The Plastomes of Two Species in the Endoparasite Genus *Pilostyles* (Apodanthaceae) Each Retain Just Five or Six Possibly Functional Genes

Sidonie Bellot[1,*] and Susanne S. Renner[1]

[1]Department of Biology, Ludwig Maximilian University, Munich, Germany

*Corresponding author: E-mail: sido.bellot@gmail.com.

Accepted: December 6, 2015

Data deposition: The plastomes of *Pilostyles aethiopica* and *Pilostyles hamiltonii* have been deposited at GenBank under the accessions KT981955 and KT981956.

Abstract

The 23 species of mycoheterotrophic or exoparasitic land plants (from 15 genera and 6 families) studied so far all retain a minimal set of 17 of the normally 116 plastome genes. Only *Rafflesia lagascae*, an endoparasite concealed in its host except when flowering, has been reported as perhaps lacking a plastome, although it still possesses plastid-like compartments. We analyzed two other endoparasites, the African Apodanthaceae *Pilostyles aethiopica* and the Australian *Pilostyles hamiltonii*, both living inside Fabaceae. Illumina and 454 data and Sanger resequencing yielded circularized plastomes of 11,348 and 15,167 bp length, with both species containing five possibly functional genes (*acc*D, *rps*3, *rps*4, *rrn*16, *rrn*23) and two/three pseudogenes (*rpo*C2 in *P. aethiopica* and *rpl*2 and *rps*12 in both species; *rps*12 may be functional in *P. hamiltonii*). Previously known smallest land plant plastomes contain 27–29 genes, making these Apodanthaceae plastomes the most reduced in size and gene content. A similar extent of divergence might have caused the plastome of *Rafflesia* to escape detection. The higher plastome degeneration in both these families of endoparasites, Rafflesiaceae and Apodanthaceae, of similar high age, compared with exoparasites points to a difference of plastome function between those two modes of parasitic life.

Key words: endoparasite, chloroplast genome, photosynthesis, gene loss, minimal plastome.

Introduction

The loss of photosynthesis has occurred in parasitic lineages ranging from algae to angiosperms, but is not necessarily accompanied by a loss of plastid genomes (reviewed in Krause 2012). Which plastid genes may persist in nonphotosynthetic land plants has been studied in 24 parasitic or mycoheterotrophic *Cuscuta*, Orobanchaceae, Rafflesiaceae, Aneuraceae, Orchidaceae, Petrosaviaceae, and Triuridaceae (Wolfe et al. 1992; Funk et al. 2007; McNeal et al. 2007; Wickett et al. 2008; Delannoy et al. 2011; Logacheva et al. 2011; Barrett and Davis 2012; Wicke et al. 2013, Barrett et al. 2014; Logacheva et al. 2014; Molina et al. 2014; Lam et al. 2015; Schelkunov et al. 2015; table 1). Comparison of these 24 plastomes (table 1) shows that 23 of them still contain the same 17 genes (ten ribosomal proteins, four ribosomal RNAs, and three transfer RNAs), either because of global or lineage-dependant selective pressure or by inertia. The exception is *R. lagascae* in which no putative plastid sequences have intact reading frames (Molina et al. 2014; and see Smith and Lee 2014 for a possible loss of the plastome in the algae *Polytomella*). This species is the only endoparasitic land plant investigated so far and seems to completely lack a functional or pseudogenized plastome (Molina et al. 2014). All the other parasitic land plants analyzed are exoparasites, meaning they connect to the host via haustoria from the outside or via fungal hyphae. Endoparasites, such as *Rafflesia*, instead live permanently inside the host as a network of parenchyma cells (Heide-Jorgensen 2008; Molina et al. 2014). Endoparasitism has evolved four times in land plants, namely in Rafflesiaceae (34 species in three genera), Cytinaceae (~10 species in two genera), Mitrastemonaceae (one or two species in one genus), and Apodanthaceae (10 species in two genera, *Apodanthes* and *Pilostyles*). The stem lineage of Rafflesiaceae is about 95 Myr old (Bendiksby et al. 2010), those of Cytinaceae 72 Myr,

Table 1
Plastome Content of Nonphotosynthetic Land Plants, and Age of Parasitism/Mycoheterotrophy

Name	Type of Parasitism	Number of Functional Genes/RNAs[a]	Source	Age of Parasitism or Mycoheterotrophy[b](Myr)
Pilostyles aethiopica (Apodanthaceae)	Endo-holoparasite	5?	This study	< 81 (62-98), Bellot and Renner 2014b
Pilostyles hamiltonii (Apodanthaceae)	Endo-holoparasite	5?	This study	< 81 (62-98), Bellot and Renner 2014b
Rafflesia lagascae (Rafflesiaceae)	Endo-holoparasite	Plastome not found	Molina et al. 2014	< 95 (83-109), Bendiksby et al. 2010
Aneura mirabilis (Aneuraceae, Liverworts)	Exo-holomycotroph	92	Wickett et al. 2008	?
Corallorhiza var. *maculata* (Orchidaceae)	Exo-holomycotroph	89	Barrett et al. 2014	≪ 49, Epidendroideae; Lovisa et al. 2010
Corallorhiza var. *occidentalis* (Orchidaceae)	Exo-holomycotroph	88	Barrett et al. 2014	≪ 49, Epidendroideae; Lovisa et al. 2010
Corallorhiza mertensiana (Orchidaceae)	Exo-holomycotroph	90	Barrett et al. 2014	≪ 49, Epidendroideae; Lovisa et al. 2010
Corallorhiza striata (Orchidaceae)	Exo-holomycotroph	82	Barrett and Davis 2012	≪ 49, Epidendroideae; Lovisa et al. 2010
Epipogium aphyllum (Orchidaceae)	Exo-holomycotroph	38	Schelkunov et al. 2015	≪ 49, Epidendroideae; Lovisa et al. 2010
Epipogium roseum (Orchidaceae)	Exo-holomycotroph	30	Schelkunov et al. 2015	≪ 49, Epidendroideae; Lovisa et al. 2010
Neottia nidus-avis (Orchidaceae)	Exo-holomycotroph	59	Logacheva et al. 2011	≪ 49, Epidendroideae; Lovisa et al. 2010
Petrosavia stellaris (Petrosaviaceae)	Exo-holomycotroph	72	Logacheva et al. 2014	≪ 49, Epidendroideae; Lovisa et al. 2010
Rhizantella gardneri (Orchidaceae)	Exo-holomycotroph	32	Delannoy et al. 2011	≪ 49, Epidendroideae; Lovisa et al. 2010
Sciaphila densiflora (Triuridaceae)	Exo-holomycotroph	28	Lam et al. 2015	≪ 90-50 Triuridaceae; Mennes et al. 2013
Boulardia latisquama (Orobanchaceae)	Exo-holoparasite	54	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013
Cistanche deserticola (Orobanchaceae)	Exo-holoparasite	62	Li et al. 2013	< 32 (13-52) Naumann et al. 2013
Cistanche phelypaea (Orobanchaceae)	Exo-holoparasite	60	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013
Conopholis americana (Orobanchaceae)	Exo-holoparasite	49	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013
Cuscuta gronovii (Convolvulaceae)	Exo-holoparasite	89	Funk et al. 2007	< 35 (13-57) Naumann et al. 2013
Cuscuta obtusiflora (Convolvulaceae)	Exo-holoparasite	92	McNeal et al. 2007	< 35 (13-57) Naumann et al. 2013
Epifagus virginiana (Orobanchaceae)	Exo-holoparasite	51	Wolfe et al. 1992	< 32 (13-52) Naumann et al. 2013
Myzorrhiza californica (Orobanchaceae)	Exo-holoparasite	78	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013
Orobanche crenata (Orobanchaceae)	Exo-holoparasite	63	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013
Orobanche gracilis (Orobanchaceae)	Exo-holoparasite	58	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013
Phelipanche purpurea (Orobanchaceae)	Exo-holoparasite	60	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013
Phelipanche ramose (Orobanchaceae)	Exo-holoparasite	57	Wicke et al. 2013	< 32 (13-52) Naumann et al. 2013

[a]The number of genes in a typical angiosperm is 116 (from Barrett et al. 2014), whereas for photosynthetic *Aneura* it is 121 (Wickett et al. 2008).
[b]The age of parasitism/mycoheterotrophy is younger than the stem age of the respective parasite/mycoheterotrophic clade.

Mitrastemonaceae 78 Myr, and Apodanthaceae about 100 Myr (Naumann et al. 2013).

