Research Paper:

**Differential gene retention in plastids of common recent origin**

Adrian Reyes-Prieto\(^a\)\(^,\)\(^e\)*, Hwan Su Yoon\(^b\)\(^,\)*, Ahmed Moustafa\(^a\), Eun Chan Yang\(^b\), Robert A. Andersen\(^b\), Sung Min Boo\(^c\), Takuro Nakayama\(^d\), Ken-ichiro Ishida\(^d\), and Debashish Bhattacharya\(^a\)\(^,\)\(^l\)

\(^a\) **Department of Ecology, Evolution and Natural Resources and Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey, USA**

\(^b\) **Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine, USA**

\(^c\) **Department of Biology, Chungnam National University, Daejeon, Korea**

\(^d\) **Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan**

\(^e\) **Present address. Department of Biology, University of New Brunswick, Fredericton, New Brunswick, Canada**

*These authors contributed equally to this work

To whom correspondence should be addressed. E-mail: bhattacharya@aesop.rutgers.edu

Key words: endosymbiosis, primary plastids, photosynthesis, *Paulinella*

Running head: *Paulinella* plastid evolution
Abstract

The cyanobacterium-derived plastids of algae and plants have supported the diversification of much of extant eukaryotic life. Inferences about early events in plastid evolution must rely on reconstructing events that occurred over a billion years ago. In contrast, the photosynthetic amoeba *Paulinella chromatophora* provides an exceptional model to study organelle evolution in a prokaryote-eukaryote (primary) endosymbiosis that occurred ca. 60 million years ago. Here we sequenced the plastid genome (0.977 Mb) from the recently described *Paulinella* FK01 and compared the sequence to the existing data from the sister taxon *Paulinella* M0880/a. Alignment of the two plastid genomes shows significant conservation of gene order and only a handful of minor gene rearrangements. Analysis of gene content reveals 66 differential gene losses that appear to be outright gene deletions rather than endosymbiotic gene transfers (EGTs) to the host nuclear genome. Phylogenomic analysis validates the plastid ancestor as a member of the *Synechococcus-Prochlorococcus* group and the cyanobacterial provenance of all plastid genes suggests these organelles were not targets of interphylum gene transfers after endosymbiosis. Inspection of 681 DNA alignments of protein-encoding genes shows that the vast majority have dN/dS ratios <<1, providing evidence for purifying selection. Our study demonstrates that plastid genomes in sister taxa are strongly constrained by selection but follow distinct trajectories during the earlier phases of organelle evolution.
Introduction

The ancient origins of mitochondria and plastids explains the fundamentally chimeric nature of eukaryotes (Sagan 1967). Photosynthesis entered the eukaryotic domain via primary endosymbiosis, whereby a cyanobacterium was captured by a heterotrophic protist and converted into a photosynthetic organelle. This pivotal event occurred more than a billion years ago and laid the foundation for many food webs on our planet (Falkowski et al. 2004; Reyes-Prieto, Weber, and Bhattacharya 2007). The host lineage for the endosymbiosis is the putative ancestor of the Plantae that subsequently split into the glaucophyte, red, and green algae (including land plants, Cavalier-Smith 1992; Bhattacharya and Medlin 1995; Palmer 2003). The canonical Plantae plastid spread via secondary and tertiary endosymbiosis to other lineages such as chromalveolates and euglenids (Palmer 2003; Bhattacharya, Yoon, and Hackett 2004; Reyes-Prieto, Weber, and Bhattacharya 2007). All extant plastids, whether of primary, secondary, or tertiary origin are specialized organelles with highly reduced genomes (100-200 Kb) leaving us to speculate about the pattern and process of gene loss early in their evolution. Given this situation, there is much interest in identifying a more recent case of organelle establishment via primary endosymbiosis. This need appears to have been recently fulfilled with molecular studies of the thecate amoeba *Paulinella chromatophora* M0880/a (Marin, Nowack, and Melkonian 2005; Yoon et al. 2006; Marin et al. 2007; Nowack, Melkonian, and Glockner 2008). *Paulinella* is a member of the supergroup Rhizaria, yet it contains two blue-green “chromatophores” (Figs. 1A, 1B) that resulted from a novel plastid acquisition (Marin, Nowack, and Melkonian 2005; Yoon et al. 2006; Nowack, Melkonian, and Glockner 2008) about 60 million years (My) ago (Nowack, Melkonian, and Glockner 2008). Several lines of evidence support the hypothesis that *Paulinella* contains bona fide photosynthetic organelles. These include a constant plastid number
(i.e., two per cell) following each round of coordinated cell division (Kies 1974), a plastid
genome that is 1/3 the size of chromosomal DNA in putative free-living cyanobacterial donors
(Nowack, Melkonian, and Glockner 2008), and evidence for gene transfer to the amoeba nuclear
genome via EGT (i.e., the cyanobacterium-derived psaE gene; Nakayama and Ishida 2009).

