Aliso: A Journal of Systematic and Floristic Botany

Volume 32 | Issue 1 Article 3

2014

The Complete Plastid Genome Sequence of Iris gatesii (Section Oncocyclus), a Bearded Species from Southeastern Turkey

Carol A. Wilson
Rancho Santa Ana Botanic Garden, Claremont, California

Aliso, 32(1), pp. 47–54
ISSN 0065-6275 (print), 2327-2929 (online)

THE COMPLETE PLASTID GENOME SEQUENCE OF *IRIS GATESII* (SECTION *ONCOCYCLUS*), A BEARDED SPECIES FROM SOUTHEASTERN TURKEY

Carol A. Wilson

Rancho Santa Ana Botanic Garden and Claremont Graduate University, 1500 North College Avenue, Claremont, California 91711
(carol.wilson@cgu.edu)

ABSTRACT

Iris gatesii is a rare bearded species in subgenus *Iris* section *Oncocyclus* that occurs in steppe communities of southeastern Turkey. This species is not commonly cultivated, but related species in section *Iris* are economically important horticultural plants. The complete plastid genome is reported for *I. gatesii* based on data generated using the Illumina HiSeq platform and is compared to genomes of 16 species selected from across the monocotyledons. This *Iris* genome is the only known plastid genome available for order Asparagales that is not from Orchidaceae. The *I. gatesii* plastid genome, unlike orchid genomes, has little gene loss and rearrangement and is likely to be similar to other genomes from Asparagales. The plastid genome of *I. gatesii* demonstrates expansion of the inverted repeat, loss of 95% of the *rps19–rpl22* intergenic spacer, the presence of introns in several protein-coding regions, and alternate start codons. Potentially variable regions are identified for further study.

Key words: Asparagales, chloroplast genome, divergence hotspots, Iridaceae, monocotyledons.

Advances in sequencing methodologies have resulted in the publication of whole plastid genomes across Angiosperms. Early work concentrated on the model plant species *Arabidopsis thaliana* (L.) Heynh. and crop species in the grass family. As costs declined, genomic sequencing was expanded to species that were not widely studied, resulting in the deposition of approximately 350 plastid genomes in Genbank. Approximately 125 species of monocotyledons, representing all orders except Commelinales and Pandanales, have plastid genomes deposited in Genbank, and 55 of these genomes are species of Poaceae. The plastid genome in angiosperms is generally conserved in organization, gene order, and gene content. A quadripartite structure is composed of two copies of an inverted repeat (IR) and two unique regions, a large single copy (LSC), and a smaller single copy (SSC). In spite of this overall similarity, there are some striking differences among plastid genomes. Jansen et al. (2007) investigated 81 plastid genes across angiosperms and reported gene losses in monocotyledons including *accD* (*Acorus* L. and Poaceae), *rsp16* (*Dioscorea* L.), *ndh* genes (*Phalaenopsis* Blume), *rpl32* (*Yucca* L.), and *ycf1* and *ycf2* (Poaceae). More recent studies using next-generation sequencing (NGS) have added to our knowledge of the plastid genomes of monocotyledons. Studies by Guisinger et al. (2010) showed that while Poaceae genera either lack *accD*, *ycf1*, and *ycf2* genes entirely or have lost all except remnants of these genes, *Typha latifolia* (Typhaceae, Poales) retains them. Yang et al. (2013) sequenced five *Cymbidium* Sw. species and found that unlike *Phalaenopsis* the *Cymbidium* species retained most of their *ndh* genes although only a few of these retained genes were functional. Studies on the palm family (Arecaceae, Arecales) have revealed a plastid genome that is typical for angiosperms, although *ycf1* was determined to be a pseudogene in the family (Huang et al. 2013). In *Colocasia esculenta* (Araceae, Alismatales) *infA* occurs as a pseudogene (Ahmed et al. 2012).

Although Asparagales are strongly supported their sister group is uncertain. APG III (2009) suggested that Asparagales are sister to commelinids (Dasypogonaceae + Arecales, Poales, Zingiberales, and Commelinales) with Liliales as sister to Asparagales + commelinids. Asparagales are diverse, with 14 families and about 1120 genera (APG 2009) yet plastid genomes are only available for approximately 12 species and all of these are from one family, Orchidaceae. Orchidaceae are early diverging within Asparagales (Pires et al. 2006; Seberg et al. 2012; Chen et al. 2013) and are estimated to have a stem node age of 95.7 MYA and a crown node age of 51.6 MYA (Chen et al. 2013). The orchid plastid genomes investigated to date lack all or some of their *ndh* genes or coding functions (Chang et al. 2006; Wu et al. 2010; Pan et al. 2012; Yang et al. 2013) and are unlikely to be typical for other members of Asparagales such as Iridaceae.

