

SCIENTIFIC REPORTS



OPEN

Evolutionary dynamics of the chromatophore genome in three photosynthetic *Paulinella* species

Duckhyun Lhee¹, Ji-San Ha¹, Sunju Kim², Myung Gil Park³, Debashish Bhattacharya⁴ & Hwan Su Yoon¹

The thecate amoeba *Paulinella* is a valuable model for understanding plastid organogenesis because this lineage has independently gained plastids (termed chromatophores) of alpha-cyanobacterial provenance. Plastid primary endosymbiosis in *Paulinella* occurred relatively recently (90–140 million years ago, Mya), whereas the origin of the canonical Archaeplastida plastid occurred >1,500 Mya. Therefore, these two events provide independent perspectives on plastid formation on vastly different timescales. Here we generated the complete chromatophore genome sequence from *P. longichromatophora* (979,356 bp, GC-content = 38.8%, 915 predicted genes) and *P. micropora* NZ27 (977,190 bp, GC-content = 39.9%, 911 predicted genes) and compared these data to that from existing chromatophore genomes. Our analysis suggests that when a basal split occurred among photosynthetic *Paulinella* species ca. 60 Mya, only 35% of the ancestral orthologous gene families from the cyanobacterial endosymbiont remained in chromatophore DNA. Following major gene losses during the early stages of endosymbiosis, this process slowed down significantly, resulting in a conserved gene content across extant taxa. Chromatophore genes faced relaxed selection when compared to homologs in free-living alpha-cyanobacteria, likely reflecting the homogeneous intracellular environment of the *Paulinella* host. Comparison of nucleotide substitution and insertion/deletion events among different *P. micropora* strains demonstrates that increases in AT-content and genome reduction are ongoing and dynamic processes in chromatophore evolution.

The endosymbiotic integration of a cyanobacterium into a eukaryotic host cell gave rise to the common ancestor of the Archaeplastida that subsequently split into the three primary plastid lineages, Glaucophyta, Viridiplantae, and Rhodophyta^{1–3}. This primary endosymbiosis is estimated to have occurred >1.5 billion years ago^{4,5}. Primary plastid origin is however exceedingly rare, with the only other known case being the relatively recent (i.e., 90–140 Mya⁶) independent origin of this organelle in photosynthetic *Paulinella* species^{7,8}. Photosynthetic *Paulinella*, a lineage of thecate filose amoebae is a member of the Cercozoa (Rhizaria)^{9,10}, has two sausage-shaped or U-shaped plastids termed “chromatophores”, which were derived from a “*Synechococcus*-like” alpha-cyanobacterium^{11,12}.

Three photosynthetic *Paulinella* species have thus far been reported, two that are freshwater, *P. chromatophora*⁷ and *P. micropora*^{13,14}, and one marine, *P. longichromatophora*¹⁵. These lineages are expected to provide invaluable information because comparing the genome of free-living cyanobacterial relatives to that of the endosymbionts offers fundamental insights into organogenesis. Three chromatophore genome sequences exist, from *P. chromatophora* strain CCAC 0185¹⁶ and two *P. micropora* strains FK01 and KR01^{14,17}. Compared to the free-living cyanobacterium, *Synechococcus* sp. WH 5701 that is closely related to the donor lineage, the genomes of chromatophores encode relatively more genes with assigned functions than those that lack annotation, suggesting that the latter set may be involved in responses to environmental change that were lost due to the shift to a homogeneous intracellular environment¹⁶. However, many genes with known and essential functions have also been lost from the chromatophore genome, suggesting that this organelle is dependent on its host for growth and survival. This idea was supported by the inability of isolated chromatophores to divide autonomously¹⁸ and more recently, by the finding of specific targeting sequences at the N-terminus of large cytosolic proteins to allow

¹Department of Biological Sciences, Sungkyunkwan University, Suwon, 16419, Republic of Korea. ²Department of Oceanography, Pukyong National University, Busan, 48513, Republic of Korea. ³Department of Oceanography, Chonnam National University, Gwangju, 61186, Republic of Korea. ⁴Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, 08901, USA. Correspondence and requests for materials should be addressed to H.S.Y. (email: hsyoon2011@skku.edu)

	Length (bp)	GC-content (%)	CDS	rRNA	tRNA	Pseudogene	GenBank Accession Number
<i>P. longichromatophora</i>	9,79,356	38.8	867	6	42	5	MG264610
<i>P. micropora</i> NZ27	9,77,190	39.9	863	6	42	3	MG976688
<i>P. micropora</i> KR01	9,76,991	39.9	860	6	42	4	KX897545
<i>P. micropora</i> FK01	9,77,204	39.9	860	6	42	4	KY124271
<i>P. chromatophora</i> CCAC 0815	1,021,616	38	878	6	42	11	NC_011087

Table 1. A comparison of chromatophore genomes of *Paulinella longichromatophora*, *P. micropora* NZ27, *P. micropora* KR01, *P. micropora* FK01, and *P. chromatophora* CCAC 0185.

their entry into the chromatophore¹⁹. When comparing the chromatophore genomes of *P. chromatophora* CCAC 0185 and *P. micropora* FK01, significant conservation in gene order was found with only a handful of minor gene rearrangements¹⁷. Moreover, inspection of 681 DNA alignments showed that the vast majority had dN/dS ratios $\ll 1$, implying that chromatophore genomes of *P. chromatophora* and *P. micropora* are strongly constrained by selection.

As described above, the chromatophore genomes of *Paulinella* species have experienced reductive genome evolution. Compared to free-living *Synechococcus* species, the chromatophore genome (ca. 1 Mbp) was reduced one-third in size, the GC-content (39%) fell when compared to the donor lineages (~57%), and the substitution rate was elevated¹³. Similar processes have been reported in other host-dependent bacteria, leading to a relatively small genome size, high sequence substitution rate, enriched AT-rich base composition, and metabolic simplification^{20–23}. These reductive processes can be roughly divided into two phases. At the onset, many genes that are not adaptive to the new environment in the host are inactivated, resulting in the formation of pseudogenes or the complete loss of genomic DNA through large deletion events. Inactivation and recombination events are frequent, giving rise to genomes with abundant pseudogenes and rearrangements. Examples of genomes in this stage are *Mycobacterium leprae*^{24,25}, *Bordetella pertussis*²⁶, *Serratia symbiotica*²⁷, *Polynucleobacter necessarius*^{28,29}, *Sitophilus oryzae*³⁰, *Richelia intracellularis*³¹, and the *Epihemia turgida* spheroid body³². In the more advanced stages of genome degradation, most of the unnecessary information has been completely lost, resulting in highly reduced genomes with a scarcity of pseudogenes. Some of the genes related to recombination and repair processes may have been lost, impacting genome stability. The speed of degradation slows down significantly in the final stages of genome reduction. Examples of endosymbiotic bacterial genomes in this stage are *Buchnera aphidicola*^{33–36}, *Blochmannia* spp.^{37,38}, *Calyptogenia okutanii*, *C. magnifica*³⁹, and *Candidatus Atelocyanobacterium thalassa*⁴⁰.

Here we sequenced the chromatophore genome from *Paulinella micropora* NZ27 and from *Paulinella longichromatophora*, the latter being the only known marine photosynthetic *Paulinella* species. Available chromatophore genomes were compared to identify reductive genome dynamics in this clade. We also compared these data to the genomes of free-living cyanobacteria and Archaeplastida plastids to better understand the evolutionary history of photosynthetic organelles of primary endosymbiotic origin.

