

RESEARCH ARTICLE

Open Access



Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family Orobanchaceae

Daniel C. Frailey, Srinivasa R. Chaluvadi, Justin N. Vaughn, Caroline G. Coatney and Jeffrey L. Bennetzen*

Abstract

Background: The chloroplast genomes (plastome) of most plants are highly conserved in structure, gene content, and gene order. Parasitic plants, including those that are fully photosynthetic, often contain plastome rearrangements. These most notably include gene deletions that result in a smaller plastome size. The nature of gene loss and genome structural rearrangement has been investigated in several parasitic plants, but their timing and contributions to the adaptation of these parasites requires further investigation, especially among the under-studied hemi-parasites.

Results: De novo sequencing, assembly and annotation of the chloroplast genomes of five photosynthetic parasites from the family Orobanchaceae were employed to investigate plastome dynamics. Four had major structural rearrangements, including gene duplications and gene losses, that differentiated the taxa. The facultative parasite *Aureolaria virginica* had the most similar genome content to its close non-parasitic relative, *Lindenbergia philippensis*, with similar genome size and organization, and no differences in gene content. In contrast, the facultative parasite *Buchnera americana* and three obligate parasites in the genus *Striga* all had enlargements of their plastomes, primarily caused by expansion within the large inverted repeats (IRs) that are a standard plastome feature. Some of these IR increases were shared by multiple investigated species, but others were unique to particular lineages. Gene deletions and pseudogenization were also both shared and lineage-specific, with particularly frequent and independent loss of the *ndh* genes involved in electron recycling.

Conclusions: Five new plastid genomes were fully assembled and compared. The results indicate that plastome instability is common in parasitic plants, even those that retain the need to perform essential plastid functions like photosynthesis. Gene losses were slow and not identical across taxa, suggesting that different lineages had different uses or needs for some of their plastome gene content, including genes involved in some aspects of photosynthesis. Recent repeat region extensions, some unique to terminal species branches, were observed after the divergence of the *Buchnera/Striga* clade, suggesting that this otherwise rare event has some special value in this lineage.

Keywords: Chloroplast, Chromosome rearrangement, Gene deletion, Parasite, Plastome, *Striga*

Background

Beyond its essential roles in photosynthesis, the chloroplast is also the specialized site for synthesis of some pigments, lipids, amino acids, and sulfur compounds [1, 2]. The chloroplast contains its own genome, the plastome. The plastome in vascular plants is believed to have a single endocytotic origin, more than a billion years ago, from a cyanobacterium-like organism [3–5]. Over time, most

genes of the cyanobacterium have either been lost or have migrated to the nuclear and/or mitochondrial genomes, so that modern plant plastomes contain approximately 10% of the ancestral endosymbiont genes [6].

Plastomes are under strong selective pressure in most plants and therefore tend to be highly conserved in terms of gene content and order, nucleotide substitution rates, structure, and size [7]. Most plastomes range between 120 to 160 kb in size [1]. Originally, plastomes were thought to be arranged as single circular molecules. Although a percent of plastomes do have this structure,

* Correspondence: maize@uga.edu

Department of Genetics, University of Georgia, Athens, GA 30677, USA



they are also present in dimers and higher order concatemers, in both circular and linear configurations [2]. Plastomes have a quadripartite structure consisting of a large single copy (LSC) region and a small single copy region (SSC) separated by two virtually identical inverted repeat (IR_A and IR_B) regions [8]. Recombination between the two inverted repeats plays a role in stabilizing the plastome [9]. In angiosperms, LSCs are typically 80–90 kb, SSCs 16–27 kb, and inverted repeats 20–30 kb [2]. The plastome contains genes encoding proteins involved in photosynthesis as well as the other biochemical pathways carried out in the chloroplast. It also encodes 30 tRNAs and several other structural RNAs [10–12].

Unlike most species of plants, parasitic plants often contain highly divergent plastomes [13–28]. Non-photosynthetic plants that obtain all their nutrients from the host plant are called holoparasites, while hemiparasites are photosynthetic plants that obtain some nutrients from the host and the rest from photosynthesis. Hemiparasites can be further broken down into obligate parasites that require a host plant during at least part of their life cycle or facultative parasites that are capable of completing their entire life cycle without a host plant, but can parasitize if a host is available. Parasitic plant plastomes can have lower selective pressure due to less reliance on photosynthesis. Not surprisingly, holoparasites have the potential to lose all of their photosynthetic genes, or perhaps the entire chloroplast genome [24], while hemiparasites are expected to maintain most or all of their photosynthetic genes. Lack of or less selection for retained photosynthesis and other vital chloroplast functions can allow higher nucleotide substitution rates and rearrangements in the plastome. These changes can include functional gene loss through pseudogenization or physical loss through gene deletion, resulting in a smaller plastome size [14, 16, 17, 19–23, 26]. In some cases, an entire copy of one of the inverted repeats has been lost [19, 20, 29]. Since the repeats play a stabilizing role in plastome structure, this leads to further destabilization and rearrangements in the plastome [30].

The Orobanchaceae are the largest family of parasitic plants. They contain a single non-parasitic genus, *Lindenbergia* [31], in addition to holoparasites and both facultative and obligate hemiparasites [32]. In this study, we looked at two species of facultative hemiparasites, *Aureolaria virginica* and *Buchnera americana*, and three species of obligate hemiparasites all in the genus *Striga*: *S. hermonthica*, *S. forbessii*, and *S. aspera*. The plastome of the autotroph *Lindenbergia philippensis* had already been fully sequenced and annotated by Wicke et al. 2013, so it provided a valuable outgroup to compare to our parasitic species. Despite the fact that all of our studied species rely on

photosynthesis, we discovered numerous cases of gene loss and other forms of rearrangement in their plastomes.

Methods

Sample collection, DNA extraction and DNA sequencing

DNA preparations from leaf tissue of *Striga* species were the same as those previously described [33]. *B. americana* and *A. virginica* leaf tissues were collected from live specimens in the Plant Biology Teaching collection at the University of Georgia. High molecular weight DNAs were isolated from these samples by a modified cetyltrimethylammonium bromide procedure [34] and were used for Illumina library preparation with the NEB-Next® Ultra™ II DNA Library Prep Kit as per manufacturer's instructions. Illumina MiSeq PE250 runs were performed at the Georgia Genomics Facility.

Plastome assembly

Plastid-homologous sequences were selected from the total cellular DNA reads of *Striga hermonthica* and then assembled with a de novo assembly pipeline, both as previously described [35], to create ten contigs. To fill the ten gaps in this assembly, primers were designed for the end of each contig and PCR was set up for each primer pair. Each PCR reaction consisted of 5–20 ng template DNA, 10 µL Q5 Reaction Buffer, 10 µL betaine, 2.5 µL DMSO, 1 µL 10 mM dNTPs, 2.5 µL forward primer, 2.5 µL reverse primer, .5 µL Q5 High-Fidelity DNA Polymerase, and 20 µL water. Touchdown PCR was performed on each reaction on an MJ Research PTC-200 Peltier Thermal Cycler. The program used was 98 °C for 30s; 10 cycles consisting of denaturation at 98 °C for 20 s, annealing at 68 °C - 1 °C each cycle for 15 s, elongation at 72 °C for 1 min; 25 cycles consisting of denaturation at 98 °C for 15 s, 58 °C for 15 s, elongation at 72 °C for 1 min; and a final elongation step at 72 °C for 10 min. PCR products were sent to Macrogen USA (Rockville, MD) for Sanger sequencing. Sequences were manually assembled to the original contig ends in order to fill gaps.

The assembled *S. hermonthica* plastome was aligned to the *Lindenbergia philippensis* plastome using Mauve 2.3.1 [36]. All junctions to predicted rearrangements in the *S. hermonthica* plastome were PCR amplified and Sanger sequenced (Macrogen USA) to confirm that the assembly was correct.