Phylogenetic studies of *Iris* L. based on plastid data have resulted in a more complete understanding of subgeneric relationships and circumscription (Wilson 2009, 2011; Guo and Wilson 2013) and provide a framework to further explore relationships within major clades. This data has, however, been inadequate to determine relationships of species within some recently divergent clades, such as section *Oncocyclus* (Siems.) Baker in subgenus *Iris* (Wilson 2011; Wilson unpubl. data) and series *Californicae* (Diels) G.H.M. Lawr. in subgenus *Limniris* Tausch (Wilson 2009) even though species are easily identified based on morphology. As a result, further exploration of the plastid genome may help explain the low taxonomic signal present and provide information on regions within the genome that may be more informative. *Iris gatesii* Foster was selected for further study because it (1) has been recognized as distinct for more than a century, (2) is geographically isolated from other members of the section,

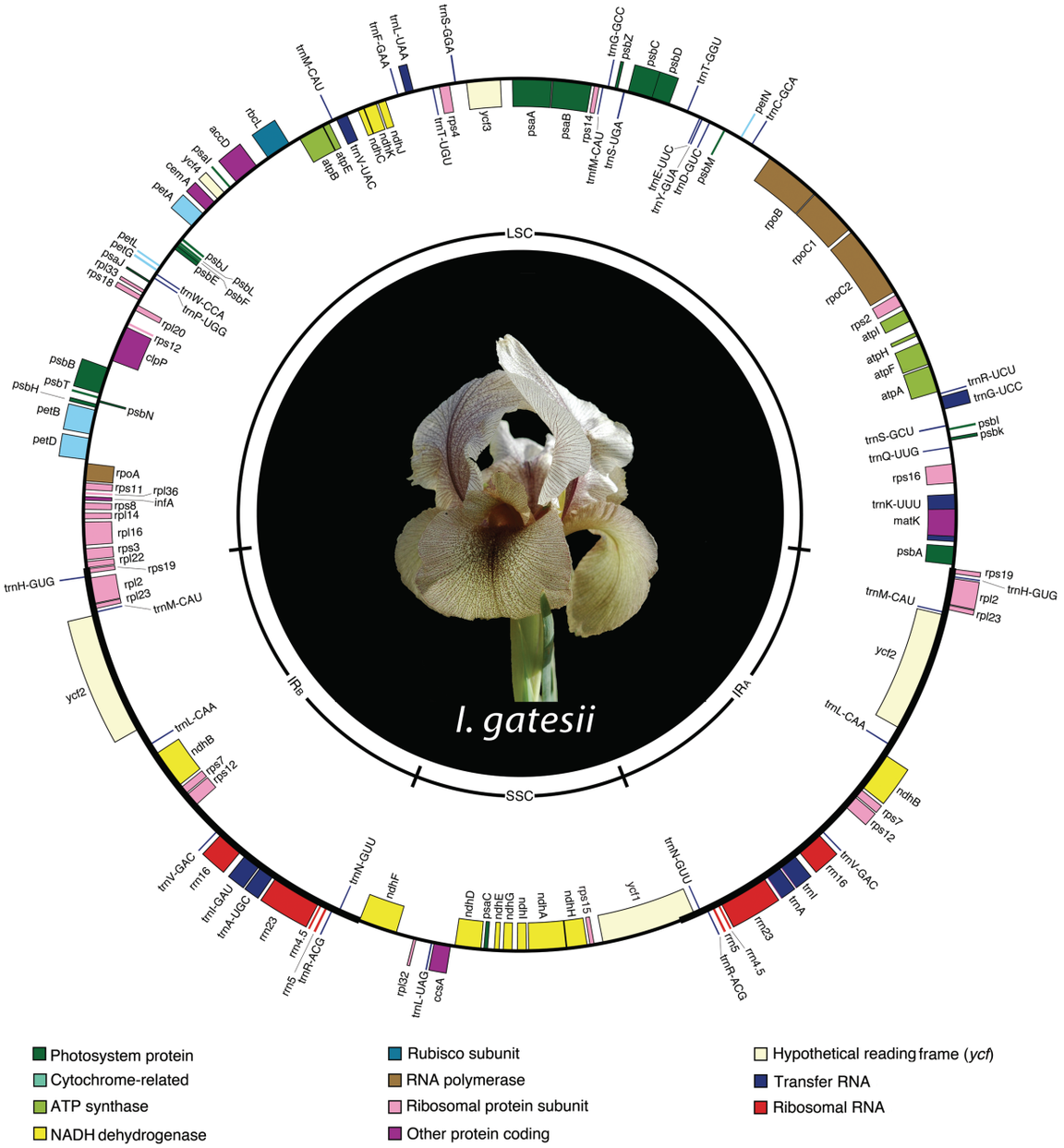

Fig. 1. Gene map of the complete plastid genome of *Iris gatesii* showing IR, SSC, and LSC regions. Genes belonging to different functional groups are color coded according to key. Genes drawn inside the circle are transcribed clockwise and those outside the circle anti-clockwise. Plant image by C. Wilson, taken east of Siirt, Turkey.

(3) has the typical chromosome number for section *Oncocyclus* ($n = 10$), a number that is relatively low for *Iris*, (4) has a limited distribution, and (5) is morphologically distinct. *Iris gatesii* is a rhizomatous, perennial species endemic to rocky hillsides in southeastern Turkey. The species has a single, large, cream-to-yellow flower sprinkled with purplish spots, that is borne above a fan of distichous leaves (Fig. 1). Each sepal has a concentration of purplish spots near the base of its free portion that is considered a signal spot. The signal spot is adjacent and distal to a relatively sparse beard of long purple hairs that extend into the perianth tube. The species was described in 1890 based on a plant collected near Mardin, Turkey, and is known only from a few sites in the surrounding region.

