Current Biology *18*, 410–418, March 25, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2008.02.051

Article

Chromatophore Genome Sequence of *Paulinella* Sheds Light on Acquisition of Photosynthesis by Eukaryotes

Eva C.M. Nowack,[1] Michael Melkonian,[1,*] and Gernot Glöckner[2,*]
[1]Botanical Institute
University of Cologne
Gyrhofstr. 15
D-50931 Köln
Germany
[2]Leibniz Institute for Age Research–Fritz Lipmann Institute
Beutenbergstr. 11
D-07745 Jena
Germany

Summary

Background: It is commonly accepted that a single primary endosymbiosis gave rise to the photosynthetic organelles of plants, the plastids. Recently, we presented evidence that photosynthetic inclusions, termed "chromatophores," present in the filose thecamoeba *Paulinella chromatophora* originated from an independent, more recent primary endosymbiotic event. To clarify metabolic capabilities of the chromatophore and its state of integration into the host, we present here the complete genome sequence of the chromatophore.
Results: Our data reveal a fundamental reduction of the chromatophore genome. The single, circular chromosome of 1.02 Mb encodes 867 protein-coding genes and is, therewith, the smallest cyanobacterial genome reported to date. Compared to *Synechococcus* WH5701, a free-living relative of the chromatophore, only 26% of the genes were retained. Eleven putative pseudogenes were identified, indicating that reductive genome evolution is ongoing. Although the chromatophore genome contains a complete set of photosynthesis genes, it lacks not only genes thought to be dispensable for an intracellular lifestyle but also genes of essential pathways for amino acid and cofactor synthesis.
Conclusions: Our data characterize the chromatophore as a photosynthetic entity that is absolutely dependent on its host for growth and survival. Thus, the chromatophores of *P. chromatophora* are the only known cyanobacterial descendants besides plastids with a significantly reduced genome that confer photosynthesis to their eukaryotic host. Their comparison with plastids and bacterial endosymbionts of invertebrates sheds light on early steps of the integration of a photosynthetic prokaryote into a eukaryotic cell.

Introduction

The establishment of photosynthesis in eukaryotes undoubtedly has been one of the most important events in the evolution of life on this planet. It is commonly accepted that the photosynthetic organelles of plants, the plastids, were acquired by phagocytotic uptake of a cyanobacterium into a unicellular eukaryote through a process termed "primary endosymbiosis" [1]. Analysis of dozens of extant plastid genomes revealed that plastids had a monophyletic origin within cyanobacteria and most likely were acquired through a single primary endosymbiotic event [2–4]. During this process, the endosymbiont was transformed to a photosynthetic organelle accompanied by massive genome reduction, transfer of genes to the host, and reimport of host-translated proteins through a dedicated protein-import pathway [5]. Yet the reasons for the apparent singularity of plastid acquisition through endosymbiosis, which is thought to have taken place more than a billion years ago [6], have remained enigmatic. Recently, we presented evidence for the occurrence of a second, more recent primary endosymbiotic event leading to a photosynthetically active eukaryote [7]: the filose thecamoeba *Paulinella chromatophora.* The freshwater thecamoeba of cercozoan affiliation [8, 9] was first described by Robert Lauterborn in the 19th century [10]. Its most distinctive characteristic is the presence of one or two large, sausage-shaped, intracellular photosynthetic entities termed "chromatophores" by Lauterborn. In contrast to plastids that evolved from a β-cyanobacterial ancestor, the *Paulinella* chromatophore is monophyletic with α-cyanobacteria (i.e., the *Prochlorococcus/Synechococcus* clade of cyanobacteria) [11]. The chromatophores confer phototrophy to *P. chromatophora*, which has dispensed with phagotrophic nutrition [12], unlike their closest marine relatives that feed on cyanobacteria [13, 14]. Thus, providing energy in the form of reduced carbon compounds to the host by using its photosynthetic machinery seems to be the main function of the chromatophore [15]. Chromatophores released from their host cell retain their structural integrity due to the possession of a bacterial peptidoglycan cell wall but are not capable of further development [16]. In nature, *P. chromatophora* has never been observed without its chromatophores [10, 17], and in experimental culture the association between host and chromatophore has been stable for more than 15 years, suggesting that this relationship is mutual and obligatory. *P. chromatophora* is the only known case of a primary endosymbiosis leading to a phototrophic eukaryote independent from the Plantae. Its more recent emergence places *Paulinella* in a pivotal position for understanding early events in the evolution of a photoautotrophic eukaryote. To what extent the chromatophore is integrated into the host's metabolism (e.g., whether significant reduction of the chromatophore genome, gene transfer to the nucleus, and re-import of proteins into the chromatophore evolved in *P. chromatophora*) is unknown but nevertheless has recently been hotly debated, based on limited information [18–21]. To clarify the chromatophore's metabolic role and its state of integration into the host, we present here the complete genome sequence of the chromatophore.

Results and Discussion

Genome Sequence and Annotation

DNA from purified chromatophores was sequenced by a whole-genome, random-sequencing approach, resulting in a single chromosome of 1,021,616 bp with a low G + C content of 38% (Table 1). Thus, the chromatophore genome is more than one-third smaller than the smallest known genome of

*Correspondence: michael.melkonian@uni-koeln.de (M.M.), gernot@fli-leibniz.de (G.G.)

Table 1. Overview of Genome Properties of the *Paulinella* Chromatophore Compared to *Synechococcus* Strain WH5701 Draft Assembly

	Chromatophore	WH5701
Genome size (bp)	1,021,616	3,043,834
GC content (%)	38	65
tRNAs	42	55
rRNA cluster	2	3?
Protein coding genes	867	3346
Hypothetical proteins (% of total)	223 (25.7)	1553 (46.4)

a free-living oxygenic phototroph, *Prochlorococcus marinus* MED4 [22]. Reductive genome evolution in intracellular bacteria is usually accompanied by the rapid acquisition of a nucleotide bias, presumably due to loss of repair genes [23]. The chromatophore genome is no exception to this rule. The chromatophore genome encodes 867 protein-coding genes. In agreement with its reduced genome size, multicopy genes are rare. Only two protein-coding genes are present as identical duplicates or triplicates (three photosystem II D1 genes [*psbA*] and two photosystem II D2 genes [*psbD*]). Furthermore, we identified two rRNA gene clusters and 42 tRNA genes for a full set of amino acids. Interestingly, the two rRNA gene clusters are located nearly mirror symmetrical in equal distance to the putative origin of replication (Figure 1). This organization mimics partition of almost all plastid genomes into small and large single-copy regions but also is present in some cyanobacteria (e.g., [24]). No transposable or insertion elements have been found in the chromatophore genome.

Reduction of genome size is the main feature linked to an intracellular lifestyle in obligate endosymbiotic bacteria and reflects a lack of selection for gene retention rather than selection for a small genome size [25]. All endosymbionts retain syntenic regions with their free-living relatives to some extent.

