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

Source: Systematic Botany, 40(1):336-351.
Published By: The American Society of Plant Taxonomists
URL: http://www.bioone.org/doi/full/10.1600/036364415X686620

Systematic Botany (2015), 40(1): pp. 336–351

DOI 10.1600/036364415X686620
Date of publication February 12, 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

Stephen R. Downie[1,4] and Robert K. Jansen[2,3]

[1]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, U. S. A.
[2]Department of Integrative Biology, The University of Texas at Austin, Austin, Texas 78712, U. S. A.
[3]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, P. O. Box 80141, Jeddah 21589, Saudi Arabia
[4]Author for correspondence (sdownie@life.illinois.edu)

Communicating Editor: Mark P. Simmons

Abstract—Previous mapping studies have revealed that the frequency and large size of inverted repeat junction shifts in Apiaceae plastomes are unusual among angiosperms. To further examine plastome structural organization and inverted repeat evolution in the Apiales (Apiaceae + Araliaceae), we have determined the complete plastid genome sequences of five taxa, namely *Anthriscus cerefolium* (154,719 base pairs), *Crithmum maritimum* (158,355 base pairs), *Hydrocotyle verticillata* (153,207 base pairs), *Petroselinum crispum* (152,890 base pairs), and *Tiedemannia filiformis* subsp. *greenmanii* (154,737 base pairs), and compared the results obtained to previously published plastomes of *Daucus carota* subsp. *sativus* and *Panax schin-seng*. We also compared the five Apiaceae plastomes to identify highly variable noncoding loci for future molecular evolutionary and systematic studies at low taxonomic levels. With the exceptions of *Crithmum* and *Petroselinum*, which each demonstrate a ~1.5 kilobase shift of its LSC-IR_B junction (J_{LB}), all plastomes are typical of most other non-monocot angiosperm plastid DNAs in their structural organization, gene arrangement, and gene content. *Crithmum* and *Petroselinum* also incorporate novel DNA in the LSC region adjacent to the LSC-IR_A junction (J_{LA}). These insertions (of 1,463 and 345 base pairs, respectively) show no sequence similarity to any other region of their plastid genomes, and BLAST searches of the *Petroselinum* insert resulted in multiple hits to angiosperm mitochondrial genome sequences, indicative of a mitochondrial to plastid transfer of DNA. A comparison of pairwise sequence divergence values and numbers of variable and parsimony-informative alignment positions (among other sequence characteristics) across all introns and intergenic spacers >150 base pairs in the five Apiaceae plastomes revealed that the *rpl32-trnL*, *trnE-trnT*, *ndhF-rpl32*, *5′rps16-trnQ*, and *trnT-psbD* intergenic spacers are among the most fast-evolving loci, with the *trnD-trnY-trnE-trnT* combined region presenting the greatest number of potentially informative characters overall. These regions are therefore likely to be the best choices for molecular evolutionary and systematic studies at low taxonomic levels. Repeat analysis revealed direct and inverted dispersed repeats of 30 base pairs or more that may be useful in population-level studies. These structural and sequence analyses contribute to a better understanding of plastid genome evolution in the Apiales and provide valuable new information on the phylogenetic utility of plastid noncoding loci, enabling further molecular evolutionary and phylogenetic studies on this economically, ecologically, and medicinally important group of flowering plants.

Keywords—Apiaceae, Araliaceae, intergenic spacer regions, intracellular transfer, plastid DNA.

The plastid genomes of the majority of photosynthetic angiosperms are highly conserved in structural organization, gene arrangement, and gene content (Palmer 1991; Raubeson and Jansen 2005; Wicke et al. 2011; Jansen and Ruhlman 2012; Ruhlman and Jansen 2014). Their hallmark is the presence of two large duplicate regions in reverse orientation known as the inverted repeat (IR), which separate the remainder of the genome into large single copy (LSC) and small single copy (SSC) regions. Variation in size of this molecule is due most typically to the expansion or contraction of the IR into or out of adjacent single-copy regions and/or changes in sequence complexity due to insertions or deletions of novel sequences. Of the two equimolar structural isomers existing for plastid DNA (ptDNA; Palmer 1983), the structure most commonly presented follows the convention used for *Nicotiana tabacum* L. (tobacco) in which one copy of the IR is flanked by genes *ycf1* and *trnH*-GUG (and is designated as IR_A) and the other copy is flanked by genes *rps19* and *ndhF* (and is designated as IR_B; Shinozaki et al. 1986; Yukawa et al. 2005). The junctions between the LSC region and each of these IR copies are designated as J_{LA} (LSC/IR_A) and J_{LB} (LSC/IR_B), and the junctions flanking the SSC region are designated as J_{SA} (SSC/IR_A) and J_{SB} (SSC/IR_B; Shinozaki et al. 1986). In most non-monocot angiosperm ptDNAs, J_{LB} lies within the ribosomal protein *S10* operon in a more or less fixed position within or near the *rps19* gene and J_{LA} lies just downstream of LSC gene *trnH*-GUG. The IRs of angiosperm plastomes can fluctuate greatly in size, although small contractions and expansions of <100 base pairs (bp) are most frequent, and the positions of all four IR/single-copy junctions can vary even among closely related species (Goulding et al. 1996; Plunkett and Downie 2000). Large IR expansions (>1,000 bp) occur less frequently and outnumber large contractions (Plunkett and Downie 2000; Raubeson and Jansen 2005; Hansen et al. 2007). Because major changes in position of IR junctions can accompany structural rearrangements elsewhere in the plastid genome (Palmer 1991; Boudreau and Turmel 1995; Aii et al. 1997; Cosner et al. 1997; Perry et al. 2002; Chumley et al. 2006; Haberle et al. 2008; Guisinger et al. 2011; Wicke et al. 2011), the availability of sequence data for these genomes can help elucidate the mechanisms responsible for these large-scale IR junction shifts and other major genomic changes.

Previous ptDNA restriction site mapping studies have shown that Apiaceae (Umbelliferae) exhibit unprecedented variation in position of their LSC/IR boundary regions (Palmer 1985a; Plunkett and Downie 1999, 2000). While most umbellifer species surveyed possess a J_{LB} indistinguishable from that of *Nicotiana tabacum* and the vast majority of other eudicots, at least one expansion and seven different contractions of the IR relative to the *N. tabacum* J_{LB} were detected in 55 species, each ranging in size from ~1–16 kilobase pairs (kb; Plunkett and Downie 2000). As examples, all examined members of the "*Aegopodium* group," representatives of tribes Careae and Pyramidoptereae (Downie et al. 2010), have a ~1.1 kb expansion of J_{LB} relative to that of *N. tabacum*. Careae and Pyramidoptereae are monophyletic sister groups (Banasiak et al. 2013), indicating that IR junction shifts have the potential to demarcate major clades within the family. *Coriandrum sativum* L. and *Bifora radians* M. Bieb., both of tribe Coriandreae, have the most contracted IRs, approximately

16.1 kb smaller than that of *N. tabacum*. *Coriandrum* also has a ~5.7 kb insertion of unknown composition in the vicinity of the 16S rRNA gene near the terminus of the IR (Plunkett and Downie 2000), a region subsequently identified through entire plastome sequencing as a duplication of genes *trnH*-GUG and *psbA* (Peery et al. 2006). Several of these junction shifts in Apiaceae are parsimony-informative and all are restricted to the "apioid superclade" of Apiaceae subfamily Apioideae, a large, morphologically heterogeneous group of umbellifers comprising 14 tribes and other major clades of dubious relationship (Plunkett and Downie 1999, 2000; Downie et al. 2010). Plastid genome sequencing has also revealed unique, noncoding DNA in the region bounded by J_{LA} and *trnH*-GUG in several related species of Apiaceae that does not show significant similarity to any other known ptDNA sequence (Peery et al. 2007, 2011), and BLAST searches of these sequences suggested a mitochondrial to plastid transfer of DNA (Peery et al. 2011). Subsequent studies have confirmed other instances of mitochondrial DNA transfer into the plastome of Apiaceae (Goremykin et al. 2009; Iorizzo et al. 2012a,b). Thus, the Apiaceae provide a model system in which to study the mechanisms leading to large-scale expansions and contractions of the IR, with concomitant novel insertions near J_{LA} either the cause or consequence of these major IR junction shifts.

The Apiaceae have been the subject of numerous phylogenetic studies using ptDNA and while these studies have contributed greatly to a broad understanding of its evolutionary history, uncertainties remain with regard to many species-level relationships, particularly those of the largest genera within the "apioid superclade" (reviewed in Downie et al. 2001, 2010). These uncertainties are the result of a paucity of variable DNA sequence characters obtained from the relatively few plastid loci examined. The earliest of these studies focused on gene (*matK* exons, *rbcL*) and intron (*rpl16*, *rps16*, *rpoC1*) sequences (Downie et al. 1996, 2000; Plunkett et al. 1996a,b; Downie and Katz-Downie 1999), whereas subsequent studies examined primarily the five noncoding regions between genes *trnK* and *psbI*, and intergenic spacer regions *trnH-psbA*, *trnD-trnT*, and *trnF-trnL-trnT* (Calviño and Downie 2007; Downie et al. 2008; Degtjareva et al. 2009; Nicolas and Plunkett 2009; Sun and Downie 2010). Rates of nucleotide change between these regions, as well as between other plastid loci, can vary tremendously within a given taxonomic group (Shaw et al. 2005); therefore, it is useful to investigate the relative utility of previously unexplored or underutilized noncoding regions among Apiaceae plastomes for their potential in resolving phylogenetic relationships at low taxonomic levels within the family. The most variable noncoding regions uncovered can also be considered for intraspecific phylogeographic and population genetic studies.

To characterize plastome structural organization and IR evolution in the Apiales, we report the complete genome sequences of *Anthriscus cerefolium* (L.) Hoffm. (chervil), *Crithmum maritimum* L. (sea samphire), *Hydrocotyle verticillata* Thunb. (whorled marshpennywort), *Petroselinum crispum* (Mill.) Fuss (parsley), and *Tiedemannia filiformis* (Walter) Feist & S. R. Downie subsp. *greenmanii* (Mathias & Constance) Feist & S. R. Downie (giant water cowbane), and compare the results obtained to previously published plastomes of *Daucus carota* L. subsp. *sativus* Schübl. & G. Martens (domesticated carrot; Ruhlman et al. 2006) and *Panax schin-seng* T. Nees (Korean ginseng; Kim and Lee 2004). *Anthriscus* and *Daucus* represent tribe Scandiceae subtribes Scandicinae and Daucinae, respectively, *Crithmum* is classified in tribe Pyramidoptereae, *Tiedemannia* represents tribe Oenantheae (Feist et al. 2012), and *Petroselinum* is placed in tribe Apieae (Downie et al. 2010). *Crithmum* and *Petroselinum* are members of the "apioid superclade" (Downie et al. 2010), and *Hydrocotyle* and *Panax* are classified in the closely related family Araliaceae (Nicolas and Plunkett 2009). To identify the most variable noncoding loci, we compare the plastomes of the Apiaceae genera *Anthriscus*, *Crithmum*, *Daucus*, *Petroselinum*, and *Tiedemannia*. The seven species considered herein represent disparate clades throughout the Apiales; they also represent plastomes exhibiting major IR expansions and contractions (as assessed previously through restriction site mapping studies), as well as those possessing IR junctions in positions very similar to those of many other unrearranged angiosperm plastid genomes. This information on Apiales plastome structural organization, IR junction shifts, and noncoding loci variability will help guide subsequent systematic and evolutionary studies of this economically, ecologically, and medicinally important group of flowering plants.

Materials and Methods

Plastid Isolation, Amplification, and Sequencing—Fresh leaves of *Anthriscus* and *Petroselinum* were obtained at a local grocery store, and living plants of *Crithmum* (UCONN 198501242), *Hydrocotyle* (UCONN 198501441), and *Tiedemannia* (as *Oxypolis greenmanii* Mathias & Constance; UCONN 200202456) were obtained from the Ecology and Evolutionary Biology Greenhouse Collections at the University of Connecticut. Leaves from multiple individual plants were combined to provide enough material for the plastid isolations. Vouchers are deposited at the University of Illinois herbarium (ILL).

Approximately 5–10 g of fresh, young leaf tissue was used in each plastid genome isolation. Plastids were isolated using the sucrose step-gradient centrifugation method (Palmer 1986) and the extraction procedures outlined in Jansen et al. (2005). The plastid pellet was resuspended in wash buffer to a final volume of 2 ml. Whole plastid genome amplification was performed with rolling circle amplification (RCA) using a REPLI-g midi kit (Qiagen Inc., Valencia, California). The manufacturer's protocol for amplification of genomic DNA from cells was followed, with the following modification to improve ptDNA amplification (M. Guisinger, unpubl.): 1.5 µl of alkaline lysis solution (10 µl of solution A in Jansen et al. 2005, activated with 1.0 µl 1 M DTT) was added to 1.0 µl of isolated plastids and 4.0 µl of PBS and incubated on ice for 10 min. Subsequently, 3.5 µl of stop solution was added to this mixture. Each RCA reaction consisted of 0.5–1.0 µl of lysate, 23.0–23.5 µl of REPLI-g midi reaction buffer, and 1.0 µl of REPLI-g midi DNA polymerase. The reaction was incubated at 30°C for 16 hr and terminated at 65° for 3 min. After confirmation that the RCA was successful by running out 2 µl of the product on a minigel, another 2 µl of RCA product was then digested with 20 units of *EcoRI* in a 20-µl reaction, run out on a 1% agarose gel with a low DNA mass ladder, and stained and visualized with ethidium bromide. The bright, discrete banding patterns relative to the nuclear DNA background ensured that the amplification product contained sufficient quality and quantity of ptDNA to proceed with genome sequencing.

