A Plastid in the Making: Evidence for a Second Primary Endosymbiosis

Birger Marin, Eva C. M. Nowack, and Michael Melkonian

Botanisches Institut, Lehrstuhl I, Universität zu Köln, Gyrhofstr. 15, 50931 Köln, Germany

Submitted August 18, 2005; Accepted September 14, 2005
Monitoring Editor: Keith Gull

One of the major steps in the evolution of life was the origin of photosynthesis in nucleated cells underpinning the evolution of plants. It is well accepted that this evolutionary process was initiated when a photosynthetic bacterium (a cyanobacterium) was taken up by a colorless host cell, probably more than a billion years ago, and transformed into a photosynthetic organelle (a plastid) during a process known as primary endosymbiosis. Here, we use sequence comparisons and phylogenetic analyses of the prokaryotic rDNA operon to show that the thecate, filose amoeba *Paulinella chromatophora* Lauterborn obtained its photosynthetic organelles by a similar but more recent process, which involved a different cyanobacterium, indicating that the evolution of photosynthetic organelles from cyanobacteria was not a unique event, as is commonly believed, but may be an ongoing process.

© 2005 Elsevier GmbH. All rights reserved.

Key words: cyanobacteria; molecular synapomorphies; *Paulinella chromatophora*; plastid phylogeny; primary endosymbiosis; rDNA operon.

Introduction

The thecate, filose amoeba *Paulinella chromatophora* was discovered by Robert Lauterborn in 1895 in wetlands of the upper Rhine valley and apparently has a global distribution in sediments of ponds and lakes (Melkonian and Mollenhauer 2005). Its most distinctive feature are (one or two) sausage-shaped blue-green inclusions (Fig. 1a) that support the photoautrophic existence of the amoeba (unlike its marine relatives, this *Paulinella* species is not known to feed on bacteria; Hannah et al. 1996; Johnson et al. 1988; Vørs 1993). The nature of the photosynthetic inclusions of *P. chromatophora* has been debated for 100 years (Geitler 1927; Keeling 2004a; Lauterborn 1895; Lederberg 1952; Pascher 1929); were they just food particles, intracellular cyanobacterial symbionts, or genuine photosynthetic organelles? Whereas the first alternative can now be refuted (a clonal, photoautotrophic culture of *P. chromatophora* has been growing in the authors’ laboratory for more than 10 years), the last two alternatives still need to be addressed. For convenience sake, we will refer to these inclusions as photosynthetic organelles. Convincing evidence has been obtained that plastids shared a monophyletic origin among cyanobacteria (e.g. Helmchen et al. 1995; Keeling 2004a; Rodriguez-Ezpeleta et al. 2005), often...
equated with a single primary endosymbiotic event. One type of plastid (the cyanelles of the Glaucoplantae) retains remnants of the peptidoglycan cell wall of its prokaryote ancestor making cyanelles the only osmotically stable plastids. Interestingly, the photosynthetic organelles of *P. chromatophora* also contain a residual peptidoglycan cell wall (sandwiched between two envelope membranes; Kies 1974) raising the possibility that both organelles had a monophyletic origin. The two hosts (the Glaucoplantae and *P. chromatophora*), however, are not monophyletic and belong to two different eukaryote ‘supergroups’ (Plantae and Rhizaria; Bhattacharya et al. 1995; Cavalier-Smith 2002; Keeling 2004a).

To clarify the phylogenetic position of the photosynthetic organelles of *P. chromatophora*, we determined almost complete sequences of the prokaryotic rDNA operon of *P. chromatophora*, and the plastids of two glaucophytes and two structurally simple red algae, and performed phylogenetic analyses of the rDNA operon including diverse eubacterial/cyanobacterial and plastid sequences.

**Results and Discussion**

1. The phylogenetic tree (Fig. 1a) shows that all plastid sequences are monophyletic (with 100%/1.00 support in all analyses; clade 4). (Fig. 1 a) displays cyanobacteria with Form 1A carboxysomes. (Fig. 1 a) shows that all plastid sequences are monophyletic (with 100%/1.00 support in all analyses; clade 4).

2. The *P. chromatophora* sequence does not belong to the plastid clade but forms a monophyletic lineage (clade 6) with a group of cyanobacteria (strains of *Prochlorococcus* and *Synechococcus*). The latter clade is unambiguously (100%/1.00) supported by all methods of analysis, and is further character-ized by two unique signatures (non-homoplasticous synapomorphies; Marin et al. 2003) in the SSU and LSU rRNA (16S rRNA and 23S rRNA) molecule, respectively (Fig. 1b).