Structural Genome Comparison with *Synechococcus* WH5701

Currently, 39 cyanobacterial genomes have been annotated or are available as draft assemblies (http://www.ncbi.nlm.nih.gov/ as of Dec. 27, 2007). BLAST searches using chromatophore-encoded proteins usually gave best matches to *Synechococcus* WH5701 (hereafter referred to as WH5701). WH5701 is a free-living member of the α-cyanobacteria and the only completely sequenced strain of the Cyanobium clade

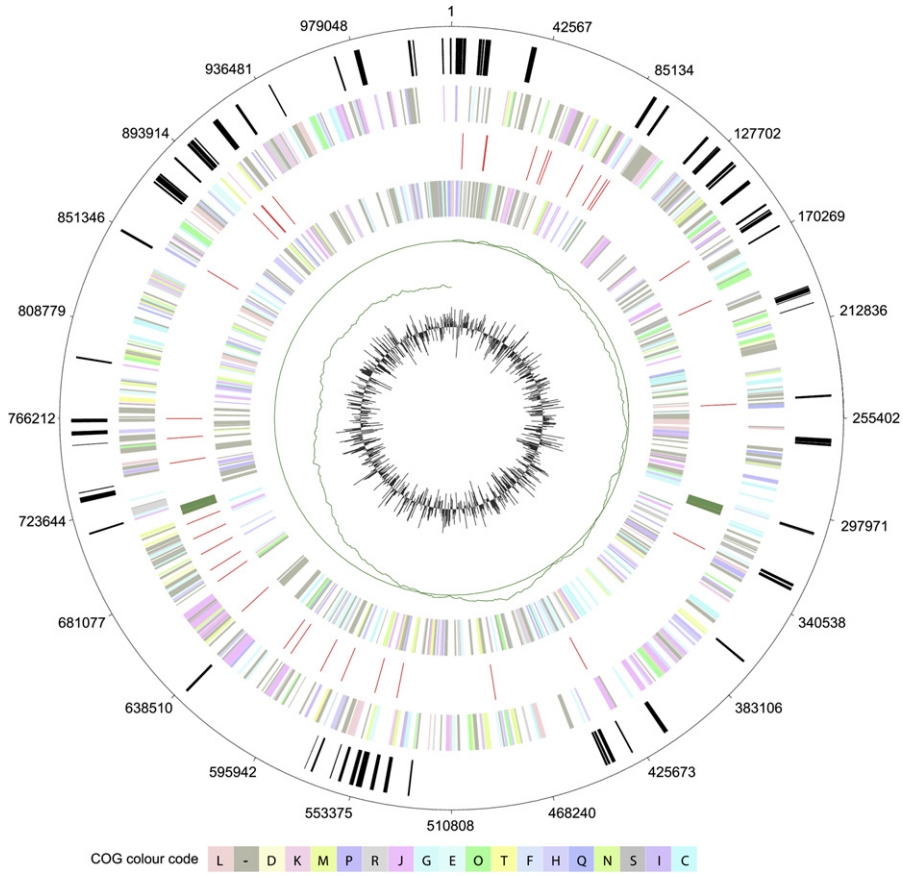

Figure 1. Graphical Representation of the *Paulinella* Chromatophore Genome

From inner to outer circle are depicted: G + C content in a window of 1000 bases, purine excess in a window of 1000 bases, minus-strand protein-coding sequences, tRNAs (red) and rRNA clusters (green), plus-strand coding sequences, and synteny breakpoints compared to *Synechococcus* WH5701. Color coding for the coding sequences is according to COG functional categories: replication, recombination, and repair (L); unclassified (−); cell cycle control, mitosis, and meiosis (D); transcription (K); cell wall/membrane biogenesis (M); inorganic ion transport and metabolism (P); general function prediction only (R); translation (J); carbohydrate transport and metabolism (G); amino acid transport and metabolism (E) posttranslational modification, protein turnover, and chaperones (O); signal transduction mechanisms (T); nucleotide transport and metabolism (F); coenzyme transport and metabolism (H); secondary metabolites biosynthesis, transport, and catabolism (Q); cell motility (N); function unknown (S); lipid transport and metabolism (I); and energy production and conversion (C).

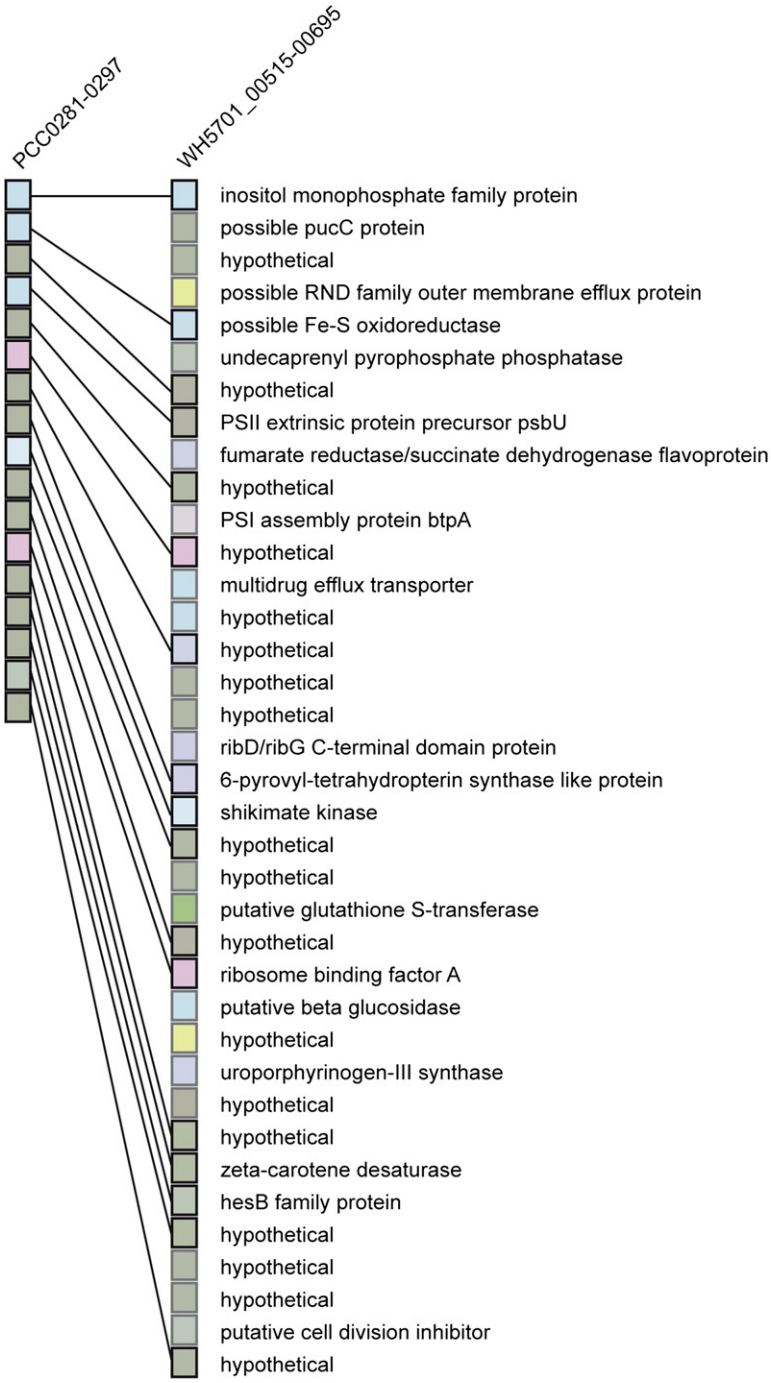

Figure 2. Example of a Synteny Group

Shown is a typical synteny group (group 531) between the *Paulinella* chromatophore genome and *Synechococcus* WH5701.