Results and Discussion

Chromatophore genome structure in *P. longichromatophora* and *P. micropora*. The circular chromatophore genome of *P. longichromatophora* was 979,356 bp in length and comprised 867 protein coding sequences (CDS) with six rRNAs and 42 tRNAs. The GC-content of this genome was 38.8% (Table 1). The length, GC-content, and number of CDS in *P. longichromatophora* were intermediate to those of *P. chromatophora* and *P. micropora*. Among these *Paulinella* chromatophore genomes, we found 12 syntenic blocks (Fig. 1a). Synteny comparison showed that there were six inversions between *P. longichromatophora* and *P. chromatophora* and only three inversions between *P. longichromatophora* and *P. micropora*. This result is consistent with the multi-gene phylogeny (Supplementary Fig. S1), which showed that *P. longichromatophora* is more closely related to *P. micropora* than to *P. chromatophora*. In the case of *Paulinella micropora* NZ27, the chromatophore genome was 977,190 bp in length, the GC-content was 39.9%, and 863 CDSs were identified. The genome structure was conserved among the three *P. micropora* strains. Overall, we find that genome structure of photosynthetic *Paulinella* species is highly constrained.

Between these synteny blocks, some inversion breakpoints were located at the duplicated genes (i.e., *psbA* and *psbD*, respectively). Few repeats are found in the late stages of genome reduction and duplicated genes (e.g., *psbA*, *psbD*) at some inversion breakpoints in chromatophore genomes appear to mediate inversions due to intra-molecular recombination^{14,41,42}. In addition, using the *Synechococcus* sp. WH 5701 genome to infer the common ancestral state, we calculated, using parsimony, the number of chromatophore inversion events (Fig. 1b). *P. micropora* contained one inversion, *P. longichromatophora* had two inversions, whereas *P. chromatophora* had six inversion events when compared to the putative ancestral state.

Genomic changes from cyanobacteria to chromatophores. We compared chromatophore genomes with cyanobacterial and other plastid genomes. The clustering of proteins from three chromatophore genomes, 17 Archaeplastida plastid genomes, and 20 cyanobacterial genomes yielded 3,348 Orthologous Gene Families (OGF) that are shared by at least five taxa. Among these, 584 OGFs, with some additional orthologs were examined in detail (Fig. 2, Supplementary Fig. S2). The plastids of the green lineage (Viridiplantae), red algae, and glaucophytes lost most OGFs, with the exception of photosynthesis and ribosome-related genes. Red algal plastids contained more OGFs in other metabolic pathways when compared to its two sister taxa. In the evolutionary

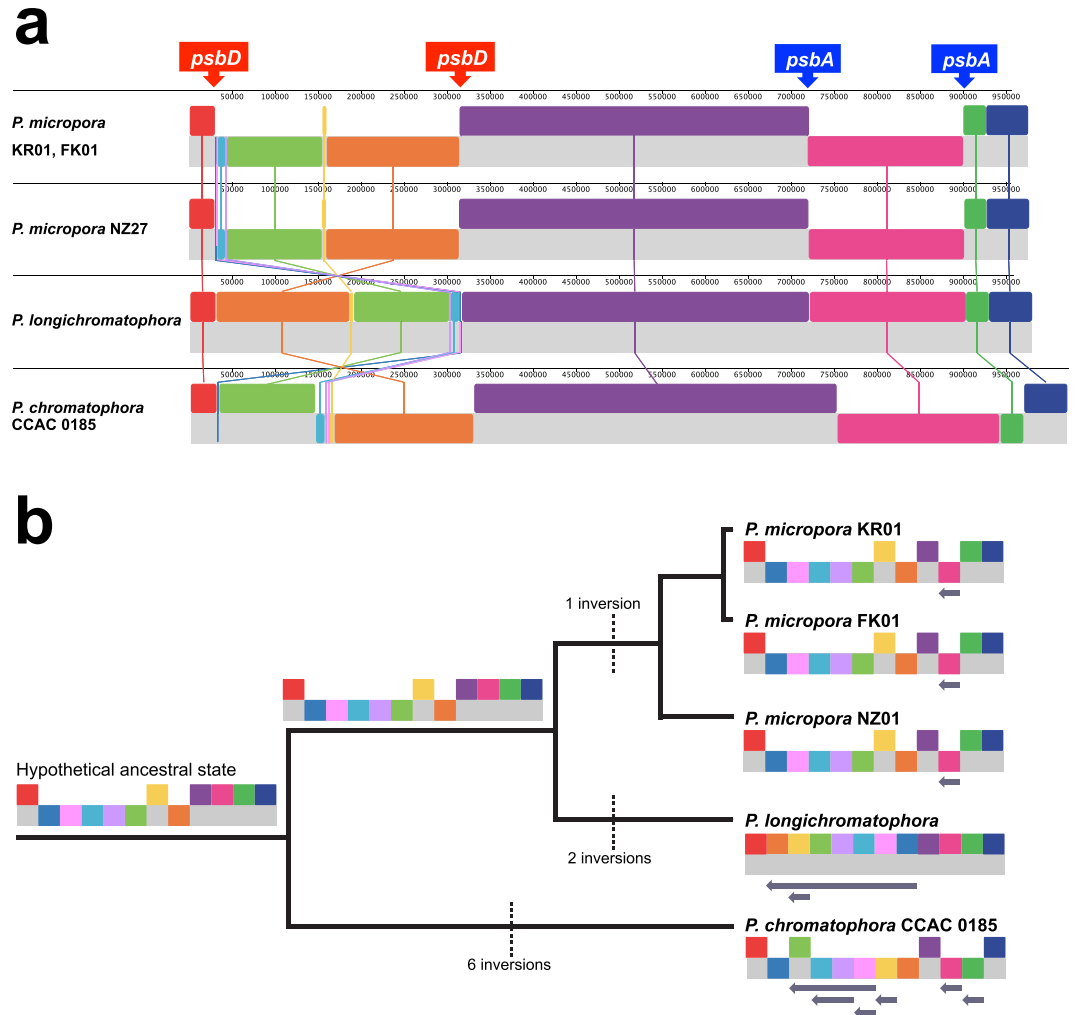


Figure 1. (a) Gene synteny comparison of the chromatophore genomes from *Paulinella longichromatophora*, *P. micropora* NZ27, *P. micropora* KR01, *P. micropora* FK01, and *P. chromatophora* CCAC 0185. (b) Synteny blocks are simplified on the tree and inversion events were calculated using parsimony. Hypothetical ancestral states were constructed by mapping chromatophore genes to the *Synechococcus* sp. WH 5701 genome.

history of Archaeplastida, red algae and the green lineage underwent differential gene loss after their split from a single common ancestor⁴³. Moreover, endosymbiotic gene transfer (EGT) has played an important role in the establishment of primary plastids^{44,45}, whereby a total of 956 plastid-derived genes have been relocated to the nuclear genomes of green plants (886 genes), glaucophytes (48 genes), and red algae (22 genes)⁴³.

As reported in a previous study¹⁶, genome reduction has progressed to a smaller extent in the chromatophore genome of *Paulinella* when compared to Archaeplastida plastids (Fig. 3, Supplementary Table S1). In particular, the citric acid cycle (TCA), environmental information processing, and ABC transporters genes underwent loss. Several amino acid biosynthetic pathways (Glu, Arg, His, Try, Met), cofactors (thiamine, riboflavin, NAD, pantothenate, coenzyme A), and lipopolysaccharides metabolism were also completely lost. In addition, some pathways of amino acid biosynthesis (Gln, Asp, Asn, Lys, Cys, Thr, Tyr, Phe), lipid metabolism, biotin biosynthesis, and the Sec (secretion) system were partially deleted. In contrast, photosynthesis, genetic information processing, and fatty acid biosynthesis functions were retained in chromatophores, as in the plastids of Archaeplastida.