Here we generated the plastid genome sequence from a sister taxon of Paulinella M0880/a
that was recently isolated in Japan (Yoon et al. 2009). This second isolate, Paulinella FK01,
provides an ideal tool to understand plastid genome evolution using a homologous organelle of
recent origin. A strategy utilizing fluorescence-activated cell sorting (FACS) to isolate organelles
was followed by single-cell genomics and ‘454’ pyrosequencing to generate a draft genome of
FK01 that was closed using targeted PCR.

Materials and Methods

Cell isolates

Paulinella FK01 was established from a single cell collected at Daigo-machi, Ibaraki prefecture,
Japan (Yoon et al. 2009). Paulinella chromatophora M0880/a was kindly provided by Michael
Melkonian (University of Cologne, Germany). Both isolates are maintained at the Bigelow
Laboratory for Ocean Sciences using DY-V medium at 20°C with a 14/10 hr light/dark cycle.

Plastid isolation and whole genome amplification

After cell disruption with glass-bead beating, single plastids were isolated using FACS (see Fig.
S1, Supplementary Material online). DNA derived from 50 isolated plastids was used for
genome amplification using the Repli-G mini kit (Qiagen), which applies multiple displacement
amplification (MDA) methods (Stepanauskas and Sieracki 2007). This resulted in ca. 10 µg of
DNA per reaction with an A260/280 ratio of 1.85. After the de-branching step with S1 nuclease to reduce chimeric sequences during MDA, a PCR survey was done using several gene markers for nuclear and plastid (pt)DNA (16S rDNA, 18S rDNA, rbcL, ftsZ, beta-tubulin) to validate the source of the nucleic acids; i.e., the nuclear gene amplifications acted as negative controls for this procedure. The PCR products were sent to the DNA Facility at the University of Iowa for Sanger sequencing, whereas the amplified plastid DNA was sent to Macrogen (Seoul, Korea) and to the University of Iowa for 454 (Roche) pyrosequencing.

**Genome sequencing, assembly, and annotation**

Combination of two separate runs of a ¼ plate each from Macrogen and the University of Iowa using the GS-FLX standard chemistry generated a total of ~2.4 million sequences with an average length of 230 bases, providing >55X theoretical coverage of the FK01 plastid genome. The filtered reads were assembled using the GS De Novo Assembler software (Roche Diagnostics Corporation), resulting in 11 contigs. Using as reference both the M0880/a plastid genome and the termini of the FK01 contigs, we designed specific PCR primers to complete the sequence of the missing regions in FK01. The final assembly was done using Sequencher. A total of 888 protein-coding genes were predicted in FK01 ptDNA using GeneMarkS (Besemer, Lomsadze, and Borodovsky 2001) and 912 proteins, using RAST (Aziz et al. 2008). Visual inspection and manual refinement substantiated a total of 841 likely protein-encoding gene models. After annotation, alignment of the *Paulinella* M0880/a and FK01 ptDNA was done under Mauve (Darling et al. 2004) using the progressive algorithm with default parameters.