[11] that contains mostly freshwater organisms. Therefore, WH5701 was chosen for a detailed genome comparison. The genome of WH5701 is available only as a draft assembly with preliminary annotation. Despite the largely differing G + C content between the two genomes (Table 1), an orthologous counterpart for 855 of the chromatophore protein-coding genes was found in the WH5701 genome (Table S2 available online). The ortholog pairs differ with respect to their synonymous versus nonsynonymous substitutions, presumably reflecting differing constraints on the genes (Supplemental Results and Table S2). Among the remaining 12 genes with no obvious ortholog in WH5701 (Table S1) were identical copies of genes (duplicates or triplicates: PCC0033, PCC0547, and PCC0686), and a single gene split (PCC0136 and PCC0137 are derived from WH5701_14511). For an additional eight genes, putative orthologs could only be identified based on synteny analysis because their sequence identity was below the threshold used. The nearly one-to-one orthology indicated that neither horizontal gene transfer nor invention of new genes took place in the chromatophore genome after the two genomes diverged from their last common ancestor. Synteny analysis revealed that in 77 segments, gene order and orientation is conserved between the two genomes (Figure 1 and Table S3). In 23 of these segments, more than five consecutive genes were lost during evolution of the chromatophore. Within synteny regions no gene rearrangements occurred irrespective of the number of genes missing within segments (Figure 2).

Twenty-one genes on the chromatophore genome were found to have undergone severe changes during evolution compared to WH5701 (Table 2); 11 of those genes, which either formed no product due to several in-frame stop codons or a product of less than half the original length, were classified as potential pseudogenes. The other ten severely altered genes yield N- or C terminally truncated amino acid sequences upon translation. In all of these cases, no traces of the missing sequences could be detected up- or downstream of the respective genes, excluding the possibility that simple frame shifts or nonsense mutations caused the truncations. Some of these severely altered genes also could be pseudogenes.

Average intergenic distances in the chromatophore genome are considerably longer than in two free-living relatives analyzed (223 bp in the chromatophore versus 106 bp and 136 bp in *Synechococcus elongatus* PCC7942 and *Prochlorococcus marinus* NATL2A, respectively; comparison to WH5701 is not possible due to the status of its genome project as draft assembly), suggesting the presence of DNA sequences that lost function but are too derived to be recognized as pseudogenes. The comparison of the chromatophore genome with the genome of WH5701 reveals that evolution of the chromatophore took mainly place through gene loss. The presence of several putative pseudogenes indicates that genome reduction in the chromatophore is ongoing.

Loss of Genes

Obligatory endosymbionts tend to lack whole metabolic pathways essential for free-living organisms; plastids have additionally distributed genes essential for their function between host and organelle genomes. To address the question whether the significant reduction of the genome size of the chromatophore was accompanied by loss of essential genes and/or

Table 2. Genes Structurally Severely Altered and Genes Classified as Potential Pseudogenes on the Chromatophore Genome Compared to *Synechococcus* WH5701

Locus Tag	Gene Product	Characteristic
Putative Pseudogenes		
PCC0002	putative CaCA family, sodium/calcium exchanger	stop codons
PCC0079	similar to WH5701_07331	stop codons
PCC0493	rubredoxin	stop codons
PCC0503	tRNA pseudogene	defined as pseudogene by tRNAscan-SE
PCC0533	similar to WH5701_01280	C-terminal half truncation
PCC0538	aspartate carbamoyltransferase catalytic subunit	degraded
PCC0622	possible rare lipoprotein A	stop codons
PCC0676	chromosome segregation protein	no start, 600 aa missing
PCC0694	cell wall hydrolase/autolysin	N-terminal half missing
PCC0716	multidrug eflux transporter	stop codons
PCC0877	similar to WH5701_12643	stop codons
Severely Altered Genes		
PCC0572	tRNA pseudouridine synthase A	55 aa truncation C-terminal (80%)*
PCC0348	metal dependent phosphohydrolase	180 aa truncation C-terminal (74%)*
PCC0410	signal recognition particle-docking protein FtsY	several N-terminal deletions (82%)*
PCC0432	TPR repeat protein	36 aa truncation C-terminal (83%)*
PCC0486	DNA mismatch repair protein	100 aa N-terminal differing + 7 aa insertion (90%)*
PCC0888	transaldolase	67 aa truncation C-terminal (82%)*
PCC0040	probable ribonuclease II	193 aa truncation N-terminal (71%)*
PCC0070	putative penicillin binding protein	132 aa truncation C-terminal (82%)*
PCC0116	similar to WH5701_09570	161 aa truncation N-terminal (72%)*
PCC0126	cell division protein Ftn2	124 aa truncation N-terminal, fusion with ORF similar to WH5701_14471

Abbreviations: aa, amino acid; ORF, open reading frame.
* Percentage of length of the gene compared to the respective ortholog in WH5701.

pathways, the set of proteins encoded on the chromatophore genome was compared with the genomes of WH5701 and 13 other free-living cyanobacteria by BLAST searches and by pathway analyses using the KEGG database [26] (Table S4A).

Compared to WH5701, far more genes without assigned functions were lost than genes with known functions (Table 1), suggesting that the former constitute a gene reservoir for functional responses to environmental changes and, thus, may no longer be needed in the homogeneous intracellular environment. However, many genes with known and essential functions also were not detected on the chromatophore genome, characterizing the chromatophore as a photosynthetic entity that is absolutely dependent on its host for growth and survival (Figure 3). We suggest that lost essential functions need to be compensated for by the host.

