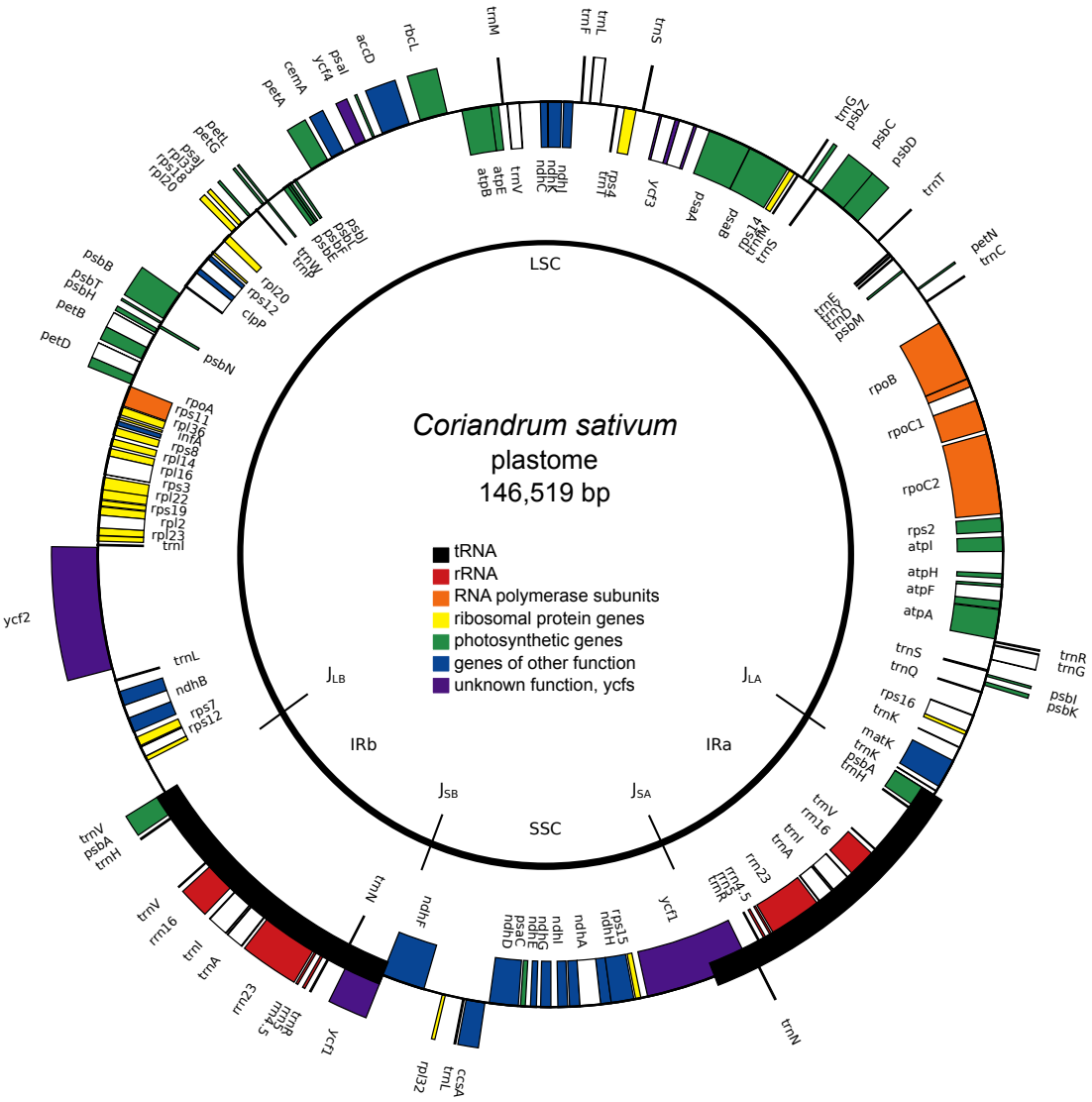


Fig. 3.3 Circular plastome map of *Coriandrum sativum*. Genes are represented by boxes; those outside the circle are transcribed clockwise and those inside the circle are transcribed counterclockwise.

Supplementary Tables and Figures

Table S3.1 Primers used in the amplification and sequencing of J_{LA} and mitochondrial DNA in representative members of the apioid superclade in Apiaceae subfamily Apioideae.

Primer name	Location	Sequence	Reference (if applicable)
mt.cob.3f	Mitochondria, *cob*	GCG GAT YGC TTA CTA CTA GG	
mt.orf25.3r	Mitochondria, *atp4*	GTC TTC AGG ACG ATC TAG TCA	
Kubo1	Mitochondria, 5' *nad4L*	CTC TTA CAT TCT ACG TTC CCG	Kubo et al. 2000
Kubo6	Mitochondria, 5' *atp4*	TCT TCT TCG AAC TTG ATG CAC	Kubo et al. 2000
Kubo5mod	Mitochondria, 3' *nad4L*	GTT ATT ACT TTC CGA GTC CG	Modified from Kubo et al. 2000
Api.mt.GSPa	Mitochondria, *atp4*–*cob* spacer	CTT CGA ACT TGA TGC ACA ATA GAT GG	
Api.mt.GSPb	Mitochondria, *atp4*–*cob* spacer	GCA GCA AAT AGC ATC TTT CTA GCC T	
Api.mt.GSPc	Mitochondria, *atp4*–*cob* spacer	GGT TTA GGA AAG GAC TTT AGA ATC GGA T	
fragShortR	Mitochondria, plastid novel DNA	ARA GGM CCT GAC CTG CCA A	
psbA.3f	Plastid, *psbA*	GCT AAC CTT GGT ATG GAA GT	
6.1r	Plastid, *trnH*	GTA GSC AAG TGG AYY AGG GC	Raubeson unpublished
rps3.3f	Plastid, *rps3*		
9	Plastid, *trnI*	GCA TCC ATG GCT GAA TGG TTA AAG C	Raubeson unpublished
JLaID	Plastid, 5' *rpl2*	TCT GTC CCA TAA TAG GTC CC	
ycf2.2004	Plastid, *ycf2*	AAT ATC GAT TGC TTG TTG AA	
ycf2.840r	Plastid, *ycf2*	TTC CGG AAG CAG ATG ATT A	
ycf2.3700r	Plastid, *ycf2*	TCT TAG AAC GTA TTG ATT TGA C	
ycf2.5800r	Plastid, *ycf2*	CTC GTG TCT GGT ACT GCA T	
ycf2.6100r	Plastid, *ycf2*	ACT GAT AAC TCT CGG ATA GA	
trnLcaa5f	Plastid, *trnL*	ATG GTA GAC ACG CGA GAC TC	
rps12.3f	Plastid, *rps12*	GAT CGT CAA CAA GGG CGT TC	
rps7.3f	Plastid, *rps7*	CCG AAT TAG TGG ATG CTG CC	
trnV	Plastid, *trnV*	TCT ACC GCT GAG TTA TAT CCC	
rrn16.trnV.igs	Plastid, *trnV*–*rrn16* spacer	AGGA TTC GGA ATT GTC TTT CA	
rrn16r	Plastid, *rrn16*	AGC GTT CAT CCT GAG CCT GG	Raubeson unpublished

Table S3.2 Simple sequence repeat DNA with found in plastomes with no LSC–IR boundary change (*Daucus* and *Helianthus*), no IR present (*Erodium* and *Pisum*), with LSC–IR boundary changes (*Anethum*, *Carum*, *Coriandrum*, *Foeniculum*, *Illicium*, *Jacobaea*, *Pelargonium*, and *Ranunculus*). All genomes were analyzed with only one copy of the IR if an IR was present.

Species[1]	Single copy genome size	Motif	Repeat size	Location of repeat
Anethum graveolens	128,691	ATAT	16	*trnR–atpA* IGS[2]
		T	19	*atpF* intron
		TTTTA	15	*trnfM–rps14* IGS
		A	22	*atpB–rbcL* IGS
		A	16	*ndhE–ndhG* IGS
Carum carvi	128,337	ATTCA	15	*matK*
		TTTTA	15	*trnfM–rps14* IGS
Coriandrum sativum		ATAT	16	*atpF–atpH* IGS
		ATTAG	15	*atpH–atpI* IGS
		TTTTA	15	*trnfM–rps14* IGS
		TATTT	15	*trnT–trnL* IGS
		ATAT	16	*trnW–trnP* IGS
		T	15	*clpP–psbB* IGS
		T	15	rps12–*trnV* IGS
Daucus carota	125,057	A	15	*trnK–rps16* IGS
		T	17	*rps19–rpl2* IGS
Erodium carvifolium[3]	116,935			
Foeniculum vulgare	128,880	ATAT	16	*trnR–atpA* IGS
		TTTTA	15	*trnfM–rps14* IGS
		A	15	*atpB–rbcL* IGS
		A	19	*atpB–rbcL* IGS
		TATAA	15	*accD–psaI* IGS
		A	20	*rpl14–rpl16* IGS
Helianthus annuus	126,471	T	15	*trnY–rpoB* IGS
		T	16	*atpF–atpA* IGS
		A	18	*trnS–psbZ* IGS
		A	28	*psaA–ycf3* IGS
		A	15	*trnT–trnL* IGS
		T	25	*atpB–rbcL* IGS
		A	16	*atpB–rbcL* IGS
		T	22	*petA–psbJ* IGS
		A	23	*psbE–petL* IGS
		T	16	*petG–trnW* IGS
		A	15	*clpP* intron
		T	22	*rps8–rpl14* IGS
		T	23	*rpl14–rpl16* IGS
		A	31	*rrn16*
		GAA	15	*ycf1*
		A	27	*ndhA* intron
		A	16	*ndhD–ccsA* IGS

Table S3.2 (cont.)

Illicium oligandrum	129,203	A	15	*rps16* intron
		A	15	*ycf3* intron
		TAT	15	*trnT–trnL* IGS
		T	15	*atpB–rbcL* IGS
		A	16	*rps18–rpl20* IGS
		T	16	*clpP* intron
		T	20	*rpl14–rpl16* IGS
		A	15	*ycf2*
		A	15	*ndhF–rpl32* IGS
Jacobaea vulgaris	125,901	A	17	*atpI–atpH* IGS
		T	15	*atpF–atpA* IGS
		T	18	*atpB–rbcL* IGS
		T	17	*rps11*
		TTAT	16	*rpl16–rps3* IGS
Pelargonium x *hortorum*	142,201	A	15	*rpl33*
		T	15	*trnfM–psbD* IGS
		A	17	*rps19*
		A	15	*petB–petD* IGS
		A	16	*petB*–IGS
Pisum sativum	122,169	A	15	*ycf1*
		A	15	*rps2–rpoC2* IGS
		T	16	*rps18*
		T	15	*rps18–rpl33* IGS
Ranunculus macranthus	129,341	T	15	*rps2–rpoC2* IGS
		CAAAT	15	*trnS–rps4* IGS
		T	16	*ndhC–trnV* IGS
		TAA	15	*petA–psbJ* IGS
		ATAT	16	*rpl16* intron
		TTATA	15	*rps15–ndhH* IGS

[1] Accession information for species can be found in Table 3.2.
[2] IGS is an abbreviation for intergenic spacer (the DNA between coding regions).
[3] *Erodium* did not have any SSRs meeting the minimum requirement of 15 bp.

Table S3.3 Location of repeat DNA (SDRs and SSRs) used to generate Fig. 3.3.

	Total number of SDRs				Total number of SSRs			
Region	*Anethum*	*Carum*	*Coriandrum*	*Foeniculum*	*Anethum*	*Carum*	*Coriandrum*	*Foeniculum*
rpl2–trnH						1	2	
trnH–psbA or trnH–trnV			41					
psbA–trnK					1		1	1
trnK–matK								
matK–trnK								
trnK–rps16					5	1		4
rps16 intron								
rps16–trnQ					3			2
trnQ–psbK								
psbK–psbI						2		1
psbI–trnS								
trnS–trnG						1	2	1
trnG intron				1				1
trnG–trnR					1	1		1
trnR–atpA					1	1	1	1
atpA–atpF					1		1	1
atpF intron					1		1	
atpF–atpH						2	1	
atpH–atpI						1	1	
atpI–rps2							1	
rps2–rpoC2			1	1		2	2	
rpoC2					2	2	2	2
rpoC2–rpoC1			2					
rpoC1 intron					1		1	1
rpoC1–rpoB								
rpoB						1	1	1
rpoB–trnC					2		1	1
trnC–petN								
petN–psbM or psbM–trnE	1	2	1	1		2	1	
psbM–trnD or trnD–trnT	1	2	1			2	1	1
trnD–trnY								
trnY–trnE					1			
trnE–trnT	1		1		1		1	1
trnT–psbD		2			1	1	1	3
psbD–psbC								
psbC–trnS		1				1		
trnS–psbZ								
psbZ–trnG							1	

Table S3.3 (cont.)

trnG–trnfM								
trnfM–rps14								
rps14–psaB								
psaB–psaA								
psaA–ycf3							1	
ycf3 intron	1	1	1	1				
ycf3 intron					1			2
ycf3–trnS					2	1	1	1
trnS–rps4					1			1
rps4–trnT	2			2	1	1	1	1
trnT–trnL		4			1		1	1
trnL intron								
trnL–trnF								
trnF–ndhJ								
ndhJ					1		1	1
ndhJ–ndhK						1		
ndhK–ndhC								1
ndhC–trnV								
trnV intron								
trnV–trnM								
trnM–atpE								
atpE–atpB								
atpB					1	1		1
atpB–rbcL				1	3	2	3	3
rbcL–accD						1		
accD–psaI			1			1	1	
psaI–ycf4					1		1	
ycf4–cemA							1	
cemA–petA						1		
petA–psbJ					1	1	1	
psbJ–psbL								
psbL–psbF								
psbF–psbE								
psbE–petL	2		1	2				
petL–petG			1					
petG–trnW								
trnW–trnP							1	
trnP–psaJ								
psaJ–rpl33								
rpl33–rps18								
rps18–rpl20								1

Table S3.3 (cont.)

rpl20–rps12						1			1
rps12–clpP									1
clpP intron					1	1	1		1
clpP intron					3	1	1		1
clpP–psbB					1	1	1		
psbB–psbT									
psbT–psbN	1	1	1	1					
psbN–psbH									
psbH–petB						1			
petB intron		1							
petB–petD									
petD intron									
petD–rpoA									
rpoA					1				
rpoA–rps11									1
rps11–rpl36							1		
rpl36–infA									
infA–rps8									
rps8–rpl14									
rpl14–rpl16						1	1		1
rpl16 intron					2		1		1
rpl16–rps3									
rps3–rpl22									
rpl22–rps19					1	1	1		1
rps19–rpl2						1			1
rpl2 intron									
rpl2–rpl23									
rpl23–trnI									
trnI–ycf2									
ycf2						1	1		
ycf2–trnL					1	1			
trnL–ndhB									1
ndhB intron	1	1	1	1					
ndhB–rps7									
rps7–rps12									
rps12 intron									
rps12–trnV	1	2	2	2	1		1		1
trnV–rrn16									
rrn16–trnI									
trnI intron						1	1		1
trnI–trnA									

Table S3.3 (cont.)

trnA intron					1		1	1	
trnA–rrn23									
rrn23–rrn4.5									
rrn4.5–rrn5	3	2	2	2					
rrn5–trnR				1					
trnR–trnN					1		2	1	
trnN–ycf1									
ycf1–ndhF									
ndhF–rpl32				2	1	2	1		
rpl32–trnL					2	2	2	1	
trnL–ccsA									
ccsA							1		
ccsA–ndhD									
ndhD–psaC									
psaC–ndhE									
ndhE–ndhG					1	1	1	1	
ndhG–ndhI					2	2	1	2	
ndhI–ndhA									
ndhA exon *1*					1			1	
ndhA intron	1	2	2	1					
ndhA–ndhH									
ndhH–rps15									
rps15–ycf1									
ycf1					2	2	2	2	

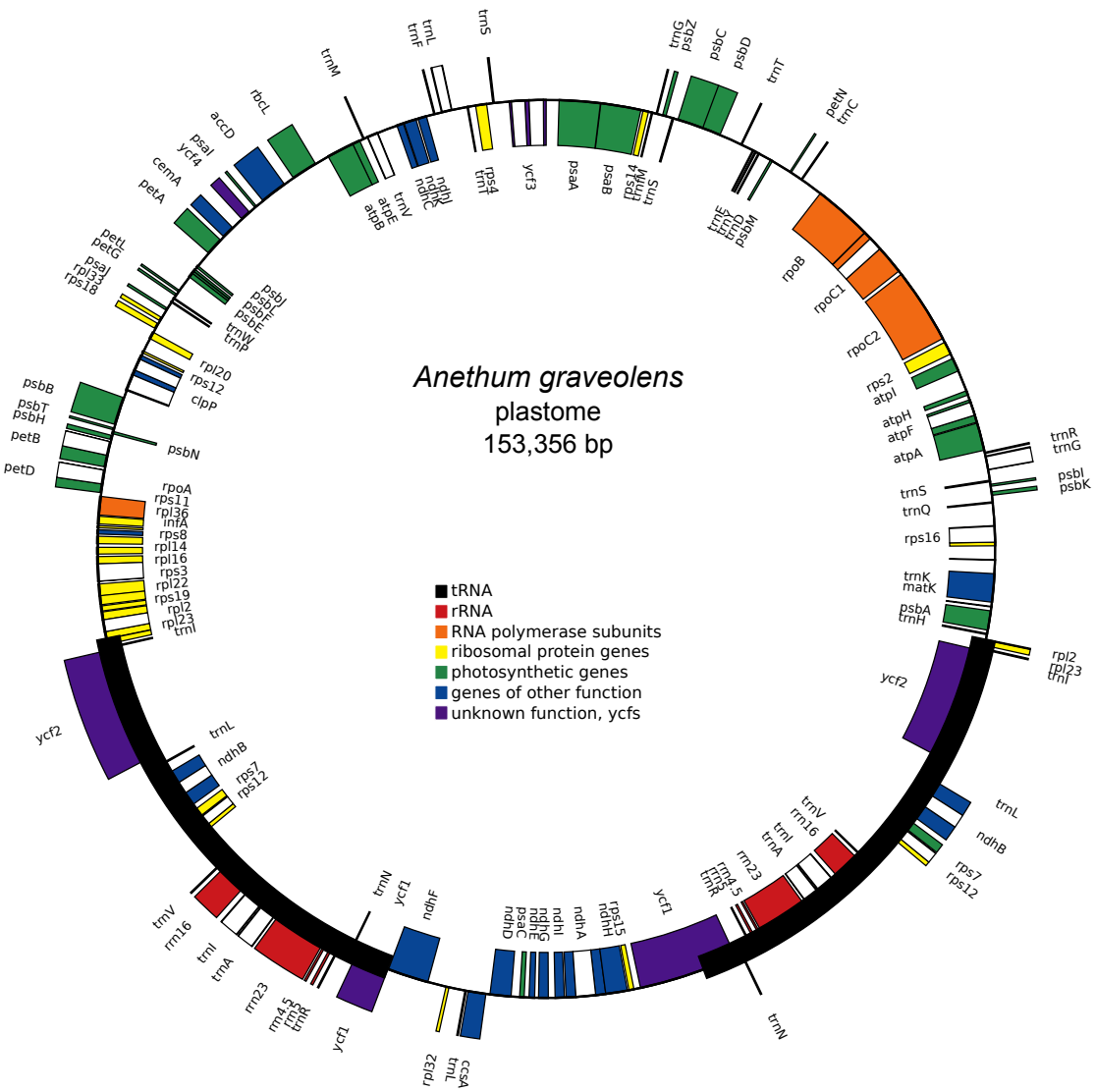


Fig. S3.1 Circular plastome map of *Anethum graveolens*. Genes are represented by boxes; those outside the circle are transcribed clockwise and those inside the circle are transcribed counterclockwise.

