Reconstructing plastome evolution across the phylogenetic backbone of the parasitic plant genus Cuscuta (Convolvulaceae)

ARJAN BANERJEE1,2,* and SAŠA STEFANOVIĆ1

1Department of Biology, University of Toronto Mississauga, Mississauga, ON L5L 1C6, Canada
2Ecology and Evolutionary Biology, University of Toronto, Toronto, ON M5S 2Z9, Canada

Received 2 December 2019; revised 10 June 2020; accepted for publication 16 June 2020

Parasitic plants have evolved to have reduced or completely lost ability to conduct photosynthesis and are usually characterized by sweeping morphological, physiological and genomic changes. The plastid genome (or plastome) is highly conserved in autotrophic plants and houses many key photosynthesis genes. This molecule is thus a useful system for documenting the genomic effects of a loss of autotrophy. Cuscuta (dodders) represents one of 12 independent transitions to a parasitic lifestyle in angiosperms. This near-cosmopolitan genus contains > 200 obligate parasitic species circumscribed in four subgenera: Grammica, Pachystigma, Cuscuta and Monogynella. With respect to photosynthesis, Cuscuta is a heterogeneous group, containing both hemi- and holoparasitic members that are, respectively, partially or entirely reliant on parasitism to meet their carbon budget. Plastomes in this genus have been reported to show a substantial degree of diversification in terms of length and gene composition. Considered together with well-understood phylogenetic relationships, this genus presents an opportunity for fine-scale comparisons among closely related species of heterotrophic plants. This research documents changes in sequence composition and structure that occurred as these plants evolved along the trophic spectrum by using multiple whole-plastome assemblies from each of the four subgenera. By ‘triangulating’ the positions of genomic changes, we construct a step-by-step model of plastome evolution across the phylogenetic backbone of Cuscuta and highlight the remarkable retention of most photosynthetic genes in these parasitic plants.


INTRODUCTION

Although green plants are often collectively seen as primary producers, the process of photosynthesis has been lost repeatedly, most often in flowering plants. True parasitism, which requires the forging of direct vascular connections between heterotrophic plants and their hosts, has evolved independently at least 12 times across the phylogenetic tree of angiosperms (Nickrent, 2020). Their evolution from photosynthetic ancestors is often accompanied by sweeping morphological, physiological and genomic changes brought about by a decreased (or absent) reliance on the photosynthetic apparatus (Kuijt, 1969; Colwell, 1994; Barkman et al., 2007; Westwood et al., 2010). This ‘parasitic reduction syndrome’ defines a set of convergent evolutionary changes shared among different groups of heterotrophic plants and provides a fertile system for studying the genomic effects of a life history change away from primary production.

The genomic changes are expected to be particularly conspicuous in plastid genomes (or plastomes) given the importance of plastids as the site of photosynthesis in the cell. Plastomes in autotrophic plants are subjected to a high degree of purifying selection and are also conserved in terms of gene order and composition, typically presenting as circular, four-part molecules with large and small single-copy regions separated by two inverted repeats (Downie & Palmer, 1992). They are usually 140–160 kbp in length (e.g. Nicotiana tabacum L.: 155 939 bp, Arabidopsis thaliana (L.) Heynh.: 154 478 bp) and encode mainly housekeeping genes and genes controlling key portions of the photosynthetic apparatus (Shinozaki et al., 1986; Sato et al., 1999). As a consequence of their reduced reliance on photosynthesis, plastomes in heterotrophic plants
have been observed to accrue evolutionary changes that lead to their reduction in size and gene content relative to their green counterparts (e.g. Westwood et al., 2010; Wicke et al., 2011; Barrett & Davis, 2012; Molina et al., 2014; Graham et al., 2017). These loss-of-function changes are expected to be irreversible. Thus, over time, as these plants adapt to life without photosynthesis, their plastomes are expected to slide down an evolutionary 'slippery slope' and accumulate more reduction reflecting gene loss (Stefanović & Olmstead, 2005). As ever-greater numbers of heterotrophic plastomes are being reported (we now have at least one plastid genome reported for 11 of the 12 lineages that have independently evolved to become parasitic), the overall pattern of sequence loss linked to heterotrophy is emerging across the angiosperms: ndh genes (primarily responsible for mitigating the effects of photo-oxidative stress) are usually the first family of genes to be lost, followed by groups of photosynthesis-related genes (psa, psb, pet etc.), whereas a core group of non-bioenergetic genes (accD, ycf1, ycf2, trnE and clpP) seems to be retained in the majority of even severely diminished genomes (Wicke et al., 2011; Barrett & Davis, 2012; Graham et al., 2017). These initial observations, however, still need broader support from plastomes in heterotrophic lineages that are underrepresented in current research.