Completely missing are pathways for the biosynthesis of several amino acids (Glu, Arg, His, Try, and Met) and cofactors (nicotinamide adenine dinucleotides, riboflavine, thiamine, biotin, cobalamine, pantothenate, and coenzyme A). Moreover, all genes coding for enzymes of the TCA cycle, which provides primary building blocks (oxaloacetate, α-ketoglutarate) for the synthesis of important cellular compounds, are absent from the chromatophore genome. For other amino acids such as Gln, Asp, Asn, Lys, Cys, Thr, Tyr, and Phe, precursors cannot be formed, although the biosynthetic pathways are apparently retained. In addition, the biosynthetic pathway leading to AICAR (5′-phospho-ribosyl-5-amino-4-imidazole carboxyamide), the precursor of purine nucleotides, is absent.

Interestingly, in some other biosynthetic pathways only single genes are missing. For example, *hemD*, which encodes the uroporphyrinogen III synthase of the chlorophyll-heme-phycocyanobilin pathway and is present in all free-living cyanobacteria analyzed in this study, is absent from the chromatophore genome. The lack of *hemD* leads to the spontaneous formation of uroporphyrinogen I [27] that cannot be further metabolized. If the function of such enzymes cannot be taken over by other, not-yet-characterized proteins, two possibilities remain: (1) either the enzyme required for the single enzymatic step is provided by the host, which requires a protein import system, or (2) a swift exchange of metabolites needs to be postulated with export of substrate from the chromatophore and re-import of product from the host. In both scenarios, the gene encoding the respective enzyme could have been either derived from the host (if an equivalent metabolic pathway exists) or acquired by intracellular gene transfer from the cyanobacterial endosymbiont.

Similar levels of gene losses in metabolic pathways that are essential for a free-living existence occur in bacterial endosymbionts of invertebrates (e.g., [28–31]). For *Wolbachia pipientis w*Mel, an obligate intracellular bacterium of *Drosophila melanogaster*, it was shown that reductive genome evolution was not accompanied by intracellular gene transfer, but missing metabolic functions seem to be compensated at the metabolite level [32]. Conversely, the genomes of four insect and four nematode species recently were found to contain genome fragments of their *Wolbachia* endosymbionts that have access to the host's germ line due to their presence in developing gametes [33]. However, the majority of horizontal gene transfers into eukaryotes have been described in phagotrophic, unicellular eukaryotes [34]. Because unicellular organisms have no sequestered germ line, integration of DNA from their food organisms appears to be rather frequent [35]. Preceding endosymbiosis, the unicellular thecamoeba presumably fed on cyanobacteria and other prokaryotes, as do other thecamoebae. Thus, genetic integration of a cyanobacterium as a chromatophore might have been accompanied by intracellular gene transfer.

Photosynthesis and Carbon Fixation

The main function of the chromatophore is undoubtedly photosynthesis [15]. In contrast to the situation in plastids, all genes

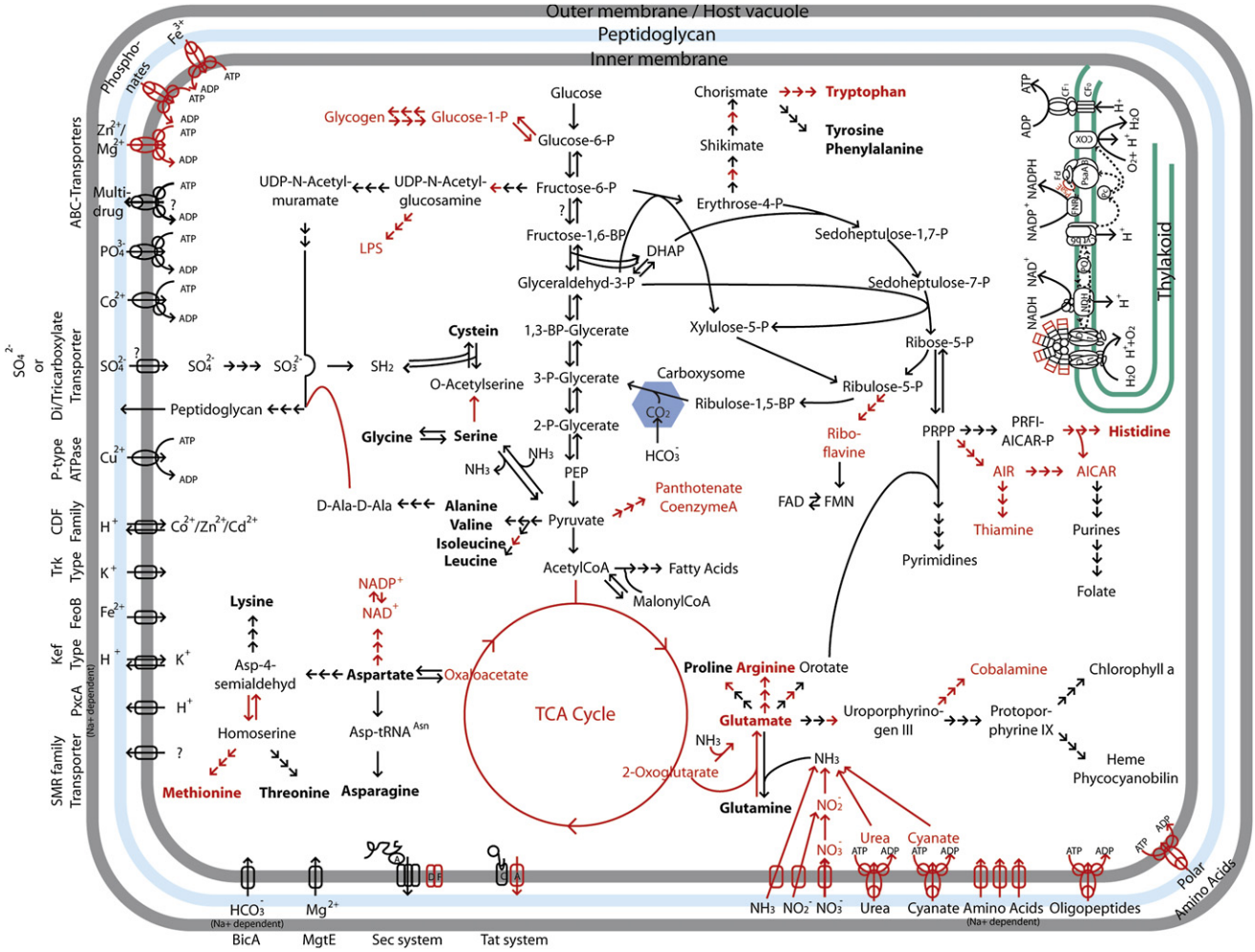

Figure 3. Metabolism of the *Paulinella* Chromatophore

Schematic representation of the main α-cyanobacterial metabolic pathways and transport systems present (black) or missing (red) in the *Paulinella* chromatophore as deduced from the genes identified in the chromatophore genome. Pathways that involve multiple reactions are represented by triple arrows. A single, red arrow in a triple arrow indicates the lack of a single enzyme in this metabolic pathway. A red triple arrow indicates that all or the majority of genes for the pathway are missing. Question marks denote uncertainties in function or identity of a protein. Amino acids are in bold. AICAR, 5′-phosphoribosyl-5-amino-4-imidazole carboxamide; AIR, 5′-phosphoribosyl-5-aminoimidazole; COX, cytochrome-c oxidase; DHAP, dihydroxyacetone phosphate; FNR, ferredoxin:NADP$^+$ reductase; NDH, NADH dehydrogenase; PEP, phosphoenolpyruvate; PQ, plastoquinone; PRFI-AICAR-P, phosphoribosyl-formimino-AICAR-phosphate; and PRPP, 5-phosphoribosyl 1-pyrophosphate.

