Mitochondrial Phylogenomics of *Cuscuta* (Convolvulaceae) Reveals a Potentially Functional Horizontal Gene Transfer from the Host

Qianshi Lin [1,2,*], Arjan Banerjee [1,2], and Saša Stefanović [1]

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

*Corresponding author: E-mail: qianshi.lin@utoronto.ca.

Accepted: 09 June 2022

Abstract

Horizontal gene transfers (HGTs) from host or other organisms have been reported in mitochondrial genomes of parasitic plants. Genes transferred in this fashion have usually been found nonfunctional. Several examples of HGT from the mitochondrial genome of parasitic *Cuscuta* (Convolvulaceae) to its hosts have been reported, but not vice versa. Here we used 31 protein-coding mitochondrial genes to infer the phylogeny of *Cuscuta*, and compared it with previous nuclear and plastid phylogenetic estimates. We also investigated the presence of HGTs within these lineages. Unlike in plastid genomes, we did not find extensive gene loss in their mitochondrial counterparts. Our results reveal the first example of organellar HGT from host to *Cuscuta*. Mitochondrial *atp1* genes of South African subgenus *Pachystigma* were inferred to be transferred from Lamiales, with high support. Moreover, the horizontally transferred *atp1* gene has functionally replaced the native, vertically transmitted copy, has an intact open reading frame, and is under strong purifying selection, all of which suggests that this xenolog remains functional. The mitochondrial phylogeny of *Cuscuta* is generally consistent with previous plastid and nuclear phylogenies, except for the misplacement of *Pachystigma* when *atp1* is included. This incongruence may be caused by the HGT mentioned earlier. No example of HGT was found within mitochondrial genes of three other, independently evolved parasitic lineages we sampled: *Cassytha*/Laurales, *Krameria*/Zygophyllales, and Lennooideae/Boraginales.

Key words: hemiparasitic, heterotrophy, holoparasitic, mitochondrion, Solanales.

Significance

Horizontal gene transfers (HGTs) from the mitochondrial genome of *Cuscuta* to host have been reported, but not vice versa. Our results show mitochondrial *atp1* genes of South African subgenus *Pachystigma* were inferred to be transferred from Lamiales, with high support; this horizontally transferred *atp1* gene has replaced the native copy and may remain functional; the incongruence between mitochondrial phylogeny of *Cuscuta* and previous plastid/nuclear phylogeny may be caused by the HGT. We reveal the first example of organellar HGT from host to *Cuscuta*, and indicate that HGTs within mitochondrial genomes of *Cuscuta* are likely to be rare events.

Introduction

The vast majority of land plants conduct photosynthesis to fix carbon, and thus do not rely on other organisms for their energy. However, heterotrophic plants obtain part or even all of their nutrients from their host plants. Haustorial (or "true") parasitism involves a direct tissue link between the vasculature of parasites and autotrophs via a haustorial connection (Kuijt 1969). This type of heterotrophy has

evolved independently at least 12 times in angiosperms (Nickrent 2020). Photosynthetic ability in parasitic plants is affected along a trophic spectrum, whereby hemiparasitic plant can still conduct photosynthesis at some level, whereas holoparasitic plants have completely lost the ability to do so, and hence acquire all nutrients from their hosts. The trophic shift to holoparasitism is usually accompanied with reduction of vegetative features, diminishing quantities of chlorophyll, and loss of plastid genes involved in photosynthesis (e.g., Kuijt 1969; Westwood et al. 2010; Wicke et al. 2011; Molina et al. 2014).

Cuscuta (dodders), the sole parasitic genus in Convolvulaceae (the morning-glory family), comprises about 200 species of stem parasites (Costea et al. 2015). This genus has nearly cosmopolitan distribution, and is ecologically and economically important in many regions (Yuncker 1932). Several *Cuscuta* species can cause massive losses to agriculture (Dawson et al. 1994), whereas some of them have been used in traditional Chinese medicine over hundreds of years (Zheng et al. 1998). Recent phylogenetic studies based on sequences from several nuclear and plastid regions subdivided *Cuscuta* into four subgenera: *Cuscuta*, *Grammica*, *Monogynella*, and *Pachystigma* (García et al. 2014; Costea et al. 2015). Monophyly of each subgenus and relationships among them are all well supported, with *Monogynella* sister to the rest of this genus (García et al. 2014). Relationships among different sections in subg. *Grammica*, the most species-rich subgenus (~150 spp.) are also well resolved (Stefanović et al. 2007; García et al. 2014). However, attempts to retrieve any plastid sequences from section Subulatae have failed (Braukmann et al. 2013; García et al. 2014). Thus, the placement of this section as sister to the rest of subg. *Grammica* currently relies only on nuclear ribosomal DNA sequence data (García et al. 2014).

Unlike most other parasitic lineages, *Cuscuta* spans the trophic spectrum, comprising both hemi- and holoparasitic species (Dawson et al. 1994). As such, this genus has become an excellent system to study genomic evolution of parasitism. Species of the subg. *Monogynella* are chlorophyllous and considered to be hemiparasitic (Dawson et al. 1994). Holoparasitic *Cuscuta* has been identified from the whole sect. Subulatae and three closely related species in sect. Ceratophorae (Braukmann et al. 2013; Banerjee and Stefanović 2019). The remaining *Cuscuta* species have been deemed "cryptically" photosynthetic (McNeal et al. 2007), that is, still able of conducting very limited and localized photosynthesis (Dawson et al. 1994; van der Kooij et al. 2000).

Recent phylogenomic studies of *Cuscuta* generally focused on their plastid genome and revealed extensive levels of gene loss, especially regarding the *ndh* gene family (Braukmann et al. 2013; Banerjee and Stefanović 2019, 2020). Compared with plastid genomes, mitochondrial genomes of plants evolve more slowly at the DNA substitution level, and generally do not experience substantial gene losses (Knoop 2004; Yurina and Odintsova 2016; Lin 2020). However, plant mitochondrial genomes undergo frequent fusion and recombination, and often incorporate sequences from their own plastid or nuclear genomes, obtained through intracellular gene transfer (e.g., Stern and Lonsdale 1982; Schuster and Brennicke 1987). As a corollary of these phenomena, sometimes genes are also acquired from other species, via horizontal gene transfer (HGT; e.g., Bergthorsson et al. 2004; Qiu et al. 2014). Cases of HGT involving mitochondrial genomes of parasitic plants have been reported repeatedly over the last couple of decades (reviewed by Petersen et al. 2020). For example, multiple host-to-parasite HGTs of *atp1* have been reported in *Cytinus* (Cytinaceae), *Mitrastemon* (Mitrastemonaceae), *Pilostyles* (Apodanthaceae), and Rafflesiaceae (Barkman et al. 2007). Cases of mitochondrial HGT from *Cuscuta* to *Plantago* (Plantaginaceae; *atp1*, *atp6*, and *matR*) and *Geranium* (Geraniaceae; *cox2*, *rpl5*, and *rps4*) have also been recorded (Mower et al. 2004, 2010; Park et al. 2015) as well as those from nuclear genomes of hosts to *Cuscuta* (Vogel et al. 2018; Yang et al. 2019). However, an example of HGT from the mitochondrial genome of a host to *Cuscuta* has not been reported so far. For example, two recent studies reported a lack of evidence for HGT in mitochondrial genomes of two closely related *Cuscuta* species from subg. *Grammica* (Anderson et al. 2021), and of seven species from three subgenera of *Cuscuta* except *Pachystigma* (Lin et al. 2022).