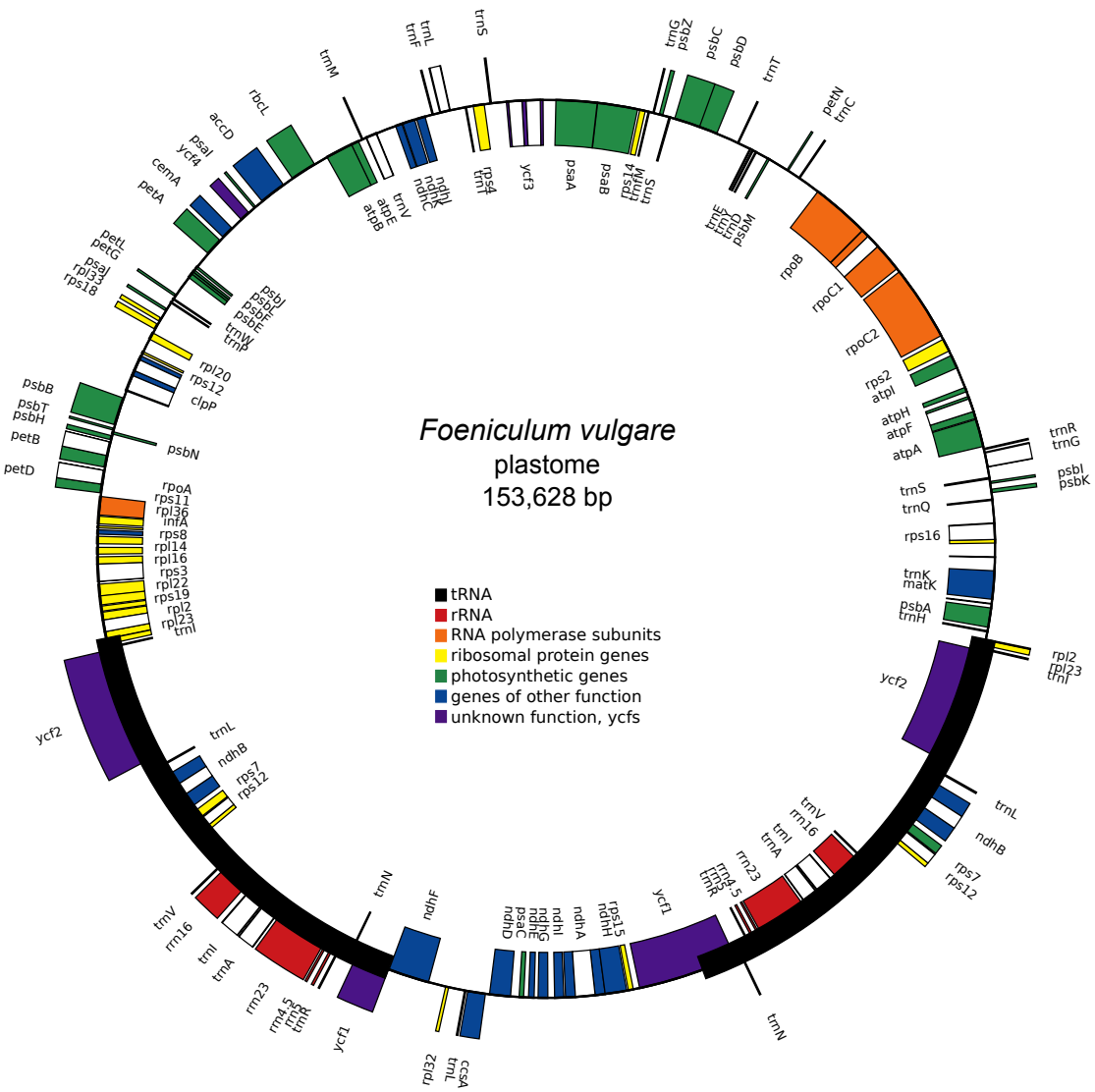


Fig. S3.2 Circular plastome map of *Foeniculum vulgare*. Genes are represented by boxes; those outside the circle are transcribed clockwise and those inside the circle are transcribed counterclockwise.

```
Foeniculum      AGCCCCGTATCAATGGGTGCCTTAATATGCATTATGCTATTCCGATTAGTCTTTCTTGGG      60
Anethum         ------------------------------------------------------------      0

Foeniculum      TTTACGATCAGATCCCATTTCGTGTTCATGAAAAACTAGTATCTTTCGGACATAGGCCAC      120
Anethum         ------------------------------------------------------------      0

Foeniculum      CCCCTTTATGGATGATAACGAGTACTTTTGGGAAAAAGTAGCGACAATCTATAAATTACC      180
Anethum         ----------------------------TGGGAAAAAGTAGCGACAATCTATAAATTACC      32
                                            ********************************

Foeniculum      CCTCTCGTATCTCGTAAAACACGAACAACCTAGAGAGAAGGGCGTGAATCTGTAGGCGGG      240
Anethum         CCTCTCGTATCTCGTAAAACACGAACAACCTAGAGAGAAGGGCGTGAATCTGTAGGCGGG      92
                ************************************************************

Foeniculum      GGAGACGACGTTAGGTTTTTCTGTATTTCAAGCAATGACTTCCTCCTTCATTACTTCATT      300
Anethum         GGAGACGACGTTAGGTTTTTCTGTATTTCAAGCAATGACTTCCTCCTTCATTACTTCATT      152
                ************************************************************

Foeniculum      CTTTTCATATACCTATGAAGGACTTTCACTCTCCTTTGTTCTCTTCTGTCTTTTTTTTTA      360
Anethum         CTTTTCATATACCTATGAAGGACTTTCACTCTCCTTTGTTCTCTTCTGTCTTTTTTTTTA      212
                ************************************************************

Foeniculum      CTTGGTTGGCAGGGTCAGGGCCTTTCTCGCTG     392
Anethum         CTTGGTTGGCAGGGTCAGGGCCTTTCTCGCTG     244
                ********************************
```

Fig. S3.3 CLUSTAL O(1.2.1) multiple sequence alignment (http://www.ebi.ac.uk/Tools /msa/clustalo/) of the novel J_{LA} fragment in *Foeniculum* and *Anethum*. *Foeniculum* is 148 bp longer, otherwise the fragments are identical (indicated with asterisks).

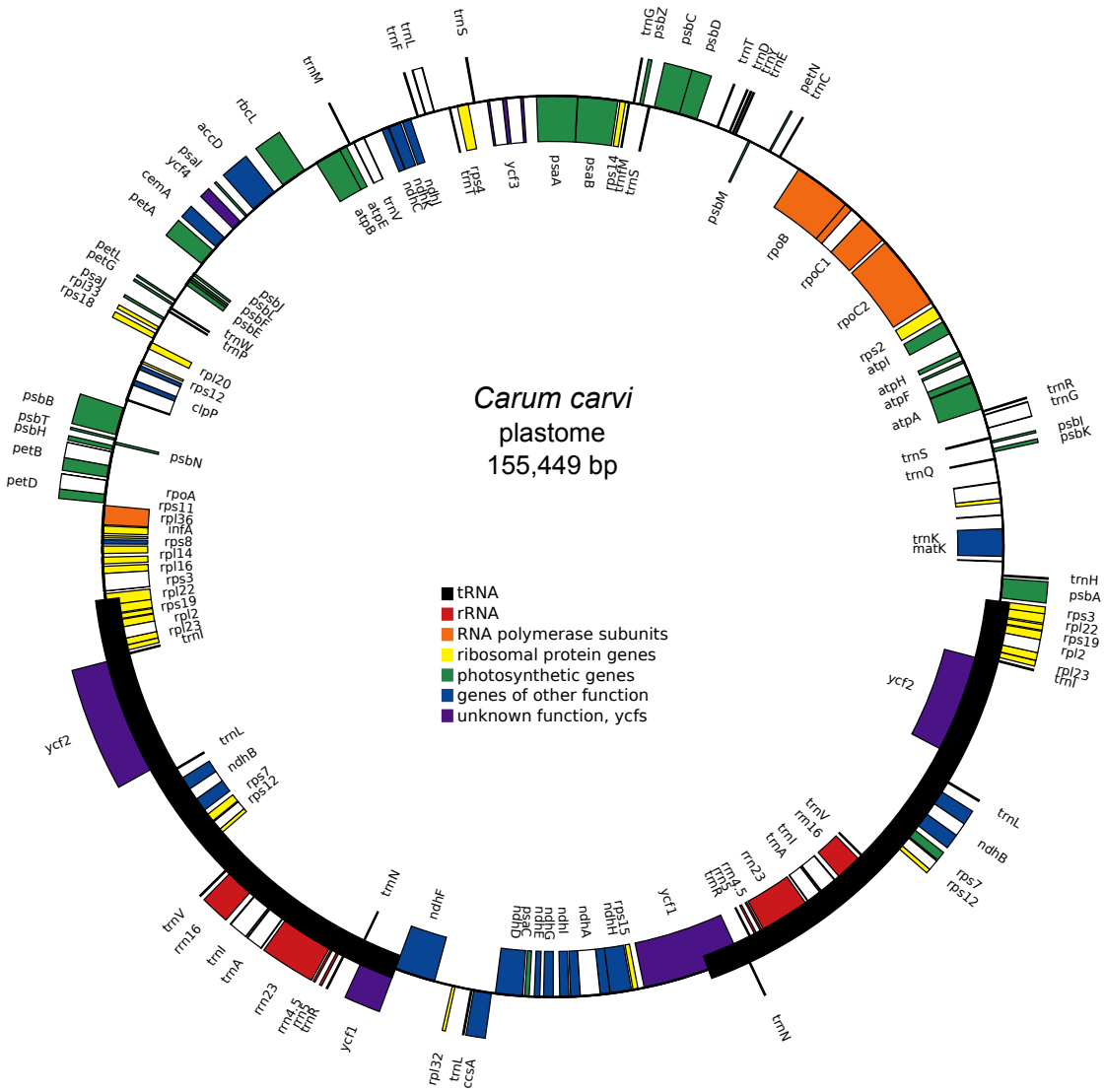


Fig. S3.4 Circular plastome map of *Carum carvi*. Genes are represented by boxes; those outside the circle are transcribed clockwise and those inside the circle are transcribed counterclockwise.

```
Naufraga_balearica          ------------------------------------------------------------      0
Opopanax_persicus           ------------------------------------------------------------      0
Pimpinella_major            ------------------------------------------------------------      0
Pimpinella_peregrina        ------------------------------------------------------------      0
Ridolfia_segetum            ATTCCCGTATCAATTATCAATGGGTGCCTTAATATGCATTATGCTATTCCGTTTAGTCTT      60
Stoibrax_dichotomum         ------------------------------------------------------------      0
Deverra_triradiata          ---------------------------------ATGCATTATGCTATTCCGATTAATCTT      27
Deverra_burchellii          ---------------------------------AAATGCATTATGCTATTCCGATTAATCTT      29
Haussknechtia_elymaitica    ------------------------------------------------------------      0
Cachrys_libanotis           ----------------------GGTGCCTTAATATGCATTATGCTATTCCGATTAGTCTT      38
Ferulago_nodosa             ------------------------------------------------------------      0
Azilia_eryngioides          ------------------------------------------------------------      0
Diplotaenia_cachrydifolia   ----------------------GGTGCCTTAATATGCATTATGCTATTCCGATTAGTCTT      38
Prangos_goniocarpa          ----------------------GGTGCCTTAATATGCATTATGCTATTCCGATTAGTCTT      38
Apium_graveolens            ------------------------------------------------------------      0
Petroselinum_crispum        ------------------------------------------------------------      0
Apium_prostratum            ------------------------------------------------------------      0
Conium_maculatum            ------------------------------------------------------------      0
Anethum_graveolens          ------------------------------------------------------------      0
Foeniculum_vulgare          -------AGCCCCGTATCAATGGGTGCCTTAATATGCATTATGCTATTCCGATTAGTCTT      53
Ammi_majus                  -------ATTCCCATATCAATGGGTGCCTTAATATGCATTATGCTATTCCGATTAGCCTT      53
Seseli_webbii               ------------------------------------------------------------      0
Sison_segetum               ------------------------------------------------------------      0


Naufraga_balearica          ------------------------------------------------------------      0
Opopanax_persicus           ------------------------------------------------------------      0
Pimpinella_major            ------------------------------------------------------------      0
Pimpinella_peregrina        ------------------------------------------------------------      0
Ridolfia_segetum            TCTTGGGTTTAGGATCAGATCCCATTTCGTGTTCATGAAAAACTAGTATCTTTCGGACAT      120
Stoibrax_dichotomum         ----------------------------------------------------------CA      2
Deverra_triradiata          TCTTGAATTTACGATCAGATCCCATTTCGTGTTCATGAAAAACTAGTATCTTTCGGACAT      87
Deverra_burchellii          TCTTGAATTTACGATCAGATCCCATTTCGTGTTCATGAAAAACTAGTATCTTTCGGACAT      89
Haussknechtia_elymaitica    ------------------------------------------------------------      0
Cachrys_libanotis           TCTTGGGTTTACGATCAGATCCCATTTCGTGTTCATGAAAAACGAGTATCTTTCGGACAT      98
Ferulago_nodosa             ------------------------------------------------------------      0
Azilia_eryngioides          ----------------------------------------------------TTAAT      5
Diplotaenia_cachrydifolia   TCTTGGGTTTACGATCAGATCCCATTTCGTGTTCATGAAAAACGAGTATCTTTCGGACAT      98
Prangos_goniocarpa          TCTTGGGTTTACGATCAGATCCCATTTCGTGTTCATGAAAAACGAGTATCTTTCGGACAT      98
Apium_graveolens            ------------------------------------------------------------      0
Petroselinum_crispum        ------------------------------------------------------------      0
Apium_prostratum            ------------------------------------------------------------      0
Conium_maculatum            ------------------------------------------------------------      0
Anethum_graveolens          ------------------------------------------------------------      0
Foeniculum_vulgare          TCTTGGGTTTACGATCAGATCCCATTTCGTGTTCATGAAAAACTAGTATCTTTCGGACAT      113
Ammi_majus                  TCTTGGGTTTACGATCAGATCCCATTTCGTGTTCATGAAAAACTAGTATCTTTCGGACAT      113
Seseli_webbii               ------------------------------------------------------------      0
Sison_segetum               ------------------------------------------------------------      0
```

Figure S3.5 (cont. on next page)

```
Naufraga_balearica          ------------------------------------------------------------      0
Opopanax_persicus           ------------------------------------------------------------      0
Pimpinella_major            ------------------------------------------------------------      0
Pimpinella_peregrina        ------------------------------------------------------------      0
Ridolfia_segetum            AG-GCCATCCCCCTCTATG--------------------------------------      138
Stoibrax_dichotomum         TTCGCCACCCCTCTCTATG--------------------------------------      21
Deverra_triradiata          TGGGGCACCCCCCTCTATG--------------------------------------      106
Deverra_burchellii          TGGGACACCCCCCTCTATG--------------------------------------      108
Haussknechtia_elymaitica    ------------------------------------------------------------      0
Cachrys_libanotis           ATTTAATATTGGCAGCGGGTGATACAACGGGGCCCGGAGGGAGTTCGCCCGATCCTTCGG      158
Ferulago_nodosa             ------------------------------------------------------------      0
Azilia_eryngioides          ATTTAATATTGTCAGCGGGTGATACAACGGGGCCCGGAGGGAGTTCGCCTGATCCTTCGG      65
Diplotaenia_cachrydifolia   ATTTAATATTGGCAGCGGGTGATACAACGGGGCCCGGAGGGAGTTCGCCTGATCCTTCGG      158
Prangos_goniocarpa          ATTTAATATTGGCAGCGGGTGATACAACGGGGCCCGGAGGGAGTTCGCCTGATCCTTCGG      158
Apium_graveolens            ------------------------------------------------------------      0
Petroselinum_crispum        ------------------------------------------------------------      0
Apium_prostratum            ------------------------------------------------------------      0
Conium_maculatum            ------------------------------------------------------------      0
Anethum_graveolens          ------------------------------------------------------------      0
Foeniculum_vulgare          AGGCCACCCC-CTTTATGGAT------------------------------------      133
Ammi_majus                  AGGCCACCACCCTCTATGGGT------------------------------------      134
Seseli_webbii               ------------------------------------------------------------      0
Sison_segetum               ------------------------------------------------------------      0


Naufraga_balearica          ------------------------------------------------------------      0
Opopanax_persicus           ---------------------------------------------------------TTT      3
Pimpinella_major            ------------------------------------------------------------      0
Pimpinella_peregrina        ------------------------------------------------------------      0
Ridolfia_segetum            ------------------------------------------GATGATAACG-AGTACTTT      156
Stoibrax_dichotomum         ------------------------------------------GATGATAACG-AGTACTTT      39
Deverra_triradiata          ------------------------------------------GATGATAACG-AAAACTTT      124
Deverra_burchellii          ------------------------------------------GATGATAACG-AAAACTTT      126
Haussknechtia_elymaitica    ------------------------------------------------------------      0
Cachrys_libanotis           AGGAAGGGCCTGTCTTTCCCTTATTGGCCAAAAACCATATGGATAATAATAAGCTCTTTT      218
Ferulago_nodosa             ------------------------------------------------------------      0
Azilia_eryngioides          GGGAAGGGCCTGTCTTTCCCTTATTGGCCAAAAACCATATGGATGATAATAAGTTCTTTT      125
Diplotaenia_cachrydifolia   GGGAAGGGCCTGTCTTTCCCTTATTGGCCAAAAACCATATGGATGATAATAAGCTATTTT      218
Prangos_goniocarpa          GGGAAGGGCCTGTCTTTCCCTTATTGGCCAAAAACCATATGGATGATAATAAGCTCTTTT      218
Apium_graveolens            ------------------------------------------------------------      0
Petroselinum_crispum        ------------------------------------------------------------      0
Apium_prostratum            ------------------------------------------------------------      0
Conium_maculatum            ------------------------------------------------------------      0
Anethum_graveolens          ------------------------------------------------------------      0
Foeniculum_vulgare          ----------------------------------------------GATAACGAGT-ACTTT      148
Ammi_majus                  ----------------------------------------------GATAACGAGT-ACTTT      149
Seseli_webbii               ------------------------------------------------------------      0
Sison_segetum               ------------------------------------------------------------      0
```

Figure S3.5 (cont. on next page)

```
Naufraga_balearica            --------------------TTATAAATTTCAAATAACCCCTCTCATAAAACACGAATAA      40
Opopanax_persicus             TGAGATAAAGTAGCGACAA--------TCAAAAATGACCCCTATCGTAAAACACGGGCAA      55
Pimpinella_major              ------------------------------------------------------------      0
Pimpinella_peregrina          ------------------------------------------------------------      0
Ridolfia_segetum              TGGGAAAAAATAGCGACAATCTATAAATTACCCCTCTCGTATCTCGTAAAACACGAACAA      216
Stoibrax_dichotomum           TGGGATCAAGTAGTGACAA--------TTACACCTCT-------CATAAAACACGAACAA      84
Deverra_triradiata            TGGGATAAAGTAGCGACAATCTTTAAATTAACCCTCT-------CGTAAAACACGAACAA      177
Deverra_burchellii            TGGGATAAAGTAGCGACAATCTTTAAATTAACCCTCT-------CGTAAAACACGAACAA      179
Haussknechtia_elymaitica      ------------------------------------------------------------      0
Cachrys_libanotis             GGTAACAA-------AGTAGCGACAGTCGAAAAATTACCCCTCTCGTAAAACACGGGCAA      271
Ferulago_nodosa               ------AA-------AGTAGCGACGGTCTAAAAATTACCCCTCTCGTAAAACACGGGCAA      47
Azilia_eryngioides            GAG-ATAA-------AGTAGCGACAATCGAAAAATTACCCCTCTCGTAAAACACGGGCAA      177
Diplotaenia_cachrydifolia     GGG-AAAA-------AGTAGCGACAGTCGAAAAATTACCCCTCTCATAAAACACGGGCAA      270
Prangos_goniocarpa            GAG-ATAA-------AGTAGCGACAGTCGAAAAATTACCCCTCTCGTAAAACACGGGCAA      270
Apium_graveolens              ----------------------------------TCTCGTAAAACTAAAACACGAACAA      25
Petroselinum_crispum          ----------------------------------TCTCGTAAAACTAAAACACGAACAA      25
Apium_prostratum              ----------------------------------TCTCGTAAAACTAAAACACGAACAA      25
Conium_maculatum              --------------------CGACAATCGAAAAATAGCCCCTCTCGTAAAACACGGGCAA      40
Anethum_graveolens            TGGGAAAAAGTAGCGACAATCTATAAATTACCCCTCTCGTATCTCGTAAAACACGAACAA      60
Foeniculum_vulgare            TGGGAAAAAGTAGCGACAATCTATAAATTACCCCTCTCGTATCTCGTAAAACACGAACAA      208
Ammi_majus                    TGGGACAACGTAGCGACAATCTAAAAATTACCCCTCT-------CGTAAAACACGAACAA      202
Seseli_webbii                 ------------------------------------------------------------      0
Sison_segetum                 ------------------------------------------------------------      0


Naufraga_balearica            CCTAGAGAAAAGGGTATGAATCTGGAGGCAGGGGAGACGAGGTTAGGTTTTTCTGTATTT      100
Opopanax_persicus             CCTAGAGAGAAGGGCGTGAATCTGGAGGCGGGGGAAACGACGTTAGGTTTTTCTGTATTT      115
Pimpinella_major              -----------------GTATCGGGAGGCGGGGGAGACGACGTTAGGTTTTTCCGTATTT      43
Pimpinella_peregrina          ------------------------------------------------------------      0
Ridolfia_segetum              CCTAGAGAGAAGGGCTTGAATCTGTAGGCGGGGGAGACGACGTTAGGTTTTTCTGTATTT      276
Stoibrax_dichotomum           CGTAGAGAGAAAGGCGTGAATTTGGAAACG-------------------------------      114
Deverra_triradiata            CCTAGAGAGAGGGGCGTGAATCTGGAGGCGGGGGAGACGACGTTAGGTTTTTCTGTATTT      237
Deverra_burchellii            CCTAGAGAGAGGGGCGTGAATCTGGAGGCGGGGGAGACGACGTTAGGTTTTTCTGTATTT      239
Haussknechtia_elymaitica      ------------------------------------------------------------      0
Cachrys_libanotis             CCTAGAGAGAAGGGCGTGAATCTGGAGGCGGGGGAGACGACGTTAGGTTTTTATGTATTT      331
Ferulago_nodosa               CCTAGAGAGAAGGGCGTGAATCTGGAGGCGGGGGAGACGACGTTAGGTTTTTATGTATTT      107
Azilia_eryngioides            CCTAGAGAGAAGAGCGTGAATCTGGAGGCGGGGGAGACGCCGTTAGGTTTTTATGTATTT      237
Diplotaenia_cachrydifolia     CCTAGAGAGAAGGGCGTGAATCTGGAGGCGGGGGAGACGACGTTAGGTTTTTATGTATTT      330
Prangos_goniocarpa            CCTAGAGAGAAGGGCGTGAATCTGGAGGCGGGGGAGACGACGTTAGGTTTTTATGTATTT      330
Apium_graveolens              CCTAGAGAGAAGGGCATGAATCTGGAGGCAGGGGAGACGAGGTTAGGTTTTTATGTATTT      85
Petroselinum_crispum          CCTAGAGAGAAGGGCATGAATCTGGAGGCAGGGGAGACGAGGTTAGGTTTTTATGTATTT      85
Apium_prostratum              CCTAGAGAGAAGGGCATGAATCTGGAGGCAGGGGAGACGAGGTTAGGTTTTTATGTATTT      85
Conium_maculatum              CCTAGACAGAAGGGCGTGAATCTGGAGGCGGGGGAAATGACGTTAGGTTTTTCTGTATTT      100
Anethum_graveolens            CCTAGAGAGAAGGGCGTGAATCTGTAGGCGGGGGAGACGACGTTAGGTTTTTCTGTATTT      120
Foeniculum_vulgare            CCTAGAGAGAAGGGCGTGAATCTGTAGGCGGGGGAGACGACGTTAGGTTTTTCTGTATTT      268
Ammi_majus                    CCTAGAGAGAAGGGCATGAATCTGAAGGCGGGGGAGACGACGTTAGGTTTTTCTGTATTT      262
Seseli_webbii                 ------------------------------------------------------------TTT      3
Sison_segetum                 ------------------------------------------------------------TTT      3
```

