Possible Loss of the Chloroplast Genome in the Parasitic Flowering Plant Rafflesia lagascae (Rafflesiaceae)

Jeanmaire Molina,*1,2 Khaled M. Hazzouri,3 Daniel Nickrent,4 Matthew Geisler,4 Rachel S. Meyer,2 Melissa M. Pentony,5 Jonathan M. Flowers,2,3 Pieter Pelser,6 Julie Barcelona,6 Samuel Alan Inovejas,7 Iris Uy,8 Wei Yuan, Olivia Wilkins,2 Claire-Iphanise Michel,1 Selina LockLear,1 Gisela P. Concepcion,8 and Michael D. Purugganan*2,3

1Department of Biology, Long Island University, Brooklyn
2Center for Genomics and Systems Biology, New York University
3Center for Genomics and Systems Biology, NYU Abu Dhabi, Abu Dhabi, United Arab Emirates
4Department of Plant Biology, Southern Illinois University, Carbondale
5Computational Genomics Core, Department of Genetics, Albert Einstein College of Medicine, Bronx, New York
6School of Biological Sciences, University of Canterbury, Christchurch, New Zealand
7Electron Microscope Facility, St. Luke’s Medical Center, Quezon City, Philippines
8Philippine Genome Center, University of the Philippines, Diliman, Quezon City, Philippines

*Corresponding author: E-mail: jeanmaire.molina@liu.edu; mp132@nyu.edu.

Abstract

Rafflesia is a genus of holoparasitic plants endemic to Southeast Asia that has lost the ability to undertake photosynthesis. With short-read sequencing technology, we assembled a draft sequence of the mitochondrial genome of Rafflesia lagascae Blanco, a species endemic to the Philippine island of Luzon, with ~350 × sequencing depth coverage. Using multiple approaches, however, we were only able to identify small fragments of plastid sequences at low coverage depth (<2 ×) and could not recover any substantial portion of a chloroplast genome. The gene fragments we identified included photosynthesis and energy production genes (atp, ndh, pet, psa, psb, rbcl), ribosomal RNA genes (rrn16, rrn23), ribosomal protein genes (rps7, rps11, rps16), transfer RNA genes, as well as matK, accD, ycf2, and multiple nongenic regions from the inverted repeats. None of the identified plastid gene sequences had intact reading frames. Phylogenetic analysis suggests that ~33% of these remnant plastid genes may have been horizontally transferred from the host plant genus Tetrastigma with the rest having ambiguous phylogenetic positions (<50% bootstrap support), except for psaB that was strongly allied with the plastid homolog in Nicotiana. Our inability to identify substantial plastid genome sequences from R. lagascae using multiple approaches—despite success in identifying and developing a draft assembly of the much larger mitochondrial genome—suggests that the parasitic plant genus Rafflesia may be the first plant group for which there is no recognizable plastid genome, or if present is found in cryptic form at very low levels.

Key words: holoparasite, Tetrastigma, plastid, gene loss, horizontal gene transfer, NUPTs.

Introduction

The ability to conduct photosynthesis is one of the defining features of plants. They owe this capacity to endosymbiotic chloroplasts, once free-living cyanobacteria that were assimilated by an ancestral protist ~1.5 billion years ago (Gould et al. 2008). Chloroplasts, one of several plastids that develop from meristematic proplastids, are the site of production and storage of key plant metabolites (Wise 2006). Plastid organelles—which include chromoplasts and amyloplasts—are also involved in fatty acid synthesis, production of tetrapyrroles and aromatic substances, and pigment and starch storage (Neuhau and Emes 2000).

Chloroplasts possess circular DNA genomes, which range size from ~107 to 217 kb (mean of ~152 kb) in the 220 photosynthetic angiosperms that have been examined to date (http://www.ncbi.nlm.nih.gov/genome, last accessed February 8, 2014). Chloroplast genomes (or plastomes) typically encode ~85 proteins and ~45 transfer RNA (tRNA) and ribosomal RNA (rRNA). The genome is a relic of the endosymbiotic origin of this organelle and is reduced in size from its bacterial ancestor, having lost some genes as well as transferring others to the nucleus (Martin 2003). As a consequence of this evolutionary relocation of genes, most proteins required for chloroplast function (~2500–3500) are encoded by nuclear loci (Blanchard and Schmidt 1995).

Chloroplast genome structure is highly conserved across flowering plants, with the interesting exception of parasitic plants. Plant parasitism is an interesting evolutionary adaptation, arising independently at least 12–13 times in flowering plants, with about 1% of all known angiosperm species being parasitic plants (Barkman et al. 2007; Westwood et al. 2010). Hemiparasites, which depend on their hosts only for water
and inorganic nutrients, still retain much of their chloroplast genomes (Braukmann and Stefanovic 2012). Achlorophyllous holoparasites, however, have undergone evolutionary reduction in genome size associated with gene loss. Epifagus virginiana, for example, has a much smaller plastid genome (~70 kb) and is about half of the expected plastome size in autotrophic land plants (Wolfe et al. 1992). This reduced chloroplast (or plastid) genome contains only 42 genes and has lost the loci for photosynthesis and chlororespiration (Wolfe et al. 1992). Despite these reductions in genome size, all plants examined to date, including the nonphotosynthetic parasitic plants, as well as apicomplexan parasites such as Plasmodium, continue to retain even a vestige of a plastid genome (McFadden et al. 1996; Maréchal and Cesbron-Delauw 2001; Krause 2008; Li et al. 2013).