essential for the function of photosystems I and II, the cytochrome b_6/f complex, the F-type ATPase, and photosynthetic electron transport are present on the chromatophore genome, as is a complete set of genes for antenna proteins (only phycoerythrin is lacking). However, the genes *psaK*, *petJ*, and *psaE*, whose products are clearly involved in cyanobacterial photosynthesis, are not present on the chromatophore genome. PsaK, a small protein of unknown function, also is absent in the free-living *Gloeobacter violaceus*, and cyanobacterial *psaK* knockout mutants showed no severe effects on growth and photosynthetic performance [36]. *PetJ* encodes the electron transfer protein cytochrome c_6 that can be functionally replaced by plastocyanin [37] and is also absent in *Prochlorococcus marinus* CCMP1986. In contrast, *psaE* is conserved in all cyanobacterial genomes analyzed in this study. Deletion of *psaE* was reported to have only minor effects on photoautotrophic growth of cyanobacterial strains [38], but in *psaE* mutants high-affinity HCO_3^- uptake could not be induced nor could the mutants be grown under low-CO_2 conditions [39]. In accordance with this observation, only the low-affinity C_i-transport systems (BicA and the NDH-1_4 complex), but not the high-affinity C_i transporters (SbtA), are encoded on the chromatophore genome. We conclude that the *Paulinella* chromatophore, which is exposed to an intracellular, high-CO_2 environment, is under no selection pressure to retain genes functional in a low-CO_2 environment.

C_i taken up into the chromatophore can be fixed via the Calvin cycle. All Calvin cycle enzymes are encoded on the chromatophore genome. However, the ability to store photosynthetic products in the form of glycogen and to synthesize sucrose, which are both well conserved in free-living cyanobacteria, were lost. This is in agreement with the notion that the chromatophores provide reduced carbon in the form of low-molecular-weight assimilates to the host [15].

Other Biosynthetic Capabilities Encoded on the Chromatophore Genome

Besides its photosynthetic machinery, the chromatophore retained other biosynthetic capabilities that could be beneficial

to the host. Amino acids such as the pyruvate-derived Ala, Val, Ile, and Leu can be synthesized and, if appropriate precursors are provided, additional amino acids (see above). The chromatophore genome also encodes a set of enzymes for assimilatory sulfate reduction for type II fatty acid biosynthesis, and for biosynthesis of the essential cofactors lipoic acid and folate (Figure 3). It is tempting to hypothesize that only those biosynthetic capabilities were retained by the chromatophore that complemented requirements of the host previously met by phagotrophic nutrition. Similar mechanisms are found in other endosymbiotic systems: In the symbiosis between *Buchnera* and its aphid host, the host depends on its endosymbionts for the delivery of essential amino acids (lacking in phloem sap, the exclusive diet of the host), whereas the host supplies nonessential amino acids to the endosymbiont [29]. In the tsetse fly-*Wigglesworthia* symbiosis, the host depends on the endosymbionts for the synthesis of vitamins of the B group that are lacking in the meals of the blood-sucking fly, whereas the fly supplies the endosymbionts with the majority of amino acids [31].

Transport

The metabolic interdependence, deduced from the gene complement encoded on the chromatophore genome, shows how intimate the association between host and chromatophore is, requiring efficient transport systems in both directions. However, the active transport capabilities of the chromatophore are strongly reduced, as our analyses revealed (Figure 3). Only phosphate and cobalt appear to be taken up actively over the chromatophore membrane via an ABC-transporter system, whereas ABC transporters for polar amino acids, urea, phosphonates, oligopeptide/nickel, zinc/manganese, and iron that are present in nearly all α-cyanobacteria are not encoded on the chromatophore genome. Most strikingly, we found no uptake system for nitrogen compounds encoded on the chromatophore genome. There are several secondary transport systems for small inorganic ions. Larger molecules may be exported by an ABC transporter annotated as a multidrug efflux system of unknown specificity. Proteins can be exported via the Sec system, although the proteins SecD and SecF, which facilitate protein export and are associated with the Sec system in all free-living cyanobacteria, are missing. The Tat system, which catalyzes the transport of folded proteins in all cyanobacteria analyzed as well as in plant thylakoids [40], however, lacks essential components.

The paucity of chromatophore-encoded membrane transport systems parallels the situation in genomes of bacterial endosymbionts [41]. It has been suggested that the host digests some of its endosymbionts to acquire specific metabolites [30]. This can be excluded for *P. chromatophora* because unlike invertebrate hosts, each *Paulinella* cell contains only one or two chromatophores. This raises the question whether host-derived transport systems could instead be involved in metabolite or even protein exchange between host and chromatophore. In agreement with this idea, it has recently been reported that the majority of plastid solute transporters in Plantae were derived from the host and hypothesized that the insertion of host-derived transporters provided a mechanism to rapidly establish metabolic control over the endosymbiont [42].

Cell Wall and Outer Membrane

Peptidoglycan synthesis by the chromatophore allows the formation of a rigid cell wall, which confers osmotic stability to the chromatophore. However, the pathway analysis revealed that two genes involved in the biosynthesis of the building blocks of peptidoglycan are missing on the chromatophore genome; namely, the bifunctional protein GlmU, with UDP-N-acetylglucosamine pyrophosphorylase and glucosamine-1-phosphate N-acetyltransferase activity, and MurF, the ligase catalyzing ligation of D-ala-D-ala to the growing peptide chain linked to UDP-N-acetylmuramic acid. GlmU seems to be missing on the genome of *Prochlorococcus marinus* NATL1A as well, and its function may be taken over by another not-yet-characterized protein. In contrast, MurF is encoded in all free-living cyanobacteria analyzed. For the assembly of the peptidoglycan polymer, there is a minimal set of four penicillin-binding proteins (PBPs), as described for other α-cyanobacteria [43]. The class A high-molecular-weight PBP (MrcB) is C terminally truncated compared to WH5701 (Table 2); however, the length of MrcB is not well conserved across cyanobacteria. The overall number of genes involved in cell wall and membrane biogenesis is strongly reduced on the chromatophore genome compared to WH5701, as revealed by the analysis of clusters of orthologous groups [44] (see Table S4B). Particularly, the ability to synthesize lipopolysaccharides (LPS) was lost completely. Interestingly, in the aphid endosymbiont *Buchnera aphidicola*, which resides in host cell vacuoles, LPS synthesis also is impaired, whereas in other endosymbiotic bacteria, which lie free in the host's cytoplasm, LPS biosynthesis has been retained [41]. Because no genes encoding porins, typical outer-membrane channel proteins, could also be identified on the chromatophore genome, it is possible that the original outer (LPS) membrane was lost and that chromatophores are enclosed in a host-derived vacuole, as had been proposed earlier based on ultrastructural observations [12]. In contrast, the outer membrane of primary plastids is thought to have been derived from the cyanobacterial outer membrane [45].