3. *Paulinella chromatophora* is the earliest divergence in this cyanobacterial clade and is sister to a well-supported subclade (no. 7; Fig. 1a) containing *Prochlorococcus* and marine strains of *Synechococcus*. The *Prochlorococcus/Synechococcus* subclade is also supported by a non-homoplasticous molecular synapomorphy in the LSU rRNA molecule (Fig. 2).

4. The phylogenetic resolution within the cyanobacteria is generally improved over phylogenies using SSU rDNA alone (e.g. Fewer et al. 2002; Honda et al. 1999; Turner et al. 1999; Wilmotte and Herdman 2001) or a concatenated data set of 50 plastid-encoded proteins (Rodrı´guez-Ezepeleta et al. 2005) although taxon sampling is still limited. In particular, we note that a lineage of cyanobacteria (clade no. 3; Fig. 1a) is moderately supported that may represent the sister group of the plastid lineage (Fig. 1a). This lineage contains only cyanobacteria with Form 1B RubisCo (and \(\beta\)-carboxysomes; Badger et al. 2002), whereas the sister lineage of the photosynthetic organelle of *P. chromatophora* (subclade no. 7; Fig. 1a) displays cyanobacteria with Form 1A RubisCo and \(x\)-carboxysomes.

5. Cyanelles and rhodoplasts are moderately supported as sister groups, while both are sister to the chloroplasts (Fig. 1a). The phylogenetic relationship between the three types of simple (primary) plastids is still unsettled (e.g. Bhattacharya et al. 2003; Keeling 2004b; McFadden 2001; Moreira and Philippe 2001; Palmer 2003; Palmer et al. 2004).

*Figure 1.* a. Phylogenetic evidence for an independent cyanobacterial origin of the photosynthetic organelle of *Paulinella chromatophora*, significantly separated from the plastid lineage. The maximum-likelihood tree (GTR+I+\(\Gamma\); \(I = 0.2661; \Gamma = 0.7513\)) was based on almost-complete rDNA operon sequences (genes encoding SSU rRNA, tRNA-Ile, tRNA-Ala, and LSU rRNA; 4104 aligned characters); numbers at branches are bootstrap values [neighbor joining and maximum parsimony (>60%)] and Bayesian posterior probabilities >0.90 (branches in bold have 100/100/1.00 support). Strain designations when available (for abbreviations, see Methods) and EMBL/GENBANK accession numbers are given. Sequences determined new to this study are in bold. Encircled numbers (1–7) denote clades analyzed by single gene- and partitioned analyses (see Results and Discussion, and *Table 1*). The inset shows a light micrograph of a cell of *Paulinella chromatophora*, displaying the theca consisting of silica scales, and a single photosynthetic organelle undergoing division. b. Unique molecular signatures (non-homoplasticous synapomorphies according to Marin et al. 2003) in highly conserved regions of the SSU and LSU rRNA molecules. Synapomorphies characteristic for the clade (no. 6 in Fig. 1a) comprising *P. chromatophora*, *Synechococcus* and *Prochlorococcus* were found in the 3′-terminus of the SSU rRNA (second to last nt), and in the GTPase-associated center of the LSU rRNA (Helix 1082; Cannone et al. 2002). Secondary structure drawings based on the *Paulinella* sequence.
Rodríguez-Ezpeleta et al. (2005). Whereas phylogenetic analyses of protein-coding genes or SSU rDNA alone often led to the conclusion that the cyanelles are the earliest diverging plastids, other topologies have also been recovered; e.g., Palmer et al. (2004) cited unpublished observations of Turner et al. suggesting a strong support for a cyanelle/rhodoplast sister group in LSU rDNA phylogenies (although *Cyanophora paradoxa* was the only glaucoplant used in these analyses). Here, we report similar results (Fig. 1a), but note that by adding more cyanelle rDNA sequences (in total three), the support values for a cyanelle/rhodoplast clade were somewhat lower compared to trees containing fewer cyanelle and rhodoplast sequences (unpublished observations).

