On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae

Thomas W. A. Braukmann1*, Michael B. Broe2*, Saša Stefanović1* and John V. Freudenstein2*

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

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 mycoheterotropic (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 (orf) genes (Ravi et al., 2008; Wicke et al., 2011). In land plants, one notable

*These authors contributed equally to this work.

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

Author for correspondence:
Saša Stefanović
Tel: +1 905 569 4305
Email: sasa.stefanovic@utoronto.ca
Received: 24 February 2017
Accepted: 29 May 2017

doi: 10.1111/nph.14681

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust
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 plants (Neuhaus & Emes, 2000). Plastome degradation in plastid genes outside photosynthesis, study patterns of gene loss, evolutionary degradation that allows us to dissect the function of extreme deviation from this canonical norm, with a wide range 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 Pyroloideae (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 monoptera (DC) Nutt., Monotropa uniflora L., Monotropis 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* (Zerbino & 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 *Discover de novo* assembler (Broad Institute, Cambridge, MA, USA) which is optimized for long-read, paired-end Illumina data. Reads were mapped back to the *Discover* 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 *Hemithamn*, where a consensus was assembled.

Annotation of the protein-coding and tRNA (*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 delicosa* 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 Gruzd 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 (*dS*) and nonsynonymous (*dN*) nucleotide substitution rates for each gene by pairwise comparison using the Yang & Nielsen (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. delicosa*, *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 *HvPhyl* (Pond & Frost, 2005a; Pond et al., 2005) using the Data-monkey webserver (Pond & Frost, 2005b) for each gene. All three models are site specific and categorize the ratio of *dN* to *dS* (*dN/dS*) as being under purifying (*dN/dS* < 1), neutral (*dN/dS* = 1), or positive (*dN/dS* > 1) selection in the context of a codon substitution model. The single likelihood ancestor counting method (SLAC) is a modified Suzuki–Gojobori counting-based approach whereas the less conservative fixed effects likelihood model (FEL) fits an independent *dN* and *dS* 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 (\(\theta\)) 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 putative 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 Monotropis. 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 \(accD-clpP-rnn16-rps14\)–partial \(rps4\) and its inverse) with a small 798-bp single copy region containing \(rnnF-GAA\). However, all of these IR genes except \(rnn16\) 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 Monotropis. In Monotropa, \(rnn16\) is moved from its canonical position adjacent to \(rnn23\) 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 \(rnn16\) and \(rnn23\). The gene order differs for the two Hypopitys plastomes, with the region spanning \(rps14\) to \(rpl14\) located adjacent to \(rnn16\) in the Russian sample (Fig. 4). Monotropis 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 \(rnn16\) and \(rnn23\).

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. Monotropis 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
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.
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 *rpm19* (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, trnL-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 *F*) 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 *dN/dS* (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 *dN* and *dS* 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 *dN/dS* 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 (*dS > dN*) across the remaining protein coding genes especially *matK, rpl2, rps2, rps3,* and *rpm8* (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, rpm8* and *rpm18*; Table 2) as well as the remaining *accD ORF*. We observed intensification of selection acting on a number of loci including *infA, rpm2, rpm14,* and *rpm20* (*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 *Scaphia 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, ccmA, 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 (*rrn* 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).

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

<table>
<thead>
<tr>
<th>Species</th>
<th>Length (bp)</th>
<th>Fragments recovered</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Allotropa virgata</em>&lt;sup&gt;WA&lt;/sup&gt;</td>
<td>35 983</td>
<td>Full circle</td>
</tr>
<tr>
<td><em>Allotropa virgata</em>&lt;sup&gt;CA&lt;/sup&gt;</td>
<td>35 944</td>
<td>Full circle</td>
</tr>
<tr>
<td><em>Hemitomes congestum</em></td>
<td>35 569</td>
<td>Full circle</td>
</tr>
<tr>
<td><em>Hypopitys monotropa</em>&lt;sup&gt;OH&lt;/sup&gt;</td>
<td>33 557</td>
<td>Full circle</td>
</tr>
<tr>
<td><em>Monotropa uniflora</em>&lt;sup&gt;WA&lt;/sup&gt;</td>
<td>45 520</td>
<td>1 linear contig</td>
</tr>
<tr>
<td><em>Monotropa uniflora</em>&lt;sup&gt;FL&lt;/sup&gt;</td>
<td>43 531</td>
<td>Full circle</td>
</tr>
<tr>
<td><em>Monotropa odorata</em></td>
<td>32 922</td>
<td>1 linear contig</td>
</tr>
<tr>
<td><em>Pityopus californicus</em>&lt;sup&gt;CA&lt;/sup&gt;</td>
<td>35 926</td>
<td>Full circle</td>
</tr>
<tr>
<td><em>Pityopus californicus</em>&lt;sup&gt;OR&lt;/sup&gt;</td>
<td>38 188</td>
<td>Full circle</td>
</tr>
<tr>
<td><em>Pterospora andromedea</em></td>
<td>27 365 and 12 955</td>
<td>2 contigs</td>
</tr>
</tbody>
</table>

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

Another gene commonly found in the plastomes of heterotrophs is \textit{clpP}, which encodes a protease subunit involved in ATP-dependent degradation of photosynthetic complexes (Peltier \textit{et al.}, 2001; Lam \textit{et al.}, 2016). \textit{clpP} is found as a
pseudogene in *Arbutus unedo*, the closest autotrophic relative to MH Ericaceae (Martinez-Alberola et al., 2013; Freudenstein et al., 2016; Lallemant et al., 2016), which suggests that a *clpP* pseudogene is a pleomorphic 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 (*chrome* 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
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 (Aponanthaceae) 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

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 rml5, 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.
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

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

Tests that are highly significant are indicated in bold. Indicated is the direction of selection, relaxation coefficient, P value, likelihood ratio (LR), \(\omega\) (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.

Molecular evolution

Rates of nucleotide substitution are increased in all MH Ericaceae protein-coding genes, with Monotropis 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 \(d_{S}/d_{S}\) 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.

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 plasmomorphic 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.
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 organelar 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 (Chiu et al., 2000; Chang et al., 2008; 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 organelar genomes (Rowan et al., 2010; Guisinger et al., 2011).

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


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.


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


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.


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.


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, Monotropopsis uniflora, Monotropopsis odorata, and Pterospora andromedea. Reads were mapped in GENEIOUS or using BOWTIE2.

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

Fig. S3 Phylograms of dN and dS 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 monophora based on plastid rrn16, rrn23 and rps2.

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

Table S2 Detailed comparisons of dS/dN, dS, and dS 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.