Plastome reads for each of our target species were extracted from raw total DNA reads by identifying reads homologous to our *S. hermonthica* assembly and used in de novo assemblies with Geneious 8.1.6 [37]. Hence, each de novo assembly was performed independently for each species.

The repeat region of each assembly was identified by the region of the plastome having approximately two times the read coverage compared to the rest of the

plastome. In the case of *S. hermonthica*, we ran PCR on all the junctions between the repeats and the LSC and SSC. Sequencing of these PCR products confirmed the assembly of these junctions.

Plastome annotation and alignment

The genes from the annotated *L. philippensis* plastome were downloaded from NCBI along with the plastome genes from all plant species from Chloroplast Genome DB [38]. We ran BLAST on all genes against each assembled plastome. All genes identified in the plastome were then mapped to the plastome using Geneious. Reads were also mapped to each gene individually to confirm that the sequence assembly was correct and that there were no genes missing in our assemblies that had reads map to them. Individual genes were aligned across all five parasitic species as well as *L. philippensis* using Clustal W [39] to identify mutations, insertions, and deletions. Genes that contained one or more frameshift mutations or premature stop codons were considered potential pseudogenes. The raw reads for the most 5' predicted stop codons or frameshift mutations that could have created a pseudogene were visually inspected to evaluate the accuracy of the read call. Read depth varied from 34X to 548X across these sites, with 18 cases giving 100% agreement with the mutational call, and the others showing > 95% agreement (data not shown).

The plastomes were aligned using Mauve 2.3.1. The phylogenetic tree was generated with sequence data from all intact genes shared by all five species. These genes were concatenated into a single sequence, which were aligned using Clustal W in Mega 7.02 [40]. *L. philippensis* was included as an outgroup. Bayesian phylogenetic analyses were done on MrBayes 3.2.1 [41] using Generalized Time Reversible + Γ + I model of evolution. We ran an initial 2.0×10^7 trees with sampling every 10,000 generations. The first 5.0×10^6 trees were not included in the final analysis. The remaining trees were used to generate a consensus tree. The same data were used to generate a phylogenetic tree using MEGA [40], and this generated nearly identical results, so only the MrBayes tree is presented.

Results

Plastome assembly

Illumina sequence analysis of total DNA from leaves provided sufficient plastome-homologous sequence data for full genome assembly of all five studied species. All of the initial contiguous sequences were confirmed in their order, and gaps filled, by PCR and Sanger sequencing of the PCR product. Hence, complete assemblies were obtained, and any predicted rearrangements by comparisons of these species genomes were also confirmed by PCR and Sanger sequencing.

Our assemblies indicated that the plastome of *A. virginica* is 153,547 bp, close in size to the 155,103 bp plastome of *L. philippensis*. The lengths of the LSC, SSC, and IR (84,317 bp, 17,168 bp, and 26,031 bp, respectively) in *A. virginica* were all similar in size to those in *L. philippensis* (85,584 bp, 17,885 bp, and 25,812 bp, respectively). The plastomes from the other four species (*S. hermonthica*, *S. aspera*, *S. forbessii* and *B. americana*) showed major differences in both total plastome size and the size of each region.

Repeat expansion

The three *Striga* species and *B. americana* all exhibited larger plastome sizes, ranging from 166,596 bp in *B. americana* to 190,233 bp in *S. forbessii*. The larger plastome sizes correlate with the larger IR lengths, ranging from 43,864 bp in *B. americana* to 63,240 bp in *S. forbessii*. Conversely, the LSC and SSC regions in all four species showed a reduction in size. SSC lengths ranged from 3377 bp in *Buchnera* to 11,191 bp in *S. forbessii*, while the LSC lengths ranged from 51,628 bp in *S. hermonthica* to 75,491 bp in *B. americana*. All plastome component sizes are shown in Table 1, and are compared to those for some previously studied parasitic plants.

In *B. americana*, repeat expansion has occurred in both the LSC and SSC regions. Approximately 8 kb of LSC sequence flanking the repeat in *L. philippensis* is now included in the repeat region in *B. americana*. This region contains the genes *trnH*, *psbA*, *matK*, *trnK*, *rps16*, *trnQ*, and *psbK*. The expansion also occurred into both flanking regions of the SSC. On one end of the SSC, *ndhF* has been lost in *B. americana* while the adjacent 6 kb region is now part of the repeat region. *ndhE*, *psaC*, *ndhD*, *ccsA*, *trnL*, and *rpl32* are all contained within this region. On the other end of the SSC, an approximately 5 kb region containing *rpl32* and *ycf1* has been shifted into the repeat in *B. americana*.

The three *Striga* species show an almost identical pattern of IR expansion, with the majority of expansion occurring in the LSC region. The two LSC-flanking regions as well as an internal region are now part of the repeat in all three *Striga* species. An approximately 12.5 kb region flanking the IR in *L. philippensis* is part of the IR in *Striga*. This region contains *trnH*, *psbA*, *matK*, *trnK*, *rps16*, *trnQ*, *psbK*, *psbI*, *trnS*, *trnG*, *trnR*, and *atpA*. On the other end of the LSC, the IR region has expanded into approximately 18.8 kb of adjacent sequence. This region contains 15 genes: *clpP*, *psbB*, *psbT*, *psbN*, *psbH*, *rpoA*, *rps11*, *rpl36*, *infA*, *rps8*, *rpl14*, *rpl16*, *rps3*, *rpl22*, and *rps19*. The internal region is 5.5 kb and contains *ycf4*, *cemA*, *petA*, *psbJ*, *psbL*, *psbF*, and *psbE*. In *L. philippensis*, there are 61 kb between this region and the IR on one side and 18 kb between this region and the IR on the other side. Approximately 4.5 kb of the

Table 1 Plastome composition in *Lindenbergia philippensis* and eight hemiparasites of the Orobanchaceae

	Plastome Length (bp)	LSC Length (bp)	SSC Length (bp)	IR Length (bp)	Expansion of IR (bp)	IR from LSC (bp)	IR from SSC (bp)	IR Genes from LSC	IR Genes from SSC
<i>L. philippensis</i>	155,103	85,606	17,885	25,800	NA	NA	NA	NA	NA
<i>S. americana</i>	160,910	84,756	6517	34,818	9018	0	6334	0	2
<i>P. cheilanthifolia</i>	155,159	85,256	17,971	25,966	0	0	0	0	0
<i>A. virginica</i>	153,547	84,317	17,168	26,031	0	0	0	0	0
<i>T. versicolor</i>	152,448	83,650	17,520	25,639	0	0	0	0	0
<i>B. americana</i>	166,596	75,491	3377	43,864	18,744	8050	10,694	7	8
<i>S. forbessii</i>	190,233	52,563	11,191	63,240	37,218	32,681	4536	34	1
<i>S. hermonthica</i>	186,418	51,628	9884	62,453	36,799	32,299	4500	34	1
<i>S. aspera</i>	185,932	51,706	10,504	61,861	36,201	31,986	4215	34	1

The table shows the size of the IR, how much the IR has expanded compared to *L. philippensis*, the amount of the expanded repeat sequence that came from the LSC and from the SSC, the number of originally single-copy genes now found in the IR, and the number of genes originally found in the LSC and SSC now contained in the IR. The *Aureolaria*, *Buchnera*, and *Striga* results are from our analysis while the other data were previously published for *Lindenbergia philippensis*, *Schwalbea americana* [19], *Pedicularis cheilanthifolia* [28], *Triphysaria versicolor* [20]

SSC region adjacent to the repeat is now part of the repeat region in all three *Striga* species. This region contains the *ycf1* gene. All details of repeat expansion are shown in Table 1.