**Phylogenomic analysis**
We prepared a local database that included more than 500 genome sequences and EST libraries, with more than 6,000,000 protein sequences (e.g., Moustafa et al. 2009). This database comprised all completely sequenced bacterial and cyanobacterial genomes, representatives of the six major eukaryotic supergroups, and organelle-encoded proteins including the M0880/a plastid data. The 841 predicted proteins from the FK01 plastid genome were used as BLAST queries against the genome database with e-value thresholds of 1E-5 and 1E-10 to address non-conserved and conserved queries, respectively (see Moustafa et al. 2009). Each query sequence and its significant homologs were aligned using MAFFT (Katoh et al. 2002). Sequence alignments were used to infer phylogenetic trees using PhyML (Guindon and Gascuel 2003) with the approximate likelihood-ratio test (aLRT; Anisimova and Gascuel 2006). The phylogenetic trees were grouped based on their topological patterns using PhyloSort (Moustafa and Bhattacharya 2008) with a minimum threshold set at 75% aLRT support values for groups of interest.

**Nucleotide substitution rate estimation**

We used the 841 predicted nucleotide-coding regions from FK01 as blastn (e-value threshold 1E-5) queries to identify the corresponding orthologous DNA regions from M0880/a ptDNA. Using this threshold we identified 681 orthologous pairs (75% of the total predicted protein coding genes in FK01) that could be reliably aligned using DNA data. Each ortholog pair was aligned with MAFFT (Katoh et al. 2002) using a maximum of 1000 iterations for alignment refinement. Individual alignments were visually inspected and edited to conserve the codon structure encoding the corresponding predicted protein. We estimated the ratio of non-synonymous (dN) to synonymous (dS) substitutions for all 681 codon-based DNA alignments using CONSEL, that
is included in the PAML suite (Yang 2007).

**PCR reactions**

Total genome DNA from both *Paulinella* species was extracted using the DNeasy Plant Mini Kit (Qiagen). PCR reactions to determine possible cases of EGT were done using specific primer pairs (see Table S1) for 27 genes encoded only in the FK01 plastid genome and for 19 genes exclusively present in M0880/a plastid DNA. In all 46 cases, total DNA from each species was used as the template in independent PCR reactions. All reactions were done with an initial denaturation step at 94°C for 10 min, followed by 35 cycles of 94°C for 1 min, 53–57°C for 1 min, and 72°C for 2 min, concluding with a 10 min extension at 72°C. PCR products were purified and either directly sequenced or cloned prior to sequencing.

**Results and Discussion**

**Genome sequencing**

We used flow cytometric single-cell sorting followed by whole genome amplification and pyrosequencing (Yoon et al. 2009) to determine the plastid genome sequence from the recently described taxon *Paulinella* FK01 (Yoon et al. 2009; Fig. 1B). We then compared its genome structure and gene composition to the previously reported ptDNA from *Paulinella* M0880/a (Nowack, Melkonian, and Glockner 2008). The ptDNA in FK01 is a circular molecule of 977,329 bases that encodes 841 predicted proteins (Table S2) and 48 structural RNAs.

Alignment of FK01 and M0880/a ptDNA (Fig. 2) reveals overall conservation of gene order, but there are five genome inversions involving fragments of sizes 110.3 Kb, 24.9 Kb, 3.9 Kb, 2.1 Kb, and 569 bp and a single 9.2 Kb translocation. Phylogenomic analysis of the predicted plastid
proteins indicates they are all derived from cyanobacteria and in most trees (820/841, 97.5%) *Paulinella* genes branch as sister to the cyanobacterial clade containing *Synechococcus-* *Prochlorococcus* species (Badger, Hanson, and Price 2002; Marin, Nowack, and Melkonian 2005; Yoon et al. 2006; Marin et al. 2007; Yoon et al. 2009). The majority (814/841, 96.7%) of FK01 protein-encoding genes have orthologs in M0880/a, and in most cases (>98%), these sequences branch as sisters in the phylogenetic trees (results not shown [all trees and alignments are available from http://dblab.rutgers.edu/paulinella]).