There are two possible routes for organelle genome reduction, outright gene loss or relocation to the nuclear genome via EGT. Lipopolysaccharide synthesis genes may provide an example of complete loss because these were presumably no longer needed in the homogeneous intracellular environment. Lipopolysaccharides are large molecules present in the outer membrane of Gram-negative bacteria⁴⁶. The original outer membrane of the chromatophore is proposed to have been lost⁸. Some of the lost or incomplete pathways in chromatophores were compensated for by nuclear genes in the case of *P. chromatophora* CCAC 0185⁴⁷. For example, missing components of the Met, Ser, Gly, Arg, His, aromatic amino acid, and purine biosynthesis pathways were filled by proteins encoded on the nuclear genome. Interestingly, these nuclear genes were primarily not derived from chromatophore DNA. At least 229 nuclear genes were acquired via horizontal gene transfer (HGT) from various bacteria, and 25% putatively arose through EGT from the chromatophore genome⁴⁷. In a more recent analysis of the chromatophore proteome, Singer *et al.*¹⁹ found that the large majority of imported proteins are host (i.e.,

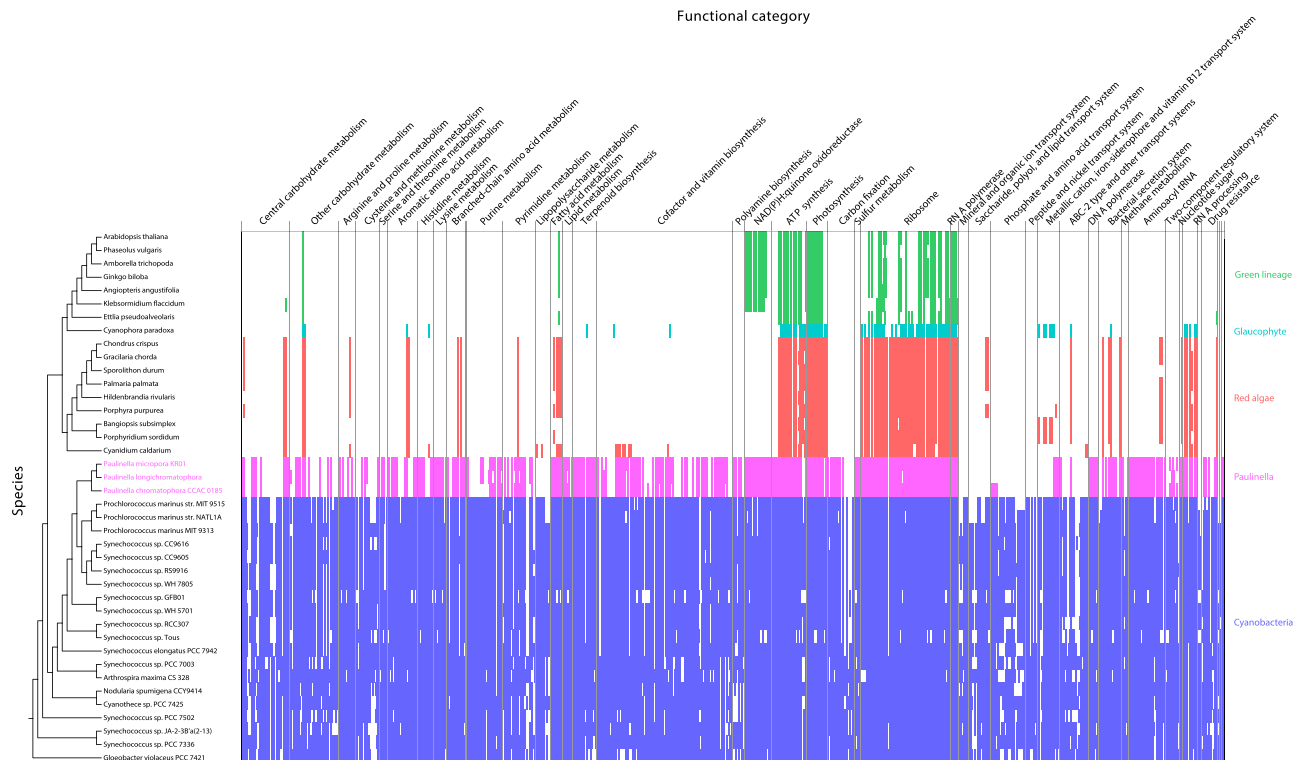


Figure 2. Heatmap with genomes of chromatophores, cyanobacteria, and Archaeplastida plastids categorized by KEGG pathways annotated as Orthologous Gene Families (OGFs). A total of 401 OGFs (x-axis) are ordered according to their KEGG annotation category. Each tick indicates genes present in the OGFs; red indicates red algae, green are green lineage species, cyan indicates glaucophytes, pink indicates *Paulinella*, and blue indicates cyanobacteria. On the left side of heatmap, the tree was constructed in a concatenated phylogenomic analysis of amino acids data using 20 orthologous genes from a total of 40 representative genomes in Archaeplastida plastids, cyanobacteria, and chromatophore genomes of *Paulinella* lineages (Supplementary Table S6).

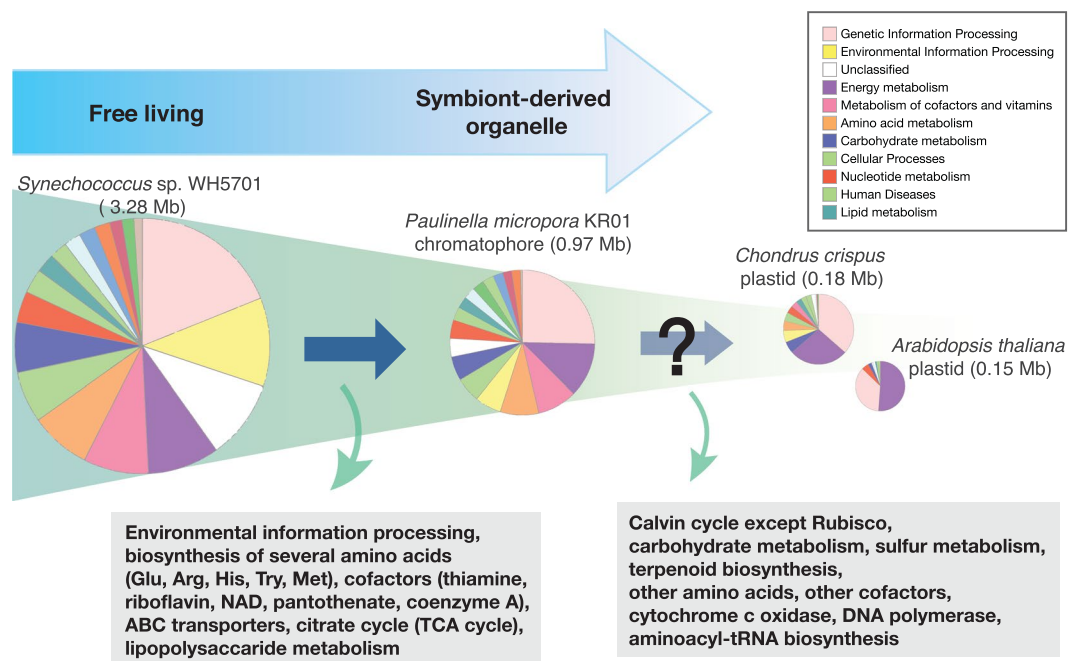


Figure 3. Schematic figure showing the process of organellogenesis in *Paulinella*, starting from a free-living cyanobacterium to the permanent chromatophore. Lost KEGG pathways from cyanobacteria to chromatophore, and different pathways between chromatophore and Archaeplastida plastids are listed.

nuclear gene)-derived or of unknown origin, suggesting that a major redeployment of *Paulinella* genetic resources was the primary means of domesticating the organelle, in addition to HGT and EGT.

Even though lost genes in the endosymbiont can be counteracted by the host, some interesting trends are apparent within the retained inventory. In the case of the obligate endosymbiont *Polynucleobacter*, genes related to DNA, RNA and protein metabolism or energy production were lost less frequently than transporters or genes involved in growth, sensing or regulation²⁹. The chromatophore genome of *Paulinella* also shows this pattern. Genome comparisons among primary endosymbionts (i.e., UCYN-A1, the *Epithemia turgida* spheroid body and the chromatophore of *P. chromatophora*) showed that 62% of chromatophore encoded genes was shared with both of these endosymbionts⁴⁸, suggesting that some core genes are retained despite independent evolution.