Cell Division

In the *Paulinella* symbiosis, division of chromatophore and host cell are strictly coupled: During cell division one of the two chromatophores is transferred to the daughter cell, then the original state is restored by division of the chromatophore in mother and daughter cells [46]. Additionally, the size and shape of the chromatophores differ profoundly from those of all other members of the *Synechococcus*/*Prochlorococcus* clade in that the chromatophores are about 20 times longer and sausage shaped. This begs the questions regarding how division of the chromatophore is regulated and how the host exerts control over it.

In plants, several of the plastid division proteins are of cyanobacterial origin: FtsZ1 and FtsZ2, ARC6 (a homolog of the cyanobacterial Ftn2), ARC3 (a chimeric protein with N-terminal homology to prokaryotic FtsZ), MinD, E, and GC-1 (a protein with low homology to cyanobacterial SulA). All of these proteins are encoded on the nuclear genome and targeted to the plastid by N-terminal transit peptides; thus division of the organelle is controlled by the nucleus [47].

On the chromatophore genome we identified most of the cell-division genes commonly found in α-cyanobacteria (*ftsZ*, *ftn2*, *minC*, *D* and *E, cdv1*, and *cdv2*) [48]. However, no *sulA* homolog was identified. In *E. coli*, SulA functions as a cell-division inhibitor blocking polymerization of FtsZ [49]. Although SulA is functionally not well characterized in cyanobacteria, SulA was determined to be essential for cytokinesis in *Synechocystis* [50] and is encoded on all cyanobacterial genomes analyzed. Therewith, *sulA* represents a candidate gene that if

transferred to the nucleus could enable the host to exert control over chromatophore division. Furthermore, *ftn2* is altered severely with respect to its homolog in free-living *Synechococcus* spp. by fusion to the neighboring ORF (PCC0126) (Table 2) and may be under positive selection (see Table S2). Most remarkably, *ftn2* is the only gene fusion that we observed on the chromatophore genome compared to WH5701. Ftn2 is involved in cytokinesis of cyanobacteria and has a conserved N-terminal DnaJ domain [51] that interacts with specific Hsp70 chaperones. Ftn2 has been proposed to be part of a chaperone system that promotes the formation of FtsZ polymers during cell division [52]. Truncation at both ends of *ftn2* in *Anabaena* sp. PCC 7120 yielded mutants with strongly elongated vegetative cells [51]. Because in the chromatophore Ftn2 is heavily truncated at the N-terminal end, lacking the otherwise highly conserved DnaJ domain, the distinctive phenotype of the chromatophores could be caused by the expression of an altered Ftn2.

DNA Replication and Repair

The *Paulinella* chromatophore has retained the main functions of DNA replication, encoding DNA helicase (*dnaB*), the single-strand binding protein SSB, DNA primase (*dnaG*), gyrase (*gyrA* and *gyrB*), topoisomerase (*topA*), the replication initiation protein DnaA, and an α-cyanobactrial set of DNA polymerase III subunits. Unexpectedly, the NAD-dependent DNA ligase (*ligA*) is missing.

In contrast, capabilities for regulation of transcription as well as DNA repair are limited as in many intracellular bacterial symbionts [28–31]. Only 6 of 35 genes encoding transcription factors identified in WH5701 survived reductive genome evolution in the chromatophore. DNA polymerase I (*polA*), with a main function in repair DNA synthesis, is missing on the chromatophore genome as well as the genes for UV-excision *uvrABCD*, the repair ATPase *recN*, and the base excision repair gene *mutY*. Only the recombinational repair pathway seems to be functional in the *Paulinella* chromatophores: RecA, RecFOR as well as RuvABC are encoded; however, RecBCD as well as the alternative initiator proteins RecQ, RecJ or UvrD are missing.

Time Scale of the Symbiosis

Due to the lack of a fossil record for *Paulinella chromatophora*, it is difficult to estimate the evolutionary age of the endosymbiosis. The mode and tempo of genome reduction has been studied in *Buchnera aphidicola* in a recent evolutionary time scale [53]. A typical pattern starts with the inactivation of a gene producing a pseudogene, followed by progressive loss of its DNA. Comparison with the fossil record revealed that an inactivated gene required 40–60 Myr to disintegrate almost completely. The gene losses encountered in the chromatophore genome, compared to a free-living relative, suggest similar modes of genome reduction during evolution. Presumably, genetic material between synteny groups was lost at an early stage of endosymbiosis followed by excision of dispensable genes from syntenic regions. The pseudogenes found on the chromatophore genome indicate that gene loss is an ongoing process. If the conclusions drawn from *B. aphidicola* can be transferred to *P. chromatophora*, the minimum age of this second primary endosymbiosis between a cyanobacterial symbiont and a eukaryotic host is 60 Myr.

Conclusions

Like plastids chromatophores are photosynthetic entities that are continuously inherited by progeny cells and depend on their host for growth and survival. However, chromatophores differ from plastids not only in their phylogenetic origin but also in the extent of genome reduction and the lack of partitioning of photosynthesis genes between plastid and host genomes. Due to the unicellularity of the host, genetic integration of the chromatophore into the host's metabolism by endosymbiotic gene transfer is plausible. If indeed and to what extent gene transfer occurred remains to be determined and requires sequence analysis of the *Paulinella* nuclear genome and its expressed genes. Until this information becomes available, we prefer to use the term "chromatophore" over "organelle" or "endosymbiont" for the photosynthetic inclusion of *Paulinella chromatophora*, as originally proposed by its discoverer, Robert Lauterborn [10].

Our results provide insight into possible early steps in the evolution of a photosynthetic organelle. Integration of a cyanobacterial symbiont into a eukaryotic host is apparently facilitated by the loss of coding capacity for essential biosynthetic functions unrelated to photosynthesis; i.e., synthesis of amino acids and cofactors. It is likely that provision of such functions by the host rendered genes on the chromatophore genome coding for these functions superfluous. If future evolution of the chromatophore will result in a canonical organelle and to what extent this organelle will resemble extant plastids is a matter of speculation. However, also finding differences to plastid evolution should be of value for understanding why the origin of a photosynthetic eukaryote has been such a rare event.