Overall gene content

No tRNA or rRNA genes have been lost in any of the five species. All of the plastomes encode 30 different tRNAs and 4 different rRNAs. Compared to the out-group *Lindenbergia*, *Aureolaria* has not lost any protein-coding genes and contains a total of 79 single copy genes and six more that are contained within the repeat region and therefore present in two copies. *Buchnera* has a total of 77 single copy genes, plus 17 in the repeat regions, with 9 genes being possible pseudogenes. *S. forbessii* contains 79 single copy genes, plus 35 contained within the repeats, including 14 potential pseudogenes. *S. hermonthica* has 76 single copy protein-coding genes, 35 in the repeats and 10 potential pseudogenes. *S. aspera* contains 77 single copy genes, plus 35 in the repeat regions, with 12 potential pseudogenes (Table 2).

Gene losses

In *L. philippensis* and *A. virginica*, *ndhA* is a 2.2 kb gene containing two exons, 539 bp and 553 bp, separated by a 1.1 kb intron. *S. hermonthica* is missing the second exon of *ndhA*, while *S. forbessii* is missing the first exon. *B. americana* and *S. aspera* have both exons, but have a predicted frameshift mutations and premature stop codons in the first exon. All *ndh* genes are present with the structure of functional genes in *A. virginica*. *B. americana* and the three *Striga* species all have one or more stop codons or frameshifts in *ndhB*, starting at amino acid 446 in *B. americana* and 448 in *S. forbessii*. *S. hermonthica* and *S. aspera* both have a frameshift at amino acid 24, suggesting a shared event

in their common ancestor, followed by multiple frameshifts and stop codons. *ndhC* is present in *B. americana* but contains a stop codon at amino acid 82. *S. forbessii* contains *ndhC* with a frameshift starting at amino acid 23 followed by multiple stop codons. *ndhC* is missing from both *S. hermonthica* and *S. aspera*. All four species contain multiple frameshifts and stop codons throughout *ndhD* starting at amino acid 36 in *B. americana*, 25 in *S. forbessii*, 4 in *S. hermonthica* and 40 in *S. aspera*. *S. forbessii* contains a stop codon in *ndhE* at amino acid 101. *B. americana* has a stop codon at amino acid 11. *S. hermonthica* has a stop codon at amino acid 40. *B. americana*, *S. hermonthica*, and *S. aspera* all contain 5' frameshift mutations, at amino acid 34 in *B. americana* and amino acid 32 in the *Striga* species, but all are independent events. All three species have additional stop codons following the beginning of the frameshift. *ndhF* is missing from *B. americana* and contains a frameshift at amino acid 20 in *S. forbessii* followed by multiple stop codons. *S. hermonthica* and *S. aspera* both contain a stop codon at the third amino acid followed by additional stop codons and frameshifts. *ndhG* contains a frameshift and multiple stop codons starting at amino acid 63 in *B. americana* and amino acid 3 in *S. forbessii*. *ndhG* is missing from *S. hermonthica* and *S. aspera*. *ndhH* is missing from *B. americana*, *S. hermonthica*, and *S. aspera*. *ndhH* is present in *S. forbessii* but contains frameshifts and multiple stop codons starting at amino acid 11. *ndhI* contains a stop codon in *B. americana* at amino acid 90 and in *S. forbessii* at amino acid 17, followed by a frameshift and additional stop codons. *S. hermonthica* contains a frameshift starting at amino acid 47 and *S. aspera* contains a frameshift starting at amino acid 42, both followed by multiple stop codons. *ndhJ* contains stop codons starting at amino acid 140 in *B. americana*.

Table 2 Plastome gene content in *Lindenbergia philippensis* and eight hemiparasites of the Orobanchaceae

	Total Protein-encoding Genes (Duplicated)	Potential Pseudogenes	tRNA	rRNA
<i>L. philippensis</i>	91 (12)	0	30	4
<i>S. americana</i>	86 (12)	5	30	4
<i>P. cheilanthifolia</i>	87 (8)	0	30	4
<i>A. virginica</i>	91 (12)	0	30	4
<i>T. versicolor</i>	91 (12)	0	30	4
<i>B. americana</i>	111 (17)	9	30	4
<i>S. forbesii</i>	149 (70)	14	30	4
<i>S. hermonthica</i>	146 (70)	10	30	4
<i>S. aspera</i>	147 (70)	12	30	4

The table shows the total number of genes that are predicted to encode proteins, the number of unique coding genes, the number of genes that are potentially pseudogenes, and the number of types of tRNA and rRNA genes. The most leftward column shows the total number of predicted genes, including pseudogenes, while the number in parentheses is the number of these genes that are duplicated because of their presence in the IR. Information sources for this Table are as described for Table 1

In *S. forbesii*, *ndhJ* contains a frameshift at amino acid 71 followed by multiple stop codons. *ndhJ* in *S. hermonthica* contains stop codons starting at amino acid 102. *ndhJ* contains a frameshift in *S. aspera* starting at amino acid 44 followed by stop codons. *ndhK* contains a stop codon at amino acid 120 in *B. americana* and amino acid 14 in *S. forbesii*. Both species contain additional stop codons and *S. forbesii* contains a frameshift. *ndhK* in both *S. hermonthica* and *S. aspera* contain a frameshift at amino acid 4 followed by stop codons. Pertinent frameshifts and stop codons are shown in Fig. 1. Table 3 shows mutations shared by two or more species at the identical amino acid that resulted in frameshifts or stop codons.

accD is present in all five species. In *A. virginica*, it appears functional with very few differences in comparison to *L. philippensis*. In the other four species, there are a numerous predicted mutations including multiple frameshifts. However, the 3' end of this gene is considerably more conserved than the 5' end of the gene. It's possible that *accD* has a conserved protein domain encoded by the 3' end of the gene. *ycf1* and *ycf2* also seem to have multiple frameshift mutations in all species, except *ycf2* in *A. virginica*.

Plastome alignments and phylogenetic analysis

All six plastomes were aligned with Mauve 2.3.1 and are shown in Fig. 2a. *A. virginica* is largely identical to *L.*

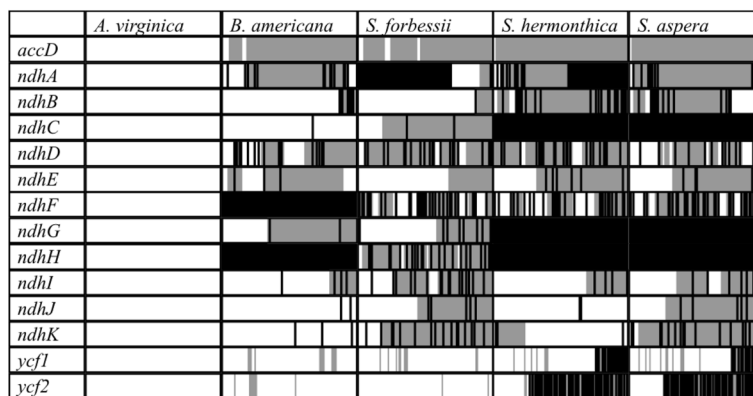


Fig. 1 Stop codons and frameshifts in potential pseudogenes. Genes that are missing or predicted as non-functional in at least one of the five species. Solid white bars indicate the gene is present and appears functional in that species, while bars with gray shading or black lines may be pseudogenes. The structure of the protein-encoding portion of the gene is indicated within each box, with the 5' end being at the far left and the 3' end at the far right. The size of each box is 100% of the protein-encoding region, so they are not corrected for the different sizes of the coding regions of each gene. Gray shading indicates portions of the genes with a shifted frame, while black lines indicate stop codons. Solid black bars are genes or gene segments that are completely missing from that species. For instance, the figure shows that the 5' end of *ndhA* is missing in *S. forbesii* and the 3' end is missing in *S. hermonthica*, but the multiple stop codons and frameshifts in the 5' end in *S. hermonthica* should guarantee that it is a non-functional gene