Apodanthaceae occur in North and South America, Africa, Iran, and Australia (Bellot and Renner 2014b), and based on mitochondrial and nuclear sequences they belong in the Cucurbitales (Filipowicz and Renner 2010). Like other endoparasites, they lack leaves and stems, emerging only as small flowers that break through the host's bark once a year (fig. 1). No chloroplasts have ever been observed in their tissues (Rutherford 1970: *Pilostyles thurberi*; Dell et al. 1982; *P. hamiltonii*), although plastid-like compartments have been reported for *P. hamiltonii* (Dell et al. 1982). The plant's body consists of cell clusters inside their hosts, which are Fabaceae for *Pilostyles* in the Americas, Africa, Iran, and Australia, but Salicaceae for *Apodanthes* in South America.

In this study, we investigate the African *Pilostyles aethiopica* and the Australian *P. hamiltonii*, endoparasites that diverged from each other about 23–33 Myr ago and that live inside Fabaceae (Bellot and Renner 2014a), to test if they retain a

Fig. 1.—Flowers of *Pilostyles aethiopica* emerging from the host *Julbernardia globiflora* (Fabaceae) in Harare, Zimbabwe. Scale bar is 5 mm. Photo S. Bellot.

functional plastome. We used three kinds of evidence to infer the genomic location of any plastome-like DNA region: BLAST searches, flanking regions, and read depth differences. In normal photosynthetically active plants, copy number of plastome sequences is expected to be one to two orders of magnitude higher than that of mitochondrial sequences and two to four orders higher than that of nuclear sequences (Zoschke et al. 2007), although those proportions may change in nonphotosynthetic tissues.

Materials and Methods

Taxon Sampling, DNA Sequencing, and Genome Size Measurements

Flower tissue from a female individual of *P. aethiopica* (voucher S. Bellot 28, deposited in the herbarium of Munich) was collected in the Mukuvisi woodlands of Harare, Zimbabwe on February 29, 2012 and kept frozen until DNA isolation with the kit DNeasy Plant Maxi Kit (Qiagen), following the manufacturer's protocol. One microliter of the DNA (17.2 ng/μl) was sent to Eurofins MWG Operon for precipitation and sequencing. One genomic shotgun library of insert sizes 160–310 bp was sequenced in one channel of Illumina HiSeq 2000, yielding 236,404,172 paired-end reads of 101 bp. The same DNA sample was also submitted by the same company to a run of pyrosequencing using a GS FLX+ sequencer (454 Life Science, Roche), yielding 864,474 reads of 640 bp in average, already trimmed based on quality.

DNA from silica-dried flowers of male and female *P. hamiltonii* collected near Perth in October 2010 (voucher K. Dixon 1039 in the herbarium PERTH) was isolated using the same approach, and the DNA was sent to the University of Vienna (C. Schlötterer's lab) for Illumina sequencing on a Genome Analyzer IIx platform. This yielded 80,223,076 paired-end reads of 101 bp.

The 1C values of *P. aethiopica* and *P. hamiltonii* were determined with flow cytometry, using the same batch of flowers as used for DNA sequencing. Flow cytometry relied on propidium iodide as the DNA stain and *Solanum pseudocapsicum* as the standard, following the protocol of Temsch et al. (2010). A CyFlow ML flow cytometer (Partec, Muenster, Germany) equipped with a green laser (100 mW, 532 nm, Cobolt Samba, Cobolt, Stockholm, Sweden) was used for the fluorescence measurements, with 5,000 particles measured per run and three runs performed per plant preparation. The C value was calculated according to the formula: 1C valueObject = (mean G1 nuclei fluorescence intensityObject/mean G1 nuclei fluorescence intensity Standard)*1C valueStandard.

Quality Control, Preprocessing, and Assembly of Reads

For *P. aethiopica*, following quality control of the Illumina reads with PRINSEQ (Schmieder and Edwards 2011) and FASTQC (Andrews, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed December 27, 2015), adaptors were removed when necessary, using fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/, last accessed December 27, 2015), and sequences were trimmed at both ends using PRINSEQ to remove polyA/T tails greater than 5 nt and all bases with a quality score less than 20, stopping at the first base with a quality greater than 20. The few sequences shorter than 60 bp were then removed as were sequences of a mean quality score less than 30, or greater than 1% of Ns, or an entropy less than 70. This left 180,443,106 reads. For *P. hamiltonii*, adaptors were removed from the reads using Cutadapt (Martin 2011), and the reads were then filtered and trimmed using PRINSEQ with similar stringency thresholds as for *P. aethiopica*. This left 69,000,257 reads.

De novo assemblies of the cleaned total reads of *P. aethiopica* and *P. hamiltonii* were performed on the CLC Genomics Workbench 7 (http://www.clcbio.com, last accessed December 27, 2015) using different word and bubble sizes, of which the automatic ones provided the best results, and improved with SSPACE (Boetzer et al. 2011), which remaps reads using the included Bowtie assembler and scaffolds the resulting contigs using the paired-end information. This produced 952,874 de novo contigs for *P. aethiopica* and 270,940 for *P. hamiltonii*, with N50 of 601 and 446 bp, and a maximum contig size of 61,798 and 33,472 bp. The 454 reads of *P. aethiopica* were also assembled with CLC, producing 54,793 contigs with a N50 of 558 bp and a maximum size of 37,200 bp.

Isolating and Assembling Reads with a Reference Plastid Genome

The Illumina reads of *P. aethiopica* and *P. hamiltonii* were mapped with low stringency against the plastome of *Cucumis sativus* (supplementary table S1, Supplementary Material online, lists all plastid genomes used in this study). For each gene, rRNA, or tRNA, the corresponding reads were extracted, as well as reads matching adjacent intergenic regions until the first gap in the mapping. Reads were then de novo assembled with Geneious R7 (Biomatters, http://www.geneious.com/, last accessed December 27, 2015), using the highest stringency to allow the separation of similar sequences mapping to the same region. Contigs were checked for ambiguities, and weakly covered consensuses (with at least 3, 20, 40, or 80 reads, depending on the length of the region) were retained to avoid losing plastome-like nuclear regions or a plastid genome that might be present in a low copy number. In many cases, multiple contigs corresponded to a given gene. All retained contigs were aligned against the plastome of *C. sativus*, keeping only contigs with an e-value <0.00001 (BLASTn command of BLAST+ version

29; Camacho et al. 2008; ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.29/, last accessed December 27, 2015). This reduced the number of plastome-like contigs from 3,755 to 313 for *P. aethiopica* and from 873 to 215 for *P. hamiltonii*. Finally, we kept the de novo contigs (previous section) overlapping those reference-based contigs as possible candidates for forming a plastome.

Finding More Divergent Plastome-Like Sequences

Because the reference genome approach did not recover a *Pilostyles* plastid genome (Results), we used a second strategy, which consisted in blasting the de novo contigs of *Pilostyles* against the plastid genes of 701 organisms (including the nonphotosynthetic apicomplexan *Babesia microti*) for which plastomes were available on GenBank in January 2015, using low stringency (maximum e-value of 10, maximum number of target sequences equal to total number of genes, and "-task" option set to BLASTn with default word-size of 11, as was the case for the BLASTn analyses described below) and keeping all contigs to which at least one gene (or a tRNA or a rRNA) matched along $\geq$50% of its length, sometimes in multiple high scoring pairs (gapped alignment). This allowed us to retrieve 60 more de novo contigs as plastome gene candidates in *P. aethiopica* and two more in *P. hamiltonii*, resulting in 176 and 121 candidates, respectively. In an even more conservative approach, the same analysis was conducted using tBLASTx and keeping all contigs matching with an e-value $\leq$0.01, no matter the length of the match. The flanking regions of the plastome-like regions were retrieved using custom scripts, BEDtools (Quinlan and Hall 2010), and SAMtools (Li et al. 2009). Mean read depth was retrieved for every de novo contig after stringent (99% overlap, 99% identity) remapping of the reads using CLC and following removal of reads possibly resulting from polymerase chain reaction (PCR) duplicates (as inferred from their mapping coordinates) using SAMtools (command "rmdup" option "–S"). To confirm that borders between flanking and plastome-like regions were not unusually covered compared with the rest of the contig, we used the "Coverage analysis" option available in CLC, using a conservative minimum length (6 bp) and *P* value (0.05), which allowed identification of even small regions of low/high coverage (normal coverage being as the global coverage of the contig). From this analysis, we pooled the reads belonging to possibly misassembled contigs containing plastome-like regions with the reads from contigs selected as plastome candidates (Results), and reassembled them de novo using Geneious R7 at low, medium, and high stringency to check for alternative assemblies that would produce different plastome candidates. For *P. aethiopica*, the 454 reads were mapped to the borders when available, and they all matched perfectly whenever CLC inferred a normal coverage (data not shown).

