New Phytologist

On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae

Thomas W. A. Braukmann[1]*, Michael B. Broe[2]*, Saša Stefanović[1]* and John V. Freudenstein[2]*

[1]Department of Biology, University of Toronto Mississauga, Mississauga, ON L5L 1C6, Canada; [2]Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH 43212-1157, USA

Author for correspondence:
Saša Stefanović
Tel: +1 905 569 4305
Email: sasa.stefanvoic@utoronto.ca

Received: *24 February 2017*
Accepted: *29 May 2017*

New Phytologist (2017) **216:** 254–266
doi: 10.1111/nph.14681

Key words: Ericaceae, gene loss mycoheterotrophs, plastome, relaxed selection.

Summary

• Ericaceae (the heather family) is a large and diverse group of plants that forms elaborate symbiotic relationships with mycorrhizal fungi, and includes several nonphotosynthetic lineages. Using an extensive sample of fully mycoheterotrophic (MH) species, we explored inter- and intraspecific variation as well as selective constraints acting on the plastomes of these unusual plants.

• The plastomes of seven MH genera were analysed in a phylogenetic context with two geographically disparate individuals sequenced for *Allotropa*, *Monotropa*, and *Pityopus*.

• The plastomes of nonphotosynthetic Ericaceae are highly reduced in size (*c*. 33–41 kbp) and content, having lost all photosynthesis-related genes, and are reduced to encoding housekeeping genes as well as a protease subunit (*clpP*)-like and acetyl-CoA carboxylase subunit D (*accD*)-like open reading frames. Despite an increase in the rate of their nucleotide substitutions, the remaining protein-coding genes are typically under purifying selection in full MHs. We also identified ribosomal proteins under relaxed or neutral selection. These plastomes also exhibit striking structural rearrangements.

• Intraspecific variation within MH Ericaceae ranges from a few differences (*Allotropa*) to extensive population divergences (*Monotropa, Hypopitys*), which indicates that cryptic speciation may be occurring in several lineages. The pattern of gene loss within fully MH Ericaceae plastomes suggests an advanced state of degradation.

Introduction

Photosynthesis is a hallmark of plant evolution as the vast majority of plants are autotrophic and rely entirely on this process to fix carbon. However, *c*. 1% of flowering plants have lost all or some of their photosynthetic ability and obtain their carbon budget from other plants. Such heterotrophic plants are broadly classified into two morphologically distinct groups: haustorial parasites and mycoheterotrophs (MHs). Haustorial parasites typically use a specialized organ, the haustorium, for transport of nutrients and/or water directly from their autotrophic hosts whereas fully MH plants epiparasitize mycorrhizal fungi (or rarely saprophytic fungi) that are themselves mutualistic with their ultimate autotrophic hosts. Plants that maintain some degree of photosynthetic ability and depend partially on their hosts for water and/or nutrition are known as partial heterotrophs. By contrast, full heterotrophs rely solely on their hosts to supply their water, fixed carbon, and nutrients. The transition from autotrophy to heterotrophy is associated with drastic changes in morphology (loss/reduction of vegetative organs; Leake, 1994; Manen *et al*., 2004), physiology (near loss of chlorophyll and high stomatal conductivity; Stewart & Press, 1990; Leake, 1994; Manen *et al*., 2004), and genome (rampant sequence divergence and loss of genes; McNeal *et al*., 2007a; Braukmann & Stefanović, 2012; Wicke *et al*., 2016; Graham *et al*., 2017).

In most photosynthetic land plants, plastid genomes (plastomes) are conserved in size, gene content, linear order of genes, and structure. They range from 140 to 160 kbp, containing *c*. 113 genes, and are typically partitioned into four regions: a large single copy (LSC), a small single copy (SSC), and two inverted repeats (IRs) (Ravi *et al*., 2008; Wicke *et al*., 2011). Plastome genes are usually subdivided into four functional classes: (1) photosynthetic apparatus genes (e.g. photosystem a (*psa*), photosystem b (*psb*), ATP synthase (*atp*), cytochrome b6/f (*pet*), large subunit of RuBisCO (*rbcL*) and NADH dehydrogenase (*ndh*)), (2) housekeeping genes (e.g. small ribosomal proteins (*rps*), large ribosomal proteins (*rpl*), RNA polymerase (*rpo*), plastid tRNA (*trn*) and translation initiation factor IF-1 (*infA*)), (3) genes with other functions (e.g. a protease subunit (*clpP*), acetyl-CoA carboxylase subunit D (*accD*), and an intron maturase (*matK*)), and (4) open reading frames (ORFs), some with as yet unknown or unconfirmed function (*ycf* genes) (Ravi *et al*., 2008; Wicke *et al*., 2011). In land plants, one notable

*These authors contributed equally to this work.

characteristic of plastomes is their relatively low rate of molecular evolution compared with the nuclear genome (e.g. Wolfe *et al.*, 1992; Drouin *et al.*, 2008). Heterotrophic plants represent an extreme deviation from this canonical norm, with a wide range of evolutionary degradation that allows us to dissect the function of plastid genes outside photosynthesis, study patterns of gene loss, and elucidate the role of the plastid within nonphotosynthetic plants (Neuhaus & Emes, 2000). Plastome degradation in heterotrophs is predominantly based upon two well-studied groups of haustorial parasites: the dodders (*Cuscuta*; McNeal *et al.*, 2007a,b; Braukmann *et al.*, 2013) and the broomrape family (Orobanchaceae; Wicke *et al.*, 2013, 2016). Among MH taxa, only the plastomes of monocots are well studied, in particular Orchidaceae (Graham *et al.*, 2017). The degradation of the plastome is marked by three intertwined processes relating to: (1) the formation of pseudogenes and their eventual loss (e.g. Wicke *et al.*, 2016; Graham *et al.*, 2017); (2) structural changes to the plastome (e.g. Wicke *et al.*, 2016; Graham *et al.*, 2017); and (3) increased rates of nucleotide substitution (McNeal *et al.*, 2007a; Lam *et al.*, 2015; Wicke *et al.*, 2016).

Broad-scale molecular phylogenetic analyses estimate a minimum of nine flowering plant families where MH has originated, some with several independent origins of full or partial MH (Bidartondo, 2005; Merckx & Freudenstein, 2010). A comparative analysis of plastomes across the trophic spectrum (from autotrophy to partial MH to full MH) from different lineages can provide valuable insights into the pattern and tempo of evolutionary loss, as well as the function of the genes that persist through the transition from autotrophy to full heterotrophy. To date, there have been very few in-depth molecular analyses of the plastomes of fully MH eudicot plastomes. A well-known group of eudicot MH is found within the heather family, Ericaceae. This family is a large and diverse group of flowering plants, with a nearly world-wide distribution (Kron *et al.*, 2002). They are known to form elaborate relationships with fungi (Kron *et al.*, 2002; van der Heijden *et al.*, 2015) and contain a number of partial MHs and fully MH species. Ericaceae s.l. is composed of nine subfamilies, of which seven are exclusively autotrophic. Fully MH species are confined to subfamily Monotropoideae (except *Pyrola aphylla* in Pyroloideae), and all partially MH taxa are circumscribed within Pyroleae (summarized by Kron *et al.*, 2002). Recent evidence shows that Monotropoideae form a monophyletic group, and that the Pyroloideae represent a second independent origin of MH in Ericaceae (Braukmann & Stefanović, 2012; Freudenstein *et al.*, 2016; Lallemand *et al.*, 2016). Previous hybridization-based approaches by Braukmann & Stefanović (2012) characterized the drastic reduction in protein-coding genes in the plastomes of fully MH Ericaceae. Four recently published complete plastome sequences of fully MH species confirmed the hybridization findings. Specifically, accessions of *Hypopitys* from the Russian Federation (Gruzdev *et al.*, 2016; Ravin *et al.*, 2016) exhibited a size of 35 kbp and reduction to a small subset of 21 protein-coding genes and Logacheva *et al.* (2016) reported similar results for *Hypopitys* and *Monotropa*.

To further explore interspecific and intraspecific variation within fully MH Ericaceae, we sequenced the plastomes of seven representative fully MH Ericaceae using a next-generation sequencing approach. The specific objectives of this study were: (1) to evaluate plastome size and structure for fully MH Ericaceae (including intraspecific variation); (2) to assess gene content, identify genes of potentially essential function, and characterize the pattern of gene loss; (3) to conduct tests of relative rates of molecular evolution of MH Ericaceae with their closest autotrophic relatives; (4) to determine if retained genes exhibit signatures of relaxed or positive selection; and (5) to conduct an in-depth comparison of plastome gene content between MH Ericaceae and other heterotrophic lineages to shed light on their unique or common patterns of evolution.

Materials and Methods

Taxon sampling and sequencing approach

We collected tissue from seven fully MH species, *Allotropa virgata* Torr. & A. Gray, *Hemitomes congestum* A. Gray, *Hypopitys monotropa* (DC) Nutt., *Monotropa uniflora* L., *Monotropsis odorata* Schwein. ex Elliot, *Pityopus californicus* (Eastw.) Copeland, and *Pterospora andromedea* Nutt. To assess intraspecific variation, we sampled two populations of *Allotropa virgata* (states of Washington and California), *M. uniflora* (Michigan and Florida), and *Pityopus californicus* (Oregon and California).