Pairwise comparisons of the ratio of substitution rates at non-synonymous and synonymous sites (dN/dS) in free-living *Synechococcus* showed that genes remaining in the chromatophore have lower dN/dS values. *Synechococcus* genes were grouped into three different categories (Supplementary Fig. S3, Table S2): genes retained in all chromatophores (GROUP-1), genes present in the common ancestor of all chromatophores but lost in some lineages (GROUP-2), and genes lost during the transition to chromatophore (GROUP-3). A significant increase in the average value of dN/dS from GROUP-1 (0.405) through GROUP-2 (0.508) to GROUP-3 (0.567) was observed (t-test, p -value < 0.05), suggesting that retained genes were under stronger purifying selection. The average Codon Adaptation Index (CAI) values of *Synechococcus* species were significantly greater in GROUP-1 (0.754) than GROUP-3 (0.732) (t-test, p -value < 0.001), implying that the low CAI genes were more readily lost than high CAI genes. Adaptive codon bias is correlated with gene expression level^{49–51}, thus *Synechococcus* genes remaining in the chromatophore may have higher expression. This finding of gene loss reflecting lower CAI has also been reported in *Buchnera* strains that are insect endosymbiotic bacteria⁵². Although dN/dS and CAI values of free-living *Synechococcus* species do not represent directly the evolutionary dynamics of intracellular chromatophores, we assume this result conveys useful insights into the evolutionary trajectory of chromatophore gene loss. That is, the impact of selection and codon usage bias in ancestral free-living cyanobacteria are broadly applicable to (and conserved) in captured endosymbiont genomes.

In order to elucidate the dynamics of mutation and selection in chromatophore genes, the dN/dS ratio was calculated using data from free-living *Synechococcus* strains and chromatophores. It turned out however that the synonymous sites were largely saturated, thereby weakening this analysis. To address this issue, the RELAX program⁵³ was used to estimate the relaxation or intensification of selection in chromatophore genes when compared to free-living alpha-cyanobacteria (Supplementary Table S3). Among 542 OGF that were well conserved among all species, 325 returned significant results. Among these, only 22 OGFs had intensified selection, whereas the remainder (303 OGFs) showed relaxed selection in chromatophore genes. This category encompassed a variety of functional categories, implying that relaxed selection is occurring at the genome level. This result is consistent with analysis of bacterial endosymbionts in insect species⁵³.

The transition to a host-dependent lifestyle is often coupled with changes in how evolution proceeds^{23,54,55}. Intracellular environments cause relaxed selection on many metabolic functions that are no longer required (or are detrimental [e.g., a competing TCA cycle in the chromatophore]) in a host cellular environment. Population size is also critical because reduction in population size lowers the efficacy of selection, favoring a greater role for random genetic drift. In the case of obligate endosymbionts, restriction to hosts and bottlenecks at the stage of inoculation of new hosts will enhance this process due to Muller's ratchet^{56–59}. Previous studies^{60,61} have shown however that although separate subpopulations will accumulate deleterious mutations rapidly, the symbiont population as a whole is protected from extinction due to strong selection for host survival. In the case of *Paulinella*, population expansion is integrated because the chromatophores divide synchronously with the host⁶². Regardless of the details of this complex host-organelle interaction, what is clear is that Muller's ratchet does not lead to genome extinction in chromatophores (as also found in other plastids) because the genome size has stabilized (see below, Fig. 4b) and many of the retained genes remain essential for host survival. The observation of relaxed selection on chromatophore genes when compared to free-living alpha-cyanobacteria may, therefore, be explained by the homogeneous intracellular environment. This aspect is balanced by very strong selection for the retention of a core set of genes that are critical for organelle function (photosynthesis). If these processes play out in an expanding population, then rare beneficial mutations (including EGT or HGT), may rise to dominance (i.e., through a selective sweep) despite the intracellular, asexual environment that the organelles reside in.

Gene content differences among chromatophore genomes. Because of the presence of duplicated genes, gene content was compared using orthologous gene families (OGFs) rather than gene numbers. More than 90% of the genes (799 OGFs) were shared between the chromatophore genomes of *Paulinella* species. Some of the differentially lost genes were assigned to specific pathways, as found in the case of the *ruvA/ruvB/ruvC* genes that were lost only in *P. longichromatophora*. The chromatophore genome of *P. longichromatophora* shared more OGFs with *P. micropora* (834) than with *P. chromatophora* (819), supporting the sister group relationship of these former species (Supplementary Fig. S1). The novel genome from *P. micropora* NZ27 strain had only two different genes (i.e., oxidoreductase and hypothetical protein) when compared to other *P. micropora* strains (KR01 and FK01). Unique genes in chromatophore genome of each species were less than 3% (~21 OGFs) of total genes and showed no specific functional trends in KEGG pathway analysis (Supplementary Table S4).

To investigate the history of chromatophore genome reduction, we constructed OGFs from chromatophore proteins and from phylogenetically closely related *Synechococcus/Prochlorococcus* species. Ancestral OGF numbers with gain and loss were estimated (Fig. 4a). In addition, divergence times of photosynthetic *Paulinella* species were estimated using a lognormal relaxed clock with 18S rRNA (Supplementary Fig. S4). A previous estimate for the origin of *P. chromatophora* was 90–140 Mya⁶. Our estimate (124 Mya), after adding three additional photosynthetic *Paulinella*, was consistent with the previous finding. The estimate of the ancestral OGF number and the molecular clock analysis were used to derive a model of gene family reduction (Fig. 4b). After a former