Genomic and Phylogenetic Origin of Plastome-Like DNA Fragments

To infer the genomic placement of the candidate plastome contigs of both species of *Pilostyles*, each plastome-like region was aligned against the NCBI nucleotide database using BLASTn, with a maximum e-value of 10 and a maximum number of target sequences set to 5,000. Custom Python scripts were written to record hits bitscores and genomic compartments. Flanking regions of the plastome-like regions were submitted to the same analysis, and results were double-checked manually. The genomic location of plastome-like contigs was also inferred from their read depth in comparison to that of contigs known to belong to the mitochondrial and nuclear genomes. Mitochondrial contigs were identified by blasting all contigs against the mitochondrial genes of *Citrullus lanatus* (NC_014043), *Cucurbita pepo* (NC_014050; Alverson et al. 2010), and *C. sativus* (NC_016004, NC_016005, and NC_016006; Alverson et al. 2011), and keeping for each gene the contig matching with the highest bitscore, some of them carrying multiple mitochondrial genes. Nuclear contigs were identified the same way, using as a reference the nuclear protein-coding genes of *C. sativus* published by Li et al. (2011) and available in GenBank (BioProject number PRJNA80169).

The high divergence and short length of most sequences made their phylogenetic placement difficult, regardless whether using BLAST or alignments and trees searches. We used a script (http://seqanswers.com/forums/showthread.php?t=40975, last accessed December 27, 2015) to extract GenBank's taxonomic classification information for each sequence and then recorded the bitscore of its first hit to Cucurbitales (the order of the parasite family Apodanthaceae), Fabales (the order of the host), Fagales or Rosales (the four orders form a monophylum), or to other orders. Contigs with parts most similar to Fabales were kept even when they were globally more similar to another order (including Cucurbitales) because they could represent cases of horizontal gene transfer from the hosts. Contigs that completely matched bacteria or Fabales (the order to which all African and Australian host species belong) were removed from further analyses after we had made sure that they could not represent parts of the *Pilostyles* plastome by looking at their gene content. We are confident that these removed contigs do not form part of any *Pilostyles* plastome because all their reads were included in the iterative remapping analyses performed to extend plastome candidates (next section); none of them allowed further extension. The genes found in the plastomes of *P. aethiopica* and *P. hamiltonii* (Results) were aligned, and we then performed maximum likelihood phylogenetic analyses to infer their relationships with other Viridiplantae, using 37–39 representative lineages of land plants and green algae as outgroups (supplementary table S1, Supplementary Material online). Alignments were generated with Geneious

R7 and MAFFT 7.017 (Katoh et al. 2002), and tree searches were conducted in RAxML 7.2.8 (Stamatakis 2006) with the GTR + G substitution model and 100 bootstrap replicates, all through Geneious R7.

Extension and Concatenation of the Plastome-Like Contigs

For each *Pilostyles* species, all reads were remapped iteratively at low stringency (requiring only 30% of length to match at 100%) onto the candidate plastome contigs (Results), using CLC. The ends of each contig were checked manually for possible further extension and/or alternative ends that would allow concatenation. After each extension, contigs were blasted against the pooled de novo contigs, to ensure that no plastid contig was missed, especially divergent noncoding regions.

To confirm the sequence of a few low-complexity repeated regions and the closure of the plastome of *P. aethiopica*, the genomic DNA used for the next-generation sequencing was used for PCRs using standard protocols (as in Bellot and Renner 2014a) and Sanger sequencing relying on the Big Dye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) and an ABI 3130-4 automated capillary sequencer. The same procedure was also applied to join the plastome contigs of *P. hamiltonii*, on a DNA sample (voucher: K. Thiele 4527, herbarium PERTH) collected at the same locality as the sample used for next-generation sequencing. We also designed a pair of primers to amplify the whole plastome of *P. aethiopica* by long-range PCR performed on the DNA used for Illumina sequencing, using the following conditions: 25 µl of LongAmp Hot Start Taq 2X Master Mix (New England BioLabs), 2 µl of each primer, 1 µl (120 ng) of DNA, and 20 µl of H_2O yielding an individual reaction volume of 50 µl. Termocycling conditions were 30s at 94 °C + 30 * (30 s at 94 °C + 60 s at 46 °C + 9 min at 65 °C) + 10 min at 65 °C. The primers were designed using Primer3Plus v. 2.3.6 (Untergasser et al. 2012) and are listed at the end of supplementary table S1, Supplementary Material online.

In Silico Assessment of Plastid Gene Functionality

Contigs bearing plastome-like regions were annotated using DOGMA (Wyman et al. 2004; http://dogma.ccbb.utexas.edu/html/cite.html, last accessed December 27, 2015) at the lowest possible stringency (threshold of 25% identity, gapped and nongapped alignment), and amino acid or nucleotide sequences of all gene and rRNA fragments identified were aligned with the corresponding genes of *Cucumis*, manually compared for identity and searched for frameshifts and stop-codons (for protein-coding genes). For the plastomes, the results of DOGMA were refined using ORF Finder online (Tatusov and Tatusov; http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi, last accessed December 27, 2015) as well as manual BLASTn, tBLASTx, and BLASTp against Genbank. In addition, the two plastomes were searched for tRNAs using tRNAscan-SE 1.21 (Lowe and Eddy 1997) at low stringency (Cove score 15), and secondary structure of the candidate tRNAs was investigated using the online program RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi, last accessed December 27, 2015).

Results

Genome Size, Sequencing Coverage, and Genomic Location of Plastome-Like Regions

The 1C value of *P. aethiopica* is 1.5 pg, that of *P. hamiltonii* 5 pg, which corresponds to, respectively, 1.467 ($1.50.978*10^9$) and 4.89 Gb. *Pilostyles aethiopica* has a chromosome number of $n = 29$ (Bellot 2015); the chromosome number of *P. hamiltonii* is unknown. The results from blasting the *de novo* contigs of *P. aethiopica* and *P. hamiltonii* against Genbank (Material and Methods) are presented in table 2 and tables S2 and S3, Supplementary Material online. To confirm or clarify the genomic locations inferred from blasting, we used coverage information (fig. 2). The length of the assembled *P. aethiopica* contigs is 498,766,256 bp, that of *P. hamiltonii* 117,704,350 bp, representing ca. 34% and 2.4% of their respective genomes; the expected read depth is thus 36 and 59 discounting copy number variation. Figure 2 shows the read depths of the mitochondrial and nuclear contigs as well as that of plastome-like contigs located in the mitochondrial genome as inferred from their flanking regions. For both species, the latter have a read-depth similar to that of contigs known to be mitochondrial. In a second step, the contigs of unclear location were assigned to a genomic compartment on the basis of expected differences in read depths and by performing further read mappings (supplementary tables S2 and S3, Supplementary Material online, provide details about the assignment of particular contigs). In *P. aethiopica*, 48 plastome-like regions are located in the chondriome, nine in the nucleus, and one contig may be part of the plastome. In *P. hamiltonii*, 55 plastome-like regions are located in the chondriome, seven in the nucleus, three may be in either the chondriome or the nuclear genome, and three may form a plastome (table 2).

Do the Plastome-Like Contigs in the Two Species of *Pilostyles* Form a Plastome?