Our seven of 13 fully MH Monotropoideae species represent a comprehensive sampling of nonphotosynthetic Ericaceae. Tissue was extracted and enriched for organellar DNA using the differential centrifugation protocol described by Palmer (1992). In brief, 50–100 g of fresh tissue were homogenized in a Waring blender (Waring, New Hartford, CT, USA) and filtered through four layers of cheesecloth. The filtrate was run through a single layer of Miracloth™ (EMD Millipore, Billerica, MA, USA) and centrifuged at 500 ***g*** for 10 min to pellet nuclei and cell debris. The supernatant was centrifuged at 2500 ***g*** for 15 min, yielding a plastid-enriched pellet. DNA was extracted using a modified hexa-decyltrimethylammonium bromide (CTAB) technique (Doyle & Doyle, 1987). Isolated DNA from *Allotropa virgata* (Washington sample), *M. uniflora* (Michigan sample), and *Pterospora andromedea* was further purified by centrifugation using a CsCl/ethidium bromide equilibrium density gradient. Library construction allowed the multiplexing of the DNAs from all six species onto a single lane of the Illumina HiSeq 2000/2500 platform (Genome Quebec, Montreal, Canada) or MiSeq platform (Molecular and Cellular Imaging Center, The Ohio State University, OH, USA). Reads were comprised of 100- or 300-bp length paired ends, respectively, with insert sizes between 400 and 500 bp.

Sequence assembly, finishing, and annotation

Two approaches were used to assemble plastomes based on platform and read length. For the 100-bp paired-end reads (Illumina HiSeq), we trimmed and filtered the different libraries using the program SICKLE v.1.29 (Joshi & Fass, 2011) at both ends with a

minimum PHRED quality score of 25 and a minimum length of 70 bp. Reads containing ambiguous base pairs and singletons were excluded. *De novo* assembly was performed with VELVET v.1.2.08 (Zerbino & Birney, 2008) and RAY v.2.1.0 (Boisvert *et al.*, 2012). Runs were optimized for hash length (71–79), minimum coverage (5), coverage cut-off (500), expected cut-off (50–100), and minimum contig length (500 bp). To minimize spurious chimeric assemblies and bias the assembly towards organellar contigs, additional parameters modified in VELVET (Zerbina & Birney, 2008) were maximum branch length (100), maximum divergence (0.1), and maximum gap count (1). For species with incomplete plastome assemblies, gaps between scaffolds were closed with the iterative read mapper IMAGE (Tsai *et al.*, 2010). IMAGE runs were carried out with hash lengths between 65 and 75. Scaffolds were merged only if overlaps between them were > 200 bp. To verify and improve assemblies, reads were aligned to draft contigs using BWA v.0.5.9 (Li & Durbin, 2010) and then reassembled using the SPADES genome assembler v.3.1.1 (Bankevich *et al.*, 2012). Post-assembly, plastid contigs were identified using BLAST and then assembled *de novo* using the GENEIOUS assembler v.6.1.7 (Kearse *et al.*, 2012) under high stringency. The 300-bp paired-end read sets (Illumina MiSeq) were assembled using the DISCOVAR *de novo* assembler (Broad Institute, Cambridge, MA, USA) which is optimized for long-read, paired-end Illumina data. Reads were mapped back to the DISCOVAR contigs in GENEIOUS (Kearse *et al.*, 2012) and re-assembled with the GENEIOUS (Kearse *et al.*, 2012) *de novo* assembler under medium sensitivity to validate the assembly. Mapped reads were also assessed for even coverage and agreement. For each accession, five independent runs were performed, and each run generated complete plastomes in full agreement with one another, with the exception of *Hemitomes*, where a consensus was assembled.

Annotation of the protein-coding and rRNA (*rrn*) genes was performed in DOGMA (Wyman *et al.*, 2004) with an identity cut-off of 40% for protein coding genes, and tRNAs were identified using tRNA SCAN (Lowe & Eddy, 1997). To verify annotations, extracted genes from *Arbutus unedo* (JQ067650) and *Vaccinium macrocarpon* (JQ757046) were mapped against the draft plastomes. Additionally, regions without annotations were manually extracted and searched against the BLAST nonredundant database to check that no gene annotations were excluded by the other two approaches. These three approaches provide a robust method for identifying divergent plastid genes, thereby avoiding potential false negatives. Potential open reading frames were identified using the annotation tool in GENEIOUS v.6.1.7 (Kearse *et al.*, 2012). All complete plastome sequences were deposited in GenBank (accession nos. MF120261–MF120270).

Comparative genomic analyses

Actinidia deliciosa is the closest autotrophic relative to MH Ericaceae whose plastome sequence is currently available (NC_026691) and possesses a canonical gene content and order that are typical for most angiosperms (Yao *et al.*, 2015). Because a complete circular genome was not constructed for *Pterospora*, there is the potential for a few genes to be present in the remaining gap(s). However, these areas are probably composed of AT-rich repeats and/or long homopolymer runs, and therefore the chance of false negatives is low even for these two samples. Taxa used for comparisons are a mix of photosynthetic and nonphotosynthetic species. We included photosynthetic relatives *V. macrocarpon* (Ericaceae; NC_019616) and *Arbutus unedo* (Ericaceae; JQ067650). We also compared our results to the sequenced plastomes of heterotrophic plants. These comparisons allow us to observe states before and after loss of photosynthesis to identify common patterns of gene retention across a range of nonphotosynthetic taxa. We also included the plastid gene sequences reported by Gruzdev *et al.* (2016), Logacheva *et al.* (2016), and Ravin *et al.* (2016) to evaluate intra-specific variation between North American and European *Hypopitys*. To examine changes in linear gene order in the plastid genomes of MH Ericaceae, we visualized changes to gene order using progressive Mauve in GENEIOUS v.6.1.7 (Darling *et al.*, 2004; Kearse *et al.*, 2012).

Gene-by-gene alignments were performed using the default alignment algorithm in GENEIOUS v.6.1.7 (Kearse *et al.*, 2012) and were subsequently adjusted by eye. *Arbutus unedo*, the closest autotrophic relative to fully MH Ericaceae with a sequenced plastome (but without a canonical gene content and order), was used to estimate synonymous (d_S) and nonsynonymous (d_N) nucleotide substitution rates for each gene by pairwise comparison using the Yang & Nielson (2000) method via the YN00 program in the PAML package (Yang, 2007). To avoid sampling bias of heterotrophic plants and provide a more robust comparison, we used *Camellia sinensis*, *A. deliciosa*, *V. macrocarpon*, and *A. unedo* as photosynthetic relatives in these tests. To exclude bias introduced by excessive sampling of *Hypopitys*, we included only two accessions, our North American *Hypopitys* and a European *Hypopitys* (KU640958). An unrooted topology based on a combined data set of plastid protein-coding genes was used in the analysis. The unrooted topology was estimated using maximum likelihood under a general time reversible model in PHYML (Guindon *et al.*, 2010). We excluded *clpP* because the extreme sequence divergence (and poor alignment) would lead to potentially spurious and misleading results.

To test if selection is acting on the remaining plastid genes in MH Ericaceae, we used three models (SLAC, FEL and REL) in HYPHY (Pond & Frost, 2005a; Pond *et al.*, 2005) using the Datamonkey webserver (Pond & Frost, 2005b) for each gene. All three models are site specific and categorize the ratio of d_N to d_S (d_N/d_S) as being under purifying ($d_N/d_S < 1$), neutral ($d_N/d_S = 1$), or positive ($d_N/d_S > 1$) selection in the context of a codon substitution model. The single likelihood ancestor counting method (SLAC) is a modified Suzuki−Gojorbi counting-based approach whereas the less conservative fixed effects likelihood model (FEL) fits an independent d_N and d_S to every site. Both approaches are more conservative (and less powerful) than the random effects likelihood model (REL), which is susceptible to type I errors. Sites were considered to be under selection if at least two of these approaches were significant (Pond & Frost, 2005a; Pond *et al.*, 2005). To account for multiple testing, we used Bonferroni error

correction analyses (corrected alphas: $P = 0.0024$ for REL and FEL; adjusted $P = 0.0048$ for SLAC).

We also tested for relaxed selection in MH Ericaceae using RELAX (Wertheim *et al.*, 2015) in the HYPHY package (Pond *et al.*, 2005). RELAX calculates a selection intensity parameter (k) and tests if selection is relaxed or intensified in a subset of branches (i.e. MH) relative to reference branches (i.e. autotrophs) in a tree. A likelihood ratio test determines if one rejects or accepts intensification or relaxation of selection. All these analyses were executed on the Datamonkey webserver (Pond & Frost, 2005b).

