Potential causes and consequences of rapid mitochondrial genome evolution in thermoacidophilic Galdieria (Rhodophyta)

Chung Hyun Cho
Sungkyunkwan University - Suwon Campus
Seung In Park
Sungkyunkwan University - Suwon Campus
Claudia Ciniglia
Universita degli Studi della Campania Luigi Vanvitelli
Eun Chan Yang
Korea Institute of Ocean Science and Technology
Louis Graf
Sungkyunkwan University
Debashish Bhattacharya
Rutgers The State University of New Jersey
Hwan Su Yoon (hsyoon2011@skku.edu)
Sungkyunkwan University https://orcid.org/0000-0001-9507-0105

Research article

Keywords: Cyanidiophyceae, extremophile, mitogenome evolution, protein divergence, mitochondrial DNA replication

Posted Date: June 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36820/v1

Version of Record: A version of this preprint was published on September 7th, 2020. See the published version at https://doi.org/10.1186/s12862-020-01677-6.

Abstract

The Cyanidiophyceae is an early-diverged red algal class that thrives in extreme conditions around acidic hot springs. Although this lineage has been highlighted as a model for understanding the biology of extremophilic eukaryotes, little is known about the evolutionary history of their mitochondrial genomes. To fill this knowledge gap, we sequenced five mitogenomes from representative clades of Cyanidiophyceae and identified two major groups, here referred to as *Galdieria*-type (*G*-type) and *Cyanidium*-type (*C*-type). *G*-type mitogenomes exhibit the following three features: (i) reduction in genome size and gene inventory, (ii) evolution of unique protein properties including charge, hydropathy, stability, amino acid composition, and protein size, and (iii) distinctive GC-content and skewness of nucleotides. Based on GC-skew-associated characteristics, we postulate that unidirectional DNA replication may have resulted in the rapid evolution of *G*-type mitogenomes. This high divergence was likely driven by natural selection in the multiple extreme environments *Galdieri*a species inhabit, their highly flexible heterotrophic metabolism, and the impacts of population size reduction. We speculate that the interplay between mitogenome divergence and adaptation may help explain the dominance of *Galdieri*a species in diverse extreme habitats.

Background

Life in hot springs environments can place severe stresses on cells due to the high temperature, acidic, and heavy-metal rich conditions [1]. Most biodiversity in such extreme environments is comprised of prokaryotes, with only a few eukaryotes able to compete successfully in these habitats [2, 3]. The unicellular red algal class, Cyanidiophyceae, is a well-known group of extremophilic eukaryotes that thrives in acidic (pH 0.5-3.0) and high temperature (50–55 °C) habitats [4, 5]. Cyanidiophyceae have traditionally been reported from volcanic regions around the world [6–8], whereas some mesophilic species (e.g., *Cyanidium chilense*) are found in moderately acidic caves (pH 5–7) around volcanic regions in Chile, Italy, France, Israel, and Turkey [4, 9–12]. Cyanidiophyceae is the earliest diverging red algae lineage, having split from other Rhodophyta about 1.5 billion years ago [5, 13]. Due to limited ultrastructural differences and the “simple” morphology of cyanidiophycean cells, it is difficult to classify them using microscopy. This has led to the hypothesis of cryptic diversity based on molecular phylogenetic studies [4, 14, 15]. Moreover, incongruent topologies or unresolved relationships of Cyanidiophyceae are reported in most molecular phylogenetic studies using single (e.g., *rbc*L, SSU rRNA) or a few genes (e.g., *rbc*L + *psa*A + *psb*A) [4, 10, 14, 16–18].

Most Cyanidiophyceae are photoautotrophs and inhabit niches exposed to sulfur fumes, but the genus *Galdieria* also occupies endolithic and interlithic habitats [4, 16] and is therefore more exposed to microenvironmental fluctuations [4, 19, 20]. Members of this genus are very effective at mixotrophic growth, able to utilize > 50 carbon sources [21]. From a genomic point of view, previous studies have demonstrated that *Galdieria* species contain a variety of horizontally transferred genes that confer adaptation for specific habitats (e.g., endolithic) [22–24]. Although genomes are significantly reduced in size (< 20 Mbp) in Cyanidiophyceae, *Galdieria* species retain a spliceosomal machinery and have a relatively large number of introns that likely play a role in regulating the stress response [20].