The only plastid contig found in *P. aethiopica* (contig 177; arrow in fig. 2) had a mean read depth of 174, higher than any of the other de novo contigs. Contig 177 matches first to an *rrn*16 region (supplementary table S2, Supplementary Material online) with high identity to those amplified by PCR from *P. hamiltonii* (Thiele et al. 2008) and *P. thurberi* (Nickrent et al. 1997). Read remapping as well as blasting of contig 177 at low stringency to all *de novo* contigs did not allow further extension, and showed homogeneous coverage except at the

Table 2

Contigs of *Pilostyles hamiltonii* and *Pilostyles aethiopica* Containing Plastome-Like Regions, with Their Genomic Location, Read Depth, and Gene Content

De Novo Contig	Genomic Location	Mean Coverage (in reads by bp)	Protein-Coding Genes	*De Novo* Contig	Genomic Location	Mean Coverage (in reads by bp)	Protein-Coding Genes
Pham_scaffold_3	**MT**	**67.6**		**Paet_scaffold_1**	**MT**	**129.7**	***orf*42***
Pham_scaffold_5	**MT**	**68.9**		**Paet_scaffold_4**	**MT**	**131.6**	***acc*D*, *atp*A*, *psb*T***
Pham_scaffold_9	MT	87.1	*psb*D*	**Paet_scaffold_5**	**MT**	**125.7**	
Pham_scaffold_10	**MT**	**87.1**	***rpl*23***	**Paet_scaffold_8**	**MT**	**127.3**	
Pham_scaffold_16	MT	73.5		**Paet_scaffold_10**	**MT**	**129.7**	***psb*E**
Pham_scaffold_17	**MT**	**86.5**		**Paet_scaffold_13**	**MT**	**126.0**	
Pham_scaffold_22	MT	79.4	*rpo*C1*	**Paet_scaffold_17**	**MT**	**127.9**	
Pham_scaffold_23	MT	81.8		**Paet_scaffold_20**	**MT**	**128.7**	
Pham_scaffold_25	**MT**	**65.0**		**Paet_scaffold_24**	**MT**	**126.2**	
Pham_scaffold_26	**MT**	**84.5**		**Paet_scaffold_28**	**MT**	**128.7**	
Pham_scaffold_29	MT	60.7		**Paet_scaffold_29**	**MT**	**130.0**	
Pham_scaffold_32	**MT**	**79.5**	***ndh*C*, *ndh*K***	**Paet_scaffold_31**	**MT**	**128.6**	
Pham_scaffold_34	**MT**	**116.3**		**Paet_scaffold_37**	**MT**	**128.4**	***rpl*23***
Pham_scaffold_37	MT	72.6	*inf*A*	**Paet_scaffold_42**	**MT**	**122.5**	
Pham_scaffold_38	MT	84.5	*pet*B*, *pet*D*, *psb*H, *rpo*A*, *rps*11	**Paet_scaffold_50**	**MT**	**133.8**	***ccs*A*, *cem*A***
Pham_scaffold_40	**MT**	**79.5**		**Paet_scaffold_51**	**MT**	**120.0**	***atp*B, *psb*C*, *psb*Z, *rps*7***
Pham_scaffold_41	**MT**	**95.6**		**Paet_scaffold_67**	**MT**	**131.5**	***acc*D*, *atp*A*, *atp*F*, *atp*I***
Pham_scaffold_42	**MT**	**77.1**		**Paet_scaffold_70**	**MT**	**127.8**	***atp*B, *atp*E***
Pham_scaffold_44	**MT**	**74.8**		**Paet_scaffold_72**	**MT**	**129.1**	
Pham_scaffold_46	MT	79.5	*rbc*L*	**Paet_scaffold_73**	**MT**	**117.3**	
Pham_scaffold_50	**MT**	**84.8**	***rpo*C1***	**Paet_scaffold_84**	**MT**	**126.5**	
Pham_scaffold_58	MT	88.3		**Paet_scaffold_85**	**MT**	**122.7**	
Pham_scaffold_62	**MT**	**67.3**	***cem*A***	**Paet_scaffold_93**	**MT**	**131.7**	
Pham_scaffold_65	**MT**	**74.9**		Paet_scaffold_95	MT	132.2	*rps*2*
Pham_scaffold_66	**MT**	**65.4**		**Paet_scaffold_106**	**MT**	**129.5**	**accD***
Pham_scaffold_69	MT	95.3	*psb*E*, psbF*, *psb*J, *psb*N*, *rpl*20*, *ycf*1*	**Paet_scaffold_107**	**MT**	**136.9**	
Pham_scaffold_70	**MT**	**76.5**		Paet_scaffold_108	MT	134.6	
Pham_scaffold_72	MT	100.1		**Paet_scaffold_122**	**MT**	**136.4**	
Pham_scaffold_78	**MT**	**67.1**	***orf*56**	Paet_scaffold_125	MT	138.8	*acc*D*, *atp*B, *ndh*J*, *ndh*K*, *rbc*L*, *rps*4*
Pham_scaffold_80	MT	79.5	*psa*A*	**Paet_scaffold_126**	**MT**	**134.9**	
Pham_scaffold_81	**MT**	**87.5**		**Paet_scaffold_136**	**MT**	**133.7**	***pet*A***
Pham_scaffold_83	**MT**	**73.8**	***atp*H***	**Paet_scaffold_157**	**MT**	**125.8**	
Pham_scaffold_88	**MT**	**60.1**		**Paet_scaffold_166**	**MT**	**123.3**	***pet*G**
Pham_scaffold_113	MT	79.1		Paet_scaffold_177	CP?		*acc*D, *rpl*2*, *rps*3,*rps*4, *rps*12*, *rrn*16, *rrn*23
Pham_scaffold_158	**MT**	**76.7**		**Paet_scaffold_189**	**MT**	**126.1**	
Pham_scaffold_160	**MT**	**68.3**	***psa*C**	**Paet_scaffold_213**	**MT**	**136.5**	
Pham_scaffold_163	**MT**	**59.4**	***psa*A***	Paet_scaffold_226	MT	126.8	*rpo*C1*
Pham_scaffold_175	MT	69.3	*psb*B*, *psb*T*	**Paet_scaffold_243**	**MT**	**124.6**	
Pham_scaffold_180	**MT**	**90.0**	***rpo*C2***	**Paet_scaffold_245**	**MT**	**113.8**	
Pham_scaffold_198	MT	48.9		**Paet_scaffold_267**	**MT**	**120.5**	***psb*A***
Pham_scaffold_203	MT or NC	45.9	*ndh*J*	Paet_scaffold_286	MT	117.2	
Pham_scaffold_205	**MT**	**80.0**		Paet_scaffold_299	MT	130.6	*psb*N*
Pham_scaffold_207	CP?	111.6	*rpl*2*, *rrn*23	**Paet_scaffold_319**	**MT**	**123.4**	

(continued)

Table 2 Continued

De Novo Contig	Genomic Location	Mean Coverage (in reads by bp)	Protein-Coding Genes	*De Novo* Contig	Genomic Location	Mean Coverage (in reads by bp)	Protein-Coding Genes
Pham_scaffold_241	**MT**	**61.4**		**Paet_scaffold_346**	**MT**	**137.6**	
Pham_scaffold_244	MT or NC	52.7		**Paet_scaffold_377**	**MT**	**129.9**	***atp*A*, *ycf*1***
Pham_scaffold_251	MT or NC	59.8		Paet_scaffold_420	MT	133.1	
Pham_scaffold_261	CP?	89.3	*rrn*16	Paet_scaffold_512	MT	118.8	
Pham_scaffold_266	**MT**	**71.2**		Paet_scaffold_908	MT	118.6	*psb*D*, *rpo*B*
Pham_scaffold_270	MT	68.5	*clp*P*	Paet_scaffold_1807	NC	7.3	
Pham_scaffold_315	**MT**	**67.8**	***rps*12***	Paet_scaffold_2390	MT	128.9	
Pham_scaffold_462	MT	71.7		Paet_scaffold_19077	NC	5.4	
Pham_scaffold_491	MT	75.6		Paet_scaffold_53779	NC	7.2	*psa*A*
Pham_scaffold_503	MT	60.0		Paet_scaffold_77278	NC	7.6	
Pham_scaffold_506	MT	42.9		Paet_scaffold_113062	NC	5.3	accD*
Pham_scaffold_508	MT	48.7		Paet_scaffold_149204	NC	22.9	
Pham_scaffold_706	CP?	116.9	*acc*D	Paet_scaffold_164148	NC	3.0	*atp*H*
Pham_scaffold_775	MT	92.5	*psa*B*, *rps*14*	Paet_scaffold_277960	NC	4.7	
Pham_scaffold_2613	**NC**	**64.5**		Paet_scaffold_605245	NC	4.4	
Pham_scaffold_14490	**MT**	**63.6**					
Pham_scaffold_22113	MT	134.5	*psb*A*				
Pham_scaffold_33324	NC	8.4					
Pham_scaffold_49221	NC	5.6					
Pham_scaffold_50886	NC	4.9	*ycf*4*				
Pham_scaffold_79191	NC	1.6					
Pham_scaffold_90224	NC	4.8					
Pham_scaffold_148584	MT	89.7	*rpl*36				
Pham_scaffold_220160	MT	78.2					
Pham_scaffold_223261	NC	4.6					