*Cuscuta* L. (dodders, Convolvulaceae) is one of the most extensively studied lineages of parasitic plants because of its nearly cosmopolitan distribution, species richness, agricultural and ecological importance and the fact that it represents one of the 12 independent origins of parasitism among angiosperms. This lineage comprises c. 200 stem parasites characterized by their scale-like leaves, twining, slender stems and an absence of roots (Yuncker, 1932). Their pale, often orange-brown colour is attributed to reduced or absent accumulation of chlorophylls (van der Kooij et al., 2000) and some species in this genus are obligate parasites, unable to survive without contact with their hosts (Hibberd et al., 1998; Heide-Jorgensen, 2008). Holoparasitic dodders, apparently having lost the ability to photosynthesize, have been identified from two sections in *Cuscuta* subgenus *Grammica* (Lour.) Peter based on the loss of most/all of their photosynthetic plastid genes (Braukmann et al., 2013; Banerjee & Stefanović, 2019). Other dodders are, however, capable of limited and localized photosynthesis (Dawson et al., 1994; Hibberd et al., 1998) and have been referred to as 'cryptically photosynthetic' (McNeal et al., 2007a). This makes *Cuscuta* one of only two lineages of haustorial parasites to be trophically heterogeneous and contain both ‘hemi’- and holoparasitic species. The other such group is Orobanchaceae, a family which has more than ten times the number of species (Westwood et al., 2010). *Cuscuta* is thus an excellent system for studying the genomic effects of a transition to parasitism at a relatively low (species) phylogenetic level. This genus also has a well-resolved and strongly supported phylogeny, both at the backbone level, subdivided into four subgenera (*Monogynella* (Des Moul.) Peter, *Cuscuta*, *Pachystigma* (Engelm.) Baker & C.H.Wright and *Grammica*; their phylogenetic relationships are shown in Fig. 1) and nearer the tips, with species circumscribed into 19 sections (Garcia et al., 2014; Costea et al., 2015). Although well studied with molecular phylogenetic data, there are too few *Cuscuta* plastid genomes to make substantial inferences about their evolution across the genus. Whole plastomes have been reported for only three of the 19 sections in the genus (Funk et al., 2007; McNeal et al., 2007b; Banerjee & Stefanović, 2019) and two of the four subgenera. The two published genomes from subgenus *Monogynella* are c. 120 kbp long with c. 103 genes (Funk et al., 2007; McNeal et al., 2007b), whereas the ten published genomes from subgenus *Grammica* are 60–85 kbp long containing 61–92 genes (Funk et al., 2007; McNeal et al., 2007b; Banerjee & Stefanović, 2019). The state of plastomes is unknown for subgenera *Cuscuta* and *Pachystigma*, as is the tempo and mode of plastome evolution across the backbone of the phylogenetic tree of the genus. Two notable studies have attempted to explore plastid evolution across *Cuscuta* without whole-plastome assemblies. McNeal et al. (2007a) used targeted amplification and sequencing of specific genes and created a useful, phylogenetically informative model of plastid evolution in *Cuscuta*, but their results were incomplete, and they were unable to triangulate the precise locations of several of the changes they detected (i.e. loss of psal, rpl32, ycf15 etc.) based on the data they had at the time. Braukmann et al. (2013) conducted a Southern hybridization survey of a phylogenetically representative sample of 112 species testing for the presence and absence of 48 protein-coding plastid genes using probes designed mostly from tobacco (*Nicotiana tabacum*). They were able to comment on general trends of gene loss in *Cuscuta*, but their survey was incomplete in terms of gene coverage and contained no sequence data, and was thus unable to provide information at the most basic, sequence level, including none for elements like ribosomal and transfer RNA genes, introns, promoters etc. Both studies provided an initial assessment of plastome evolution across *Cuscuta*, but lacked extensive sequence data and/or comprehensive taxon sampling, leaving multiple unanswered questions.

To bridge this gap, we assembled plastid genomes spanning the crown nodes of subgenera *Pachystigma* and *Cuscuta* with the following objectives in mind: (1)
Figure 1. Annotated plastid genomes from six species sampled from across the subgenera of Cuscuta, four of which are newly assembled in this research (highlighted in bold): two species each from subgena Pachystigma (C. nitida and C. africana) and Cuscuta (C. pedicellata and C. approximata). Previously published C. exaltata (from subgenus Monogynella;
to elucidate the state of the plastid genome in these clades using whole genome sequence data; (2) to create a step-by-step reconstruction of plastome evolution across the backbone of the phylogenetic tree for *Cuscuta* using new and previously available data; (3) to compare the patterns of sequence changes observed along the backbone of this genus with plastome evolution patterns deduced for other lineages of heterotrophic plants and (4) to gain insight into on the tempo of evolution in *Cuscuta*.

**MATERIAL AND METHODS**

**TAXON SAMPLING, DNA EXTRACTION AND SEQUENCING**

Based on the broad-scale phylogenetic analysis of *Cuscuta* (Garcia et al., 2014), representatives from subgenera *Cuscuta* and *Pachystigma* sampled for this study were strategically chosen from each group in such a way as to span their respective crown nodes and hence to capture maximum diversity from each clade. Subgenus *Cuscuta* is represented by *C. approximata* Bab. (section *Cuscuta*) and *C. pedicellata* Ledeb. (section *Epistigma* Engelm.) and subgenus *Pachystigma* by *C. nitida* E.Mey. and *C. africana* Wildl. Voucher information for each of these is listed in Table 1 with NCBI accession numbers. Total genomic DNA was isolated from herbarium (*C. pedicellata*) or silica-dried tissue (remaining species) using the modified cetyltrimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987) and checked for quantity and quality using a Nano Drop 1000 Spectrophotometer (Thermo Fisher Scientific). Extractions were sent for sequencing facility.

**PLASTOME ASSEMBLY, ANNOTATION AND COMPUTATIONAL METHODS**

Reads were trimmed using Sickle v.1.33 (Joshi & Fass, 2011) with minimum post-trim read lengths set at 99 bp and the threshold for quality set at a minimum PHRED score of 27 at each site. Several separate assemblies were conducted *de novo* using different subsamples of the HiSeq reads (using between 3 948 776 and 8 005 563 reads for each assembly) on Geneious R10 (Biomatters, Auckland, New Zealand) with the ‘produce scaffolds’ and ‘don’t merge variants’ boxes unchecked. The assembled contigs were aligned and joined (where required), with remaining gaps closed manually on Geneious.