Results

Diversity, size, and content of plastomes

We recovered complete plastomes for the MH Ericaceae with the exception of one *Monotropa* sample (100-bp assembly; one gap) and *Pterospora* (two gaps). The plastomes of fully MH Ericaceae are all significantly reduced in size compared to typical plastomes, and range from 33 557 bp in *Hypopitys* to 41 360 bp in *Pterospora* (Fig. 1; Table 1). The two plastomes of *Allotropa* contain the same gene complement and are nearly identical in both size (35 983 bp vs 35 944 bp) and gene order. The two plastome copies of *Pityopus* also encode the same set of plastid genes but are substantially different in size (Oregon sample, 38 188 bp vs California sample, 35 926); all the differences are located in intergenic spacers. We observed a similar gene order for both *Monotropa* plastomes but substantial sequence divergence between the two samples; similar sequence differences were seen in both *Hypopitys* sequences (Fig. 2). The consistent recovery of *Allotropa* and *Monotropa* plastomes from 100-bp and 300-bp reads confirms that the short read assemblies (100 bp) were not spurious (Supporting Information Fig. S1). Although *Pterospora* was only assembled using 100-bp reads, it has a similar gene order to other fully MH species, suggesting that it is not the product of misassembly.

There are common patterns of structural changes and gene loss among the different MH species. All photosystem (*psa* and *psb*), *ndh*, *pet*, *atp* and *rbcL* genes are absent (except a *psbA* pseudogene in *Pterospora*) and represent a single loss of photosynthesis for MH Ericaceae. Additional losses include plastid hypothetical ORFs (*ycf* genes), *ycf1* and *ycf2*. Both *matK* and *infA* are present as ORFs in the plastomes of MH Ericaceae (Fig. 3) with the exception of a putative *matK* pseudogene in *Monotropsis*. The putative *matK* pseudogene is caused by a single nucleotide deletion 255 bp from the 3′ end of the gene (coverage 51 X; 100% read identity) that leads to a premature stop codon 231 bp from the end of the gene.

For all but one species, MH Ericaceae have lost a copy of the IR region. *Monotropa* does contain a small IR (partial *accD-clpP-rrn16–rps14*–partial *rps4* and its inverse) with a small 798-bp single copy region containing *trnF-GAA*. However, all of these IR genes except *rrn16* are located in the large single copy region in *Arbutus* and *Camellia*. This same gene region is also involved in structural changes in *Hemitomes* and *Monotropsis*. In *Monotropa*, *rrn16* is moved from its canonical position adjacent to *rrn23* and is located between the *clpP*-like ORF and *rps14* (and is then presumably repeated and inverted). In *Hemitomes*, *rps14* is moved from its canonical position adjacent to *rps4* and is located between *rrn16* and *rrn23*. The gene order differs for the two *Hypopitys* plastomes, with the region spanning *rps14* to *rpl14* located adjacent to *rrn16* in the Russian sample (Fig. 4). *Monotropsis* is notable for the presence of two direct repeats of 1002 bp. Each repeat contains a stretch of 491 identical bases with the remaining 511 bp exhibiting a small number of single nucleotide polymorphisms (SNPs) and short indels. One repeat is located between *rps4* and *rps14*, the other between *rrn16* and *rrn23*.

The *accD* gene is involved in fatty acid synthesis in the cell and is presumed to be a plastid-encoded gene with a function in heterotrophic plastomes. However, *accD* exists as a highly divergent ORF in all MH Ericaceae with little sequence conservation between species, which corroborates the previous observations in MH Ericaceae (Gruzdev *et al.*, 2016; Logacheva *et al.*, 2016). These long *accD*-like ORFs are typically between 1700 and 2000 bp (compared with 1542 bp in *Camellia*) and have only relatively short segments (*c.* 300–500 bp) at the 3′ end that can be recognized as plastid *accD*. The remaining portions of the *accD* ORFs yield no similarity hits when compared to the databases via BLAST (at both nuclear and amino acid levels). In *Monotropa* there are two long *accD* fragments (*c.* 4200 bp and *c.* 2100 bp) extending beyond the end of the IR where agreement between the two copies is disrupted by the presence of tandem repeats. The shorter fragment contains little sequence similarity with the other *accD* fragments. *Monotropsis* possesses a truncated version of the *accD*-like ORF that is only 1164 bp in length. Nucleotide alignment of the *accD* fragments of MH indicates a high level of sequence divergence among species (only 850 of 2495 sites are identical, excluding one fragment from *Monotropa*) and within species (pairwise identity 94.5% in *Allotropa*, 89.3% in *Pityopus*, and 88% in *Hypopitys*). Amino acid sequence divergence is also rather high within species (pairwise similarity 94.3% in *Allotropa*, 88.8% in *Hypopitys*, and 85.8% in *Pityopus*; Table S1). The substantial divergence within and between species suggests that these *accD*-like ORFs are either under positive selection or evolving neutrally (see Discussion below).

Similar to *accD*, we also observed a highly divergent ORF for *clpP* in MH Ericaceae. *clpP* encodes a protease subunit involved in ATP-dependent degradation of photosynthetic complexes. The *clpP*-like ORF retains little sequence identity between species, with only 140 of the 816 being identical but with high intra-specific sequence similarity (100% in *Allotropa*, 96.2% in *Monotropa*, and 91.9% in *Pityopus*; Table S1). The *clpP*-like ORF is disrupted by a stop codon in the Russian *Hypopitys* accession which implies a nonfunctional status. Most intriguingly, we also identified an additional large ORF unique to nonphotosynthetic Ericaceae between *clpP* and *accD*. This ORF is between 900 and 1100 bp with no significant similarity to any sequence in GenBank, but maintains a relatively high level of inter-specific sequence similarity with 422 identical sites (of 928) when the

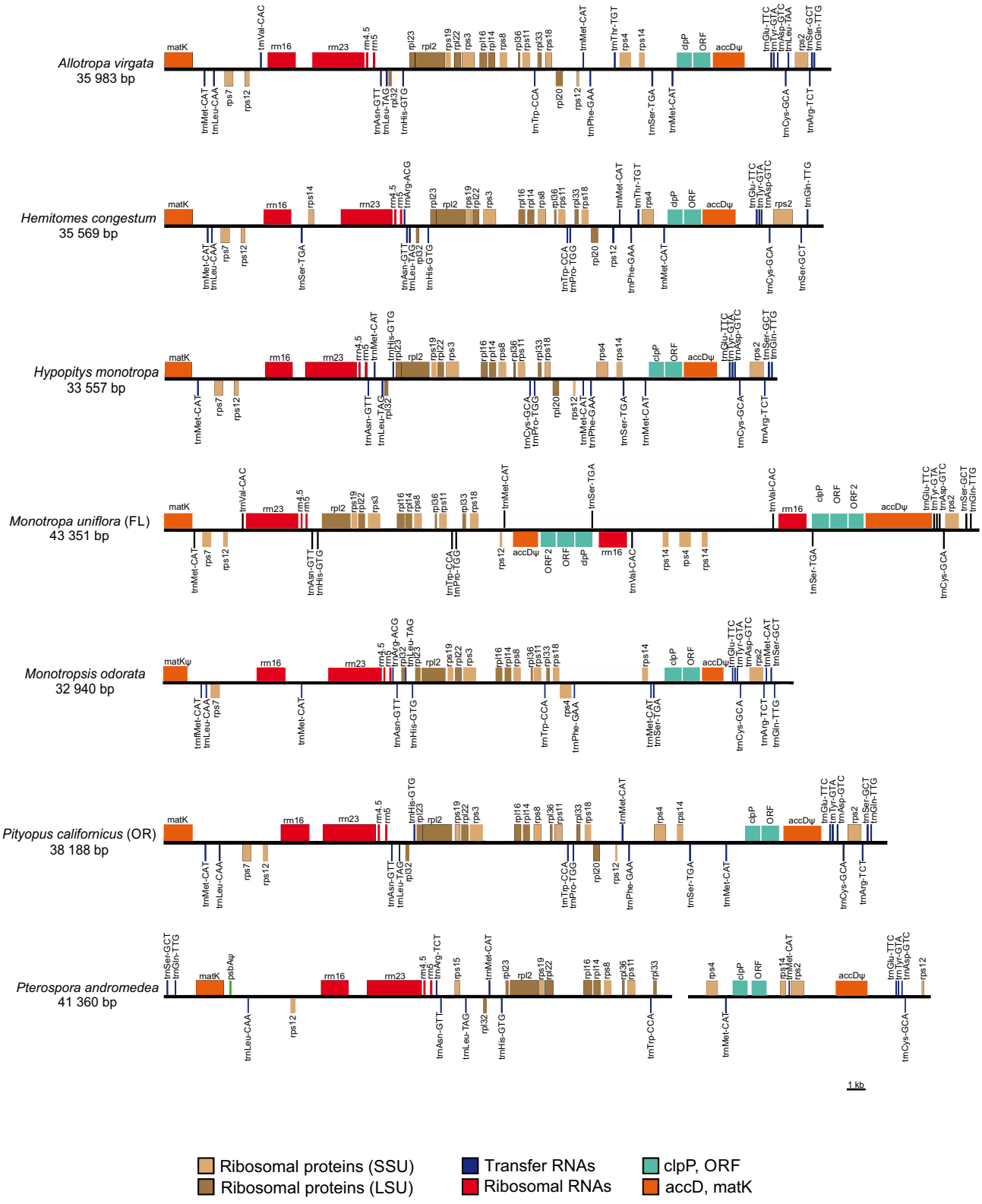

Fig. 1 Linear map of mycoheterotrophic (MH) Ericaceae. Present are plastid tRNA (*trn*) genes (purple), small ribosomal proteins (*rps*) genes (light brown), large ribosomal proteins (*rpl*) genes (brown), *rrn* genes (red), a protease subunit (*clpP*) (turquoise), open reading frames (ORFs) (turquoise), an intron maturase (*matK*) (orange), and acetyl-CoA carboxylase subunit D (*accD*) (orange). Maps for plastomes were generated in OGDRAW v.1.2 (Lohse *et al.*, 2007).