Figure S3.5 (cont. on next page)

```
Naufraga_balearica              CGGGGCAATGATTTCCTCCTTCATTACTTCATTCTTTTCAATATACCTATGAAGGACTTT      160
Opopanax_persicus               CAAGCAA-TGACTTACTCCTTCATTTCTTAATTC-TTTCCATATACCTATGAAGGACTTT      173
Pimpinella_major                CAAGCAA-TGACTTCCTCCTTCATTTCTTCATTC-TTTCCATATACCTATGAAGGGCTTT      101
Pimpinella_peregrina            ------------------------------------------------------------      0
Ridolfia_segetum                CAAGCAA-TGACTTCCTCCTTCATT--------C-TTTTCATATACCTATGAAGGACTTT      326
Stoibrax_dichotomum             -------------------------------------------GAGGAGACAAGGACTTT      131
Deverra_triradiata              CAAGCAA-TGACTTCCTCCTCCATTACTTCATTC-TTTTCATATACCTATGAAGGACTTT      295
Deverra_burchellii              CAAGCAA-TGACTTCCTCCTCCATTACTTCATTC-TTTTCATATACCTATGAAGGACTTT      297
Haussknechtia_elymaitica        ------------------------------------------------------------      0
Cachrys_libanotis               CAAGCCA-TGACTCCCTCCTTC--------ATTC-TTTCCATATACCTATGAAGGACTTT      381
Ferulago_nodosa                 CAAGCAA-TGACTCCCTCCTGC--------ATTC-TTTTCATATACCTATGAAGGACTTT      157
Azilia_eryngioides              CAAGCAA-TGACTTCCTCCTTC--------ATTC-TTTCCATATACCTATGAAGGACTTT      287
Diplotaenia_cachrydifolia       CAAGCAA-TGACTCCCTCCTTC--------ATTC-TTTCCATATACCTATGAAGGACTTT      380
Prangos_goniocarpa              CAAGCAA-TGACTCCCTCCTTC--------ATTC-TTTCCATATACCTATGAAGGACTTT      380
Apium_graveolens                CAGGCAA-TGATTTCCTCCTTCATTACTTCATTA-TTTTCATATACCTATGAAGGACTTT      143
Petroselinum_crispum            CAGGCAA-TGATTTCCTCCTTCATTACTTCATTA-TTTTCATATACCTATGAAGGACTTT      143
Apium_prostratum                CAGGCAA-TGATTTCCTCCTTCATTTCTTCATTA-TTTTCATATACCTATGAAGGACTTT      143
Conium_maculatum                CAAGCAA-TGACTTCCTCCTTCATTTCTTCATTC-TTTCCATATACCTATGAAGGACTTT      158
Anethum_graveolens              CAAGCAA-TGACTTCCTCCTTCATTACTTCATTC-TTTTCATATACCTATGAAGGACTTT      178
Foeniculum_vulgare              CAAGCAA-TGACTTCCTCCTTCATTACTTCATTC-TTTTCATATACCTATGAAGGACTTT      326
Ammi_majus                      CAAGCAA-TGACTTCCTCCTTCATTACTTCATTC-TTTTCATATACCTATGAAAGACTTT      320
Seseli_webbii                   CAAGCAA-TGACTTCCTCCTTCATTACTTCATTC-TTTTCATATACCTATGAAAGACTTT      61
Sison_segetum                   CAAGCAA-TGACTTCCTCCTTCATTACTTCATTC-TTTTCATATACCTATGAAAGACTTT      61


Naufraga_balearica              CACTCTCCTTTGTTCTCTTCTGTCTTTTTTTTTT----CCTTGGTTGGCAGGATCGGGTC      216
Opopanax_persicus               CACTCTCCTTTGTTCTCTTCTGTCTTTTTTCTTT--TGGCAGGGTTGGCAGGGTCAGGGC      231
Pimpinella_major                CACTCTCCTTTGTTCTCTTTTATCTTT---------TGACTTCGTTGGCAGGGTCAGGGC      152
Pimpinella_peregrina            -ACTCTCCTTTGTTCTCTTTTATCTTT---------TGACTTCGTTGGCAGGGTCAGGGC      50
Ridolfia_segetum                CACTCTCCTTTGTTCTCTTCTGTCTTTTTTT------TACTTGGTTGGCAGGGTCAGGGC      380
Stoibrax_dichotomum             CACTCTCCTTTGTTCTCTTCTGTCTGTTTTCTTT--TGACTTGGTTGGCAGGATCAGGGC      189
Deverra_triradiata              CACTCTCCTTTGTTCTCTTCTGTCTTTTTTCTTT--TGACTTGGTTGGCAGGGTCAGGGC      353
Deverra_burchellii              CACTCTCCTTTGTTCTCTTCTGTCTTTTTTCTTT--TGACTTGGTTGGCAGGGTCAGGGC      355
Haussknechtia_elymaitica        ----------------------------TCTTT--TGACTTGGTTGGCAGGGTCAGGGC      29
Cachrys_libanotis               CACTCTCCTTTGTTCTCTTCTGTCTTTTTTTTTC-----TTTGGTTGGCAGGGTCAGGGC      436
Ferulago_nodosa                 CACTCTCCTTTGTTCTCTTCTGTCTTTTTTC--------TTTGGTTGGCAGGGTCAGAGC      209
Azilia_eryngioides              CACTCTCCTTTGTTCTCTTCTGTCTTTTTTC--------TTTGGTTGGCAGGGTCAGGGC      339
Diplotaenia_cachrydifolia       CACTCTCCTTTGTTCTCTTCTGTCTTTTTTC--------TTTGGTTGGCAGGGTCAAGGC      432
Prangos_goniocarpa              AACTCTCCTTTGTTATCTTCTGTCTTTTTTC--------TTTGGTTGGCAGGGTCAGGGC      432
Apium_graveolens                CACTCTCCTTTGTTCTCTTCTGTCTTTTTTTTTTTTTTTACTTTGTTGGCAGGATCAGGTC      203
Petroselinum_crispum            CACTCTCCTTTGTTCTCTTCTGTCTTTTTTTTTTTTTTTACTTTGTTGGCAGGATCAGGTC      203
Apium_prostratum                CACTCTCCTTTGTTCTCTTCTGTTTTTTT--TTTTTTTACTTTGTTGGCAGGATCAGGTC      201
Conium_maculatum                CACTCTCCTTTGTTCTCTTCTGTCTTTTTTTTTTTCTTTGG----TTGGCAGGGTCAGGGC      214
Anethum_graveolens              CACTCTCCTTTGTTCTCTTCTGTCTTTTTTT-----TTACTTGGTTGGCAGGGTCAGGGC      233
Foeniculum_vulgare              CACTCTCCTTTGTTCTCTTCTGTCTTTTTTT-----TTACTTGGTTGGCAGGGTCAGGGC      381
Ammi_majus                      CACTCTACTTTGTTCTCTTCTGTCTTTTTTTTTTTTTTTACTTGGTTGGCAGGGTCAGGGC      380
Seseli_webbii                   CACTCTCCTTTGTTCTTTC-----TTTTTTTTTTTTTTTACTTGGTTGGCAGGGTCAGGGC      116
Sison_segetum                   CACTCTCCTTTGTTCTTTC-----TTTTTTTTTTTTTTTACTTGGTTGGCAGGGTCAGGGC      116
                                                                        ******** **    *
```

Figure S3.5 (cont. on next page)

```
Naufraga_balearica            CTTTCTCGCTG   227
Opopanax_persicus             CTTTCTCGCTG   242
Pimpinella_major              CTTTCTCGCTG   163
Pimpinella_peregrina          CTTTCTCGCTG   61
Ridolfia_segetum              CTTTCTCGCTG   391
Stoibrax_dichotomum           CTTTCTCGCTG   200
Deverra_triradiata            CTTTCTCGCTG   364
Deverra_burchellii            CTTTCTCGCTG   366
Haussknechtia_elymaitica      CTTTCTCGCTG   40
Cachrys_libanotis             CTTTCTCGCTG   447
Ferulago_nodosa               CTTTCTCGCTG   220
Azilia_eryngioides            CTTTCTCGCTG   350
Diplotaenia_cachrydifolia     CTTTCTCGCTG   443
Prangos_goniocarpa            CTTTCTCGCTG   443
Apium_graveolens              CTTTCTCGCTG   214
Petroselinum_crispum          CTTTCTCGCTG   214
Apium_prostratum              CTTTCTCGCTG   212
Conium_maculatum              CTCTCTCGCTG   225
Anethum_graveolens            CTTTCTCGCTG   244
Foeniculum_vulgare            CTTTCTCGCTG   392
Ammi_majus                    CTTTCTCGCTG   391
Seseli_webbii                 CTTTCTCGCTG   127
Sison_segetum                 CTTTCTCGCTG   127
                              ** ********
```

Fig. S3.5 CLUSTAL O(1.2.1) multiple sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalo/) of the novel J_{LA} fragment in all species with a *rpl2* IR boundary except *Oedibasis platycarpa*, which had no similarity to the other taxa. Identical bases are indicated below the alignment with asterisks.

CHAPTER 4: THE PHYLOGENETIC UTILITY OF PLASTOME RARE GENOMIC CHANGES, PLASTID GENE REGIONS *PSBM–PSBD* AND *PSBA–TRNH*, AND NUCLEAR GENE *PHYA* IN RESOLVING RELATIONSHIPS WITHIN THE APIOID SUPERCLADE OF APIACEAE SUBFAMILY APIOIDEAE

Abstract

Relationships among the 14 tribes and other major clades comprising the apioid superclade of Apioideae (Apiaceae) are unclear, with previous studies of primarily nrDNA ITS sequence data resolving either a large polytomy or poorly supported clades. In an effort to better elucidate higher-level relationships within the group and to determine the phylogenetic utility and limitations of the four rare genomic changes (RGCs) detected in previous studies, the plastid regions *psbM–psbD and psbA–trnH* and the nuclear gene *PHYA* were sequenced. These loci were analyzed separately and in combination with previously available ITS data and the four RGCs. Maximum likelihood and Bayesian analyses of partitioned and variously combined data matrices yielded largely consistent results, with resolution of some higher-level relationships achieved. The *psbA–trnH* region does not contain enough parsimony informative characters and did not yield any resolution of higher-level relationships. *PHYA* was also uninformative at the tribal level, but did add resolution at some lower taxonomic levels. The *psbM–psbD* region provided the strongest support for relationships among major lineages. Results of Bayesian analysis of combined ITS and *psbM–psbD* data recovered the most tribes and other major clades and resolved the most intertribal relationships. These data supported the monophyly of tribes Apieae, Careae, Echinophoreae, Pimpinelleae, Selineae, and Tordylieae and the *Cachrys* and *Sinodielsia* clades. Tribe Pyramidoptereae was resolved as paraphyletic, with Careae arising from within. The two examined species of the *Opopanax* clade also did not comprise a monophyletic group. Tribe Coriandreae is monophyletic upon the exclusion of *Bifora testiculata*. RGCs did not improve resolution when analyzed with UPGMA or when included as a partition in a matrix with combined sequence data. When RGCs were mapped onto the phylogeny, some

are homoplasious while others provide support for recovered topologies. The inversion of *psbA* and *trnH*, mitochondrial DNA at the large single copy – inverted repeat boundary, and boundary types B and D are all RGCs that support intertribal relationships. The other boundary types (A, D'-I') and the presence of filler DNA at the large single copy – inverted repeat boundary have occurred independently multiple times. Most of the uninformative RGCs occur within tribe Selineae and subtribe Tordyliinae, which also have low overall intratribal resolution. While some intertribal relationships are resolved by these data, further study of the apioid superclade is necessary to resolve all relationships and produce a stable classification of its major lineages.

Introduction

The plant family Apiaceae (or Umbelliferae) contains many economically, medicinally, and ecologically important species, such as carrot, caraway, coriander, dill, fennel, and parsnip, as well as highly toxic plants such as poison hemlock. Apiaceae are a large family, with over 400 genera and 3,200 species recognized (The Plant List 2013). The largest of its four subfamilies, Apioideae, contains 41 major clades, many of which are recognized at the rank of tribe (Downie et al. 2010). Within subfamily Apioideae is a large, morphologically heterogeneous group of umbellifers comprising 14 tribes and other major clades of dubious relationship referred to as the apioid superclade (Plunkett and Downie 1999, 2000). These lineages include tribes Apieae, Careae, Coriandreae, Echinophoreae, Pimpinelleae, Pyramidoptereae, Selineae, and Tordylieae (including three subclades) and the *Cachrys*, *Conium*, *Opopanax, and Sinodielsia* clades (Fig. 4.1; Downie et al. 2010).

Although Apioideae phylogenetics has received much attention (e.g., Downie et al. 1996, 1998, 2000, 2001, 2010; Downie and Katz-Downie 1996, 1999; Katz-Downie et al. 1999; Zhou et al. 2008, 2009; Magee et al. 2010), studies focused explicitly on resolving the higher-level relationships of the apioid superclade have been few (e.g., Downie et al. 2000). The first plastid gene used to infer Apiaceae phylogenetic history was *matK* and the resultant gene tree supported the apioid superclade as a monophyletic group, although sampling was limited (Plunkett et al. 1996; Plunkett and Downie 1999). Downie et al. (1996, 1998, 2000) considered introns from plastid genes *rpoC1*, *rpl16*, and *rps16*, and while each study recovered a strongly supported apioid superclade, the relationships among its constituent major clades were either not resolved or if resolved not well supported despite an ever-increasing taxon sampling. A study of chloroplast DNA (cpDNA) restriction site data also failed to show robust relationships within the group, although the frequency and large size of the inverted repeat (IR) junction shifts detected showed great promise in circumscribing major clades (Plunkett and Downie 1999,

2000). To date, nrDNA internal transcribed spacer (ITS) sequences comprise the most comprehensive database for Apioideae phylogenetic study. Greatest resolution of higher-level relationships within the apioid superclade was obtained by Zhou et al. (2008, 2009) in their studies of Chinese Apioideae based on ITS and cpDNA *rpl16* and *rps16* intron sequences. Well-resolved phylogenies are critical in addressing patterns and processes of evolution and, to date, resolution of relationships among the tribes and other major clades comprising the apioid superclade remains poor.

Previously used molecular markers are either too conserved (plastid gene and intron sequences) or have a high mutation rate causing saturation (ITS) and are, therefore, unable to adequately resolve taxonomic relationships among apioid superclade lineages. Resolving such relationships requires additional data to increase the number of informative characters, as well as markers that can unambiguously define monophyletic groups, such as plastome rare genomic changes (RGCs; Downie and Palmer 1992; Plunkett and Downie 1999, 2000; Rokas and Holland 2000; Raubeson and Jansen 2005).

In this chapter I examine the utility of two plastid DNA regions (*psbM–psbD* and *psbA–trnH*) and a single copy nuclear gene (*PHYA*), loci that have not previously been used in Apiaceae phylogenetic study, to resolve the higher-level phylogenetic relationships of the apioid superclade. I also consider the plastome RGCs identified in earlier studies as additional markers (Chapter 3). The plastid *psbM–psbD* (*psbMD*) region includes the tRNA genes *trnD*, *trnY*, *trnE*, and *trnT*. This locus was deemed highly variable by Shaw et al. (2005, 2007); furthermore, Downie and Jansen (2015) identified it as the most variable region in their comparison of five Apiales plastomes. Likewise, the non-coding region between *psbA* and *trnH* is also highly variable and has been proposed as a potential barcoding locus in Apiaceae and other angiosperm families (CBOL Plant Working Group 2009; Liu et al. 2014). The nuclear single copy gene *PHYA* is one of several genes in the phytochrome gene family (Mathews and Sharrock

1997; Mathews 2010). To date, *PHYA* has been used to resolve phylogenies in Orobanchaceae (Bennett and Mathews 2006), Brassicaceae (Beilstein et al. 2008), and Magnoliaceae (Nie et al. 2008). The plastome RGCs identified in earlier studies of the apioid superclade (Chapter 3) include gene synteny changes at single copy–IR boundaries, inversions, and insertions of novel DNA through intracellular gene transfer (IGT).

Resolution of evolutionary relationships among the major lineages comprising the apioid superclade is the last major problem of Apiaceae higher-level systematics, but work to date has been thwarted because the molecular markers that have been used are too conserved to discern relationships. The major aim of this paper is to assess the phylogenetic utility of new plastid, nuclear, and RGC markers to resolve these relationships and to use a combined DNA sequence analysis approach to understand the distribution of RGCs within the group.

Methods

Markers

The RGCs matrix was constructed using four plastome structural characters: 1) a 571 bp inversion between *psbM* and *trnT*, resulting in the inversion of genes *trnD-trnY-trnE*; 2) a 2178 bp inversion from the large single copy (LSC)–IR boundary to the *3' trnK* exon, resulting in the inversion of genes *trnH* and *psbA*; 3) gene adjacency changes at the plastid LSC–IR boundary; and 4) the presence of novel DNA between the LSC–IR boundary and 3' *trnH*. The ancestral gene synteny for *psbM* through *trnT* was scored as 0 and the inversion of *trnD-trnY-trnE* was scored as 1. Similarly, the ancestral gene synteny of *trnH* adjacent to the IR in the LSC region followed by *psbA* then 3' *trnK* was scored as 0 (no inversion) and the inversion placing *psbA* adjacent to the IR was scored as 1. Twelve different gene adjacencies have been detected at J_{LA} (Chapter 1; Fig. 1.1) in species of the apioid superclade (Chapter 3; Fig. 3.1). These gene adjacency data were scored as 12 binary characters (as opposed to one character having 12

states). For each state, the presence of a specific gene adjacency was indicated by 1 and its absence by 0, such that for each taxon 11 of the characters would be coded as 0 and one would be coded as 1. The two different novel DNA insertions at J_{LA} (Chapter 3) were each scored separately. The absence of putative mitochondrial DNA (mtDNA) at J_{LA}, the ancestral state, was scored as 0 and its presence was scored as 1. The absence of "filler DNA" from double-strand break repair at J_{LA}, the ancestral state, was scored as 0 and its presence was scored as 1.

Taxon Sampling

Species were chosen because of their inclusion in previous phylogenetic studies of Apiaceae subfamily Apioideae. Whenever possible, new data were obtained from precisely the same accessions as used previously (Table S4.1). If PCR amplification failed using standard or high-fidelity polymerases, or if DNA from a given accession was unavailable, then alternative accessions of that species or alternative species were substituted. I sampled from all of the 14 major clades comprising the apioid superclade (Fig. 4.1), including the three subclades of tribe Tordylieae (i.e., subtribe Tordyliinae and the *Cymbocarpum* and *Lefebvrea* clades). The list of genera comprising each of these tribes/major clades provided in Downie et al. (2010) was used as a sampling guide, although that list is not comprehensive because it does not include taxa not yet considered in molecular phylogenetic study. In total, 143 ingroup species representing 123 genera were considered herein (Table S4.1). As outgroups, I chose 11 species from subfamily Apioideae outside of the apioid superclade but closely related to it based on previous higher-level studies of the subfamily. These include representatives from Scandiceae, Oenantheae, and the *Acronema* clade (Downie et al. 2010).

PCR, Cloning, and Sequencing

Three primer pairs and one internal primer were used for PCR and sequencing of *PHYA*, one primer pair was used for both PCR and sequencing of *psbA–trnH*, and several primer pairs and interal primers were used for PCR and sequencing of *psbMD* (Table 4.1). The two plastid markers were amplified with GoTaq polymerase® (Promega) in a volume of 25 μl with the following component concentrations: 1X GoTaq buffer, 2.5 mM $MgCl_2$, 200 μM each dNTP, 0.5 μM of each primer, 1 U polymerase, and 0.5 μL of unquantified genomic DNA. Thermal cycler conditions for *psbA–trnH* are as follows: initial denaturation of 1 min at 94˚C; 29 cycles of 94˚C for 30 sec, 55˚C for 30 sec, and 72˚C for 2 min; and a final extension of 10 min at 72˚C. If this initial reaction failed amplification was attempted a second time with Phusion High-Fidelity DNA Polymerase (Life Technologies). Phusion reactions had the following component concentrations: 1X HF Phusion buffer, 200 μM each dNTP, 0.5 μM each primer, 3% DMSO, 0.625 U polymerase, and 0.5 μL of unquantified genomic DNA. Thermal cycler conditions for the Phusion polymerase are an initial denaturation of 1 min at 98˚C, followed by 34 cycles of 10 sec at 98˚C, 30 sec at 50.7˚C, and 1 min at 72˚C, and a final extenstion of 10 min at 72˚C. If neither of these amplification conditions produced a product an alternative accession was chosen and the same protocols applied. Thermal cycler conditions followed those of Shaw et al. (2007) for amplification of *psbMD*. For a few accessions, all attempts at amplification of the *psbMD* region failed. Alternative primer combinations were used to amplify shorter fragments and to account for the inversion of *trnD-trnY-trnE*. If these also failed alternative accessions of the same species were tried.