Initial annotations for each of the four new plastid genomes were conducted on Geneious with *C. exaltata* Engelm. (McNeal et al., 2007b) and *C. costaricensis* Yunck. (Banerjee & Stefanović, 2019) as references. Annotations were then refined and confirmed manually using nucleotide BLAST (Altschul et al., 1990), BLASTx (Altschul et al., 1990) and tRNAscan-SE 2.0 (Lowe & Chan, 2016) to confirm rRNA gene sequences, establish all open-reading frames and to determine the boundaries of tRNA genes. Putative pseudogenes were identified based on BLASTx alignments. The genomes were aligned using the ‘translation align’ tool on Geneious and using progressiveMauve (Darling et al., 2010) for the identification of sequence and structural differences. The previously published plastomes from *C. reflexa* Roxb. (NC_009766), *C. exaltata* (EU189132), *C. obtusiflora* Kunth (EU189133) and *C. costaricensis* (MK881072) (Funk et al., 2007; McNeal et al., 2007b; Banerjee & Stefanović, 2019) were used for comparisons in the genus, and the closely related *Ipomoea nil* (L.) Roth (NC_031159) (Hoshino et al., 2016), also belonging to Convolvulaceae (the morning-glory family), was used as an outgroup.

Selection analyses were conducted for ten protein coding genes (**accD**, **atpA**, **petA**, **psaA**, **psbA**, **rbcL**, **rpl20**, **rps8**, **ycf1** and **ycf2** chosen as representatives from the major functional groups that are retained in the plastomes of the four new species and of the four species used for comparison. Sequences extracted from each species for these genes were aligned using the ‘translation align’ tool on Geneious. The type of selection acting on each of them was assessed by the ratio of substitution rates (dN/dS) using the Analysis of Phylogenetic Evolution (APE) package on R v3.6.1 (Paradis et al., 2004; Popescu et al., 2012). *Ipomoea nil* from Convolvulaceae was used as the photosynthetic outlier for dN/dS calculation. As a secondary test, a codon-by-codon estimation of selection using a fast, unconstrained Bayesian approximation for inferring selection (FUBAR) (Murrell et al., 2013) was conducted as well on the platform HyPhy (Pond et al., 2005) using five independent MCMC chains to obtain posterior samples of grid point weights inferred from 2 000 000 total steps (discarding 1 000 000 as burn-in).

McNeal et al., 2007b) and *C. costaricensis* (from subgenus *Grammica*; Banerjee & Stefanović, 2019) have been included for comparison. For legibility, only one of the two inverted repeat regions are represented. Pseudogenes are represented by the ψ symbol beside the gene labels. The phylogenetic relationships are based on those inferred in Garcia et al. (2014). The loss of IRα in *C. pedicellata* and *C. approximata* are denoted by the ‘X’ symbol.

PLASTOME EVOLUTION IN CUSCUTA

RESULTS

PLASTOMES IN CUSCUTA SUBGENUS CUSCUTA

Closed plastomes were assembled for each of the two species sampled from subgenus Cuscuta. Cuscuta pedicellata (section Epistigma) was found to have a plastome 97,091 bp long and C. approximata (section Cuscuta) 98,380 bp in length, both with 64 protein-coding genes and four rRNA genes (Table 1). The major reason for the reduction in the size of the plastome in subgenus Cuscuta is not further gene loss but instead the loss of the inverted repeat (IR) component of the molecule (Fig. 2). In both these species, no IR sequence was found in the assemblies.

INVERTED REPEAT SURVEY

To confirm the loss of IR in C. approximata and C. pedicellata (discussed later, Fig. 1), and to survey the extent of potential loss across subgenus Cuscuta, a survey of potential plastome loss was conducted. The plastid genome of Ipomoea nil and Cuscuta subgenus Monogynella are 98,866 and 97,091 base pairs, respectively. To assess the extent of plastome loss across subgenus Cuscuta, a polymerase chain reaction (PCR) based test was conducted. Based on the results of the PCR test, it was determined that the IR component was lost in C. approximata and C. pedicellata, but retained in Cuscuta subgenus Monogynella.

Table 1. Plastid genome size and structure information for the four newly assembled species from Cuscuta subgenus Pachystigma (C. nitida and C. africana) and subgenus Cuscuta (C. pedicellata and C. approximata).

<table>
<thead>
<tr>
<th>Species</th>
<th>PLASTOME SIZE (bp)</th>
<th>GENES (PROTEIN/TRNA/RNA)</th>
<th>GC (%)</th>
<th>SSC (bp %)</th>
<th>IR (bp %)</th>
<th>IR boundaries (bp)</th>
<th>NCBI accession numbers</th>
<th>Voucher</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. pedicellata</td>
<td>97,091</td>
<td>64/27/4</td>
<td>35.4</td>
<td>7,716 (7.95)</td>
<td>N/A</td>
<td>N/A</td>
<td>MN464181</td>
<td>Humbles 10061</td>
</tr>
<tr>
<td>C. approximata</td>
<td>98,380</td>
<td>64/27/4</td>
<td>35.0</td>
<td>7,369 (7.49)</td>
<td>N/A</td>
<td>N/A</td>
<td>MN464180</td>
<td>Stefanović 09-47</td>
</tr>
<tr>
<td>C. africana</td>
<td>105,066</td>
<td>59/28/4</td>
<td>37.5</td>
<td>7,580 (7.21)</td>
<td>18,328</td>
<td>5,782 → 25,909</td>
<td>MN464179</td>
<td>Oliver 11852</td>
</tr>
<tr>
<td>C. nitida</td>
<td>113,762</td>
<td>65/27/4</td>
<td>37.5</td>
<td>8,012 (7.04)</td>
<td>18,135</td>
<td>8,013 → 26,147</td>
<td>MN464178</td>
<td>Verboom 174</td>
</tr>
</tbody>
</table>
not even a residual one as observed in Pinaceae (Tsudzuki et al., 1992) or one reduced in length as in Geranium L. or Monsonia L. (Guisinger et al., 2011). This represents the first instance of IR loss recorded in Cuscuta and in Convolvulaceae.