The amplified genomic DNA was sent to the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign, sheared by nebulization, subjected to 454 library preparation, and then sequenced using the Roche/454 Genome Sequencer (GS) FLX platform (454 Life Sciences Corp., Branford, Connecticut), following the manufacturer's protocol. For two genomes that were sequenced at two later dates (*Hydrocotyle*, *Tiedemannia*), GS FLX Titanium series reagents and plates were used.

Genome Assembly, Finishing, and Annotation—The obtained nucleotide sequence reads were assembled de novo into contiguous sequences ("contigs") using Newbler 2.3, a DNA sequence assembly software package designed specifically for the 454 GS FLX sequencing platform, using default settings (90% minimum overlap identity). To identify contigs that were plastid sequences, all contigs were searched against NCBI's Nucleotide Collection database using BLAST's megablast option (Altschul et al. 1990). Putative gene identifications in each plastid contig were then made

using the annotation program DOGMA (Dual Organellar GenoMe Annotator; Wyman et al. 2004). These contigs were then compared with previously published reference sequences from the complete plastid genomes of *Daucus carota* subsp. *sativus* (GenBank accession number DQ898156) and *Panax schin-seng* (AY582139) and assembled using Sequencher 4.7 (Gene Codes Corp., Ann Arbor, Michigan). The lack of genome rearrangements, low frequency of repeated sequences, and availability of gene mapping data for three of the five Apiales plastid genomes (Plunkett and Downie 1999; Lee and Downie 2000) greatly facilitated assemblies. Gaps between contigs were filled in (or the contigs determined to abut, because there was no missing sequence between adjoining contigs) by identifying sequences at both ends of adjacent contigs and using these regions to pull out high-quality individual sequence reads from the 454 FASTA-formatted output. A minimum of 20 overlapping sequence reads in each direction was aligned using ClustalX (Jeanmougin et al. 1998) and their consensus sequence used to link the contigs. This same method was used to identify IR-single copy junctions. Sequences between a dozen pairs of contigs were confirmed by designing PCR primer pairs (18–20 bp in size) flanking gaps and Sanger sequencing using an ABI Prism BigDye Terminator v3.1 ready reaction cycle sequencing kit (Applied Biosystems, Foster City, California) and an ABI 3730xl DNA Analyzer. Because breaks between contigs generally occurred at or near IR-single copy junctions and in noncoding regions characterized by homopolymer runs, this Sanger sequencing verified IR boundaries and confirmed homopolymeric repeats in these regions. Genomic data for three taxa (*Anthriscus, Petroselinum*, *Crithmum*) were also assembled into contigs using MIRA (Mimicking Intelligent Read Assembly), under slightly less stringent alignment settings (Chevreux et al. 2000). DOGMA was implemented to assist in annotating all genes and to identify coding sequences, tRNAs, and rRNAs using the plastid/bacterial genetic code; start and stop codons were added manually, as were small exons. Exon-intron boundaries were determined by detailed comparison with other annotated genomes and individual gene sequences, especially that of *N. tabacum* (Shinozaki et al. 1986). The identification of tRNAs was confirmed using tRNAscan-SE (Lowe and Eddy 1997; Schattner et al. 2005). Following convention and for annotation purposes, numbering begins with the first base of the LSC region following J_{LA} and proceeds counterclockwise around the circular genome.

Plastome and Sequence Comparisons—Plastid genome sequences (minus one copy of the IR) of *Anthriscus, Crithmum, Petroselinum*, and *Tiedemannia* were compared to the reference *Daucus* using MultiPipMaker (Schwartz et al. 2003). The options "Search both strands" and "Show all matches" were chosen. REPuter (Kurtz and Schleiermacher 1999) was used to identify and locate forward (direct) and inverted (palindromic) repeat sequences. Searches of plastomes (upon the removal of one copy of the IR) were carried out with minimal repeat size ≥30 bp and a Hamming distance of 3. Because REPuter overestimates the number of repetitive elements in a given sequence by recognizing nested or overlapping repeats within a given region containing multiple repeats (Chumley et al. 2006; Timme et al. 2007), all redundant repeats were identified manually and excluded. Circular genomic maps were constructed using OGDraw (Organellar Genome Draw; Lohse et al. 2007).

The "Extract Sequences" option of DOGMA was used to extract intergenic spacers and introns from the annotated sequences in order to construct data matrices to identify highly divergent regions. The more divergent Araliaceae representatives, *Panax* and *Hydrocotyle*, were excluded from these analyses. A total of 70 noncoding ptDNA regions were compared, representing all intergenic spacers >150 bp in *Daucus* and introns from both LSC and SSC regions. Coding regions were excluded because they tend to provide fewer variable characters than noncoding regions; similarly, all regions contained in the IR were also excluded because many of them diverge at slower rates compared to sequences located in their adjacent single-copy regions (Wolfe et al. 1987; Perry and Wolfe 2002; Kim and Lee 2004). Four regions of combined noncoding and coding loci were also compared, as these loci are small enough to be PCR-amplified and sequenced together (i.e. *trnD-trnY-trnE-trnT*; *trnS-psbZ-trnG-trnfM*; *psbB-psbT-psbN-psbH*; *rpl36-infA-rps8-rpl14*). Sequences from the five Apiaceae genomes were aligned using ClustalX and, if necessary, manually adjusted using PAUP* 4.0b10 (Swofford 2002) to produce an alignment with the fewest number of changes (indels or nucleotide substitutions). The comparison of these 74 regions included numbers of constant, uninformative, parsimony-informative, and variable alignment positions, range of sequence divergence in pairwise comparisons (uncorrected "p" distance), and numbers of total and parsimony-informative indels. The number of nucleotide substitutions (uninformative + informative alignment positions) and indels for each ptDNA region was tallied; this number divided by the total aligned length resulted in the proportion of observed mutational events (or percent variability) for each region compared (Shaw et al. 2005).

Results

Plastome Assemblies—Newbler assemblies of the nucleotide sequence reads resulted in 54 (*Anthriscus*) to 333 (*Hydrocotyle*) non-redundant contigs, a portion of each of which blasted to published plastid genomes. Average read lengths varied from about 225 bp (*Anthriscus, Crithmum, Petroselinum*) to 336 and 381 bp (in *Hydrocotyle* and *Tiedemannia*, respectively). Average sequencing depth ranged from about 90× (*Anthriscus, Tiedemannia*) to 350× (*Hydrocotyle*). A comparison of assemblies of the five Apiales plastid genomes is presented in Table 1.

Plastid contigs assembled using MIRA were substantially longer than the Newbler assemblies and, in most cases, confirmed the consensus sequences that were assembled manually from the individual 454 reads. As one example, a MIRA re-assembly of *Crithmum* under the "accurate setting" resulted in one 68-kb contig spanning many AT-rich intergenic spacers, whereas the Newbler-generated assembly resulted in 18 smaller contigs covering the same ptDNA region. In general, Newbler generated more contigs by breaking at homopolymer runs, even though high-quality reads existed at the ends of all contigs. In only a few instances did the MIRA assemblies differ from those constructed using Newbler and these were all single nucleotide differences in regions of poly-A/T uncertainties in intergenic spacers.

Plastome Structural Characteristics—With the exceptions of *Crithmum* and *Petroselinum*, where J_{LB} and J_{LA} differ and unique DNA has been incorporated into the LSC region adjacent to the J_{LA} boundary, all plastomes possess the ancestral angiosperm plastid genome organization. These genomes ranged in size from 152,890 bp (*Petroselinum*) to 158,355 bp (*Crithmum*), with size extremes caused by contraction or expansion of the IR at J_{LB} (Table 2). All plastomes except *Crithmum* and *Petroselinum* share identical complements of coding genes, each containing 30 unique tRNA genes, four unique rRNA genes, and 79 unique protein-coding genes. Allowing for duplication of genes in the IR, there are a total of 130 complete, predicted coding regions in these plastomes. Eighteen genes have introns, 16 of which contain a single intron, while two (*ycf3, clpP*) have two introns. All maintain

Table 1. Comparison of assemblies of five Apiales plastid genomes.

Feature	*Anthriscus cerefolium*	*Crithmum maritimum*	*Hydrocotyle verticillata*	*Petroselinum crispum*	*Tiedemannia filiformis* subsp. *greenamnii*
Total no. of non-redundant contigs	54	142	333	162	219
No. of plastid contigs	13	50	63	14	8
Plastid contig size (bp)	241–68,210	120–19,171	110–26,905	242–75,583	3,130–72,953
Average plastid contig size (bp)	9,891	2,596	2,049	9,144	16,019
No. of reads assembled	53,255	129,820	131,886	65,608	29,867

TABLE 2. Comparison of major structural features of seven Apiales plastid genomes. Features of *Panax schin-seng* are provided by Kim and Lee (2004) and those of *Daucus carota* are provided by Ruhlman et al. (2006).

Feature	*Panax schin-seng*	*Hydrocotyle verticllata*	*Daucus carota* subsp. *sativus*	*Anthriscus cerefolium*	*Tiedemania filiformis* subsp. *greenmanii*	*Petroselinum crispum*	*Crithmum maritimum*
GenBank accession number	AY582139	HM596070	DQ898156	GU456628	HM596071	HM596073	HM596072
Entire plastid genome size (bp)	156,318	153,207	155,911	154,719	154,737	152,890	158,355
Large single copy region size (bp)	86,106	84,352	84,242	84,774	84,585	86,116	85,230
Small single copy region size (bp)	18,070	18,739	17,567	17,551	17,140	17,508	17,139
Inverted repeat size (bp)	26,071	25,058	27,051	26,197	26,506	24,633	27,993
A + T content (%)	61.9	62.4	62.3	62.6	62.7	62.2	62.4
Number of repeats ≥ 30 bp (maximum size in bp)	11 (57)	9 (41)	14 (70)	12 (41)	20 (39)	18 (41)	29 (71)
Number of homopolymers ≥ 7 bp (A or T; G or C)	270 (244; 26)	283 (257; 26)	303 (281; 22)	320 (293; 27)	291 (262; 29)	298 (271; 27)	326 (300; 26)
Homopolymer length (bp) (A or T; G or C)	7–13 (7–13; 7–11)	7–19 (7–19; 7–11)	7–17 (7–17; 7–11)	7–21 (7–21; 7–11)	7–14 (7–14; 7–11)	7–16 (7–15; 7–16)	7–20 (7–20; 7–12)

conserved intron boundaries, and with the exception of the sole group I intron found in *trnL*-UAA, all are group II introns (Michel et al. 1989). The intron-containing gene *rps12* is *trans*-spliced; its 5′ end exon (5′*rps12*) occurs in the LSC region and its second and third exons (3′*rps12*) reside some 27 kb away in IR_B separated by an intron (an additional copy of 3′*rps12* with intron exists in IR_A). Similar to other angiosperm plastid genomes, Apiales plastomes are AT-rich.

A comparison of the major structural features of each of these five plastid genomes, along with comparable data for the previously published *Daucus* and *Panax* plastomes, is provided (Table 2, Fig. 1). In the *Anthriscus cerefolium* plastome J_{LB} occurs within gene *rps19,* resulting in the duplication of a portion of this gene (99 bp) in IR_A at J_{LA}. There are 2 bp of noncoding sequence between J_{LA} and the 3′ end of gene *trnH*-GUG in the LSC region. The *Tiedemannia filiformis* subsp. *greenmanii* plastid genome contains the same complement of uniquely occurring and duplicated genes as in *Anthriscus*. In *Tiedemannia*, J_{LB} occurs within *rps19* resulting in the duplication of a portion of this gene (56 bp) in IR_A, and there are 2 bp of noncoding sequence between J_{LA} and *trnH*-GUG. The *Hydrocotyle verticillata* plastome contains the same complement of uniquely occurring and duplicated genes as in *Anthriscus* and *Tiedemannia*. Here, J_{LB} extends into *rps19* resulting in the duplication of a portion of this gene (62 bp) in IR_A, and there are 6 bp of noncoding sequence between J_{LA} and *trnH*-GUG. Relative to the previous three plastomes, the IR of *Petroselinum crispum* has contracted ~1.5 kb such that all of *rps19* and most of *rpl2* are now single-copy. J_{LB} is contained within *rpl2* and a small portion of this gene (37 bp) is duplicated at the end of IR_A. In *Petroselinum*, there are now only four intron-containing genes duplicated in the IR, not five. Between J_{LA} and the 3′ end of *trnH*-GUG in the LSC region there are 345 bp of noncoding sequence. Unlike the previous four genomes, the IR of *Crithmum maritimum* has expanded outwards ~1.5 kb such that all of *rps19* and previously single-copy genes *rpl22* and