In addition, a previous study reported that a *Galdieria* species has a highly reduced mitochondrial genome and these genes have the fastest substitution among all red algae due to the polyextremophilic lifestyle [25]. However, only two published (*Cyanidioschyzon merolae* 10D and *Galdieria sulphuraria* 074W) and one unpublished (Cyanidiophyceae sp. MX-AZ01) mitochondrial genomes are available to date [25, 26]. To better understand cyanidiophycean mitochondrial genome evolution, we generated five complete mitogenomes that represent all major clades of Cyanidiophyceae. Among them, three *Galdieria* mitogenomes are not only highly reduced in size, but also substantially differ in protein characteristics compared to Cyanidiophyceae and other red algae. Based on GC-skewness and other mitogenomic features, we postulate that a unique replication system may exist in *Galdieria* species.

Results And Discussion

To confirm the phylogenetic position of the eight strains under study, we used the *rbc*L phylogeny to identify the major groups of Cyanidiophyceae (Figure S2). The taxon-rich (269 taxa) *rbc*L tree showed five major cyanidiophycean groups: 1) *Cyanidium chilense* assemblage (*Cd. chilense*; known as mesophilic *Cyanidium* sp.) [12], 2) *Cyanidium caldarium* (*Cd. caldarium*), 3) *Galdieria sulphuraria* assemblage (*G. sulphuraria*), 4) *Cyanidioschyzon merolae* (*Cz. merolae*), and 5) *Cyanidiococcus*

yangmingshanensis assemblage (*Cc. yangmingshanensis*; known as *Galdieria maxima*) [27], which is consistent with previous studies [14, 18]. Our eight strains represent well the diversity of the five major groups of Cyanidiophyceae (see in Figure S2).

The general characteristics of the eight mitogenomes were compared including five new and three published datasets (Table 1) [25, 26]. These mitogenomes are clearly divided into two types as *Cyanidium*-type and *Galdieria*-type based on mitogenome features (e.g., genome size, genome structure, number of genes, skewness of nucleotides; see Supplementary Information 1). *Cyanidium*-type (*C*-type) is comprised of five taxa (mesophilic *Cd. chilense* Sybil Cave, *Cd. caldarium, Cz. merolae*, *Cc. yangmingshanensis*, and Cyanidiophyceae sp. MX-AZ01), and the *Galdieria*-type (*G*-type) is comprised of three taxa (*G. phlegrea*, *G. sulphuraria* 108.79 E11 and 074W). Based on our observations and previous work, *C*-type and *G*-type are recognized not only using mitogenome characteristics, but also on the basis of morphological characteristics, cellular features, and ecological habitats. Cells of *G*-type species are generally larger than *C*-type and have a simpler morphology, compared to the more diverse cell shapes (e.g., spherical, oval, club-shaped) in *C*-type species (Fig. 1A). According to a previous analysis, *G*-type *Galdieria sulphuraria* contains several mitochondria per cell that have a net-like structure, whereas *C*-type *Cd. caldarium* contains a single mitochondrion in a spheroid cell [28]. Likewise, multiple mitochondria are identified in *G*-type *G. sulphuraria* 108.79 E11, whereas a single spheroid mitochondrion is found in *C*-type *Cc. yangmingshanensis* 8.1.23 F7 in transmission electron microscopic observation (Figs. 1A).

Table 1
General characteristics of Cyanidiophyceae mitogenomes. *abbreviations: CZME (*Cyanidioschyzon merolae*), CYSP (Cyanidiophyceae sp.), CCYA (*Cyanidiococcus yangmingshanensis*), CDCA (*Cyanidium caldarium*), CDCH (*Cyanidium chilensis*), GAPH (*Galdieria phlegrea*), GASU (*Galdieria sulphuraria*).

Type	***Cyanidium*-type (*C*-type)**					***Galdieria*-type (*G*-type)**		
Species*	**CZME 10D**	**CYSP MX-AZ01**	**CCYA 8.1.23 F7**	**CDCA ACUF 019**	**CDCH Sybil Cave**	**GAPH DBV 009**	**GASU 074W**	**GASU 108.79 E11**
Genome Size (bp)	32,211	32,620	32,387	34,207	33,039	21,792	21,428	21,611
GC-content (%)	27.1	26.7	26.4	25.9	44.5	41.4	44.0	41.8
GC-skew	0.06	0.03	0.03	0.02	0.01	0.71	0.74	0.66
AT-skew	0.01	0.02	0.03	0.03	0.03	0.25	0.25	0.29
Number of Genes	64	61	62	61	61	26	27	27
Non-coding Region (%)	5.20	6.50	5.13	10.64	6.37	17.55	15.55	16.49
NCBI Accession Number	NC_000887	KJ569774	MT270119 (this study)	MT270118 (this study)	MT270117 (this study)	MT270116 (this study)	NC_024666	MT270115 (this study)