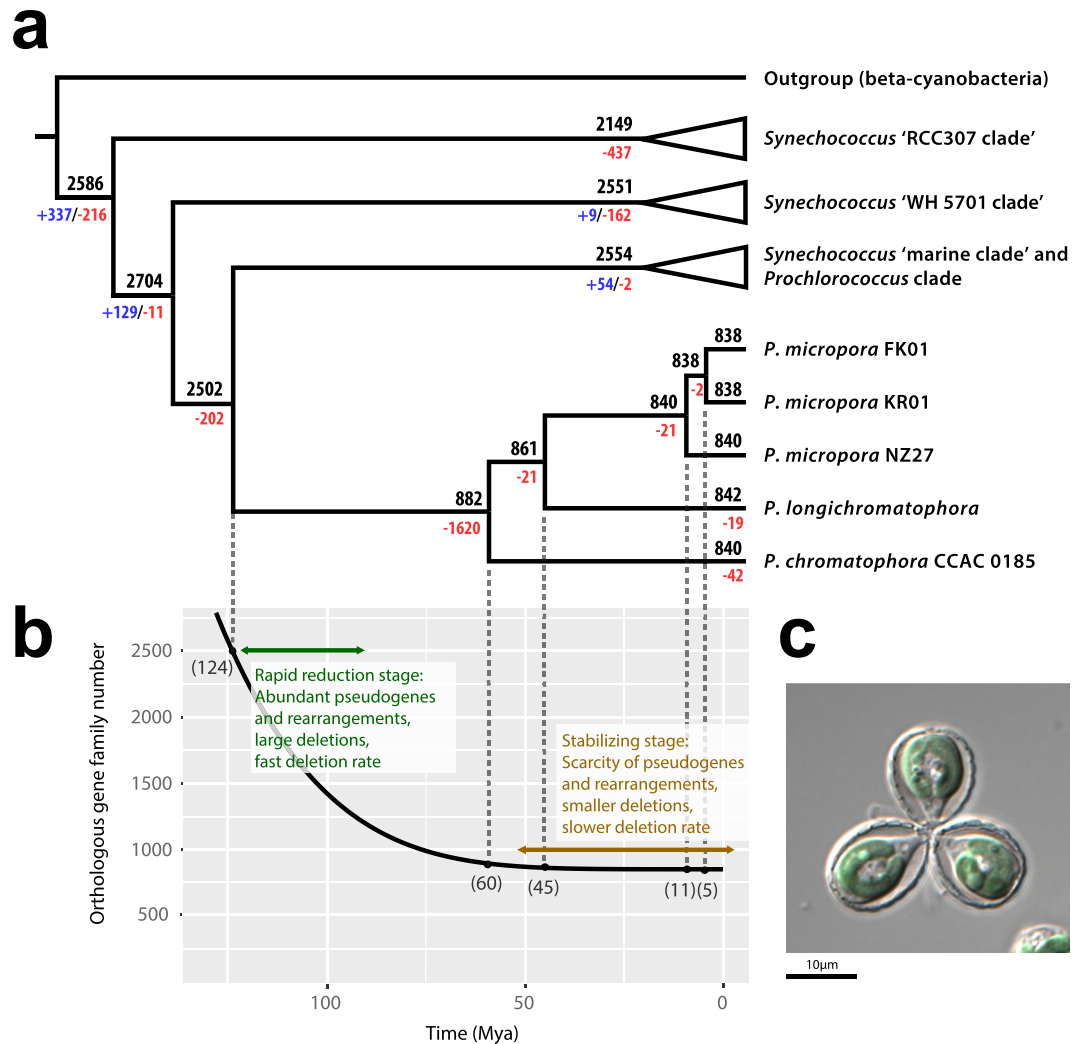


Figure 4. (a) Summary of orthologous gene families (OGF) loss events across the phylogenetic tree of photosynthetic *Paulinella*. This figure presents the total number of OGFs in each species or node estimated by Dollo parsimony. The numbers on branches of the phylogenetic tree correspond to expanded (blue), contracted (red), or inferred ancestral (black) OGFs along each lineage. (b) Proposed evolutionary model of OGFs in the chromatophore genome of *Paulinella*. The split time was estimated according to SSU rRNA lognormal molecular clock calculations. Numbers inside the parentheses indicate the estimated split time (Mya). (c) Light microscopy image of *Paulinella micropora* KR01.

free-living cyanobacterium became the endosymbiont (124 Mya), genes were lost very rapidly during the early stages of plastid evolution. When photosynthetic *Paulinella* species diverged from a common ancestor (60 Mya), 1,620 OGFs were lost, leaving 882 OGFs in the chromatophore; i.e., 65% of the gene inventory in the ancestral free-living cyanobacterium was lost outright. Subsequently, gene loss slowed down and only 40–44 OGFs were lost during the past ca. 60 Mya. Therefore, a small proportion (<5%) of coding sequences was differentially lost and rearrangement occurred rarely among these three species. It should be noted that the molecular analysis usually provides a minimal age for the divergence, hence our estimated divergence times are the minimal ages, and actual divergence time could be much older. Moreover, there is a possibility that extinct alpha-cyanobacteria are the closest relatives of chromatophores. Nonetheless, the current genome reduction model for the chromatophore using available data suggests that the rate of genome reduction has slowed down and genome size has stabilized after diversification of *Paulinella* species. In addition to this result, a high proportion of conserved coding regions is found among species and there is a scarcity of pseudogenes, suggesting that chromatophores in *Paulinella* are in the “stabilizing stage” of endosymbiont genome reduction (see below, Fig. 4).

When comparing genomes of endosymbiont lineages, rearrangement and gene content comparisons should be considered in the context of their divergence time. In the early stages of genome reduction, lost genes can vary conspicuously among all categories and considerable rearrangement may occur among species. When endosymbionts are in the “stabilizing stage” of genome reduction, fewer rearrangements will occur and most genes will be conserved among species. It should be kept in mind that common genes lost among species could have been purged before divergence or independently lost after the species split from each other. Most of the losses in the

chromatophore genome of *Paulinella* were likely to have occurred before the basal split of photosynthetic species. Nevertheless, independent loss after speciation also happened as in the case of oxidoreductase (see below, Supplementary Fig. S5).

Genomic difference among *Paulinella micropora* strains. We searched for recently derived genomic differences between *P. micropora* strains (KR01 isolated from Korea, FK01 isolated from Japan, and NZ27 isolated from New Zealand). There was no difference among these strains with regard to gene content, except for two additional genes being present in NZ27, which have been pseudogenized in KR01 and FK01 (Supplementary Table S4). One of these pseudogenized genes is an oxidoreductase that provides an excellent example of progressive gene loss (Supplementary Fig. S5). This gene is intact in NZ27 and in *P. longichromatophora*, but has been pseudogenized in the other three *Paulinella* lineages and is at different stages of loss: i.e., there are stop codons at position 673 due to a single substitution (CAA to TAA) in both KR01 and FK01. Additionally, there is a large deletion (181 bp) only in KR01, suggesting future complete loss. In *P. chromatophora* CCAC 0185, it was challenging to uncover the history of this gene: i.e., two flanking conserved hypothetical proteins are conserved in all *Paulinella*, but this intergenic space was difficult to align with other oxidoreductases, suggesting complete gene loss. It is interesting that oxidoreductase gene inactivation resulted from two independent events in the ancestor of KR01 and FK01, and *P. chromatophora* CCAC 0185. The ortholog of this oxidoreductase was pyridoxal reductase involved in vitamin B6 metabolism^{63,64}. Furthermore, one-half of the genes involved in this pathway have been shed in the chromatophore, suggesting that this gene is no longer under selective constraints leading to its loss.

Nucleotide substitution and insertion/deletion events were analyzed between *P. micropora* NZ27 and KR01 strains. A total of 22,190 positions (22.7 SNPs/kb) showed single nucleotide polymorphisms (SNPs). These SNPs were more concentrated in noncoding regions (35.9 SNPs/kb) than in coding regions (19.7 SNPs/kb). We also detected a total of 211 insertion or deletion events (indels): 138 single-base indels, 55 indels of 2 to 15 bases, 14 indels of 16 to 42 bases, and 4 large deletions (97 to 136 bases). Compared to SNPs, indels were also more common in noncoding regions (1.05 indels/kb) than in coding regions (0.0263 indels/kb), demonstrating that frameshifts caused by indels could lead to loss of function in proteins. Moreover, the distribution of indels was analyzed among mononucleotide runs (“homopolymers”) (Supplementary Table S5) and found to be more frequent in longer homopolymers.

To determine what types of changes accrued when compared to the ancestral chromatophore, SNPs and indels between KR01 and FK01 were compared to NZ27 and the ancestral state was reconstructed using parsimony (Supplementary Fig. S6). As expected, transitions were more frequent than transversions. However, even among transitions, GC to AT changes were significantly greater than AT to GC (Chi-square test, p -value < 0.001). Using these directional base changes, we calculated the relative per-site mutation bias that quantifies the underlying mutational biases²³. If the genome is under mutational equilibrium, this relative per-site mutational bias (33%) should be equal to the current GC-content (39.9%). However, this was not the case for chromatophore genomes, suggesting that selection on base composition could play a role in the process. This result suggests that AT-enrichment is an ongoing process in chromatophore genomes. This change in GC-content can be explained by the universally biased mutation towards AT in bacteria⁶⁵ and continued relaxed selection on endosymbiont DNA. In the case of indels, insertion events (26) outnumbered deletion events (8), however, the total change was –224 bp because of large deletions, suggesting that genome reduction is also ongoing.

Different mechanisms cause genome reduction in endosymbionts. Insertion of sequences and other mobile DNA can act as molecular mechanisms that contribute to early gene inactivation and deletion in endosymbionts²³. Moran *et al.*⁶⁶ provided a model of stepwise symbiont genome erosion in *Buchnera aphidicola*. The shift toward high AT-content leads to an increased occurrence of A/T homopolymers, and replication slippage occurs, which results in small indels. These indels lead to frameshifts and gene inactivation, and eventually loss of DNA via large deletions. The chromatophore genome also shows a shift toward high AT-content and the frequency of indels was higher in longer homopolymers, which were mostly A or T. However, this was not as pronounced as in the *Buchnera aphidicola* case, implying that other mechanisms might play bigger role in chromatophore genomes. As in the case of oxidoreductase in photosynthetic *Paulinella*, indels and SNPs without homopolymers can also cause inactivation of genes. After inactivation, DNA removal occurs because mutational bias favors deletions over insertions in bacterial genomes⁶⁷.