Note.—Genes marked with an asterisk are pseudogenes, whereas the others may be functional as inferred only from the DNA sequence (see supplementary table S4, Supplementary Material online, for details). Contigs without plastid genes show similarities with very small plastid gene fragments or with uncoding plastid regions. Genomic location of the plastid regions in contigs in bold was inferred from their flanking regions whereas for the others, it required coverage analysis and read remappings (see supplementary tables S2 and S3, Supplementary Material online, for details). MT, mitochondrial; NC, nuclear; CP?, possibly part of a chloroplast genome.

beginning and end of the contig and in two other small sections (supplementary fig. S2*A*, Supplementary Material online). Circularization of contig 177 of *P. aethiopica* is supported in silico by eight Illumina reads and 24 paired-end reads having one mate mapping to the start of the contig and one mate mapping to its end. Sanger sequencing using primers matching the ends of this contig was successful, confirming the existence of plastomes in which the ends are adjacent (supplementary fig. S2*B*, Supplementary Material online). Finally, long-range PCR using a primer pair designed on contig 177 (fig. 2A) yielded a product size of approximately 11 kb, corresponding to the size of contig 177 (supplementary fig. S2*B*, Supplementary Material online, inset). The contig with the highest read depth obtained from the assembly of the 454 data was identical to contig 177, with the same pattern of coverage distribution, albeit not opened (linearized) at the same place but instead in a well-covered repeated region. By Sanger resequencing we checked the assembly of contig 177 (supplementary fig. S2*C*, Supplementary Material online). These different types of evidence support that contig 177 forms the complete plastome of *P. aethiopica*, although it is probably not always circular.

The plastid contigs found in *P. hamiltonii* (contigs 207, 261, and 706) align to different regions of the plastome of *P. aethiopica* and have read depths of, respectively, 112, 89, and 117, about the same as mitochondrial contigs (table 2 and fig. 2). Contigs 207 and 706 match to *Plasmodium* and green algae and then to angiosperm plastomes, and contig 261 matches the *rrn*16 sequences of *P. hamiltonii* found in GenBank (EF446141 to EF446144 and EU512418 to EU512420; Thiele et al. 2008). Blasting of the de novo contigs of *P. hamiltonii* to the plastome of *P. aethiopica*, as well as iterative remappings of all the reads (Material and Methods), revealed six more contigs that overlapped contigs 207, 261, and 706 by at least a few reads. Those six contigs had mean read depths between 31 and 77, and two of them showed similarities with known plastid genes (below). PCR amplification and Sanger sequencing (supplementary fig. S2*D*, Supplementary Material online) suggest that all nine contigs form a circular plastome, collinear to that of *P. aethiopica*.

Possibly Functional Genes in the Plastomes of *Pilostyles*

The plastomes of *P. aethiopica* and *P. hamiltonii* are represented in figure 3. Mapping of the reads to the plastomes

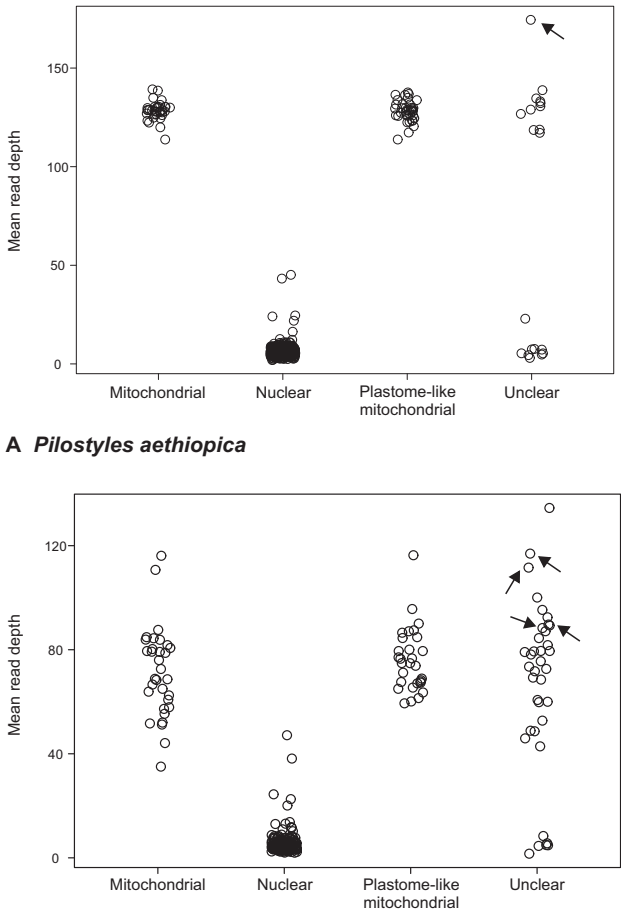

Fig. 2.—Read depth of *Pilostyles* contigs with plastome-like regions. Unclear: Contigs with plastome-like regions whose flanking regions were too short to infer their genomic location. Arrows indicate contigs that are part of plastomes (Results). The means for *P. aethiopica* are based on 34 contigs located in the chondriome, 670 located in the nuclear genome, 38 inferred to be in the chondriome from their flanks, and 19 of unclear location. The means for *P. hamiltonii* are based on 32 contigs located in the chondriome, 369 located in the nuclear genome, 31 inferred to be in the chondriome from their flanks, and 37 of unclear location.

at various stringencies did not reveal the presence of the typical inverted repeat. In total, the plastome of *P. aethiopica* is 11,348 bp long, of which 52% (5,901 bp) are recognizable as functional or pseudogenized genes and rRNAs. The plastome of *P. hamiltonii* is 15,167 bp long, of which 41.6% (6,314 bp) are functional or pseudogenized genes; its larger size is thus mostly due to an increase of non-coding DNA. The mean GC content of the plastome of *P. aethiopica* is 24.2% and that of *P. hamiltonii* is 22.7% (fig. 3).

We found the genes *acc*D, *rpl*2, *rps*3, *rps*4, *rps*12, *rrn*16, and *rrn*23 in both species, and *rpo*C2 only in *P. aethiopica*. At least *rpl*2 and *rpo*C2 are pseudogenes judging from their short size and lack of start codons, whereas *acc*D, *rps*3 and *rps*4 may be functional because they have long open reading frames (supplementary table S4, Supplementary Material online, provides details on the inference of gene function). The rRNAs are as long as in other plants, and BLAST searches of the *rrn*23 and *rrn*16 of *P. hamiltonii*, and *rrn*23 of *P. aethiopica* against *P. thurberi* transcriptome data (Matasci et al. 2014; https://www.bioinfodata.org/Blast4OneKP/, last accessed December 27, 2015) recovered a hit for each of them. The *rrn*16 of *P. aethiopica*, however, failed to match anything despite its similarity to *P. hamiltonii* (fig. 3). The *rps*12 gene from both plastomes is as long as in photosynthetic plants, but consists of one piece instead of having three exons; it has a start codon in *P. hamiltonii* but apparently not in *P. aethiopica*.

The tRNA-like sequences found in the plastomes of *P. aethiopica* and *P. hamiltonii* by comparing them to *Cucumis melo* and the apicomplexan *B. microti* do not seem capable of forming a typical cloverleaf structure. When blasting (BLASTn, word size=7) the *trn*E genes of *Cucumis* and *Babesia* against the plastomes of *P. hamiltonii* and *P. aethiopica*, there was no match with a score greater than 25. The same searches against all de novo contigs hit a mitochondrial contig of *P. hamiltonii* (number 198; table 2) with a bitscore of 104, but again it does not seem capable of forming a cloverleaf secondary structure. No hit with a bitscore greater than 54 was found in *P. aethiopica*.

The Mitochondrial and Nuclear Genomes of *Pilostyles* Contain Plastid Pseudogenes

Supplementary table S4, Supplementary Material online, summarizes the presence/absence of plastid genes in the total de novo contig sets of *P. aethiopica* and *P. hamiltonii* compared with *C. sativus*. Of the 81 protein-coding genes present in *Cucumis*, 30 had detectable traces in *P. aethiopica* and 34 in *P. hamiltonii*, with 14 shared among them (not considering the genes located in the plastomes). In *P. aethiopica*, five mitochondrial contigs contained possibly functional plastid genes (three copies of *atp*B, and *pet*G, *psb*E, and *psb*Z), as inferred from their length and absence of internal stop codons (supplementary table S4, Supplementary Material online). In *P. hamiltonii*, we also found five mitochondrial contigs containing possibly functional *orf*56, *psa*C, *psb*H, *psb*J, *rpl*36, and *rps*11 (table 2). However, none of these genes matched the transcriptome data from the American *P. thurberi* available online (Matasci et al. 2014; https://www.bioinfodata.org/Blast4OneKP/, last accessed December 27, 2015).