Compared with plastomes, mitogenomes of parasitic plants have not been nearly as well studied yet, as pointed out by Petersen et al. (2020). Mitochondrial genomes of angiosperms usually contain genes for subunits of the respiratory chain complexes I (*nad* genes), II (*sdh*), III (*cob*), IV (*cox*), and V (*atp*), large and small ribosomal protein (*rpl* and *rps*), subunits of a cytochrome c maturation pathway (*ccm*), matурase (*matR*), and a subunit of a twin-arginine translocase (*tatC* or *mttB*). This represents about 40 protein-coding genes in total, with another three rDNA genes, and about 20 tRNA genes (e.g., Stern and Palmer 1986). However, this number can vary between different taxa, by as much as 20 genes (Knoop 2004; Skippington et al. 2015; Mower 2020). For example, the number of *sdh* and ribosomal proteins genes have been known to vary greatly between different plant lineages (Adams et al. 2002). Here we present a mitochondrial phylogenomic analysis of 26 *Cuscuta* species, including multiple representatives from three out of four subgenera. The primary goal was to characterize these genomes and confirm or reassess phylogeny of *Cuscuta* based on 31 mitochondrial genes with the previous phylogenetic estimates for this parasitic genus. Also, to further expand available mitochondrial data and to establish if we can find any repeated patterns

in mitochondrial evolution across independently evolved groups of parasitic plants, we sequenced three additional parasitic lineages, from three different orders: *Cassytha*/Laurales, *Krameria*/Zygophyllales, and Lennooideae/Boraginales. Possible examples of HGT in the mitochondrial genome of *Cuscuta* and other sampled groups are also investigated.

Results

Gene Recovery

Frequent recombination makes the linear order of most plant mitochondrial genomes very variable (Gualberto et al. 2014), and causes substantial problems when trying to assemble entire mitochondrial genomes. We did not attempt to assemble complete mitochondrial genomes from next generation sequencing data here, but instead obtained individual mitochondrial genes for our phylogenomic study. We did not find evidence of substantial gene loss in mitochondrial genomes, neither for *Cuscuta* nor for the other parasitic plants we sampled in this study, when compared with their respective autotrophic relatives. Occasional absences noted in supplementary material figure S1, Supplementary Material online may represent gene loss or a failure of gene recovery due to the relatively low coverage (between 10× to 1,000×) of mitochondrial data. Except for *atp1* in two species of *Cuscuta* subg. *Pachystigma*, we did not find any other case of HGT in our sampling.

Phylogeny and Analyses of *atp1*

In our *atp1* phylogeny, all *Cuscuta* species except those belonging to subg. *Pachystigma* are grouped together with moderate support (89% bootstrap), and are nested within Convolvulaceae (87% bootstrap) and Solanales (fig. 1). However, samples from subg. *Pachystigma* are nested within Lamiales with high support (99% bootstrap), and are sister to all Lamiales except *Plocosperma* (Plocospermataceae) and *Olea* (Oleaceae; fig. 1). The constraint test shows that the phylogeny shown in figure 1 is significantly better than the alternative which includes all *Cuscuta atp1* in one clade (approximately unbiased [AU] and Shimodaira–Hasegawa [SH] tests comparing trees from best unconstrained vs. constrained maximum likelihood [ML] analyses: $P < 0.03$). *Atp1* sequences of all *Cuscuta* individuals sampled possess a complete open reading frame, including those of subg. *Pachystigma*. For other sampled parasitic plant lineages, *atp1* of two (*Krameria* and *Cassytha*) are grouped together with their autotrophic relatives (Zygophyllales and Laurales, respectively) with high support (98–100% bootstrap). An exception is the position of the third lineage, Lennooideae, which is not nested within Boraginales, as expected. This relationship received only poor support, along with other backbone relationships (fig. 1). We predicted 19 RNA editing sites in total within the DNA matrix (supplementary material table S1, Supplementary Material online). The deletion of RNA editing sites does not influence the phylogenetical topology and support values for any of the major clades (supplementary material fig. S2, Supplementary Material online).

The putative HGT *atp1* of *C. africana* was found as a part of a long assembly contig (119,092 bp, generated from 22,016 reads; supplementary material fig. S3, Supplementary Material online), together with 13 other mitochondrial genes, including *atp4*, *ccmFc*, *nad2*, *rps12*, etc. The quantity of *C. nitida* data was not as ample as that for *C. africana*; hence, the contig with foreign *atp1* (generated from 209 reads) did not include other mitochondrial genes. After mapping back all raw reads of *C. africana* and *C. nitida* to the sequence of native *Cuscuta atp1*, with a total of 16,009 (*C. africana*, 828.4 average coverage) and 5,023 (*C. nitida*, 247.8 average coverage) reads, respectively, we found no reads that represented the potential native copy. No part of the HGT *atp1* copy is found to be chimeric, that is, a product of recombination between the orthologous and xenologous copies of *atp1* after the HGT event. The polymerase chain reaction (PCR) survey recovered only the native, vertically transmitted *Cuscuta atp1* copy from the control, and only the HGT xenologous copy of *atp1*, phylogenetically related to Lamiales, from all samples of subg. *Pachystigma* (supplementary material fig. S4, Supplementary Material online). The selection test shows that when compared with other *Cuscuta* species, both native and HGT *atp1* copies are under strong purifying selection ($\omega = 0.166$ and 0.201, respectively), and are not significantly different ($P = 0.583$). Similarly, when compared with Lamiales, both native and HGT *atp1* copies are also under strong purifying selection ($\omega = 0.077$ and 0.072, respectively), and are not significantly different from each other ($P = 0.929$).