In terms of gene composition, all ndh genes are absent (with no detectable non-functional remnants) from the plastomes of this subgenus, as are rpl23, rpl16, ycf15 and psaD. In addition, the tRNA genes trnK-UUU, trnV-UAC and trnG-UCC are also missing. The open-reading frame for rpoC2 has been lost with the gene presenting as a putative pseudogene in both species. Structurally, there is a c. 1 kb inversion in the large single-copy region of these plastomes between trnT-UGU and trnF-GAA, as has been previously noted (McNeal et al., 2007a). There is also a structural change in the small single copy region, with rpl32 translocating to in between rps15 and ycf1 and inverted in both species.

**PLASTOMES IN CUSCUTA SUBGENUS PACHYSTIGMA**

Closed circular plastomes were also assembled for each of the two species sampled from subgenus Pachystigma. The plastome of C. nitida is 108 971 bp in length with 65 protein-coding, 27 tRNA and four rRNA genes, and that of C. africana is 100 207 bp in length with 59 protein-coding, 28 tRNA and four rRNA genes.
These genomes are c. 61–67% the length of the plastome in Ipomoea nil and c. 85% the length of those in Cuscuta subgenus Monogynella. The quadripartite structure of the molecule is retained with two inverted repeat sections (c. 18 kb long) separated by small (c. 8 kb) and large single-copy regions. There are no structural changes in these plastomes, aside from gene loss, compared to the autotrophic outgroup. In terms of gene composition, all ndh genes are absent from both species, as are rpl23, rps16, ycf15 and psaD. Both species are also missing trnK-UUU and trnV-UAC, and the introns in rpl2, atpF, trnA-UGC and trnI-GAU. Both introns in ycf3 have also been lost.

In addition to these changes in common, there are also several differences between the plastomes in C. nitida and C. africana. The tRNA gene trnA-UGC was found to be missing in C. nitida. In C. africana, the protein-coding genes psbZ and rpoC1 have been rendered putative pseudogenes owing to frame-shift mutations, and a stop-codon substitution was observed in the second (middle) exon of clpP. All other rpo genes have been lost in C. africana, although they have been maintained in C. nitida.

**Selection on genes in the genus Cuscuta**

The results of the pairwise dN/dS tests conducted for each of the four new plastomes assembled in this research and one plastome each from subgenera Grammica and Monogynella are shown in Fig. 3. The photosynthetic Ipomoea nil was used as the outgroup for these analyses. The only gene that showed a degree of neutral or positive selection was accD which has a substitution ratio of 1.18 in C. pedicellata (the only > 1 value found in all species) and > 0.80 in all species. The dN/dS values for ycf1 and ycf2 are between 0.60 and 0.85 for all species, which the other genes show strong purifying selection in Cuscuta, especially the photosynthetically important atpA, petA, psaA, psbA and rbcL genes (Fig. 3).

**Figure 3.** Bar charts showing the substitution ratio (dN/dS) values for the ten genes sampled from across the genus Cuscuta. A phylogenetic tree has been added to illustrate how these species are related. The four species for which plastomes have been newly assembled (C. pedicellata and C. approximata from subgenus Cuscuta; C. nitida and C. africana from subgenus Pachystigma) have been underlined. One species each from the remaining subgenera Monogynella (C. exaltata; McNeal *et al*., 2007b) and Grammica (C. costaricensis; Banerjee & Stefanović, 2019) have been included for comparison. The outgroup used for dN/dS analyses was the photosynthetic Ipomoea nil (Hoshino *et al*., 2016).
The number of codons under positive and purifying selection (with greater than/equal to 90% probability) for each of the ten genes analysed based on alignments constrained by previously published phylogenies for the genus are shown in Table S1. There are no codons under positive selection for seven of those genes; for the remaining three, only 0.08, 0.29 and 0.31% of codons are under positive selection for psbA, petA, and atpA, respectively. Five of the genes (atpA, petA, psaA, psbA and rbcL) appear to be under strong purifying selection with 28.13 to 43.98% of their codons found to be under purifying selection. The remaining five genes (accD, rpl20, rps8, ycf1 and ycf2) appear to be under weaker purifying selection with 4.06–9.70% of their codons found to be under purifying selection.

DISCUSSION

A reconstruction of plastome evolution in genus Cuscuta

The two pairs of newly assembled plastid genomes, one set from subgenus Cuscuta and the other from subgenus Pachystigma (Fig. 1), provide detailed information regarding the state of the plastome in these subgenera for the first time. Because in both cases the species have been chosen to span the crown nodes of these clades, they are likely to capture their overall diversity and the most important features. Along with the previously published plastomes from subgenus Grammica and subgenus Monogynella, we now have a full snapshot that allows us to create a step-by-step reconstruction of plastome evolution along the backbone of the genus (Fig. 4). In turn, this allows us to compare the patterns of change in plastid genomes in Cuscuta to other lineages that represent independent origins of parasitism in the angiosperms. The modified phylogeny shown in Fig. 4 and discussed below is based on the generally accepted assumptions that plastid genomes accrue only irreversible loss-of-function changes and undergo no gains in function.

As the only parasitic lineage in Convolvulaceae, it is not surprising that the branch leading to Cuscuta from its closest fully photosynthetic relatives in Convolvulaceae (Stefanović et al., 2002; Stefanović & Olmstead, 2004) has accumulated a large number of plastome sequence losses. All ndh genes other than ndhB and ndhD have been lost in all Cuscuta spp., and the remaining two ndh genes exist only as putative pseudogenes, and only in subgenus Monogynella (in which the ndhB intron has been lost). trnK-UUU has also been lost on this branch, and ycf15, rpl23 and rps16 have been pseudogenized (with rps16 subsequently having lost its residual intron).