Many of the plastome-like sequences found in the nuclear and the mitochondrial genomes of *Pilostyles* have their first BLAST hit to Fabales (supplementary tables S2 and S3, Supplementary Material online), suggesting horizontal transfers from the host, but their shortness and high divergence prevented reliable assessment of their origin.

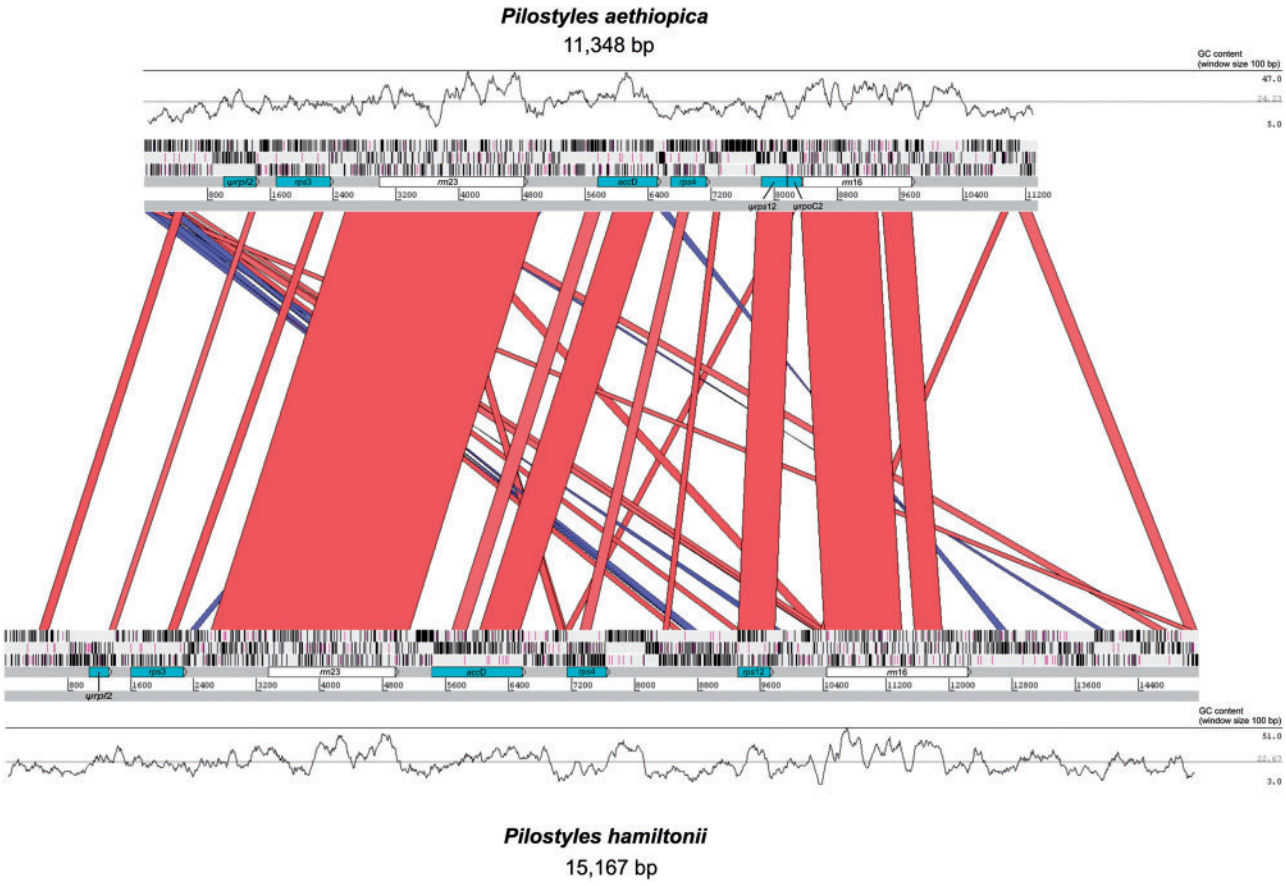

Fig. 3.—Map of the plastid genomes of *P. aethiopica* and *P. hamiltonii*. The skyline graphs represent the GC% with the minimum, mean, and maximum values indicated on the right. The blue and red bands indicate identity greater than 70% for bitscores greater than 100, red bands show a match in the same orientation whereas blue bands symbolize reversed-complement matches. The three bars above the gene labels refer to the reading frames; stop codons are represented by vertical black bars, and start codons (methionine) by purple vertical bars. Ψ means the gene is pseudogenized. Visualization obtained with the Artemis Comparison Tool (Carver et al. 2005).

No Detectable Horizontal and Internal Gene Transfers towards the *Pilostyles* Plastomes

The phylogenetic placement in Viridiplantae of each possibly functional plastome gene was assessed by maximum likelihood inference; results are shown in supplementary figure S1, Supplementary Material online. In all trees, the two *Pilostyles* species were sisters, with 100% or 90% (*rps*4) bootstrap support (BP), and the branches leading to *Pilostyles* are longer by at least one order of magnitude than other branches in the respective phylogenies. The genes *rps*3, *rps*4, or *rrn*23 placed *Pilostyles* as sister to *Silene*, *Welwitschia*, or *Psilotum*, always with low support (≤32% BP). The tree obtained from the *rrn*23 alignment does not fit currently accepted angiosperm relationships, while that from *rps*12 shows *Pilostyles* as sister to land plants (with 76% BP) in a topology that otherwise more or less fits expected relationships. The *rrn*16 gene places *Pilostyles* closer to *Cicer* (Fabales; with 62% BP), and together *Pilostyles* and *Cicer* are sister to the remaining angiosperms (in which relationships are not well resolved). When we blasted the *rrn16* of *Pilostyles* against GenBank (BLASTn at http://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed April 2015), Fabales were not among the first hits, but instead other sequences of *Pilostyles*, then *Balanophora* (a holoparasite belonging to Santalales), and then *Plasmodium* (a nonphotosynthetic apicomplexan). The *acc*D tree, finally, shows *Pilostyles* as sister to *Cucumis* (Cucurbitales; with 63% BP), and all three are located among other Rosales in a well-resolved angiosperm phylogeny. BLASTn of each gene against GenBank revealed the high divergence of the *Pilostyles* plastome genes, with every gene except *acc*D having its first hit outside of angiosperms, always with a relatively low bitscore compared with the length of the gene (results available on request).

BLAST searches (BLASTn at http://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed April 2015) of the non-coding regions of the plastome of *P. aethiopica* revealed a 509 bp-long low-complexity region (12% GC) matching an

un-annotated region of the genome of the worm *Anisakis simplex* and then (only in 325 bp) an un-annotated nuclear region of *C. melo*. It did not match mitochondrial genomes. BLASTn searches for *P. hamiltonii* revealed a 377 bp-long low complexity (15% GC) region matching the mitochondrial *nad*5 of the hymenoptera *Paraligoneurus* and a 479 bp-long low-complexity region (14% GC) matching a noncoding region of the mitochondrial genome of *Ditaxis biseriata* (Euphorbiaceae).

Discussion

Distinguishing a Plastome from Plastid-Like Regions Located in the Nuclear or Mitochondrial Genomes, and from Contaminations

Because of its high coverage and circularization, supported by PCR amplification and Sanger resequencing in different individuals (supplementary fig. S2*A–C*, Supplementary Material online), we are confident that the plastome of *P. aethiopica* indeed forms a separate genomic compartment. The plastome of *P. hamiltonii*, the circularization of which was also supported by resequencing (supplementary fig. S2*D*, Supplementary Material online), resembles the plastome of *P. aethiopica* in structure and gene content despite these species coming from different continents and having been sequenced separately (and by different companies). The relatively low coverage of both plastomes (a bit higher than, or of the same order of magnitude as, that of mitochondrial contigs, fig. 2) may be due to the relatively old flowers used for DNA extraction (in the case of *P. hamiltonii*). Coverage ratios can also be modified by endopolyploidization (Barow 2006), ontogeny (Preuten et al. 2010), and may differ in nonphotosynthetic tissues (Isono et al. 1997). In a less parsimonious explanation suggested by a reviewer of this article, the circularized plastomes of *P. aethiopica* and *P. hamiltonii* might be artifacts generated by plastid regions repeated in tandem and integrated into the mitochondrial genome. Cytological studies could distinguish between these alternatives, but must await fresh anthers for pollen mother cell counts (since there are no root tips).