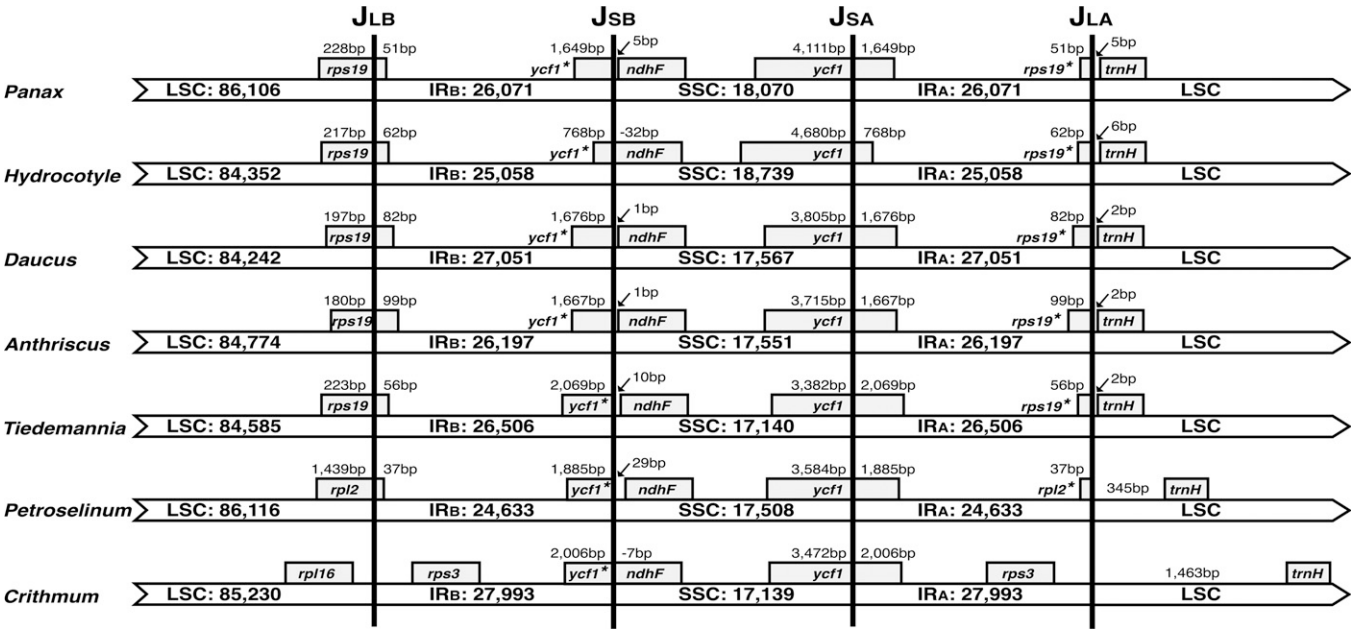

FIG. 1. Comparison of IR-single copy border positions across seven Apiales plastid genomes (vertical lines), with sizes of each of the four major plastid genome components indicated (LSC, IR_B, SSC, IR_A). The various lengths of complete genes (*rps19, rpl2, ycf1*), pseudogenes denoted by asterisks (*ycf1*, rps19*, rpl2**), and intergenic spacer regions (J_{SB}-*ndhF*, J_{LA}-*trnH*) adjacent to J_{LB}, J_{SB}, J_{SA}, and J_{LA} are also indicated. In *Crithmum*, J_{LB} occurs between genes *rpl16* and *rps3*, whereas in *Petroselinum*, J_{LB} occurs in the *rpl2* 5′exon. In all other Apiales examined, J_{LB} occurs at different positions within *rps19*. In *Hydrocotyle* and *Crithmum*, the final 32 and 7 nucleotides, respectively, of IR_B adjacent to J_{SB} are shared by the genes *ycf1** and *ndhF*, in opposite transcriptional orientations. *Crithmum* and *Petroselinum* have incorporated large regions of novel DNA (of 1,463 and 345 bp, respectively) into J_{LA}-*trnH*.

rps3 are now contained within the IR. Nine protein-coding genes are contained entirely within the IR, not six as in the other plastomes, and between J_{LA} and *trnH*-GUG there are 1,463 bp of noncoding sequence. The circular plastome gene map of *Crithmum* can be viewed as online supplementary data (supplementary Fig. S1).

Junctions J_{SB} and J_{SA} are in the same relative gene positions in all seven Apiales plastomes (Fig. 1). J_{SB} occurs within the 3′ end of *ndhF* (*Hydrocotyle*, *Crithmum*) or between genes *ndhF* and *ycf1**, the latter existing as a pseudogene in IR_B (*Panax, Daucus, Anthriscus, Tiedemannia, Petroselinum*). In *Hydrocotyle* and *Crithmum*, the final 32 and 7 nucleotides, respectively, of IR_B are shared by genes *ycf1** and *ndhF*, albeit transcribed in opposite directions (Fig. 1). *Ycf1* varies in size from 5,382–5,760 bp, and possesses numerous indels. In all seven plastomes, J_{SA} occurs in *ycf1*, but at different positions within this large and variably sized gene. As examples, in the *Tiedemannia* plastome, 2,069 bp of *ycf1* are duplicated in the IR, whereas in *Hydrocotyle* only 768 bp of *ycf1* are duplicated. Gene *rps19*, at 279 bp in size in all plastomes, overlaps J_{LB} in *Panax, Hydrocotyle, Daucus, Anthriscus*, and *Tiedemannia*, and as a result various lengths (51–99 bp) of *rps19* pseudogenes are located adjacent to J_{LA} in IR_A (Fig. 1).

An examination of DNA sequences flanking J_{SB} and J_{SA} across all seven species reveals that, other than *Daucus* and *Anthriscus* (both Apiaceae tribe Scandiceae) whose IR-SSC junctions are in identical positions, there are no significant sequence motifs or repetitive elements in common at or near these junctions which may have instigated IR boundary changes. Alignment of DNA sequences flanking J_{LA} across *Anthriscus, Daucus, Hydrocotyle, Tiedemannia, Panax*, and Araliaceae species *Tetrapanax papyrifer* (Wang et al. 2008) and *Eleutherococcus senticosus* (Yi et al. 2012), also revealed no common sequence on either side of J_{LA}. Differences among these seven species are due to length rather than point mutations, and variable poly(A) tracts are apparent between genes *rpl2* and *rps19**, ranging in length from A_8 in *Eleutherococcus, Panax*, and *Tetrapanax*, to A_{21} in *Anthriscus*. The sequences flanking J_{LA} in *Crithmum* and *Petroselinum* plastomes also each appear to be unique, although in these species novel, noncoding DNA is apparent between J_{LA} and 3′*trnH*-GUG.

The positions and percent identities of gap-free segments of pairwise alignments of *Anthriscus, Crithmum, Tiedemannia*, and *Petroselinum* with *Daucus* are shown as a MultiPip (Fig. 2). The positions of the genes shown on the MultiPip are relative to the *Daucus* plastome coordinates, the reference sequence. Most of the *Daucus* sequence aligns with that of the other four species of Apiaceae. Coding regions are represented as mostly unbroken aligning segments of relatively high percent identity (95–98%); noncoding regions, on the other hand, show more mismatches and indels, with percent identity values much lower. Highly variable noncoding regions are scattered throughout the single-copy regions and include intergenic spacers *ndhF-rpl32* and *rpl32-trnL* in the SSC region and 5′*trnK*-3′*rps16*, 5′*rps16-trnQ*, and *trnD-trnY-trnE-trnT* in the LSC region. A large segment of the *Daucus* plastome is not present in any of the secondary sequences. This region occurs in the IR between genes 3′*rps12* and *trnV*-GAC at *Daucus* coordinates ~99,300–100,800 and has been identified previously as a putative mitochondrion to plastid transfer region (Goremykin et al. 2009; Iorizzo et al. 2012a,b).

***Repeat Analysis*—**Repeat analysis identified 9–29 direct and inverted repeats of 30 bp or longer with a sequence identity of ≥90% among the five newly sequenced plastomes (Table 2). These repeated sequences were detected in coding regions (primarily *ycf2* and the three serine transfer-RNA [*trnS*] genes that recognize different codons, but also in *psaA* and *psaB*), introns (*ycf3, petB, petD, ndhB,* and *ndhA*), and most commonly, within 21 intergenic spacer regions scattered throughout the entire plastomes. The majority of these repetitive elements were direct repeats. Maximum repeat sizes ranged from 39 bp (*Tiedemannia*) to 71 bp (*Crithmum*). Several of these repeats occurred in the same locations and were shared by all five Apiaceae plastomes, either within genes, introns, or intergenic spacers.

Repeat analysis of the 345-bp region between J_{LA} and *trnH* of *Petroselinum* failed to detect any direct or inverted repeats ≥30 bp. In contrast, a series of direct repeats characterize the 1,463-bp *Crithmum* J_{LA}-*trnH* region, with the four largest of these being perfect or nearly perfect repeats of 71, 67, 49, and 39-bp. Many smaller repeats were also identified through manual examination, with the largest of these being 27- and 21-bp in size, plus an 18-bp repeating unit occurring three times in tandem. All of these repeats were restricted to positions 4–943 of the *Crithmum* plastome.

The number of homopolymers ≥7 bp in each of the five newly sequenced plastid genomes ranged between 283 and 326, with A or T polymers outnumbering G or C polymers by ratios of 9.0–11.5 to 1 (Table 2). Generally, the largest homopolymers were composed of A's or T's, the largest being 21-bp in size (Table 2). In each genome, the majority of homopolymers was 7 bp long (173–186), followed by 8 bp long (42–65), then 9 bp long (28–40). As the length of the homopolymer increased, fewer numbers were identified. Homopolymer sequences of >13 bp occurred infrequently in each genome (0–6). Among the largest homopolymers, the single 21-bp poly-A homopolymer (*Anthriscus*) occurred between IR genes *rpl2* and *rps19**; the single 20-bp poly-A homopolymer (*Crithmum*) occurred between LSC genes *trnT*-GGU and *psbD*; the single 19-bp poly-T homopolymer (*Hydrocotyle*) occurred between LSC genes *rps8* and *rpl14*; and the single 18-bp poly-A homopolymer (*Hydrocotyle*) occurred in SSC gene *ccsA*. The vast majority of homopolymers, and all but one of the largest ones (>13 bp), occurred in intergenic spacers; several others occurred in the variable and large hypothetical coding frames *ycf1* and *ycf2*.

***Novel DNA*—**Both *Petroselinum* and *Crithmum* contain novel DNA sequences between J_{LA} and LSC gene *trnH*-GUG of 345 and 1,463 bp, respectively (Fig. 1). In contrast, the other three newly sequenced plastomes are similar to *Daucus* and *Panax* in having only 2–6 bp of noncoding sequence in this region. These two large insertions show no sequence similarity to any other portions of their respective plastid genomes (by comparing these insert sequences to their truncated plastid genomes under varying degrees of stringency), nor do they match any ptDNA data currently available in GenBank. A BLAST search querying the 345-bp *Petroselinum* sequence resulted in multiple hits to several angiosperm mitochondrial genome sequences. The best alignment score, with 35% query coverage (positions 224–345, numbering beginning the first bp in the LSC region following J_{LA} and proceeding counter-clockwise), showed 91% sequence similarity to a mitochondrial DNA (mtDNA) intergenic spacer between genes cytochrome b (*cob)* and *ORF25* in several *Daucus carota* subsp. *sativus* cultivars and breeding lines (Bach et al. 2002). The same high match was obtained against the

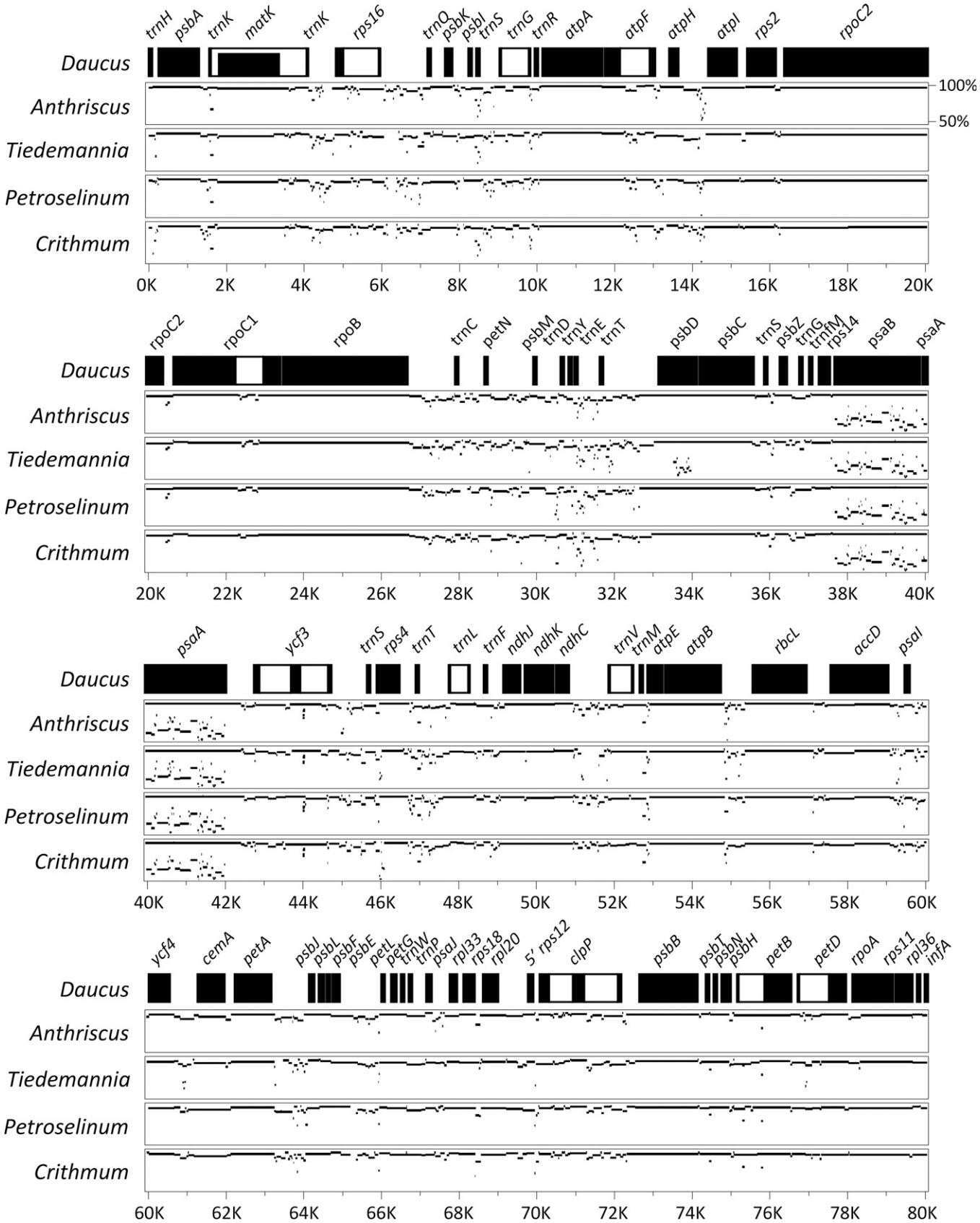

FIG. 2. Multiple percent identity plots (MultiPip) comparing the *Daucus* plastid genome (as the reference sequence) to *Anthriscus*, *Tiedemannia*, *Petroselinum*, and *Crithmum* plastid genomes using MultipPipMaker (Schwartz et al. 2003). IR_A is excluded. The top line is a gene map of *Daucus* (Ruhlman et al. 2006); introns are depicted by white boxes. Sequence similarity of aligned regions in *Anthriscus*, *Tiedemannia*, *Petroselinum*, and *Crithmum* is shown as horizontal bars in each plot indicating average percent identity between 50–100% (vertical axis). The horizontal axis represents the coordinates in the *Daucus* plastid genome. Parallel lines, as in *psaA* and *psaB*, indicate repeated sequence.