Mitochondrial Phylogenomic Inference for *Cuscuta*

In our phylogenetic tree based on all retrieved mitochondrial genes except *atp1*, all sampled parasitic lineages are grouped together with their respective autotrophic relatives with high support (fig. 2). We predicted 803 RNA editing sites in total within the DNA matrix (supplementary material table S1, Supplementary Material online). The deletion of RNA editing sites does not influence the phylogenetical topology and support values for any of the major clades (supplementary material fig. S5, Supplementary Material online). Within *Cuscuta*, subg. *Cuscuta* species are inferred to be the sister group to the rest of sampled *Cuscuta*, consistent with plastid- and nuclear-derived phylogenies (García et al. 2014), but only with poor support

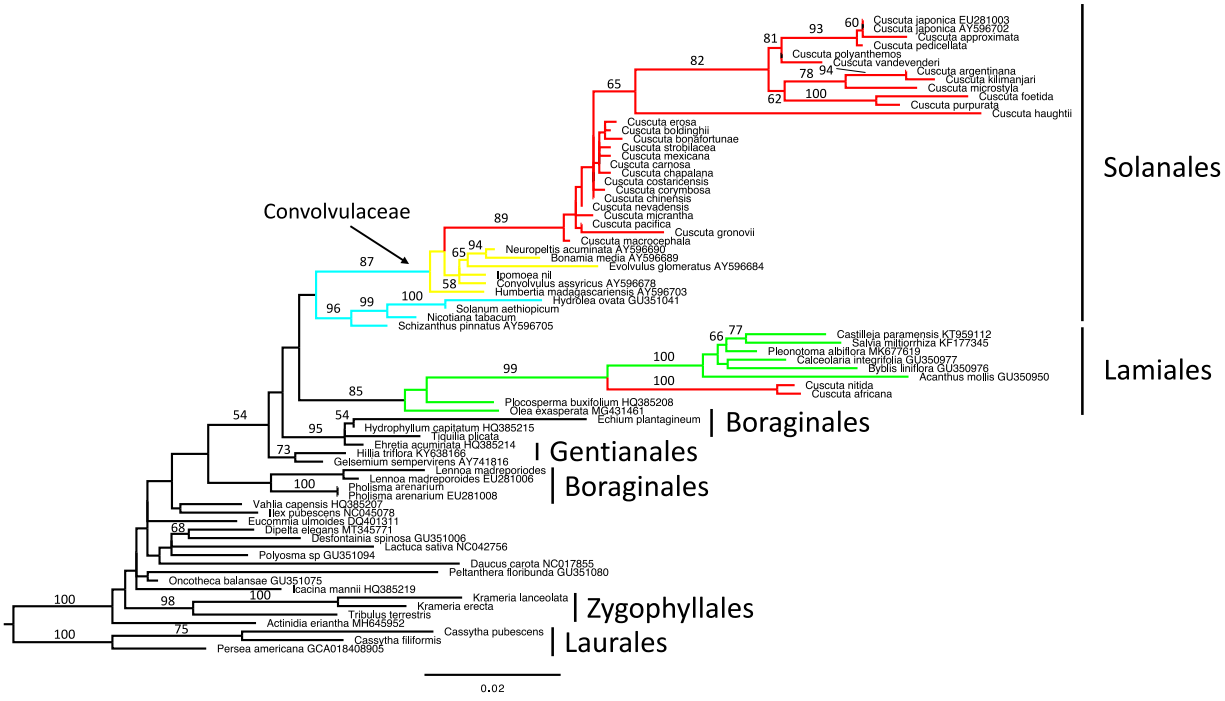

Fig. 1.—Maximum likelihood phylogeny of the mitochondrial *atp1* in *Cuscuta* and three other parasitic lineages sampled in this study, in a broader phylogenetic context representing the diversity across dicots. Bootstrap values >50% are indicated beside branches. Scale bar indicates estimated substitutions per site. (Red lineages: *Cuscuta*; yellow lineages: other taxa in Convolvulaceae; blue lineages: other taxa in Solanales; green lineages: Lamiales.)

(fig. 2). However, the phylogeny with *atp1* included places subg. *Pachystigma* as the sister group to the rest, with strong support (supplementary material fig. S6, Supplementary Material online). The constraint test on the alignment without *atp1* shows that the original topology is not significantly better compared with the alternative one, in which the relative positions of subg. *Pachystigma* and *Cuscuta* are switched around (AU and SH tests comparing trees from best unconstrained vs. constrained ML analyses: $P = 0.451$ and 0.466, respectively). Within subg. *Grammica*, five samples representing sect. Subulatae are found monophyletic, and placed as the sister to the rest of this subgenus with high support (fig. 2), in accordance with nuclear-derived phylogenies (Stefanović et al. 2007; García et al. 2014). Other relationships within closely allied species of subg. *Grammica* generally received only weak to moderate support (fig. 2).

Discussion

Mitochondrial Genes of Parasitic Plants and the Effect of RNA Editing Sites

Unlike plastomes, mitochondrial genomes of parasitic plants are generally much more conservative in gene content and substitutional level (Petersen et al. 2020). Previous studies found little evidence of gene loss in the mitochondrial genome of parasitic plants, except in Viscaceae, which have functionally lost all nine *nad* genes (Petersen et al. 2015; Skippington et al. 2015, 2017). Our results show a similar pattern, where none of our sampled parasitic plant lineages experienced extensive levels of gene loss compared with their autotrophic relatives (supplementary material fig. S1, Supplementary Material online). We also find that all *Cuscuta* we sampled possess fragmented *ccmFc* genes, in which exon 1 splits into two pieces, consistent with the observation in recent studies (Anderson et al. 2021; Lin et al. 2022). Based on the phylogeny, each parasitic lineage is also grouped together with their autotrophic relatives with high support (fig. 2). Compared with mitogenomes, plastomes of many parasitic plants, especially holoparasitic plants, experienced structural changes and extensive level of gene loss, due to the degradation or even complete loss of photosynthetic function (e.g., Wicke et al. 2011; Braukmann et al. 2013; Molina et al. 2014). For example, all *ndh* genes (responsible for regulating photosynthetic electron flow, mitigating effects of PS I inhibition, and fine-tuning photosynthesis in dynamic light conditions) are lost or pseudogenized in *Cuscuta* (Braukmann et al. 2013; Banerjee and Stefanović 2020).