As previously reported by McNeal et al. (2007a), the only changes that can be attributed to the branch leading to subgenus Monogynella are two inversion events: a large 13-kb inversion between trnV-UAC and psbE in the large singly copy region, and a smaller 2-kb inversion between trnL-UAG and ccsA (Fig. 4). At the tips in this subgenus, C. exaltata and C. reflexa show some autapomorphic differences from one...
another. Specifically, five genes that are non-functional in C. exaltata (ndhB, ndhD, rpl23, rps16 and ycf15) have been lost entirely from the smaller plastome in C. reflexa (Funk et al., 2007; McNeal et al., 2007b). We predict that the branch leading to the rest of the genus (and containing the three other subgenera) experienced the loss of the same five genes, suggesting that these sequences were functionally lost at the common ancestor to all Cuscuta (Fig. 4), along with trnV-UAC and the intron in trnG-UCC. In terms of gene content, the two newly obtained plastomes from subgenus Cuscuta, C. approximata and C. pedicellata, are identical to one another. In addition, they share the absence of trnG-UCC, the pseudogenization of rpoC2 and three interesting structural features: (1) a 1-kb inversion between trnT-UGU and trnF-GAA in the large single copy region; (2) the translocation of rpl32 from between trnN-GUU and ccsA to between rps15 and ycf1 (in the opposite reading frame) and (3) the loss of the plastid inverted repeat. The observed translocation of rpl32 may have been the result of multiple structural changes over time rather than one conventional translocation event.

Subgenera Pachystigma and Grammica share the absence of five introns, namely the ones in atpF, trnA-UGC, trnI-GAU and both introns in ycf3. The branch leading to subgenus Pachystigma is the only internal branch in the model reported in Figure 4 to have not accumulated any synapomorphic changes. However, there are several tip-specific (autapomorphic) losses in C. africana and C. nitida. trnA-UGC is missing from C. nitida, and rpoA, rpoB and rpoC2 are absent in C. africana, with rpoC1 and psbZ present as putative pseudogenes in this species. A stop-codon was also observed in the second exon of clpP in C. africana causing the severe truncation of the reading frame and a presumed loss-of-function in this gene. This is particularly noteworthy because clpP has been identified as a core non-bioenergetic plastome gene that is usually retained even in some of the most diminished plastomes (Graham et al., 2017).

Numerous sequence loss events are noted on the branch leading to subgenus Grammica, as predicted by earlier studies (McNeal et al., 2007a; Braukmann et al., 2013). All rpo genes are absent from species of subgenus Grammica, as are matK, rpl32, psaI, trnG-UGC, trnR-ACG, trnI-GAU and the 3’ intron in rps12 (Funk et al., 2007; McNeal et al., 2007b; McNeal et al., 2009; Banerjee & Stefanović, 2019). Cuscuta obtusiflora has also lost trnA-UGC (McNeal et al., 2007b). Additional reductions in gene composition have been reported for plastomes in section Ceratophorae (Yunck.) Costea & Stefanović (Banerjee & Stefanović, 2019) and predicted for those in section Subulatae Costea & Stefanović (Engelm.) (Braukmann et al., 2013), both in subgenus Grammica, but these changes are isolated to those two sections and are not informative for the broader-scale model discussed here.

When inferring the ‘loss’ of coding sequences from the plastid genome, movement of genes from the plastome to the nuclear genome is also possible (Ayliffe & Timmis, 1992; Martin & Herrmann, 1998; Huang et al., 2003; Shahmuradov et al., 2003; Matsuo et al., 2005), although functional transfer is unlikely in heterotrophic plants. In fact, we know that most of the genes that used to be in the genomes of cyanobacteria at the time of endosymbiosis have been moved to the nucleus (Martin & Herrmann, 1998; Martin et al., 1998). Without the nuclear genome sequence for these species, we cannot say whether genes missing from the plastomes we have assembled here are entirely missing from the cell. The role of plastid genes may also be subsumed by the nucleus, as in the case of heterotrophic plants that have lost their rpo genes (which produce plastid-encoded polymerase and function in the expression of plastid genes) and have to rely on nuclear-encoded polymerase for gene expression instead (Krause et al., 2003). A mechanism like this may explain how C. africana can compensate for having a non-functional plastid clpP (Fig. 4) when its protein product is an ATP-dependent proteolytic protease which has been shown to have vital functions in plastid protein synthesis, folding and quality control, features needed for plastome function beyond photosynthesis (Sjogren et al., 2006).

**LOSS OF THE INVERTED REPEAT IN CUSCUTA SUBGENUS CUSCUTA**

No inverted repeat region was observed in the plastomes assembled for C. approximata and C. pedicellata, and a natural joint was found between the psbA end of the large single-copy region and the small single-copy region, usually separated by IRb. A PCR test was conducted to confirm this observation and check other species in subgenus Cuscuta for the presence of IRb (Fig. 2). Fragments of predicted length (c. 550–650 bp) were recovered for all species sampled from subgenus Cuscuta (from all three sections) but for none of the representative species sampled from each of the other three subgenera, indicating that the loss of IRb occurred in the common ancestor of subgenus Cuscuta only (Fig. 2).

Inverted repeat regions have been found to be lost completely in some fully photosynthetic plants, including cupressophytes (Guo et al., 2014), some Fabaceae (Medicago L., Pisum L. and Vicia L.) (Palmer et al., 1987) and in Erodium L’Hér. ex Aiton (Guisinger et al., 2011). In heterotrophs, the loss of an IR is observed much more often, especially in severely reduced plastomes (e.g. in Pilostyles Guill.) (Bellot & Renner, 2015) and sometimes also
in more intact genomes, such as those in *Orobanche* L., *Phelipanche* Pomel or *Cassytha* L. (Wicke et al., 2013; Wu et al., 2017; Petersen et al., 2019). Often a missing IR is associated with increased instances of rearrangements in a plastome (Palmer & Thompson, 1982), and this prediction correlates well with two small rearrangement events that we have reported here in the plastomes from subgenus *Cuscuta* (Fig. 4).