Conclusion

Our genome reduction model demonstrates that chromatophore genomes in *Paulinella* species are in the “stabilizing stage” of DNA loss (Fig. 4). Even though this process has slowed down, many chromatophore genes remain under relaxed selection when compared to homologs in free-living cyanobacterial relatives. Moreover, chromatophore genomes are much larger than Archaeplastida plastids (see Fig. 3). Will chromatophore genome size eventually reduce to that of plastids? We have no way of predicting so far into the future even though the major “rules” of permanent photosynthetic endosymbiont evolution seem to be conserved among the different lineages. The focus of comparative studies should now turn to the nuclear genome of *Paulinella* species to elucidate their role in organelle formation.

Materials and Methods

Sampling, isolation, DNA extraction, genome amplification. *Paulinella longichromatophora* was collected from Gomso Bay on the western coast of Korea (35° 31′ 11.40″ N, 126° 29′ 0.65″ E) on March 8, 2015. Using a capillary pipet, each cell was individually isolated and washed several times with sterile DY-V medium. For shotgun genome sequencing, DNA from ~3–15 isolated cells were amplified using the REPLI-g Mini Kit (Qiagen, Santa Clarita, California, USA) and MALBAC kit (Yikon Genomics, Taizhou, Jiangsu, China).

Paulinella micropora strain NZ27 was collected from Nga Manu Nature Reserve, Waikanae, Wellington, New Zealand (40°51'39.96"S 175°03'36.63"E) on April 4, 2009. Cells were isolated into culture and maintained in the DY-V medium at room temperature in flat plastic culture flasks. The DNeasy Plant Mini Kit (Qiagen) was used for DNA extraction.

Chromatophore genome sequencing, assembly, and annotation. In the case of *Paulinella longichromatophora*, next-generation sequencing with the Ion Torrent PGM platform (Thermo Fisher Scientific, Santa Clarita, California, USA) was used to generate data. A sequencing library (400 bp size selection) was constructed using the Ion Xpress Plus gDNA Fragment Library Kit, and genome sequencing was conducted with the Ion PGM Template OT2 400 Kit and the Ion PGM Sequencing 400 Kit following manuals provided by Thermo Fisher Scientific. *Paulinella micropora* strain NZ27 was sequenced using a HiSeq2500 platform (Illumina) at a commercial sequencing company (DNA Link Inc., Seoul, Korea). A sequencing library was constructed using the Truseq DNA PCR-Free kit (Insert 550 bp). The chromatophore genome was assembled using CLC Genomics Workbench 5.5.1 (CLC bio, Aarhus, Denmark) and a MIRA assembler that was incorporated in Ion Server (Thermo Fisher Scientific). Genes were identified using the BLASTx algorithm⁶⁸ by comparing the predicted ORFs to the NCBI GenBank database. Gene annotation was conducted using Geneious 6.1.6 (<http://www.geneious.com/>), RNAmmer 1.2 Server⁶⁹, and tRNAscan-SE 1.21⁷⁰. For gap filling, polymerase chain reaction (PCR) amplification was done with an initial denaturation step at 95 °C for 10 min, followed by 35 cycles of 95 °C for 1 min, 50–55 °C for 1 min, and 72 °C for 1 min, concluding with a 10 min extension at 72 °C. Amplified products were purified by QIAquick PCR Purification Kit (Qiagen) and QIAquick Gel Extraction Kit (Qiagen), and sent to Macrogen Inc. (Seoul, Korea) for sequencing.

Phylogenomic analysis. A total of 35 representative genomes from marine, euryhaline, and freshwater *Synechococcus* and *Prochlorococcus* species were included with the five photosynthetic *Paulinella* strains for the phylogenetic analysis based on concatenated amino acids sequence data (Supplementary Table S6). *Gloeobacter violaceus* PCC 7421 was used as the outgroup. The orthologous genes were clustered using OrthoFinder⁷¹ and alignment was done using MAFFT v7.305b⁷². A total of 406 conserved genes were used to reconstruct the phylogeny. The maximum likelihood phylogeny was reconstructed with IQ-TREE version 1.4.4⁷³ using the evolutionary model selected with the IQ-TREE build-in model selection function. For the concatenated tree reconstruction, an independent model for each partition of the concatenated data was used. Support values were estimated using 1,000 replicates of the ultrafast bootstrap implemented in IQ-TREE⁷⁴. For the tree shown in Fig. 2, the same method was used using 20 orthologous genes concatenated amino acids alignment from a total of 40 representative genomes in Archaeplastida plastids, cyanobacteria, and chromatophore genomes of *Paulinella* lineages (Supplementary Table S6).

Genome structure comparison. Synteny comparison was done using the progressive Mauve alignment algorithm in Geneious 6.1.6 (<http://www.geneious.com/>). Hypothetical ancestral states were constructed by mapping chromatophore genes to the *Synechococcus* sp. WH 5701 genome. Despite the presence of massive rearrangements, genes flanking the break points were identified and assumed to represent the ancestral state. Genome rearrangement events were counted by UniMoG: a unifying framework for genomic distance calculation and sorting based on DCJ⁷⁵.

Orthologous Group Families (OGF) analysis. The Orthologous Group Families (OGF) of proteins in cyanobacteria, plastids from Archaeplastida, and chromatophores were clustered using OrthoFinder⁷¹ (Supplementary Table S6). KEGG pathways⁷⁶ were annotated using BlastKOALA⁷⁷ and a heatmap was constructed using a Python script and Microsoft Excel (Microsoft Corporation, Redmond, Washington, USA). Ancestral OGF numbers with gain and loss were estimated using the Dollo parsimony principle⁷⁸.

Estimation of substitution rates at non-synonymous and synonymous sites (dN/dS) and codon adaptation index. For calculating the ratio of substitution rates at non-synonymous and synonymous sites (dN/dS), each ortholog was aligned with Clustal Omega⁷⁹ and then re-aligned with PAL2NAL⁸⁰. Maximum likelihood estimates of dS and dN were computed using codeml from the PAML package (runmode = 2)⁸¹. To avoid the problem of insufficient sequence divergence or saturation⁸², genes with dS values >2 and <0.01 were discarded. The dN/dS value was calculated among three free-living *Synechococcus* sp. (CC9616, RS9916, WH 7805) and averaged dN/dS values were used. Codon Adaptation Index (CAI) values of *Synechococcus* sp. WH 5701 were calculated using CAIcal⁸³. The codon usage table of *Synechococcus* sp. CC9311 was used to calculate the CAI value. For detecting relaxed selection, a general hypothesis testing framework (RELAX)⁵³ on the webserver Datamonkey⁸⁴ was used. Given two subsets of branches in a phylogeny, RELAX can determine whether selective strength was relaxed or intensified in one of these subsets relative to the other.

Molecular clock analysis. We added the 18S rDNA sequences of *P. longichromatophora*, *P. micropora* NZ27, and KR01 to the xml data file provided by Delaye *et al.*⁶. Small subunit rRNA sequences from Rhizaria and Stramenopila from the SILVA database were used. Sequences were aligned with MUSCLE⁸⁵. BEAUTi was used to revise the original xml file. Time trees were inferred by using a lognormal relaxed molecular clock as implemented in BEAST 2⁸⁶. We followed the previous work⁶ priors and calibration scheme: a) the origin of rhizosolenid diatoms, which is known with high confidence (91.5 ± 1.5 Mya); b) a minimal time divergence of pennate diatoms (80 Mya); c) a minimal time divergence for diatoms (133.9 Mya); d) a minimal time divergence of Euglyphidae 40 Mya; and e.1) a time estimation of the divergence of Rhizaria from Stramenopila ~1232 Mya.