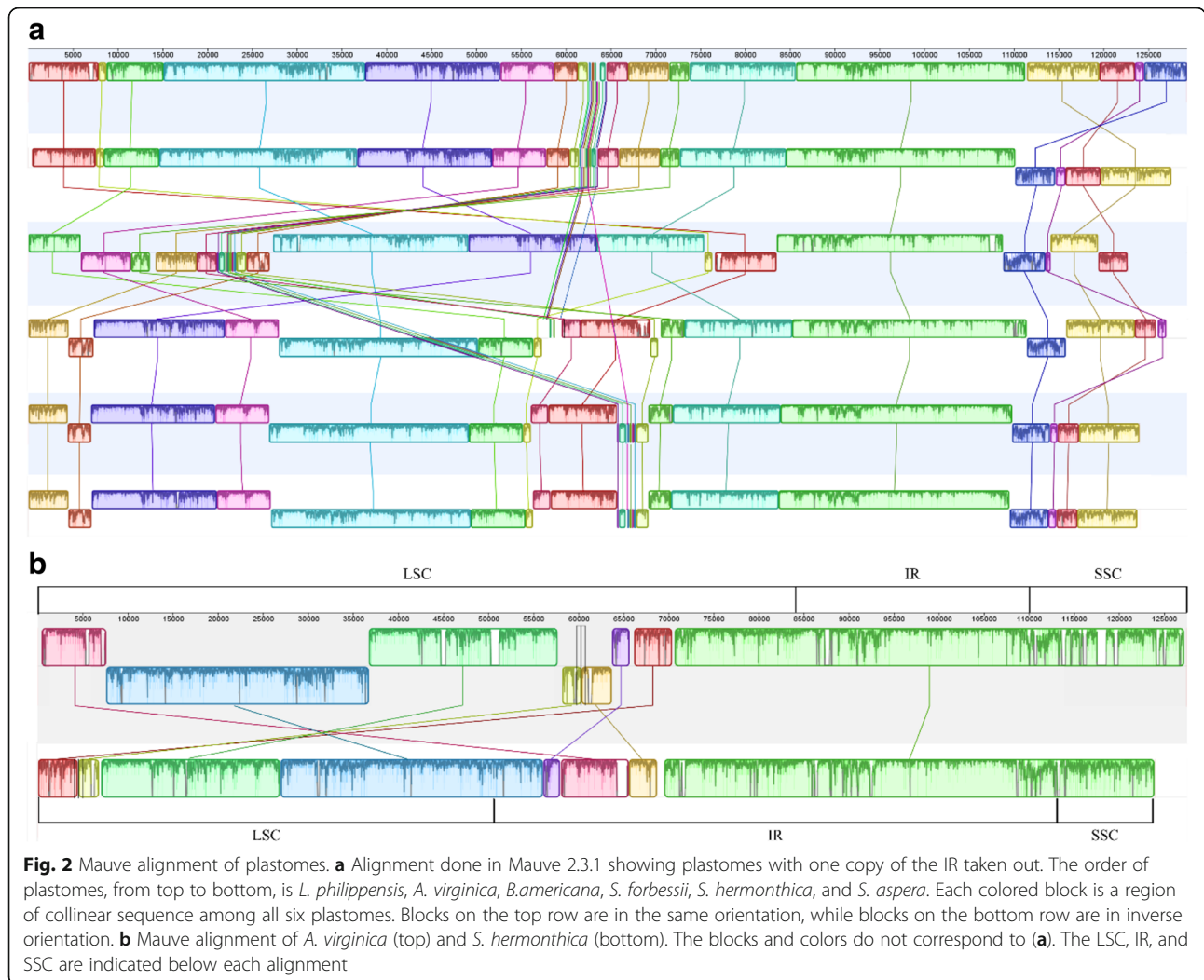
Table 3 List of shared frameshift mutations and stop codons

Gene	Species	Amino Acid	SC or FS
<i>ndhA</i>	<i>B. americana</i>	19	SC
	<i>S. hermonthica</i>	20	SC
	<i>S. aspera</i>	20	SC
<i>ndhB</i>	<i>S. hermonthica</i>	24	FS
	<i>S. aspera</i>	24	FS
	<i>S. hermonthica</i>	118	SC + FS
<i>ndhD</i>	<i>S. aspera</i>	118	SC + FS
	<i>S. hermonthica</i>	94	FS
	<i>S. aspera</i>	94	FS
	<i>S. hermonthica</i>	116	FS
	<i>S. aspera</i>	116	FS
	<i>S. hermonthica</i>	152	FS
	<i>S. aspera</i>	152	FS
	<i>S. forbessii</i>	269	FS
	<i>S. hermonthica</i>	269	FS
	<i>S. aspera</i>	269	FS
	<i>S. hermonthica</i>	319	FS
	<i>S. aspera</i>	317	FS
	<i>S. hermonthica</i>	358	FS
	<i>S. aspera</i>	356	FS
	<i>S. hermonthica</i>	372	FS
	<i>S. aspera</i>	370	FS
	<i>S. hermonthica</i>	391	FS
	<i>S. aspera</i>	389	FS
	<i>S. hermonthica</i>	425	FS
	<i>S. aspera</i>	424	FS
<i>ndhF</i>	<i>S. hermonthica</i>	3	SC
	<i>S. aspera</i>	3	SC
	<i>S. forbessii</i>	20	SC
	<i>S. hermonthica</i>	20	SC
	<i>S. aspera</i>	20	SC
	<i>S. hermonthica</i>	81	FS
	<i>S. aspera</i>	91	FS
	<i>S. forbessii</i>	617	FS
<i>ndhI</i>	<i>S. aspera</i>	593	FS
	<i>B. americana</i>	90	SC
	<i>S. forbessii</i>	89	SC
	<i>S. hermonthica</i>	90	SC
	<i>S. aspera</i>	90	SC
	<i>S. hermonthica</i>	112	FS
<i>ndhK</i>	<i>S. aspera</i>	112	FS
	<i>S. hermonthica</i>	3	FS
<i>accD</i>	<i>S. aspera</i>	3	FS
	<i>B. americana</i>	57	FS

Table 3 List of shared frameshift mutations and stop codons (Continued)

Gene	Species	Amino Acid	SC or FS
<i>ycf1</i>	<i>S. forbessii</i>	12	FS
	<i>S. hermonthica</i>	12	FS
	<i>S. aspera</i>	59	FS
	<i>B. americana</i>	92	FS
	<i>S. forbessii</i>	47	FS
	<i>S. hermonthica</i>	47	FS
	<i>S. aspera</i>	129	FS
	<i>B. americana</i>	251	FS
	<i>S. forbessii</i>	256	FS
	<i>S. hermonthica</i>	258	FS
	<i>S. aspera</i>	254	FS
	<i>B. americana</i>	360	FS
	<i>S. forbessii</i>	386	FS
	<i>S. hermonthica</i>	392	FS
	<i>S. aspera</i>	391	FS
	<i>S. forbessii</i>	467	FS
	<i>S. hermonthica</i>	491	FS
<i>ycf2</i>	<i>S. aspera</i>	475	FS
	<i>B. americana</i>	674	FS
	<i>S. forbessii</i>	648	FS
	<i>S. hermonthica</i>	661	FS
	<i>S. aspera</i>	660	FS
	<i>S. hermonthica</i>	1288	FS
	<i>S. aspera</i>	1287	FS
	<i>S. hermonthica</i>	1452	SC
	<i>S. aspera</i>	1451	SC
	<i>S. hermonthica</i>	1462	FS
	<i>S. aspera</i>	1461	FS
	<i>S. hermonthica</i>	482	FS
	<i>S. aspera</i>	489	FS
	<i>B. americana</i>	848	FS
	<i>S. forbessii</i>	1024	FS
<i>S. hermonthica</i>	987	FS	
<i>S. aspera</i>	994	FS	

List of frameshift mutations (FS) and stop codons (SC) that are shared by two or more species and the amino acid position in each species. FS and SC in the same box are the result of the identical nucleotide mutation in each species. Stop codons created by frameshift mutations are not included. *S. forbessii* is missing the first exon of *ndhA*, which contains the shared stop codon at the 20th amino acid in the other three species. *ndhE* does not have any shared mutations, but *B. americana*, *S. hermonthica*, and *S. aspera* have independent insertions at the same aligned amino acid position (35 in *B. americana* and 33 in the two *Striga* species)



philippensis, except that the SSC is in an inverted orientation. All other plastomes contained multiple rearrangements. Within the *Striga* genus, *S. hermonthica* and *S. aspera* have no rearrangements between them, while *S. forbessii* has a single rearrangement within the repeat region. At a higher resolution, Fig. 2b shows the example of the Mauve alignment of just *S. hermonthica* and *A. virginica*.