**Selection on plastome genes in Cuscuta**

Beyond examining the presence and absence of genes in plastid genomes of *Cuscuta*, analyses of sequence divergence and evolutionary rate variation suggest that the functional groups of genes that are retained in all the sampled species remain under varying degrees of purifying selection (Fig. 3). It has been reported that the intensity of selection, in a positive or purifying direction, tends to be elevated in heterotrophic plastomes (Barrett et al., 2019) in response to the accelerated evolution these genomes experience. There also tends to be an increase in numbers of individual sites within plastid genes exhibiting variable substitution rates (synonymous and non-synonymous) in heterotrophs generally (Barrett et al., 2019), and this also appears to be the case for the genes sampled in this study in *Cuscuta* (Supporting Information Table S1). Notwithstanding these facts, the five genes tested that perform important photosynthetic functions (*atpA*, *petA*, *psaA*, *psbA* and *rbcL*) exhibit dN/dS values (< 0.3) that indicate strong purifying selection, the housekeeping genes tested (*rpl20* and *rps8*) seem to be under moderate purifying selection (0.35 < dN/dS < 0.58) and the three non-bioenergetic genes examined (*accD*, *ycf1* and *ycf2*) show relatively weak purifying selection (dN/dS > 0.65; Fig. 3).

Fig. 3 shows that *accD* is the only gene in this study for which a dN/dS value > 1.00 was observed (indicative of positive selection), and that was the case only in one species, *C. pedicellata* (subgenus *Cuscuta*). In all the other species, *accD* shows dN/dS values between 0.80 and 0.95, markedly elevated relative to the other genes sampled (Fig. 3). This gene, which codes for a key carboxylase enzyme, responsible for facilitating fatty acid biosynthesis (Neuhaus & Emes, 2010; Wicke et al., 2011), has been previously noted to show weaker purifying selection in *Cuscuta* (Banerjee & Stefanović, 2019) and to have a higher rate of sequence divergence in general compared to other plastid genes (Logacheva et al., 2016). Still, *accD* is retained in even highly reduced plastomes and is considered to be an essential non-bioenergetic gene (Graham et al., 2017). In general, caution is warranted when reporting observations of positive selection (as with *C. pedicellata* in this study) because of the high proportion of false positives in selection analyses (Mallick et al., 2009).

The genes *ycf1* and *ycf2* code for large proteins that have recently been shown to have a role in the transport of other proteins into the plastid (de Vries et al., 2015; Kikuchi et al., 2018). These sequences seem to be under weak purifying selection in *Cuscuta* (Fig. 3), consistent with our earlier findings that *ycf1* and *ycf2* accumulate large amounts of non-synonymous change (Banerjee & Stefanović, 2019). These genes incur relatively high sequence divergence across the heterotrophic plants (Wicke et al., 2011; Barrett et al., 2019) and in photosynthetic plants (Li et al., 2013; Barnard-Kubow et al., 2014). Both *ycf1* and *ycf2* have dN/dS values greater than one in the autotrophic *Campanulastrum americanum* (L.) Small (Barnard-Kubow et al., 2014), and a pairwise comparison for *ycf2* between the photosynthetic *Nicotiana tabacum* and *Olea europaea* L. revealed a dN/dS value between 0.6 and 0.7, substantially elevated in comparison to most other plastid genes in the same study (Li et al., 2013). Along with *accD*, *ycf1* and *ycf2* have even been lost from photosynthetic plastomes in Poaceae (Guisinger et al., 2010).

The housekeeping genes *rpl20* and *rps8* appear to be under moderate purifying selection, with dN/dS values between 0.35 and 0.58 (Fig. 3), similar to dN/dS values for *rpl20* in the photosynthetic Poaceae (Guisinger et al., 2010), but elevated when compared to the same genes in most autotrophic eudicots (Li et al., 2013). However, small ribosomal protein (*rps*) genes have been shown to have elevated nucleotide substitution rates in heterotrophic (Barrett et al., 2019) and photosynthetic (Barnard-Kubow et al., 2014) lineages, potentially associated with increased plastome rearrangement events.

The photosynthetically important *atp*, *pet*, *psa* and *psb* genes exhibit low dN/dS values in autotrophic plants, indicative of strong purifying selection (Guisinger et al., 2010; Li et al., 2013; Barnard-Kubow et al., 2014). The fact that the representatives of these gene families shown in Fig. 3 are also under strong purifying selection in plastomes of *Cuscuta* reinforces the idea that many species across this genus are ‘cryptically photosynthetic’ and capable of limited, localized photosynthesis. The *rbcL* gene codes for the large subunit of the enzyme RuBisCO which plays a crucial role in the fixation of carbon in the first step of photosynthesis and so is an essential gene in autotrophic plastomes. Nonetheless, rates of nucleotide substitution in *rbcL* are highly variable, and this gene has been found to be under positive selection in the majority of land plants (Kapralov & Filatov, 2007), probably reflecting the need for optimization of the performance of RuBisCO in various gaseous and thermal conditions (Kapralov & Filatov, 2007). Still, *rbcL* is maintained under strong purifying selection in *Cuscuta* (Fig. 3) and is retained in the plastomes of other lineages of heterotrophic plants (Wolfe &...
dePamphilis, 1998; Wicke et al., 2011), indicative of some important secondary function in these plants (discussed further below).