highly divergent *Pterospora* is excluded (otherwise 451 identical in 1293 sites). For this ORF, we also observed high within-species identity (pairwise identity of 99.9% in *Allotropa*, 97.1% in *Pityopus*, 96.6% in *Monotropa*, and 89.5% in *Hypopitys*; Table S1). There is an additional ORF in the same region unique to *M. uniflora* that is 795 bp in length in the Florida accession

Table 1 Summary of plastid genomes recovered in this study with length, and contigs recovered

Species	Length (bp)	Fragments recovered
Allotropa virgata^{WA}	35 983	Full circle
Allotropa virgata^{CA}	35 944	Full circle
Hemitomes congestum	35 569	Full circle
Hypopitys monotropa^{OH}	33 557	Full circle
Monotropa uniflora^{MI}*	45 520	1 linear contig
Monotropa uniflora^{FL}	43 531	Full circle
Monotropsis odorata	32 922	1 linear contig
Pityopus californicus^{CA}	35 926	Full circle
Pityopus californicus^{OR}	38 188	Full circle
Pterospora andromedea	27 365 and 12 955	2 contigs

Abbreviation for locations of mycoheterotrophic (MH) species with multiple representatives are: CA, California; FL, Florida; MI, Michigan; OH, Ohio; OR, Oregon; WA, Washington.
*Estimated plastome size.

and 1584 bp in the Michigan sample. It is unclear if these ORFs are spurious and more evidence is needed to determine the status of these large ORFs.

Allotropa exhibits minimal variation in other coding regions between populations (99.8–100%) but these regions are less conserved in *Monotropa* (92.6–99.7%), *Pityopus* (95.4–98.9%), and *Hypopitys* (86.5–97.3%) (Table S1). Aside from *clpP* and *accD*, the least conserved gene varies for each species (Table S1) but we observed low pairwise similarity for *matK* (87.7%), *rps2* (88.3%), and *rps19* (86.5%) in *Hypopitys*.

Both small (*rps*) and large (*rpl*) ribosomal subunit proteins, along with transfer RNA (*trn*) genes, compose the bulk of the remaining gene content for all three fully MH Ericaceae (Fig. 4). Losses tend to be idiosyncratic (i.e. *rpl20*), with a small subset of genes being lost in all MH Ericaceae sequenced (*rps16*, *trnA-UGC*, *trnG-GCC*, *trnI-GAU*, *trnK-UUU*, *trnT-GGU*, and *trnV-UAC*). Otherwise, *Monotropa* exhibits the greatest number of losses (18 genes; see Fig. 3 for details), with the majority of losses occurring among tRNA genes. Despite the extensive loss of plastid genes, 17 protein coding genes and 10 tRNA genes are shared among all MH Ericaceae. The plastomes of other nonphotosynthetic plants also predominantly retain functional copies of *rps*, *rpl*, and *trn* genes. Among fully heterotrophic plants, several *rps* (*2*, *4*, *7*, *8*, *11*, *18* and *19*), *rpl* (*2*, *14*, *16* and *36*), and *trn* (*E* and *fM*) genes are ubiquitously present.

When this study was in review, the plastomes of *Pyrola rotundifolia*, *M. uniflora*, and another *Hypopitys* (as *Monotropa hypopitys*) were published by Logacheva *et al.* (2016) and their data largely corroborate our findings, except that we observed an additional ORF in all MH Ericaceae and a second ORF in *M. uniflora*.

Comparative genome analysis

The protein-coding genes that remain in fully MH Ericaceae are under purifying selection, with d_N/d_S (ω) ratios well below 0.5 for most plastid genes (Fig. S2; Table S2). Among fully MH Ericaceae, *Monotropa* and *Monotropsis* exhibit the most elevated rates of nucleotide substitution for most plastid genes (Figs S2, S3; Table S2). This is reflected in the long branches leading to a clade composed of *Monotropa* and *Monotropsis* (Fig. 2, S3). If anything, autotrophic *Camellia* has a higher ω than MH Ericaceae (Fig. S2) for several genes (*infA*, *rps7*, *rps8* and *rps14*) relative to *Arbutus*, although the d_N and d_S are substantially lower (Fig. S3). There is almost no sequence variation among the plastid genes of the two *Allotropa* plastomes but there is substantially more variation between the two *Pityopus*, *Monotropa*, *and Hypopitys* samples. However, the intraspecific variation for all species is less than the interspecific variation (see Fig. 2; Table S2).

The d_N/d_S predicts that only a few genes are under positive or relaxed selection (except *rpl33*; the short sequence length of *rpl33* makes it susceptible to artifacts and results for this locus should be interpreted with caution). This is well supported by the number of negatively selected sites ($d_S > d_N$) across the remaining protein coding genes especially *matK*, *rpl2*, *rps2*, *rps3*, and *rps8* (Table S3). Only the REL analysis detected sites under positive selection (Table S3) but this approach is susceptible to type I errors (false positives; Pond and Frost, 2005a). RELAX detected significant relaxation of selection ($P < 0.05$) among MH for numerous ribosomal proteins (*rpl33*, *rps3*, *rps4*, *rps7*, *rps8* and *rps18*; Table 2) as well as the remaining *accD* ORF. We observed intensification of selection acting on a number of loci including *infA*, *rps2*, *rps14*, and *rpl20* ($P < 0.05$; Table 2).

Discussion

Plastome size, content, and organization

Fully MH Ericaceae have highly reduced plastomes in both their size and coding content. The extent of plastid gene losses in fully MH Ericaceae is similar to that in the highly reduced plastomes of *Sciaphila densiflora* (*c.* 19 kbp) and *Epipogium roseum* (*c.* 18 kbp), where the majority of the remaining protein-coding genes are also ribosomal proteins (Lam *et al.*, 2015; Schelkunov *et al.*, 2015). The plastid genomes of the fully MH Ericaceae are in the final stages of plastid reduction proposed by Barrett *et al.* (2014), Wicke *et al.* (2016), and Graham *et al.* (2017), who hypothesize that plastid gene loss in heterotrophic plants happens sequentially, starting with the NAD(P)H complex (ndh genes), followed by photosynthetic genes (*rbcL*, *psa*, *psb*, *pet*, *ycf3*, *cemA*, *ccsA*, and *ycf4* genes), plastid-encoded RNA polymerase (*rpo* genes), and ATP synthase (*atp* genes). The last genes to be lost include the ribosomal protein genes (*rpl* and *rps*), ribosomal RNA genes (*rrn*), plastid tRNAs (*trn* genes), an initiation factor (*infA*), an intron maturase (*matK*), acetyl CoA-carboxylase (*accD*), a protease subunit involved in ATP-dependent degradation of photosynthetic complexes (*clpP*), and two genes involved in a number of cellular process (*ycf1* and *ycf2*). Plastome degradation in heterotrophic plants is associated with the elevation of nucleotide substitutions, genomic rearrangements, a decrease in GC content, and a weakening of selection (Wicke *et al.*, 2016).