Experimental Procedures

Culture of *Paulinella chromatophora*

Paulinella chromatophora is difficult to cultivate. To our knowledge there exists only one isolate in culture (strain M0880, Culture collection Melkonian, University of Cologne, Germany), which contains at least 18 different bacterial contaminants [21]. Therefore, *Paulinella* cells from strain M0880 were washed by serial transfer of single cells through drops of sterile culture medium. In this way a new clonal culture was established (strain CCAC 0185, available through the Culture Collection of Algae at the University of Cologne; http://www.ccac.uni-koeln.de). Examination of this culture (see Supplemental Experimental Procedures) revealed some remaining bacterial contaminants that were identified as members of the Caulobactales (*Asticcacaulis* sp. and *Caulobacter* sp.) and Sphingomondales.

The newly prepared *Paulinella* culture was grown at 14 ± 2°C in Fernbach flasks containing 0.75 l WARIS-H [54] modified by addition of 0.75 mM Na_2SiO_3. Light intensity was adjusted to 1.5–3 $\mu E\ m^{-2}\ s^{-1}$ by using a light/dark cycle of 14/10 hr.

Genome Sequencing and Annotation

DNA from isolated chromatophores (procedure as described in the Supplemental Experimental Procedures) was used for traditional Sanger sequencing on a small insert library to a depth of 3× (methods essentially as described in [55]). Material extracted in the same way was sequenced by using a GS20 machine (454, Roche), which yielded 15 contigs from 30 Mb raw reads.

Both assemblies were merged and assembled. This yielded one large contig (average G + C content, 38%) representing the chromatophore genome and a number of small contigs (average G + C content, 47.4%), which were considered as derived from the host nuclear genome or bacterial contaminants (Table S5). The remaining gap in the chromatophore sequence was closed by generating a PCR product connecting both ends of the large contig by using custom primers. Three results indicate that the whole chromatophore genome was assembled: (1) the assembly yielded only one large contig, (2) this contig could be circularized, and (3) the sequences not included in the assembly differed significantly in their G + C content from that of the chromatophore genome. The bacterial contaminants were identified by BLAST searches as species similar to *Caulobacter* or *E. coli*.

A marked change in the fraction of purine residues at the origin of replication (Figure 1) indicated correct assembly [56]. Bacterial chromosomal

origins are commonly found in the vicinity of the *dnaA* gene. Thus, we annotated the genome starting from the *dnaA* gene. An initial gene prediction was performed by using GeneMarkS [57]. Of the 892 potential protein-coding genes defined, 25 were either removed as potential false positives or tagged as potential pseudogenes after manual inspection. Genes encoding tRNAs were identified by using tRNAscan-SE [58]. Procedures for determination of gene orthology as well as potential positive selection are presented in the Supplemental Experimental Procedures.

Accession Numbers

The complete chromatophore genome sequence has been deposited in GenBank under the accession number CP000815.

Supplemental Data

Supplemental Results, Experimental Procedures, and five tables are available at http://www.current-biology.com/cgi/content/full/18/6/410/DC1/.

Acknowledgments

E.C.M.N. was supported by a grant from the International Graduate School in Genetics and Functional Genomics at the University of Cologne, funded by the State of North Rhine-Westphalia.

Received: January 17, 2008
Revised: February 14, 2008
Accepted: February 15, 2008
Published online: March 20, 2008

References

1. Keeling, P.J. (2004). Diversity and evolutionary history of plastids and their hosts. Am. J. Bot. *91*, 1481–1493.
2. Moreira, D., Le Guyader, H., and Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. Nature *405*, 69–72.
3. Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. Curr. Biol. *15*, 1325–1330.
4. Palmer, J.D. (2003). The symbiotic birth and spread of plastids: How many times and whodunit? J. Phycol. *39*, 4–11.
5. Dyall, S.D., Brown, M.T., and Johnson, P.J. (2004). Ancient invasions: From endosymbionts to organelles. Science *304*, 253–257.
6. Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., and Philippe, H. (2004). The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. USA *101*, 15386–15391.
7. Marin, B., Nowack, E.C., and Melkonian, M. (2005). A plastid in the making: Evidence for a second primary endosymbiosis. Protist *156*, 425–432.
8. Bhattacharya, D., Helmchen, T., and Melkonian, M. (1995). Molecular evolutionary analyses of nuclear-encoded small-subunit ribosomal RNA identify an independent rhizopod lineage containing the Euglyphina and the Chlorarachniophyta. J. Eukaryot. Microbiol. *42*, 65–69.
9. Cavalier-Smith, T., and Chao, E.E.Y. (2003). Phylogeny and classification of phylum Cercozoa (Protozoa). Protist *154*, 341–358.
10. Lauterborn, R. (1895). Protozoenstudien. Z. Wiss. Zool. *59*, 537–544.
11. Marin, B., Nowack, E.C.M., Glöckner, G., and Melkonian, M. (2007). The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like gammaproteobacterium. BMC Evol. Biol. *7*, 85. 10.1186/1471-2148-7-85.
12. Kies, L. (1973). Electron microscopical investigations on *Paulinella chromatophora* Lauterborn, a thecamoeba containing blue-green endosymbionts (cyanelles). Protoplasma *80*, 69–89.
13. Hannah, F., Rogerson, A., and Anderson, O.R. (1996). A description of *Paulinella indentata* n. sp. (Filosea: Euglyphina) from subtidal coastal benthic sediments. J. Eukaryot. Microbiol. *43*, 1–4.
14. Johnson, P.W., Hargraves, P.E., and Sieburth, J.M. (1988). Ultrastructure and ecology of *Calycomonas ovalis* Wulff, 1919 (Chrysophyceae), and its redescription as a testate rhizopod, *Paulinella ovalis* n. comb. (Filosea: Euglyphina). J. Protozool. *35*, 618–626.
15. Kies, L., and Kremer, B.P. (1979). Function of cyanelles in the thecamoeba *Paulinella chromatophora*. Naturwissenschaften *66*, 578–579.
16. Pascher, A. (1929). Über die Natur der blaugrünen Chromatophoren des Rhizopoden *Paulinella chromatophora*. Zool. Anz. *81*, 189–194.
17. Penard, E. (1905). Notes sur quelques Sarcodinés - 12. *Paulinella chromatophora* Lauterborn. Rev. Suisse Zool. *13*, 585–616.
18. Bodył, A., Mackiewicz, P., and Stiller, J.W. (2007). The intracellular cyanobacteria of *Paulinella chromatophora*: Endosymbionts or organelles? Trends Microbiol. *15*, 295–296.
19. Bhattacharya, D., and Archibald, J.M. (2006). The difference between organelles and endosymbionts - response to Theissen and Martin. Curr. Biol. *16*, R1017–R1018.
20. Theissen, U., and Martin, W. (2006). The difference between organelles and endosymbionts. Curr. Biol. *16*, R1016–R1017.
21. Yoon, H.S., Reyes-Prieto, A., Melkonian, M., and Bhattacharya, D. (2006). Minimal plastid genome evolution in the *Paulinella* endosymbiont. Curr. Biol. *16*, R670–R672.
22. Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., et al. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature *424*, 1042–1047.
23. Wernegreen, J.J. (2002). Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. *3*, 850–861.
24. Kaneko, T., Matsubayashi, T., Sugita, M., and Sugiura, M. (1996). Physical and gene maps of the unicellular cyanobacterium *Synechococcus* sp. strain PCC6301 genome. Plant Mol. Biol. *31*, 193–201.
25. Mira, A., Ochman, H., and Moran, N.A. (2001). Deletional bias and the evolution of bacterial genomes. Trends Genet. *17*, 589–596.
26. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35* (*Suppl 2*), W182–W185.
27. Battersby, A.R., Fookes, C.J.R., Matcham, G.W.J., McDonald, E., and Gustafson-Potter, K.E. (1979). Biosynthesis of the natural porphyrins: Experiments on the ring-closure steps and with the hydroxy-analogue of porphobilinogen. J. Chem. Soc. Chem. Commun. *7*, 316–319.
28. Gil, R., Silva, F.J., Zientz, E., Delmotte, F., Gonzalez-Candelas, F., Latorre, A., Rausell, C., Kamerbeek, J., Gadau, J., Holldobler, B., et al. (2003). The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. Proc. Natl. Acad. Sci. USA *100*, 9388–9393.
29. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature *407*, 81–86.
30. Kuwahara, H., Yoshida, T., Takaki, Y., Shimamura, S., Nishi, S., Harada, M., Matsuyama, K., Takishita, K., Kawato, M., Uematsu, K., et al. (2007). Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. Curr. Biol. *17*, 881–886.
31. Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., and Aksoy, S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. *32*, 402–407.
32. Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., McGraw, E.A., Martin, W., Esser, C., Ahmadinejad, N., et al. (2004). Phylogenomics of the reproductive parasite *Wolbachia pipientis wMel*: A streamlined genome overrun by mobile genetic elements. PLoS Biol. *2*, e69. 10.1371/journal.pbio.0020069.
33. Hotopp, J.C.D., Clark, M.E., Oliveira, D., Foster, J.M., Fischer, P., Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S.L., et al. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science *317*, 1753–1756.
34. Andersson, J.O. (2005). Lateral gene transfer in eukaryotes. Cell. Mol. Life Sci. *62*, 1182–1197.
35. Doolittle, W.F. (1998). You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. *14*, 307–311.
36. Nakamoto, H., and Hasegawa, M. (1999). Targeted inactivation of the gene *psaK* encoding a subunit of photosystem I from the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol. *40*, 9–16.
37. Duran, R.V., Hervas, M., De La Rosa, M.A., and Navarro, J.A. (2004). The efficient functioning of photosynthesis and respiration in *Synechocystis* sp. PCC 6803 strictly requires the presence of either cytochrome c6 or plastocyanin. J. Biol. Chem. *279*, 7229–7233.
38. Zhao, J., Snyder, W.B., Mühlenhoff, U., Rhiel, E., Warren, P.V., Golbeck, J.H., and Bryant, D.A. (1993). Cloning and characterization of the *psaE* gene of the cyanobacterium *Synechococcus* sp. PCC