In addition to analyses of the concatenated data set, both rRNA genes (SSU rDNA and LSU rDNA) were analyzed separately, and a partitioned Bayesian analysis (see Methods) was performed to ensure that data partitions produced similar tree topologies (Table 1). Except for the branching pattern of cyanelles, rhodoplasts and chloroplasts (clade no. 5 in Fig. 1a and Table 1), single gene and partitioned analyses were congruent, especially concerning the position of *P. chromatophora* (clades nos. 6 and 7 Table 1). Support for tree topologies produced by LSU rDNA alone were comparable to those of the concatenated analysis (Fig. 1a), whereas in SSU rDNA analyses distinctly lower support values were obtained for most clades (Table 1).

To exclude the possibility that a cyanobacterial contaminant had been sequenced (the culture of *P. chromatophora* is not axenic), we (1) carefully studied chlorophyll autofluorescence in the culture by confocal laser scanning microscopy, and (2) sequenced rDNA from isolated and washed single photosynthetic organelles. The first approach

**Figure 2.** Evidence for the monophyly of *Synechococcus* and *Prochlorococcus* (clade no. 7 in Fig. 1a) to the exclusion of the photosynthetic organelle of *Paulinella chromatophora*: free-living taxa share compensatory base changes (synapomorphies) in successive positions in the LSU rRNA (Helix 837), whereas the *P. chromatophora* organelle retained the conserved (plesiomorphic) character states. One of the two synapomorphies shown is unique (pair 868—909; numbers indicate homologous positions in *Escherichia coli*). Secondary structure drawing based on *Synechococcus* WH 8102 sequence; interactions (base pairing) in gray color according to Cannone et al. (2002) but absent from models in De Rijk et al. (2000).
revealed no contaminants, and the second yielded rDNA sequences that were identical to those obtained before (results not shown). It had been previously shown by DAPI staining that the photosynthetic organelles of *P. chromatophora* contained nucleoids that resembled most closely those of a unicellular cyanobacterium (Lukavský and Cepák 1992). Using SSU rDNA alone and a larger taxon sampling (including freshwater strains of *Synechococcus*), we still recovered the photosynthetic organelle of *P. chromatophora* as a separate branch (results not shown). A close cyanobacterial relative of the photosynthetic organelle of *P. chromatophora* has thus not yet been identified (attempts to grow the photosynthetic organelle outside its host have consistently failed, and cell and organelle division are precisely coordinated; Hoogenraad 1927, and own observations). The acquisition of the photosynthetic organelle of *P. chromatophora* was likely a more recent event than the origin of plastids, since the *Paulinella* rDNA sequences are significantly more closely related to those of their cyanobacterial sister group than is the amount of sequence divergence even within plastids. To what extent the photosynthetic organelle of *P. chromatophora* is genetically reduced (its genome size is not yet known) will be of considerable interest and awaits further analysis.

**Methods**

**Culture origins and molecular methods:** DNA from *Paulinella chromatophora* Lauterborn (strain M 0880; the strain is available from the authors upon request), two glaucoplasts (*Glaucocystis nostochinearum* and *Cyanoptyche gloeocystis*), and two rhodoplasts (*Porphyridium aerugineum* and *Compsopogon coeruleus*) was extracted by using a CTAB protocol, or the DNeasy Plant Mini Kit (QIAGEN). The source of *P. chromatophora* has been described earlier (Bhattacharya et al. 1995); for origin of other cultures, see Figure 1. Culture collections: ATCC (http://www.atcc.org), CCMP (http://ccmp.bigelow.org/), M (Culture Collection Melkonian, University of Cologne, Germany), MIT (Massachusetts Institute of Technology), NCTC (http://www.ukncc.co.uk), NIES (http://www.nies.go.jp/biology/mcc/home.htm), PCC (http://www.pasteur.fr/recherche/banques/PCC/), SAG (http://www.gwdg.de/~epsag/phykologia/epsag.html), WH (Woods Hole Culture Collection), UTEX (http://www.bio.utexas.edu/research/utex/).