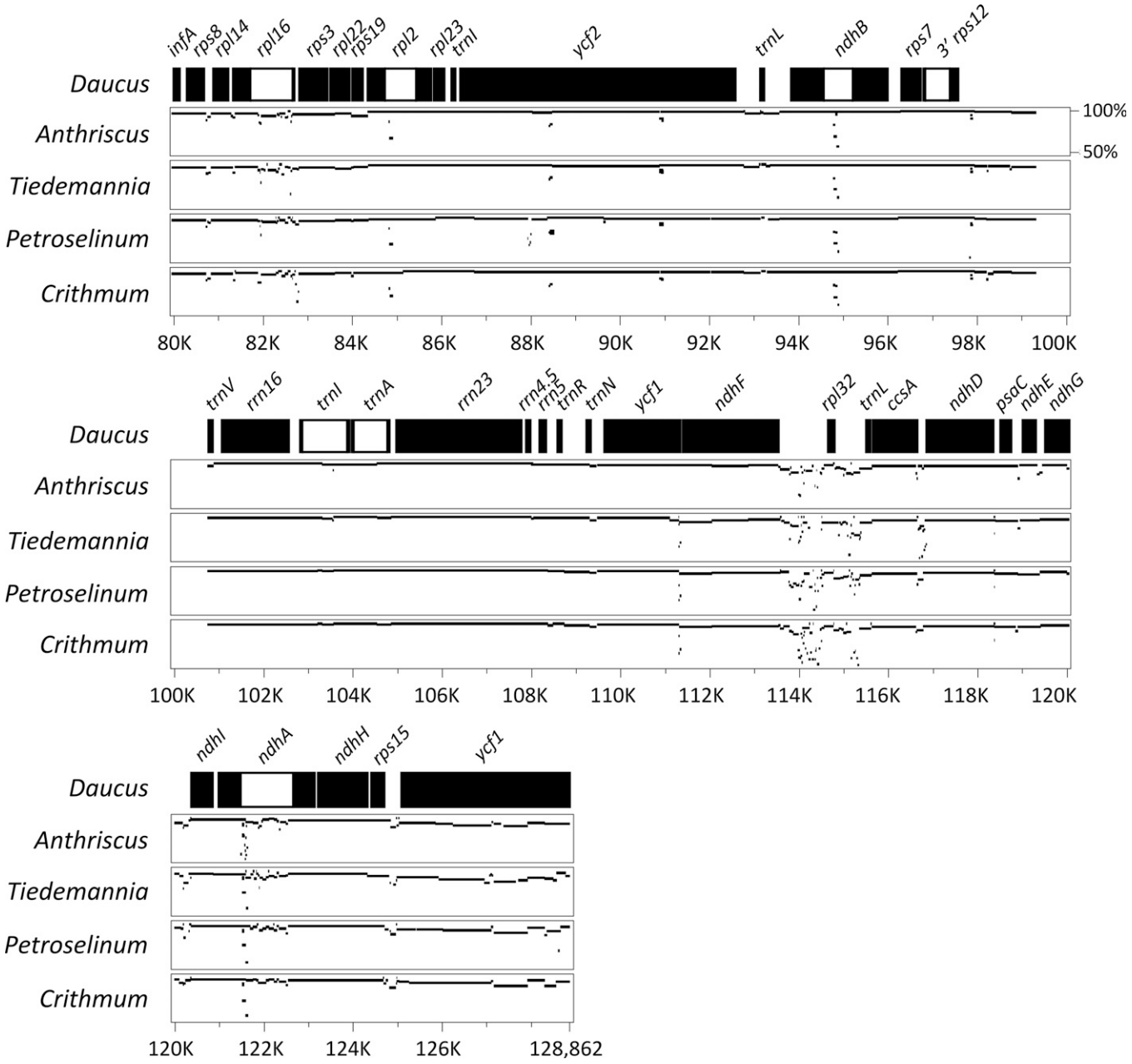

FIG. 2. continued.

complete *Daucus carota* subsp. *sativus* mitochondrial genome sequence, at positions 193,593–193,712 (Iorizzo et al. 2012b) representing an intergenic spacer also bounded by genes *cob* and *atp4* (*ORF25*). A BLAST search querying the 1,463-bp novel sequence of *Crithmum* resulted in no significant alignments; however, 39 bp of this sequence immediately adjacent to *trnH*-GUG showed 87% sequence similarity to a portion of the 345-bp *Petroselinum* insert and to the *Daucus carota* mitochondrial genome fragment. The *Crithmum trnH-psbA* region also contains novel DNA and this region was unalignable with the other Apiales species examined; a BLAST search of this sequence resulted in a single hit to the same spacer region in another accession of *Crithmum maritimum* (Degtjareva et al. 2009).

Identification of Divergent Regions—A comparison of sequence features across five Apiaceae plastid genomes and 70 noncoding ptDNA loci (>150 bp in *Daucus*) is presented in Table 3. Aligned lengths for each locus ranged from 172 (*trnfM-rps14*) to 1,545 bp (*5′rps16-trnQ*) and the number of parsimony-informative positions ranged from two (for several loci) to 73 (*trnE-trnT*). The number of variable alignment positions ranged from seven (*trnfM-rps14*) to 337 (*rpl32-trnL*). There was no obvious relationship between the length of the aligned sequence and the number of variable or parsimony-informative positions, a relationship consistent with other plastome sequence comparisons (Shaw et al. 2007). Pairwise sequence divergence values for each locus ranged from 0–3.1% for *trnfM-rps14* to 8.5–26.2% for *rpl32-trnL*. The number of indels in each region ranged from two (*rpoC2-rpoC1*) to 55 (*ndhF-rpl32*) and the number of parsimony-informative indels ranged from zero (for multiple loci) to 10 (*5′rps16-trnQ*). Percent variability ranged from 6.4 (*trnfM-rps14*) to 38.7% (*rpl32-trnL*). Of the four combined regions examined, *trnD-trnY-trnE-trnT* was the most variable, with 30.5%

TABLE 3. Comparison of 74 noncoding loci from five plastid genomes of Apiaceae (*Daucus*, *Anthriscus*, *Tiedemannia*, *Petroselinum*, and *Crithmum*). For the individual intergenic spacers and introns, the loci are listed as they occur on the *Crithmum* ptDNA map (Fig. S1), starting at J_{LA}, proceeding counterclockwise, and excluding the two IR regions. These noncoding loci represent all intergenic spacers >150 bp and introns from both LSC and SSC regions. The *trnH-psbA* locus excludes *Crithmum* because of alignment ambiguities.

Locus	Length (bp)	Aligned length (bp)	Constant positions	Uninformative positions	Parsimony-informative positions	Variable positions	Sequence divergence (%)	Total indels	Parsimony-informative indels	Percent variability
trnH-psbA	180–191	197	165	29	3	32	5.5–12.5	10	1	21.3
psbA-3′trnK	204–259	265	226	32	7	39	5.4–12.6	8	0	17.7
3′trnK-matK	221–283	284	268	13	3	16	1.8–4.4	3	0	6.7
matK-5′trnK	716–736	760	709	44	7	51	1.4–3.7	16	2	8.8
5′trnK-3′rps16	508–751	817	641	139	37	176	9.1–16.2	22	1	24.2
rps16 intron	837–875	902	809	83	10	93	2.3–6.7	16	3	12.1
5′rps16-trnQ	1,087–1,432	1,545	1,307	204	34	238	3.8–11.9	40	10	18.0
trnQ-psbK	343–362	363	341	14	8	22	0.9–4.0	6	0	7.7
psbK-psbI	316–412	434	361	59	14	73	3.6–12.8	15	3	20.3
trnS-5′trnG	510–557	605	528	63	14	77	3.8–10.0	18	3	15.7
trnG intron	698–712	726	654	50	22	72	3.9–6.6	12	2	11.6
atpF intron	721–741	770	682	79	9	88	1.1–8.4	19	5	13.9
5′atpF-atpH	275–412	420	361	49	10	59	3.0–9.6	8	1	16.0
atpH-atpI	742–1169	1,209	1,028	154	27	181	3.5–9.5	28	4	17.3
atpI-rps2	264–290	338	320	16	2	18	1.9–4.3	10	2	8.3
rps2-rpoC2	209–230	257	234	18	5	23	0.5–7.3	8	1	12.1
rpoC2-rpoC1	211–214	216	202	10	4	14	1.9–4.3	2	0	7.4
rpoC1 intron	742–751	775	716	47	12	59	2.2–4.2	15	2	9.5
rpoB-trnC	1,189–1,272	1,360	1,155	178	27	205	4.3–10.5	42	6	18.2
trnC-petN	475–690	755	662	82	11	93	3.7–8.8	25	4	15.6
petN-psbM	1,073–1,131	1,213	1,031	150	32	182	3.8–9.0	33	6	17.7
psbM-trnD	613–689	704	603	86	15	101	4.4–9.9	12	1	16.1
trnE-trnT	599–794	850	575	202	73	275	13.9–23.9	42	3	37.3
trnT-psbD	1,235–1,465	1,526	1,298	187	41	228	3.7–10.0	35	7	17.2
psbC-trnS	229–254	262	242	18	2	20	2.1–5.3	7	2	10.3
trnS-psbZ	340–394	400	357	33	10	43	3.7–6.6	12	2	13.8
psbZ-trnG	276–299	317	285	28	4	32	2.1–6.4	12	1	13.9
trnfM-rps14	159–166	172	165	5	2	7	0–3.1	4	2	6.4
psaA-ycf3 exon3	697–721	749	683	56	10	66	2.1–5.6	15	0	10.8
ycf3 intron2	735–784	794	730	55	9	64	1.3–5.3	11	1	9.4
ycf3 intron1	697–731	755	695	53	7	60	2.1–4.3	13	3	9.7
ycf3 exon1-trnS	693–877	903	759	122	22	144	3.1–11.2	28	1	19.0
trnS-rps4	186–320	328	278	45	5	50	4.9–8.9	9	2	18.0
rps4-trnT	353–382	408	347	43	18	61	4.1–10.5	16	3	18.9
trnT-5′trnL	766–803	863	730	107	26	133	3.7–11.1	29	3	18.8
trnL intron	496–518	522	492	28	2	30	0.8–4.8	5	1	6.7
3′trnL-trnF	310–375	377	334	39	4	43	2.9–7.7	12	2	14.6
trnF-ndhJ	385–399	427	365	48	14	62	3.1–9.9	12	2	17.3
ndhC-3′trnV	830–1,102	1,152	942	181	29	210	2.3–16.7	43	7	22.0
trnV intron	558–568	570	527	33	10	43	1.4–4.1	8	1	8.9
trnV-trnM	173–183	183	167	12	4	16	2.4–5.2	3	0	10.4
atpB-rbcL	742–770	839	742	86	11	97	1.3–8.1	22	0	14.2
rbcL-accD	582–605	631	573	45	13	58	2.9–5.9	15	3	11.6
accD-psaI	370–479	525	443	73	9	82	4.2–14.1	26	3	20.6
psaI-ycf4	386–422	449	397	43	9	52	3.6–7.1	18	1	15.6
ycf4-cemA	682–863	926	837	75	14	89	2.4–7.2	14	1	11.1
cemA-petA	229–267	275	253	20	2	22	1.2–6.3	7	0	10.5
petA-psbJ	733–967	1049	888	144	17	161	5.7–9.4	41	4	19.3
psbE-petL	916–1,160	1,215	1,068	124	23	147	3.2–7.5	27	1	14.3
trnP-psaJ	378–384	386	338	39	9	48	3.2–8.2	4	1	13.5
psaJ-rpl33	437–462	486	403	75	8	83	3.6–10.3	18	4	20.8
rps18-rpl20	202–230	235	193	35	7	42	1.8–10.8	9	4	21.7
rpl20–5′rps12	769–785	807	738	58	11	69	2.2–4.7	13	2	10.2
5′rps12–3′clpP	145–172	177	157	17	3	20	3.4–10.8	5	0	14.1
clpP intron2	619–660	670	610	53	7	60	2.9–4.8	27	3	13.0
clpP intron1	839–850	874	789	75	10	85	1.6–6.4	22	3	12.2
5′clpP-psbB	432–495	534	490	37	7	44	2.2–6.6	14	2	10.9
psbB-psbT	172–206	213	194	13	6	19	1.0–6.5	9	0	13.1
petB intron	748–756	757	693	54	10	64	1.7–4.9	5	0	9.1
petD intron	747–773	786	721	54	11	65	2.1–4.8	13	1	9.9
rps8-rpl14	194–203	206	179	23	4	27	1.0–10.9	5	0	15.5
rpl16 intron	883–952	1,000	883	100	17	117	2.4–7.4	32	5	14.9
ndhF-rpl32	843–1,086	1,202	945	213	44	257	7.0–15.0	55	5	26.0
rpl32-trnL	634–975	1,006	669	296	41	337	8.5–26.2	52	2	38.7
ccsA-ndhD	205–224	228	172	44	12	56	4.9–17.5	13	2	30.3
psaC-ndhE	260–286	294	249	33	12	45	3.3–10.4	5	0	17.0

(*Continued*)

TABLE 3. (CONTINUED).