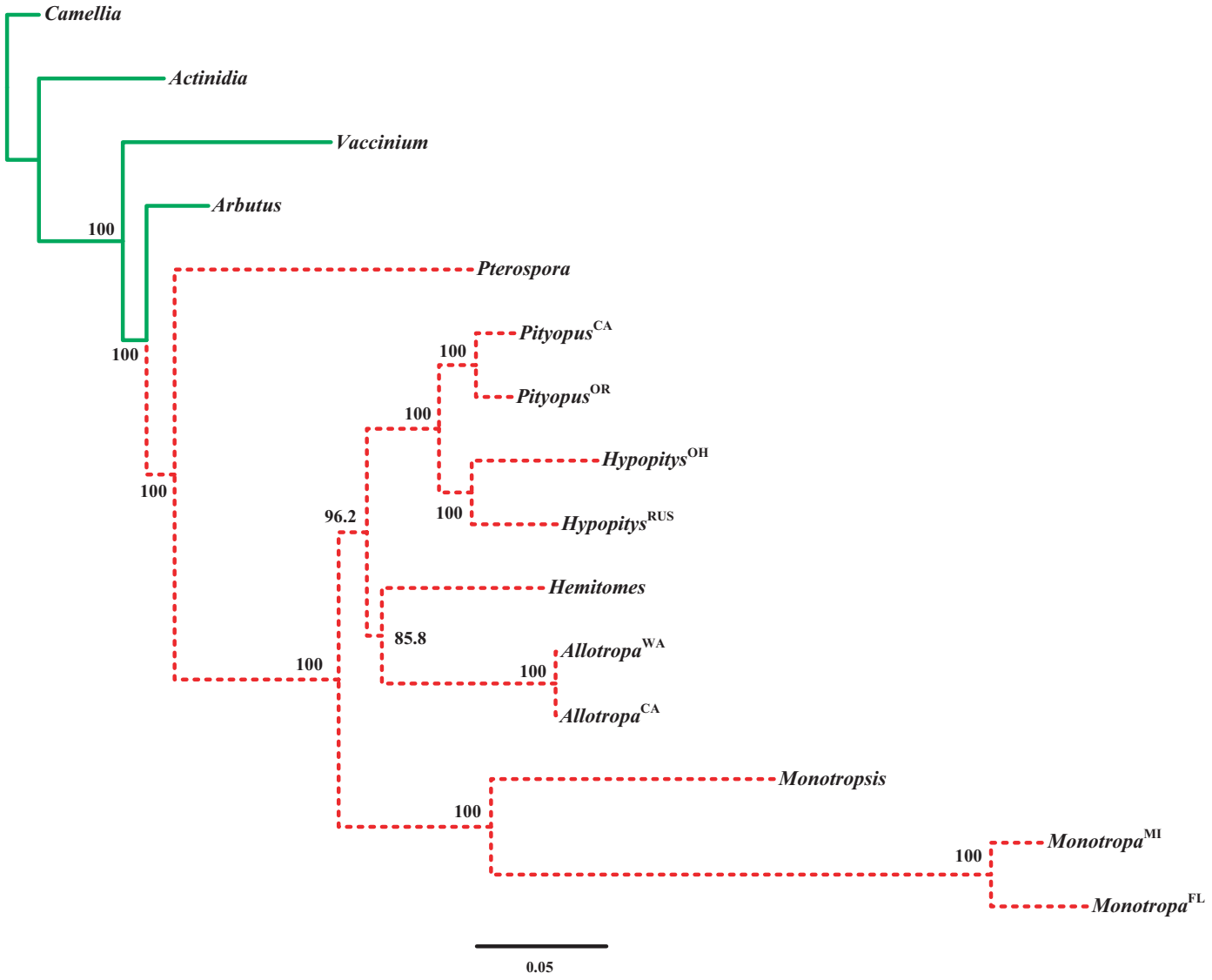

Fig. 2 Phylogenetic relationships of mycoheterotrophic (MH) Ericaceae generated in this study and selected photosynthetic Ericales based on their complete plastome sequences. The tree was generated in PHYML under a GTR+G model and bootstrap proportion for 500 bootstrap replicates is indicated at each node. Branch lengths are proportional to the inferred amount of substitutions. The GTR+G model was estimated in JMODELTEST 2 (Darriba *et al.*, 2012). Green, photosynthetic taxa; red, MH taxa. Abbreviations for locations of MH species with multiple representatives are: CA, California; FL, Florida; MI, Michigan; OH, Ohio; OR, Oregon; WA, Washington; RUS, Russia.

A striking finding is the loss of both *ycf1* and *ycf2* in fully MH Ericaceae. *ycf1* is thought to function as a translocator of the inner envelope of chloroplasts, whereas *ycf2* is linked to a wide variety of gene functions including cell division, proteolysis, and protein transport (Wicke *et al.*, 2011 and references therein). The near ubiquitous presence of *ycf1* and *ycf2* in full heterotrophs indicates that these genes have essential functions beyond the maintenance of photosynthesis in plastids. The loss of *ycf1* and *ycf2* probably preceded the transition to MH, having been lost from the plastomes of their autotrophic relatives (or rendered pseudogenes) (Braukmann & Stefanović, 2012; Fajardo *et al.*, 2013; Martinez-Alberola *et al.*, 2013). Typically, plastomes contain seven group IIA introns, all of which are spliced with the assistance of a plastid intron maturase (*matK*; Vogel *et al.*, 1997). *matK* is nearly ubiquitous in angiosperm plastomes, with the exception of *Cuscuta* subgenus *Grammica* and fully MH monocots, such as *Rhizanthella gardneri*, *Epipogium*, and *Neottia* (McNeal *et al.*, 2009; Delannoy *et al.*, 2011; Logacheva *et al.*, 2011; Schelkunov *et al.*, 2015; Feng *et al.*, 2016). We did observe a truncated version of *matK* in *Monotropsis* despite the presence of genes with group IIa introns (*rpl2* and *rps23*). The truncated *matK* in *Monotropsis* might function as a less efficient form or be a pseudogene that is preceding further loss of plastid-encoded ribosomal proteins (*rpl2* and *rps23*).

Another gene commonly found in the plastomes of heterotrophs is *clpP*, which encodes a protease subunit involved in ATP-dependent degradation of photosynthetic complexes (Peltier *et al.*, 2001; Lam *et al.*, 2016). *clpP* is found as a

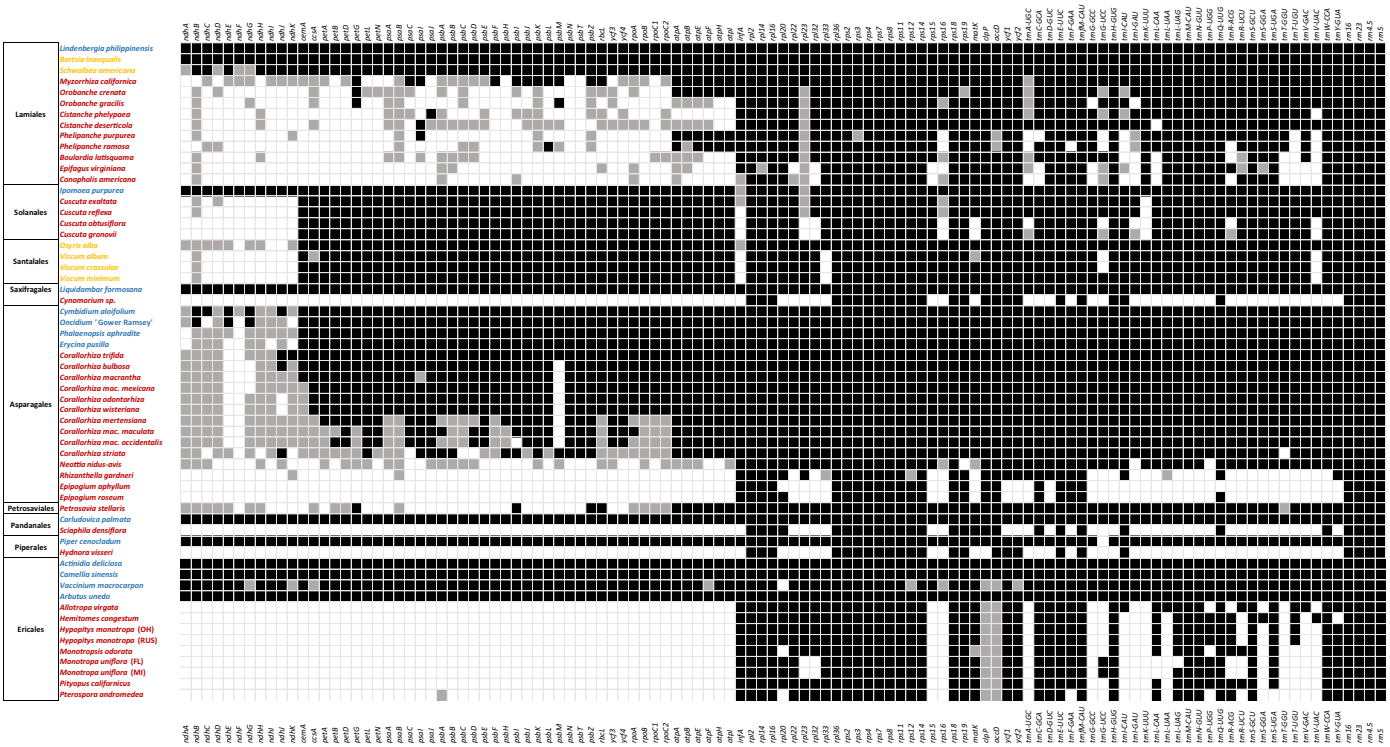

Fig. 3 Heat map of the presence or absence of plastid genes in mycoheterotrophic (MH) Ericaceae (species names in red), their photosynthetic relatives (in blue) as well as other nonphotosynthetic plants (also in red) following the paper by Naumann *et al.* (2016). Yellow, partially heterotrophic plants; black, presence of a gene; grey, presence of a pseudogene; white, absence of the gene from the plastome.

pseudogene in *Arbutus unedo*, the closest autotrophic relative to MH Ericaceae (Martinez-Alberola *et al.*, 2013; Freudenstein *et al.*, 2016; Lallemand *et al.*, 2016), which suggests that a *clpP* pseudogene is a pleisomorphic state. However, *clpP* exists as a highly divergent ORF in most MH Ericaceae. The lone exception is the disruption of the *clpP*-like ORF in a *Hypopitys* accession (Gruzdev *et al.*, 2016), which suggest that these divergent *clpP*-like ORFs are pseudogenes.