All plastid PCR products were cleaned using the ExoSAP method (Bell 2008) modified by using 5 U of Exonuclease I (New England Biolabs) and 2.5 U of Antarctic Phosphatase (New England Biolabs). Sequencing reactions were performed using the ABI Prism® BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems) in 10 μl volumes as follows: 1X BigDye buffer, 1.25% glycerol, 1 μM primer, 0.4 μl of BigDye, and 75-100 ng of template DNA.

These reactions were carried out at 98˚C for initial denaturation, followed by 34 cycles of 94˚C for 15 sec, 45˚C for 15 sec, and 60˚C for 4 min. Sequencing was done at the University of Illinois W.M. Keck Center for Comparative and Functional Genomics.

The nuclear gene *PHYA* was amplified and cloned following the protocol of Mathews et al. (1995). To develop specific primers for the apioid superclade PCR products were cloned using the TOPO TA cloning kit and TOP10® competent cells (Invitrogen). A total of five clones each from 10 accessions were sequenced. The redesigned primers were then used to amplify all remaining accessions (Table 4.1). Single band PCR products were never generated, therefore all products were run in 2% TAE gels and bands of the correct size were excised and cleaned using the QIAEX II® Gel Extraction kit (QIAGEN®) prior to direct sequencing. Sequencing was performed as previously described.

Sequences were edited and assembled into contigs using Sequencher v. 5.1 (GeneCodes, Ann Arbor, Michigan, USA) by trimming reads automatically with base confidence values set to 20 and all other parameters set at default. Contigs were assembled using a minimum percent match of 85 and a minimum overlap of 35; algorithm and other parameters were set to default. These contigs were then examined by eye to resolve discrepancies among reads and to ensure each base had minimum Phred score of Q40 if only a single read covered the region or a Q20 or above if there were at least two reads that had no mismatches. If these minimum quality levels were not met additional sequencing was done to improve confidence in base calls for those nucleotides.

Alignment and Phylogenetic Analyses

Contigs were exported from Sequencher as consensus sequences and were aligned with MUSCLE v. 3.8.31 (Edgar 2004) and the OPAL package (Wheeler and Kececioglu 2007; Wheeler and Maddison 2012) within Mesquite v. 3.01 (build 658; Maddison and Maddison

2014). The *PHYA* fragment consisted of two exons and its intervening intron region, the latter excluded from subsequent analysis due to difficulty with alignment at the generic level. The alignment of exon data was kept in frame using EST data from *Petroselinum crispum* (parsley; GenBank accession X75412; Poppe et al. 1994), a member of tribe Apieae.

Each locus was aligned separately and partitioning schemes of individual genes and combined matrices were tested using PartitionFinder v. 1.0.1 (Lanfear et al. 2012, Lanfear et al. 2014). PartitionFinder uses alignments as inputs and simultaneously discriminates amongst several user-defined partitions of the data, called schemes, to find the best evolutionary model and partitioning scheme based on the Akaike information criterion (AIC). The partitions examined were as follows: 1) each gene region separately with no partitioning within a region; 2) all combinations of genes together with no partitioning within a region (for a total of seven schemes); 3) each gene region and any potential within gene partitions separate (i.e., codons and coding/non-coding DNA); and 4) all coding sequences partitioned separately from all non-coding sequences. Coding DNA within *psbMD* includes the genes *trnD*, *trnY*, *trnE*, and *trnT,* and coding DNA within the *psbA–trnH* region includes 3' *psbA* and *trnH*. Coding regions for the plastid genes were not partitioned by codon position because only 101 bp of *psbA* and 38 bp of 5' *psbD* were sequenced. PartItionFinder, through the AIC, supports the scheme with the best likelihood as: ITS, *PHYA* codon positions, *psbA–trnH* coding sequence and non-coding sequence separate, and coding and non-coding sequence of *psbMD* separate. This partitioning scheme was used for all analyses.

Nine matrices were constructed and analyzed with maximum likelihood (RAxML; Stamatakis 2014) and Bayesian (MrBayes 3; Ronquist and Huelsenbeck 2003) inference methods (Table 4.2). These include: 1) RGCs; 2) ITS; 3) *PHYA*; 4) *psbMD*; 5) *psbA–trnH*; 6) ITS + *psbMD*; 7) ITS + *PHYA* + *psbMD* for 63 taxa; 8) ITS + *PHYA* + *psbMD* + RGCs; and 9) ITS + *PHYA* + *psbMD* for 132 taxa. The single gene matrices contained 124 (ITS), 109 (*psbMD*), 86

(*PHYA*), and 67 (*psbA–trnH*) taxa (Table 4.2; Fig. 4.2). To be included in matrices 6-8 a taxon had to have all gene regions sequenced; in matrix 9, however, each taxon was only required to have data from two of the three gene regions. Thus, all matrices vary in number of taxa included. Matrices 7 and 9 contain the same gene regions, but differ in taxon sampling. The combined matrices range from 63 (matrices 7 and 8) to 132 (matrix 9) included taxa. Matrices 6 and 7 contain no missing data. Matrix 8 has missing data because not all RGCs were scoreable in all taxa; those taxa for which RGCs were available are indicated in Fig. 4.2. Matrix 9 also contain missing data because taxa were included that did not have all gene regions sequenced.

Matrix 1, RGCs data, was run in PAUP* 4.0b (Sinauer Associates, Inc.) using the clustering methods unweighted pair group method with arithmetic mean (UPGMA) and neighbor joining (NJ). Matrices 1-8 were analyzed on the CIPRES server (http://www.phylo.org/). All maximum likelihood (ML) analyses were run using the GTR GAMMA model. Bayesian analyses were run on ITS, *psbA–trnH*, and *psbMD* data sets with NST = 2 and rates = gamma, PHYA with NST = 6 and rates = invgamma, and RGCs with rates = gamma. These were the models supported by AIC in PartitionFinder. Bayesian results were examined with Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/) to ensure that runs converged, enough burn-in was eliminated, and effective sample sizes were adequate.

RESULTS

Matrices

The aligned matrices ranged in character number from 568 in the *psbA–trnH* matrix to 7065 in the ITS + *PHYA* + *psbMD* + RGC matrix (Table 4.2). Over half of these 7065 characters are contributed by *psbMB*. ITS has the highest number of parsimony informative (PI) sites relative to its size (368 in 692 aligned characters), whereas *psbMD* contributed the greatest

number of PI characters to the study overall (775). Not all matrices were equally informative (Table 4.2).

Phylogenetic Analyses

Trees generated with the ITS matrix resolved most tribes and other major clades recognized previously in the apioid superclade, the exceptions being the *Cachrys* clade in the ML tree and the *Sinodielsia* clade in both ML and Bayesian trees (Table 4.3; Fig. 4.3). The *Sinodielsia* and *Cachrys* clades are polytomies in the ML tree. The Bayesian analysis supports some intraclade relationships in the *Sinodielsia* clade and fully resolves the *Cachrys* clade (PP = 100). In general, tribal and generic-level relationships as inferred by ITS sequences are more fully resolved in the Bayesian tree than they are in the ML tree. In both trees, Careae and Pyramidoptereae are supported as sister tribes (BS = 91, PP = 100). In addition, in the Bayesian tree Apieae is allied with Pimpinelleae, Coriandreae is basal to Selineae, and there is weak support for the *Cachrys* clade as basal to Selineae + Coriandreae. In the ML tree Pimpinelleae and Apieae are part of a larger polytomy including Coriandreae, Echinophoreae, Selineae, *Conium*, and the *Cachrys* and *Opopanax* clades.

The trees generated using *psbA-trnH* sequences were highly unresolved, with some similar clades supported in both analyses (Fig. 4.4). The Bayesian tree resolved only five nodes with high (≥95) PP values, and the ML tree resolved only six nodes with ≥80 BS support values. One anomalous, well-supported node in the Bayesian tree places a *Tordylium* species (tribe Tordylieae) as sister to *Silaum* (*Sinodielsia* clade). There are no supported intertribal relationships in either tree. This locus contained far fewer PI characters than the other data sets, and because at least one well-supported node resulted in a rather spurious relationship, it was not included in any combined analysis.

Analysis of *PHYA* provides little resolution within and among tribes as well (Fig. 4.5). The ML tree contains 19 nodes with ≥ 80 BS. The Bayesian tree has a similar low level of resolution (16 nodes with ≥ 95 PP). Tribe Coriandreae is weakly supported as monophyletic (PP = 80) in the Bayesian tree and tribe Tordylieae is recovered as two well-supported lineages. These delineate clades Tordyliinae and *Cymbocarpum* of Tordylieae. The placement of all outgroup taxa as basal to the apioid superclade does not occur in the *PHYA* trees. *Anthriscus* (tribe Scandiceae) is supported as sister to *Rhodosciadium* (tribe Selineae) in both gene trees (BS = 100, PP = 98).

The *psbMD* trees resolved tribes Apieae, Careae, Echinophoreae, Pimpinelleae, and Pyramidoptereae (Table 4.3; Fig. 4.6). The Bayesian tree offers more resolution at intratribal levels than the ML tree. Both trees separate the two genera of the *Opopanax* clade, with *Smyrniopsis* sister to *Spermolepis* and *Opopanax* falling as a branch in a large polytomy. The two examined species of *Bifora* (tribe Coriandreae) do not form a monophyletic genus, with one species (*Bifora testiculata*) allying with the *Cachrys* clade and the other (*Bifora radians*) more closely allied with Selineae. In the ML tree Careae and Pyramidoptereae are sister tribes. Other tribal relationships are not resolved.

Analyses of the ITS + *psbMD* data set recovered tribes Apieae, Careae, Coriandreae, Echinophoreae, Pimpinelleae, Pyramidoptereae, Tordylieae, and the *Cachrys* clade with strong support (Table 4.3; Fig. 4.7). In addition, the Bayesian tree resolved the *Sinodielsia* clade. In the ML tree the *Sinodielsia* clade resolves as two well-supported lineages. In both ITS + *psbMD* trees (Fig. 4.7) Coriandreae is sister to Selineae. The relationship among these major clades differs than what was inferred using ITS. As an example, the Bayesian ITS + *psbMD* tree supports the *Cachrys* clade as part of a trichotomy with *Conium* and *Opopanax* and this clade is allied with Pimpinelleae and Apieae (PP = 98), while in the ITS Bayesian tree (Fig. 4.3) the *Cachrys* clade is supported as sister to Coriandreae + Selineae (PP = 87). Unlike the ITS trees

(Fig. 4.3), the *Opopanax* clade is not recovered as monophyletic in the ITS + *psbMD* trees (Fig. 4.7). ITS + *psbMD* trees have a topology similar to that of the *psbMD* trees (Fig. 4.6) such that *Smyrniopsis* is sister to *Spermolepis* (Figs. 4.6, 4.7) and *Opopanax persicus* is allied with *Conium* and the *Cachrys* clade.

Bayesian and ML trees generated from the ITS + *PHYA* + *psbMD* matrix of 63 taxa recovered tribes Apieae, Careae, Coriandreae, and the *Sinodielsia* and *Cachrys* clades (Table 4.3; Fig. 4.8). In both trees Pyramidoptereae is paraphyletic with Careae nested within. These two allied tribes are resolved as basal with regard to the rest of the apioid superclade. Monophyly of tribes Echinophoreae and Pimpinelleae and the *Opopanax* clade could not be determined because one or no representatives from each were included in the analyses. Analysis of the ITS + *PHYA* + *psbMD* + RGCs matrix retained much of the same overall topology of the ITS + *PHYA* + *psbMD* trees but did not improve resolution among tribes and clades (Fig. 4.9).

Analyses of the ITS + *PHYA* + *psbMD* matrix comprising 124 taxa recovered Apieae, Careae, Echinophoreae, Pimpinelleae, Tordylieae, and the *Sinodielsia* clade (Table 4.3; Fig. 4.10) using both inference methods. The Bayesian tree also supports Selineae as monophyletic (PP = 100), if *Smyrniopsis* is considered misplaced within the group. Some intertribal relationships are also resolved. Pimpinelleae was resolved as sister to Apieae (BS = 95, PP = 100) and these two tribes were allied with members a trichotomy with the *Cachrys* clade, *Conium*, and *Opopanax persicus* with strong support (PP = 100) in the Bayesian tree. As with the ITS + *PHYA* + *psbMD* Bayesian tree with fewer taxa (Fig. 4.8) Pyramidoptereae is resolved as paraphyletic with Careae arising from within. There is little resolution among Coriandreae, Selineae, Echinophoreae, Tordylieae, and the *Sinodielsia* clade. The two representatives from the *Opopanax* clade are distantly placed in the trees with *Smyrniopsis* sister to *Spermolepis* and *Opopanax persicus* as one branch a polytomy with the *Cachrys* clade + *Conium* and Apieae +

Pyramidoptereae in the Bayesian tree. Both genera of *Opopanax* occur within polytomies in the ML tree.

Rare Genomic Changes

Four plastid RGCs were scored for 111 taxa. These changes include inversions of gene regions *trnD-trnY-trnE* and *trnH-psbA*, changes in gene synteny at J_{LA}, and the presence of novel DNA at J_{LA} (Chapter 3). The inversion of *trnD-trnY-trnE* occurs in *Carum* (Careae), *Spermolepis* (Selineae), and *Smyrniopsis* (*Opopanax* clade). The inversion of *trnH-psbA* occurs in tribes Careae and Tordylieae, and in the *Sinodielsia* clade. Inverted repeat boundary type A, within *rps19*, occurs only in the outgroup taxa (Chapter 3; Fig. 3.1). Boundary type B, expansion into *rps3*, occurs in Careae and Pyramidoptereae. Boundary type D, contraction into *rpl2*, occurs in tribes Apieae and Pimpinelleae and the *Cachrys*, *Conioselinum*, and *Opopanax* clades. Boundary type D' is a newly identified boundary type, adjacent to D but before E (Chapter 3; Fig. 3.1) within non-coding DNA between genes *rpl2* and *rpl23*. Boundary types D' (*rpl2* to *rpl23* intergenic sequence), E (within *ycf2*), F (*ycf2* to *trnL* intergenic sequence), and G (within *ndhB*) are dispersed throughout Selineae, Tordylieae, and the *Sinodielsia* clade. Boundary types I and I' both occur within the Coriandreae. Boundary type H occurs only in *Tordylium aegyptiacum* var. *palaestinum* (Tordylieae). There are two types of DNA insertion at J_{LA}, mtDNA and filler DNA. The mtDNA insertion is found in tribes Apieae and Pimpinelleae and the *Cachrys*, *Conioselinum*, and *Opopanax* clades. Filler DNA at J_{LA} is dispersed throughout Pyramidoptereae, Selineae, Tordylieae, and the *Sinodielsia* clade.

When these characters are mapped onto the tree with the most resolved relationships, inferences about the number of times each RGC occurred during the evolution of the apioid superclade can be made (Fig. 4.11). The inversion of *trnD-trnY-trnE* occurred at least twice and potentially three times during the evolution of the apioid superclade. The inversion is shared by

Smyrniopsis and *Spermolepis* and could have occurred in their common ancestor, if this sister relationship is correct. However, if *Smyrniopsis* is misplaced and indeed belongs to the *Opopanax* clade then the inversion would have occurred three times. The inversion of *trnH - psbA* evolved a minimum of three times (Fig. 4.11). It is present in three species of Careae, one species of the *Sinodielsia* clade, and one species of Tordylieae. Boundary types A, B, D, D', G, H, I, and I' each evolved once when considering the more resolved tree (Figs. 4.9, 4.11). Boundary types E, F, and G are paraphyletic. Boundary type E occurs 14 times, three times in the *Sinodielsia* clade and 11 times in Selineae. Five taxa have boundary type F: one Tordylieae species and four Selineae species. Boundary type G occurs three times, once in the *Sinodielsia* clade and once in Tordylieae. The insertion of mtDNA at J_{LA} occurred only once in the ancestor of the clade containing tribes Apieae and Pimpinelleae and the *Cachrys* and *Opopanax* clades. The insertion of filler DNA at J_{LA} is paraphyletic and occurred a minimum of nine times, once within the *Sinodielsia* clade, three times in Tordylieae, and five times within Selineae.

When taxon sampling is increased to 132 (matrix 8, Fig. 4.10), the minimum number of times boundary types D', E and I occur increases (Fig. 4.12). With increased taxon sampling boundary type I is supported as evolving twice if the placement of *Bifora testiculata* is correct (Fig. 4.12): once in *Bifora radians* (Coriandreae) and once in *Bifora testiculata* (*Cachrys* clade). Eighteen taxa have boundary type E, three in the *Sinodielsia* clade, and 15 in Selineae. This supports the boundary type evolving at least three times. Boundary type D' occurs six times, all within Selineae. The number of filler DNA insertions also increases evolving a minimum of 10 times, once in Pyramidoptereae, once in the *Sinodielsia* clade, three times in Tordylieae, and six times in Selineae.

RGCs can be used to discriminate amongst hypotheses of relationships. Within the ITS trees (Fig. 4.3) the two genera of the *Opopanax* clade resolve as sister taxa. However, these two taxa are placed distantly in many other trees presented herein. RGCs ally *Opopanax* with

the clade of Apieae, Pimpinelleae, *Conium maculatum,* and the *Cachrys* clade (Fig. 4.11). The members of this group all have IR boundary type D within *rpl2* and novel DNA at J_{LA}. Phylogenies generated herein support either tribes Careae and Pyramidoptereae as monophyletic sister groups or a monophyletic Careae nested within a paraphyletic Pyramidoptereae. The RGC data cannot discriminate amongst these hypotheses, because all taxa share a boundary type that only evolved once (B) and Careae has the *trnH-psbA* inversion that is not shared by Pyramidoptereae (Fig. 4.11). This inversion supports monophyly of Careae but not a close relationship to Pyramidoptereae. *Smyrniopsis* is allied with *Spermolepis* and this close relationship is supported by the sharing of the *trnD-trnY-trnE* inversion. The molecular data suggest that the genus *Bifora* may not be monophyletic. *Bifora testiculata* is placed within the *Cachrys* clade in several different phylogenies. Both *Bifora* species have IR boundary type I, which does not support the placement of *Bifora testiculata* within the *Cachrys* clade that has boundary type D.

RGCs within the *Sinodielsia* clade are not shared by all members. *Silaum silaus* has an IR boundary within 5' *ndhB* and has the inversion of genes *psbA* and *trnH.* The remaining *Sinodielsia* taxa all have their IR boundary within *ycf2* as well as novel DNA at J_{LA}. This pattern is similar to what is found in Selineae and Tordylieae, where there is homoplasy in RGCs throughout the tribes (Figs. 4.11, 4.12).

Discussion

The apioid superclade is composed of 14 major lineages of largely unknown evolutionary relationships. Although previous molecular systematic studies of cpDNA and ITS sequences, cpDNA restriction sites, and RGCs (e.g., Downie et al. 2001, 2010; Plunkett and Downie 1999, 2000) have increased understanding of intertribal relationships, all have failed to fully resolve them. The goal of this study was three-fold: 1) to determine the phylogenetic utility of *psbM*–

psbD, *psbA–trnH,* and *PHYA* sequences, and RGCs in resolving relationhsips within the apioid superclade; 2) to elucidate these intertribal relationships; and 3) to trace the evolutionary history of RGCs in the apioid superclade.

To determine the phylogenetic utility of the four markers, informativeness was assessed through comparison of PI characters, as well as number of resolved tribes and other major clades identified in previous studies. RGCs by themselves did not produce phylogenetic trees having any resolution, as assessed by UPGMA and NJ methods. This lack of signal is undoubtedly due to the low number of RGCs scored. For example, within Campanulaceae the large number of RGCs identified were able to produce a well-resolved phylogeny (Cosner et al. 2004). Furthermore, when analyzed alongside the DNA markers, the addition of RGCs did not considerably improve the delimitation of the various tribes and other clades within the apioid superclade.

Characters pertaining to boundary types B and D, mtDNA at J_{LA}, and the *trnH* and *psbA* inversion are all phylogenetically informative. Even though the inversion of *trnH* and *psbA* occurs within more than one clade, its presence supports the monophyly of at least one tribe, Careae. The filler DNA at J_{LA}, the other IR boundary types, and the inversion of *trnD-trnY-trnE* between *psbM* and *trnT* are all homoplasous and do not aid in supporting any previously identified higher-level relationships within the apioid superclade.

The *psbA–trnH* locus is not informative at the generic and tribal levels. Additionally, this locus was unable to recover any previously recognized tribes or other major clades. Liu et al. (2014) assessed ITS, *psbA–trnH*, and two additional plastid loci as potential barcoding regions in Apiaceae. While they reported that *psbA–trnH* was the most variable locus they examined, ITS performed better at species identification. The current study agrees with this conclusion – *psbA–trnH* is not a good locus for higher-level phylogenetic inference in Apiaceae.