Locus	Length (bp)	Aligned length (bp)	Constant positions	Uninformative positions	Parsimony-informative positions	Variable positions	Sequence divergence (%)	Total indels	Parsimony-informative indels	Percent variability
ndhE-ndhG	209–230	235	197	34	4	38	2.4–10.9	5	0	18.3
ndhG-ndhI	301–362	377	319	50	8	58	4.0–9.4	14	1	19.1
ndhA intron	1,055–1,104	1,142	1,005	111	26	137	2.8–7.6	30	1	14.6
rps15-ycf1	289–402	427	358	58	11	69	3.5–11.4	17	0	20.1
Combined Regions										
trnD-trnY-trnE-trnT	945–1,150	1,219	898	227	94	321	7.6–19.3	51	6	30.5
trnS-psbZ-trnG-trnfM	1,065–1,133	1,206	1,099	90	17	107	2.5–5.0	34	5	11.7
psbB-psbT-psbN-psbH	594–626	638	611	19	8	27	1.3–2.7	12	1	6.1
rpl36-infA-rps8-rpl14	1,066–1,079	1,083	994	76	13	89	2.1–5.0	8	0	9.0

variability, 7.6–19.3% sequence divergence, 321 variable alignment positions, and 51 indels.

A comparison of the 25 most divergent noncoding plastome regions, as represented by numbers of variable and parsimony-informative alignment positions, is presented in Fig. 3. Among the most variable loci include *rpl32-trnL, trnD-trnY-trnE-trnT, trnE-trnT, ndhF-rpl32, 5′rps16-trnQ, trnT-psbD, ndhC-3′trnV, rpoB-trnC, petN-psbM, atpH-atpI, 5-trnK-3′rps16, petA-psbJ*, and *psbE-petL*. These 13 regions had among the highest percent variability of all loci examined (14.3–38.7%). They also had among the highest pairwise sequence divergence values (19.3% for *trnD-trnY-trnE-trnT*, 23.9% for *trnE-trnT*, and 26.2% for *rpl32-trnL*). No intron was included in these top 13 most divergent loci, but of those included in the analysis four (*ndhA, rpl16, rps16*, and *atpF*) had the greatest numbers of variable alignment positions (88–137) and among the highest sequence divergence and % variability values (Table 3).

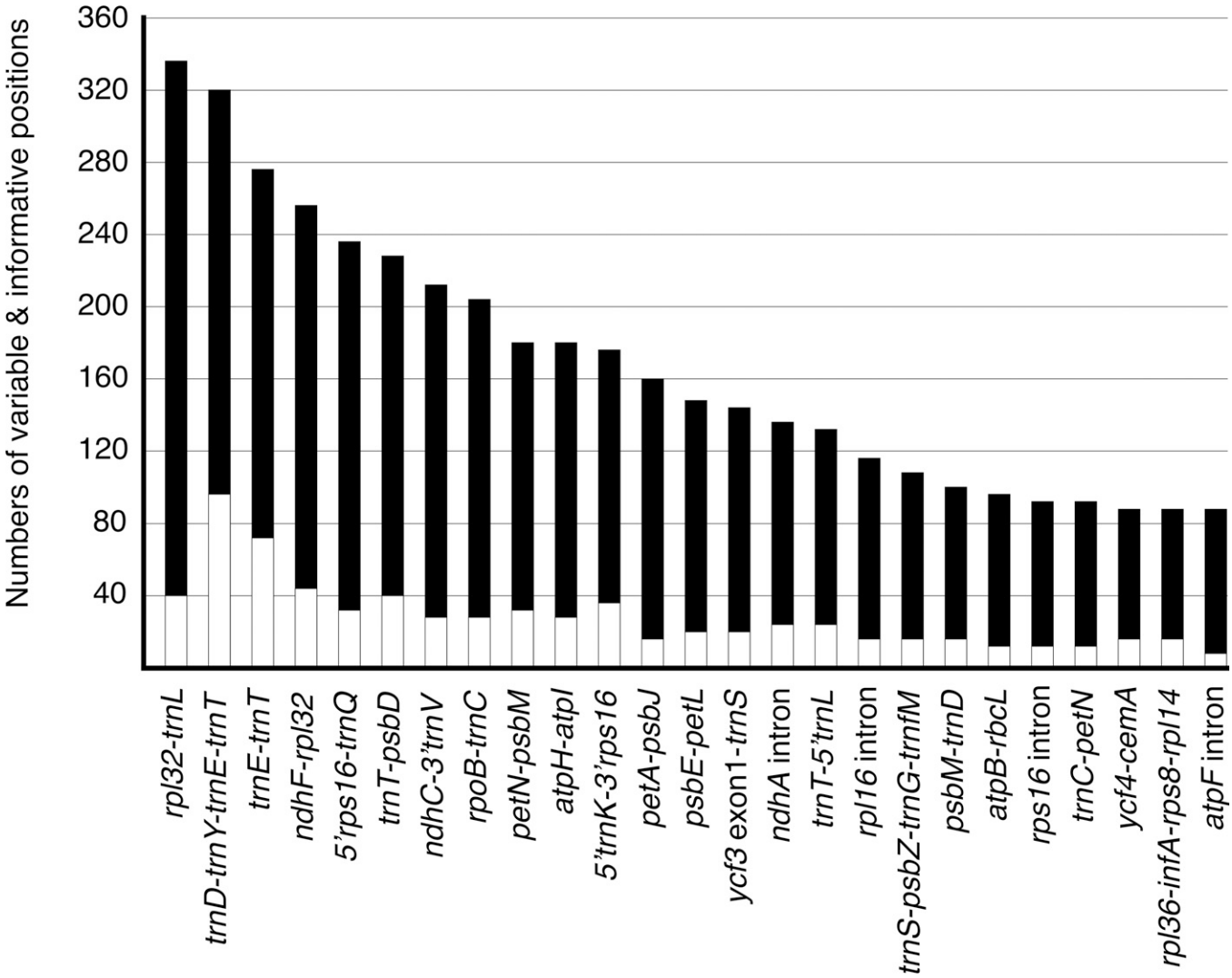

FIG. 3. Histogram showing the numbers of variable (dark boxes) and parsimony-informative (white boxes) positions in 25 of the most variable noncoding loci (intergenic spacers and introns) identified among the 74 regions compared from five Apiaceae plastid genomes.

Discussion

Plastome Comparisons—The plastomes of *Anthriscus, Crithmum, Hydrocotyle, Petroselinum*, and *Tiedemannia* are highly conserved in size, structure, and gene order and content compared to the previously published plastomes of *Panax* (Kim and Lee 2004) and *Daucus* (Ruhlman et al. 2006). They are also similar to the plastome of *Eleutherococcus senticosus* (Rupr. & Maxim.) Maxim. (Siberian ginseng), the most recent member of Araliaceae to have its plastid genome sequenced (Yi et al. 2012). Each of these Apiales plastomes shares characteristics of gene content and organization with the ancestral angiosperm plastid genome, as most frequently represented by *N. tabacum* (Raubeson et al. 2007). The ptDNAs of *Anthriscus* and *Daucus*, both Apiaceae tribe Scandiceae, are colinear, although the plastid genome map presented in Ruhlman et al. (2006) for *Daucus* indicates an additional open reading frame (*ORF80*) in the IR. Additionally, *Daucus* contains two copies of *trnG*-GCC in the LSC region, whereas in all other Apiales plastomes considered herein, one of these is replaced by *trnG*-UCC (between *trnS*-GCU and *trnR*-UCU). The sizes of these Apiales plastomes are within the known range for most angiosperms possessing an IR; however, the size of the *Crithmum* IR, at 27,993 bp, is at the higher end of its 20–30 kb average size (Plunkett and Downie 2000; Cai et al. 2006; Chumley et al. 2006; Ruhlman et al. 2006; Haberle et al. 2008; Wicke et al. 2011).

The IR boundaries of *Anthriscus, Daucus, Hydrocotyle, Tiedemannia*, *Panax*, and *Eleutherococcus* are also in the same relative positions as in many other unrearranged non-monocot angiosperm plastid genomes. In these Apiales taxa, J_{LB} occurs within *rps19*, resulting in the duplication of a portion of this gene at the end of IR_A adjacent to J_{LA}. J_{LA} occurs downstream from gene *trnH*-GUG in the adjacent LSC region. In *Petroselinum,* however, the IR has contracted ~1.5 kb at the IR_B-LSC boundary, such that J_{LB} is contained within the *rpl2* 5′exon, and in *Crithmum,* the IR has expanded ~1.5 kb at the IR_B-LSC boundary to duplicate genes *rpl22* and *rps3*. These large IR junction shifts (>1,000 bp) are unusual, for they occur less frequently in angiosperm plastomes than do IR contractions and expansions of <100 bp (Goulding et al. 1996). Another major difference between the plastomes of *Petroselinum* and *Crithmum* and those of most other angiosperms is the presence of a large insertion of novel, noncoding DNA in the region between J_{LA} and the 3′end of *trnH*-GUG, with portions of these inserts showing high sequence similarity to mitochondrial DNA.

Complete sequence data for 199 angiosperm plastid genomes (representing 134 genera) are publicly available in NCBI's Organelle Genome Resources database (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid; accessed February 1, 2013). Excluding most monocots and other angiosperm species whose plastid genomes are rearranged relative to the ancestral angiosperm plastome gene order (as exemplified by *N. tabacum*) and those species whose plastomes have lost one copy of the IR (Downie and Palmer 1992; Raubeson and Jansen 2005; Guisinger et al. 2011; Jansen and Ruhlman 2012), the J_{LA} boundaries for the remaining 108 plastomes (73 genera) typically occur between *rpl2* and *trnH*-GUG, with the latter present in its entirety only in the LSC region (Wang et al. 2008). While the *N. tabacum* plastome is frequently inferred to represent the ancestral plastid genome organization for angiosperms, it may not actually contain the exact ancestral IR boundaries (Raubeson et al. 2007). In many instances (like those of some Apiales), an *rps19* pseudogene of varying length remains in IR_A immediately adjacent to the LSC region; in three other instances, *rps19* is full-length (and therefore exists as two complete copies in the IR) and J_{LA} occurs between *rps19* and *trnH*-GUG. The majority of these plastomes (94) have between 0 and 30 nucleotides of noncoding sequence between J_{LA} and the 3′end of *trnH*-GUG. Twelve of the remaining 14 plastomes have between 34 and 71 bp of noncoding sequence in this region, one species of *Gossypium* L. has 121 bp of noncoding sequence, and *Hevea brasiliensis* (Willd. ex A. Juss.) Müll. Arg. has 198 bp of noncoding sequence in this region. Therefore, the large insertions of novel, noncoding sequence between J_{LA} and *trnH*-GUG of 345 nucleotides (*Petroselinum*) and 1,463 nucleotides (*Crithmum*) are highly unusual among angiosperms. Similar large insertions of novel DNA (of between 214 and 393 nucleotides) have been reported within this same region in other Apiaceae species and their pattern of distribution suggests that these insertions may be restricted to the "apioid superclade" of Apiaceae subfamily Apioideae (Peery et al. 2007). In most monocot plastomes, the gene *trnH* is located between *rpl2* and *rps19* in the IR (Mardanov et al. 2008; Wang et al. 2008; Wang and Messing 2011), and between J_{LA} and the first gene in the LSC region, *psbA*, there are 82–148 nucleotides of noncoding sequence. These sequences are also unique within the plastome (Hansen et al. 2007), but to the best of our knowledge these regions do not show any similarity to angiosperm mitochondrial DNA.