The genus Rafflesia, which belongs to the family Rafflesiaceae (order Malpighiales), is one of the eight known genera of plant holoparasites. Rafflesia is unique to the tropics of Southeast Asia, with some species in the genus producing the largest single flowers in the world, growing up to a meter in diameter. It has no stems, roots, or leaves, with only its massive flower protruding from the roots or stems of its sole host plant, the tropical vine Tetrastigma (Vitaceae) (Nais 2001) (see fig. 1). Nearly one-third of the 30 known Rafflesia species are endemic to the Philippines (Nickrent et al. 1997; Barcelona et al. 2009). Other members of the Rafflesiaceae family include Sapria and Rhizanthes, which are also holoparasites of Tetrastigma (Nais 2001).

Attempts to isolate highly conserved plastid genes from members of the holoparasite genus Rafflesia (Rafflesiaceae) (Nickrent et al. 1997; Davis et al. 2007) have failed, and another study has indicated the possibility of plastid genome loss in Rafflesia leonardii (Nickrent DL, Molina J, Geisler M, Bamber AR, Pelser PB, Barcelona JF, Inovejs SAB, Uy I, Purugganan MD, unpublished data). Here we provide a strong evidence that suggests that the chloroplast/plastid genome is entirely absent in R. lagascae Blanco (see fig. 1), a species found only on the Philippine island of Luzon but nevertheless the most widespread of all the Philippine Rafflesia species (Pelser et al. 2013). With data from Illumina next-generation sequencing, we employ multiple separate techniques for organellar genome assembly. We are able to assemble a draft of the R. lagascae mitochondrial genome at high coverage. We cannot, however, identify an intact plastid genome in R. lagascae, which indicates that members of the parasitic plant genus Rafflesia may be the first plant group shown to have lost its plastid genome.

**Results**

**Draft Sequence Assembly of a Mitochondrial Genome in Rafflesia**

A floral bud of R. lagascae was collected from Cagayan province in the Philippines. Attached only at its base to its Tetrastigma host, Rafflesia tissue was carefully dissected from the host plant, and genomic DNA was extracted from the disk distant from host tissue and enclosed in layers of bracts. Both 100-bp and 3-kb insert libraries were made from genomic DNA and sequenced using Illumina next-generation technology.

Of the approximately 440 million Illumina paired-end (PE) sequencing reads from R. lagascae from both insert libraries, we used two distinct methods to assemble a draft sequence of the mitochondrial genome. First, we used a bait mapping approach by employing previously published data for R. cantleyi (Xi et al. 2013) to assemble the mitochondrion from this species using SOAPdenovo (Luo et al. 2012), and used this assembled R. cantleyi sequence as bait to identify mitochondrial genome sequences from the R. lagascae Illumina reads. These identified reads were then assembled by SOAPdenovo to provide a draft genome assembly of the R. lagascae mitochondrial genome. We were able to identify and assemble ~320.3 kb of the R. lagascae mitochondrial genome (N50 = 4.45 kb); this constitutes a draft assembly with 213 gaps (see fig. 2). The mean sequencing depth coverage across the reference is 349.7×, with a standard deviation of 173.02× (see supplementary fig. S1, Supplementary Material online, for sequence depth coverage across the draft-assembled genome).

Additionally, we used a de novo assembly approach (without any reference genome) to obtain ~1,447,235 sequence contigs of the R. lagascae sequence data using CLC Genomics Workbench (CLC Bio, Aarhus, Denmark), with an N50 = 182 bp. The low N50 for the entire data set is due to either the large size of the R. lagascae nuclear genome or its high repeat content. Despite this, we were able to readily identify ~4,000 sequence contigs containing mitochondrial genome sequence through Blast analysis of assembled sequence contigs against 15 plant mitochondrial genomes. The largest 49 sequence contigs were shown to have about 210× sequence coverage. These contigs ranged from 1.0 to 17.3 kb, with an average size of 6.3 ± 5.0 kb; this represents a total sequence length of 382 kb.

![Fig. 1. Open flower of R. lagascae Blanco. The flower is 15–20 cm in diameter.](https://academic.oup.com/molbev/article-abstract/31/4/793/1110087/794)
We identified 42 known plant protein-coding mitochondrial genes, as well as 60 unique open reading frames (ORFs) (see fig. 2 and supplementary table S1, Supplementary Material online). Phylogenetic analyses of the mitochondrial protein-coding genes show that 7 of 24 genes with clear phylogenetic placement (29%) have these Rafflesia loci allied to *Vitis vinifera* (Vitaceae) or other plant groups, instead of the more closely related *Ricinus communis* (Euphorbiaceae, Malpighiales) (results not shown). The other identified mitochondrial genes show equivocal placement in the phylogenies. These confirm previous findings of rampant horizontal gene transfer (HGT) in this genus (Xi et al. 2012, 2013; Nickrent DL, Molina J, Geisler M, Bamber AR, Pelser PB, Barcelona JF, Inovejas SAB, Uy I, Purugganan MD, unpublished data).