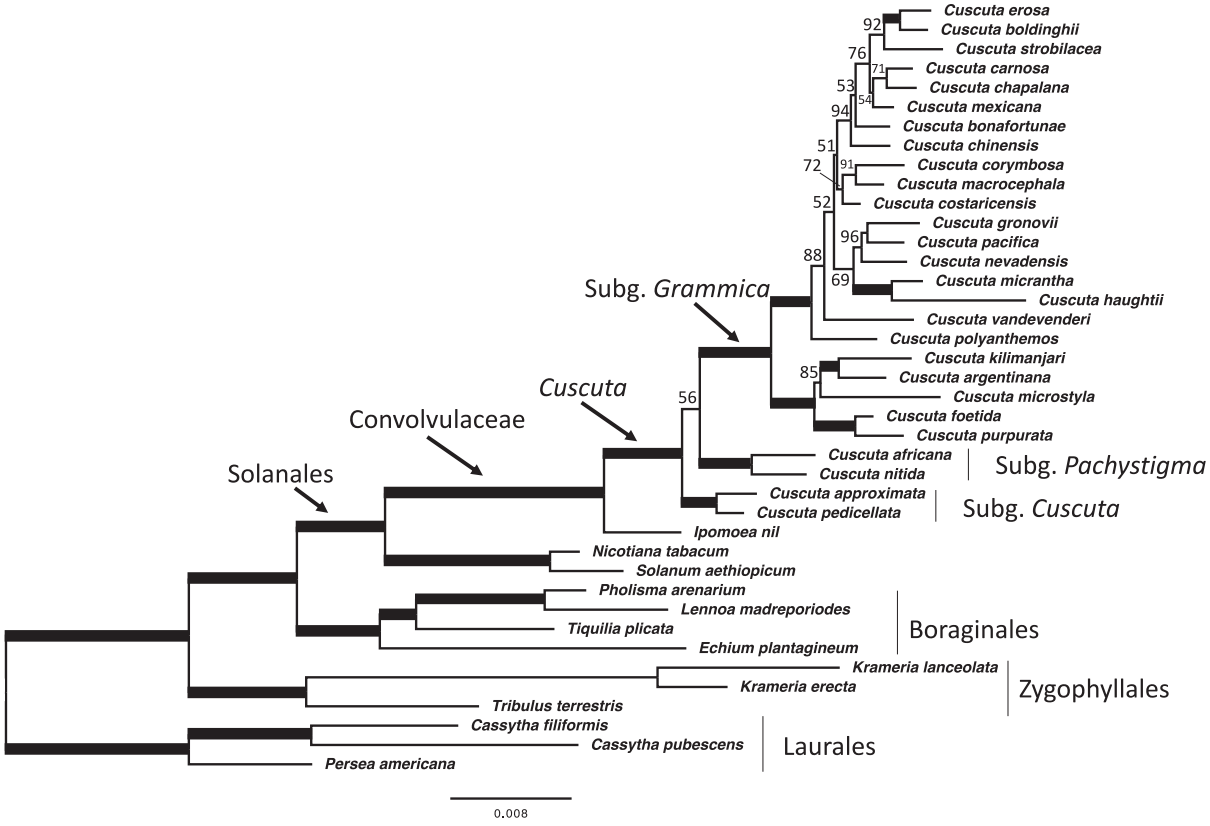

Fig. 2.—Phylogenetic placement of *Cuscuta* and three other parasitic lineages sampled in this study, inferred from partitioned maximum likelihood analyses of 30 mitochondrial genes (without *atp1*) using a "G × C" partitioning scheme (see text). Thick lines indicate 100% bootstrap; bootstrap values <100% and >50% indicated beside branches.

In *Cuscuta* subg. *Cuscuta*, the inverted repeat has also been lost (Banerjee and Stefanović 2020). Considering the essential role of mitochondrial genes in respiration and biosynthetic pathways, parasitism seems not to have influenced the evolution of those genomes as profoundly as plastomes (Petersen et al. 2020).

RNA editing sites have been shown to potentially mislead phylogeny when edited transcripts of genes are reverse transcribed and then replace the native genomic copies (Bowe and dePamphilis 1996). The deletion of all predicted RNA editing sites from our data sets does not affect the interpretation of our results and conclusion below (supplementary material figs. S2 and S4, Supplementary Material online).

HGT from Host to *Cuscuta*

The intimate association between parasitic plants and their hosts through haustorial connections makes HGT relatively easy to happen. The active DNA uptake system of the mitochondria also makes it easier to host foreign sequences (Koulintchenko et al. 2003). Although this process may be not as active as in nuclei (Vogel et al. 2018; Yang et al. 2019), several examples of HGT from the mitochondrial genome of host-to-parasitic plants, and vice versa, have been recorded in different lineages (reviewed recently by Petersen et al. 2020). However, most of those transfers lead to pseudogenization, with only a few examples of potential replacement of the original mitochondrial genes, regain of genes that had been lost, or coexistence and gene conversion with the original copy (e.g., Bergthorsson et al. 2003; Mower et al. 2012; Sanchez-Puerta et al. 2017). Within mitochondrial genes, despite two examples of HGT from *Cuscuta* to *Plantago* and *Geranium* (Mower et al. 2004, 2010; Park et al. 2015), HGT from a host to *Cuscuta* was heretofore unknown. Our results show that mitochondrial *atp1* of *Cuscuta* subg. *Pachystigma* are nested within Lamiales with high support (fig. 1). They form a deep split which is sister to core Lamiales except *Plocosperma* and *Olea*, indicating a potentially ancient transfer from a Lamiales host to *Cuscuta*. The current host range of subg. *Pachystigma* is wide, spanning many species in Asteraceae, Fabaceae, Proteaceae, etc., including some Lamiales members (e.g., *Oftia*, Scrophulariaceae; Meeuse and Welman 2000). However, the hosts of today are not necessarily the same as the host of a few millions of years ago. Targeted PCR survey and sequencing confirmed *atp1* HGT for additional samples of *C. nitida* and *C. africana*, indicating that this HGT event likely have

happened in the common ancestor of the whole subgenus. This finding needs to be confirmed with other species belonging to *Cuscuta* subg. *Pachystigma*. In addition to the connection via haustorial parasitism, other direct plant–plant contacts as well as illegitimate pollination and vector-mediated transfer have also been proposed as alternative mechanisms for HGT (Keeling and Palmer 2008; Bock 2010).