Mitochondrial DNA Transfer into the Plastid Genome—Angiosperm mitochondrial genomes readily accept DNA of plastid and nuclear origin (Knoop 2004; Kubo and Mikami 2007; Richardson and Palmer 2007; Goremykin et al. 2009; Iorizzo et al. 2012b), as well as sequences of horizontal origin from foreign genomes (Bergthorsson et al. 2003, 2004; Mower et al. 2010). In contrast, intracellular or horizontal gene transfer to the plastid genome is extremely rare, with the few reported cases restricted to some red and green algal plastids (Sheveleva and Hallick 2004; Rice and Palmer 2006; Kleine et al. 2009). Until recently, no convincing evidence of DNA transfer from either the nucleus or the mitochondrion into the angiosperm plastome has been documented (Palmer 1985b, 1990; Rice and Palmer 2006; Richardson and Palmer 2007; Bock and Timmis 2008; Bock 2010); in fact, such a transfer has generally been considered highly unlikely, possibly reflecting the lack of a DNA uptake system in plastids (Richardson and Palmer 2007; Kleine et al. 2009; Smith 2011).

Several recent papers, however, have identified DNA sequences of putative mitochondrial origin in angiosperm plastomes or invoked intracellular or horizontal transfer to explain the origin of novel sequences within plastid genomes. In the *Pelargonium* × *hortorum* L. H. Bailey (Geraniaceae) plastome, for example, *ORF56* and *ORF42* in the *trnA*-UGC intron showed high sequence similarities to the mitochondrial ACR-toxin sensitivity (*ACRS*) gene of *Citrus jambhiri* Lush. (Rutaceae) and to the mitochondrial *pvs-trnA* gene of *Phaseolus* L. (Fabaceae), respectively (Chumley et al. 2006). *ORF56* is also shared by both plastid and mitochondrial genomes in other taxa, suggesting transfer from the mitochondrion into the plastid (Chumley et al. 2006). The plastid genome of *Trifolium subterraneum* L. (Fabaceae) contains 19.5 kb of unique DNA distributed among 160 fragments ranging in size from

30–494 bp, some of which was hypothesized to represent instances of horizontal transfer from bacterial genomes (Cai et al. 2008). Other legume plastid genomes contain much repetitive sequences and unique DNA too, and speculation on their origins included intracellular transfers from the mitochondrion or nucleus, or horizontal transfers from other genomes, possibly pathogenic bacteria (Cai et al. 2008). Another example of transfer of mitochondrial DNA to the plastid genome has been identified in Apocynaceae tribe Asclepiadeae (Ku et al. 2013; Straub et al. 2013). In this case, it was suggested that homologous recombination is responsible for the transfer of the mitochondrial sequence into the plastome. Thus, it appears the transfer of mitochondrial DNA to the plastome is more common than previously thought.

Goremykin et al. (2009) uncovered two small sequences (of 74 and 126 bp) of the *Vitis vinifera* L. (Vitaceae) mitochondrial genome that were highly similar to regions of the *Daucus carota* plastome. BLAST analysis of the larger sequence revealed a high similarity to the coding region of the mitochondrial cytochrome c oxidase subunit 1 gene (*cox1*), prompting the authors to suggest that its presence in the *Daucus* plastome might possibly represent a rare transfer of DNA from the mitochondrion into the plastid. These two sequences are contained within a large 1,439-bp fragment of the *D. carota* IR (positions 99,309–100,747 and 139,407–140,845 in Ruhlman et al. 2006) that is a part of the 3′*rps12-trnV*-GAC intergenic spacer region; this fragment, however, has no similarity to any other published plastid nucleotide region (Goremykin et al. 2009). The MultiPip (Fig. 2) shows that this large fragment of the *Daucus* plastome is not present in any other Apiaceae plastome sequenced herein, including its closest ally *Anthriscus* (both Apiaceae tribe Scandiceae). Subsequently, Iorizzo et al. (2012a,b) confirmed the presence of this unique 1,439-bp fragment (identified as 1,452 bp in their study) in the *Daucus carota* plastome, while also discovering that it is present as three noncontiguous sequences in the *D. carota* mitochondrial genome. Further, the authors documented consistent conservation of a large portion of this region across all mitochondrial genomes of the diverse Apiaceae species they examined (representing seven additional species of *Daucus* and six other genera of Apiaceae, the latter including five species from the "apioid superclade" and *Eryngium planum* L. of subfamily Saniculoideae). In the plastid genome, this fragment, or a large portion thereof, was present only in *Daucus* and *Cuminum* L., both of tribe Scandiceae subtribe Daucinae (Lee et al. 2001). Iorizzo et al. (2012a) concluded that their results provided strong evidence of a mitochondrial to plastid transfer of DNA, and the presence of this putative mitochondrial to plastid transfer region in Scandiceae subtribe Daucinae but not in subtribe Scandicinae (*Anthriscus*) suggests this transfer event occurred sometime during the early evolution of the Daucinae clade. The authors also suggested that a retrotransposon-like event might have been responsible for transferring this sequence into the plastid genome (Iorizzo et al. 2012b).

Crithmum and *Petroselinum* also incorporate novel DNA into their plastomes. BLAST searches of the 345-bp *Petroselinum* unique insert resulted in 122 bp of this region having multiple hits to angiosperm mitochondrial DNAs from a variety of angiosperms, with the best alignment score matching a portion of the *cob-atp4*(*ORF25*) intergenic spacer in the *Daucus carota* mitochondrial genome. This high sequence similarity is suggestive of a possible intracellular transfer of DNA into the plastid from the *Petroselinum* mitochondrial genome. BLAST searches of the 1,463-bp *Crithmum* insert resulted in no significant hits, except for a small (39-bp) region that also matched a portion of the *Petroselinum* insert and the *D. carota* mitochondrial genome fragment. The origin of the large number of repetitive elements within the *Crithmum* insert remains unclear. We interpret these results cautiously, for we have not firmly demonstrated the mitochondrial provenance of this novel DNA, and suggest that the sequencing of whole mitochondrial genomes from the Apiaceae, specifically from *Petroselinum*, may eventually shed light on its origin.

Repeated Sequences—For the five newly sequenced plastomes, repeat analysis identified 9–29 dispersed repeats of 30 bp or longer with a sequence identity of 90% or greater. In the previously published *Daucus* plastome, repeat analysis identified 12 direct and two inverted repeats of ≥30 bp, four of which were direct repeats (of 30–70 bp) occurring in *ycf2* (Ruhlman et al. 2006). In *Panax*, nine direct and two inverted repeats ≥30 bp were identified, the largest of these being four tandem repeats of 57-bp in *ycf1* (Kim and Lee 2004). In the *Eleutherococcus* plastome, a total of 23 repeats ≥30 bp were located, the largest being 79 bp in size and also occurring in *ycf2* (Yi et al. 2012). Repeated sequences, other than the IR, are considered to be uncommon in plastid genomes and are more prevalent in those plastomes that have undergone major changes in genome organization (Palmer 1985a, 1991; Chumley et al. 2006; Haberle et al. 2008; Weng et al. 2014). Considering all Apiales plastomes investigated to date, repeats occurred most often in five introns and numerous intergenic spacer regions scattered throughout the genome; within coding regions, repeats were prevalent in genes *ycf2, trnS* (3 genes), *psaA,* and *psaB*. The repeated elements identified in Apiales are similar in number, size, and location to those of other angiosperms whose plastomes are structurally unrearranged (Ruhlman et al. 2006; Yi et al. 2012). The origin of these repeats is not known, although replication slippage could be responsible for generating direct repeats (Palmer 1991).

Similarly, homopolymer sequences of >13 bp occurred infrequently in each plastid genome. In *Panax*, 18 homopolymers ≥10 bp were identified previously (the largest being 13-bp in size), the majority being composed of multiple A's or T's (Kim and Lee 2004). In *Eleutherococcus*, 27 homopolymers ≥10 bp were identified, of which 22 were composed of multiple A or T bases and the largest were also of 13 bp in size (Yi et al. 2012). These repetitive sequences, of which there are many in the Apiales, may be potentially useful for population-level genetic studies, as sequence length polymorphisms can be surveyed among multiple accessions of a species revealing haplotype variation (Yi et al. 2012, and references cited therein). Homopolymer runs are difficult to resolve using the Roche/454 platform (and other sequencing methods) and contribute to sequencing errors (Moore et al. 2006). In our study, Sanger sequencing was used for independent confirmation of only a few homopolymeric repeats at or near IR-single copy junctions. However, with the low depth of coverage at some contig ends and the fact that very few homopolymer runs were confirmed by Sanger sequencing, we caution that sequencing errors are likely present, especially within the largest homopolymeric regions.

All major shifts in J_{LB} position documented in Apiaceae are restricted to members of the "apioid superclade" (Plunkett and Downie 1999, 2000). These include both major IR expansions and contractions, such as those presented herein for *Crithmum* and *Petroselinum*. The prevalence of IR junction

shifts in this one (albeit large) clade suggests that some common mechanism is responsible for flux in position of J_{LB} and that this mechanism likely originated in the immediate common ancestor of the group (Plunkett and Downie 2000). The underlying mechanisms responsible for causing such large IR junction shifts are not yet known, but may involve homologous recombination between small, dispersed repeated sequences, adjacency to tRNA genes, or double-stranded breakage followed by recombination between poly(A) tracts (Palmer 1985a; Goulding et al. 1996; Plunkett and Downie 2000; Wang et al. 2008). Small IR expansions (<100 bp), such as those found in many *Nicotiana* species, are likely caused by gene conversion (Goulding et al. 1996; Wang et al. 2008). Comparisons of DNA sequences flanking IR-single copy regions in Apiales reveal that there are no dispersed repeats at or near these junctions, although this doesn't preclude such repetitive elements from instigating IR boundary changes in other apioid taxa. However, the absence of any significant sequence motifs or repetitive elements at or near IR-single copy junctions in other angiosperms exhibiting IR boundary fluxes suggests that these types of sequences may not be the typical cause of IR spreading (Palmer 1985b; Goulding et al. 1996), although Wang et al. (2008) observed that IR-LSC junctions are indeed found at either poly(A) tracts or in A-rich regions. In the "apioid superclade," the presence of both large-scale expansions and contractions of the IR at J_{LB} and large, novel insertions in the LSC region adjacent to J_{LA} may be coincidental, or the latter may be mediating, in some way, these IR boundary changes, or vice versa.

Expansion/Contraction of the IR as a Phylogenetic Marker—Because of their infrequent occurrence, major structural rearrangements of the plastid genome usually can provide strong evidence of monophyly (Downie and Palmer 1992; Raubeson and Jansen 2005; but see Downie et al. 1991, for an exception). These rare, large-scale mutational changes provide complementary markers that can be used alongside nucleotide substitutions in molecular systematic studies. Variation in overall size of the IR is common in angiosperms, but such a character cannot readily be used in a phylogenetic analysis because length mutations can occur anywhere within the IR, making homologous size variants difficult to assess. Specific expansions and contractions of border positions of the IR, however, can be used to demarcate monophyletic groups, and these markers have been used previously to delimit major clades in Ranunculaceae (Johansson and Jansen 1993; Hoot and Palmer 1994; Johansson 1998), Berberidaceae (Kim and Jansen 1994), *Nicotiana* (Goulding et al. 1996), Campanulales (Knox and Palmer 1999), Poaceae (Guisinger et al. 2010), and monocots (Wang et al. 2008). Minor changes in IR junction positions (<100 bp) are more common (Goulding et al. 1996), but this increased frequency suggests that they would not make very reliable phylogenetic markers.

In Apiaceae, the IR has both expanded and contracted significantly, with successive contractions identified within the "apioid superclade," five of which are parsimony-informative (Plunkett and Downie 2000). Members of Apiaceae tribes Careae and Pyramidoptereae, both previously attributable to the "*Aegopodium* group," share a large expansion of J_{LB} relative to the *Nicotiana* plastome and are monophyletic sister groups based on a nuclear ribosomal DNA internal transcribed spacer (nrDNA ITS) sequence phylogeny (Plunkett and Downie 2000; Banasiak et al. 2013). Members of Apiaceae tribe Apieae (the "*Apium* group"), as represented herein by *Petroselinum*, but also by *Ammi* L., *Anethum* L., *Apium* L., *Foeniculum* Hill., and *Ridolfia* Moris. in Plunkett and Downie (2000), are all characterized by a ~1.5 kb J_{LB} contraction. *Conium* L. (*Conium* clade) and *Pimpinella* L. (tribe Pimpinelleae) also possess a similar-sized IR contraction as that found in tribe Apieae (Plunkett and Downie 2000), but whether this contraction serves as a synapomorphy uniting these three major clades of taxa remains to be seen. In the absence of sequence data flanking their IR junctions, ambiguity remains in assessing the precise location of J_{LB}.

To date, all major IR expansions and contractions uncovered in the Apiales are restricted to the "apioid superclade," a large, distally branching group within Apiaceae subfamily Apioideae whose higher-level relationships are unclear (Plunkett and Downie 2000). Resolution of relationships among its constituent 14 tribes and major clades (Downie et al. 2010), as well as the formal recognition of these major clades, must await supporting data, such as that provided by the distribution of well-characterized, coincident IR junction locations. Such rearrangements are rare enough that they have the potential to resolve with confidence a particular branching point in a phylogeny and define monophyletic groups (Downie and Palmer 1992). We are currently surveying for, characterizing, and circumscribing the distribution of major IR junction shifts in the "apioid superclade" to more rigorously evaluate the extent of IR junction flux in the group (R. Peery et al. unpubl. data). Even though these rearrangements alone are unlikely to provide a comprehensive framework of relationships, when used in conjunction with other molecular and traditional approaches, they have the power to illuminate phylogeny or to provide additional support for otherwise weakly-supported clades. Of course, there is also the possibility that IR fluxes are completely random, as they appear to be among closely related *Nicotiana* species (Goulding et al. 1996), or successive expansion-contraction events have occurred, as has been suggested in adzuki bean (Perry et al. 2002) and *Pelargonium* × *hortorum* (Chumley et al. 2006), which would confound interpretation of phylogeny.