Here I present the complete nucleotide sequence of the plastid genome of *I. gatesii*. This study provides a reference genome for Asparagales outside Orchidaceae. This study also compares the *Iris* genome to other monocotyledonous genomes and explores areas of the plastid genome that might be informative in resolving the *Iris* phylogeny. Finally, this study is a first step in a more extensive comparison of plastid genome diversity within *Iris*.

Table 1. Order and family of *Iris gatesii* and 16 comparator species with their Genbank numbers.

Order	Family	Species	Source	Genbank
Acorales	Acoraceae	*Acorus calamus* L.	Goremykin et al. 2005	NC_007407
Alismatales	Araceae	*Colocasia esculenta* (L.) Schott	Ahmed et al. 2012	NC_016753
		Spirodela polyrhiza (L.) Schleid.	Wang and Messing 2011	NC_015891
Dioscoreales	Dioscoreaceae	*Dioscorea elephantipes* (L'Hér.) Engl.	Hansen et al. 2007	NC_009601
Liliales	Liliaceae	*Fritillaria taipaiensis* P.Y. Li	Li et al. unpubl.	NC_023247
	Melanthiaceae	*Veratrum patulum* Loes.	Do et al. 2013	NC_022715
	Smilacaceae	*Smilax china* Walter	Lu unpubl.	HM536959
Asparagales	Iridaceae	*Iris gatesii* Foster	This study	KM014691
	Orchidaceae	*Cymbidium tracyanum* Rolfe	Yang et al. 2013	NC_021432
		Erycina pusilla (L.) N.H. Williams & M.W. Chase	Pan et al. 2012	NC_018114
Arecales	Arecaceae	*Calamus caryotoides* A. Cunn. ex Mart.	Barrett et al. 2013	NC_020365
		Phoenix dactylifera L.	Yang et al. 2010	NC_013991
Unplaced	Dasypogonaceae	*Dasypogon bromeliifolius* R. Br.	Barrett et al. 2013	NC_020367
Poales	Typhaceae	*Typha latifolia* L.	Guisinger et al. 2010	NC_013823
Zingiberales	Heliconiaceae	*Heliconia collinsiana* Griggs	Barrett et al. 2013	NC_020362
	Musaceae	*Musa textilis* Née	Barrett et al. 2013	NC_022926
	Zingiberaceae	*Zingiber spectabile* Griff.	Barrett et al. 2013	NC_020363

MATERIALS AND METHODS

DNA Extraction, Sequencing, and Assembly

Genomic DNA was isolated from silica-dried leaf materials of *I. gatesii* from near Mardin in southern Turkey (*C. Wilson T06-52*, 26 May 2006, RSABG) using protocols modified from the CTAB method of Doyle and Doyle (1987). Modifications of this procedure included an RNase treatment, an ethanol precipitation with ammonium acetate following the initial isopropanol precipitation, and a final ethanol rinse. Quality and quantity of the extracted DNA was determined by gel electrophoresis and spectrophotometry using a nanoVue spectrophotometer (GE Healthcare Bio-Sciences, Pittsburgh, PA). Sixty µl of DNA in sterile H_2O (3 µg total) was sent to the University of Missouri DNA Core Facility (Columbia, MO) for DNA library construction and NGS. Library construction used the Illumina TruSeq Kit (Illumina, Inc., San Diego, CA), and sequencing was performed on an Illumina HiSeq 2000. Illumina sequencing produced 54.9 million reads of approximately 100 base pair length each (27.5 million paired-end reads).

Initial de novo assembly of paired-end reads was performed at the University of Missouri Informatics Research Core Facility (Columbia, MO) using Velvet (v0.7.60; Zerbino and Birney 2008). The contig received from the Informatics Research Core Facility was blasted against the date palm, *Phoenix dactylifera*, plastid genome (NC_013991.2) to determine homology. The inverted repeat was identified and the final sequence was assembled in Geneious vers. 6.1 (Biomatters, Auckland, New Zealand) based on comparison with the *Phoenix dactylifera* plastid genome. Targeted sampling of the raw paired-end reads, especially in regions near the IRs, was used to confirm the resulting *I. gatesii* plastid genome sequence. Sampling included alignment of raw reads to sequence segments spanning the potential boundaries between IRs and single copy areas. Single and duplicated gene regions near the IRs were also verified by alignment of reads to these gene regions and known single and duplicated gene regions of similar length. Determination of gene copy was based on the relative abundance of aligned reads that resulted. Using Geneious, the assembled plastid sequence was annotated by comparison to the *Calamus caryotoides* (NC_020365) plastid genome and annotated protein-coding sequences were checked for start and stop codons.

The final annotated plastid genome was compared to genomes of 16 other monocotyledons (Table 1). Comparator species were selected based on searches, conducted during 10–14 Feb 2014, of Genbank (*www.ncbi.nlm.nih.gov*) including general searches of genomes for each order and a survey of the plastid genome resources page (*www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid*). Plastid genomes available in Genbank represented nine of 11 orders recognized by APG (2009) and one unplaced family, Dasypogonaceae. Species were selected to represent diversity within each order except Petrosaviales and are from 13 families. Genbank held plastid genomes from only one family each for five orders: Acorales, Alismatales, Asparagales, Arecales, and Petrosaviales. For Poales, sequences were available for Typhaceae and the economically important grass family, Poaceae. A Poaceae plastid genome was not included in final comparisons because initial comparisons with *Leersia tisserantii* (A. Chev.) Launert revealed rearrangements within the sequence that were divergent when compared to those of *I. gatesii* and the other monocots selected. Similarly, the plastid genome of the mycoheterotrophic *Petrosavia stellaris* Becc. (Petrosaviaceae, Petrosaviales) was not included because initial comparisons indicated several rearrangements within its genome. A discussion of highly rearranged plastid genomes is beyond the scope of this study.