Within the nuclear genome of *C. campestris*, multiple HGT events from a variety of different hosts have been identified, including six HGT events from Lamiales (Vogel et al. 2018; Yang et al. 2019). In contrast, more recent studies found no evidence of HGT in mitochondrial genes of seven *Cuscuta* species across subg. *Cuscuta*, *Grammica*, and *Monogynella* (Anderson et al. 2021; Lin et al. 2022). *Grammica* is the largest subgenus of *Cuscuta*, comprising about three-fourth of the species diversity of this genus, from which we included 23 samples, representing nearly all of its 15 sections (supplementary material table S1, Supplementary Material online). Despite this thorough and phylogenetically diverse sampling strategy, no additional HGT has been found in our mitochondrial results, supporting the notion that the xenologous copy of *atp1* in subg. *Pachystigma* most likely represents an independent event from those in the nuclear genome in *C. campestris* (Vogel et al. 2018). Anderson et al. (2021) concluded that the apparent lack of mitochondrial HGT in *Cuscuta* may be due to the relatively small size of their mitogenomes, which limits the capacity to integrate foreign DNA. Supporting this notion is a recent finding of "HGT-like" sequences that have been reported within the mitogenome of *C. japonica* (Lin et al. 2022). In this species, belonging to subg. *Monogynella*, the mitochondrial genome is substantially inflated, exceeding 800 kbp in size, and is twice as large as the next largest mitogenomes in *Cuscuta*, those in subg. *Cuscuta*. However, none of its putatively horizontally transferred sequences were found in coding regions (Lin et al. 2022). Our finding indicates that although HGTs overall may be easier to happen in large mitogenomes, the HGTs of coding genes are still likely to be rare events within mitochondrial genomes of *Cuscuta*.

The HGT *atp1* of *C. africana* was found as a part of a long assembly contig together with other 13 mitochondrial genes (supplementary material fig. S3, Supplementary Material online). Along with its high coverage (23.2 average), this strongly supports the notion that the HGT copy of *atp1* is found within the mitochondrial genome of *C. africana*. From the next generation sequencing data of subg. *Pachystigma*, with high total reads (>5,000) and average coverage (>200×), we could not recover any raw reads that represent the original, vertically transmitted copy of *atp1*. We also could not recover an orthologous *atp1* copy from the genome of subg. *Pachystigma* by PCR, despite the fact that the same primers and PCR condition easily amplified the native copy of this gene in the control. This result is consistent with a possible functional replacement of the original gene by the HGT copy. A scenario like this has been reported in other parasitic plants (Sanchez-Puerta et al. 2017). Further supporting this notion, horizontally transferred *atp1* genes have intact open reading frames in both sampled species, and are under strong purifying selection, which strongly suggests that they remain functional. However, this finding needs to be confirmed through expression studies, with reverse transcription-PCR and/or the western blot.

We have not found any example of HGT in other sampled parasitic plants—Lennoaceae, Krameriaceae, and *Cassytha* (Lauraceae)—congruent with former studies using mitochondrial genes/genomes (Barkman et al. 2007; Zervas et al. 2019). However, our sampling of these lineages and their autotrophic relatives are not as dense as *Cuscuta*.

Phylogeny of *Cuscuta*

Despite the unclear placement of *Cuscuta* within Convolvulaceae (Stefanović and Olmstead 2004), the backbone phylogeny within *Cuscuta* has been well resolved by plastid and nuclear genes (García et al. 2014; Costea et al. 2015). Subg. *Monogynella* is the sister to the rest of *Cuscuta*, followed by subg. *Cuscuta* sister to a clade containing *Pachystigma* and *Grammica*, all with high support in the plastid (*rbcL*) phylogeny, and moderate to high support in the nuclear (nrLSU) phylogeny (García et al. 2014). Our phylogeny excluding *atp1* is congruent with this topology, although with poor support (fig. 2). However, the phylogeny based on all mitochondrial genes, including *atp1*, provides a strongly supported alternative topology, with subg. *Pachystigma* sister to subgenera *Cuscuta* and *Grammica* (supplementary material fig. S6, Supplementary Material online). As mentioned earlier, the *atp1* HGT in subg. *Pachystigma* contributes substantially to this topological conflict. This conflict further illustrates how HGT can strongly influence the phylogenetic placement of heterotrophic plants, and HGT should be considered in all studies when utilizing mitochondrial genes for phylogenetic purposes.

We also recovered mitochondrial data from sect. Subulatae, a holoparasitic lineage of *Cuscuta* from which no plastid genes have ever been recovered (Braukmann et al. 2013; García et al. 2014). We confirm that this section is the sister group to the rest of subg. *Grammica*, which is congruent with the nuclear phylogeny (García et al. 2014). However, other relationships within subg. *Grammica* are generally without high support. The few informative sites provided by the mitochondrial genome due to its conservative nature may contribute to the low support for those shallow nodes in the phylogeny.

Materials and Methods

Taxon Sampling, DNA Isolation, and Sequencing

We generated mitochondrial genome sequences for 26 *Cuscuta* species and retrieved sequences of an additional one (*Cuscuta gronovii*) from GenBank. In total, this represents 14 (out of 18) sections, and 3 (out of 4) subgenera (see supplementary material table S2, Supplementary Material online for details). Within this comprehensive sampling, 19 species are considered cryptically photosynthetic, whereas 8 species are holoparasitic (3 species in sect. Ceratophorae and all 5 species in sect. Subulatae; supplementary material table S2, Supplementary Material online). We also retrieved sequences of one autotrophic relative from the same family as *Cuscuta* (*Ipomoea nil*; Convolvulaceae), and two autotrophic relatives in the same order (*Nicotiana tabacum* and *Solanum aethiopicum*; Solanaceae) from GenBank (supplementary material table S1, Supplementary Material online).

Our sampling for three other lineages of independently evolved parasitic plants was representative but less detailed. In Boraginales, we generated mitochondrial genome sequences for two parasitic taxa from the holoparasitic subfamily Lennooideae in Boraginaceae (*Lennoa madreporoides* and *Pholisma arenarium*), and one autotrophic close phylogenetic relative from the same family (*Tiquilia plicata*). Furthermore, we retrieved sequences of one additional autotrophic relative (*Echium plantagineum*; Boraginaceae) from GenBank (supplementary material table S2, Supplementary Material online). In Zygophyllales, we generated mitochondrial genome sequences for hemiparasitic *Krameria erecta* (Krameriaceae) and its autotrophic relative from the same order (*Tribulus terrestris*; Zygophyllaceae). We also retrieved sequences of another hemiparasitic *K. lanceolata* (Krameriaceae) from GenBank (supplementary material table S2, Supplementary Material online). Finally, in Laurales, we generated mitochondrial genome sequences for hemiparasitic *Cassytha filiformis*, retrieved sequences of another hemiparasitic *Cassytha* (*Cassytha pubescens*) and their autotrophic relative from the same family (*Persea americana*; Lauraceae) from GenBank (supplementary material table S2, Supplementary Material online).