A total of 33 genes were apparently acquired by the cyanobacterial donor of the *Paulinella* plastid *via* horizontal gene transfer (HGT) from other bacterial sources, prior to endosymbiosis (Table S3). This group comprises 26 genes (including eight that were previously reported; Marin et al. 2007) of proteobacterial origin and seven of unresolved affiliation. Therefore, as observed in Plantae plastids (e.g., Rice and Palmer 2006), the *Paulinella* photosynthetic organelle is protected from HGT; i.e., gene loss, not gain characterizes this earlier phase in organelle evolution. Estimations of non-synonymous (dN) and synonymous substitution (dS) rates for 681 ortholog pairs from both *Paulinella* isolates reveal that all dN/dS ratios are < 1 with the vast majority (629) having a dN/dS < 0.1 Table S4). Reduced dN/dS values reflect an elevation in the dS value (i.e., in 97% of gene pairs, dS >1.1; Table S4), which is generally found for endosymbiont-encoded genes (Moran 2002; Moran, McLaughlin, and Sorek 2009). These results indicate that even under elevated rates of synonymous substitutions in *Paulinella* plastid genes (Nowack, Melkonian, and Glockner 2008; Yoon et al. 2009) and the tendency for endosymbiont genomes to accumulate deleterious mutations, purifying selection is strongly constraining amino acid substitutions in most of the proteins encoded by M0880/a and FK01 ptDNA. A similar complex pattern of genome evolution is also found in Plantae plastids and differs fundamentally
from insect endosymbionts that may be highly streamlined (e.g., *Carsonella ruddii*; Nakabachi et al. 2006) to only a handful of essential functions (e.g., amino acid synthesis).

*Differential gene loss*

Gene-by-gene comparison of FK01 and M0880/a ptDNA reveals 27 genes encoded in FK01 that are absent from M0880/a, whereas 39 genes in M0880/a are absent from FK01 (Table S5). These 66 genes trace their origin to the *Synechococcus-Prochlorococcus* group of cyanobacteria suggesting they were present in the ancestor of these photosynthetic *Paulinella* and are examples of lineage-specific losses. We investigated whether these plastid gene losses may be explained by EGT to the host nucleus (e.g., as for *psaE*) or by outright loss. A total of 46 lineage-specific genes (27 specific to FK01, and 19 specific to M0880/a) were targeted by PCR with gene-specific primers and total amoeba DNA. Here if the gene was absent in ptDNA but provided a product when analyzing total DNA, we interpreted this as a putative case of EGT. Positive controls were provided by total DNA from the lineage that we knew contained a copy of the gene in the plastid (although additional nuclear copies could be present in these cases). This approach provided unambiguous results demonstrating that none of the 46 lineage-specific plastid genes are present in total DNA when absent from the plastid, but result in PCR fragments from positive control DNA (results not shown); i.e., they are not candidates for EGT and likely constitute complete losses. It is of course possible that high sequence divergence or the insertion of spliceosomal introns that interrupt the PCR primer sites explain some of these results. However given the overall trend, we postulate that gene loss, not EGT, explains much of the differential plastid gene loss in these *Paulinella* isolates. We expect however that many endosymbiont genes that encode critical plastid functions have been transferred to the nucleus of
these photosynthetic amoebae and would only be uncovered by analysis of the nuclear genome sequence.