RESULTS

Size, Organization, and Coding Regions

The complete plastid genome of *I. gatesii* is 153,441 base pairs (bp) in length and has a GC content of 39.7% (Fig. 1; Table 2). The LSC region is 82,659 bp, SSC region 18,376 bp, and IRb is 26,220 bp in length. There are 132 genes, of which 86 are predicted to be protein-coding DNA sequence (CDS). Of these, 72 CDSs are in single copy regions and 7 are

Table 2. Comparison of plastid genome size and protein-coding DNA sequence content for *Iris gatesii* and comparator monocot species. The first four columns show sizes in base pairs of the complete plastid genome and LSC, SSC, and IR regions; the remaining columns show the number of protein-coding (PC) genes, tRNA genes, and rRNA genes, and the % GC content for the complete plastid genome. The comparisons of PC genes are based on the 79 unique genes common to most angiosperms. The lengths of plastid regions that were estimated based on comparisons across genomes are in parentheses.

	Complete genome	LSC	SSC	IRb	PC	tRNA	rRNA	GC
Iris gatesii	153,441	82,659	18,376	26,220	79	30	4	39.7
Cymbidium tracyanum	156,286	84,968	17,929	26,695	68	30	4	46.1
Erycina pusilla	143,164	84,189	12,097	23,439	65	29	4	36.7
Acorus calamus	153,821	(83,136)	(18,194)	(26,469)	78	30	4	38.6
Colocasia esculenta	162,424	89,670	22,208	25,273	78	30	4	36.1
Spirodela polyrhiza	168,788	91,222	14,056	31,755	77	29	4	35.7
Dioscorea elephantipes	152,609	82,777	18,806	25,513	78	30	4	38.9
Fritillaria taipaiensis	151,693	(81,744)	(17,883)	(26,062)	78	30	4	37.0
Veratrum patulum	153,699	83,372	17,697	26,360	77	30	4	37.7
Smilax china	157,878	84,608	18,536	27,367	75	30	4	37.3
Calamus caryotoides	157,270	85,525	17,595	27,075	79	30	4	37.4
Phoenix dactylifera	158,462	86,198	17,712	27,276	79	30	4	37.2
Dasypogon bromeliifolius	157,858	(86,177)	(18,376)	(53,305)	79	30	4	37.2
Typha latifolia	161,572	89,140	19,652	26,390	79	30	4	33.8
Heliconia collinsiana	161,907	(90,002)	(26,741)	(18,699)	79	30	4	37.3
Musa textilis	161,347	88,338	10,768	35,433	79	30	4	36.8
Zingiber spectabile	155,890	(86,045)	(18,496)	(25,649)	79	30	4	36.3

duplicated in the IR. No pseudogenes were detected. Coding regions comprise 59% of the *I. gatesii* plastid genome, have a GC content of 40%, and include 79 unique protein-coding genes, 30 tRNAs, and 4 rRNAs. The IR extends into the LSC with duplications of the *trnH*-GUG and *rps19* genes and loss of 95% of the *rps19–rpl22* intergenic spacer (Fig. 1). As is common among angiosperms, *ycf1* spans the SSC/IRa junction and the 5′-end of the gene is duplicated at the SSC end of the IRA.

Comparisons Among the Plastid Genomes of Iris gatesii *and Other Monocotyledons*

Of the 16 comparator species only three (*Erycina pusilla*, *Dioscorea elephantipes*, and *Fritillaria taipaiensis*) had plastid genomes shorter than that of *I. gatesii* (Table 2). The shortest genome was *Erycina pusilla* with 143,164 bp and the longest was *Spirodela polyrhiza* with 168,788 bp. Comparator genera from Asparagales, *Erycina pusilla* and *Cymbidium tracyanum* (Orchidaceae), had the greatest loss of protein-coding genes or their function (Table 2) with both plastid genomes lacking genes or coding regions of *ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhF*, *ndhG*, *ndhH*, *ndhK*, and *ndhL*. In addition, the plastid genome of *Erycina pusilla* lacked the *trnG*-UCC tRNA and CDSs for *ndhE*, *ndhJ*, and *ycf1*. Plastid genomes representing Acorales, Alismatales, Dioscoreales, and Liliales also lacked coding regions (Table 2) although they retained each of the 11 *ndh* genes. Two commonly lost protein-coding regions, *infA* that codes for initiation-factor 1 and *accD* that encodes the beta-carboxyl transferase subunit of acetyl- CoA carboxylase, were present in plastid genomes of *I. gatesii* and the two orchids. *InfA* is a relatively small gene (234 bp for taxa in this study) that has been lost from the plastid genome multiple times across angiosperms (Millen et al. 2001) including the comparator species *Acorus calamus* (Acorales) and *Smilax china* (Liliales). *AccD* is a larger gene (ca. 1480 bp for taxa in this study) that has been lost from plastid genomes of some angiosperm families (Rousseau-Gueutin et al. 2013) including the two species representing Alismatales (*Spirodela polyrhiza* and *Colocasia esculenta*) and the three species included in this study from Liliales (*Smilax china*, *Veratrum patulum*, and *Fritillaria taipaiensis*). In some cases *accD* and *infA* have been reported as transferred to the nuclear genome (Millen et al. 2001; Rousseau-Gueutin et al. 2013). Other coding regions missing from the plastid genome of comparator species included *psaJ* and *ycf1* (*Smilax china*), *rps12*, *rpl20*, and *trnH*-GUG tRNA (*Spirodela polyrhiza*), and *rps16* (*Veratrum patulum* and *Dioscorea elephantipes*).