PHYA is a single copy (in apioid superclade species) nuclear gene and belongs to the phytochrome gene family. *PHYA* has more PI characters that ITS, however, it did not perform well in the phylogentic analysis. This locus was unable to recover any previously designated tribes or other major clades within the apioid superclade, nor did it resolve any tribal-level relationships. There are two reasons why *PHYA* did not perform well. One, the gene may not have coalesced. Pillon et al. (2013) reported, in their study of island plants, that single copy nuclear genes, including *PHYA*, may not be an ideal choice for phylogenetics of young lineages. Banasiak et al. (2013) dated the divergence of the apioid superclade at 24-30 mya, potentially making the group too young for the coalescence of *PHYA*. Two, the gene may have recently been duplicated such that its copies were not in fact homologous. Duplications of *PHYA* are reported from individual species in some lineages (Bennett and Mathes 2006; Turner et al. 2013), and are readily identifiable. If the duplication was recent a non-homologous copy would not be divergent enough to be apparent during alignment and therefore missed. This may explain the odd placements of some outgroup taxa within the apioid superclade. Overall this locus is not suitable for resolving relationships at deep levels within the apioid superclade.

The *psbMD* region was identified by Downie and Jansen (2015) as the most variable plastid region in a comparison of five Apiales plastomes. Indeed, *psbMD* has the most PI characters of the regions considered herein and produces trees with much resolution at the generic- and tribal-levels. The combination of ITS and *psbMD* produced trees with the greatest resolution of all new loci examined based on overall resolution, both in strong support for previously recognized tribes and major clades, and illuminating more intertribal relationships than any other matrix analyzed thus far.

The ITS + *psbMD* trees resolve Careae and Pyramidoptereae as sister tribes; this relationship is also supported by RGC data. This sister relationship has been inferred in other studies using ITS and plastid intron sequences (Ajani et al. 2008; Zhou et al. 2008; Spalik et al.

2010; Banasiak et al. 2013). In addition, in all analyses with resolution, Careae + Pyramidoptereae appear basal to all other members of the apioid superclade, a position also supported by ITS and plastid intron sequences (Ajani et al. 2008; Zhou et al. 2008, 2009; Spalik et al. 2010; Banasiak et al. 2013).

Consistent with other studies, ITS + *psbMD* resolves tribe Coriandreae (*Bifora* and *Coriandrum*) as monophyletic and basal to Selineae (Ajani et al. 2008; Banasiak et al. 2013), when *Bifora testiculata* is not considered. *B*ifora *testiculata* has not been included in previous molecular studies. Bayesian analysis of *PHYA* groups *B. testiculata*, *B. radians*, and *Coriandrum* (PP = 80), while the *psbMD* Bayesian tree places B. *testiculata* within a paraphyletic *Cachrys* clade sister to *Azilia* (PP = 88). Both *Bifora* species have boundary type I, however, *Coriandrum* has boundary type I'. None of these taxa have inversions or insertions to help with placement. Additional studies assessing the monophyly of *Bifora*, which consists of three species (Pimenov and Leonov 1993), are necessary.

The ITS + *psbMD* trees support the clade comprising tribes Apieae and Pimpinelleae, the *Cachrys* and *Conium* clades, and *Opopanax* as basal to Coriandreae, Echinophoreae, the *Sinodielsia* clade, Selineae, and *Smyrniopsis*. This clade conflicts with ITS results, but is consistent with relationships in the *psbMD* trees. Previous analyses resolved the *Cachrys* clade as basal to Coriandreae + Selineae (Zhou et al. 2008; Banasiak et al. 2013). In addition, ITS phylogenies placed Pimpinelleae basal to Apieae (Zhou et al. 2008, 2009; Banasiak et al. 2013). Apieae is basal to the *Opopanax* clade followed by the *Conium* clade in some ITS trees (Banasiak et al. 2013), while Apieae is basal to Selineae in other trees (Spalik et al. 2010), or relationships are unresolved (Zhou et al. 2008). The RGCs data, boundary type D and mitochondrial DNA at J_{LA}, support the close relationship of Apieae, Pimpinelleae, *Cachrys* clade, *Conium* clade, and *Opopanax persicus*. No intertribal relationships among Echinophoreae, Tordylieae, and the *Sinodielsia* clade were recovered.

The *Opopanax* clade requires revision. The two examined members of this clade, *Opopanax persicus* and *Smyrniopsis aucheri*, have inconsistent relationships. In previous molecular studies *Smyrniopsis* is resolved as sister to *Opopanax persicus* (Spalik et al. 2004; Ajani et al. 2008). Ajani et al. (2008) reported that *Opopanax* (three species) and the monotypic *Smyrniopsis* were sister clades in the apioid superclade. In the current study, phylogenetic signal from *psbMD* overwhelmed that of ITS and led to *Smyrniopsis* allying with *Spermolepis* and not with *Opopanax persicus. Opopanax persicus* and *Smyrniopsis* do not share plastid RGCs. In this study the placement of *Opopanax persicus* is supported by two RGCs, boundary type D and the insertion of DNA at J_{LA}, and is allied with Apieae, the *Cachrys* clade, Pyramidoptereae, and *Conium*. *Smyrniopsis* and *Spermolepis* are supported as being closely related by sharing the *trnD-trnY-trnE* inversion.

Conium maculatum, poison hemlock, is perhaps the most infamous apioid superclade species. It is also one of the most difficult to place. Resolution of *Conium* ranges from no supported placement (Winter et al. 2008), weakly supported as an ally to the *Cachrys* clade (Logacheva 2010), allied with Pimpinella (Downie et al. 1996), basal to Tordylieae (Ajani et al. 2008; Zhou et al. 2008), to sister to an expanded *Apium* clade (Downie et al. 2001, 2002). In the current study *Conium* falls basal to the *Cachys* clade and is allied with Apieae and Pimpinelleae. This relationship is supported by DNA data and two RGCs characters – IR boundary location and the presence of putative novel DNA at J_{LA}.

While additional sequence data from *psbMD* and RGCs have illuminated inconsistencies in the placement of genera within the *Opopanax* clade and the monophyly of *Bifora*, these data have helped to clarify some relationships among the tribes and major clades of the apioid superclade. In addition, increased resolution among these lineages has provided context for studying the evolution of plastome RGCs. Apiaceae plastomes have dynamic synteny changes and novel DNA insertions that make for an ideal study system for plastome evolution. Further

work to delineate relationships among within the apioid superclade needs to be done to illuminate the frequency of these RGCs.

References

Ajani Y, Ajani A, Cordes JM, Watson MF, Downie SR (2008) Phylogenetic analysis of nrDNA ITS sequences reveals relationships within five groups of Iranian Apiaceae subfamily Apioideae. Taxon 57:383–401.

Banasiak Ł, Piwczyński M, Uliński T, Downie SR, Watson MF, Shakya B, Spalik K (2013) Dispersal patterns in space and time: a case study of Apiaceae subfamily Apioideae. J Biogeog 40:1324-1335.

Beilstein MA, Al-Shehbaz IA, Mathews S, Kellogg EA (2008) Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. Am J Bot 95:1307-1327.

Bell JR (2008) A Simple Way to Treat PCR Products Prior to Sequencing Using ExoSAP-IT®. BioTechniques 44:834.

Bennett JR, Mathews S (2006) Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. Am J Bot 93:1039-1051.

Calviño CI, Tilney PM, van Wyk B-E, Downie SR (2006) A molecular phylogenetic study of southern African Apiaceae. Am J Bot 93:1828-1847.

CBOL Plant Working Group, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim K-J, Kress WJ, Schneider H, van AlphenStahl J, Barrett SCH, van den Berg C, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacón J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TAJ, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim Y-D, Lahaye R, Lee H-L, Long DG, Madriñán S, Maurin O, Meusnier I, Newmaster SG, Park C-W, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi D-K, Little DP (2009) A DNA barcode for land plants. PNAS 106:12794-12797.

Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol Biol 4:27.

Downie SR, Hartman RL, Sun F-J, Katz-Downie DS (2002) Polyphyly of the spring-parsleys (*Cymopterus*): molecular and morphological evidence suggests complex relationships among the perennial endemic genera of western North American Apiaceae. Can J Bot 80:1295-1324.

Downie SR, Jansen RK (2015) A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot 40:336-351.

Downie SR, Katz-Downie DS (1996) A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal transcribed spacer sequences. Am J Bot 83:234–251.

Downie SR, Katz-Downie DS (1999) Phylogenetic analysis of chloroplast *rps16* intron sequences reveals relationships within the woody southern African Apiaceae subfamily Apioideae. Can J Bot 77:1120–1135.

Downie SR, Katz-Downie DS, Watson MF (2000) A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpoC1* intron sequences: towards a suprageneric classification of subfamily Apioideae. Am J Bot 87:273–292.
Downie SR, Palmer JD (1992) Use of chloroplast DNA rearrangements in reconstructing plant physiology. In: Molec Syst Plants, Eds Soltis PS, Soltis DE, Doyle JJ pp 14-35.
Downie SR, Plunkett GM, Watson MF, Spalik K, Katz-Downie DS, Valiejo-Roman CM, Terentieva EI, Troitsky AV, Lee B-Y, Lahham J, El-Oqlah A (2001) Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. Edinb J Bot 58:301–330.
Downie SR, Ramanath S, Katz-Downie DS, Llanas E (1998) Molecular systematics of Apiaceae subfamily Apioideae: phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer and plastid rpoC1 sequences. Am J Bot 85:563–591.
Downie SR, Spalik K, Katz-Downie DS, Reduron J-P (2010) Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. Plant Div Eol 128:111-136.
Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.
Katz-Downie DS, Valiejo-Roman CM, Terentieva EI, Troitsky AV, Pimenov MG, Lee B, Downie SR (1999) Towards a molecular phylogeny of Apiaceae subfamily Apioideae: additional information from nuclear ribosomal DNA ITS sequences. Plant Syst Evol 216:167–195.
Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695-1701.
Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A (2014) Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol Biol 14:82.
Logacheva MD, Valiejo-Roman CM, Degtjareva CV, Stratton JM, Downie SR, Samigullin TH, Pimenov MG (2010) A comparison of nrDNA ITS and ETS loci for phylogenetic inference in the Umbelliferae: an example from tribe Tordylieae. Mol Phylogenet Evol 57:471-476.
Lui J, Shi L, Han J, Li G, Lu H, Hou J, Zhou XI, Meng F, Downie SR (2014) Identification of species in the angiosperm family Apiaceae using DNA barcodes. Mol Ecol Resources 14:1231-1238.
Maddison WP, Maddison DR (2014) Mesquite: a modular system for evolutionary analysis. Version 3.01 http://mesquiteproject.org.
Magee AR, Calviño CI, Liu M, Downie SR, Tilney PM, van Wyk B-E (2010) New tribal delimitations for the early diverging lineages of Apiaceae subfamily Apioideae. Taxon 59:567-580.
Magee AR, van Wyk B-E, Tilney PM, Sales F, Hedge I, Downie SR (2009) *Billburttia*, a new genus of Apiaceae (tribe Apieae) endemic to Madagascar. Plant Syst Evol 283:237-245.
Mathews S (2010) Evolutionary studies illuminate the structural-functional model of plant phytochromes. The Plant Cell 22:4-16.
Mathews S, Lavin M, Sharrock RA (1995) Evolution of the phytochrome gene family and its utility for phylogenetic analyses of angiosperms. Annals Missouri Bot Garden 82:296–321.
Mathews S, Sharrock RA (1997) Phytochrome gene diversity. Plant Cell Env 20:666-671.
Nie Z-L, Wen J, Zauma H, Qiu Y-L, Sun H, Meng Y, Sun W-B, Zimmer EA (2008) Phylogenetic and biogeographic complexity of Magnoliaceae in the Northern Hemisphere inferred from three nuclear data sets. Mol Phylogenet Evol 48:1027-1040.

Pillon Y, Johansen J, Sakishima T, Chamala S, Barbazuk WB, Raolson EH, Price DK, Stacey E (2013) Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. BMC Evol Biol 13:35.
Pimenov MG, Kljuykov EV, Degtjareva GV (2011) Survey of the genus *Diplotaenia* (Umbelliferae), with description of two new species from Turkey. Willdenowia 41:67-74.
Plunkett GM, Downie SR (1999) Major lineages within Apiaceae subfamily Apioideae: a comparison of chloroplast restriction site and DNA sequence data. Am J Bot 86:1014–1026.
Plunkett GM, Downie SR (2000) Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Systematic Botany: 25-648-667.
Plunkett GM, Soltis DE, Soltis PS (1996) Evolutionsary patterns in Apiaceae: inferences based on *matK* sequence data. Syst Bot 83:499-515.
Poppe C, Ehmann B, Frohnmeyer H, Furuya M, Schafer E (1994) Regulation of phytochrome A mRNA abundance in parsley seedlings and cell-suspension cultures. Plant Mol Biol 26: 481-486.
Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ, editor. Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants. London: CAB International.
Reduron J-P, Mathez J, Downie SR, Danderson CA, Ostroumova T (2009) *Pseudoridolfia*, nouveau genre d'Apiaceae découvert au Maroc. Acta Botanica Gallica 156:487-500.
Rokas A, Holland PWH (2000) Rare genomic changes as tools for phylogenetics. Trends Ecol Evol 15:454-459.
Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
Ruhlman T, Lee S, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H (2006) Complete plastid genome sequence of *Daucus carota*: Implications for biotechnology and phylogeny of angiosperms. BMC Genomics 7: 222.
Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92:142–166.
Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275-288.
Spalik K, Downie SR (2007) Intercontinental disjunctions in *Cryptotaenia* (Apiaceae, Oenantheae): an appraisal using molecular data. J Biogeog 34:2039-2054.
Spalik K, Piwczynski M, Danderson CA, Kurzyna-Mlynik R, Bone TS, Downie SR (2010) Amphitropic amphiantarctic disjunctions in Apiaceae subfamily Apioideae. J Biogeogr 37:1977-1994.
Spalik K, Reduron J-P, Downie SR (2004) The phylogenetic position of *Peucedanum* sensu lato and allied genera and their placement in tribe Selineae (Apiaceae, subfamily Apioideae). Plant Syst Evol 243:189-210.
Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics 30:1312-1313.
Sun F-J, Downie SR, Hartman RL (2004) An ITS-based phylogenetic analysis of the perennial, endemic Apiaceae subfamily Apioideae of western North America. Syst Bot 29:419-431.
Terentieva EI, Valiejo-Roman CM, Samigullin TH, Pimenov MG, Tilney PM (2015) Molecular phylogenetic and morphological analyses of the traditional tribe Coriandreae (Umbelliferae-Apioideae). Phytotaxa 195:251–271.

The Plant List (2013). Version 1.1. Published on the Internet; http://www.theplantlist.org/ (accessed 27 March 2015).
Turner B, Munzinger J, Duangjai S, Temsch EM, Stockenhuber R, Barfuss MHJ, Chase MW, Samuel R (2013) Molecular phylogenetics of New Caledonian *Diospyros* (Ebenaceae) using plastid and nuclear markers. Mol Phylogenet Evol 69:740-763.
Valiejo-Roman CM, Pimenov MG, Terentieva EI, Downie SR, Katz-Downie DS, Troitsky AV (1998) Molecular systematics of the Umbelliferae: using nuclear rDNA internal transcribed spacer sequences to resolve issues of evolutionary relationships. Botanicheskii Zhurnal 83:1-22.
Valiejo-Roman CM, Shneyer VS, Samigullin TH, Terentieva EI, Pimenov MG (2006) An attempt to clarify taxonomic relationships in “Verwandtschaftskreis der Gattung *Ligusticum*” (Umbelliferae-Apioideae) by molecular analysis. Plant Syst Evol 257:25-43.
Wheeler TJ, Kececioglu JD (2007) Multiple alignments by aligning alignments. Bioinformatics, 23:i559-i568.
Wheeler TJ, Maddison DR (2012) Opalescent: A Mesquite package providing access to Opal within Mesquite v. 2.10.
Winter PJD, Magee AR, Phephu N, Tilney PM, Downie SR, and van Wyk B-E (2008) A new generic classification for African peucedanoid species (Apiaceae). Taxon 57:347-364.
Zhou J, Gong X, Downie SR, Peng H (2009) Towards a more robust molecular phylogeny of Chinese Apiaceae subfamily Apioideae: additional evidence from nrDNA ITS and cpDNA intron (*rpl16* and *rps16*) sequences. Mol Phyl Evol 53:56-68.
Zhou J, Peng H, Downie SR, Liu Z-W, Gong X (2008) A molecular phylogeny of Chinese Apiaceae subfamily Apioideae inferred from nuclear ribosomal DNA internal transcribed spacer sequences. Taxon 57:402-416.

Tables and Figures

Table 4.1 Primer name, location, sequence, and reference to previous publication, if applicable.

Name	Location	Sequence	Reference
a152f.1	phytochrome A	ACN ATG GTN AGY CAY GCN GTN CC	Mathews et al. 1995
a156f.api	phytochrome A	CAY GCT GTT CCA AGT GTN GGY G	modified from Mathews et al. 1995
a230f.api	phytochrome A	GAC TTY GAR CCB GTB ARG CCT TAY G	modified from Mathews et al. 1995
a832r	phytochrome A	RTT CCA YTC NGA RCA CCA NCC	Mathews et al. 1995
a840r.api	phytochrome A	CCA TCC AGA YAA YTC TGT CAT AGC	modified from Mathews et al. 1995
a2241r.api	phytochrome A	TGG ARC YRA GTY TTC CCT RGA	
psbA3f	photosystem II protein D1	GCT AAC CTT GGT ATG GAA GT	
trnHr	tRNA-His	GCC TTR RTC CAC TTG SCT AC	
psbMf	photosystem II protein M	AGC AAT AAA TGC AAG AAT ATT TAC TTC	
trnDf	tRNA-Asp	ACC AAT TGA ACT ACA ATC CC	
trnDr	tRNA-Asp	GGG ATT GTA GTT CAA TTG GT	
trnEf	tRNA-Glu	CTC CTT GAA AGA GAG ATG TCC T	
trnT	tRNA-Thr	CCC TTT TAA CTC AGT GGT AG	
trnTr	tRNA-Thr	CTA CCA CTG AGT TAA AAG GG	
psbD	photosystem II protein D2	CTC CGT ARC CAG TCA TCC ATA	

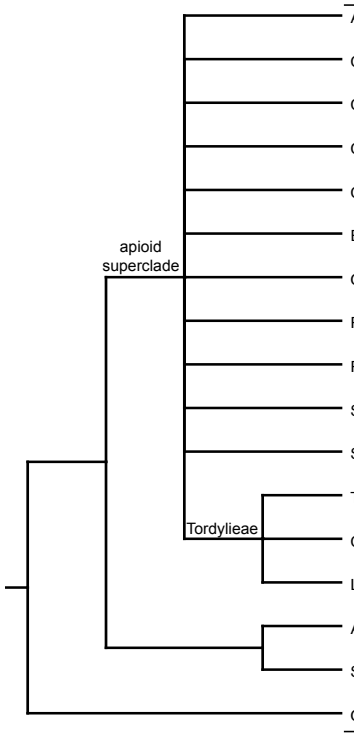
Table 4.2 Number of genera, taxa, aligned characters, and informativeness (calculated as parsimony informative characters, PI) in each data matrix. Sums of genera and taxa do not include outgroup species.

Matrix number and marker	No. of genera	No. of taxa	No. of aligned characters	No. of constant characters	No. of variable character that are not PI	No. of PI characters
1) ITS	105	124	692	235	89	368
2) *PHYA*	77	86	1961	1209	336	416
3) *psbMD*	90	109	4353	2705	873	775
4) *psbA–trnH*	64	67	568	320	126	122
5) ITS + *psbMD*	86	99	5087	3058	990	1039
6) ITS + *PHYA* + *psbMD*	58	63	7048	4730	1188	1130
7) ITS + *PHYA* + *psbMD* + RGCs	58	63	7065			
8) ITS + *PHYA* + *psbMD*	110	132	7006	4240	1248	1518

Table 4.3 Bootstrap and posterior probabilities supporting previously designated tribes and major clades in the apioid superclade. If a node is not well supported (BS ≥ 80; PP ≥ 95) values are not reported.

	0) RGCs		1) ITS		2) *PHYA*		3) *psbMD*		4) *psbA–trnH*		5) ITS + psbMD		6) ITS + *PHYA* + psbMD		7) ITS + *PHYA* + psbMD + RGCs		8) ITS + *PHYA* + psbMD	
Tribe/Clade	ML	Bayes[1]	ML	Bayes	ML	Bayes	ML	Bayes	ML	Bayes	ML	Bayes	ML	Bayes	ML	Bayes	ML	Bayes
Apieae			99	100			100	100			100	100	100	100	100	100	100	100
Cachrys clade				100							100	100	98	100	99	100	98	100
Careae			100	100			100	100			100	100	100	100	100	100	98	100
Coriandreae			96	100							96	100	100	100	98	100	100	100
Echinophoreae			100	100			90	99			100	100	NA[2]	NA	NA	NA	100	90
Opopanax clade			100	100									NA	NA	NA	NA		
Pimpinelleae			100	100			100	100			100	100	NA	NA	NA	NA	98	100
Pyramidoptereae			100	100			100	100			100	100						
Selineae			99	100														100
Sinodielsia clade						94						97	98	100	98	100	100	100
Tordylieae			92	100							97	100					93	96

[1]Bayes = Bayesian inference
[2]NA = no taxa or not enough taxa were included from this tribe/clade to determine monophyly.