In aggregate, we generated a 40-taxon mitochondrial matrix that represents three out of four *Cuscuta* subgenera, four parasitic lineages in different orders (*Cuscuta*/Solanales; Lennooideae/Boraginales; *Krameria*/Zygophyllales; *Cassytha*/Laurales), and ensured each parasitic lineage has at least one autotrophic close relative within the same family (or the same order when the whole family is parasitic) to compare with. We also retrieved an additional 35 *atp1* sequences from GenBank (supplementary material table S2, Supplementary Material online) to represent the diversity of dicots, and thereby provide a broad phylogenetic context in which to study the evolution of *atp1* and potential cases of HGT. To this end, we generated a 75-taxon *atp1* matrix, including all four *Cuscuta* subgenera.

Total genomic DNA was isolated from herbarium or silica-dried tissue using the modified cetyltrimethylammonium bromide method (Doyle and Doyle 1987) and checked for quantity and quality using a Nano Drop 1000 Spectrophotometer (Thermo Fisher Scientific). DNA extractions from *P. arenarium*, *C. strobilacea*, and *C. kilimanjari* were sent to Genome Quebec at McGill University in Montreal, Quebec for library preparation and high-throughput sequencing on their Illumina HiSeq 2000 platform using a 2×100 paired-end read format. All other extractions were sent for high-throughput sequencing on an Illumina Hi-Seq 2500 platform (2×126 bp paired-end reads; The Centre for Applied Genomics, SickKids Hospital, Toronto, Ontario). Demultiplexing of raw reads and the removal of indexing barcodes were performed at the sequencing facility.

Contig Assembly, Sequence Alignment, and Data Matrix Construction

We obtained contigs by performing de novo assemblies using Geneious R10 (Biomatters, Auckland, New Zealand) with default settings, using 20% of raw reads, and selected contigs with an average of >10× coverage. We retrieved mitochondrial genes using the blastn program from BLAST+ NCBI (version NCBI-BLAST-2.2.30+, Camacho et al. 2009), using *C. gronovii* (KP940494–514) as query. We recovered data for the 31 protein-coding genes common to most *Cuscuta*, and set up individual gene files for these, each with 40 taxa. We recovered nearly all mitochondrial genes from most taxa (see supplementary material fig. S1, Supplementary Material online), without evidence of substantial gene loss compared with estimates from southern hybridization survey of mitochondrial gene content in angiosperms (Adams et al. 2002). Unrecovered genes were coded as missing data (supplementary material fig. S1, Supplementary Material online).

We conducted initial alignment of individual mitochondrial genes as nucleotides using the MUSCLE (Edgar 2004a, b) online portal, with default settings, and concatenated gene files manually. We then used Mesquite v.3.15 (Maddison and Maddison 2018) to perform manual adjustments of the matrix, following Graham and Olmstead (2000). We also removed several regions hard to align in case they might influence phylogenetic analyses. The final alignment is a 30,222-bp matrix for the concatenated protein-coding genes (for reference, derived from 28,325 bp of unaligned data in *C. africana*). The *atp1* alignment is 1,731-bp matrix long (for reference, derived from 1,521 bp of unaligned data in *C. africana*). We also generated a concatenated alignment without *atp1*. In order to avoid potential misleading effects of RNA editing sites on

phylogenetic inference (Bowe and dePamphilis 1996), we used PREPACT 3.12.0 (http://www.prepact.de/prepact-main.php, Lenz et al. 2010; Lenz and Knoop 2013; Lenz et al. 2018) to predict RNA editing sites in our concatenated DNA matrix (with *N. tabacum* as the reference), and constructed new matrices without those sites (see supplementary material table S1, Supplementary Material online, for locations of predicted sites in matrices). We deposited new sequence data in GenBank (OM749746-OM751323), and alignments on FigShare (https://figshare.com/s/9de6a034d89c5ade148e).

Phylogenetic Analyses

We performed ML analyses using RAxML v7.4.2 (Stamatakis 2006) with a graphical interface (Silvestro and Michalak 2012), considering partitioned likelihood analysis of DNA data. We partitioned nucleotides using a gene by codon ("G × C") partitioning scheme, starting with initial individual partitions derived from first, second, and third codon positions of each gene, representing 93 initial partitions for the complete concatenated DNA matrix, 90 initial partitions for the concatenated DNA matrix without *atp1*, and 3 initial partitions for the *atp1* matrix. We combined partitions that did not have significantly different substitution models by using PartitionFinder2 (Lanfear et al. 2016) with the relaxed hierarchical clustering algorithm (r-clustering) and the corrected Akaike Information Criterion, limiting the DNA substitution models under consideration to those implemented in RAxML v. 7.4.2 (see supplementary material table S3, Supplementary Material online, for final partitions and models). The final partitioning schemes (supplementary material table S3, Supplementary Material online) include 48 partitions for the complete concatenated DNA matrix, 49 partitions for the concatenated DNA matrix without *atp1*, and 3 partitions for the *atp1* matrix. GTR + G or GTR + I + G DNA substitution models were inferred to be the optimal fit for the DNA version of the matrix; we used the GTR + G model for all partitions as the "I" parameter for invariant sites may be accommodated by the gamma-distribution shape parameter "alpha" (Yang 2006). For all analyses, we ran 20 independent searches for the best tree and estimated branch support using 500 bootstrap replicates (Felsenstein 1985). We only ran unpartitioned ML analyses for matrices without RNA editing sites. We also conducted unpartitioned ML searches for the best tree on each individual gene (GTR + G model, five independent searches) to detect potential HGT (see supplementary material fig. S7, Supplementary Material online for each individual gene tree). To find potential cases of HGT, we performed blast searches in GenBank for all sequences with >20× average coverage, and paid especial attention to genes with unexpected phylogenetic placement (supplementary material fig. S7, Supplementary Material online). This allowed us to assess if any parasitic lineage was unexpectedly more similar to some of the more distant taxa, indicative of HGT. We considered highly supported branches to have ≥95% bootstrap support and poorly supported branches to have <70% support, following Soltis and Soltis (2003). We also performed constraint tests on: (1) concatenated matrix without *atp1* to evaluate if the placement of subg. *Pachystigma* in figure 2 is significantly different compared with its placement obtained from the alignment with *atp1* (supplementary material fig. S6, Supplementary Material online); and (2) *atp1* matrix to see if the phylogeny which places all *Cuscuta atp1* in one clade is significantly different compared with the phylogeny shown in figure 1. We found the shortest likelihood tree that satisfied constraint of monophyly of subg. *Cuscuta* and *Grammica*, and monophyly of all *Cuscuta atp1* in RaxML, and compared the resulting tree sets to test if the constrained trees were significantly worse, using the AU (Shimodaira 2002) and SH tests in CONSEL (Shimodaira and Hasegawa 2001), considering site-likelihoods from unpartitioned ML analysis.