While most initiation codons conformed to the commonly used ATG (methionine), several alternate start codons were detected in the plastid genome of *I. gatesii*, including ACG (threonine) for *ndhD* and *rpl2* and GTG (valine) for *rps19* (Table 3). Alternate start codons were also present in genomes of several of the comparator taxa (Table 3). The most frequent alternate start codon across all genomes was ACG (threonine), a start codon common for the *rpl2* coding region. All except three genomes had GTG (valine) as an alternate start codon for *rps19*.

A comparison of CDS lengths revealed nine regions where the number of nucleotides for *I. gatesii* differed by at least 2% from either the mean CDS length across the 14 genomes of comparator species outside Asparagales or from the CDS length of the two orchid species (Table 4). The coding regions *rpl22*, *psaJ*, *accD*, *matK*, *ycf2*, and *ycf1* of *I. gatesii* had one, one, two, three, seven, and 20 indels, respectively, relative to most comparator species. Twenty-six of these indels were relatively short (3, 6, or 9 bp in length). Four indels were of moderate length (15, 18, 21, and 21 bp). The *rpl22* CDS for *I. gatesii* had only one indel and it was 65 bp in length while the *ycf1* had 17 short indels and three long indels of 39, 42, and 147 bp. Differences in CDS lengths for *petD* and *rpl20* were due to frame shifts resulting in positional changes of stop codons across species. The coding region of the *rps16* gene was similar across most species including *I. gatesii* but was

Table 3. Alternate start codons detected in four protein-coding DNA sequence regions of the *Iris gatesii* and comparator genomes. Alternate three-letter start codons are followed by the single-letter code for the corresponding amino acid.

	rpl16	*rps19*	*rpl2*	*ndhD*
Iris gatesii	ATG = M	GTG = V	ACG = T	ACG = T
Cymbidium tracyanum	ATG = M	GTG = V	ATA = I	Lacks CDS
Erycina pusilla	ATG = M	GTG = V	ATA = I	Lacks CDS
Acorus calamus	ATG = M	GTG = V	ACG = T	ACG = T
Colocasia esculenta	ATG = M	GTG = V	ACG = T	ACG = T
Spirodela polyrhiza	ATC = I	GTG = V	ATG = M	ATG = M
Dioscorea elephantipes	ATG = M	GTG = V	ACG = T	ACG = T
Fritillaria taipaiensis	ATG = M	GTG = V	ACG = T	ATC = I
Veratrum patulum	ATG = M	GTG = V	ACG = T	ACG = T
Smilax china	ATG = M	GTG = V	ATA = I	ATC = I
Calamus caryotoides	ATG = M	GTG = V	ACG = T	ATC = I
Phoenix dactylifera	ATG = M	GTG = V	ACG = T	ACG = T
Dasypogon bromeliifolius	ATG = M	GTG = V	ACG = T	ATC = I
Typha latifolia	ATG = M	GTG = V	ACG = T	ATC = I
Heliconia collinsiana	ATG = M	ATG = M	ACG = T	ATC = I
Musa textilis	ATG = M	ATG = M	ATA = I	ATC = I
Zingiber spectabile	ATG = M	ATG = M	ACG = T	ATC = I

considerably longer in the two orchid species because of a shared 33 bp insertion.

Duplications of the *trnH*-GUG and *rps19* coding regions reported above for *I. gatesii* were shared with the orchids *Erycina pusilla* and *Cymbidium tracyanum* and six comparator species representing Dasypogonaceae, Poales, Musaceae, Arecaceae, and Smilacaceae. In these species, unlike *I. gatesii* where 95% of the *rps19–rpl22* intergenic spacer region was lost, the intergenic spacer region was retained although it varied in length across species. In some species this spacer region was also duplicated along with the *trnH*-GUG and *rps19* coding regions. The *trnS*-UGA tRNA was reversed in *I. gatesii* and *Calamus caryotoides* relative to other comparator species.