The results of the FUBAR codon-by-codon selection analyses (Supporting Information Table S1) conducted on the same genes across the broad phylogenetic tree for Cuscuta subgenera also suggest that all these sequences (including accD) seem to be under purifying selection. The finding that accD has only two codons under positive selection in the FUBAR results (compared to 43 codons under purifying selection) reinforces the argument that the results of the substitution ratio test for this gene must be treated cautiously, especially as the FUBAR approach has been shown to be more robust against false positives than other selection analysis tools (Murrell et al., 2013). None of the ten genes analysed in Supplementary Table S1 has > 1% of codons under positive selection; hence, there seems to be no selective pressure acting toward the removal of these genes from plastomes of Cuscuta. Instead, accD, rpl20, rps8, ycf1 and ycf2 appear to be under weak purifying selection, with between 4.06 and 9.70% of their codons experiencing purifying selective pressure. The other five genes (atpA, petA, psaA, psbA and rbcL) all have 28.13–43.98% of their codons under purifying selection, and thus the selective pressures on them appear to be stronger. These results corroborate the findings from the dN/dS test shown in Figure 3.

PATTERNS OF PLASTOME REDUCTION IN CUSCUTA COMPARED TO OTHER HETEROTROPIC PLANTS

As mentioned before, Cuscuta appears to be particularly promising for studying the genomic effects of a transition to parasitism because it contains trophically transitional lineages. The reconstruction of plastome evolution discussed in this research confirms the early loss of the ndh family of genes and the parallel loss of rpo genes in several species that still appear to be at least partially photosynthetically (i.e. species in subgenus Grammica; Figs 3, 4). These observations have been observed repeatedly in other lineages of heterotrophic plants, although rpo genes are sometimes retained until later on in the transition process (Barrett & Davis, 2012; Graham et al., 2017). Some of the other losses that seem to be idiosyncratic in Figure 4 are actually non-random and have been observed frequently elsewhere, specifically rpl23, rps16 and trnA-UGC (Graham et al., 2017). These genes may be missing, even though they have housekeeping functions, because they are easily moved to the nuclear genome, or because their roles are readily fulfilled by nuclear genes (Cusimano & Wicke, 2016).

A major step in the evolution of parasitic plastomes is the wholesale loss of photosynthesis-related genes (Barrett & Davis, 2012; Graham et al., 2017) and this has not happened yet in the species discussed in this research (Fig. 5). All eight species included in the reconstruction shown in Figure 4 not only retain all their atp and pet genes, but they are still under purifying selection. Hence, we assume that their ATP-synthase and cytochrome b6/f complex subunits encoded by these genes are still expressed and functional. The retention of atp genes in holoparasitic lineages, even beyond the point at which photosynthesis is thought to be entirely lost, has been observed before, specifically in Orobanchaceae, elsewhere in Cuscuta and in several mycotrophic orchids (Kohzuma et al., 2012), but whether they maintain this particular function in heterotrophs is not known (Graham et al., 2017). However, the fact that they remain apparently intact in holoparasitic plastomes suggests an essential non-photosynthetic function.

The majority of psa and psb genes are also maintained, with the notable exceptions of psaI in subgenus Grammica and psbZ in Cuscuta africana. These initial losses in the two gene families may indicate that some photosystem I and II genes are non-essential (psaI and psbZ are particularly small genes, for instance, being c. 100 and 189 bp long, respectively) and may be lost moving forward, although the fact that psaA and psbA are under strong purifying selection in these species (as shown in Fig. 3) suggests the opposite. Another photosynthetically important gene that is retained under strong purifying selection in all four new plastomes is rbcL. This gene, which codes for the large subunit of RuBisCO (Kellogg & Juliano, 1997), has a variable presence in the more intact holoparasitic plastome, being present and putatively functional in some non-photosynthetic Orobanchaceae (Wolfe & dePamphilis, 1998; Wicke et al., 2011) and pseudogenized in others (Wicke et al., 2013). Its retention has been previously explained by a secondary function in facilitating rapid and efficient lipid synthesis (Schwender et al., 2004). Elsewhere in Cuscuta, rbcL has been completely lost in some holoparasitic members of subgenus Grammica (Braukmann et al., 2013; Banerjee & Stefanović, 2019) and has been shown to be non-functional and a potential pseudogene in C. Mexicana Yunck., a partially photosynthetic species in the same subgenus (Banerjee & Stefanović, 2019).

The state of introns in Cuscuta is reported in Table 3. All group IIA introns (except the second intron in clpP, which is considered a group IIA intron even though it is self-splicing) are missing in subgenus Grammica, and this is accompanied by the simultaneous loss of matK, which encodes a maturation responsible for splicing group IIA introns (McNeal et al., 2009). However, all but one of these introns
are also lost in subgenus *Pachystigma* (in which the 3′ intron in *rps12* remains) which still expresses a functional *matK* gene (Table 3). This seems to be an intermediate step between the retention of most group IIA introns (except *trnV-UAC* and *trnK-UUU*, the gene for which is missing in all species in this genus) in subgenus *Cuscuta* and their wholesale loss in subgenus *Grammica* (Fig. 4, Table 3). One group IIA intron (the one in *trnV-UAC*) has also been lost in subgenus *Cuscuta*, although the rest are maintained. Other group II introns show idiosyncratic losses in the genus, although those in *clpP* (intron 1), *petB*, *petD* and *rpl16* are present for all species, along with the solitary group I intron *trnL-UAA* (Table 3).
Plastome Evolution in Cuscuta

Table 2. tRNA presence (+) and absence (-) in the two pairs of newly assembled plastid genomes from Cuscuta subgenus Pachystigma (C. nitida and C. africana) and subgenus Cuscuta (C. pedicellata and C. approximata), with the same information for two previously published plastomes from each of the other two subgenera for comparison: subgenus Grammica (C. costaricensis and C. obtusiflora) and subgenus Monogynella (C. exaltata and C. reflexa). Species for which plastomes were assembled in this research have been underlined.