Taxonomic group	Number of genera	Number sampled
Apieae	13	12
Cachrys clade	8	5
Careae	12	7
Conium clade	1	1
Coriandreae	2	2
Echinophoreae	4	3
Opopanax clade	2	2
Pimpinelleae	16	6
Pyramidoptereae	32	6
Selineae	57	50
Sinodielsia clade	16	6
Tordyliinae	16	6
Cymbocarpum clade	2	2
Lefebvrea clade	10	2
Acronema clade	18	1
Scandiceae	31	9
Oenantheae	18	1

Fig. 4.1 Summary of relationships among the tribes and other major clades of the apioid superclade inferred by phylogenetic analysis of molecular data (modified from Downie et al. 2010). Also included are the *Acronema* clade, and tribes Scandiceae and Oenantheae as outgroups. The number of genera per clade (Downie et al. 2010) and the number of taxa sampled in this study are also indicated.

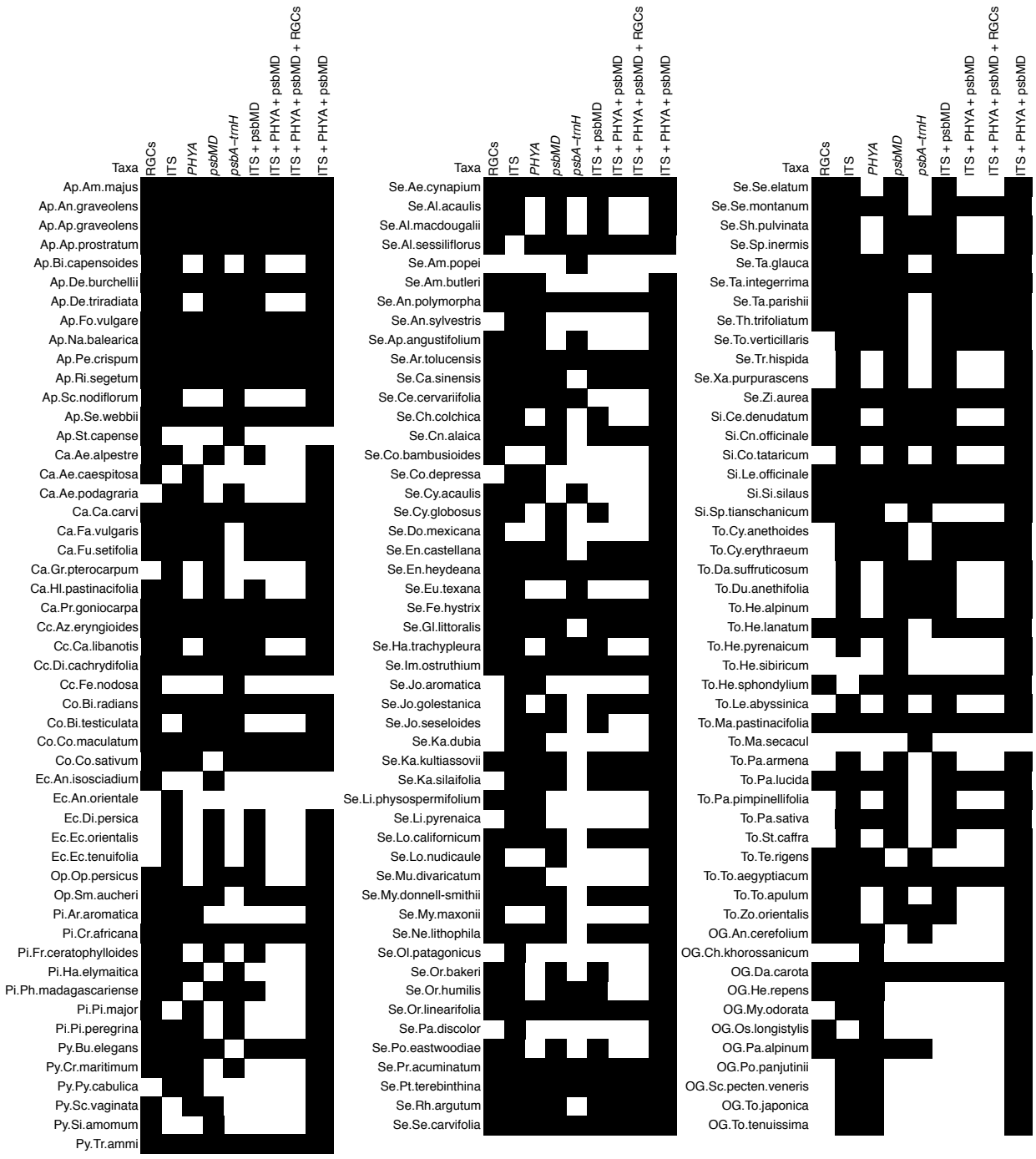


Fig. 4.2 Apioid superclade and outgroup taxa included in each of the eight data matrices analyzed herein. The inclusion of a taxon in a dataset is indicated by a black cell while its absence is indicated by a blank cell.

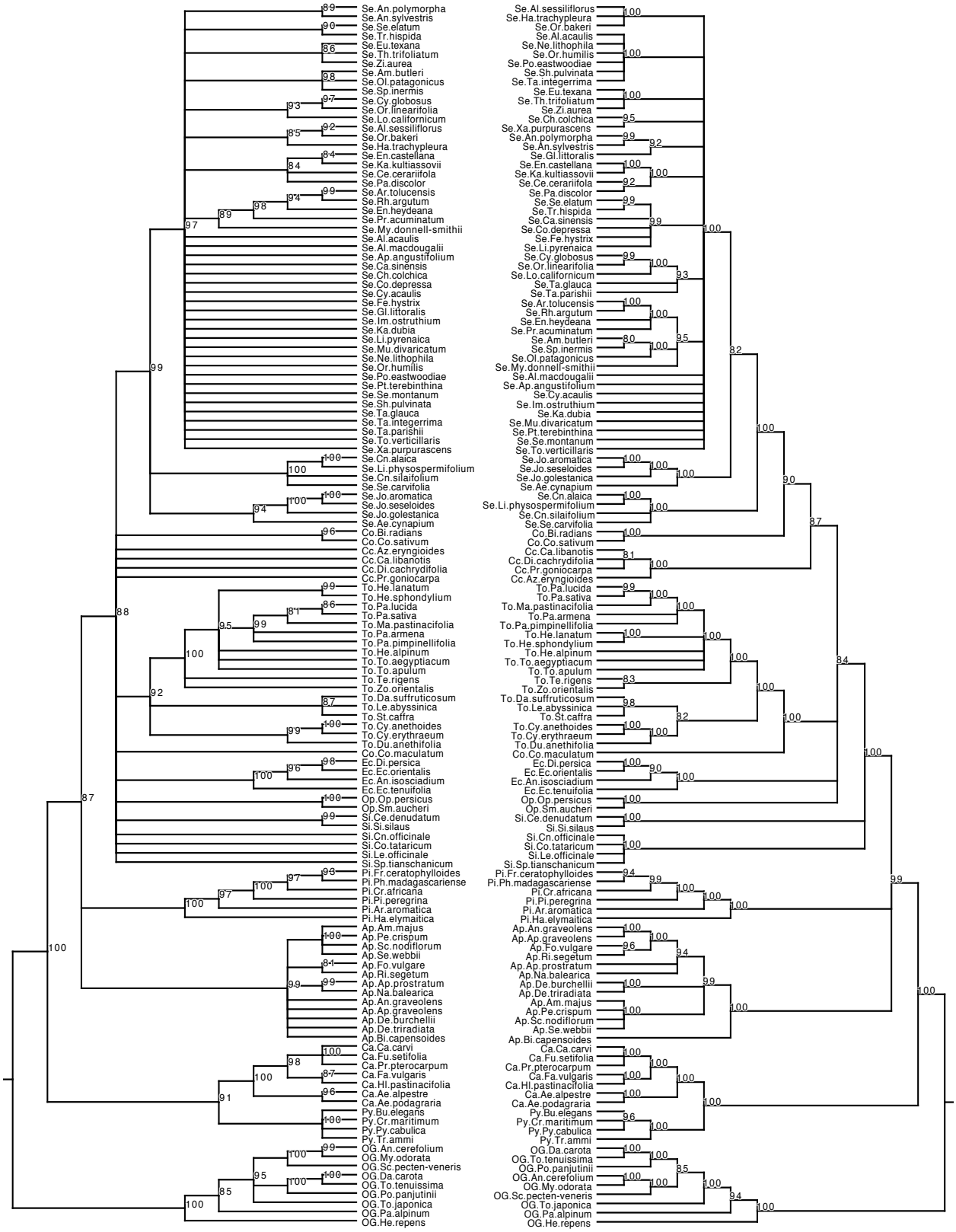


Fig. 4.3 Phylogenies generated from ITS matrix. The ML tree is on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

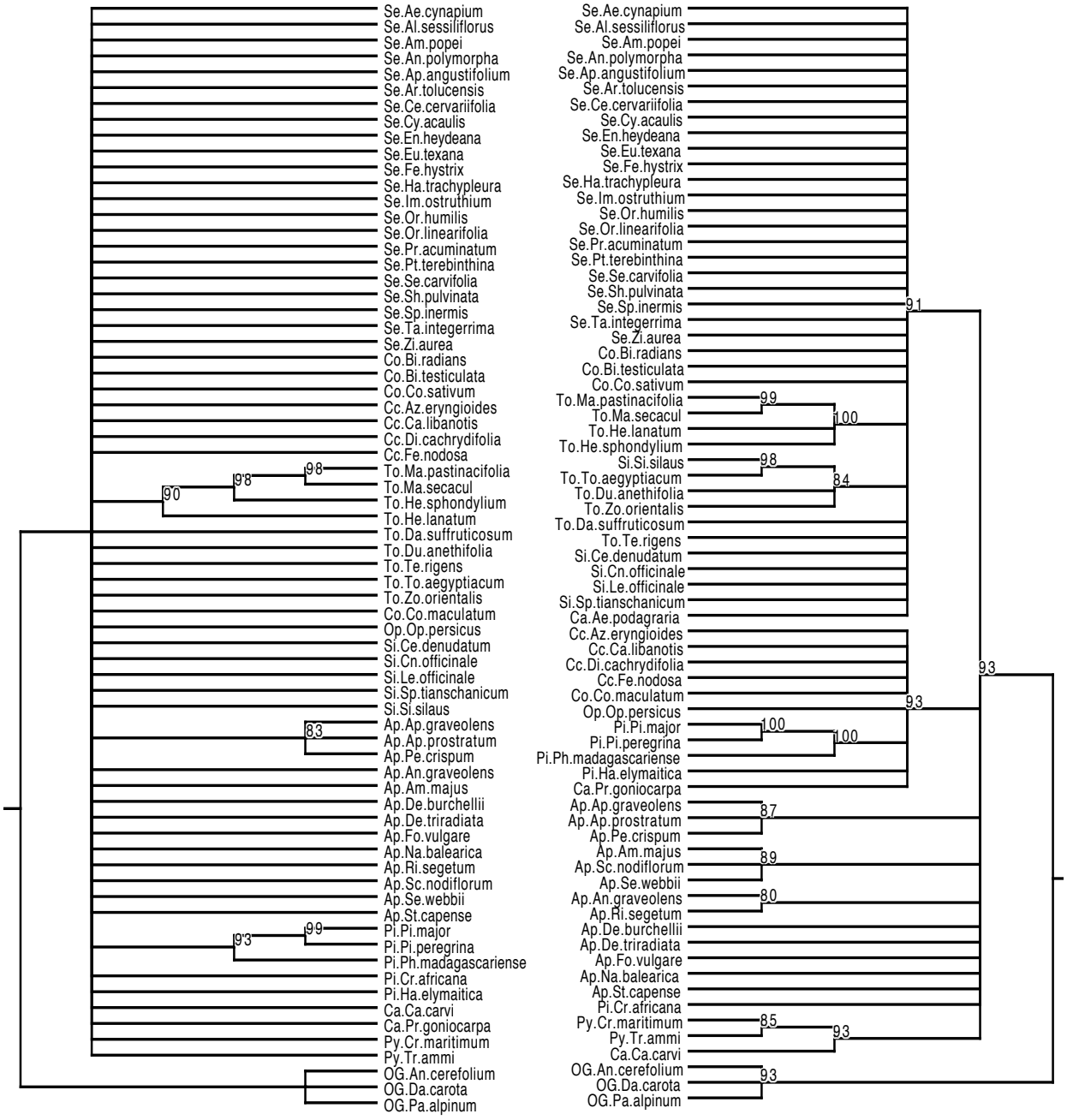


Fig. 4.4 Phylogenies generated from *psbA–trnH* matrix. The ML tree is on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

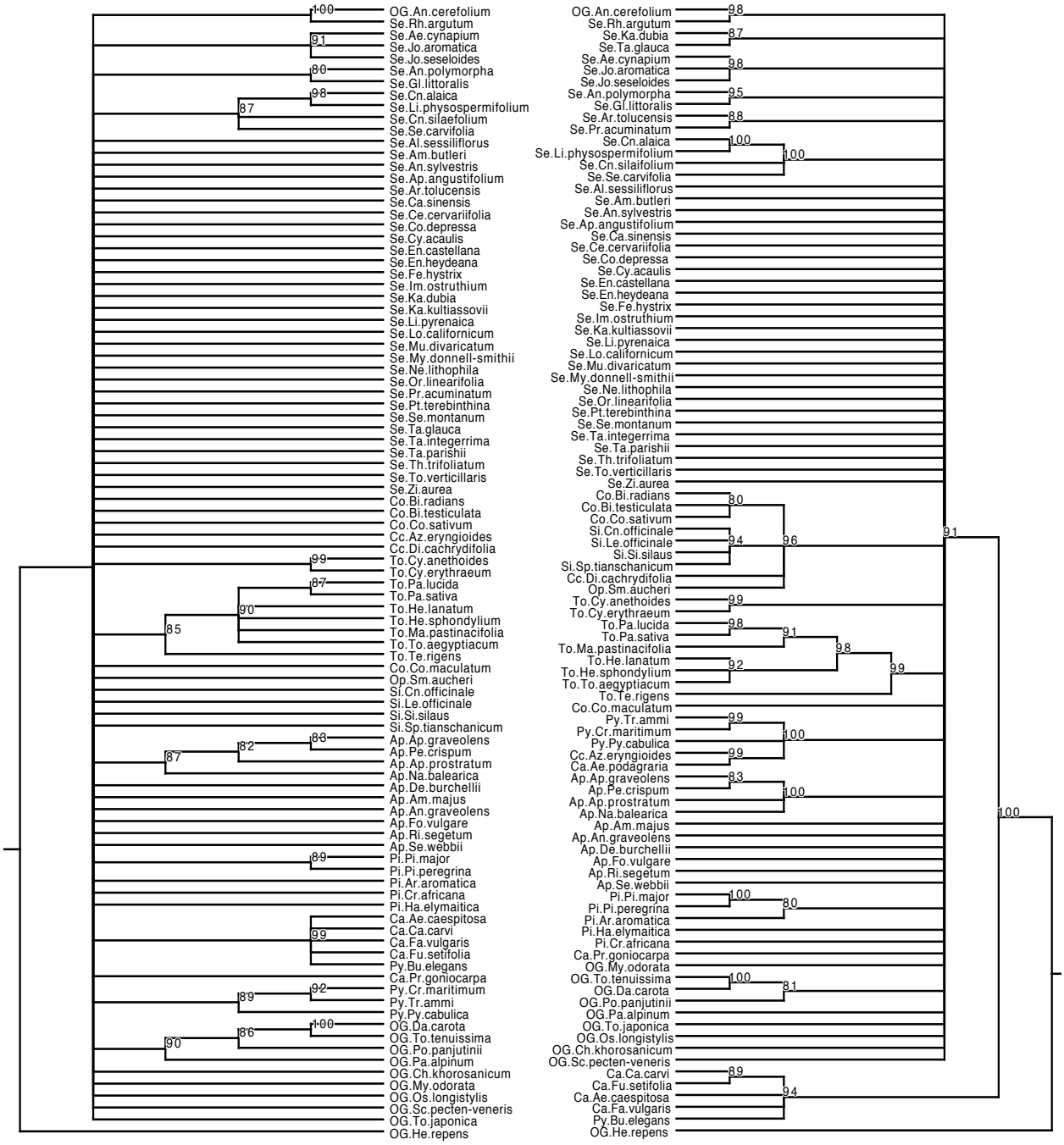


Fig. 4.5 Phylogenies generated from *PHYA* matrix. The ML tree is on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

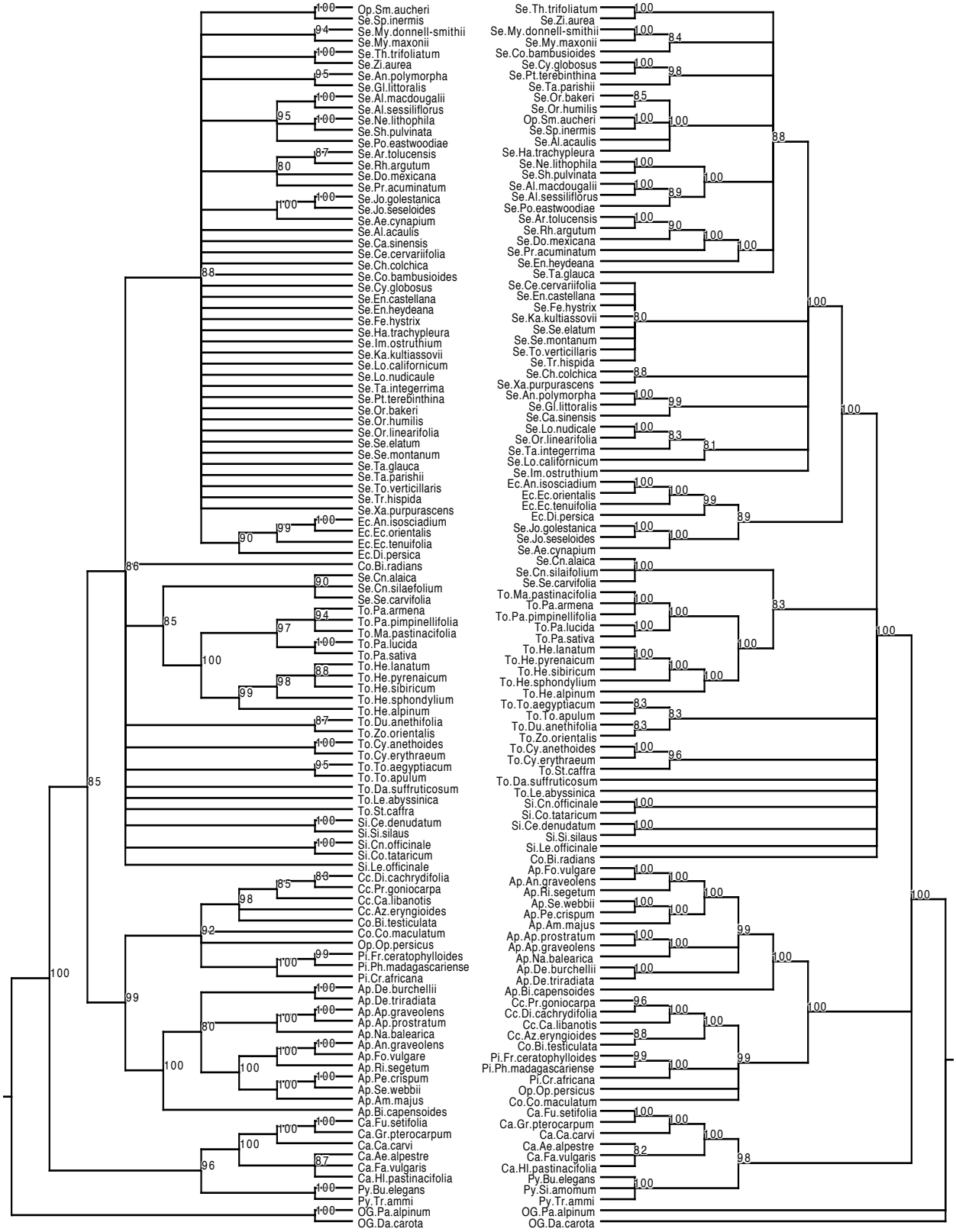


Fig. 4.6 Phylogenies generated from *psbMD* matrix. The ML tree is on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