Phylogenetic trees built from single plastid genes of *Pilostyles* (supplementary fig. S1, Supplementary Material online) reject the possibility that the plastomes retrieved here result from contamination. Clear contaminations we detected were the complete genome of a *Pantoea* species, a Gram-negative bacterium of the family Enterobacteriaceae and some host DNA. The distant affinity of the *rrn16* of *Pilostyles* to *Plasmodium* was not supported by further reassembly. Such unexpected similarity illustrates how lack of function can result in loss of complexity (low GC%) and nucleotide convergence in unrelated organisms.

The plastomes of *P. aethiopica* and *P. hamiltonii* do not harbor mitochondrial or nuclear regions; the only three matches (one in *P. aethiopica* and two in *P. hamiltonii*) to such regions were in low complexity stretches (<15%GC) and probably represent noise. Incorporation of mitochondrial or nuclear sequences into plastid genomes is rare (Iorizzo et al. 2012; Knox 2014), whereas the inverse is common (Notsu et al. 2002; Goremykin et al. 2009; Rice et al. 2013). Fitting with this, *Pilostyles* mitochondrial genomes have absorbed at least 64–71 plastid regions (table 2). Mitochondrial genomes of *Cucurbita pepo*, *Citrullus lanatus*, and *Cucumis sativus* also contain numerous plastid and nuclear-derived sequences (Alverson et al. 2010, 2011), and the dynamics of the

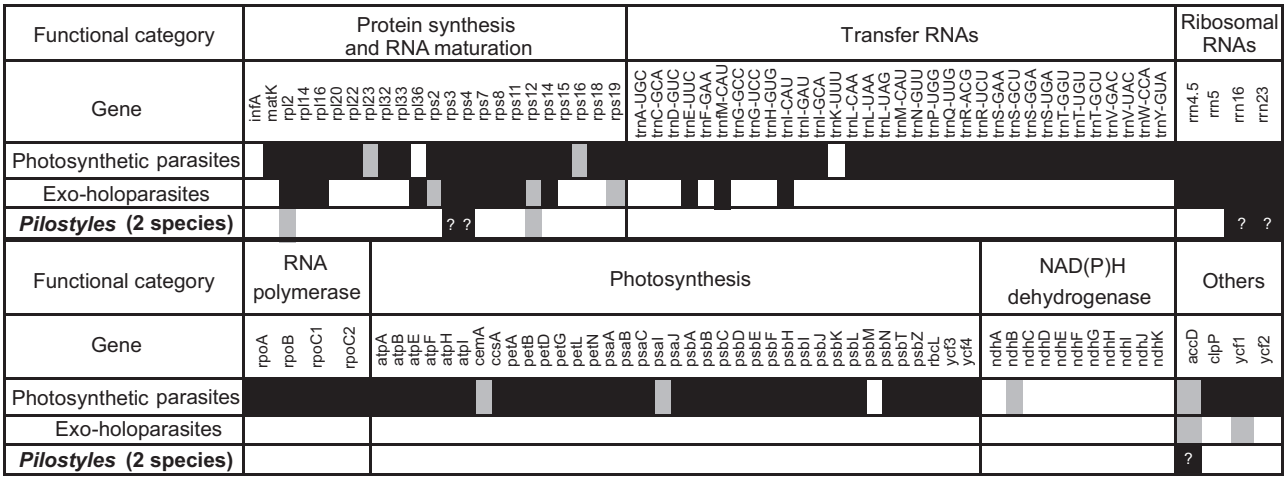

Fig. 4.—The plastome genes retained in parasitic and mycoheterotrophic land plants, based on Li et al. (2013), Barrett et al. (2014), Schelkunov et al. (2015), and Lam et al. (2015). White/grey fields indicate that a gene is lost/pseudogenized in at least one lineage of a given type of parasitism. Black fields indicate that a gene is functional in all lineages of the respective parasitism category. For *Pilostyles*, we did not take pseudogenized tRNAs into account, and the functioning of the genes remains to be confirmed (question marks).

mitochondrial genomes of *Pilostyles* may thus reflect a Cucurbitales heritage rather than result from a parasitic way of life.

Structure and Function of the Plastome of the Two *Pilostyles* Species

Prior to this study, the most reduced plastomes known from nonphotosynthetic angiosperms where those of *Sciaphila densiflora* (Triuridaceae) and *Epipogium aphyllum* and *E. roseum* (Orchidaceae), with 27–29 functional genes and transfer RNAs (Lam et al. 2015; Schelkunov et al. 2015). The plastomes of *P. aethiopica* and *P. hamiltonii*, with five or six potentially functional genes (*acc*D, *rps*3, *rps*4, *rrn*16, *rrn*23 and *rps*12 for *P. hamiltonii*) and two to three pseudogenes (*rpl*2, *rpo*C2, and *rps*12 for *P. aethiopica*), retain only about one-fifth of the genes of these previous smallest plastomes. We were able to find these extremely reduced plastomes by combining contig blasting at low stringencies, read-depth analysis, analysis of flanking regions, and a range of assembly strategies. That Molina et al. (2014) failed to identify a plastid genome in *R. lagascae*, the only land plant so far reported as possibly having lost its plastome, could be due to a similarly divergent and small plastome that escaped detection; their inference was based on coverage information and blasting at relatively high stringencies. Successful PCR amplification of the *rrn*16 gene from *Rafflesia* (Bendiksby et al. 2010) indeed points to the presence of a plastome. Under the microscope, cells of both Rafflesiaceae and Apodanthaceae have plastome-like compartments, suggesting a metabolic function (Dell et al. 1982; Molina et al. 2014), and in the American *Pilostyles thurberi*, the *rrn*16 and the *rrn*23 are expressed, despite their unusual secondary structure (Nickrent et al. 1997; Matasci et al. 2014).

Because the five or six potentially functional genes remaining in the *Pilostyles* plastome can all be transcribed by a nuclear-encoded polymerase (NEP), at least in tobacco (Liere et al. 2011), the loss of the *rpo* genes for the plastid-encoded polymerase in *Pilostyles* probably does not prevent their transcription. There is no consensus sequence for most types of NEP promoters so they are not easily identifiable from the DNA sequence only, and those of *Pilostyles* remain to be characterized. Because we found no trace of functional tRNAs in the *Pilostyles* plastomes, the tRNAs required for the expression of the five or six retained plastome genes must be imported from the cytosol, as seems to be the case for other parasites, such as *Epifagus virginiana* (Wolfe et al. 1992).

Do the Plastomes of Endo- and Exoparasites Differ in Their Function?

Plastids fulfill metabolic functions other than photosynthesis, and of the approximately 116 land plant plastome genes, only approximately 50 are involved in photosynthesis while most of the remainder are involved in the modification of the RNAs or proteins encoded by the first 50 (Bock 2007; Wicke et al. 2011). Including the two Apodanthaceae species studied here, the plastid genomes of 26 nonphotosynthetic plants from seven families have now been investigated (table 1). Of these, 23 are exoparasites that still form vegetative shoots, whereas three are endoparasites without any shoots, namely the one *Rafflesia* (Molina et al. 2014) and the two *Pilostyles* species investigated here.

The plastomes of the exoparasites still contain 27 to 92 protein-coding genes and RNAs, depending on the lineage (fig. 4 adapted from Barrett et al. 2014), with 17 genes retained in all of them: three tRNAs (*trn*E-UUC, *trn*fM-CAU, and*trn*I-CAU), ten ribosomal proteins, and four ribosomal RNAs. A plastid location of the *trn*E might be essential for biosynthesis of mitochondrial haem components (Barbrook et al. 2006), but why the other two tRNAs are retained is unclear. Other genes, namely *acc*D, *clp*P, *ycf*1, and *ycf*2 (Krause 2012), have also been considered the *raison d'être* of a plastome in nonphotosynthetic plants, even though they have been lost in some (Li et al. 2013; Wicke et al. 2013). Of these, *acc*D functions in fatty-acids biosynthesis, *clp*P is likely a protease and also involved in the import of proteins into the plastid (Krause 2012), *ycf*1 functions in "photosynthetic protein import, and [is] therefore essential for plant viability" (Kikuchi et al. 2013. p. 573), and *ycf*2 has an unknown function.