Furthermore, when we consider genome data from many endosymbiotic bacteria (Moran 1996; Perez-Brocal et al. 2006; Hosokawa et al. 2006; Moran, McLaughlin, and Sorek 2009) we can assume that during organellogenesis the effective population size of the endosymbiont is small (e.g., two plastids per amoeba), entailing restricted (or absence of) DNA recombination. This would lead to the accumulation of deleterious mutations and shifts in base composition (Moran 1996). As a consequence, it is plausible that genetic drift (Moran 2002; Marais 2008) has been driving reduction of the *Paulinella* plastid genome. Consistent with this model (Moran 2002) both the AT content (>60%) and the nucleotide substitution rate of the *Paulinella* plastid genomes are relatively elevated (Nowack, Melkonian, and Glockner 2008; Yoon et al. 2009). The assumption that a subset of endosymbiont genes have little or no consequences for host fitness would explain why the size of endosymbiont genomes are significantly reduced over short evolutionary periods (Marais 2008; Moran, McLaughlin, and Sorek 2009). In *Paulinella*, we observe that the relative distribution of plastid gene losses in FK01 and M0880/a (i.e., 39/27, respectively, versus the null expectation of 33/33) is consistent with genetic drift as the underlying explanation (Fisher’s exact test, *P* = 0.382). The independent trajectory of gene retention (versus loss) apparent in FK01 and M0880/a is also likely to be explained by selective forces operating under local ecological conditions although it should be noted that taxa closely related to both *Paulinella* isolates analyzed here have been collected from the same pond (unpublished results).

*Coordinated plastid gene loss in Paulinella*
A striking example of plastid gene loss in *Paulinella* is ferredoxin-dependent glutamate synthase (Fd-GOGAT; *glsF/gltS*, 4.6 Kb) that is present in FK01 but absent in M0880/a. Fd-GOGAT is involved in ammonium assimilation (Kameya et al. 2007) and its absence results in nitrogen-deficiency phenotypes in some cyanobacteria (see Okuhara et al. 1999). The genome alignment indicates that the two loci (lioyl synthase, *lipA*, and hypothetical protein PCC_0558) flanking the *glsF/gltS* gene in FK01 ptDNA are contiguous in M0880/a, separated by 980 bp (Fig. 3A). Fd-GOGAT and glutamine synthetase type I (GSI, plastid-encoded in both *Paulinella* species) are essential for ammonium assimilation in cyanobacteria and plastids (i.e., the GS-GOGAT cycle; Marques, Florencio, and Candau 1992; Muro-Pastor, Reyes and Florencio 2005; Kameya et al. 2007). The signaling molecule regulating this pathway, PII, is present in both plastid genomes suggesting strongly that the Fd-GOGAT locus has been transferred to the nucleus in M0880/a. Apart from the absence of this key sequence, genes encoding ammonium transporters (*amt*), isocitrate dehydrogenases (*icd*), transcriptional regulators of the Crp/Fnr family (*ntcA*), and glutamate dehydrogenases (*gdhA*) are absent from M0880/a ptDNA. These latter genes that are encoded in the genome of *Synechococcus* sp. WH 5701 (Marin et al. 2005; Yoon et al. 2006) have surprisingly also been lost from FK01 ptDNA. Given the importance of the ammonium assimilation pathway, our working hypothesis is that some or all of the ‘missing’ genes in FK01 and M0880/a ptDNA have been differentially transferred to the *Paulinella* nucleus, although it is possible that some of these functions may have been substituted by existing host proteins that are now plastid targeted.

Another example of gene loss is the four (*pstS* and a 3-gene cluster comprised of *pstA*, *pstB*, and *pstC*) members encoding different subunits of the ABC phosphate transporter *pstSACB* (Raymond, Trautman, and Larkum 2001) that are present in M0880/a ptDNA but
coordinately lost in FK01. Conservation of the homologous flanking genes in FK01 demonstrates the precise nature of gene deletions for these regions (Fig. 3B). The question that remains is: how does the FK01 plastid internalize phosphate after loss of this high-affinity transporter? There is for example, no evidence of genes encoding putative low-affinity phosphate (e.g., pitA permeases) or phosphonate (phnCDE) ABC transporters in either plastid genome. Therefore, if all cyanobacterial pstSACB genes have been lost in FK01, as our total DNA PCR results suggest (although this clearly provisional), we hypothesize that phosphate (or phosphorous compounds) uptake by this organelle relies on transport mechanisms that evolved after the endosymbiosis; e.g., the co-option of host-derived transporters to support plastid metabolism.