Noncoding Regions as a Source of Phylogenetic Information in Apiaceae—At present, nrDNA ITS sequences comprise the most comprehensive database for Apiaceae phylogenetic study (Downie et al. 2010). These data have often been supplemented with sequences from a variety of ptDNA introns and intergenic spacers, yet for those taxa where radiations have been relatively recent it has been difficult to generate sufficient phylogenetic signal because of the relatively slow rate of evolution of the various plastid loci being compared. Within the "apioid superclade," phylogenetic resolution is particularly poor within several groups of genera, such as the clade of perennial, endemic Apiaceae subfamily Apioideae of western North America (Sun et al. 2004; Sun and Downie 2010), *Bunium* (Degtjareva et al. 2009), *Conium* (Cordes 2009), *Pimpinella* (Magee et al. 2010), the *Arracacia* clade (Danderson 2011), *Heracleum* (Yu et al. 2011), and *Angelica* (Liao et al. 2013). Each of these groups contains unresolved terminal clades of presumably closely related species, which would benefit considerably by the inclusion of additional phylogenetically informative characters.

The development of genetic markers to infer evolutionary relationships among recently diverged taxa remains an important challenge in molecular systematic studies and, to this end, considerable advances have been made in the identification

and development of markers from the angiosperm plastid genome (Shaw et al. 2005, 2007). Despite these efforts, a plastid region identified as highly variable in one group may not actually be phylogenetically informative in another. As an example, the *Crithmum trnH-psbA* intergenic spacer region shows extensive sequence divergence when compared with the other four Apiaceae species and could not be aligned unambiguously with these sequences. This region is substantially shorter in *Crithmum* ptDNA (127 bp) than that of the other species (180–191 bp); it also contains novel DNA, many small, repeated sequences in direct orientation, and is AT-rich (73%). This spacer region is highly variable in angiosperms (Aldrich et al. 1988; Shaw et al. 2007) and has been proposed for DNA barcoding (Kress et al. 2005), yet in our study this locus ranked 59th out of 74 in its number of variable alignment positions. Furthermore, in *Heracleum* and its allies (Apiaceae tribe Tordylieae), the locus is quite conserved, with one small internal region characterized by an inversion and duplication (Logacheva et al. 2008). Repeats, large deletions, and small inversions in *trnH-psbA* have been reported for other species of Apiales, severely confounding the use of this locus in phylogenetic analyses (Degtjareva et al. 2012).

Practical information to emerge from our study is the identification of divergent plastid regions for enhancing resolution of low-level phylogenetic relationships in Apiaceae. A comparison of levels of DNA sequence divergence across five Apiaceae plastomes, representing species from four disparate lineages within the family (tribes Apieae, Oenantheae, Pyramidoptereae, and Scandiceae), revealed that intergenic spacer regions *rpl32-trnL, trnE-trnT, ndhF-rpl32, 5′rps16-trnQ*, and *trnT-psbD* are among the fastest-evolving loci, as measured by numbers of variable and parsimony-informative characters, and also by total indels, pairwise sequence divergence estimates, and percent variability. Aligned lengths of each of these regions ranged from 850–1,545 positions, and these sizes are of sufficient length for ease of PCR-amplification and DNA sequencing. These five spacers appear to be good candidates for resolving recent divergences. Consideration of two spacers simultaneously because of gene adjacency and co-amplification as a single contiguous unit (i.e. *ndhF-rpl32* and *rpl32-trnL*, and *trnE-trnT* and *trnT-psbD*) offers even better regions for phylogenetic analysis. The *trnD-trnY-trnE-trnT* combined region is substantially longer and presents the greatest number of parsimony-informative characters overall. Shaw et al. (2007) identified nine noncoding regions as the best possible choices for low-level phylogenetic studies of angiosperms. All nine of these regions are within our top 13 most divergent loci. To these nine, we also include *trnE-trnT, rpoB-trnC, petN(ycf6)-psbM*, and the *trnD-trnY-trnE-trnT* combined region; these regions, too, were also noted as highly variable or in the "Tier 1" category of Shaw et al. (2005, 2007). Although we acknowledge that a plastid region identified as highly variable in one group may not be as useful phylogenetically in another group, it is interesting that the Shaw et al. (2005, 2007) studies identified the same highly variable loci for several disparate angiosperm lineages as we did for the Apiaceae. The resolution of recent divergences in Apiaceae would benefit considerably by the inclusion of any or all of these highly variable loci.

No intron was included in the top 13 most divergent noncoding loci, although four (*ndhA, rpl16, rps16* and *atpF*) made the top 25. Of these, the *rpl16* and *rps16* introns have been used extensively in Apiaceae phylogenetic studies, whereas we are unaware of any studies of the family incorporating *ndhA* or *atpF* intron sequences. Group II introns of the plastid genomes of land plants, which includes the aforementioned four introns, show a strong relationship between the functional importance of its secondary structural features and the likelihood of mutational change, with those domains and subdomains essential for intron-associated functions most conserved evolutionarily (Downie et al. 1996, 2000; Kelchner 2002). Intergenic spacers are less functionally constrained than introns and therefore more likely to vary, thereby providing more variable and phylogenetically informative characters for phylogenetic analyses. The greater variability of intergenic spacer regions over introns in plastome comparisons has been reported previously (Shaw et al. 2007). In Apiaceae, intron sequences do not provide enough informative characters to resolve relationships below generic levels (Downie et al. 1996, 2000, 2008; Downie and Katz-Downie 1999).

In conclusion, a comparative analysis of seven complete plastid genomes from the Apiales, five of which were sequenced during the course of this investigation, has provided a plethora of information on their structural organization and sequence evolution. The features of five of these genomes are consistent structurally with typical plastomes of other non-monocot angiosperm species, while two, *Crithmum* and *Petroselinum*, each demonstrate a ~1.5 kb expansion or contraction of its IR. These two species also incorporate novel DNA adjacent to J_{LA}, indicative of a mitochondrial to plastid transfer of DNA. A comparison of 74 loci, representing all noncoding regions >150 bp from throughout the single-copy portions of five Apiaceae plastomes, demonstrated a wide range of sequence divergence in different regions. Several regions were identified that yielded greater numbers of variable and parsimony-informative characters than many of those regions heretofore commonly employed in Apiaceae molecular systematic studies, providing new information on the potential phylogenetic utility of plastid noncoding loci. Therefore, these highly variable loci should be more useful in future interspecific phylogenetic, intraspecific phylogeographic, and population-level genetic studies of these plants. The distribution of specific IR junction shifts, on the other hand, has the potential to demarcate major clades within the family, particularly those within the "apioid superclade" whose higher-level relationships remain elusive.

The frequency and large size of IR junction shifts in Apiaceae plastomes appear to be unusual in angiosperm families whose plastid genomes are otherwise unrearranged, indicating that the group represents a model system in which to study the mechanisms leading to large-scale expansions and contractions of the IR. Concomitant with these IR shifts is the incorporation of novel DNA in the LSC region adjacent to J_{LA}, some of which represents transfer of DNA from the mitochondrial genome into the plastid. With other reports of a putative mitochondrial to plastid transfer region elsewhere in the plastomes of *Daucus* and *Cuminum* (Goremykin et al. 2009; Iorizzo et al. 2012a) it will be well worthwhile to examine additional Apiaceae plastomes for instances of intracellular transfer of DNA. To improve understanding of the mechanisms causing IR junction shifts and to further characterize and circumscribe the putative mitochondrial DNA insert, the sequences spanning the four IR-single-copy junctions are being compared in additional species of Apiaceae (R. Peery et al. unpubl. data). By combining these data and analyzing them within the context of a ptDNA-derived phylogeny of the

group, it will be possible to determine when the first IR junction shift occurred and the extent of IR junction mobility during evolution of the group. Complete plastome sequences of other members of the "apioid superclade" will further advance understanding of the evolutionary events in the plastome that accompanied umbellifer diversification.

Acknowledgments. The authors thank Clinton Morse of the University of Connecticut Plant Growth Facilities for providing plant material, Murray Henwood of the University of Sydney for identifying the *Hydrocotyle* species, J. Chris Blazier and Mary M. Guisinger for providing technical assistance, Rhiannon M. Peery for discussions, and two anonymous reviewers for constructive comments. This work was initiated while SRD was on sabbatical leave in the laboratory of RKJ. This research was supported by a grant to RKJ from the National Science Foundation (IOS-1027259).

Literature Cited

Aii, J., Y. Kishima, T. Mikami, and T. Adachi. 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. *Current Genetics* 31: 276–279.

Aldrich, J., B. W. Cherney, E. Merlin, and L. Christopherson. 1988. The role of insertions/deletions in the evolution of the intergenic region between *psbA* and *trnH* in the chloroplast genome. *Current Genetics* 14: 137–146.

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

Bach, I. C., A. Olesen, and P. W. Simon. 2002. PCR-based markers to differentiate the mitochondrial genomes of petaloid and male fertile carrot (*Daucus carota* L.). *Euphytica* 127: 353–365.

Banasiak, Ł., M. Piwczynski, T. Ulinski, S. R. Downie, M. F. Watson, B. Shakya, and K. Spalik. 2013. Dispersal patterns in space and time: a case study of Apiaceae subfamily Apioideae. *Journal of Biogeography* 40: 1324–1335.

Bergthorsson, U., K. L. Adams, B. Thomason, and J. D. Palmer. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424: 197–201.

Bergthorsson, U., A. O. Richardson, G. J. Young, L. R. Goertzen, and J. D. Palmer. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proceedings of the National Academy of Sciences USA* 101: 17747–17752.

Bock, R. 2010. The give-and-take of DNA: horizontal gene transfer in plants. *Trends in Plant Science* 15: 11–22.

Bock, R. and J. N. Timmis. 2008. Reconstructing evolution: gene transfer from plastids to the nucleus. *BioEssays* 30: 556–566.

Boudreau, E. and M. Turmel. 1995. Gene rearrangements in *Chlamydomonas* chloroplast DNAs are accounted for by inversions and by the expansion/contraction of the inverted repeat. *Plant Molecular Biology* 27: 351–364.

Cai, Z., M. Guisinger, H. G. Kim, E. Ruck, J. C. Blazier, V. McMurtry, J. V. Kuehl, J. Boore, and R. K. Jansen. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *Journal of Molecular Evolution* 67: 696–704.

Cai, Z., C. Penaflor, J. V. Kuehl, J. Leebens-Mack, J. E. Carlson, C. W. dePamphilis, J. L. Boore, and R. K. Jansen. 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evolutionary Biology* 6: 77.

Calviño, C. I. and S. R. Downie. 2007. Circumscription and phylogeny of Apiaceae subfamily Saniculoideae based on chloroplast DNA sequences. *Molecular Phylogenetics and Evolution* 44: 175–191.

Chevreux, B., T. Pfisterer, and S. Suhai. 2000. Automatic assembly and editing of genomic sequences. In: Suhai S. (ed), Genomics and proteomics—functional and computational aspects. New York: Kluwer Academic/Plenum Publishers.

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

Cordes, J. M. 2009. *A systematic study of poison hemlock (Conium, Apiaceae)*. M. S. thesis. Urbana, Illinois: University of Illinois at Urbana-Champaign.

Cosner, M. E., R. K. Jansen, J. D. Palmer, and S. R. Downie. 1997. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Current Genetics* 31: 419–429.

Danderson, C. A. 2011. *A phylogenetic study of the Arracacia clade (Apiaceae)*. Ph. D. thesis. Urbana, Illinois: University of Illinois at Urbana-Champaign.

Degtjareva, G. V., E. V. Kljuykov, T. H. Samigullin, C. M. Valiejo-Roman, and M. G. Pimenov. 2009. Molecular appraisal of *Bunium* and some related arid and subarid geophilic Apiaceae-Apioideae taxa of the ancient Mediterranean. *Botanical Journal of the Linnean Society* 160: 149–170.

Degtjareva, G. V., M. D. Logacheva, T. H. Samigullin, E. I. Terentieva, and C. M. Valiejo-Roman. 2012. Organization of chloroplast *psbA-trnH* intergenic spacer in dicotyledonous angiosperms of the family Umbelliferae. *Biochemistry (Moscow)* 77: 1056–1064.

Downie, S. R. and D. S. Katz-Downie. 1999. Phylogenetic analysis of chloroplast *rps16* intron sequences reveals relationships within the woody southern African Apiaceae subfamily Apioideae. *Canadian Journal of Botany* 77: 1120–1135.

Downie, S. R. and J. D. Palmer. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. Pp. 14–35 in *Molecular Systematics of Plants*, eds. P. S. Soltis, D. E. Soltis, and J. J. *Doyle*. New York: Chapman and Hall.

Downie, S. R., D. S. Katz-Downie, and K.-J. Cho. 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpoC1* intron. *Molecular Phylogenetics and Evolution* 6: 1–18.

Downie, S. R., D. S. Katz-Downie, and M. F. Watson. 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. *American Journal of Botany* 87: 273–292.

Downie, S. R., D. S. Katz-Downie, F.-J. Sun, and C.-S. Lee. 2008. Phylogeny and biogeography of Apiaceae tribe Oenantheae inferred from nuclear rDNA ITS and cpDNA *psbI-5′trnK*$^{(UUU)}$ sequences, with emphasis on the North American endemics clade. *Botany* 86: 1039–1064.