The plastid/organelle-encoded ribosomal DNA operon was amplified by polymerase chain reaction (PCR), using an initial step of 3 min at 95 °C, followed by 30 cycles of 95 °C (1 min), 55 or 60 °C (2 min) and 72 °C (3 min), and a final step of 5 min at 72 °C. To avoid unintended amplification of a bacterial rDNA operon (all cultures contained bacteria), the 23S rRNA gene was screened for molecular synapomorphies that uniquely characterized the Cyanobacteria lineage (including plastids) to the exclusion of the domain Bacteria (for search strategy, see below). Two of the synapomorphies found (one located in the spacer between domains C and D, the other in Helix B19; nomenclature after De Rijk et al. 2000) have been used to design cyanobacterial/plastid-specific primers ptLSU C-D-rev (5′-GCGGCC-TCATTTCTCAAC-3′) and ptLSU B19-fowr (5′-CACGTGRAATYCCGTGTAATCWGC-3′) with specific

### Table 1. Confidence measures for seven clades (encircled numbers in Fig. 1 a), inferred by single-gene analyses of SSU rDNA and LSU rDNA (support values as in Fig. 1 a), and posterior probabilities resulting from a Bayesian analysis of the complete alignment with three data partitions corresponding to SSU rDNA (1438 positions), both tRNA genes (148 positions), and LSU rDNA (2518 positions), using a mixed model (GTR+I+Γ) with independent parameter estimates for each partition.

<table>
<thead>
<tr>
<th>Clade</th>
<th>SSU rDNA (1438 positions)</th>
<th>LSU rDNA (2518 positions)</th>
<th>rDNA-operon (4104 positions); partitioned Bayesian analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71/78/1.00</td>
<td>99/98/1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>68/68/0.99</td>
<td>94/77/1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>76/58/0.97</td>
<td>96/55/0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>96/87/1.00</td>
<td>100/100/1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>69/60/0.9</td>
<td>90/93/1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>100/100/0.98</td>
<td>100/100/1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>81/79/0.92</td>
<td>100/98/1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*These analyses favored chloroplasts and rhodoplasts as sister clades with low significance (61/53/0.92).
positions at the 3’ ends. For PCR reactions, these specific primers were combined with universal SSU rRNA (SSU-4-forw, 5’-GATGCTCGATGCTAGGCTGGC-GATKAAGCCTGGC-3’) and LSU rDNA primers (LSU G20-rev, CACCGGATGGACCRAACTGTC; and LSU H1-4-rev, 5’-ACTYATCTTRGGTRGCGTCC-3’) to obtain overlapping PCR products that together covered the almost complete rDNA operon (length: 4290—4934 nt). Purified PCR products were sequenced directly (see Marin et al. 1998) with 15 sequencing primers (not shown) using a bidirectional LI-COR sequencer (LONG READ IR 4200). The five new rDNA sequences are available under accession numbers AM084273—AM084277.

As a control, partial rDNA sequences from single photosynthetic organelles of Paulinella chromatophora were obtained as follows: several single photosynthetic organelles of Paulinella chromatophora were isolated from a sterile culture medium, and transferred into a sterile culture medium, and transferred into a PCR-tube with distilled water, immediately followed by a freezing step (liquid nitrogen) and a short incubation at 95° C (3 min). At 4°C, PCR-components and primers SSU-4-forw and ptLSU C-D-rev were added for a primary PCR with 60°C as annealing temperature. Since no visible PCR products were obtained, 0.5—2 μl of the primary PCR were added to a secondary PCR with SSU-4-forw and the reverse primer SG2 (5’-CACGGATCCAGGAGGATCCANGCCNCACC-3’). Secondary PCR-products were used to determine partial SSU rDNA sequences of single photosynthetic organelles.

**Alignments and phylogenetic analyses:** The new rDNA sequences were combined with database sequences to construct an alignment that contained the genes encoding SSU rRNA, tRNA-Ile, tRNA-Ala, and LSU rRNA. Most sequences came from completed or ongoing (‘shotgun’) genome projects, whereas in some cases, separate entries for SSU and LSU rRNA genes from the same strain were combined (for EMBL/GENBANK accession numbers, see Fig. 1 a). The conserved rRNA and tRNA secondary structure (available at http://www.rna.icmb.utexas.edu/ and http://www.psbrug.ac.be/RNA/index.html) determined the architecture of the alignment.