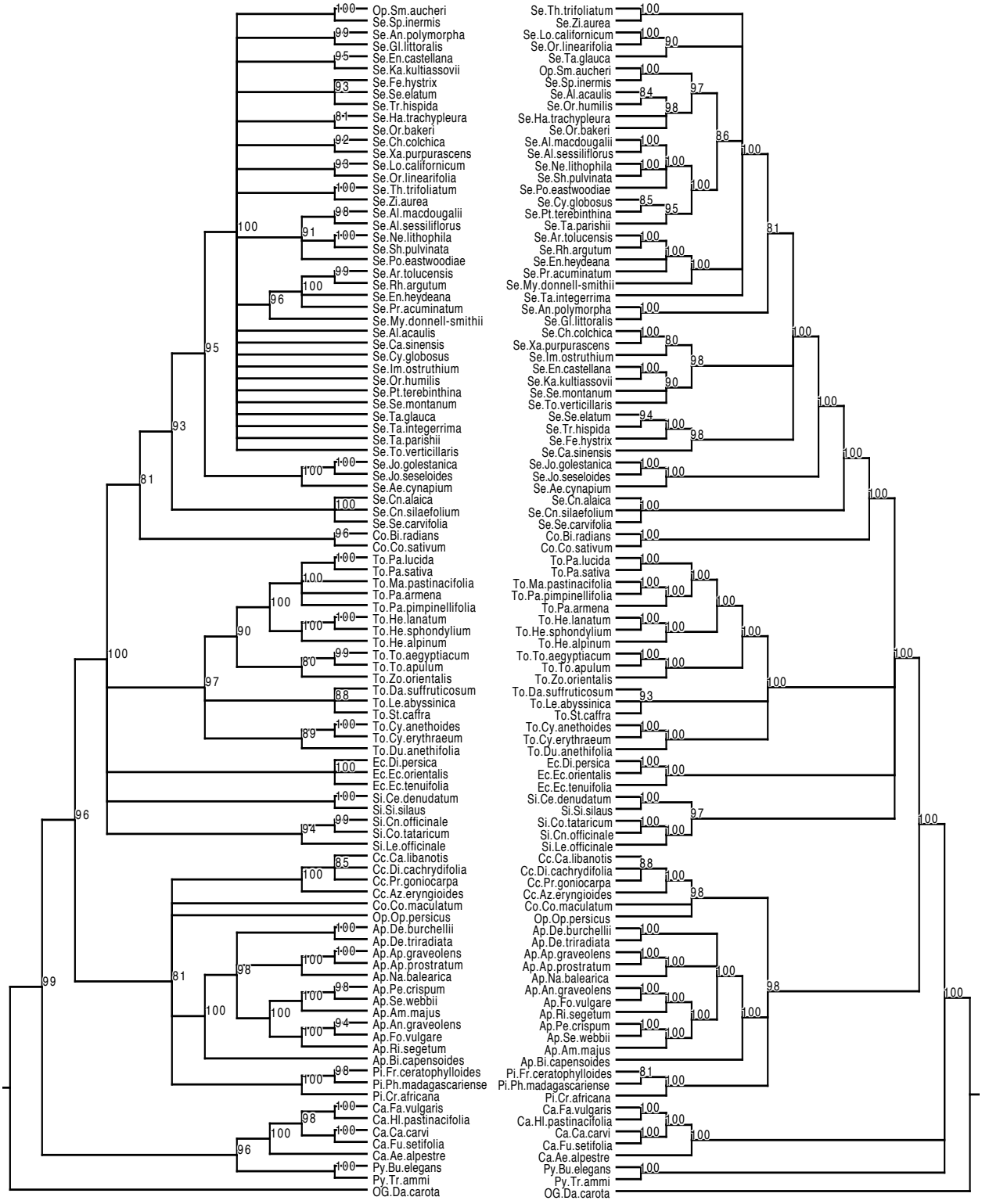


Fig. 4.7 Phylogenies generated from ITS + *psbMD* matrix. The ML tree is on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

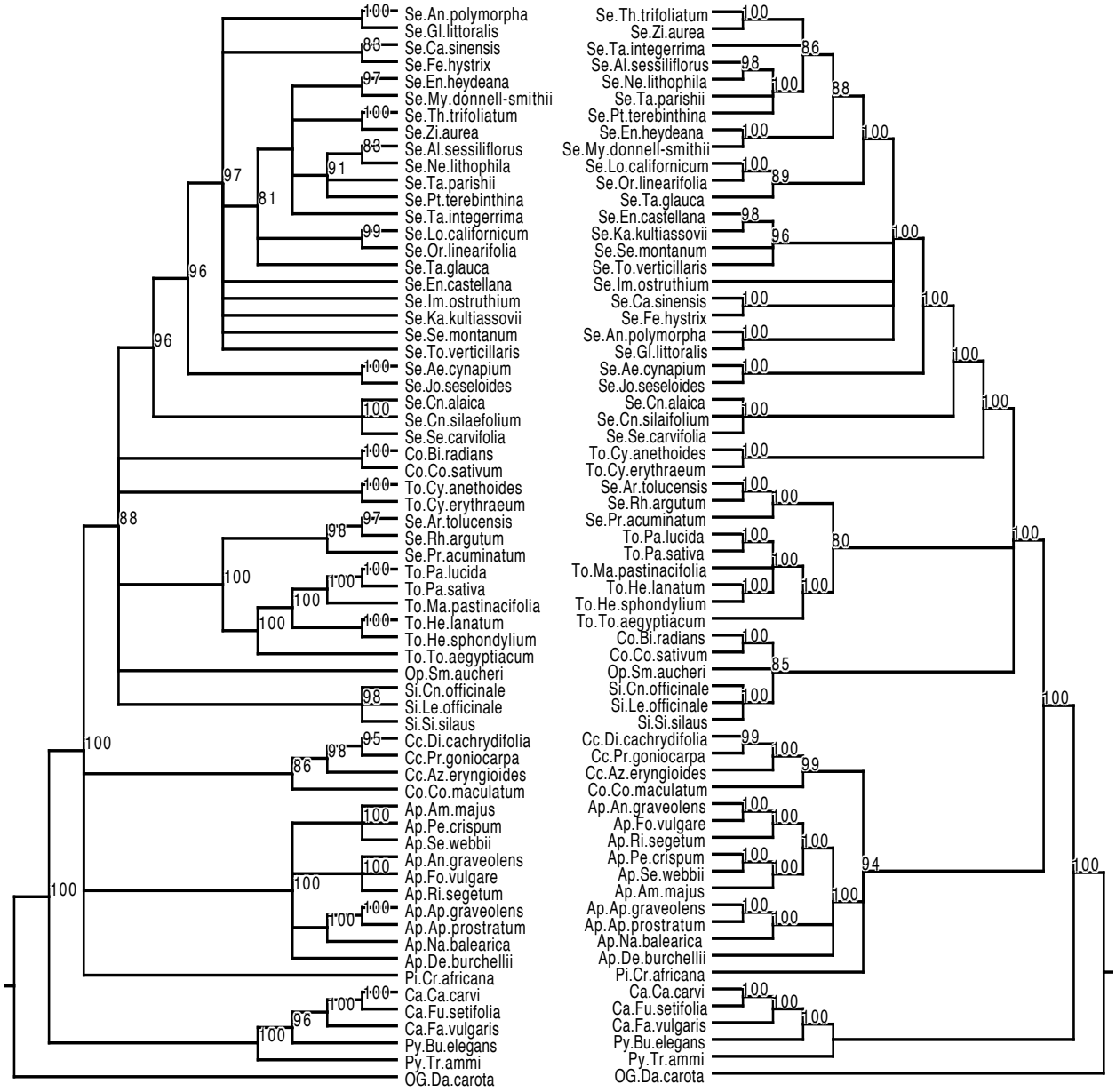


Fig. 4.8 Phylogenies generated from ITS + *PHYA* + *psbMD* matrix. The ML tree is on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

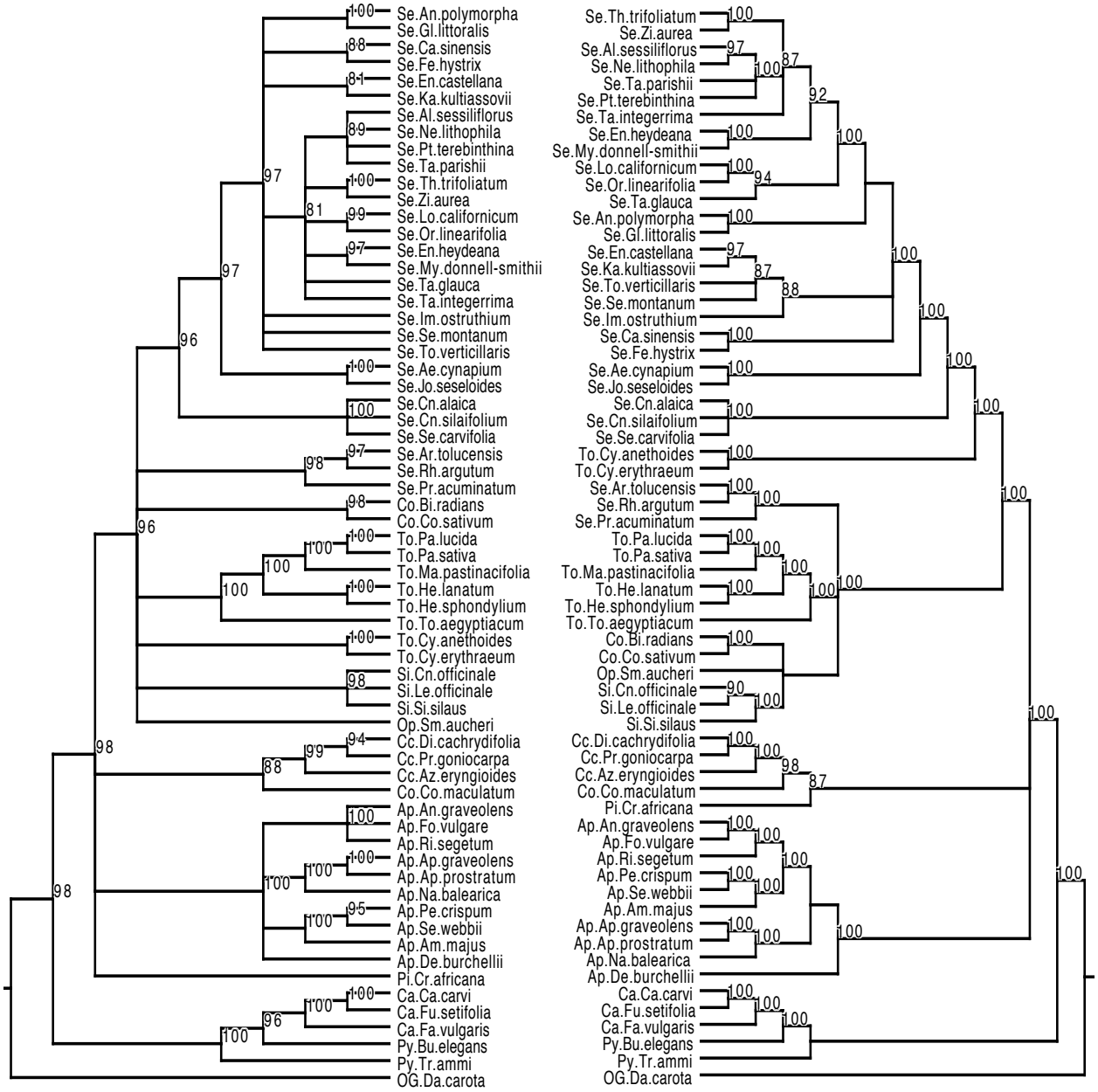


Fig. 4.9 Phylogenies generated from ITS + *PHYA* + *psbMD* + RGCs matrix. The ML tree is on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

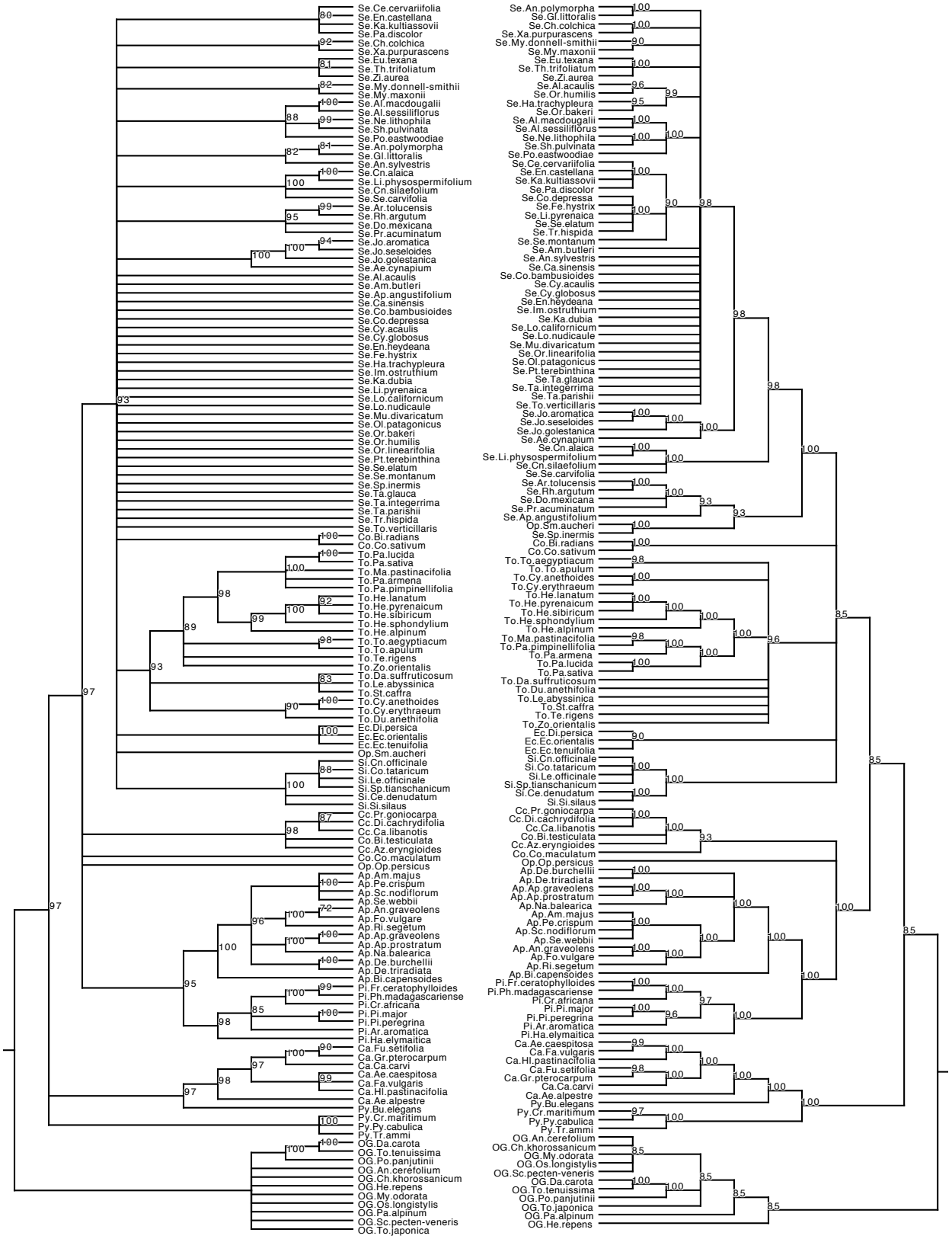


Fig. 4.10 Phylogenies generated from the 143 taxa ITS + *PHYA* + *psbMD* combined matrix. ML tree on the left and Bayesian tree on the right, numbers at nodes are BS and PP respectively.

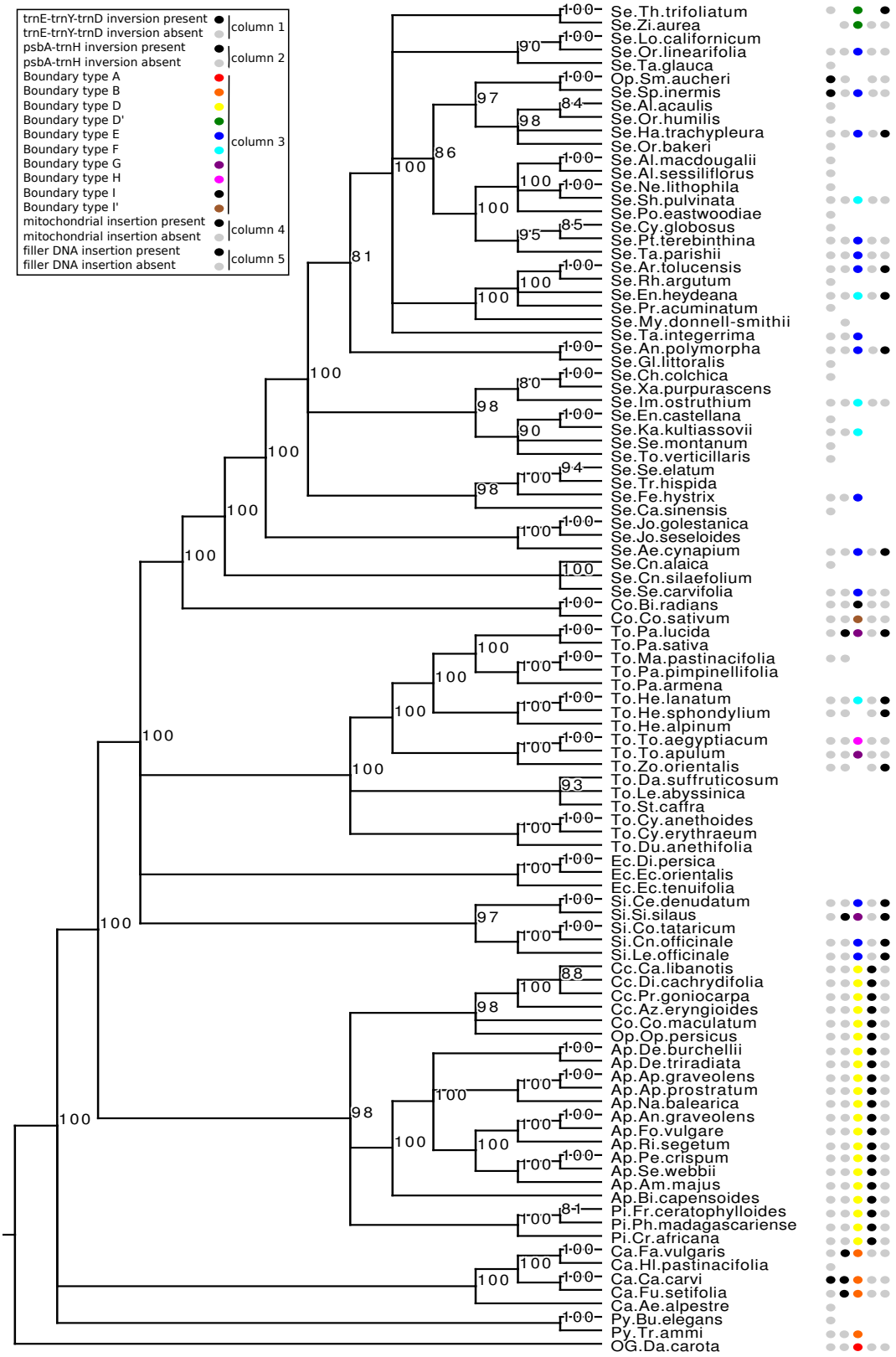


Fig. 4.11 The same Bayesian tree as presented in Fig. 4.7. RGCs are indicated by circles adjacent to taxa. If no circle occurs at a position there is no data for that RGC.

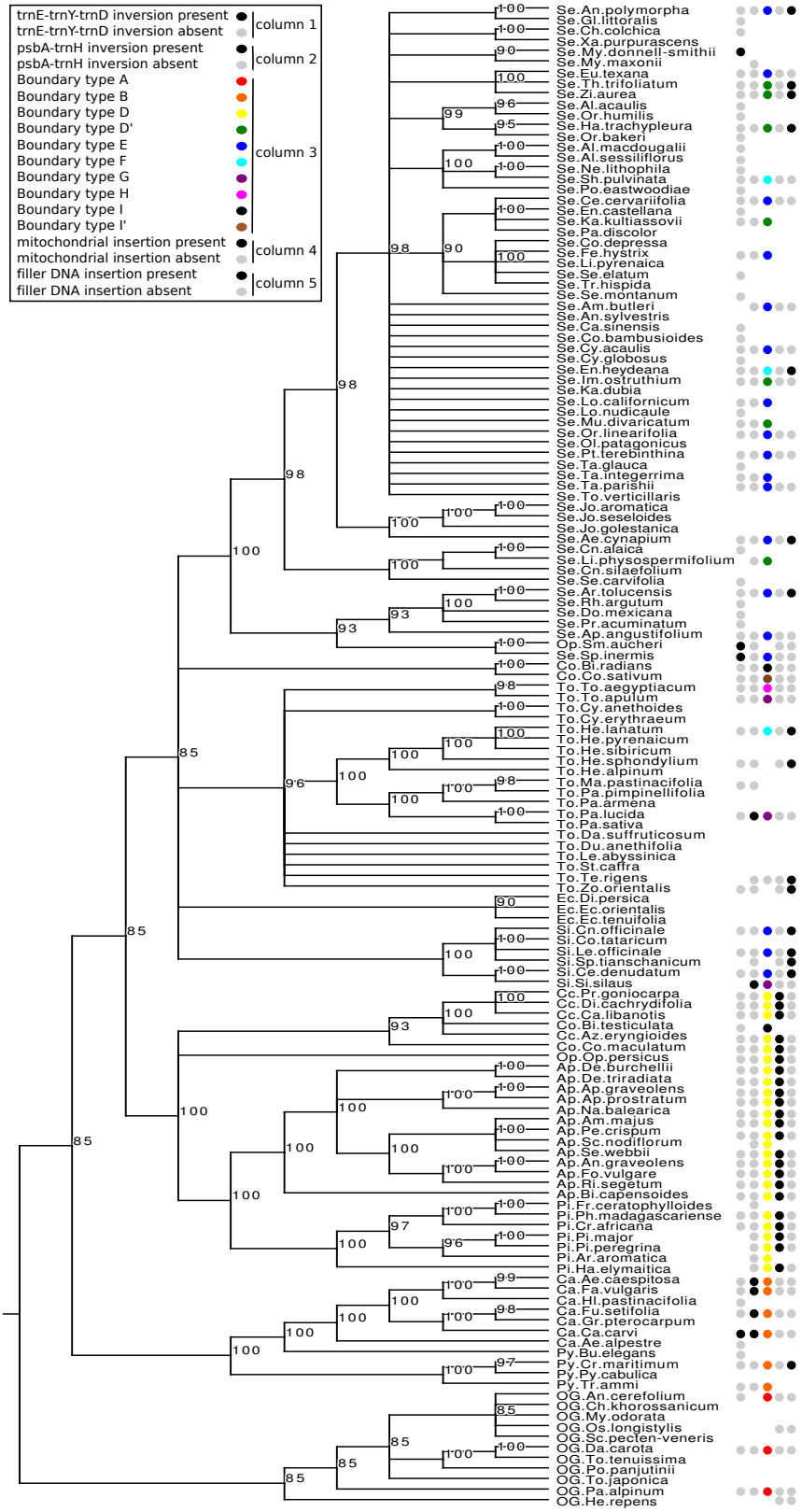


Fig. 4.12 The same Bayesian tree as presented in Fig. 4.10. RGCs are indicated by circles adjacent to taxa. If no circle occurs at a position there is no data for that RGC.