Nucleotide substitution and insertion or deletion events. Nucleotide substitution, insertion/deletion events, and homopolymers were analyzed using Geneious 6.1.6 (<http://www.geneious.com/>) and customized Python scripts. To reconstruct the ancestral state of *P. micropora* KR01 and FK01, the chromatophore genomes of *P. micropora* KR01, FK01, and NZ27 strains were aligned, and single nucleotide substitution and insertion/deletion events between KR01 and FK01 were searched in NZ27. Sequence shared with NZ27 was considered as the ancestral state of KR01 and FK01. To minimize some bias in the substitution changes, we calculated the number of directional nucleotide change in each region and divided each value by the frequency of the original nucleotide⁶⁶. After obtaining the relative frequency of each class of nucleotide substitution, we transformed these values to make the total sum of the relative frequencies equal to 100%. Finally, we grouped the changes that could not be differentiated because the mutated and complementary strands could not be distinguished.

Data Availability

Chromatophore genomes of *Paulinella longichromatophora* (MG264610) and *Paulinella micropora* NZ27 (MG976688) are deposited at the NCBI archive. The phylogenetic data supporting this article have been uploaded as part of the Supplementary Material.

References

- Bhattacharya, D., Yoon, H. S. & Hackett, J. D. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* **26**, 50–60, <https://doi.org/10.1002/bies.10376> (2004).
- Reyes-Prieto, A., Weber, A. P. & Bhattacharya, D. The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* **41**, 147–168, <https://doi.org/10.1146/annurev.genet.41.110306.130134> (2007).
- Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* **64**, 583–607, <https://doi.org/10.1146/annurev-arplant-050312-120144> (2013).
- Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809–818, <https://doi.org/10.1093/molbev/msh075> (2004).
- Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. & Knoll, A. H. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl. Acad. Sci. USA* **114**, E7737–E7745, <https://doi.org/10.1073/pnas.1620089114> (2017).
- Delage, L., Valadez-Cano, C. & Perez-Zamorano, B. How really ancient is *Paulinella chromatophora*? *PLoS Curr. Tree. Life*. <https://doi.org/10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c6b> (2016).
- Lauterborn, R. P. II. *Paulinella chromatophora* nov. gen., nov. spec., ein beschalter Rhizopode des Süßwassers mit blaugrünen chromatophorenartigen Einschlüssen. *Z. Wiss. Zool.* **59**, 537–544 (1895).
- Kies, L. Electron microscopical investigations on *Paulinella chromatophora* Lauterborn, a thecamoeba containing blue-green endosymbionts (cyanelles). *Protoplasma* **80**, 69–89, <https://doi.org/10.1007/bf01666352> (1974).
- Adl, S. M. *et al.* The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–493, <https://doi.org/10.1111/j.1550-7408.2012.00644.x> (2012).
- Bhattacharya, D., Helmchen, T. & Melkonian, M. 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, <https://doi.org/10.1111/j.1550-7408.1995.tb01541.x> (1995).
- Marin, B., Nowack, E. C. & Melkonian, M. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* **156**, 425–432, <https://doi.org/10.1016/j.protis.2005.09.001> (2005).
- Yoon, H. S., Reyes-Prieto, A., Melkonian, M. & Bhattacharya, D. Minimal plastid genome evolution in the *Paulinella* endosymbiont. *Curr. Biol.* **16**, R670–R672, <https://doi.org/10.1016/j.cub.2006.08.018> (2006).
- Yoon, H. S. *et al.* A single origin of the photosynthetic organelle in different *Paulinella* lineages. *BMC Evol. Biol.* **9**, 98, <https://doi.org/10.1186/1471-2148-9-98> (2009).
- Lhee, D. *et al.* Diversity of the photosynthetic *Paulinella* species, with the description of *Paulinella micropora* sp. nov. and the chromatophore genome sequence for strain KR01. *Protist* **168**, 155–170, <https://doi.org/10.1016/j.protis.2017.01.003> (2017).
- Kim, S. & Park, M. G. *Paulinella longichromatophora* sp. nov., a new marine photosynthetic testate amoeba containing a chromatophore. *Protist* **167**, 1–12, <https://doi.org/10.1016/j.protis.2015.11.003> (2016).
- Nowack, E. C., Melkonian, M. & Glockner, G. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* **18**, 410–418, <https://doi.org/10.1016/j.cub.2008.02.051> (2008).
- Reyes-Prieto, A. *et al.* Differential gene retention in plastids of common recent origin. *Mol. Biol. Evol.* **27**, 1530–1537, <https://doi.org/10.1093/molbev/msq032> (2010).
- Kies, L. & Kremer, B. P. Function of cyanelles in the thecamoeba *Paulinella chromatophora*. *Naturwissenschaften* **66**, 578–579, <https://doi.org/10.1007/BF00368819> (1979).
- Singer, A. *et al.* Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba *Paulinella chromatophora*. *Curr. Biol.* **27**, 2763–2773 e2765, <https://doi.org/10.1016/j.cub.2017.08.010> (2017).
- Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**, 583–586, [https://doi.org/10.1016/S0092-8674\(02\)00665-7](https://doi.org/10.1016/S0092-8674(02)00665-7) (2002).
- McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26, <https://doi.org/10.1038/nrmicro2670> (2011).
- Moran, N. A. & Bennett, G. M. The tiniest tiny genomes. *Annu. Rev. Microbiol.* **68**, 195–215, <https://doi.org/10.1146/annurev-micro-091213-112901> (2014).
- Wernegreen, J. J. Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann. N. Y. Acad. Sci.* **1360**, 16–35, <https://doi.org/10.1111/nyas.12740> (2015).
- Cole, S. T. *et al.* Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007, <https://doi.org/10.1038/35059006> (2001).
- GÓmez-Valero, L. *et al.* Patterns and rates of nucleotide substitution, insertion and deletion in the endosymbiont of ants *Blochmannia floridanus*. *Mol. Ecol.* **17**, 4382–4392, <https://doi.org/10.1111/j.1365-294X.2008.03912.x> (2008).
- Parkhill, J. *et al.* Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**, 32–40, <https://doi.org/10.1038/ng1227> (2003).
- Burke, G. R. & Moran, N. A. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol. Evol.* **3**, 195–208, <https://doi.org/10.1093/gbe/evr002> (2011).
- Boscaro, V. *et al.* *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **110**, 18590–18595, <https://doi.org/10.1073/pnas.1316687110> (2013).
- Boscaro, V. *et al.* Parallel genome reduction in symbionts descended from closely related free-living bacteria. *Nat. Ecol. Evol.* **1**, 1160–1167, <https://doi.org/10.1038/s41559-017-0237-0> (2017).
- Oakeson, K. F. *et al.* Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol. Evol.* **6**, 76–93, <https://doi.org/10.1093/gbe/evt210> (2014).