### Searching for a Plastid Genome

We attempted to recover plastid sequences from *R. lagascae* using several approaches. First, we mapped the sequence contigs obtained from the CLC Genomics Workbench assembly method onto the chloroplast genomes of *Ricinus* and *Vitis*, using the program Geneious R6 (Biomatters, Auckland, New Zealand). This yielded 17 sequence contigs (G1–G17; table 1) that mapped to the plastid genomes of these two species, with a total length of ~3.9 kb. We also conducted a BlastN search of all conserved plastid genes available from GenBank against the assembled CLC sequence contigs, which identified additional 26 contigs with significant (e-value < $1 \times 10^{-10}$) hits (table 1). In addition, we generated profile hidden Markov models (HMMs; Henderson et al. 1997) from alignments of conserved plastid genes.

![Draft structure of *R. lagascae* mitochondrial genome. The gene positions indicated are based on the assumption of synteny with *Ricinus* (GenBank accession number HQ874649; Rivarola et al. 2011). The coordinates and encoded products of the specific genes are shown in supplementary table S1, Supplementary Material online. Although depicted as a complete circular genome, it should be stressed that this is a draft assembly with 213 gaps and that portions of the mitochondrial sequence remain unassembled.](https://academic.oup.com/molbev/article-abstract/31/4/793/1110087/795)
Table 1. Identified Plastid Sequences in Rafflesia lagascae Including Noncoding Sequences in Inverted Repeat (IR) Regions.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Size (bp)</th>
<th>Method of Recovery</th>
<th>Contig Number</th>
<th>Phylogenetic Alliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>accD</td>
<td>136</td>
<td>BlastN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>atpA</td>
<td>158</td>
<td>BlastN</td>
<td>1,105,137</td>
<td></td>
</tr>
<tr>
<td>atpB</td>
<td>127</td>
<td>BlastN</td>
<td>568,017</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>138</td>
<td>BlastN</td>
<td>690,602</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>160</td>
<td>BlastN</td>
<td>552,527</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>163</td>
<td>Geneious</td>
<td>G14</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>190</td>
<td>Geneious</td>
<td>G11</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>199</td>
<td>Geneious</td>
<td>G3</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>207</td>
<td>BlastN</td>
<td>435,549</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>219</td>
<td>BlastN</td>
<td>1,355,988</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>227</td>
<td>Geneious</td>
<td>G15</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>274</td>
<td>BlastN</td>
<td>426,447</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>299</td>
<td>Geneious</td>
<td>G6</td>
<td></td>
</tr>
<tr>
<td>IR, trnA-UGC</td>
<td>104</td>
<td>Geneious</td>
<td>G8</td>
<td></td>
</tr>
<tr>
<td>IR, trnA-UGC</td>
<td>113</td>
<td>Geneious</td>
<td>G7</td>
<td></td>
</tr>
<tr>
<td>IR, trnA-UGC</td>
<td>168</td>
<td>Geneious</td>
<td>G16</td>
<td></td>
</tr>
<tr>
<td>IR, trnA-UGC</td>
<td>330</td>
<td>BlastN</td>
<td>859,744</td>
<td></td>
</tr>
<tr>
<td>IR, trnA-UGC</td>
<td>484</td>
<td>Geneious</td>
<td>G10</td>
<td></td>
</tr>
<tr>
<td>IR, trn-GAU</td>
<td>496</td>
<td>BlastN</td>
<td>12,131</td>
<td></td>
</tr>
<tr>
<td>IR, trnV-GAC</td>
<td>145</td>
<td>Geneious</td>
<td>G13</td>
<td></td>
</tr>
<tr>
<td>IR, trnV-GAC</td>
<td>178</td>
<td>Geneious</td>
<td>G5</td>
<td></td>
</tr>
<tr>
<td>IR, trnV-GAC</td>
<td>178</td>
<td>BlastN</td>
<td>1,147,528</td>
<td></td>
</tr>
<tr>
<td>matK</td>
<td>227</td>
<td>BlastN</td>
<td>710,409</td>
<td></td>
</tr>
<tr>
<td>ndhB</td>
<td>118</td>
<td>Geneious</td>
<td>G4</td>
<td></td>
</tr>
<tr>
<td>ndhB</td>
<td>247</td>
<td>BlastN</td>
<td>1,167,865</td>
<td></td>
</tr>
<tr>
<td>petB</td>
<td>160</td>
<td>Geneious</td>
<td>G2</td>
<td></td>
</tr>
<tr>
<td>petG</td>
<td>116</td>
<td>BlastN</td>
<td>832,312</td>
<td></td>
</tr>
<tr>
<td>psbA</td>
<td>211</td>
<td>BlastN</td>
<td>76,772</td>
<td></td>
</tr>
<tr>
<td>psbA</td>
<td>315</td>
<td>BlastN</td>
<td>662,630</td>
<td></td>
</tr>
<tr>
<td>psbA</td>
<td>324</td>
<td>BlastN</td>
<td>972,383</td>
<td></td>
</tr>
<tr>
<td>psbD</td>
<td>161</td>
<td>BlastN</td>
<td>1,131,304</td>
<td></td>
</tr>
<tr>
<td>psbD</td>
<td>280</td>
<td>BlastN</td>
<td>39,796</td>
<td></td>
</tr>
<tr>
<td>psbZ</td>
<td>137</td>
<td>BlastN</td>
<td>1,038,399</td>
<td></td>
</tr>
<tr>
<td>rbcL1</td>
<td>395</td>
<td>BlastN</td>
<td>127,053</td>
<td></td>
</tr>
<tr>
<td>rbcL2</td>
<td>647</td>
<td>BlastN</td>
<td>170,975</td>
<td></td>
</tr>
<tr>
<td>rps11</td>
<td>113</td>
<td>HMM</td>
<td>322,006</td>
<td></td>
</tr>
<tr>
<td>rps16</td>
<td>168</td>
<td>BlastN</td>
<td>641,235</td>
<td></td>
</tr>
<tr>
<td>rps7</td>
<td>119</td>
<td>HMM</td>
<td>864,624</td>
<td></td>
</tr>
<tr>
<td>rrm16 (IR)</td>
<td>191</td>
<td>Geneious</td>
<td>G12</td>
<td></td>
</tr>
<tr>
<td>rnm23 (IR)</td>
<td>139</td>
<td>Geneious</td>
<td>G17</td>
<td></td>
</tr>
<tr>
<td>ycf2 (IR)</td>
<td>203</td>
<td>HMM</td>
<td>113,164</td>
<td></td>
</tr>
</tbody>
</table>