DISCUSSION

Currently, the only Asparagales plastid genomes available in Genbank are twelve orchid sequences. The plastid genome of species in the early-diverging but highly derived orchid lineage differs from that of most monocots by the shared loss of at least some photosynthesis genes. The two orchid species included in this study, *Cymbidium tracyanum* and *Erycina pusilla*, are photosynthetic yet demonstrate significant loss of the 11 *ndh* genes or their coding regions. Similar losses of *ndh* genes have been reported in the photosynthetic orchids, *Oncidium* Gower Ramsey (Wu et al. 2010) and *Phalaenopsis* (Chang et al. 2006), while the complete loss of all photosynthetic genes or at least their function have been reported from the non-photosynthetic species, *Neottia nidus-avis* (L.) Rich (Logacheva et al. 2011) and *Rhizanthella gardneri* R.S. Rogers (Delannoy et al. 2011). The loss of photosynthesis genes in non-photosynthetic parasitic plants and orchid mycoheterotrophs is not surprising. It is less obvious why fully photosynthetic plants have lost their photosynthesis genes. Gene losses in the plastid genome that are associated with current and ongoing plant functions are often attributed to transfer of these genes to the nuclear genome (Wakasugi et al. 1994; Wu et al. 2010; Pan et al. 2012). Transfer of orchid photosynthesis genes to the nuclear genome is suggested by studies of Chang et al. (2006) who recovered intact *ndhA*, *ndhF*, and *ndhH* CDSs from total DNA of *Phalaenopsis aphrodite* Rchb. f. even though these coding regions were not present in its plastid genome. *Erycina pusilla* also lacks the *ycf1* gene and *trnG*-UCC tRNA while in *Cymbidium tracyanum* the positions of *petM* and *petN* genes are reversed. These considerable changes in gene content and organization within

Table 4. A comparison of length variation (base pairs) of nine coding regions (CDS) among plastid genomes of *Iris gatesii* and all 16 comparator species. Ranges and means refer to all comparator species except the orchids. Actual lengths are given for *I. gatesii* and each of the two orchid species.

CDS	Range	Mean	*Iris gatesii*	*Erycina pusilla*	*Cymbidium tracyanum*
accD	1464–1799	1512	1473	1467	1455
matK	1407–1560	1531	1569	1557	1554
petD	482–527	499	489	492	489
psaJ	129–135	131	129	135	135
rpl20	354–429	363	354	363	360
rpl22	306–420	371	309	366	360
rps16	234–261	252	258	279	285
ycf1	5348–5972	5599	5406	Lacks CDS	5409
ycf2	6231–7011	6805	6876	6678	6849

the Orchidaceae indicate that plastid evolution within this lineage is divergent relative to most monocots and, based on *I. gatesii*, presumably most Asparagales.

This study demonstrates that unlike the orchid plastid genome, the gene content and organization of the *I. gatesii* genome is similar to most other monocot plastid genomes and is likely to be similar to genomes from other Asparagales families. Each of the commonly occurring protein-coding genes, tRNAs, and rRNAs are present in *I. gatesii* with no evidence of gene loss or pseudogene development. Expansion of the IR, contraction of the LSC, and alternate start codons are present in *I. gatesii* but these organizational changes are common across monocots and do not represent significant reorganizations of the genome. *Iris* is a diverse genus with bearded and non-bearded species; species with rhizomes, bulbs, or tuberous roots; mesic and xeric species; and species with unifacial or dorsiventral leaves that may be large and obvious or reduced and only a few centimeters tall. Researchers have found divergent plastid genomes that may be related to habitat or plant life form (Delannoy et al. 2011; Logacheva et al. 2011; Wang and Messing 2011; Peredo et al. 2013). The diversity of habitats utilized and morphological forms present in *Iris* provides an opportunity to explore plastid genome diversity between sister clades and among more distantly related clades across diverse plant forms and modes of habitat utilization. Preliminary studies (Wilson 2013) indicate that plastid genomes of rhizomatous *Iris* species are quite different from genomes of some bulbous species.

The expansion of the IR at the 3′ LSC boundary to include *trnH*-GUG is typical for monocots and also occurs in some early diverging angiosperms (Wang et al. 2008). Wang et al. (2008) also demonstrated a further expansion of the IR at the 3′ LSC boundary to include *rps19* in plastid genomes of Asparagales and commelinid (Commelinales, Zingiberales, Arecales, Dasypogonaceae, and Poales) species included in their study. They concluded that plastid genomes of the more derived monocots, including Asparagales, and members of the commelinid clade, are all likely to share an expansion of the IR to include *rps19*. They further concluded that the endpoint of the IR is highly conserved because across Asparagales and commelinids only 35–99 bp were duplicated beyond *rps*19. The Wang et al. (2008) study is based on results of Sanger sequencing at the IR-LSC boundaries to identify genes present and reverse transcriptase-polymerase chain reaction (RT-PCR) assays to verify transcription of the *trnH*-GUG and/or *rps19* genes. The results of this study are not entirely concordant with the Wang et al. study with respect to the duplication of *rps19*. In support of Wang et al. (2008), this study detected a duplication of *rps19* in each of the three Asparagales genomes, including *I. gatesii*, and five of seven genomes representing commelinid comparator species. In this study, duplication of *rps19* was not detected in the plastid genomes of two commelinid species, *Zingiber spectabile* and *Heliconia collinsiana*, but was detected in the genome of *Smilax china*, one of three comparator species from Liliales. Wang et al. (2008) did not detect a duplication of *rps19* in the *Lilium formosanum* Wallace genome, the only Liliales included in their study. In the current study, conclusions about gene content of the plastid genome of comparator species are based on annotated plastid genomes available from Genbank rather than experimental studies and are dependent, in part, on decisions made by other researchers. In addition, species included in each study have little overlap. An interesting finding for the *I. gatesii* genome is that most of the *rps19–rpl22* intergenic spacer from the LSC region has been lost. This loss was not detected in genomes of comparator species nor mentioned for genomes included in Wang et al. (2008), indicating that expansion of the IR and contraction of the LSC regions may be more dynamic in the *I. gatesii* genome than in many other monocot plastid genomes. The present study demonstrates that a more complete understanding of the evolution of IR expansion and LSC contraction across monocot genomes requires additional studies that include more representative species from each order.