The plastomes of the two endoparasites studied here contain a seemingly functional copy of the *acc*D gene and lack a functional nuclear copy, supporting that *acc*D may be essential to plastome maintenance; the gene appears to play a role in fatty-acid synthesis and leaf development (Kode et al. 2005). We could not detect a cloverleaf-forming *trn*E in the plastome of any *Pilostyles*, which leaves open the mechanism of haem synthesis in these endoparasites. No *trn*E has been detected in the only other endoparasite studied, *R. lagascae* (Molina et al. 2014).

Living completely embedded in a photosynthetic host might have reduced selection on plastome genes more than is the case in exoparasites. The four families of endoparasites, Apodanthaceae, Cytinaceae, Mitrastemonaceae, and Rafflesiaceae, are not closely related to each other or to any exoparasites, and there is so far no scenario for how endoparasitism evolved. Parasitism in the six families of exoparasites so far studied happens to be younger than it is in the two lineages of endoparasites (table 1), and so age per se might explain the less reduced plastomes in the former. Studies of the plastomes of ancient exo-holoparasites, such as Balanophoraceae with a stem age of 110 Myr, Cynomoriaceae with a stem age of 100 Myr, or Hydnoraceae with a stem age of 101 Myr (Naumann et al. 2013), are required to test if such old exoparasites have similarly reduced plastomes as Apodanthaceae.

Supplementary Material

Supplementary figures S1 and S2 and tables S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank M. Silber and A. Sousa for support in the lab; E. Temsch, Department of Systematic Botany, University of Vienna, for C value measurements in March and May 2012; K. Dixon, Kings Park and Botanic Garden West Perth, Western Australia, K. Thiele, Western Australian Herbarium, and M. Hyde and D. Plowes, Zimbabwe, for help with collecting plant material; M. Piednoël for teaching the first author Python; S. Wicke, University of Münster, for bioinformatics help in 2011, and N. Cusimano for discussion. Financial support came from the German Science Foundation (RE 603/9-1).

Literature Cited

Alverson AJ, et al. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol.. 27:1436–1448.

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell 23:2499–2513.

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

Barow M. 2006. Endopolyploidy in seed plants. Bioessays 28:271–281.

Barrett CF, Davis JI. 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. Am J Bot. 99:1513–1523.

Barrett CF, et al. 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. Mol Biol Evol.. 31:3095–3112.

Bellot S. 2015. Natural history, taxonomy, biogeography and genome evolution of the worldwide endoparasite family Apodanthaceae (Cucurbitales). [Ph.D. thesis], Ludwig-Maximilians University, Munich, Germany.

Bellot S, Renner SS. 2014a. Exploring new dating approaches for parasites: the worldwide Apodanthaceae (Cucurbitales) as an example. Mol Phylogenet Evol.. 80:1–10.

Bellot S, Renner SS. 2014b. The systematics of the worldwide endoparasite family Apodanthaceae (Cucurbitales), with a key, a map, and color photos of most species. Phytokeys 36:41–57.

Bendiksby M, et al. 2010. Elucidating the evolutionary history of the Southeast Asian, holoparasitic, giant-flowered Rafflesiaceae: Pliocene vicariance, morphological convergence and character displacement. Mol Phylogenet Evol. 57:620–633.

Bock R. 2007. Structure, function, and inheritance of plastid genomes. In: Bock R, editor. Cell and molecular biology of plants. Topics in Current Genetics XIX. Berlin, Heidelberg: Springer-Verlag. p. 29–63.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579.

Camacho C, et al. 2008. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Carver TJ, et al. 2005. ACT: the Artemis comparison tool. Bioinformatics 21:3422–3423.

Delannoy E, Fujii S, des Francs CC, Brundrett M, Small I. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. Mol Biol Evol. 28:2077–2086.

Dell B, Kuo J, Burbidge AH. 1982. Anatomy of *Pilostyles hamiltonii* C. A. Gardner (Rafflesiaceae) in stems of *Daviesia*. Aust J Bot. 30:1–9.

Filipowicz N, Renner SS. 2010. The worldwide holoparasitic Apodanthaceae confidently placed in the Cucurbitales by nuclear and mitochondrial gene trees. BMC Evol Biol. 10:219

Funk H, Berg S, Krupinska K, Maier U, Krause K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. BMC Plant Biol. 7:45

Goremykin VV, Salamini F, Velasco R, Viola R. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol. 26:99–110.

Heide-Jorgensen HS. 2008. Parasitic flowering plants. The Netherlands: Brill.

Iorizzo M, et al. 2012. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 12:61

Isono K, Niwa Y, Satoh K, Kobayashi H. 1997. Evidence for transcriptional regulation of plastid photosynthesis genes in *Arabidopsis thaliana* roots. Plant Physiol. 114:623–630.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Kikuchi S, et al. 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science 339:571–574.

Knox EB. 2014. The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. Proc Natl Acad Sci U S A. 111:11097–11102.

Kode V, Mudd EA, Iamtham S, Day A. 2005. The tobacco plastid *accD* gene is essential and is required for leaf development. Plant J. 44:237–244.

Krause K. 2012. Plastid genomes of parasitic plants: a trail of reductions and losses. In: Bullerwell CE, editor. Organelle genetics. Berlin, Heidelberg: Springer-Verlag. p. 79–103.

Lam VKY, Gomez MS, Graham SW. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. Genome Biol Evol. 7:2220–2236.

Li H, et al. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079.

Li X, et al. 2013. Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). PLoS One 8:e58747

Li Z, et al. 2011. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. BMC Genomics 12:540

Liere K, Weihe A, Börner T. 2011. The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. J Plant Physiol. 168:1345–1360.

Logacheva MD, Schelkunov MI, Nuraliev MS, Samigullin TH, Penin AA. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. Genome Biol Evol. 6:238–246.

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

Lovisa A, Gustafsson S, Verola CF, Antonelli A. 2010. Reassessing the temporal evolution of orchids with new fossils and a Bayesian relaxed clock, with implications for the diversification of the rare South American genus *Hoffmannseggella* (Orchidaceae: Epidendroideae). BMC Evol Biol. 10:177

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10–12.
Matasci N, et al. 2014. Data access for the 1,000 plants (1KP) project. GigaScience 3:17.
McNeal JR, Kuehl J, Boore J, dePamphilis CW. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. BMC Plant. Biol. 7:57
Mennes CB, Smets EF, Moses SN, Merckx VSFT. 2013. New insights into the long-debated evolutionary history of Triuridaceae (Pandanales). Mol Phylogenet Evol. 69:994–1004.
Molina J, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). Mol Biol Evol. 31:793–803.
Naumann J, et al. 2013. Single-copy nuclear genes place haustorial Hydnoraceae within Piperales and reveal a Cretaceous origin of multiple parasitic angiosperm lineages. PLoS One 8:e79204
Nickrent DL, Duff RJ, Konings DAM. 1997. Structural analyses of plastid-derived 16S rRNAs in holoparasitic angiosperms. Plant Mol Biol. 34:731–743.
Notsu Y, et al. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics. 268:434–445.
Preuten T, et al. 2010. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. Plant J. 64:948–959.
Quinlan R, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
Rice DW, et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science 342:1468–1473.
Rutherford RJ. 1970. The anatomy and cytology of *Pilostyles thurberi* Gray (Rafflesiaceae). Aliso 7:263–288.
Schelkunov MI, et al. 2015. Exploring the limits for reduction of plastid genomes: a case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. Genome Biol Evol. 7:1179–1191.
Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864.
Smith DR, Lee RW. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. Plant Physiol. 164:1812–1819.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
Temsch EM, Greilhuber J, Krisai R. 2010. Genome size in liverworts. Preslia 82:63–80.
Thiele KR, Wylie SJ, Maccarone L, Hollick P, McComb JA. 2008. *Pilostyles coccoidea* (Apodanthaceae), a new species from Western Australia described from morphological and molecular evidence. Nuytsia 18:273–284.
Untergasser A, et al. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res. 40:e115
Wicke S, et al. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and non-photosynthetic parasitic plants of the broomrape family. Plant Cell 25:3711–3725.
Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76:273–297.
Wickett NJ, et al. 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol Biol Evol. 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. J Mol Evol. 35:304–317.
Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci U S A. 89:10648–10652.
Wyman SK, Jansen RKRK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255.
Zoschke R, Liere K, Börner T. 2007. From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. Plant J. 50:710–722.

Associate editor: Sarah Schaack