Note.—Certain IR sequences may appear multiple times because they belong to the inverted repeats region of the plastid genome, whereas genes like ndhJ were identified by different methods.

*aContig number is the specific contig out of the approximately 1.4 million contigs from CLC, except for those prefixed with “G,” which were derived from Geneious.

*bTaxa and genome/s (mt, mitochondria; cp, chloroplast; nuc, nuclear) to which Rafflesia was shown strongly associated with (>50% BS, compare with fig. 4) in phylogenetic analyses; sequences marked with “—” had ambiguous phylogenetic positions. Multiple taxa/genome associations represent polytomous nodes (>50% BS in which Rafflesia is embedded in.

*cCorresponding phylogenies for these sequence contigs that show Rafflesia in unequivocal positions are provided as Supplementary Material (taxa in the rps11 phylogeny represent the only significant hits recovered).

identified three more contigs with significant hits (e-value < 1 e−10) (table 1). Together, these approaches identified a total of 46 putative plastid sequence contigs with an average length of 242 bp and a total length of 11.5 kb.

To ensure that our failure to identify a plastid genome was not due to a problem of the genome assembly method, we identified 925 Illumina 100-bp PE reads (out of ~214 million reads) that directly mapped to plastid sequences found in GenBank. These sequences were then de novo assembled using SOAPdenovo, another sequence assembly program for short-read sequencing data (Luo et al. 2012). Aside from those sequences already identified by the previous methods, we found additional five sequence fragments, but they appear embedded in the assembled mitochondrial genome of R. lagascae.

Unlike the high sequence coverage for the draft assembly of the mitochondrial genome (~350 x), the mean sequence coverage for the plastid sequence fragments we identified was substantially lower at 1.48 x ± 1.26 x reads. The anomalously low sequencing read depth coverage contrasts with the several hundred copies of plastid genomes that should exist within plant cells (Bock 2007). As a comparison with the normal expectation in different species, we used Illumina whole genome re-sequencing data to demonstrate that chloroplast/plastid genome sequencing depth coverage was nearly equal to or exceeded that of mitochondrial genome in leaf tissue in Oryza, Phoenix, and Arabidopsis, and light-grown single-cell culture in the algae Chlamydomonas (see fig. 3). The sequencing reads in these species covered >98% of the bases in these organellar genome sequences.

Because the plastid copy number is reduced in nonphotosynthetic tissues like roots (Isono et al. 1997), it is possible that lower levels of plastid DNA in the nonphotosynthetic Rafflesia may be the source of our inability to identify a plastid genome in this species. As a further positive control, we therefore obtained ~62 and ~104 million 100-bp PE Illumina sequencing reads from root DNA in Oryza glaberrima and Arabidopsis thaliana, respectively, and examined relative levels of plastid and mitochondrial genome copies in these nonphotosynthetic tissues. Based on sequencing depth coverage from root genomic DNA (see fig. 3), the plastid genome in O. glaberrima is ~66% that of mitochondrial genome levels, whereas in A. thaliana the plastid genome is found at ~3.5-fold greater levels than mitochondrial genome (down from an ~17-fold greater level in leaves). Our data demonstrate a reduction in plastid genome levels in the nonphotosynthetic root compared with leaf tissues but nevertheless show that plastid genome levels can be substantial in nonphotosynthetic tissues.