7002: Characterization of a *psaE* mutant and overproduction of the protein in *Escherichia coli*. Mol. Microbiol. *9*, 183–194.
39. Sültemeyer, D., Price, G.D., Bryant, D.A., and Badger, M.R. (1997). PsaE- and NdhF-mediated electron transport affect bicarbonate transport rather than carbon dioxide uptake in the cyanobacterium *Synechococcus* sp. PCC7002. Planta *201*, 36–42.
40. Berks, B.C., Palmer, T., and Sargent, F. (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. Curr. Opin. Microbiol. *8*, 174–181.
41. Zientz, E., Dandekar, T., and Gross, R. (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol. Mol. Biol. Rev. *68*, 745–770.
42. Tyra, H.M., Linka, M., Weber, A.P.M., and Bhattacharya, D. (2007). Host origin of plastid solute transporters in the first photosynthetic eukaryotes. Genome Biol. *8*, R212.
43. Leganés, F., Blanco-Rivero, A., Fernández-Piñas, F., Redondo, M., Fernández-Valiente, E., Fan, Q., Lechno-Yossef, S., and Wolk, C.P. (2005). Wide variation in the cyanobacterial complement of presumptive penicillin-binding proteins. Arch. Microbiol. *184*, 234–248.
44. Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., et al. (2007). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *35*, D5–D12.
45. Duy, D., Soll, J., and Philippar, K. (2007). Solute channels of the outer membrane: from bacteria to chloroplasts. Biol. Chem. *388*, 879–889.
46. Hoogenraad, H.R. (1927). Zur Kenntnis der Fortpflanzung von *Paulinella chromatophora* Lauterb. Zool. Anz. *72*, 140–150.
47. Maple, J., and Møller, S.G. (2007). Plastid division: Evolution, mechanism and complexity. Ann. Bot. (Lond.) *99*, 565–579.
48. Miyagishima, S.Y., Wolk, C.P., and Osteryoung, K.W. (2005). Identification of cyanobacterial cell division genes by comparative and mutational analyses. Mol. Microbiol. *56*, 126–143.
49. Justice, S.S., García-Lara, J., and Rothfield, L.I. (2000). Cell division inhibitors SulA and MinC/MinD block septum formation at different steps in the assembly of the *Escherichia coli* division machinery. Mol. Microbiol. *37*, 410–423.
50. Raynaud, C., Cassier-Chauvat, C., Perennes, C., and Bergounioux, C. (2004). An Arabidopsis homolog of the bacterial cell division inhibitor SulA is involved in plastid division. Plant Cell *16*, 1801–1811.
51. Koksharova, O.A., and Wolk, C.P. (2002). A novel gene that bears a DnaJ motif influences cyanobacterial cell division. J. Bacteriol. *184*, 5524–5528.
52. Osteryoung, K.W., and Nunnari, J. (2003). The division of endosymbiotic organelles. Science *302*, 1698–1704.
53. Gómez-Valero, L., Latorre, A., and Silva, F.J. (2004). The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. Mol. Biol. Evol. *21*, 2172–2181.
54. McFadden, G.I., and Melkonian, M. (1986). Use of Hepes buffer for microalgal culture media and fixation for electron microscopy. Phycologia *25*, 551–557.
55. Roe, B.A. (2004). Shotgun library construction for DNA sequencing. Methods Mol. Biol. *255*, 171–187.
56. Freeman, J.M., Plasterer, T.N., Smith, T.F., and Mohr, S.C. (1998). Patterns of genome organization in bacteria. Science *279*, 1827.
57. Besemer, J., and Borodovsky, M. (2005). GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. *33*, W451–W454.
58. Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.