The timing of gene loss events

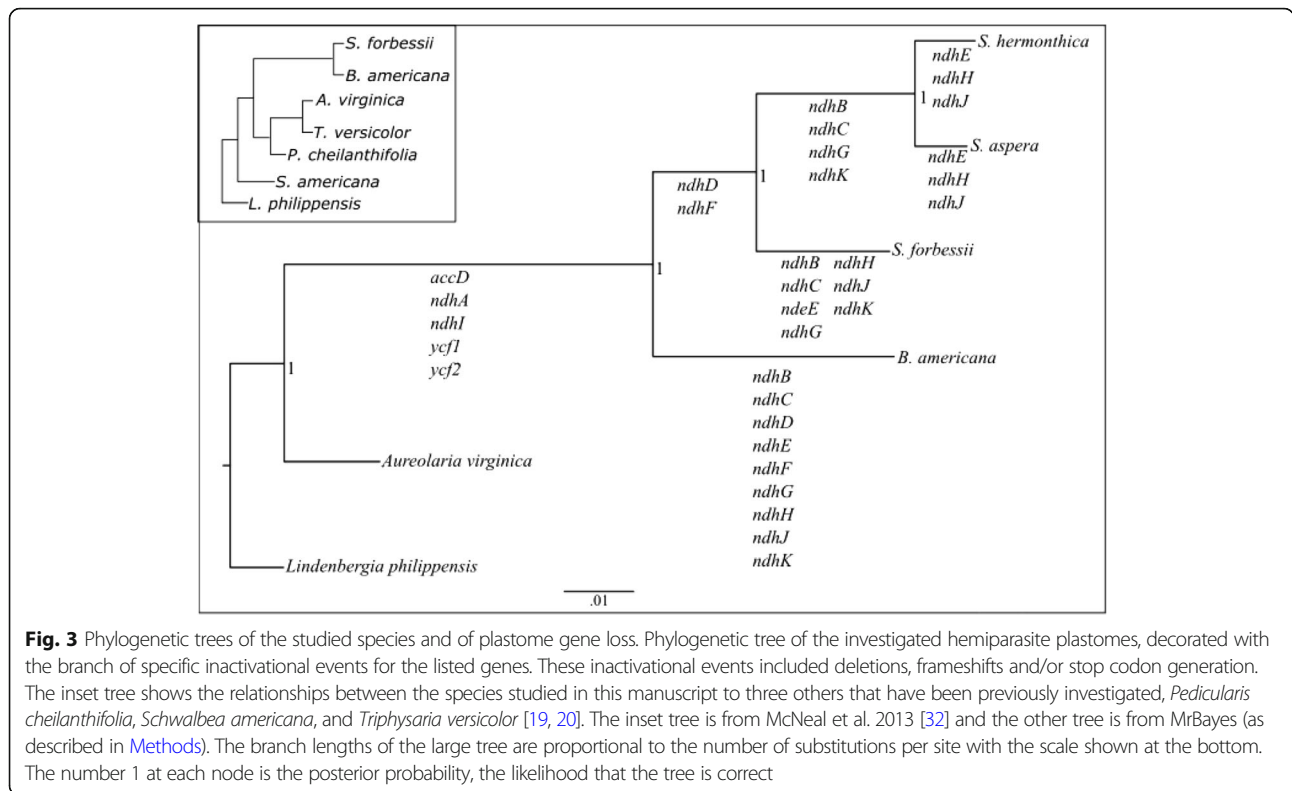
In many cases, a specific nucleotide location of a sequence change that is predicted to inactivate a gene was shared by more than one species. Hence, it is most likely that such events occurred in a common ancestor of the species that share the event. Because these deletions, frameshifts and/or stop codon creations are predicted to inactivate the gene, then it is appropriate to plot the lineage and relative time of inactivation of each gene onto a phylogenetic tree of the compared species. The phylogenetic tree, with *L. philippensis* as an outgroup, is

shown in Fig. 3, with predicted shared gene inactivations indicated as gene names along specific branches.

Discussion

Plastome conservation and change

Plastomes perform common and vital functions in virtually all vascular plants, and thus are very powerful tools for phylogenetic analysis. The parasitic plants that are non-photosynthetic lose strong selection for the many plastome genes involved in photosynthesis, but are expected to retain coding capacity for genes that are involved in other essential chloroplast roles. In previous studies of non-photosynthetic plants [13, 14, 18–21], most or all photosynthetic genes are seen to be pseudogenized or lost. This process appears to be caused by random mutation, and is thus thought to be a gradual process that will only be completed at some distant time after the parasitic/non-photosynthetic lifestyle is established. Our observations, akin to those observed by Wicke et al. [20] in a broader but less detailed survey of



parasitic plants, indicate that some chloroplast genes are lost even in fully photosynthetic parasites, suggesting that chloroplast roles other than generation of photosynthate may now be provided by the host. Hence, extensive gene loss from the plastome does not link exclusively with a holoparasitic lifestyle. If hemiparasitism is a routine step toward the evolution of holoparasitism, then it may be that the plastome gene loss process is as much a cause of holoparasitism as it is an outcome.

Parasitism has evolved independently at least 12 to 13 times within angiosperms [42]. Parasitism within Orobanchaceae is believed to have evolved once and is estimated to have occurred 32–64 mya [43–45]. It is interesting that these different plastomes have come to somewhat different final gene contents and that many sequences are still present despite their loss of apparent function. This suggests that the gene loss process is very slow, especially compared to nuclear DNA loss [46], and raises the possibility that the different adaptations of these different species has led to different plastome gene loss outcomes. For instance, most *ndh* genes were lost early in the lineage leading to the genus *Striga*, but *ndhE*, *ndhH* and *ndhJ* have only been lost in terminal branches, suggesting a possible transitional role for these genes even when the other *ndh* genes were no longer functional. Similarly, *Pedicularis cheilanthifolia* has apparently lost only one plastome gene to pseudogenization, *ndhF*, a gene that has been independently lost in

many lineages [28]. All these results confirm the observation that, in hemiparasites, the *ndh* pathway is unnecessary or possibly even selected against. Further analysis of the nature of plastome genome change should provide additional insights into which genetic (and, thus, physiological) functions are lost, in what lineages and at what evolutionary rates and times.