HGT Confirmation and Tests

To try to recover the native *atp1* copy in the genomes of *C. africana* and *C. nitida*, if it existed, we first mapped back all next generation sequencing raw reads of these two species to the orthologous *atp1* sequence of *C. gronovii* in Geneious R10. To confirm the HGT of *atp1* in *Cuscuta* subg. *Pachystigma*, we also conducted PCRs on additional samples of *C. africana* (*Stefanović SS-17-141*) and *C. nitida* (*Stefanović SS-17-126*, *Stefanović SS-17-143*), with *C. psorothamnensis* (*Stefanović SS-13-07*) as the control outside of subg. *Pachystigma*. All vouchers are deposited in the Herbarium of the University of Toronto Mississauga. We used atp1-4 (5′- CATTCGATCACAGAAKCCRTT-3′) and atp1-5 (5′- GAGCTGCGGAACTMACVAVTC-3′) as primers. Under routine PCR conditions (1 unit of regular Taq DNA polymerase per 50 μl reaction, final concentration of 1 × reaction buffer, 200 μM dNTPs, 200 nM primers, 55 °C primer annealing temperature, 90 s extension at 72°C, 35 cycles), this primer combination is expected to be able to amplify both HGT *atp1* xenolog as well as native, vertically transmitted *Cuscuta atp1* copy.

After finding no evidence of the native, orthologous copy of *atp1* in the representatives of subg. *Pachystigma*, either by bioinformatics means or by molecular techniques, we wanted to investigate the possibility that the HGT *atp1* sequence found in these two species could have been chimeric in nature. To that end, we conducted a recombination test using GENECONV v. 1.81 (Sawyer 1989).

We also used the CodeML module in PAML4.8 (Yang 2007) to assess changes in selective regime in HGT *atp1* gene in *Cuscuta* subg. *Pachystigma*. The objective was to

test hypotheses of different ω values (ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site) for the HGT *atp1* copy, compared separately with the original *Cuscuta* or Lamiales *atp1* homologs shown in figure 1. We built two codon-based "branch" models, which can detect differences in selection regimes in particular lineages (Yang 2007). In the simpler model (M0, one ratio), all branches evolve under one ω-ratio. In the alternative model (M1, two ratios), HGT *atp1* was allowed to evolve under a different ω-ratio than the original *Cuscuta* or Lamiales *atp1*. We removed taxa in alignments lacking >10% sequence data and regions with indels that resulted in missing data for 90% or more of the taxa and built a matrix with 22 *Cuscuta* taxa, and a matrix with eight Lamiales taxa shown in figure 1 and two *Cuscuta* subg. *Pachystigma*. We used the likelihood ratio test statistic $-2(\ln L\ \mathrm{M1} - \ln L\ \mathrm{M0})$ to compare the fit of the null versus alternative model, and calculated *P*-values based on a χ^2 test with 1 degree of freedom.

Supplementary Material

Supplementary figures and supplementary tables are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Olga Alimowska for conducting PCR as well as Mihai Costea, Alison Colwell, Eleazer Carranza, Ignacio García Ruiz, Marc Johnson, Tom Van Devender, and the directors and curators of herbaria for providing tissue crucial to this project. 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. They also thank Benjamin M. Anderson and one anonymous reviewer for their constructive feedback on earlier versions of this article.

Author Contributions

Q.L. and S.S. designed the research. S.S. provided materials. Q.L. and A.B. did the data collection. Q.L. performed most of the data analysis. Q.L. and S.S. performed the data interpretation and wrote the manuscript. A.B. participated in manuscript writing and editing.

Data Availability

All data supporting the findings of this study are available within the paper, its supplementary materials published online, or are openly available in GenBank and FigShare data repository (https://figshare.com/; https://figshare.com/s/9de6a034d89c5ade148e).

Literature Cited

Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U S A. 99:9905–9912.

Anderson BM, Krause K, Peterson G. 2021. Mitochondrial genomes of two parasitic *Cuscuta* species lack clear evidence of horizontal gene transfer and retain unusually fragmented *ccmFc* genes. BMC Genom. 22:816.

Banerjee A, Stefanović S. 2019. Caught in action: Fine-scale plastome evolution in the parasitic plants of *Cuscuta* section Ceratophorae (Convolvulaceae). Plant Mol Biol. 100:621–634.

Banerjee A, Stefanović S. 2020. Reconstructing plastome evolution across the phylogenetic backbone of the parasitic plant genus *Cuscuta* (Convolvulaceae). Bot J Linn Soc. 194:423–438.

Barkman TJ, et al. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol Biol. 7:248.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 424:197–201.

Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc Natl Acad Sci U S A. 101:17747–17752.

Bock R. 2010. The give-and-take of DNA: Horizontal gene transfer in plants. Trends Plant Sci. 15:11–22.

Bowe LM, dePamphilis CW. 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. Mol Biol Evol. 13:1159–1166.

Braukmann T, Kuzmina M, Stefanović S. 2013. Plastid genome evolution across the genus *Cuscuta* (Convolvulaceae): Two clades within subgenus Grammica exhibit extensive gene loss. J Exp Bot. 64:977–989.

Camacho C, et al. 2009. BLAST+: Architecture and applications. BMC Bioinform. 10:421.

Costea M, Garcia MA, Stefanović S. 2015. A phylogenetically based infrageneric classification of the parasitic plant genus *Cuscuta* (dodders, Convolvulaceae). Syst Bot. 40:269–285.

Dawson JH, Musselman LJ, Wolswinkel P, Dorr I. 1994. Biology and control of *Cuscuta*. Rev Weed Sci. 6:265–317.

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

Edgar RC. 2004a. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Edgar RC. 2004b. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5:113.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39:783–791.

García MA, Costea M, Kuzmina M, Stefanović S. 2014. Phylogeny, character evolution, and biogeography of *Cuscuta* (dodders; Convolvulaceae) inferred from coding plastid and nuclear sequences. Am J Bot. 101:670–690.

Graham SW, Olmstead RG. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am J Bot. 87:1712–1730.