Supplementary Table

Table S4.1 Taxa included in phylogenetic analyses of eight data matrices. Those taxa included in each analysis are presented in Fig. 4.2.

Family	Species	Abbreviation used in phylogenies	Reference/Voucher
Apieae	*Ammi majus* L.	Ap.Am.majus	Downie et al. 1998
Apieae	*Anethum graveolens* L.	Ap.An.graveolens	Downie et al. 1998
Apieae	*Apium graveolens* L.	Ap.Ap.graveolens	Downie et al. 1998
Apieae	*Apium prostratum* Labill.	Ap.Ap.prostratum	Reduron et al. 2009
Apieae	*Billburttia capensoides* Sales and Hedge	Ap.Bi.capensoides	Magee et al. 2009 (MO)
Apieae	*Deverra burchellii* Eckl. & Zeyh.	Ap.De.burchellii	Winter et al. 2008
Apieae	*Deverra triradiata* Hochst. Ex Boiss.	Ap.De.triradiata	Downie et al. 2000
Apieae	*Foeniculum vulgare* Mill.	Ap.Fo.vulgare	Downie et al. 1998; Chapter 3
Apieae	*Naufraga balearica* onstance & Cannon	Ap.Na.balearica	Downie et al. 2000
Apieae	*Petroselinum crispum* (Mill.) Mansf.	Ap.Pe.crispum	Downie et al. 1998
Apieae	*Ridolfia segetum* (L.) Moris	Ap.Ri.segetum	Downie et al. 1998
Apieae	*Sclerosciadium nodiflorum* Coss.	Ap.Sc.nodiflorum	Spalik et al. 2010
Apieae	*Seseli webbii* Coss.	Ap.Se.webbii	Spalik et al. 2004
Apieae	*Stoibrax capense* (Lam.) B.L.Burtt	Ap.St.capense	*Downie* K108
Cachrys	*Azilia eryngioides* (Pau) Hedge & Lamond	Cc.Az.eryngioides	Ajani et al. 2008
Cachrys	*Cachrys libanotis* L.	Cc.Ca.libanotis	Ajani et al. 2008
Cachrys	*Diplotaenia cachrydifolia* Boiss.	Cc.Di.cachrydifolia	Ajani et al. 2008
Cachrys	*Ferulago nodosa* (L.) Boiss.	Cc.Fe.nodosa	*Downie* 3862
Cachrys	*Prangos goniocarpa* (Boiss.) Zohary	Cc.Pr.goniocarpa	Ajani et al. 2008
Careae	*Aegokeras caespitosa* (Sibth. Sm.) Raf.	Ca.Ae.caespitosa	Plunkett and Downie 2000
Careae	*Aegopodium alpestre* Ledeb.	Ca.Ae.alpestre	Downie et al. 1998
Careae	*Aegopodium podagraria* L.	Ca.Ae.podagraria	Danderson, April 20, 2007, Champaign, cultivated, *Downie 3284*
Careae	*Carum carvi* L.	Ca.Ca.carvi	Downie et al. 1998; *Downie 3912*
Careae	*Falcaria vulgaris* Burnh.	Ca.Fa.vulgaris	Downie et al. 1998
Careae	*Fuernrohria setifolia* K.Koch	Ca.Fu.setifolia	Katz-Downie et al. 1999
Careae	*Grammosciadium pterocarpum* Boiss.	Ca.Gr.pterocarpum	Downie et al. 2000
Careae	*Hladnikia pastinacifolia*	Ca.Hl.pastinacifolia	*Gardner 2615*
Conium	*Conium maculatum* L.	Co.Co.maculatum	Downie et al. 1998
Coriandreae	*Bifora radians* M.Bieb.	Co.Bi.radians	Downie et al. 1998
Coriandreae	*Bifora testiculata* (L.) Spreng.	Co.Bi.testiculata	19970503, RBGE
Coriandreae	*Coriandrum sativum* L.	Co.Co.sativum	Downie et al. 1998
Echinophoreae	*Anisosciadium isosciadium var. idumaeum* DC.	Ec.An.isosciadium	Jordan, 13 April 1980, *Frey & Kurschner VO5151* (E); extracted by K. Spalik
Echinophoreae	*Anisosciadium orientale* DC.	Ec.An.orientale	Iran, 50 km from Lar to Jahrom; *Davis and Bokhari 56241* (RBGE E00042061)
Echinophoreae	*Dicyclophora persica* Boiss.	Ec.Di.persica	Downie et al. 2000
Echinophoreae	*Echinophora orientalis* Hedge & Lamond	Ec.Ec.orientalis	Ajani et al. 2008
Echinophoreae	*Echinophora tenuifolia* L.	Ec.Ec.tenuifolia	Downie et al. 2000

Table S4.1 (cont.)

Opopanax	*Opopanax persicus* Boiss. & Heldr.	Op.Op.persicum	Ajani et al. 2008
Opopanax	*Smyrniopsis aucheri* Boiss.	Op.Sm.aucheri	Downie et al. 1998
Pimpinelleae	*Arafoe aromatic* Pimenov & Lavrova	Pi.Ar.aromatica	Downie et al. 1998
Pimpinelleae	*Cryptotaenia africana* Drude	Pi.Cr.africana	Plunkett and Downie 1999; *Douglas 1751*, BYU 313770
Pimpinelleae	*Frommia ceratophylloides* H. Wolff	Pi.Fr.ceratophylloides	Spalik and Downie 2007; MO 2448554
Pimpinelleae	*Haussknechtia elymaitica* Boiss.	Pi.Ha.elymaitica	Ajani et al. 2008
Pimpinelleae	*Phellolophium madagascariense* Baker	Pi.Ph.madagascariense	*Phillipson 2208* (MO 3514162)
Pimpinelleae	*Pimpinella major* (L.) Huds.	Pi.Pi.major	Plunkett & Downie 2000
Pimpinelleae	*Pimpinella peregrina* Lej.	Pi.Pi.peregrina	Downie et al. 1998
Pyramidoptereae	*Bunium elegans* Grossh.	Py.Bu.elegans	Jordan, Ajlun, near the Community College, Lahham and El-Oqlah 9 (Yarmouk Univ. Herb.)
Pyramidoptereae	*Crithmum maritimum* L.	Py.Cr.maritimum	Downie et al. 1998; Downie and Jansen 2015
Pyramidoptereae	*Pyramidoptera cabulica* Boiss.	Py.Py.cabulica	Katz-Downie et al. 1999
Pyramidoptereae	*Schrenkia vaginata* (Ledeb.) Fisch. & C.A.Mey.	Py.Sc.vaginata	*Goloskokov*, 15-Jun-59, RBGE
Pyramidoptereae	*Sison amomum* L.	Py.Si.amomum	France, Val-de-Marne, Créteil, au Mont-Mesly. Reduron 19770711-01
Pyramidoptereae	*Trachyspermum ammi* (L.) Sprague	Py.Tr.ammi	Downie et al. 1998
Selineae	*Aethusa cynapium* L.	Se.Ae.cynapium	Plunkett and Downie 2000
Selineae	*Aletes acaulis* (Torr.) J.M.Coult. Rose	Se.Al.acaulis	Downie et al. 2002
Selineae	*Aletes macdougalii ssp. breviradiatus* W.L.Theob. & C.C. Tseng	Se.Al.macdougalii	(28) #49, Sun 1999 RM trip (=Oreoxix trotteri)
Selineae	*Aletes sessiliflorus* W.L.Theob. & C.C.Tseng	Se.Al.sessiliflorus	(39) #25, Sun 1999 RM trip
Selineae	*Ammoselinum butleri* (Engelm. Ex S.Watson) J.M.Coult. & Rose	Se.Am.butleri	USA, Mississippi, Leflore Co., West of Greenwood, *Cryson 13404* (MO)
Selineae	*Ammoselinum popei* Torr. & A.Gray	Se.Am.popei	USA, Oklahoma, Roger Mills Co., 25 April 2001, *Freeman & Loring 16921* (MO)
Selineae	*Angelica polymorpha* Maxim.	Se.An.polymorpha	Downie et al. 1998
Selineae	*Angelica sylvestris* L.	Se.An.sylvestris	Downie et al. 1998
Selineae	*Apiastrum angustifolium* Nutt. ex Torr. & A.Gray	Se.Ap.angustifolium	USA, California, Riverside Co., Vail Lake area; *Boyd et al. 3848* (MO 4000398)
Selineae	*Arracacia tolucensis* (Kunth) Hemsl.	Se.Ar.tolucensis	*C-2124*, University of California, Berkeley;
Selineae	*Carlesia sinensis* Dunn	Se.Ca.sinensis	Downie et al. 1998
Selineae	*Cervaria cervariifolia* (C.A.Mey.) Pimenov	Se.Ce.cervariifolia	Ajani et al. 2008
Selineae	*Chymsydia colchica* (Albov) Woronow ex Grossh.	Se.Ch.colchica	Downie et al. 1998
Selineae	*Cnidiocarpa alaica* Pimenov	Se.Cn.alaica	Katz-Downie et al. 1999
Selineae	*Cnidium silaifolium* (Jacq.) Simonkai	Se.Cn.silaifolium	Downie et al. 1998
Selineae	*Coaxana bambusioides* Mathias & Constance	Se.Co.bambusioides	*D.E. Breedlove 12248*, 27-VIII-1965, UC-1348337
Selineae	*Cortia depressa* (D.Don) C.Norman	Se.Co.depressa	29; RBGE, 19892739
Selineae	*Cymopterus acaulis* (Pursh) Raf.	Se.Cy.acaulis	50, Vanderhorst 2236

Table S4.1 (cont.)

Selineae	*Cymopterus globosus* S.Watson	Se.Cy.globosus	Downie et al. 1998
Selineae	*Donnellsmithia mexicana* (S.Watson) Mathias & Constance	Se.Do.mexicana	*D. E. Breedlove 36156*, 13–XI–1973, CAS 573904
Selineae	*Enantiophylla heydeana* J.M.Coult. & Rose	Se.En.heydeana	Downie et al. 1998
Selineae	*Endressia castellana* Coincy	Se.En.castellana	Downie et al. 1998
Selineae	*Eurytaenia texana* Torr. & A.Gray	Se.Eu.texana	*Seigler et al. 9834* (ILL)
Selineae	*Ferulopsis hystrix* (Bunge ex Ledeb.) Pimenov	Se.Fe.hystrix	Ajani et al. 2008
Selineae	*Glehnia littoralis var. leiocarpa* (Mathias) B.Boivin	Se.Gl.littoralis	*Halse 1228*, OSU 146791
Selineae	*Harbouria trachypleura* (A.Gray) J.M.Coult. & Rose	Se.Ha.trachypleura	24, Embry 56; (16) #5, Sun 1999 RM trip
Selineae	*Imperatoria ostruthium* L.	Se.Im.ostruthium	Downie et al. 1998
Selineae	*Johrenia aromatic* Rech.f.	Se.Jo.aromatica	Ajani et al. 2008
Selineae	*Johrenia golestanica* Rech.f.	Se.Jo.golestanica	Ajani et al. 2008
Selineae	*Johrenia seseloides* (Hoffm.) Koso-Pol.	Se.Jo.seseloides	Ajani et al. 2008
Selineae	*Kadenia dubia* (Schkuhr) Lavrova & V.N.Tikhom.	Se.Ka.dubia	(13) *Reduron 99160* cult.
Selineae	*Karatavia kultiassovii* (Korovin) Pimenov & Lavrova	Se.Ka.kultiassovii	Katz-Downie et al. 1999
Selineae	*Libanotis pyrenaica* Bourg. ex Nyman	Se.Li.pyrenaica	Spalik et al. 2004
Selineae	*Ligusticum physospermifolium* Albov	Se.Li.physospermifolium	Katz-Downie et al. 1999
Selineae	*Lomatium californicum* (Nutt. ex Torr. & A.Gray) Mathias & Constance	Se.Lo.californicum	Downie et al. 1998
Selineae	*Lomatium nudicaule* (Nutt.) J.M.Coult. & Rose	Se.Lo.nudicaule	2, 8, Hartman 8736
Selineae	*Musineon divaricatum* (Pursh) Nutt.	Se.Mu.divaricatum	Downie et al. 2002
Selineae	*Myrrhidendron donnell-smithii*	Se.My.donnell-smithii	Downie et al. 1998
Selineae	*Myrrhidendron maxonii* J.M.Coult. & Rose	Se.My.maxonii	*B. Hammel 2811*, 5-V-1978, MO-2903476
Selineae	*Neoparrya lithophila* Mathias	Se.Ne.lithophila	Downie et al. 2002
Selineae	*Oligocladus patagonicus* (Speg.) Pérez-Mor.	Se.Ol.patagonicus	Vanni et al 4355 9-1-2000 (CTES)
Selineae	*Oreoxis bakeri* J.M.Coult. & Rose	Se.Or.bakeri	Downie et al. 2002
Selineae	*Oreoxis humilis* Raf.	Se.Or.humilis	Downie et al. 2002
Selineae	*Orogenia linearifolia* S.Watson	Se.Or.linearifolia	Downie et al. 2002
Selineae	*Paraligusticum discolor* (Ledeb.)V.N.Tikhom	Se.Pa.discolor	Downie et al. 1998
Selineae	*Podistera eastwoodiae* (J.M.Coult. & Rose) Mathias & Canstance	Se.Po.eastwoodiae	Downie et al. 2002
Selineae	*Prionosciadium acuminatum* B.L.Rob ex J.M.Coult. & Rose	Se.Pr.acuminatum	Downie et al. 2002
Selineae	*Pteryxia terebinthina var. calcarea* (M.E.Jones) Mathias	Se.Pt.terebinthina	Downie et al. 2002
Selineae	*Rhodosciadium argutum* (Rose) Mathias & Constance	Se.Rh.argutum	Downie et al. 1998
Selineae	*Selinum carvifolia* (L.) L.	Se.Se.carvifolia	Sun et al. 2004
Selineae	*Seseli elatum* Thuill.	Se.Se.elatum	Downie et al. 1998
Selineae	*Seseli montanum* L.	Se.Se.montanum	Downie et al. 1998

Table S4.1 (cont.)

Selineae	*Shoshonea pulvinata* Evert & Constance	Se.Sh.pulvinata	Downie et al. 1998
Selineae	*Spermolepis inermis* (Nutt. ex DC.) Mathias & Constance	Se.Sp.inermis	USA, Illinois, Carroll Co., Savanna Army Depot., Green Island, 30 June 1993, Phillippe et al. 22290 (ILLS)
Selineae	*Taenidia integerrima* (L.) Drude	Se.Ta.integerrima	Downie et al. 1998
Selineae	*Tauschia glauca* (J.M.Coult. & Rose ex Rose) Mathias & Constance	Se.Ta.glauca	Downie et al. 2002
Selineae	*Tauschia parishii* (J.M.Coult. & Rose) J.F.Macbr.	Se.Ta.parishii	Downie et al. 2002
Selineae	*Thaspium trifoliatum* (L.) A.Gray	Se.Th.trifoliatum	Downie et al. 1998
Selineae	*Tommasinia verticillaris* (L.) Bertol.	Se.To.verticillaris	Katz-Downie et al. 1999
Selineae	*Trinia hispida* Hoffm.	Se.Tr.hispida	Ajani et al. 2008
Selineae	*Xanthogalum purpurascens* Avé-Lall.	Se.Xa.purpurascens	Ajani et al. 2008
Selineae	*Zizia aurea* (L.) W.D.J.Koch	Se.Zi.aurea	Downie et al. 1998
Sinodielsia	*Cenolophium denudatum* (Fisch. ex Hornem.) Tutin	Si.Ce.denudatum	Valiejo-Roman et al. 1998
Sinodielsia	*Cnidium officinale* Makino	Si.Cn.officinale	Downie et al. 1998
Sinodielsia	*Conioselinum tataricum* Hoffm.	Si.Co.tataricum	Downie et al. 1998
Sinodielsia	*Levisticum officinale* W.D.J.Koch	Si.Le.officinale	Downie et al. 1998
Sinodielsia	*Silaum silaus* (L.) Schinz & Thell.	Si.Si.silaus	Reduron specimens, March 14, 2002 [probably France, Bas-Rhin, between Herbsheim et Boofzheim, 14 August 2001, Reduron (Hb. Reduron)]; (1) UIUC 94204, greenhouse Room 1513, fresh leaf material
Sinodielsia	*Sphaenolobium tianschanicum* (Korovin) Pimenov	Si.Sp.tianschanicum	Katz-Downie et al. 1999
Tordylieae	*Cymbocarpum anethoides* DC.	To.Cy.anethoides	Ajani et al. 2008
Tordylieae	*Cymbocarpum erythraeum* Bioss.	To.Cy.erythraeum	Ajani et al. 2008
Tordylieae	*Dasispermum suffruticosum* (P.J.Bergius) B.L.Burtt	To.Da.suffruticosum	Ajani et al. 2008
Tordylieae	*Ducrosia anethifolia* (DC.) Boiss.	To.Du.anethifolia	Ajani et al. 2008
Tordylieae	*Heracleum alpinum* Siev.	To.He.alpinum	Ajani et al. 2008
Tordylieae	*Heracleum lanatum* Michx.	To.He.lanatum	Downie et al. 1998
Tordylieae	*Heracleum pyrenaicum* Lam.	To.He.pyrenaicum	Ajani et al. 2008
Tordylieae	*Heracleum sibiricum*	To.He.sibiricum	Reduron 4 Aug 2000
Tordylieae	*Heracleum sphondylium* L.	To.He.sphondylium	Downie et al. 1998
Tordylieae	*Lefebvrea abyssinica* A.Rich.	To.Le.abyssinica	Willis 168 4/4/2000
Tordylieae	*Malabaila pastinacaefolia* Boiss. & Balansa	To.Ma.pastinacifolia	Turkey, B6: Kayseri, Pinarbasi-Gurun arasi, 5 Km, 1550m, 10.07.2000, *A. Duran 5498*, Y.Menemen & M. Sagiroglu (ADO) (tube 2)
Tordylieae	*Malabaila secacul* (Mill.) Boiss.	To.Ma.secacul	Jordan, University of Science and Technology, *Lahham 26 (*Yarmouk U. Herb.) *Lee 253*
Tordylieae	*Pastinaca armena* Fisch. & C.A.Mey.	To.Pa.armena	Katz-Downie et al. 1999
Tordylieae	*Pastinaca lucida* L.	To.Pa.lucida	Ajani et al. 2008
Tordylieae	*Pastinaca pimpinellifolia* Bory & Chaub.	To.Pa.pimpinellifolia	Ajani et al. 2008
Tordylieae	*Pastinaca sativa* Thomas ex DC.	To.Pa.sativa	Downie et al. 1998

Table S4.1 (cont.)

Tordylieae	*Stenosemis caffra* Sond.	To.St.caffra	Calviño et al. 2006
Tordylieae	*Tetrataenium rigens* (DC.) Manden.	To.Te.rigens	Downie et al. 1998 as *Heracleum rigens*
Tordylieae	*Tordylium apulum* L.	To.To.apulum	Ajani et al. 2008
Tordylieae	*Tordylium aegyptiacum var. palaestinum* (Zohary) Zohary	To.To.aegyptiacum	Downie et al. 1998
Tordylieae	*Zosima orientalis* Hoffm.	To.Zo.orientalis	Ajani et al. 2008
Outgroup	*Anthriscus cerefolium* (L.) Hoffm.	OG.An.cerefolium	Downie and Jansen 2015
Outgroup	*Chaerophyllum khorossanicum* Czerniak. ex Schischk.	OG.Ch.khorossanicum	Valiejo-Roman (DNA #892)
Outgroup	*Daucus carota* L.	OG.Da.carota	Ruhlman et al. 2006
Outgroup	*Daucus carota subsp. drepanensis* (Arcang.) Heywood	OG.Da.carota	Peery, Spring 2010, Urbana, IL, cultivated from seeds;
Outgroup	*Helosciadium repens* Syme ex F.W.Schultz	OG.He.repens	Winter 2008
Outgroup	*Myrrhis odorata* (L.) Scop.	OG.My.odorata	Downie et al. 2002
Outgroup	*Osmorhiza longistylis* (Torr.) DC.	OG.Os.longistylis	Downie et al. 2002
Outgroup	*Pachypleurum alpinum* Ledeb.	OG.Pa.alpinum	Dave Murray from Alaska (Collected from Russia)
Outgroup	*Polylophium panjutinii* Manden. & Schischk.	OG.Po.panjutinii	Ajani et al. 2008
Outgroup	*Scandix pecten-veneris* (L.)	OG.Sc.pecten-veneris	Downie et al. 1998
Outgroup	*Torilis japonica* (Houtt.) DC.	OG.To.japonica	Downie et al. 2001
Outgroup	*Tornabenea tenuissima* (Chev.) O.E.Erikss.	OG.To.tenuissima	Spalik and Downie 2007