The most common start codon for protein coding genes, ATG (methionine), begins translation in almost all of the protein coding genes present in *I. gatesii* and comparator genomes. The alternative start codon, GTG (valine), is present in the *rps19* gene of *I. gatesii* and most comparator genomes. The alternate start codons ATC and ATA (isoleucine) are common in *rp12* and *ndhD* genes of comparator genomes but are not found in *I. gatesii* where instead ACG (threonine) occupies the start position of both genes. Alternative start codons have been reported in many non-monocot angiosperm plastid genomes including genomes of early diverging groups such as *Amborella* Baill. and Nymphaceae where GTG and ACG have been detected (Raubeson et al. 2007). Using RNAseq data, Lee et al. (2014) detected 30 RNA editing sites in the plastid genome of *Deschampsia antarctica* E. Desv. resulting in the conversions of C to U, U to C, A to C, A to G, G to A, G to C, U to A, A to U, and U to G. Seventeen of the substitutions in the Lee et al. (2009) study were C to U and three were G to A. It has been shown in *Nicotiana* L. (Kudla et al. 1992) and suggested for *Eucalyptus* L'Hér. (Steane 2005) that C to U editing of the ACG codon results in correct translation. It is likely that for the two alternative start codons detected in the *I. gatesii* genome (GTG, ACG), RNA editing results in the mRNA codon (AUG) for methionine and the incorporation of N-formylmethionine as the first amino acid.

Plastid genome evolution is often conservative with relatively few nucleotide substitutions and indel mutations generated during DNA replication or DNA damage. Studies have suggested that sequence regions with indels often also have nucleotide substitutions (Tian et al. 2008; Zhu et al. 2009; Ahmed et al. 2012; Sloan et al. 2014), making these regions divergence hotspots that may be particularly informative for the resolution of relationships among closely related taxa. Nine coding regions were identified as potentially divergent in the *I. gatesii* plastid genome based on differences in CDS length when compared to orchid and other comparator genomes. Length differences for seven of these regions are due to indels while changes in the position of the stop codon are responsible for length variation in two regions. Four of the CDS regions (*accD*, *matK*, *ycf1*, *ycf2*) have multiple indels in *I. gatesii* that contribute to length differences that exceed 100 bps for at least one pairwise comparison between comparators and *Iris*. In *I. gatesii* the CDS regions for *ycf2* and *ycf1* are particularly rich in indels and will be trialed for Sanger sequencing in *Iris*. Preliminary studies based on about 1000 bp of the *ycf1* gene indicate that this region is relatively variable across species in two lineages where relationships have been

difficult to resolve, subgenus *Iris* and subgenus *Limniris* series *Californicae* (Wilson unpubl. data).

In conclusion, the plastid genome of *I. gatesii* is similar to that of other monocotyledons with little gene loss and rearrangement. It is likely to be more similar to most Asparagales than the orchid genomes currently available. Similar to many monocotyledons the *I. gatesii* plastid genome demonstrates an expansion of the IR, while the genome also has an almost complete loss of the intergenic spacer *rps19–rpl22*, a loss that is not shared with genomes from comparator species. Alternate start codons occur in the plastid genome of *I. gatesii*, and it is suspected that RNA editing results in the mRNA codon for methionine. Plastid genome regions were identified that may be variable within *Iris*, and targeted Sanger sequencing is underway to determine if resolution of closely related taxa can be improved.

ACKNOWLEDGMENTS

This research was supported by grants from the American Iris Society Foundation and NSF: DEB-1020826.

LITERATURE CITED

Ahmed, I., P. J. Biggs, P. J. Matthews, L. J. Collins, M. D. Hendy, and P. J. Lockhart. 2012. Mutational dynamics of aroid chloroplast genomes. *Genome Biology and Evolution* **4**: 1316–1323.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**: 105–121.

Barrett, C. F., J. I. Davis, J. Leebens-Mack, D. Stevenson, and J. G. Conran. 2013. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics* **29**: 65–87.

Chang, C. C., H. C. Lin, I. P. Lin, T. Y. Chow, H. H. Chen, W. H. Chen, C. H. Cheng, C. Y. Lin, S. M. Liu, C. C. Chang, and S. M. Chaw. 2006. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Molec. Biol. Evol.* **23**: 279–291.

Chen, S., D.-K. Kim, M. W. Chase, and J.-H. Kim. 2013. Networks in a large-scale phylogenetic analysis: reconstructing evolutionary history of Asparagales (Lilianae) based on four plastid genes. *PLoS ONE* **8**: e59472.

Delannoy, E., S. Fujii, C. Colas des Francs-Small, M. Brundrett, and I. Small. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Molec. Biol. Evol.* **28**: 2077–2086.

Do, H. D., J. S. Kim, and J. H. Kim. 2013. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). *Gene* **10**: 229–235.

Doyle, J. J. and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**: 11–15.

Goremykin, V. V., B. Holland, K. I. Hirsch-Ernst, and F. H. Hellwig. 2005. Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Molec. Biol. Evol.* **22**: 1813–1822.