<table>
<thead>
<tr>
<th></th>
<th>C. exaltata</th>
<th>C. reflexa</th>
<th>C. pedicellata</th>
<th>C. approximata</th>
<th>C. nitida</th>
<th>C. africana</th>
<th>C. costaricensis</th>
<th>C. obtusiflora</th>
</tr>
</thead>
<tbody>
<tr>
<td>trnA-UGC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>trnG-UCC</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>trnI-GAU</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>trnK-UUU</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>trnR-ACG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>trnV-UAC</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: trnC-GCA, trnD-GUC, trnE-UUC, trnP-GAA, trnF-CAU, trnF-GCC, trnH-GUG, trnI-CAU, trnL-CAC, trnL-UAA, trnL-UGA, trnM-CAU, trnN-GU, trnP-UUG, trnQ-UUG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnW-CCA and trnY-FUA are present in all these plastomes of Cuscuta and so have been omitted from this table for brevity.

A model of plastome evolution in the genus Cuscuta, modified from and compared to the models of mycoheterotroph plastome evolution created by Barrett & Davis (2012) and Graham et al. (2017), is presented in Figure S1. Since the bulk of ATP, housekeeping non-bioenergetic genes are maintained in Cuscuta spp., the current state of the most reduced plastid genome in the genus (Banerjee & Stefanović, 2019) is considered to be partway down the ‘slippery slope’ of plastome evolution (Supporting Information Fig. S1).

Tempo and Mode of Evolution in Cuscuta

There are several instances of parallel sequence loss that can be discerned from Figure 4. Most notably, rpo genes are lost (either functionally or actually) in three separate locations: along the branch leading to subgenus Cuscuta; in C. africana; and along the branch leading to subgenus Grammica. In addition, trnG-UCC is absent from all species of subgenus Cuscuta and subgenus Grammica species, and trnA-UGC is missing in C. nitida and C. obtusiflora. These parallel reduction events that cannot be explained as phylogenetically shared (synapomorphic) are significant because they may be indicative of sequence redundancy in these species and may be predictive of future changes that could possibly occur in the genus.

Among the 12 lineages of angiosperms with independent origins of haustorial parasitism, there are only two which contain both hemi- and holoparasitic species. In Orobancheaceae, the evolution of plastid genomes was thought to have occurred in a punctuated fashion, with intense evolutionary change attributed to the branch leading to all holoparasitic species, but few reductions accumulating elsewhere (Young et al., 1999). A greatly expanded recent work has statistically shown that changes in repeat density in this family are consistent with a punctuated mode of evolution (Wicke et al., 2013). In Cuscuta, the tempo of evolution had originally been suspected to be gradual with reductions spread throughout the genus instead of being concentrated on any one branch (Stefanović et al., 2002; Stefanović & Olmstead, 2005). Recently,
however, studies have shown that there are lineages in the phylogeny of *Cuscuta* that accumulate far more changes than others (McNeal et al., 2007a; Braukmann et al., 2013; Banerjee & Stefanović, 2019). The results of the research reported here, and the reconstruction of plastome evolution depicted in Figure 4, show that although plastid evolution in *Cuscuta* is indeed not limited to any one branch, there are clearly branches where there is a concentration of evolutionary activity, notably those leading to the genus as a whole and to subgenus *Grammica*. We also now know that in subgenus *Grammica* there are two lineages (sections *Ceratophorae* and *Subulatae*) that exhibit more sequence losses than others (Braukmann et al., 2013). We may thus conclude that the evolution of plastomes in *Cuscuta* displays elements of both gradual and punctuated evolution, but that the overall tempo of change in this genus is predominantly punctuated.

**ACKNOWLEDGEMENTS**

We thank Mihai Costea and the directors and curators of the herbaria IND and SANBI for providing tissue crucial to this project. We also thank Erika Frangione and Daniel Frederic for help with preliminary assembly work and two anonymous reviewers for their detailed and constructive feedback that helped improve this manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada (grant no. 326439); the Canada Foundation for Innovation (grant no. 12810); and Ontario Research Funds.

**CONFLICT OF INTEREST STATEMENT**

The authors declare no conflicts of interest.

**REFERENCES**


Barnard-Kubow KB, Sloan DB, Galloway LF. 2014. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evolutionary Biology* 14: 268.


Extreme reconfiguration of plastid genomes in the angiosperm evolution. Genome Biology and Evolution 205:162–165.


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

**Table S1.** Analysis of selection on selected plastome genes (representative of genes present in all eight species used in this research) across all four subgenera of Cuscuta conducted using FUBAR (a fast, unconstrained Bayesian approximation for inferring selection; Murrell et al. 2013) on HyPhy (Pond et al. 2005).

**Figure S1.** An irreversible model of plastome gene reduction (top panel; Barrett and Davis, 2012) and a model of plastome gene loss (middle panel) adapted and modified from Graham et al. (2017) showing their ranges of most likely points of loss (thick lines) and possible ranges of loss (dashed lines). The bottom panel is an updated and refined model specific to Cuscuta, based on the most recent data for this genus (this research, and Banerjee & Stefanović, 2019). The current state of the most reduced sequenced plastome in Cuscuta (based on C. bodinghii Urb., C. erosa Yunck. and C. strobilacea Liebm.; Banerjee & Stefanović, 2019) is inferred to be between the loss of rpo genes and the loss of atp genes. ‘X’ denotes the loss of the final non-bioenergetic gene. All lines (solid and dashed) in the bottom panel are approximated made from our current best understanding of plastome sequence data for Cuscuta.