Plastome size

Most plant plastomes range in size from 120 to 160 kb. *L. philippensis* falls within this range. The *A. virginica* plastome is slightly smaller than that of *L. philippensis*, but still falls within the typical range. The LSC, SSC, and IR of *A. virginica* are all similar in size to those of *L. philippensis* and fall within the general size range for angiosperms. *B. americana* and the three *Striga* species, however, all show interesting size difference in both whole plastome length and the lengths of their LSCs, SSCs, and IRs. All have a larger plastome than *L. philippensis* and above the range generally seen in plants. This was unexpected as most parasitic species studied so far have smaller plastome sizes [14, 16, 17, 19–23]. The increase in size can be explained by an increase in the length of the inverted repeats. All five species have an increased inverted repeat size, although *Aureolaria*'s is close to *Lindenbergia* (26,031 bp compared to 25,812 bp) and falls within the typical range of angiosperm inverted repeat size. The other four species have considerably larger repeats above the 20 to 30 kb size

generally seen in angiosperms, ranging from 43,864 bp in *Buchnera* to 63,240 bp in *S. forbesii*. Small changes of a few hundred bp in inverted repeat size are common in land plants [47–49], but larger changes as seen here are rare [47–52]. The largest known repeat belongs to *Pelargonium hotorum*, at 75,741 bp [53]. Slight IR expansion has been seen in the parasitic plastomes of *Schwalbea americana* and *Cistanche phelypaea*. However, these expansions were much smaller in size, encompassing 2 genes and 3 genes [19], respectively, compared to the expansions we saw in *Striga* and *B. americana*.

The repeat size increase that we observed is due to an expansion into the single copy regions. All five parasitic species have smaller SSC and LSC regions compared to *L. philippensis*. The *Striga* expansions differ dramatically from the expansion in *B. americana*. In *B. americana*, most of the repeat expansion has been into the SSC region. In the three *Striga* species, there has been some repeat expansion into the SSC region, but most of the expansion has been into the LSC region. There are two similarities in the expansion patterns between the *Striga* species and *Buchnera*. The first is the 8 kb in the LSC region adjacent to the IR (spanning from *trnH* to *psbK*). This region is contained within the repeat of all four species, although the *Striga* IRs extend a further 4.5 kb region into this region. It is possible that the initial 8 kb expansion occurred before the species diverged and was followed by an additional 4.5 kb expansion in the *Striga* lineage. The second similarity is the expansion into the SSC that contains the *ycf1* gene. The *B. americana* IR includes an additional 0.5 kb containing *rpl32*, suggesting that this was an additional expansion after the lineages diverged. Other than these two similarities, the rest of the repeat expansions are unique to either *B. americana* or the *Striga* species and most likely occurred after the two lineages diverged. These several independent repeat expansions indicate either selection for this phenomenon or a mechanistic anomaly in these lineages that makes this outcome unusually likely, compared to other studied plants. Functionally, these increases in the duplicated regions do not appear to have created any new genes. Because of the process of gene conversion across these recombining repeats [3], it is not expected that any new gene functions would evolve through this IR expansion.

S. aspera and *S. hermonthica* show no rearrangements between them or differences between the repeat expansions. There is one major structural difference between the plastomes of these two and that of *S. forbesii*. Within the repeat, a 2.5 kb region containing *cemA* and *petA* is located in a different part of the region in an inverse orientation. One possibility is that repeat expansion occurred independently in the *S. forbesii* lineage and the shared *S. hermonthica/aspera* lineage after the two lineages diverged. The other possibility is that the

repeat expansion occurred before the common ancestor of these three species diverged, followed by lineage-specific rearrangement(s) after divergence.

Missing genes

The NAD(P)H-dehydrogenase complex is involved in one of the multiple pathways that recycles electrons around photosystem I [54, 55, 60]. The NAD(P)H complex functions under stressful conditions and may be essential for photosynthesis under conditions of highly variable light intensities [56, 57]. Eleven subunits are encoded by the plastome: *ndhA, B, C, D, E, F, G, H, I, J,* and *K* [58]. These genes have been lost several times in land plants and are commonly missing in parasitic plants [13, 16, 17, 19, 20, 23, 59–63].

Ycf1 was the first plastid-encoded protein identified whose presence was shown to be essential for the survival of green plants. *Ycf1* protein is an essential component of the translocon that is responsible for chloroplast protein transport in green plants [64]. However, no *ycf1* homolog was detected among Poales, except *Typha latifolia* [65]. In our five studied species, *ycf1* was observed to be highly divergent, with predicted frameshift mutations. The gene *ycf2* still has an unknown function [66] and has a functional structure in *A. virginica*, but contains multiple frameshift mutations in *B. americana* and the three *Striga* species. Both *ycf1* and *ycf2* have high substitution rates in most of the land plant, including non-parasites, and may have become pseudogenes in some lineages [67, 68]. The 5' ends of both genes tend to be relatively conserved, while the remaining portions of the genes are more divergent. We observed a similar pattern with the most conserved regions of both genes being in the 5' end. Complete losses of both *ycf1* and *ycf2* have occurred in some monocots [69].

AccD encodes the beta subunit of the multimeric acetyl-CoA carboxylase, which mediates the conversion of acetyl-CoA to malonyl-CoA during fatty acid synthesis. The *accD* gene has been lost several times in angiosperms [70], where its function is taken over by nuclear copies [71]. The 3' region of the *accD* gene in all five of our studied species is considerably more conserved than the 5' end. Higher divergence or truncation of the 5' end of the *accD* gene also has been observed in several species [19, 20]. We do not have a nuclear genome sequence for any of the species investigated in this manuscript, but we predict that a nuclear *accD* homologue is now functional in the *B. americana* and *Striga* lineages. We also predict that the 3' end of the *accD* gene may have a separate, plastid-specific, function.

Plastome alignments

A. virginica has no major genic rearrangements compared to *L. philippensis* except an inverted SSC region.

The SSC region is frequently flipped in plastomes, with opposite SSC orientations often present in a 50:50 ratio in plant cells [1]. Comparisons between *A. virginica* and *B. americana* are interesting as they are both facultative hemiparasites. Because they are both capable of living without a host, thereby depending heavily on photosynthesis, it is expected that they would have a relatively small number of rearrangements compared to autotrophic plant plastomes. This is the case with *A. virginica*. However, *B. americana* has multiple rearrangements compared to *A. virginica* and *L. philippensis*. The three *Striga* species also have several rearrangements compared to *A. virginica* and *L. philippensis*, as well as compared to *B. americana*. There are no rearrangements between *S. hermonthica* and *S. aspera*, and a single rearrangement between these two and *S. forbesii*. These patterns all fall within an appropriate phylogenetic order that predicts their time and lineage of origin.

The phylogenetic tree agrees with existing phylogenetic analysis as far as the relationship between *L. philippensis*, *A. virginica*, *B. americana*, and the *Striga* genus [72], as well as the relationship of the three *Striga* species to each other with *S. hermonthica* and *S. aspera* being more closely related to each other than to *S. forbesii* [33].

Plastome instability and parasitic plant function

Although all five of the plant species investigated in this study are obligate photosynthetic organisms, plastome instability was found to be extensive in four species. These results are in agreement with previous observations for other parasitic plants [13–21], and has also been seen more rarely in non-parasites [29, 45, 53, 65, 73–79]. Hence, plastome instability is a source of genomic variation that provides the raw material for natural selection. In any green plant, this selection is expected to lead to overall conservation of gene content, with exceptions for any chloroplast function that can be provided by the environment directly. For parasitic plants, the host plant is a primary environmental contributor, potentially for photosynthate and many of the numerous small molecules that are generated by pathways localized to the chloroplast. The *ndh* pathway is lost in many lineages, suggesting that its role in photosynthesis is often dispensable. Given the rapid and near-complete loss of this pathway independently in so many lineages, it suggests that *ndh* may actually be selected against in some plant lineages. Future studies that return this pathway to a species from which it has been naturally lost would provide an excellent opportunity to investigate the lineages and conditions under which the *ndh* gene array might benefit or debilitate a plant.

In this same vein, all of the gene changes observed in this study could be investigated for their role in plant fitness. Expression levels of the genes that differ in their copy number across these taxa, for instance due to their

inclusion or lack of inclusion in the IRs, could suggest a role for the generation and retention of such rearrangements across many plant lineages. Natural variation for plastome structure should be tested across more Orobanchaceae species, both to acquire better ideas of the lineages and rates of such rearrangements, but also to provide the raw material for comparisons of the functional outcomes of these rearrangements. With the ability to transform chloroplasts now available in many plant species [80–83], investigations of natural plastome variation and its role in organismal function could be directly compared to experiments that test these gene effects in engineered transgenics.