Perhaps the most intriguing protein-coding gene residing in heterotrophic plastomes is the one encoding a subunit of acetyl CoA-carboxylase (*accD*), a protein essential to the production of fatty acids in the cell (Bungard, 2004; Sasaki & Nagano, 2004; Wicke *et al.*, 2011). The involvement of *accD* in nonphotosynthetic pathways explains its frequent occurrence in heterotrophic plastomes and its loss or pseudogenization is rare (*Phelipanche*: Wicke *et al.*, 2013; *Cynomorium*: Bellot *et al.*, 2016). *accD* persists as a highly divergent ORF in a few members of holoparasitic groups (Wicke *et al.*, 2013; Lam *et al.*, 2016) and extreme divergence of *accD* is linked to the loss of *ycf1* from the plastome (Kikuchi *et al.*, 2013; de Vries *et al.*, 2015). An analogous, but even more extreme situation is observed in the fully MH Ericaceae where there are highly divergent *accD*-like ORFs with only a small portion retaining any sequence similarity, at either nucleotide or amino acid level. Furthermore, the *accD* gene has been reduced to very short pseudogenes (*c.* 200–300 bp) in photosynthetic Ericaceae (Fajardo *et al.*, 2013; Martinez-Alberola *et al.*, 2013), which indicates a single (functional) loss of this gene within the family. Recently, Logacheva *et al.* (2016) found evidence for expression of *accD* in *Hypopitys*; it is part of a polycistronic RNA together with *clpP* (unpublished results, 1KP project; Logacheva *et al.*, 2016), which is located next to *accD* in this species. We found that *clpP* and *accD* shared this arrangement in all our samples. We also found evidence of another ORF between *clpP* and *accD*, which is highly conserved and accurately tracks the phylogeny. In *Monotropa*, there is an additional ORF located before *accD* that is highly divergent between populations. Given the lack of sequence conservation at both the nucleotide and amino acid levels along with the number of unusual ORFs in this region, and evidence of relaxed selection among fully MH Ericaceae, it is more probable that these elements are pseudogenes. The alternative hypothesis is that *clpP* and *accD* ORFs are maintained across fully MH Ericaceae because these genes are functional or potentially undergoing neofunctionalization. Additionally, the strong purifying selection acting on ribosomal genes supports the continued presence of protein-coding genes without housekeeping function (i.e. *accD* and *clpP*). Further efforts are needed to determine if these genes are expressed and produce functional proteins. This investigation of fully MH Ericaceae plastomes has generated a number of testable hypotheses with regard to gene function and intraspecific variation in these enigmatic plants, and highlights the need for downstream studies on both transcription and translation.

The remaining ribosomal protein genes (*rps* and *rpl*) might have little to no function within fully MH Ericaceae if all other protein-coding genes are nonfunctional. If the *accD*-like ORF is functional, this would potentially maintain the translation

Fig. 4 Collinearity of circularized plastomes of mycoheterotrophic (MH) Ericaceae species. In this linearization, small ribosomal proteins 2 (*rps2*) is at the extreme right, and an intron maturase (*matK*) near the extreme left. Syntenic blocks are indicated by differently coloured boxes with lines tracking the location of these regions in the different plastomes. Gaps indicate regions absent in a plastome of a given species. The red bars in the *Hypopitys* (Russia) track represent *rrn5*, *4.5*, *23* and *16*, located at the end of the blue block. Green Blocks represent tRNA genes and clear blocks represent protein coding genes.

machinery in the plastome, but we did observe the relaxation of selection on several of the small ribosomal proteins. The number of ribosomal proteins under relaxed selection suggests that the remaining plastid genes are in the early stages of becoming pseudogenes. The intensification of selection in a few ribosomal proteins could represent historical signal from earlier losses of ribosomal proteins, or a continued role for a small set of ribosomal proteins in the plastomes (i.e. expressing *accD* and *clpP*). The 'essential tRNA hypothesis' (Barbrook *et al.*, 2006) posits that the end-point of plastome evolution in nonphotosynthetic plants would be mini-circles, encoding the plastid tRNA for glutamic acid (*trnE*). Plastid *trnE* is essential for mitochondrial and chloroplast haeme synthesis, and cytosolic *trnE* cannot functionally participate in this pathway (Barbrook *et al.*, 2006). The isolation of haeme synthesis in the plastid could lead to *trnE* being trapped in a mini-circle that would be incredibly difficult to detect with high-throughput methods and *in situ* techniques (Howe & Smith, 1991; Barbrook *et al.*, 2006; Wicke *et al.*, 2016). To date, only Rafflesiaceae and *Cuscuta* subg. *Grammica* section *Subulatae* (the 'O' clade) are suggested to have cryptic plastomes (Nickrent *et al.*, 1997; Braukmann *et al.*, 2013; Molina *et al.*, 2014). Interestingly, two recently sequenced plastomes from the endoparasite *Pilostyles* (Apondanthaceae) are reduced to a small subset of five or six functional plastid genes that do not include *trnE* but still possess *accD* (Bellot & Renner, 2015), suggesting that plastid genomes can be continuously reduced until all essential genes are transferred to the nucleus or mitochondria (Naumann *et al.*, 2016; Wicke *et al.*, 2016). Any functional mini-circle with plastid-based transcription and translation would require nuclear encoded transcriptional and translation machinery (Graham *et al.*, 2017). Given their reduction to a mere set of housekeeping genes, we hypothesize that the plastomes of fully MH Ericaceae represent the final stages of full MH evolution in eudicots, at the brink before becoming mini-circles where only an essential tRNA, *trnE*, *accD*, and *clpP* retain functions vital to the survival of the cell.

Concurrent with the loss of plastid genes is the reduction in plastome size. The plastomes of fully MH Ericaceae are highly reduced in size (between 33 and 45 kbp). Fully MH Ericaceae plastomes have lost the IR and therefore have only single copies of genes normally found in this region. The loss of the IR (except in *Monotropa*) is revealed by the consistent read depth across all

Table 2 Outcomes of the RELAX (Wertheim *et al.*, 2015) analysis for detecting relaxation or intensification of selection in mycoheterotrophic (MH) Ericaceae vs autotrophic relatives

Gene	Selection	Relaxation coefficient (*k*)	*P*	Likelihood ratio (LR)	ω^{ref}	ω^{MH}	AICc null	AICc alternative
infA	Intensification	2.6	0.0058	7.61	0.606	0.295	3070.89	3065.42
rpl20	Intensification	11.64	0.0133	6.12	0.91	0.353	3623.92	3619.92
rpl33	**Relaxation**	**0.18**	**0.0012**	**10.52**	**0.629**	**0.454**	**2401.91**	**2393.59**
rps2	**Intensification**	**2.86**	**0.0008**	**11.16**	**0.652**	**0.399**	**8479.42**	**8470.3**
rps3	**Relaxation**	**0.14**	**< 0.0001**	**21.09**	**0.721**	**0.3**	**7595.09**	**7576.05**
rps4	Relaxation	0.19	0.0033	8.63	0.543	0.281	6279.12	6272.56
rps7	**Relaxation**	**0.15**	**< 0.0001**	**24.3**	**0.864**	**0.32**	**4558.5**	**4536.23**
rps8	Relaxation	0.22	0.0254	5	0.484	0.264	4407.61	4404.71
rps14	Intensification	4.35	0.0208	5.34	0.729	0.313	2678.72	2678.93
rps18	Relaxation	0.65	0.0268	4.91	0.744	0.616	5677.4	5674.57
accD	**Relaxation**	**0.55**	**0**	**170.53**	**1.39**	**0.615**	**33721.1**	**33552.59**

Tests that are highly significant are indicated in bold. Indicated is the direction of selection, relaxation coefficient, *P* value, likelihood ratio (LR), ω (dN/dS) for the reference (ref) and test (MH) branches, and Akaike's information corrected criterion (AICc) values; *accD*, acetyl-CoA carboxylase subunit D; *infA*, translation initiation factor IF-1, *rpl*, large ribosomal proteins; *rps*, small ribosomal proteins.

the recovered plastid contigs in our study and the observation that genes found within the IR are contiguous with genes of the large single copy region. Typically, the IR is thought to stabilize the plastome through homologous recombination and induced repair mechanisms (Wicke *et al.*, 2011). The loss of the IR is associated with plastomes that are highly reduced in both size and coding content (Wicke *et al.*, 2013; Lam *et al.*, 2015; Schelkunov *et al.*, 2015), and is linked to mutational hotspots and overall increase in nucleotide substitution rates in the plastomes (Wicke *et al.*, 2011 and references therein). It is difficult to assess the impact that elimination of an IR region has on the plastome of nonphotosynthetic plants, which already have elevated rates of nucleotide substitution independent of the presence or absence of the IR. Furthermore, the increase in micro-structural mutations associated with higher mutation rates (e.g. homopolymer runs, palindromic repeats, and an increase in AT richness) is thought to promote illegitimate recombination (Wicke *et al.*, 2013; Petersen *et al.*, 2015).