Fragments of the Plastid Genome

Although we could not identify an intact plastid genome, we did find small fragments of plastid genes that ranged in size from 104 to 1,026 bp. We recovered short segments of 17 protein-coding genes (including ribosomal proteins), two rRNA and three tRNA genes, as well as ten intergenic
sequences that are found in inverted repeat regions of the chloroplast genome in other species (see table 1). No full gene sequences were identified. None of the plastid sequence contigs had intact reading frames when aligned with coding sequences of plastid genes from photosynthetic taxa (results available upon request).

To determine whether any of the recovered plastid sequences were expressed, we also mapped the sequencing reads from the R. cantleyi transcriptome library (accession SRA052224) to these R. lagascae sequence fragments (Xi et al. 2012). The greatest number of reads mapped were 23 singleton reads from the partial ribosomal rrn16 (~150 bp) fragment. These results indicate that there is no significant expression for any of these putative plastid-derived sequences, at least in the actively developing floral bud tissue of this parasitic plant.

Phylogenetic Analysis of Remnant Rafflesia Plastid Gene Sequences

Phylogenetic analysis of the 46 plastid sequence fragments that we identified show that 15 of these Rafflesia sequence fragments are allied with Vitis with >50% BS (table 1, fig. 4 and supplementary data S2, Supplementary Material online). The rest of the plastid sequences show Rafflesia in equivocal positions (BS <50%).

Four sequence fragments, including ndhJ (fig. 4A), depict Rafflesia plastid sequences associated with both Vitis’ nuclear and plastid sequences. Five other sequences such as a noncoding sequence from one of the inverted repeats were more similar to a Vitis’ nuclear sequence than they are to plastid sequences (fig. 4B; supplementary data S2, Supplementary Material online; table 1). Only Rafflesia’s rrn23 was solely with a Vitis chloroplast sequence (fig. 4C). A 136-bp fragment from the accD gene had Rafflesia grouping with nuclear, plastid and mitochondrial copies (76% BS; supplementary data S2, Supplementary Material online, and table 1).

No plastid sequence was found to be phylogenetically associated with Ricinus or Hevea, the closest relatives to Rafflesia with available organelar genome sequencies. However, the Rafflesia psaB gene fragment grouped with the Nicotiana homolog with 80% BS (fig. 4D).

The anomalous phylogenetic placement of these remnant Rafflesia plastid genes may arise from contamination from the DNA of the host plant Tetrastigma. To test for contamination of Tetrastigma in the R. lagascae DNA extract, we used barcoding primers to amplify the rbcL gene (Kress et al. 2009) in these two species. We were unable to polymerase chain reaction (PCR)-amplify rbcL from R. lagascae, although this gene was easily amplified from the host Tetrastigma and from other evolutionary divergent photosynthetic taxa from the asterid and rosid families. Interestingly, we were able to recover two nonoverlapping segments of rbcL sequence (rbcL1, rbcL2) from the Illumina sequence contigs (~1 kb in size) from R. lagascae (table 1), but these sequences are diverged in the barcoding primer sequence regions (Kress et al. 2009) that are normally conserved across multiple divergent autotrophic angiosperm taxa. Like the other recovered plastid sequences, Rafflesia’s rbcL contains premature stop codons.

Discussion

Possible Loss of the Plastid Genome in a Parasitic Plant

The parasitic plant lifestyle affords an intimate connection between a parasite and its host plant. As in many parasites, this can lead to subsequent relaxation of selection pressure to maintain key genes in the parasite as they become dependent on host plants for crucial functions (Bromham et al. 2013; Wicke et al. 2013). Moreover, the close connection to the host can also lead to genetic transfer of information to the parasite (Davis and Wurdack 2004; Xi et al. 2012, 2013).

The evolution of the chloroplast genome in parasitic plants, particularly nonphotosynthetic holoparasites, can lead to significantly reconfigured plastomes (Wicke et al. 2013). In these plants many photosystem and energy production genes are lost from the plastome (Krause 2008; Li et al. 2013; Wicke et al. 2013). The 45.6-kb plastome of Conopholis americana (Orobanchaceae) is the smallest published plastid genome to date (Colwell 1994; Wicke et al. 2013). Though devoid of genes expected to be present in autotrophic plants,
it still maintains genes that are conserved in all previously sequenced plastomes, like genes for rRNA, some genes for ribosomal proteins, tRNA (e.g., trnE and trnFM) and the essential genes clpP and ycf2 (Wicke et al. 2013). Other parasitic plants and even the protist Plasmodium, descendant of the same photosynthetic protist as plants, still show remnant plastid genomes (McFadden et al. 1996; Maréchal and Cesbron-Delaun 2001; Krause 2008; Li et al. 2013).

It is clearly challenging to prove the complete absence of a plastid genome (Keeling 2010). Nevertheless, our inability to
identify substantial plastid genome sequences from \textit{R. lagascae} using multiple approaches, despite success in identifying and developing a draft assembly of the much larger mitochondrial genome, strongly suggests that there may be no recognizable plastid genome in the parasitic plant \textit{R. lagascae}, or if present in cryptic form, is at very low levels. Moreover, a similar result has also been observed for \textit{R. leonardi} (Nickrent DL, Molina J, Geisler M, Bamber AR, Pelser PB, Barcelona JF, Inovejas SAB, Uy I, Purugganan MD, unpublished data), which our study now reinforces with multiple lines of evidence. These results together suggest that plastid genome loss may be shared across multiple \textit{Rafflesia} species, and this putative loss is likely an evolutionary consequence of the very ancient onset of parasitism in the family (Xi et al. 2013), perhaps dating back to mid-Cretaceous (Molina J, unpublished data).