Given the results described above, what evidence exists thus far for EGT in *Paulinella*? Analysis of the >3,000 expressed sequence tags (ESTs) available from *Paulinella* FK01 reveals that in addition to the previously described *psaE* gene (Nakayama and Ishida 2009), cyanobacterial *psaI*, that encodes subunit VIII of photosystem I (PSI-VIII), has been relocated to the amoeba nucleus. This gene is still identifiable in the FK01 plastid genome but has been silenced by two nonsense mutations. In contrast, plastid-encoded *psaI* in M0880/a is intact. Using FK01 total DNA as template we amplified a genome region that contains a *psaI* locus and is interrupted by a spliceosomal intron, demonstrating its nuclear derivation (Fig. 4). These results provide evidence that two photosystem I subunits are encoded in the nucleus of *Paulinella* and expressed as mRNA. Subunits produced by the genes *psaE* and *psaI* participate respectively as accessory protein (PSI-subunit IV or PsaE) modulating binding between the PSI and soluble ferredoxin during electron transport, and in the correct oligomerization (PSI-subunit VIII or PsaI) of photosystem I. Structural interactions of PsaI with the subunit PsaL (plastid
encoded in *Paulinella*) are critical to stabilize the trimeric state of PSI (Xu et al. 1995). Absence of PsaI has detrimental effects on the structural organization and activity of PSI in cyanobacteria (Xu et al. 1995). This limited evidence suggests that PsaI is produced in the FK01 cytoplasm and likely imported into the plastid to support standard PSI function. In contrast, ten other genes encoding PSI proteins (Ycf4, Ycf4, *psaC*, *psaA*, *psaB*, *psaL*, *psaD*, *psaI*, *psaF*, *psaK*) are still encoded in the *Paulinella* plastid genome. It remains to be addressed how *Paulinella* PsaE and PsaI, if in fact translated in the cytosol, enter the plastid because the ESTs do not appear to encode an organelle transit sequence (Nakayama and Ishida 2009).

Other examples of FK01 ‘exclusive’ genes are proteins (e.g., QueC, QueD, QueD) presumably involved in queuosine (a nucleoside) biosynthesis, two subunits of phosphoribosylformylglycinamidine synthase (*de novo* purine biosynthesis), the uroporphyrinogen-III synthase (tetrapyrrole biosynthesis), S-adenosylmethionine synthetase (cysteine and methionine metabolism), and a DNA helicase. In the case of the M0880/a ptDNA some examples of its set of exclusive proteins are those involved in cell (i.e., organelle) division and DNA replication, such as two ATP-dependent metalloproteases FtsH (encoded by *ftsH2* and *ftsH3*, with 56% of identity at the protein level), and chromosome duplication (DnaA ATPase) and segregation (SMC ATPase). Our results also demonstrate that genes essential for key enzymatic functions such as tetrapyrrole (e.g., chlorophyll) biosynthesis and DNA replication have been lost from ptDNA in both *Paulinella* species. The upcoming nuclear genome sequence from FK01 (HSY, RAA, and DB unpublished data) will provide definitive evidence for EGT *versus* outright loss as explanation for the missing genes in this taxon.
Conclusions

Our data provide important insights into early organelle evolution that until recently appeared intractable due to the ancient origins of the mitochondrion and the plastid. The relatively recently established *Paulinella* plastid appears to mimic some of the predicted ancestral features of the canonical Plantae plastid, including differential gene loss in descendant lineages (e.g., Martin et al. 1998), high conservation of photosynthetic and informational functions, protection from HGT (e.g., Rice and Palmer 2006), and EGT to the host nucleus (Martin and Herrmann 1998). The photosynthetic *Paulinella* lineage may therefore provide a “retelling of the tale” that took place when the Plantae ancestor, long ago, diversified into some of the most important photosynthetic eukaryotes on our planet.

Supplementary Material

Supplementary Figure S1 and supplementary Tables S1-S5 are available at MBE online.