Downie, S. R., G. M. Plunkett, M. F. Watson, K. Spalik, D. S. Katz-Downie, C. M. Valiejo-Roman, E. I. Terentieva, A. V. Troitsky, B.-Y. Lee, J. Lahham, and A. El-Oqlah. 2001. Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. *Edinburgh Journal of Botany* 58: 301–330.

Downie, S. R., R. G. Olmstead, G. Zurawski, D. E. Soltis, P. S. Soltis, J. C. Watson, and J. D. Palmer. 1991. Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: molecular and phylogenetic implications. *Evolution* 45: 1245–1259.

Downie, S. R., K. Spalik, D. S. Katz-Downie, and J.-P. Reduron. 2010. Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. *Plant Diversity and Evolution* 128: 111–136.

Feist, M. A. E., S. R. Downie, A. R. Magee, and M. Liu. 2012. Revised generic delimitations for *Oxypolis* and *Ptilimnium* (Apiaceae) based on leaf morphology, comparative fruit anatomy, and phylogenetic analysis of nuclear rDNA ITS and cpDNA *trnQ-trnK* intergeneric spacer sequence data. *Taxon* 61: 402–418.

Goulding, S. E., R. G. Olmstead, C. W. Morden, and K. H. Wolfe. 1996. Ebb and flow of the chloroplast inverted repeat. *Molecular & General Genetics* 252: 195–206.

Goremykin, V. V., F. Salamini, R. Velasco, and R. Viola. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* 26: 99–110.

Guisinger, M. M., T. W. Chumley, J. V. Kuehl, J. L. Boore, and R. K. Jansen. 2010. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *Journal of Molecular Evolution* 70: 149–166.

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

Haberle, R. C., H. M. Fourcade, J. L. Boore, and R. K. Jansen. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *Journal of Molecular Evolution* 66: 350–361.

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

Hoot, S. B. and J. D. Palmer. 1994. Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. *Journal of Molecular Evolution* 38: 274–281.

Iorizzo, M., D. Grzebelus, D. Senalik, M. Szklarczyk, D. Spooner, and P. Simon. 2012a. Against the traffic: The first evidence for mitochondrial DNA transfer into the plastid genome. *Mobile Genetic Elements* 2: 261–266.

Iorizzo, M., D. Senalik, M. Szklarczyk, D. Grzebelus, D. Spooner, and P. Simon. 2012b. *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 Biology* 12: 61.

Jansen, R. K., L. A. Raubeson, J. L. Boore, C. W. dePamphilis, T. W. Chumley, R. C. Haberle, S. K. Wyman, A. J. Alverson, R. Peery, S. J. Herman, H. M. Fourcade, J. V. Kuehl, J. R. McNeal, J. Leebens-Mack, and L. Cui. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology* 395: 348–384.

Jansen, R. K. and T. A. Ruhlman. 2012. Plastid genomes of seed plants. Pp. 103–126 in *Genomics of chloroplasts and mitochondria, advances in photosynthesis and respiration*, eds. R. Bock and V. Knoop. Dordrecht: Springer.

Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with ClustalX. *Trends in Biochemical Sciences* 23: 403–405.

Johansson, J. T. 1998. Chloroplast DNA restriction site mapping and the phylogeny of *Ranunculus* (Ranunculaceae). *Plant Systematics and Evolution* 213: 1–19.

Johansson, J. T. and R. K. Jansen. 1993. Chloroplast DNA variation and phylogeny of the Ranunculaceae. *Plant Systematics and Evolution* 187: 29–49.

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

Kim, Y.-D. and R. K. Jansen. 1994. Characterization and phylogenetic distribution of a chloroplast DNA rearrangement in the Berberidaceae. *Plant Systematics and Evolution* 193: 107–114.

Kim, K.-J. and H.-L. Lee. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* 11: 247–261.

Kleine, T., U. G. Maier, and D. Leister. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology* 60: 115–138.

Knoop, V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics* 46: 123–139.

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

Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences USA* 102: 8369–8374.

Ku, C., W.-C. Chung, L.-L. Chen, and C.-H. Kuo. 2013. The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: plastid genome evolution, molecular marker identification, and phylogenetic implications in asterids. *PLoS ONE* 8: e68518.

Kubo, T. and T. Mikami. 2007. Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum* 129: 6–13.

Kurtz, S. and C. Schleiermacher. 1999. REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* 15: 426–427.

Lee, B.-Y. and S. R. Downie. 2000. Phylogenetic analysis of cpDNA restriction sites and *rps16* intron sequences reveals relationships among Apiaceae tribes Caucalideae, Scandiceae and related taxa. *Plant Systematics and Evolution* 221: 35–60.

Lee, B.-Y., G. A. Levin, and S. R. Downie. 2001. Relationships within the spiny-fruited umbellifers (Scandiceae subtribes Daucinae and Torilidinae) as assessed by phylogenetic analysis of morphological characters. *Systematic Botany* 26: 622–642.

Liao, C., S. R. Downie, Q. Li, Y. Yu, X. He, and B. Zhou. 2013. New insights into the phylogeny of *Angelica* and its allies (Apiaceae) with emphasis on East Asian species, inferred from nrDNA, cpDNA, and morphological evidence. *Systematic Botany* 38: 266–281.

Logacheva, M. D., C. M. Valiejo-Roman, and M. G. Pimenov. 2008. ITS phylogeny of West Asian *Heracleum* species and related taxa of Umbelliferae—Tordylieae W. D. J. Koch, with notes on evolution of their *psbA-trnH* sequences. *Plant Systematics and Evolution* 270: 139–157.

Lohse, M., O. Drechsel, and R. Bock. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* 52: 267–274.

Lowe, T. M. and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955–964.

Magee, A. R., B.-E. van Wyk, P. M. Tilney, and S. R. Downie. 2010. Phylogenetic position of African and Malagasy *Pimpinella* species and related genera (Apiaceae, Pimpinelleae). *Plant Systematics and Evolution* 288: 201–211.

Mardanov, A. V., N. V. Ravin, B. B. Kuznetsov, T. H. Samigullin, A. S. Antonov, T. V. Kolganova, and K. G. Skyabin. 2008. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *Journal of Molecular Evolution* 66: 555–564.

Michel, F., K. Umesono, and H. Ozeki. 1989. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82: 5–30.

Moore, M. J., A. Dhingra, P. S. Soltis, R. Shaw, W. G. Farmerie, K. M. Folta, and D. E. Soltis. 2006. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology* 6: 17.

Mower, J. P., S. Stefanovic, W. Hao, J. S. Gummow, K. Jain, D. Ahmed, and J. D. Palmer. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biology* 8: 150.

Nicolas, A. N. and G. M. Plunkett. 2009. The demise of subfamily Hydrocotyloideae (Apiaceae) and the re-alignment of its genera across the entire order Apiales. *Molecular Phylogenetics and Evolution* 53: 134–151.

Palmer, J. D. 1983. Chloroplast DNA exists in two orientations. *Nature* 301: 92–93.

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

Palmer, J. D. 1985b. Evolution of chloroplast and mitochondrial DNA in plants and algae. Pp. 131–240 in *Monographs in evolutionary biology: Molecular evolutionary genetics*, Chapter 3, ed. R. J. MacIntyre. New York: Plenum Publishing Corporation.

Palmer, J. D. 1986. Isolation and structural analysis of chloroplast DNA. *Methods in Enzymology* 118: 167–186.

Palmer, J. D. 1990. Contrasting modes and tempos of genome evolution in land plant organelles. *Trends in Genetics* 6: 115–120.

Palmer, J. D. 1991. Plastid chromosomes: structure and evolution. Pp. 5–53 in *Cell culture and somatic cell genetics of plants*, Vol. 7A, eds. L. Bogorad and I. K. Vasil. San Diego: Academic Press.

Perry, A. S. and K. H. Wolfe. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution* 55: 501–508.

Perry, A. S., S. Brennan, D. J. Murphy, T. A. Kavanagh, and K. H. Wolfe. 2002. Evolutionary re-organization of a large operon in adzuki bean chloroplast DNA caused by inverted repeat movement. *DNA Research* 9: 157–162.

Peery, R., J. V. Kuehl, J. L. Boore, and L. Raubeson. 2006. Comparisons of three Apiaceae chloroplast genomes—coriander, dill and fennel. Abstract, Botany 2006 Meeting.

Peery, R., S. R. Downie, J. V. Kuehl, J. L. Boore, and L. Raubeson. 2007. Chloroplast genome evolution in Apiaceae. Abstract, Botany 2007 Meeting.

Peery, R., S. R. Downie, R. K. Jansen, and L. A. Raubeson. 2011. Apiaceae organellar genomes. XVIII International Botanical Congress Abstract Book, p. 481.

Plunkett, G. M., D. E. Soltis, and P. S. Soltis. 1996a. Higher level relationships of Apiales (Apiaceae and Araliaceae) based on phylogenetic analysis of *rbcL* sequences. *American Journal of Botany* 83: 499–515.

Plunkett, G. M., D. E. Soltis, and P. S. Soltis. 1996b. Evolutionary patterns in Apiaceae: inferences based on *matK* sequence data. *Systematic Botany* 21: 477–495.

Plunkett, G. M. and S. R. Downie. 1999. Major lineages within Apiaceae subfamily Apioideae: a comparison of chloroplast restriction site and DNA sequence data. *American Journal of Botany* 86: 1014–1026.

Plunkett, G. M. and S. R. Downie. 2000. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. *Systematic Botany* 25: 648–667.

Raubeson, L. A. and R. K. Jansen. 2005. Chloroplast genomes of plants. Pp. 45–68 in *Plant diversity and evolution: Genotypic and phenotypic variation in higher plants*, ed. R. J. Henry. London: CAB International.

Raubeson, L. A., R. Peery, T. W. Chumley, C. Dziubek, H. M. Fourcade, J. L. Boore, and R. K. Jansen. 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8: 174.

Rice, D. W. and J. D. Palmer. 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, A. O. and J. D. Palmer. 2007. Horizontal gene transfer in plants. *Journal of Experimental Botany* 58: 1–9.

Ruhlman, T. A. and R. K. Jansen. 2014. The plastid genomes of flowering plants. Pp. 3–38 in *Chloroplast Biotechnology: Methods and Protocols*, ed. P. Maliga. New York: Springer.

Ruhlman, T., S.-B. Lee, R. K. Jansen, J. B. Hostetler, L. J. Tallon, C. D. Town, and H. Daniell. 2006. Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. *BMC Genomics* 7: 222.

Schattner, P., A. N. Brooks, and T. M. Lowe. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* 33: W686–W689.

Schwartz, S., L. Elnitski, M. Li, M. Weirauch, C. Riemer, A. Smit, NISC Comparative Sequencing Program, E. D. Green, R. C. Hardison, and W. Miller. 2003. MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Research* 31: 3518–3524.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.

Shaw, J., E. B. Lickey, E. E. Schilling, and R. L. Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94: 275–288.

Sheveleva, E. V. and R. B. Hallick. 2004. Recent horizontal intron transfer to a chloroplast genome. *Nucleic Acids Research* 32: 803–810.

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

Smith, D. R. 2011. Extending the limited transfer window hypothesis to inter-organelle DNA migration. *Genome Biology and Evolution* 3: 743–748.

Straub, S. C. K., R. C. Cronn, C. Edwards, M. Fishbein, and A. Liston. 2013. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biology and Evolution* 5: 1872–1885.

Sun, F.-J., S. R. Downie, and R. L. Hartman. 2004. An ITS-based phylogenetic analysis of the perennial, endemic Apiaceae subfamily Apioideae of western North America. *Systematic Botany* 29: 419–431.

Sun, F.-J. and S. R. Downie. 2010. Phylogenetic relationships among the perennial, endemic Apiaceae subfamily Apioideae of western North America: Additional data from the cpDNA *trnF-trnL-trnT* region continue to support a highly polyphyletic *Cymopterus*. *Plant Diversity and Evolution* 128: 151–172.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (* and other methods), v. 4.0 beta 10. Sunderland: Sinauer Associates.

Timme, R. E., J. V. Kuehl, J. L. Boore, and R. K. Jansen. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *American Journal of Botany* 94: 302–312.

Wang, R. J., C. L. Cheng, C. C. Chang, C. L. Wu, T. M. Su, and S. M. Chaw. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology* 8: 36.

Wang, W. and J. Messing. 2011. High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. *PLoS ONE* 6: e24670.

Weng, M.-L., J. C. Blazier, M. Govindu, and R. K. Jansen. 2014. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. *Molecular Biology and Evolution*, doi: 10.1093/molbev/mst257.

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

Wolfe, K. H., W.-H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences USA* 84: 9054–9058.

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

Yi, D.-K., H.-L. Lee, B.-Y. Sun, M.-Y. Chung, and K.-J. Kim. 2012. The complete chloroplast DNA sequence of *Eleutherococcus senticosus* (Araliaceae); Comparative evolutionary analyses with other three asterids. *Molecules and Cells* 33: 497–508.

Yu, Y., S. R. Downie, X. He, X. Deng, and L. Yan. 2011. Phylogeny and biogeography of Chinese *Heracleum* (Apiaceae tribe Tordylieae) with comments on their fruit morphology. *Plant Systematics and Evolution* 296: 179–203.

Yukawa, M., T. Tsudzuki, and M. Sugiura. 2005. The 2005 version of the chloroplast DNA sequence from tobacco (*Nicotiana tabacum*). *Plant Molecular Biology Reporter* 23: 359–365.