It is possible for organelles to lose their genomes. This has already been documented in some anaerobic ciliates, trichomonads, and fungi, which possess hydrogenosomes, genomeless organelles that produce molecular hydrogen and are derived from mitochondria (Van der Giezen et al. 2005). However, the loss of the plastome has not yet been demonstrated in plants but deemed possible (Palmer 1997). Plastids have been secondarily lost outside the flowering plant lineage, including the protozoan trypanosomes (Martin and Borst 2003). However, in a recent study by Braukmann et al. (2013) on some species of the parasitic flowering plant \textit{Cuscuta} (Convulvulaceae), difficulty has been reported in detecting plastid rRNA (rrn) genes that are highly conserved elements in plant plastomes even in heterotrophic species (Krause 2011; Wicke et al. 2011; Li et al. 2013). Accordingly, Braukmann et al’s (2013, p. 9) observations led them to conclude that holoparasitic \textit{Cuscuta} may have “reached the same or similar evolutionary endpoint, where the very presence of a plastid genome is questionable.”

Despite the possible loss of the chloroplast and its genome, there still remain plastid-like structures in \textit{Rafflesia}. Images from ultrastructural analysis from the congeneric \textit{R. philippensis} using transmission electron microscopy (TEM) demonstrate that \textit{Rafflesia} (Nickrent DL, Molina J, Geisler M, Bamber AR, Pelser PB, Barcelona JF, Inovejas SAB, Uy I, Purugganan MD, unpublished data; for additional images, see fig. 5) does contain plastid-like compartments with homogeneous stroma (Renzaglia K, personal communication). None of these structures have the distinctive lamellar/endomembrane system found in all types of plastids. These suggest that \textit{Rafflesia} retains plastid compartments for certain metabolic functions, even in the apparent absence of a plastid genome.

There are two possible scenarios for the evolution of these genome-less plastids. One is that any relevant genes still necessary for metabolic function have relocated to the nucleus and/or mitochondria. Another possibility is that these \textit{Rafflesia} plastids were originally obtained from the host, with subsequent translocation of host-encoded plastid genes to the nucleus and eventually degenerating as pseudogenes. The latter possibility may explain the large proportion of remnant nonfunctional plastid gene sequences in \textit{R. lagascae} that are phylogenetically allied with \textit{Vitis} nuclear sequences (table 1; supplementary data S2, Supplementary Material online). Such host-derived plastids have been observed in certain parasitic red algae (Goff and Coleman 1995) as well as in natural grafts of sexually incompatible species of \textit{Nicotiana} (Stegemann et al. 2012). These two possibilities are not mutually exclusive, and the precise genetic basis for the maintenance of these \textit{Rafflesia} plastid-like structures must await more detailed analysis of fully assembled nuclear genomes.

Interestingly, osmiophilic plastoglobules or carotenoid bodies, which are typical inclusions of chromatplasts in colored flowers (Cama et al. 1995), were also not seen in the \textit{Rafflesia} plastid-like structures, and thus, may not be the primary source of the bright red-orange coloration characteristic of \textit{Rafflesia} species. Instead, there were very large vacuoles observable in \textit{Rafflesia} ramenta filled with osmiophilic material, which may be phenolic compounds or terpenoids that also appear as electron-dense material in other TEM studies of trichomes (Sacchetti et al. 1999; Wen-Zhe et al. 2002). These plant terpenoids can serve as precursors for a diversity of plant metabolites such as carotenoids and anthocyanins and volatiles responsible for flower color and odor (Tanaka et al. 2008).

Several hypotheses have been proposed with respect to what biochemical constraints on plastid genome size or its complete loss might exist owing to the need to retain particular genes required for metabolism (Bungard 2004; Barbrook et al. 2006). The essential tRNA hypothesis states that plastid-encoded trnE is considered essential for heme biosynthesis (a component of the mitochondrial P450 cytochromes) and could not be easily replaced by a cytosolic tRNA. It may be that \textit{Rafflesia} continues to retain plastids for various metabolic functions, but the genes that encode for these are found in the nucleus or in the mitochondria as in the case of a trnE sequence recovered in \textit{R. leonardi} (Nickrent DL, Molina J, Geisler M, Bamber AR, Pelser PB, Barcelona JF, Inovejas SAB, Uy I, Purugganan MD, unpublished data). The fates of other essential plastid genes, such as the clpP and ycf2 loci (Wicke et al. 2013), are unknown, and must await detailed genomic and biochemical studies on these parasitic plant species.