Conclusions

We sequenced and assembled the chloroplast genomes from five species of hemiparasitic plants from the Orobanchaceae. We compared the assemblies to the available plastome assembly from *Lindenbergia philippensis*, an autotroph from the Orobanchaceae. *Aureolaria virginica* was almost identical to *L. philippensis* in terms of plastome structure, gene content, and gene order. *B. americana*, *S. forbesii*, *S. hermonthica*, and *S. aspera* all showed a high number of rearrangements and both physical and functional gene loss. The most interesting result was the increase in the plastome size in these four species, since most parasites have plastomes with decreased sizes. The increase was due to the repeat region expanding into the single copy regions, although what effect this has on the plastome is not yet known. *B. americana*, *S. hermonthica*, and *S. aspera* all had missing *ndh* genes, while these three species and additionally *S. forbesii* had multiple *ndh* genes that are potentially non-functional due to predicted frameshift mutations and stop codons. No other genes were missing from any of the species, but *accD*, *ycf1*, and *ycf2* all had high levels of divergence with frameshifts in *B. americana*, *S. forbesii*, *S. hermonthica*, and *S. aspera*, as well as stop codons in *ycf1* and *ycf2* in *S. hermonthica* and *S. aspera*.

Acknowledgements

We thank Prof. J. Palmer (Indiana University) and Prof. S. Wicke (University of Muenster) for their helpful comments.

Funding

This research was supported by USDA/NIFA grant award #2012–67013–19473 and the endowment associated with the Giles Professorship in the Department of Genetics at the University of Georgia.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the GenBank repository, <https://www.ncbi.nlm.nih.gov/genbank/> with accession numbers MF780870, MF780871, MF780872, MF780873, and MF780874.

Authors' contributions

DCF worked on the plastome assembly, annotation, analysis and manuscript preparation and revision. SRC conducted experiments, assisted in plastome assembly, annotation and manuscript revision. JNV assisted in plastome

assembly. CGC assisted with plastome sequencing. JLB conceived the project, and helped with experiments, data analyses, manuscript preparation and revision. All authors contributed to revisions of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 August 2017 Accepted: 30 January 2018

Published online: 06 February 2018

References

- Palmer JD. Comparative organization of chloroplast genomes. *Annu Rev Genet.* 1985;19:325–54.
- Wicke S, Schneeweiss GM, de Pamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011;76:273–97.
- Palmer JD. Molecular evolution: a single birth of all plastids? *Nature.* 2000; 405:32–3.
- McFadden GI, van Dooren GG. Evolution: red algal genome affirms a common origin of all plastids. *Curr Biol.* 2004;R514:516.
- Keeling PJ, et al. *Philos Trans R Soc Lond Ser B Biol Sci.* 2010;365:729–48.
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A.* 2002;99:12246–51.
- Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry RJ, editor. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants.* Wallingford, UK: CABI Publishing; 2005. p. 45–68.
- Kolodner R, Tewari KK. Inverted repeats in chloroplast DNA from higher plants. *P Natl Acad Sci USA.* 1979;76:41–5.
- Maréchal A, Parent J, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2009;106:14693–8.
- Palmer JD. Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK, editors. *Cell culture and somatic genetics of plant, vol 7A. Molecular biology of plastids.* San Diego: Academic Press; 1991. p. 5–53.
- Sugiura M. The chloroplast genome. *Plant Mol Biol.* 1992;19:149–68.
- Bock R. Structure, function, and inheritance of plastid genomes. In: Bock R, editor. *Cell and molecular biology of plastids.* Berlin Heidelberg, 63: Springer. p. 29.
- dePamphilis CW, Palmer JD. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature.* 1990; 348:337–9.
- Wolfe KH, Morden CW, Palmer JD. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci U S A.* 1992;89:10648–52.
- Nickrent DL, Ouyang R, Joel D, dePamphilis CW. Do nonasterid holoparasitic flowering plants have plastid genomes? *Plant Mol Biol.* 1997;34:717–29.
- Funk H, Berg S, Krupinska K, Maier U, Krause K. Complete DNA sequences of plastid genomes of two parasitic flowering plant species *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* 2007;7:45.
- McNeal JR, Kuehl J, Boore J, dePamphilis C. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* 2007;7:57.
- Krause K. From chloroplasts to “cryptic” plasts: evolution of plastid genomes in parasitic plants. *Curr Genet.* 2008;54:111–21.
- Wicke S, Müller KF, dePamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Schneeweiss GM. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell.* 2013;25:3711–25.
- Wicke S, Müller KF, de Pamphilis CW, Quandt D, Bellot S, Schneeweiss GM. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *P Natl A Sci.* 2016;113:9045–50.
- Samigullin TH, Logacheva MD, Penin AA, Vallejo CM. Complete plastid genome of the recent holoparasites *Lathraea squamaria* reveals earliest stages of plastome reduction in Orobanchaceae. *PLoS One.* 2016; <https://doi.org/10.1371/journal.pone.0150718>.
- Delavault PM, Russo NM, Luson NA, Thaluouan P. Organization of the reduced plastid genome of *Lathraea clandestina* an achlorophyllous parasitic plant. *Physiol Plant.* 1996;96:674–82.
- Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Plock SA, Wolf PG, dePamphilis CW, Boore JL, Goffinet B. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Mol Biol Evol.* 2008;25:393–401.
- Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, Flowers JM, Pelsler P, Barcelona J, Inovejas SA, et al. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol Biol Evol.* 2014;31:793–803.
- Naumann J, Der JP, Wafula EK, Jones SS, Wagner ST, Honaas LA, Ralph PE, Bolin JF, Maass E, Neinhuis C, et al. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biol Evol.* 2016;8:345–63.
- Bellot S, Renner SS. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biol Evol.* 2015;8:189–201.
- Roquet C, Coissac É, Cruaud C, Boleda M, Boyer F, Alberti A, Gielly L, Taberlet P, Thuiller W, Van Es J, Laverigne S. Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasites *Cytinus hypocistis* (Cytinaceae). *Ann Bot.* 2016;118:885–96.
- Zhang R, Wang J, Han K, Ren T, Zeng S, Biffin E, Liu Z. Complete chloroplast genome sequence of *Pedicularis cheilanthifolia*, an alpine plant in China. *Conserv Genet Resour.* 2017;doi:<https://doi.org/10.1007/s12686-017-0740-2>
- Downie SR, Palmer JD. Restriction site mapping of the chloroplast DNA inverted repeat: a molecular phylogeny of the Asteridae. *Ann Mo Bot Gard.* 1992;79:266–83.
- Perry AS, Wolfe KH. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol.* 2002;55:501–8.
- Olmstead RG, dePamphilis CW, Wolfe AD, Young ND, Elisons WJ, Reeves PA. Disintegration of the Scrophulariaceae. *Am J Bot.* 2001;88:348–61.
- McNeal JR, Bennet JR, Wolfe AD, Mathews S. Phylogeny and origins of holoparasitism in Orobanchaceae. *Am J Bot.* 2013;100:971–83.
- Estep MC, Gowda BS, Huang K, Timko MP, Bennetzen JL. Genomic characterization for parasitic weeds of the genus by sample sequence analysis. *Plant Genome.* 2012;5:30–41.
- Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 1987;19:11–5.
- Vaughn JN, Chaluvadi SR, Tushar RL, Bennetzen JL. Whole plastome sequences from five ginger species facilitate marker development and define limits to barcode methodology. *PLoS One.* 2014; <https://doi.org/10.1371/journal.pone.0108581>.
- Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genet Res.* 2004;14:394–1403.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28:1647–9.
- Cui L, Veeraraghavan N, Richter A, Wall K, Jansen RK, Leebens-Mack J, Makalowska I, dePamphilis CW. ChloroplastDB: the chloroplast genome database. *Nucleic Acids Res.* 2006;34:D692–6.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0. *Mol Biol Evol.* 2016;33:1870–4.
- Ronquist F, Huelsenback JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19:1572–4.
- Westwood JH, Yoder J, Timko MP, dePamphilis CW. The evolution of parasitism in plants. *Trends Plant Sci.* 2010;15:227–35.
- Bremer K, Fijis EM, Bremer B. Molecular phylogenetic dating of asteroid flowering plants shows early cretaceous diversification. *Syst Biol.* 2004; 53:496–505.