The plastomes of fully MH Ericaceae lack collinearity with autotrophic plastomes. In *Hemitomes* and *Monotropa* there is a breakdown in the highly conserved gene order of plastid ribosomal RNA genes (Fig. 3). The exact impact that the breaking up of rRNA genes has on plastid functions is unclear and future studies are needed to understand plastid gene expression in these plants. The small IR in *Monotropa* is probably a novel IR that re-established recently because it is more likely that a new duplicate region arose from rearrangements rather than multiple independent IR losses. There is also evidence that a similar secondary repeat exists in other Ericaceae (*Moneses uniflora*; T. W. A. Braukmann & S. Stefanović, unpublished data).

Multiple representatives from several populations allowed us to assess intraspecific variation for *Allotropa*, *Monotropa*, *Hypopitys* and *Pityopus*. For *Monotropa*, *Hypopitys*, and *Pityopus* all genes vary between individuals, while *Allotropa* shows no variation in 14 of the remaining 21 protein-coding genes (see Table S1). Variation in plastome length (> 2 kb) was observed for *Monotropa*, *Hypopitys* and *Pityopus*. The frequent movement of *rrn16* or *rps14* in European *Hypopitys*, *Monotropa*, and *Hemitomes* indicates that this region containing these genes is highly labile. This AT-rich region has many short direct repeats in the North American *Hypopitys* that may promote recombination. We observed three gene orders in *Hypopitys* (Fig. S4), with the North American accession maintaining the plesiomorphic pattern of *Pityopus* and *Allotropa*. It is unclear if these plastome variants in *Hypopitys* represent strong drift or cryptic speciation between North American and European populations as well as within European populations (Broe, 2014). The intraspecific variation in fully MH Ericaceae demonstrates that a single plastome sequence is not representative of the species. Substantial intraspecific variation was also observed within *Epigogium roseum* (Schelkunov *et al.*, 2015), and appears to be relatively common in nonphotosynthetic plants.

Molecular evolution

Rates of nucleotide substitution are increased in all MH Ericaceae protein-coding genes, with *Monotropsis* and *Monotropa* having the highest rates (Figs 2, S2, S3; Table S2). The acceleration of nucleotide substitution is common for heterotrophic plants (Young & de Pamphilis, 2005; McNeal *et al.*, 2007a; Wickett *et al.*, 2008; Barrett & Davis, 2012). Although we see an increase in nucleotide substitution rates and the relaxation of selection on several *rps* genes, the majority of the remaining ribosomal protein genes residing in MH taxa are still under purifying selection. This is evidenced by the d_N/d_S well below 0.5 and the presence of negatively selected sites in the remaining plastid genes. We also did not observe an accumulation of insertions or deletions (indels) that would be expected under relaxed selection. Although there tend to be more indels in plastid genes relative to their autotrophic relatives, we found that several *rps* genes are evolving neutrally. This finding, combined with the rearrangement of the rRNA genes in several species, suggests that the plastid transcriptional machinery is being eroded by either loss of plastid function or its replacement with nuclear-encoded ribosomal proteins.

Relaxation of selection on nuclear loci responsible for DNA repair and plastome structure could result in changes to gene order and elevated nucleotide substitution rates despite the remaining plastid genes being under strong purifying selection (Table S1). The observed increased rates of nucleotide substitution in heterotrophic plastomes could be a result of relaxation of selection on nuclear loci rather than the relaxation of selection on remaining plastid genes. Acceleration of nucleotide substitution in heterotrophic plants could be governed by mutator genes (Marechal & Brisson, 2010). For example, in *Oenothera*, mutants of the nuclear locus *plastome mutator* (*PM*) have increased rates of indels and point mutations as a result of either enzyme slippage or recombination events. Mutants of the *PM* locus often result in variegated plants as a consequence of changes to the plastid genome (Chiu *et al.*, 1990; Chang *et al.*, 1996; Stoike & Sears, 1998; Marechal & Brisson, 2010). *PM* mutants are strongly selected against in photosynthetic plants because of disruption of photosynthesis. *PM* mutants, which are normally purged from a population, would persist in nonphotosynthetic plants as a result of a relaxation of selection on maintaining a functional photosynthetic apparatus. Other nuclear loci that could contribute to increased rates of plastome mutation are *recA* (a recombinase) homologs, which are associated with DNA repair in organellar genomes (Rowan *et al.*, 2010; Guisinger *et al.*, 2011).

With the relaxation of selection on plastome repair mechanisms, the remaining plastid genes would become highly divergent and nearly unrecognizable when compared with plastid homologs found in autotrophic relatives. This effect, combined with reduced size of the plastid genome over long evolutionary time scales, can lead to the development of cryptic plastid genomes, as suggested with parasitic Rafflesiaceae, where no plastid genome has yet been detected (Nickrent *et al.*, 1997; Molina *et al.*, 2014).

Acknowledgements

We would like to thank the anonymous reviewers whose feedback greatly improved the manuscript. We acknowledge the contribution of staff of the McGill University and Génome Québec Innovation Centre, Montréal, Canada for sequencing services. Financial support from the Natural Sciences and Engineering Research Council of Canada (grant no. 326439), the Canada Foundation for Innovation (grant no. 12810), and the Ontario Research Funds to S.S., and from the US National Science Foundation (grant no. DEB-0842076) to J.V.F., is gratefully acknowledged. We also thank the Natural Sciences and Engineering Research Council of Canada for the scholarship award provided to T.W.A.B.

Author contributions

T.W.A.B., M.B.B., S.S. and J.V.F. planned and designed the research. T.W.A.B. and M.B.B. were responsible for data analysis. T.W.A.B., M.B.B., S.S. and J.V.F. were responsible for interpretation of data and writing the manuscript.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

Barbrook AC, Howe CJ, Purton S. 2006. Why are plastid genomes retained in non-photosynthetic organisms? *Trends in Plant Science* **11**: 101–108.

Barrett CF, Davis JI. 2012. The plastid genome of mycoheterotrophic *Corallorhiza striata* Orchidaceae is in the relatively early stages of degradation. *American Journal of Botany* **99**: 1513–1523.

Barrett CF, Freudenstein JV, Li J, Mayfield-Jones DR, Perez L, Pires JC, Santos C. 2014. Investigating the path of plastid genome degradation in an early transition clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* **31**: 3095–3112.

Bellot S, Cusimano N, Luo S, Sun G, Zarre S, Gröger A, Temsch E, Renner SS. 2016. Assembled plastid and mitochondrial genomes, as well as nuclear genes, place the parasite family Cynomoriaceae in the Saxifragales. *Genome Biology and Evolution* **8**: 2214–2230.

Bellot S, Renner SS. 2015. The plastomes of two species of endoparasite genus *Pilostyles* (Apoodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution* **8**: 189–201.

Bidartondo MI. 2005. The evolutionary ecology of myco-heterotrophy. *New Phytologist* **167**: 335–352.

Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. 2012. Ray Meta: scalable *de novo* metagenome assembly and profiling. *Genome Biology* **13**: R122.

Braukmann TWA, Kuzmina M, Stefanović S. 2013. Plastid genome evolution across the genus *Cuscuta* Convolvulaceae.: two clades within subgenus *Grammica* exhibit extensive gene loss. *Journal of Experimental Botany* **64**: 977–989.

Braukmann TWA, Stefanović S. 2012. Plastid genome evolution in mycoheterotrophic Ericaceae. *Plant Molecular Biology* **79**: 5–20.

Broe MB. 2014. *Phylogenetics of the Monotropoideae (Ericaceae) with special focus on the Genus Hypopitys Hill, together with a novel approach to phylogenetic inference using lattice theory.* PhD thesis, Ohio State University, Columbus, OH, USA.

Bungard RA. 2004. Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. *BioEssays* **26**: 235–247.

Chang TL, Stoike LL, Zarka D, Schewe G, Chiu WL, Jarrell DC, Sears BB. 1996. Characterization of primary lesions caused by the plastome mutator of *Oenothera*. *Current Genetics* **30**: 522–530.

Chiu WL, Johnson EM, Kaplan SA, Blasko K, Sokalski MB, Wolfson R, Sears BB. 1990. *Oenothera* chloroplast DNA polymorphisms associated with plastome mutator activity. *Molecular Genomics and Genetics* **221**: 59–64.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* **14**: 1394–1403.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.

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

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

Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution* **49**: 827–831.

Fajardo D, Senalik D, Mercedes A, Zhu H, Steffan SA, Harbut R, Polashock J, Vorsa N, Gillespie E, Kron K *et al.* 2013. Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genetics and Genomes* **9**: 489–498.

Feng Y-L, Wicke S, Li J-W, Han Y, Lin C-S, Li D-Z, Zhou T-T, Huang W-C, Huang L-Q, Jin X-H. 2016. Lineage specific reductions of plastid genomes in

an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biology and Evolution* **8**: 2164–2175.

Freudenstein JV, Broe MB, Feldenkris ER. 2016. Phylogenetic relationships at the base of Ericaceae: implications for vegetative and mycorrhizal evolution. *Taxon* **65**: 794–804.