\textbf{Vestiges of the Missing Plastid Genome}

The small regions we could successfully identify in \textit{R. lagascae} as putative plastid sequences represent less than 10% of the chloroplast genome of photosynthetic plants, and do not have intact reading frames and so are likely nonfunctional. These results would make \textit{Rafflesia} arguably the plant with the smallest amount of plastid “genome” sequence that has been observed thus far (Krause 2008; Li et al. 2013). Given the low sequencing read coverage for these gene fragments (<2× coverage), it is likely that these remnant plastid sequences are located in the nuclear genome, similar to what has been observed for nuclear-integrated chloroplast genes and/or nuclear plastid DNAs (NUPTs) observed in other plant species (Blanchard and Schmidt 1995; Kleine et al. 2009).
Phylogenetic analyses of the remnant plastid gene fragments in *Rafflesia* show that many of these genes have anomalous but strongly supported (>50% BS support) phylogenetic placements that suggest they are the products of HGT. Most of these genes grouped phylogenetically with homologs in *Vitis* (which belongs to the same family as the host species *Tetraestigma*), rather than either *Ricinus* or *Hevea*, which together are more closely related to *Rafflesia*. There is a possibility that these may be due to contamination of the *Rafflesia* DNA, but control experiments using universal *rbcL* barcoding primers suggest that this is unlikely. The anomalous phylogenetic placement of remnant *Rafflesia* plastid sequences may also arise from a complex history of gene duplication and extinction, although we think that HGT is a more parsimonious explanation.

Although these other possibilities cannot be completely ruled out, we feel that our results suggest that ~33% of the plastid loci we have identified in *R. lagascae* may have been horizontally acquired, most likely from its host as a result of parasitism. It has already been demonstrated in other parasitic plants that the plasmodesmal continuity between host and parasite allows for molecular movement, including that of genetic material (Birschwilk et al. 2006; Roney et al. 2007; Talianova and Janousek 2011). Rampant HGT between *Rafflesia* and *Tetraestigma* involving several nuclear and mitochondrial genes has also been repeatedly shown (Xi et al. 2012, 2013). Individual phylogenies of chloroplast sequences recovered from the confamilial *Sapria* have also exhibited a closer relationship to *Vitis* than to *Ricinus* (Xi et al. 2013). A complete genome sequence of the *Rafflesia* nuclear genome will be necessary in order to conclusively determine the fate of the genes that would normally reside in the plastid as well as examine the other molecular evolutionary consequences of the obligate parasitic lifestyle of this enigmatic yet fascinating plant genus.

**Materials and Methods**

**Illumina Whole-Genome Sequencing**

A flower bud of *R. lagascae* was collected from Sitio Kanapawan, Barangay Bolos Point, Gattaran Municipality, Cagayan Province, Philippines in 2010 (collection number Nickrent 5791). All necessary collecting (Philippine Department of Environment and Natural Resources Gratuitous Permit 193, and subsequent renewals GP 202 and 217), transport, and export permits were obtained.

A sizeable *Rafflesia* bud, attached only at its base to its host, *Tetraestigma*, was carefully dissected from the host plant. Genomic DNA was extracted from a portion of the disk, which is sufficiently distant from host tissue and enclosed in layers of bracts. DNA extraction was performed following Nickrent et al. (2004). Contamination of *R. lagascae* DNA by *Tetraestigma* was tested using standard PCR amplification with degenerate *rbcL* barcoding primers (Kress et al. 2009). PCR amplification was positive in *Tetraestigma* but negative in *R. lagascae* (as well as in *R. leonardi*). The extracted *R. lagascae* DNA was submitted to Ambry Genetics (Aliso Viejo, CA) for Illumina next-generation sequencing, using both a 100-bp and a 3-kb insert size library on Illumina HiSeq 2000.
Sequencing reads are deposited with the NCBI Sequence Read Archive (SRX434531).

Mitochondrial Genome Assembly and Analysis
Illumina sequencing reads from R. lagascae were assembled using two approaches. In one approach, we used a combination of SOAPdenovo (Luo et al. 2012) and reference-assisted mapping to assemble R. lagascae’s organellar genomes. We first assembled the mitochondrial gene sequence of the closely related species, R. cantleyi (Xi et al. 2013), whose Illumina reads were previously published and available in NCBI (accession SRR629613) and used this assembly of its mitochondrial genome as reference bait sequence. We then collected all the R. lagascae 100-bp PE reads that mapped to the R. cantleyi mitochondrial genome and used these to de novo assemble R. lagascae’s mitochondrial genome in SOAPdenovo. To annotate its mitochondrial genome, we used Mitofy (Alverson et al. 2010). Annotated ORFs with their accompanying numbers are based on the NCBI GenBank database. HGT of the mitochondrial genes was detected using phylogenetic analyses in MEGA5 (Tamura et al. 2010) to the resulting contigs.