31. Hilton, J. A. *et al.* Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. *Nat. Commun.* **4**, 1767, <https://doi.org/10.1038/ncomms2748> (2013).
32. Nakayama, T. *et al.* Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle. *Proc. Natl. Acad. Sci. USA* **111**, 11407–11412, <https://doi.org/10.1073/pnas.1405222111> (2014).
33. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *APS. Nature* **407**, 81, <https://doi.org/10.1038/35024074> (2000).
34. Tamas, I. *et al.* 50 Million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376 (2002).
35. van Ham, R. C. *et al.* Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. USA* **100**, 581–586, <https://doi.org/10.1073/pnas.0235981100> (2003).
36. Pérez-Brocá, V. *et al.* A small microbial genome: the end of a long symbiotic relationship? *Science* **314**, 312 (2006).
37. Gil, R. *et al.* The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc. Natl. Acad. Sci. USA* **100**, 9388–9393, <https://doi.org/10.1073/pnas.1533499100> (2003).
38. Degnan, P. H., Lazarus, A. B., Brock, C. D. & Wernegreen, J. J. Host–symbiont stability and fast evolutionary rates in an ant–bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus* Blochmannia. *Syst. Biol.* **53**, 95–110, <https://doi.org/10.1080/10635150490264842> (2004).
39. Kuwahara, H. *et al.* Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea *Calyptogena* clams. *Extremophiles* **12**, 365–374, <https://doi.org/10.1007/s00792-008-0141-2> (2008).
40. Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J. & Zehr, J. P. Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J.* **8**, 2530–2542, <https://doi.org/10.1038/ismej.2014.167> (2014).
41. Palmer, J. D. Chloroplast DNA exists in two orientations. *Nature* **301**, 92–93 (1983).
42. Marechal, A. & Brisson, N. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* **186**, 299–317, <https://doi.org/10.1111/j.1469-8137.2010.03195.x> (2010).
43. Lee, J. *et al.* Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biol.* **14**, 75, <https://doi.org/10.1186/s12915-016-0299-5> (2016).
44. Price, D. C. *et al.* *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* **335**, 843–847, <https://doi.org/10.1126/science.1213561> (2012).
45. Huang, J. & Yue, J. Horizontal gene transfer in the evolution of photosynthetic eukaryotes. *J. Syst. Evol.* **51**, 13–29, <https://doi.org/10.1111/j.1759-6831.2012.00237.x> (2013).
46. Sperandeo, P., Martorana, A. M. & Polissi, A. Lipopolysaccharide biogenesis and transport at the outer membrane of Gram-negative bacteria. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **1862**, 1451–1460, <https://doi.org/10.1016/j.bbalip.2016.10.006> (2017).
47. Nowack, E. C. M. *et al.* Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proc. Natl. Acad. Sci. USA* **113**, 12214–12219, <https://doi.org/10.1073/pnas.1608016113> (2016).
48. Zehr, J. P., Shilova, I. N., Farnelid, H. M., Munoz-Marin, M. D. & Turk-Kubo, K. A. Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A. *Nat. Microbiol.* **2**, 16214, <https://doi.org/10.1038/nmicrobiol.2016.214> (2016).
49. Gouy, M. & Gautier, C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* **10**, 7055–7074 (1982).
50. Rispe, C., Delmotte, F., van Ham, R. C. H. J. & Moya, A. Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. *Genome Res.* **14**, 44–53, <https://doi.org/10.1101/gr.1358104> (2004).
51. Schaber, J. *et al.* Gene expression levels influence amino acid usage and evolutionary rates in endosymbiotic bacteria. *Gene* **352**, 109–117, <https://doi.org/10.1016/j.gene.2005.04.003> (2005).
52. Delmotte, F., Rispe, C., Schaber, J., Silva, F. J. & Moya, A. Tempo and mode of early gene loss in endosymbiotic bacteria from insects. *BMC Evol. Biol.* **6**, 56, <https://doi.org/10.1186/1471-2148-6-56> (2006).
53. Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832, <https://doi.org/10.1093/molbev/msu400> (2015).
54. Moran, N. A., McCutcheon, J. P. & Nakabachi, A. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190, <https://doi.org/10.1146/annurev.genet.41.110306.130119> (2008).
55. Moran, N. A. & Mira, A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* **2**, research0054.0051–research0054.0012, <https://doi.org/10.1186/gb-2001-2-12-research0054> (2001).
56. Moran, N. A. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **93**, 2873–2878 (1996).
57. Muller, H. J. Some genetic aspects of sex. *Amer. Nat.* **66**, 118–138, <https://doi.org/10.1086/280418> (1932).
58. Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res.* **106**, 2–9, [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8) (1964).
59. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
60. Rispe, C. & Moran, N. A. Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat.* **156**, 425–441, <https://doi.org/10.1086/303396> (2000).
61. Pettersson, M. E. & Berg, O. G. Muller's ratchet in symbiont populations. *Genetica* **130**, 199, <https://doi.org/10.1007/s10709-006-9007-7> (2006).
62. Nomura, M., Nakayama, T. & Ishida, K.-i Detailed process of shell construction in the photosynthetic testate amoeba *Paulinella chromatophora* (Euglyphid, Rhizaria). *J. Eukaryot. Microbiol.* **61**, 317–321, <https://doi.org/10.1111/jeu.12102> (2014).
63. Nakano, M. *et al.* Purification, molecular cloning, and catalytic activity of *Schizosaccharomyces pombe* pyridoxal reductase: A possible additional family in the aldo-keto reductase superfamily. *J. Biol. Chem.* **274**, 23185–23190, <https://doi.org/10.1074/jbc.274.33.23185> (1999).
64. Herrero, S., González, E., Gillikin, J. W., Véléz, H. & Daub, M. E. Identification and characterization of a pyridoxal reductase involved in the vitamin B6 salvage pathway in *Arabidopsis*. *Plant Mol. Biol.* **76**, 157–169, <https://doi.org/10.1007/s11103-011-9777-x> (2011).
65. Hershberg, R. & Petrov, D. A. Evidence that mutation is universally biased towards AT in bacteria. *PLOS Genetics* **6**, e1001115, <https://doi.org/10.1371/journal.pgen.1001115> (2010).
66. Moran, N. A., McLaughlin, H. J. & Sorek, R. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* **323**, 379, <https://doi.org/10.1126/science.1167140> (2009).
67. Mira, A., Ochman, H. & Moran, N. A. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596, [https://doi.org/10.1016/S0168-9525\(01\)02447-7](https://doi.org/10.1016/S0168-9525(01)02447-7) (2001).
68. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421–421, <https://doi.org/10.1186/1471-2105-10-421> (2009).
69. Lagesen, K. *et al.* RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108, <https://doi.org/10.1093/nar/gkm160> (2007).
70. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
71. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157, <https://doi.org/10.1186/s13059-015-0721-2> (2015).
72. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780, <https://doi.org/10.1093/molbev/mst010> (2013).
73. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274, <https://doi.org/10.1093/molbev/msu300> (2015).

74. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195, <https://doi.org/10.1093/molbev/mst024> (2013).
75. Hilker, R., Sickinger, C., Pedersen, C. N. S. & Stoye, J. UniMoG—a unifying framework for genomic distance calculation and sorting based on DCJ. *Bioinformatics* **28**, 2509–2511, <https://doi.org/10.1093/bioinformatics/bts440> (2012).
76. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30, <https://doi.org/10.1093/nar/28.1.27> (2000).
77. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731, <https://doi.org/10.1016/j.jmb.2015.11.006> (2016).
78. Csűös, M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910–1912, <https://doi.org/10.1093/bioinformatics/btq315> (2010).
79. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539, <https://doi.org/10.1038/msb.2011.75> (2011).
80. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–612, <https://doi.org/10.1093/nar/gkl315> (2006).
81. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591, <https://doi.org/10.1093/molbev/msm088> (2007).
82. Dos Reis, M. & Yang, Z. Why do more divergent sequences produce smaller nonsynonymous/synonymous rate ratios in pairwise sequence comparisons? *Genetics* **195**, 195–204, <https://doi.org/10.1534/genetics.113.152025> (2013).
83. Puigbò, P., Bravo, I. G. & García-Vallve, S. CAlcal: A combined set of tools to assess codon usage adaptation. *Biol. Direct* **3**, 38, <https://doi.org/10.1186/1745-6150-3-38> (2008).
84. Weaver, S. *et al.* Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **35**, 773–777, <https://doi.org/10.1093/molbev/msx335> (2018).
85. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113–113, <https://doi.org/10.1186/1471-2105-5-113> (2004).
86. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537, <https://doi.org/10.1371/journal.pcbi.1003537> (2014).

Acknowledgements

This work was supported by the grants from the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (MOF) (20180430), the National Research Foundation of Korea (NRF-2017R1A2B3001923 and NRF-2016R1A6A1A03012647), and the Next-generation BioGreen21 Program (PJ01389003) from the Rural Development Administration, Korea. We thank Giuseppe Zuccarello for helping in *Paulinella* sampling in New Zealand, Louis Graf for *P. longichromatophora*, and JunMo Lee for the bioinformatics analysis. We also thanks to Jong Im Kim for providing image of *Paulinella micropora* KR01 in Fig. 4c.

Author Contributions

D.L. collected *P. longichromatophora* samples from field and S.K. and M.G.P. provided the isolated chromatophores of *P. longichromatophora*. D.L. and J.H. assembled and annotated chromatophore genomes. D.L. wrote the draft of the manuscript. H.S.Y. and D.B. contributed to designing the work, interpreted the results, and edited the manuscript. H.S.Y. supervised the project. All authors discussed the results and commented on the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-38621-8>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019