44. Wolfe AD, Randle CP, Liu L, Steiner KE. Phylogeny and biogeography of *Orobanchaceae*. *Folia Geobot.* 2005;40:115–34.
45. Naumann J, Salomo K, Der JP, Wafula EK, Bolin JF, Maass E, Frenzke L, Samain MS, Neinhuis C, dePamphilis CW, Wanke S. Single-copy nuclear genes place haustorial Hydnoraceae within Piperales and reveal a cretaceous origin of multiple parasitic angiosperm lineages. *PLoS One.* 2013; <https://doi.org/10.1371/journal.pone.0079204>.
46. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A.* 1987;84:9054–8.
47. Goulding SE, Wolfe KH, Olmstead RG, Morden CW. Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet.* 1996;252:195–206.
48. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol.* 2008;8:36.
49. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst Bot.* 2015;40:336–51.
50. Knox EB, Palmer JD. The chloroplast genome arrangement of *Lobelia thuliniana* (*Lobeliaceae*): expansion of the inverted repeat in an ancestor of the *Campanulales*. *Plant Syst Evol.* 1999;214:49–64.
51. Dugas DV, Hernandez D, Koenen EJM, Schwarz E, Straub S, Hughes CE, Jansen RK, Nageswara-Rao M, Staats M, Trujillo JT, Hajrah NH, Alharbi NS, Al-Malki AL, Sabir JSM, Bailey CD. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci Rep.* 2015;5:16958.
52. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* 2016;209:1747–56.
53. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol.* 2006;23:2175–90.
54. Casano LM, Zapata JM, Marti M, Sabater B. Chlororespiration and poisoning of cyclic electron transport. *J Biol Chem.* 2000;275:942–8.
55. Nixon PJ. Chlororespiration. *Philos Trans R Soc B Biol Sci.* 2000;355:1541–7.
56. Martín M, Funk HT, Serrot PH, Poltnigg P, Sabater B. Functional characterization of the thylakoid Ndh complex phosphorylation by site-directed mutations in the *ndhF* gene. *BBA-Bioenergetics.* 2009;1787:920–8.
57. Martín M, Sabater B. Plastid *ndh* genes in plant evolution. *Plant Physiol Biochem.* 2010;48:636–45.
58. Friedrich T, Steinmüller K, Weiss H. The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett.* 1995;367:107–11.
59. McCoy SR, Kuehl JV, Boore JL, Raubeson LA. The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Evol Biol.* 2008;8:130.
60. Wu C, Lai Y, Lin C, Wang Y, Chaw S, et al. *Mol Phylogenet Evol.* 2009;52:115–24.
61. Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, Chaw SM. The chloroplast genome of *Phalaenopsis aphrodite* (*Orchidaceae*): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol.* 2006;23:279–91.
62. Wu F, Chan M, Liao D, Hsu C, Lee Y, Daniell H, Duvall M, Lin C. Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of molecular markers for identification and breeding in *Oncidiinae*. *BMC Plant Biol.* 2010;10:68.
63. Barrett CF, Freudenstein JV, Li J, Mayfield-Jones DR, Perez L, Pires JC, Santos C. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids and implications for heterotrophic angiosperms. *Mol Biol Evol.* 2014;31:3095–112.
64. de Vries J, Sousa FL, Bölter B, Soll J, Gould SB. YCF1: a green TIC? *Plant Cell.* 2015;27:1827–33.
65. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK. Implications of the plastid genome sequence of typha (*typhaceae*, *poales*) for understanding genome evolution in *poaceae*. *J Mol Evol.* 2010;70:146–66.
66. Logacheva MD, Schelkunov MI, Shtratnikova VY, Matveeva MV, Penin AA. Comparative analysis of plastid genomes of non-photosynthetic *Ericaceae* and their photosynthetic relatives. *Sci Report.* 2016; <https://doi.org/10.1038/srep30042>.
67. Oliver M, Murdock A, Mishler BD, Kuehl J, Boore J, Mandoli D, Everett K, Wolf PG, Duffy A, Karol KG. Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. *BMC Genomics.* 2010;11:143.
68. Wolf PG, Der J, Duffy A, Davidson J, Grusz A, Pryer KM. The evolution of chloroplast genes and genomes in ferns. *Plant Mol Biol.* 2010; <https://doi.org/10.1007/s11103-010-9706-4>.
69. Downie SR, Katz-Downie DS, Wolfe KH, Calie PJ, Palmer JD. Structure and evolution of the largest chloroplast gene (ORF2280): internal plasticity and multiple gene loss during angiosperm evolution. *Curr Genet.* 1994;25:367–3781.
70. Jansen RK, Cai Z, Raubeson LA, Daniell H, de Pamphilis CW, Leebens-Mack JH, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, JR MN, Kuehl JV, Boore JL. 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.* 2007;104:19369–74.
71. Nakkaew A, Chotigeat W, Eksomtramage T, Phongdara A. Cloning and expression of a plastid-encoded subunit betacarboxyltransferase gene (*accD*) and a nuclear-encoded subunit biotin carboxylase of acetyl-CoA carboxylase from oil palm (*Elaeis guineensis* Jacq.). *Plant Sci.* 2008;175:497–504.
72. Bennet JR, Mathews S. Phylogeny of the parasitic plant family *Orobanchaceae* inferred from phytochrome A. *Am J Bot.* 2006;93:1039–51.
73. Palmer JD. Chloroplast DNA evolution and biosystematics uses of chloroplast DNA variation. *Am Nat.* 1987;130:S6–S29.
74. Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, Wakasugi T, Sugiura M. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Mol Gen Genet.* 1992;232:206–14.
75. Plunkett GM, Downie SR. Expansion and contraction of the chloroplast inverted repeat in *Abiaceae* subfamily *Apioideae*. *Syst Bot.* 2000;64:8–67.
76. Daniell H, Lee S, Grevich J, Sasaki C, Quesada-Vargas T, Guda C, Tomkins J, Jansen RK. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other *Solanaceae* genomes. *Theor Appl Genet.* 2006;112:1503–18.
77. Wolfe PG, Roper JM, Duffy AM. The evolution of chloroplast genome structure in ferns. *Genome.* 2010;53:731–8.
78. Grewe F, Viehoveer P, Weisshaar B, Knoop V. A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* 2009;37:5093–104.
79. Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, Schwarzbach AE, Mower JP. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cypressophyte evolution. *Genome Biol Evol.* 2014;6:580–90.
80. Svab Z, Hajdukiewicz P, Maliga P. Stable transformation of plastids in higher plants. *Proc Natl Acad Sci U S A.* 1990;87:8526–30.
81. Sikdar SR, Serino G, Chaudhuri S, Maliga P. Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Rep.* 1998;18:20–4.
82. Sidorov VA, Kasten D, Pang SZ, Hajdukiewicz PTJ, Staub JM, Nehra NS. Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *Plant J.* 1999;19:209–16.
83. Wani SH, Haider N, Kumar H, Singh NB. Plant plastid engineering. *Curr Genomics.* 2010;11:500–12.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