Plastid Sequence Identification and Analysis
To identify sequence contigs comprising the plastid genome, we employed multiple approaches. First, sequence contigs from R. lagascae identified by the CLC Genomics Workbench were mapped using Geneious R6 (Biomatters Ltd., Auckland, New Zealand). These contigs were first mapped to the full mitochondrial genomes of V. vinifera (GenBank accession number FM179380) and Ri. communis (HQ874649) to eliminate mitochondrial sequences, and all unmapped reads were then collected and mapped to the chloroplast genomes of Vitis (DQ424856.1) and Ricinus (JF937588.1). Ricinus communis, like R. lagascae, is in the Malpighiales, and is the closest species to Rafflesia for which mitochondrial genome sequences are available. Vitis vinifera is the closest species to Tetrastigma for which both mitochondrial and chloroplast genome sequences are also available. Second, we conducted a BlastN search (e-value < $e^{-10}$) of all the conserved plastid genes found in angiosperms available from GenBank against the CLC-assembled sequence contigs. We also generated profile HMM (Henderson et al. 1997) from alignments of conserved plastid genes using HHMER (http://hmmr.janelia.org/, last accessed February 8, 2014), which develops probabilistic models (profile HMMs) and can detect more remote similarities in sequence searches. We then mapped all these plastid sequence contigs identified (Geneious, BlastN, HMM) to the chloroplast genome of another species in the Malpighiales, Hevea brasiliensis (GenBank accession number NC_015308), using Geneious R6 (Biomatters, Ltd.). In a third approach, we mapped the 100-bp PE reads using BWA (Burrows–Wheeler alignment) (Li and Durbin 2010) and SAMtools (Sequence Alignment/Map) (Li et al. 2009) to angiosperm plastid sequences from GenBank, and then the resulting mapped reads were de novo assembled using SOAPdenovo (Luo et al. 2012).

Putative plastid sequences identified earlier were then aligned with homologous regions from other taxa whose chloroplast genomes are available in GenBank using default parameters in the Multiple Alignment using Fast Fourier Transform (MAFFT) program (Katoh and Toh 2008) and visually checked (supplementary data S1, Supplementary Material online). The alignments were analyzed phylogenetically in MEGAS (Tamura et al. 2011) using maximum likelihood with 100 bootstrap replicates and applying the best substitution model with the lowest Bayesian information criterion scores (supplementary data S2, Supplementary Material online).

To determine whether any of these putative plastid sequences are expressed, we also mapped the reads from the R. cantleyi transcriptome library (accession SRA052224) (Xi et al. 2012) using Bowtie 1.0.0 (Langmead et al. 2009), TopHat2 (Kim et al. 2013), and Cufflinks (Trapnell et al. 2010) to the resulting contigs.

Comparative Levels of Mitochondrial and Plastid Genomes in Algal and Flowering Plant Species
To compare our results in Rafflesia with the relative levels of mitochondrial and chloroplast genome DNA in various species, we used whole genome resequencing data from three monocot species (Phoenix dactylifera, O. sativa, and O. glaberrima), one eudicot species (A. thaliana), and one photosynthetic algae (Chlamydomonas reinhardtii). DNA from these species was isolated from single-cell cultures (for C. reinhardtii), leaf or shoot tissue (P. dactylifera, O. sativa, O. glaberrima, and A. thaliana), or nonphotosynthetic root tissue (O. glaberrima and A. thaliana). Libraries were constructed with 100-bp insert sizes using Illumina Standard DNA Library or Nextera kits (Illumina, San Diego, CA) and were sequenced as either 50-bp or 100-bp PE reads using Illumina HiSeq 2000 at NYU Center for Genomics and Systems Biology in New York or Abu Dhabi to obtain between 50 and 250 million reads per sample. For all these sequences, we mapped the reads using BWA (Li and Durbin 2010) and SAMtools (Sequence Alignment/Map) (Li et al. 2009), with the species genome sequence available in GenBank as a reference sequence.

Supplementary Material
Supplementary data S1 and S2, figure S1, and table S1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
The sequencing data have been deposited in the NCBI Sequence Read Archive (accession number SRX434531). The authors are grateful to Mary Ann Cajano, Salud Pangan, Filipinas Natividad, Karen Renzaglia, Steve Rounsley, Joseph Morin, Sandra Yap, Neda Barghi, Steven Sullivan, Jane Carlton, Arturo Lluisma, and Cynthia Saloma for their help and support in various aspects of the project. The authors are also grateful for the help provided by the Department of
Environment and Natural Resources-Protected Areas and Wildlife Bureau (DENR-PAWB) of the Philippines, and thank Director Theresa Mundita S. Lim, Josefina de Leon, and Cecile Garcia for facilitating issuance of collecting permits. For assistance in collecting R. lagascae, they thank Jovencio Payba, Provincial Environment and Natural Resources Office (PENRO)/DENR-Region 2, Apolinario U. Unarce, Sumper Aresta, Tabuc, Rolando Echanique and Alfredo Gabriel and The Cagayan Valley Partners in People Development (CAVAPPED) through its President/CEO Perla A. Visororo. For assistance in collecting R. philippensis (Mt. Banahaw), the authors thank Jerry R. Mendua, Ananias M. Cahilo Sr., Romeo R. Diale, Angeles Coronado, Salud Pangan, and Mary Ann O. Cajano. This work was supported by a start-up grant from LIU and a Visiting Professorship grant from the University of the Philippines to J.M., and grants from the US National Science Foundation Plant Genome Research Program (IOS 1126971) and the NYU Abu Dhabi Institute to M.D.P. This article is dedicated to the memory of Leonard Co, a leading Philippine botanist and Rafflesia investigator, who passed away while conducting fieldwork.

References