Acknowledgements

This project was partially supported by grants from the National Science Foundation Microbial Genome Sequencing program (EF 08-27023) that was awarded to HSY, RAA, DB, from the Assembling the Tree of Life program (EF 04-31117) that was awarded to DB, and grants from the Korean Science and Engineering Foundation (MOST; R01-2006-000-10207-0, KRF-2008-357-C00148) that were awarded to SMB and ECY. AM was supported by an Institutional NRSA (T32 GM98629) from the National Institutes of Health. TN was supported by a JSPS Research Fellowship for Young Scientists (DC1).
Literature cited


Raymond, J. R., D. A. Trautman, and A. W. D. Larkum. 2001. Phosphate limited cultures of the cyanobacterium *Synechococcus* are capable of very rapid, opportunistic uptake of phosphate. New Phytologist **152**:189-201.


Mutational analysis of photosystem I polypeptides in the cyanobacterium *Synechocystis*
sp. PCC 6803. Targeted inactivation of psaI reveals the function of psaI in the structural


Yoon, H. S., T. Nakayama, A. Reyes-Prieto, R. A. Andersen, S. M. Boo, K. I. Ishida, and D.
Bhattacharya. 2009. A single origin of the photosynthetic organelle in different

Yoon, H. S., A. Reyes-Prieto, M. Melkonian, and D. Bhattacharya. 2006. Minimal plastid
FIG. 1. (A) Maximum likelihood phylogenetic tree inferred using 16S rDNA sequences from plastids, cyanobacteria, and other bacteria. Numbers above the branches nodes denote support values (when ≥ 50%) from RAxML bootstrap replicates and thick branches indicate Bayesian posterior probabilities > 0.95. Branch lengths are proportional to the number of substitutions per site (see scale bar). (B) Light microscopy and SEM images from Paulinella FK01: A, oral scales; C, chromatophores (plastids); and S, scales.

FIG. 2. Alignment of the Paulinella M0880/a and FK01 plastid genomes (see Methods). The putative origin of replication of M0880/a ptDNA (Nowack, Melkonian, and Glockner 2008) was used to align the genomes. The diagonal lines indicate genome rearrangements.

FIG. 3. Plastid genome alignment illustrating the loss of genes encoding (A) Fd-GOGAT (glsF/gltS) in Paulinella M0880/a, and (B) the four subunits of the phosphate transporter in Paulinella FK01, pstSACB (pstA, pstB, pstC, pstS). The white blocks indicate the genes that were differentially lost in each genome. The black blocks and gray diagonal lines indicate gene order conservation of the flanking genome regions.

FIG. 4. (A) Plastid genome alignment showing the presence of the psaI (subunit VIII of photosystem I, PSI-VIII) gene (white block) in Paulinella M0880/a and its apparent loss in the homologous region from Paulinella FK01 ptDNA after automated annotation (see text and Methods). (B) Protein alignment of PSI-VIII sequences from Plantae plastids, cyanobacteria, M0880/a, and conceptual translation of both the pseudogenes present in the FK01 plastid genome (FK01 ptDNAtra) and the EST from FK01 (FK01 ESTtra). The nucleotide alignment of
the \textit{psaI} genome sequence (FK01 genomic) and the EST (FK01 EST) from FK01 indicates the presence of a 198 bp spliceosomal intron in this coding region.
Euglenids
Haptophytes + Dinoflagellates
Reds
Greens + Chlorarachniophytes
Glaucophytes
Cryptophytes + Dinoflagellates
Stramenopiles

99
Gloeobacter violaceus
Synechococcus JA 2 3Ba

98
50
Glaucophytes
Greens + Chlorarachniophytes

98
66
Reds
Haptophytes + Dinoflagellates

98
84
Euglenids

99
Stramenopiles

100
Cryptophytes + Dinoflagellates


5 μm

Cyanobacteria

Diverse Bacteria

Plantae and Plantae-derived Plastids

Cyanobacteria

Plantae and Plantae-derived Plastids

Rhizaria

0.1 substitutions /site