Gualberto JM, et al. 2014. The plant mitochondrial genome: Dynamics and maintenance. Biochimie. 100:107–120.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9:605–618.
Knoop V. 2004. The mitochondrial DNA of land plants: Peculiarities in phylogenetic perspective. Curr Genet. 46:123–139.
Koulintchenko M, Konstantinov Y, Dietrich A. 2003. Plant mitochondria actively import DNA via the permeability transition pore complex. EMBO J. 22:1245–1254.
Kuijt J. 1969. The biology of parasitic flowering plants. Berkeley, CA: University of California Press.
Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Bio Evol. 34:772–773.
Lenz H, et al. 2010. Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. Curr Genet. 56:189–201.
Lenz H, Hein A, Knoop V. 2018. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. BMC Bioinform. 19:255.
Lenz H, Knoop V. 2013. PREPACT 2.0: Predicting C-to-U and U-to-C RNA editing in organelle genome sequences with multiple references and curated RNA editing annotation. Bioinform Biol Insights. 7:BBI.S11059.
Lin Q. 2020. Phylogenetics and evolution of monocot mycoheterotrophs and a newly demonstrated lineage of carnivorous monocots. PhD thesis. University of British Columbia.
Lin Y, et al. 2022. Unprecedented organelle genomic variations in morning glories reveal independent evolutionary scenarios of parasitic plants and the diversification of plant mitochondrial complexes. BMC Biol. 20:49.
Maddison WP, Maddison DR. 2018. Mesquite: A modular system for evolutionary analysis. Version 3 .51. Available from: http://www.mesquiteproject.org; Last accessed September 1, 2021.
McNeal JR, Arumugunathan K, Kuehl JV, Boore JL, Depamphilis CW. 2007. Systematics and plastid genome evolution of the cryptically photosynthetic parasitic plant genus *Cuscuta* (Convolvulaceae). BMC Biol. 5:55.
Meeuse AD, Welman WG. 2000. Flora of Southern Africa. Volume 28: Part 1: Convolvulaceae. Pretoria: Botanical Research Institute.
Molina J, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). Mol Bio Evol. 31:793–803.
Mower JP, et al. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biol. 8:150.
Mower JP. 2020. Variation in protein gene and intron content among land plant mitogenomes. Mitochondrion. 53: 203–213.
Mower JP, Jain K, Hepburn NJ. 2012. The role of horizontal transfer in shaping the plant mitochondrial genome. In: Maréchal-Drouard L, editor. Advances in Botanical Research. New York: Academic Press. p. 41–69.
Mower JP, Stefanović S, Young GJ, Palmer JD. 2004. Gene transfer from parasitic to host plants. Nature. 432:165–166.
Nickrent D. 2020. Parasitic angiosperms: How often and how many? Taxon. 69:5–27.
Park S, et al. 2015. Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol. 208:570–583.
Petersen G, Anderson B, Braun H-P, Meyer EH, Møllerd IM. 2020. Mitochondria in parasitic plants. Mitochondrion. 52:173–182.
Petersen G, Cuenca A, Møller IM, Seberg O. 2015. Massive gene loss in mistletoe (*Viscum*, Viscaceae) mitochondria. Sci Rep. 5: 17588.
Qiu Y, Filipenko SJ, Darracq A, Adams KL. 2014. Expression of a transferred nuclear gene in a mitochondrial genome. Curr Plant Biol. 1: 68–72.
Sanchez-Puerta MV, García LE, Wohlfeiler J, Ceriotti LF. 2017. Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. New Phytol. 214:376–387.
Sawyer S. 1989. Statistical tests for detecting gene conversion. Mol Biol Evol. 6:526–538.
Schuster W, Brennicke A. 1987. Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of Oenothera: Is genetic information transferred between organelles via RNA? EMBO J. 6:2857–2863.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51:492–508.
Shimodaira H, Hasegawa M. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.
Silvestro D, Michalak I. 2012. RaxmlGUI: a graphical front-end for RaxML. Org Divers Evol. 12:335–337.
Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. Proc Natl Acad Sci U S A. 112:E3515–E3524.
Skippington E, Barkman TJ, Rice DW, Palmer JD. 2017. Comparative mitogenomics indicates respiratory competence in parasitic *Viscum* despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size. BMC Plant Biol. 17:49.
Soltis PS, Soltis DE. 2003. Applying the bootstrap in phylogeny reconstruction. Stat Sci. 18:256–267.
Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Stefanović S, Kuzmina M, Costea M. 2007. Delimitation of major lineages within *Cuscuta* subgenus *Grammica* (dodders; Convolvulaceae) using plastid and nuclear DNA sequences. Am J Bot. 94:568–589.
Stefanović S, Olmstead RG. 2004. Testing the phylogenetic position of a parasitic plant (*Cuscuta*, Convolvulaceae, Asteridae): Bayesian inference and the parametric bootstrap on data drawn from three genomes. Syst Biol. 53:384–399.
Stern DB, Lonsdale DM. 1982. Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. Nature. 299:698–702.
Stern DB, Palmer JD. 1986. Tripartite mitochondrial genome of spinach: physical structure, mitochondrial gene mapping, and locations of transposed chloroplast DNA sequences. Nucleic Acids Res. 14:5651–5666.
van der Kooij TAW, Krause K, Dörr I, Krupinska K. 2000. Molecular, functional and ultrastructural characterisation of plastids from six species of the parasitic flowering plant genus *Cuscuta*. Planta. 210:701–707.
Vogel A, et al. 2018. Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. Nat Commun. 9: 2515.
Westwood JH, Yoder JI, Timko MP, dePamphilis CW. 2010. The evolution of parasitism in plants. Trends Plant Sci. 15:227–235.
Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76: 273–297.
Yang Z. 2006. Computational molecular evolution. Oxford, UK: Oxford University Press.
Yang Z. 2007. PAML4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Yang Z, et al. 2019. Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. Nat Plants. 5: 991–1001.
Yuncker TG. 1932. The genus *Cuscuta*. Memoirs of the Torrey Botanical Club. 18:113–331.
Yurina NP, Odintsova WS. 2016. Mitochondrial genome structure of photosynthetic eukaryotes. Biochemistry (Moscow). 81:101–113.
Zervas A, Petersen G, Seberg O. 2019. Mitochondrial genome evolution in parasitic plants. BMC Evol Biol. 19:87.
Zheng HZ, Dong ZH, She J. 1998. Modern study of traditional Chinese medicine. 1st ed. Beijing: Beijing Xue Yuan Press of the People's Republic of China. p. 4110–4120.

Associate editor: Daniel Sloan