Graham SWG, Lam VKY, Merckx VSFT. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotrophic plastid genomes. *New Phytologist* **214**: 48–55.

Gruzdev EV, Mardanov AV, Beletsky AV, Kochieva EZ, Ravin NV, Skryabin KG. 2016. The complete chloroplast genome of parastic flowering plant *Monotropa hypopitys*: extensive gene losses and size reduction. *Mitochondrial DNA Part B: Resources* **1**: 212–213.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307–321.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Gerianaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution* **28**: 583–600.

van der Heijden MG, Martin FM, Selosse MA, Sanders IR. 2015. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytologist* **205**: 1406–1423.

Howe CJ, Smith AG. 1991. Plants without chlorophyll. *Nature* **349**: 109.

Joshi NA, Fass FN. 2011. *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files version 1.29*. [WWW document] URL https://github.com/najoshi/sickle

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Kikuchi S, Bedard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **339**: 571–574.

Kron KA, Judd WS, Stevens PF, Crayn DM, Anderberg AA, Gadek PA, Quinn CJ, Luteyn JL. 2002. Phylogenetic classification of Ericaceae: molecular and morphological evidence. *Botanical Review* **68**: 35–423.

Lallemand F, Gaudeul M, Lambourdiére J, Matsuda Y, Hashimoto Y, Selosse MA. 2016. The elusive predisposition to mycoheterotrophy in Ericaceae. *New Phytologist* **212**: 314–319.

Lam VKY, Merckx VSFT, Graham SW. 2016. A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *American Journal of Botany* **103**: 692–708.

Lam VKY, Soto Gomez M, Graham SW. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* Triuridaceae is colinear with its green relatives and is under strong purifying selection. *Genome Biology and Evolution* **7**: 2220–2236.

Leake JR. 1994. The biology of mycoheterotrophic 'saprophytic' plants. *New Phytologist* **127**: 171–216.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–595.

Logacheva MD, Schelkunov MI, Penin AA. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biology and Evolution* **3**: 1296–1303.

Logacheva MD, Schelkunov MI, Shratnikova VY, Matveeva Penin AA. 2016. Comparative analysis of plastid genomes of non-photosynthetic Ericaceae and their photsynthetica relatives. *Scientific Reports* **6**: 30042.

Lohse M, Dreschel O, Kahlau S, Bock R. 2007. Organellar Genome DRAW OGDRAW – a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* **52**: 267–274.

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

Manen JF, Habashi C, Jeanmonod D, Park J, Schneeweiss GM. 2004. Phylogeny and intraspecific variability of holoparasites *Orobanche* Orobanchaceae inferred from plastid *rbcL* sequences. *Molecular Phylogenetics and Evolution* **33**: 482–500.

Marechal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* **186**: 299–317.

Martinez-Alberola F, del Campo EM, Lazaro-Gimeno D, Mezquita-Claramonte S, Molins A, Mateu-Andre I, Pedrola-Monfort J, Casano LM, Barreno E. 2013. Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. *PLoS ONE* **8**: e79685.

McNeal JR, Arumugunathan K, Kuehl JV, Boore JL, de Pamphilis CW. 2007b. Systematics and plastid genome evolution of the cryptically photosynthetic parasitic plant genus *Cuscuta* Convolvulaceae. *BMC Biology* **5**: 55.

McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. 2007a. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology* **7**: 57.

McNeal JR, Kuehl JV, Boore JL, Leebens-Mack J, de Pamphilis CW. 2009. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS ONE* **4**: e59825.

Merckx V, Freudenstein JV. 2010. Evolution of mycoheterotrophy in plants: a phylogenetic perspective. *New Phytologist* **185**: 605–609.

Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, Flowers JM, Pelser P, Barcelona J, Inovejas SA *et al.* 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* Rafflesiaceae. *Molecular Biology and Evolution* **31**: 793–803.

Naumann J, Der JP, Wafula EK, Jones SS, Wagner ST, Honaas LA, Ralph PE, Bolin JF, Maass E, Neinhuis C *et al.* 2016. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biology and Evolution* **8**: 345–363.

Neuhaus HE, Emes MJ. 2000. Nonphotosynthetic metabolism in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **51**: 111–140.

Nickrent DL, Ouyang Y, Duff RJ, de Pamphilis CW. 1997. Do nonasterid holoparasitic flowering plants have plastid genomes? *Plant Molecular Biology* **34**: 731–743.

Palmer JD. 1992. Organelle DNA isolation and RFLP analysis. In: Beckmann JS, Osborn TC, eds. *Plant genomes: methods for genetic and physical mapping*. Boston, MA, USA: Kluwer, 35–53.

Peltier JB, Ytterberg J, Liberles DA, Roepstorff P, van Wijk KJ. 2001. Identification of a 350 kDa ClpP and protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. *Journal of Biological Chemistry* **276**: 16318–16327.

Petersen G, Cuenca A, Seberg O. 2015. Plastome evolution in hemiparasitic Mistletoes. *Genome Biology and Evolution* **7**: 2520–2532.

Pond SL, Frost SDW. 2005a. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**: 1208–1222.

Pond SL, Frost SDW. 2005b. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**: 2531–2533.

Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679.

Ravi V, Khurana JP, Tyagi AK, Khurana P. 2008. An update on chloroplast genomes. *Plant Systematics and Evolution* **271**: 101–122.

Ravin NV, Gruzdev EV, Beletsky AV, Mazur AM, Prokhortchouk EB, Filyushin MA, Kochieva EZ, Kadnikov VV, Mardanov AV, Skryabin KG. 2016. The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. *BMC Plant Biology* **16**: 153–161.

Rowan BA, Oldenbrug D, Bendich AJ. 2010. *RecA* maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. *Journal of Experimental Botany* **61**: 2575–2588.

Sasaki Y, Nagano Y. 2004. Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Bioscience, Biotechnology, and Biochemistry* **68**: 1175–1184.

Schelkunov MI, Shtratnikova VY, Nuraliev MS, Selosse MA, Penin AA, Logacheva MD. 2015. Exploring the limits for reduction of plastid genomes: a case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution* **7**: 1179–1191.

Stewart GR, Press MC. 1990. The physiology and biochemistry of parasitic angiosperms. *Annual Review of Plant Physiology and Plant Molecular Biology* **41**: 127–151.

Stoike LL, Sears BB. 1998. Plastome mutator-induced alterations arise in *Oenothera* chloroplast DNA through template slippage. *Genetics* **149**: 347–353.

Tsai IJ, Otto TD, Berriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biology* **11**: R41.

Vogel J, Hübschmann T, Börner T, Hess WR. 1997. Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: support for *matK* as an essential splice factor. *Journal of Molecular Biology* **270**: 179–187.

de Vries J, Sousa FL, Bolter B, Soll J, Gould SB. 2015. YCF1: a green TIC? *Plant Cell* **27**: 1827–1833.

Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution* **32**: 820–832.

Wicke S, Müller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Schneeweiss GM. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the Broomrape family. *Plant Cell* **25**: 3711–3725.

Wicke S, Müller KF, dePamphilis CW, Quandt D, Bellot S, Schneeweiss GM. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences, USA* **113**: 9045–9050.

Wicke S, Schneeweiss GM, de Pamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76**: 273–297.

Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Plock SA, Wolf PG, de Pamphilis CW, Boore JL, Goffinet B. 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Molecular Biology and Evolution* **25**: 393–401.

Wolfe KH, Morden CW, Ems SC, Palmer JD. 1992. Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *Journal of Molecular Evolution* **35**: 304–317.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2017**: 3252–3255.

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yang X, Nielson R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**: 32–43.

Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. 2015. The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. *PLoS ONE* **4**: e0129347.

Young ND, de Pamphilis CW. 2005. Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. *BMC Evolutionary Biology* **5**: 16.

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

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Read coverage maps for 100- and 300-bp assemblies for *Allotropa virgata*, *Monotropa uniflora*, *Monotropsis odorata*, and *Pterospora andromedea*. Reads were mapped in GENEIOUS or using BOWTIE2.

Fig. S2 Pairwise d_N/d_S, d_N, and d_S vs *Arbutus unedo* for protein-coding genes found in fully mycoheterotrophic Ericaceae.

Fig. S3 Phylograms of d_N and d_S trees from likelihood analyses using a constrained unrooted topology. Included are plastid genes *rpl2*, *rps2*, *infA*, and *matK*.

Fig. S4 Phylogeny of European (KU640958, KU640957 and KU878156) and North American *Hypopitys monotropa* based on plastid *rrn16*, *rrn23* and *rps2*.

Table S1 Pairwise nucleotide and amino acid similarities for *Allotropa*, *Hypopitys*, *Monotropa*, and *Pityopus* (all pairwise similarities were calculated in GENEIOUS)

Table S2 Detailed comparisons of d_N/d_S, d_N, and d_S vs *Arbutus unedo* from the PAML yn00 program

Table S3 Summary of HyPhy analysis (Pond *et al.*, 2005) under REL, FEL, and SLAC models

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.