A total of 4104 unambiguously aligned positions were used to infer molecular phylogenetic trees with maximum likelihood, distance (neighbor joining), maximum parsimony, and Bayesian methods (PAUP 4.0b10; MrBayes 3.1.1; Ronquist and Huelsenbeck 2003; Swofford 2002). Except for maximum parsimony, analyses were performed under a GTR+I+Γ model (model and model parameters selected by AIC in MODELTEST; Posada and Crandall 1998). The maximum likelihood tree was calculated by two heuristic searches, with starting trees obtained by either neighbor joining or by stepwise taxon addition (both searches resulted in the same topology). The robustness of branches was tested by bootstrapping (distance and maximum parsimony), using 1000 replications (5 sequence addition replicates per bootstrap replicate in maximum parsimony). For the Bayesian analysis, two separate MCMC chains were performed with 300 000 generations; trees were sampled every 100 generations. Already after 28 000 generations, the standard deviation between the two MCMC chains was below 0.10, indicating convergence. Trees from generations 1 to 28 000 were discarded as ‘burnin’, whereas the remaining trees (plotting log-likelihoods against generations using the command ‘sump’ confirmed that the analysis reached a stationary phase) were used for calculating two 90% majority rule consensus trees (trees of the two MCMC chains were identical) to obtain posterior probabilities of branches.

To compare the phylogenetic signal of single genes, (1) separate analyses (distance and maximum parsimony bootstrap; Bayesian posterior probabilities) of the SSU rRNA- and LSU rRNA genes were performed as described before, and (2) the Bayesian analysis of the complete alignment was repeated after defining three data partitions corresponding to SSU rDNA (positions 1—1438), both tRNA genes (positions 1439—1586), and LSU rDNA (positions 1587—4104), using a mixed GTR+I+Γ model that allowed independent model parameter estimations for each partition.

**Search for unique synapomorphies:** The general procedure to find uniquely derived positions for a clade (branch) of interest has been briefly described earlier (Marin et al. 2003). First, sequence data and the treefile reflecting the maximum likelihood topology were imported into PAUP. After defining the outgroup (Bacteria) and selecting maximum parsimony as optimality criterion ( Parsimony settings; character state optimization: DELTRAN), the logfile option was activated (File -> ‘Log Output to Disk’). Sequence data were then used to obtain a labeled tree reconstruction and a complete list of synapomorphies (Trees -> ‘Describe Trees’ with
‘phylogram’, ‘labeled internal nodes’, and ‘list of synapomorphies’). The resulting logfile revealed all synapomorphies of the data set, especially those of two nested clades (nos. 6, 7 in Fig. 1 a): the *Synechococcus*/Prochlorococcus clade including, as well as excluding *Paulinella*. Over 99% of these changes were homoplasious (usually indicated by a low consistency index <0.3), i.e. showed parallel (convergent) changes. Synapomorphies with a higher consistency index (0.3—1.00) were mapped on the tree topology to visualize all character changes via PAUP (Trees -> ‘Show Reconstructions’). Only five synapomorphies had a consistency index of 1.00, and proved to be unique within the taxa analyzed (i.e. they were non-homoplasious synapomorphies [NHS] sensu Marin et al. 2003; Figs 1b, 2). For these synapomorphies, the position in the conserved rRNA secondary structure was identified (via alignment): one was located at the end of the SSU rRNA molecule (U == > A), whereas the remaining positions formed base pairs in the LSU rRNA (Helix 837, Helix 1082; Cannone et al. 2002), i.e. these synapomorphies represented unique compensatory base changes (see Figs 1b, 2).

Finally, BLAST searches were performed to test the uniqueness of synapomorphies among homologous organellar/prokaryotic sequences in the databases (http://www.ncbi.nlm.nih.gov/BLAST/). BLAST searches using the sequences GGATG-GATCACCTCCTC, TAGAAGCAGCCATCCTCA, TAGAACGAGCCATCCTCAAAG, CATTTCGGTGCGGGCTGCGA, and TGAACCTCN-GAACTCCTGT (for various synapomorphies, see Figs 1 b, 2) revealed mismatches in the synapomorphic positions for all database sequences except for rRNA genes of *Synechococcus* and *Prochlorococcus*.

References


Kies L (1974) Elektronenmikroskopische Untersuchungen an *Paulinella chromatophora* Lauterborn,
einer Thekamöbe mit blaugrünen Endosymbionten (Cyanellen). Protoplasma 80: 69—89


Vørs N (1993) Marine heterotrophic amoebae, flagellates and Heliozoa from Belize (Central America) and Tenerife (Canary Islands), with descriptions of new species, Luffisphaera bulbochaeta n. sp., L. longihastis n. sp., L. turiformis n. sp. and Paulinella intermedia n. sp. J Eukaryot Microbiol 40: 272—287


Available online at www.sciencedirect.com