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. *J. Molec. Evol.* **70**: 149–166.

Guo, J. and C. A. Wilson. 2013. Molecular phylogenetic study of the crested *Iris* based on five plastid markers. *Syst. Bot.* **38**: 987–995.

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

Huang, Y.-Y., A. J. M. Matzke, and M. Matzke. 2013. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS ONE* **8**: e74736.

Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. dePamphilis, J. Leebens-Mack, K. F. Müller, M. Guisinger-Bellian, R. C. Haberle, A. K. Hansen, T. W. Chumley, S.-B. Lee, R. Peery, J. R. McNeal, J. V. Kuehl, and J. L. Boore. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 19369–19374.

Kudla, J., G. L. Igloi, M. Metzlaff, R. Hagemann, and H. Kössel. 1992. RNA editing in tobacco chloroplasts leads to the formation of a translatable *psbL* mRNA by a C to U substitution within the initiation codon. *E. M. B. O. J.* **11**: 1099–1103.

Lee, J., Y. Kang, S. C. Shin, H. Park, and H. Lee. 2014. Combined analysis of the chloroplast genome and transcriptome of the antarctic vascular plant *Deschampsia antarctica* Desv. *PLoS ONE* **9**: e92501.

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

Millen, R. S., R. G. Olmstead, K. L. Adams, J. D. Palmer, N. T. Lao, L. Heggie, T. A. Kavanagh, J. M. Hibberd, J. C. Gray, C. W. Morden, P. J. Calie, L. S. Jermiin, and K. H. Wolfe. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Pl. Cell* **13**: 645–658.

Pan, I.-C., D.-C. Liao, F.-H. Wu, H. Daniell, N. D. Singh, C. Chang, M.-C. Shih, M.-T. Chan, and C.-S. Lin. 2012. Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical *Oncidium* breeding. *PLoS ONE* **7**: e34738.

Peredo, E. L., U. M. King, and D. H. Les. 2013. The plastid genome of *Najas flexilis*: adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. *PLoS ONE* **8**: e68591.

Pires, J. C., I. J. Maureira, T. J. Givnish, K. J. Sytsma, O. Seberg, G. Petersen, J. I. Davis, D. W. Stevenson, P. J. Rudall, M. F. Fay, and M. W. Chase. 2006. Phylogeny, genome size, and chromosome evolution of Asparagales. *Aliso* **22**: 287–304.

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*. *B. M. C. Genomics* **8**: 174.

Rousseau-Gueutin, M., X. Huang, E. Higginson, M. Ayliffe, A. Day, and J. N. Timmis. 2013. Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to the nucleus in some angiosperm lineages. *Pl. Physiol.* **161**: 1918–1929.

Seberg, O., G. Petersen, J. I. Davis, J. C. Pires, D. W. Stevenson, M. W. Chase, M. F. Fay, D. S. Devey, T. Jørgensen, K. J. Sytsma, and Y. Pillon. 2012. Phylogeny of the Asparagales based on three plastid and two mitochondrial genes. *Amer. J. Bot.* **99**: 875–889.

Sloan, D. B., D. A. Triant, N. J. Forrester, L. M. Bergner, M. Wu, and D. R. Taylor. 2014. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). *Molec. Phylogen. Evol.* **72**: 82–89.

Steane, D. A. 2005. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *D. N. A. Res.* **12**: 215–220.

Tian, D., Q. Wang, P. Zhang, H. Araki, S. Yang, M. Kreitman, T. Nagylaki, R. Hydson, J. Bergelson, and J. Q. Chen. 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature* **455**: 105–108.

Wakasugi, T., J. Tsudzuki, S. Ito, K. Nakashima, T. Tsudzuki, and M. Sugiura. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 9794–9798.

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. *B. M. C. Evol. Biol.* **8**: 36.

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

Wilson, C. A. 2009. Phylogenetic relationships among the recognized series in *Iris* section *Limniris*. *Syst. Bot.* **34**: 277–284.

———. 2011. Subgeneric classification in *Iris* re-examined using chloroplast sequence data. *Taxon* **60**: 27–35.

———. 2013. Shifts in geophytic form across non-uniform habitats are linked to plastid genome structural changes and rates of diversification in *Iris*, pp. 129–130. *In* Monocots V: 5th International Conference on Comparative Biology of Monocots, 7–13 Jul 2013, New York Botanical Garden & Fordham University, New York, USA [abstract].

Wu, F.-H., M.-T. Chan, D.-C. Liao, C.-T. Hsu, Y.-W. Lee, H. Daniell, M. R. Duvall, and C.-S. Lin. 2010. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. *B. M. C. Pl. Biol.* **10**: 68.

Yang, M., X. Zhang, G. Liu, Y. Yin, K. Chen, Q. Yun, D. Zhao, I. S. Al-Mssallem, and J. Yu. 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* **5**: e12762.

Yang, J.-B., M. Tang, H.-T. Li, Z.-R. Zhang, and D.-Z. Li. 2013. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *B. M. C. Evol. Biol.* **13**: 84.

Zerbino, D. R. and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**: 821–829.

Zhu, L., Q. Wang, P. Tang, H. Araki, and D. Tian. 2009. Genomewide association between insertions/deletions and the nucleotide diversity in bacteria. *Molec. Biol. Evol.* **26**: 2353–2361.