Cyanidium **-type and** Galdieria**-type mitogenomes resolved using phylogenetic analysis**

Phylogenetic analysis also supports the recognition of the *C*-type and *G*-type mitogenomes. The concatenated protein ML phylogeny resolves the *C*-type and *G*-type with full bootstrap support (see Fig. 1B). Within the *C*-type, the mesophilic *Cd. chilense* Sybil Cave diverged first, followed by *Cd. caldarium* ACUF 019. *Cz. merolae* 10D are grouped together with the monophyletic clade of the *Cc. yangmingshanensis* 8.1.23 F7 + Cyanidiophyceae sp. MX-AZ01. Our current mitogenome data resolve the internal relationships within the Cyanidiophyceae, in particular, the positions of mesophilic *Cd. chilense* Sybil Cave and *Cd. caldarium* ACUF 019, that were poorly resolved until now. In this mitogenome data analysis, however, we observed an extraordinarily long internal branch of the *G*-type (Fig. 1B), which implies high divergence when compared to *C*-type species or unidentified/extinct

genetic diversity in the *G*-type lineages. In addition, we tested individual gene phylogenies to see the consistency compared to concatenated gene tree in Supplementary Information 2. Applying variable mitochondrial gene datasets, we were able to resolve phylogenetic relationships among major cyanidiophycean clades.

Different CDS content between Cyanidium**-type and** Galdieria**-type mitogenomes: gene loss and transfer**

After the recognition of two different groups in Cyanidiophyceae, we focused on mitogenomes gene gains and losses. Comparison of CDS content between the two different types revealed that one-half of mitochondrion-encoded genes in *C*-type mitogenomes are missing in the *G*-type, in particular synteny of the green block in *G*-type was changed (Fig. 2A) [i.e., losses of all ribosomal protein genes (*rps*, *rpl*) and a few core genes (*ccm*A,B and *sdh*B,D)]. To examine endosymbiotic gene transfer (EGT) from mitochondrial to the nuclear genome for these missing genes in *G*-type, we searched 18 homologous genes in the two available nuclear genomes of *Cz. merolae* 10D and *G. sulphuraria* 074W [23, 29]. Ten genes were identified from the nuclear genome of *G. sulphuraria* 074W, but owe did not identify eight mitochondrial genes ('N/D' in Fig. 2A) that may indicate outright gene losses in the mitogenome, although it is not possible to rule out issues related to low-quality genome data. Other explanations for missing genes is either high diversification after gene transfer or degeneration of genes from mitochondria. Out of twelve ribosomal proteins, only *rps*12 and *rpl*20 were found in the *G. sulphuraria* 074W nuclear genome. These nuclear-encoded *rps*12 and *rpl*20 genes were not grouped together but rather located in two different scaffolds, 57 and 29, respectively (Fig. 2B), whereas most ribosomal protein encoding genes are located in a single syntenic block in *C*-type mitogenomes (*ccm*A-*nad*6; green block in Fig. 2A). We could not detect the remaining ribosomal protein encoding genes, instead, we found six nuclear-encoded homologs of ribosomal protein (*rpl*6, *rpl*14, *rpl*16, *rps*4, *rps*11, *rps*19) in the genome (see question marks in Fig. 2A). With the *ccm*F gene (see Supplementary Information 3 for details), the origins of six homologous genes of ribosomal protein were unclear based on phylogenetic analyses due to low bootstrap support values.

It is unlikely that mitochondrial translation would function properly without a complete set of ribosomal subunit proteins, therefore, the nuclear-encoded homologs could "compensate" for gene losses (e.g., 16–19 tRNAs loss in *G*-type mitogenome). Meanwhile, the homologs of *rps*13 and *rps*19 were not detected from the *Cz. merolae* 10D but were found in the early branched, mesophilic *Cd. chilense*, suggesting independent gene losses in both *C*-type and *G*-type mitogenomes. It was possible to detect a plastidial-copy, nuclear-copy (host-derived), or other (e.g., unknown sources) ribosomal proteins from homologous searches, which implies the possibility to translocate ribosomal subunits from various origins (e.g., nuclear, plastid, other bacteria) into mitochondria as suggested in previous studies [30–32] .

Changes in amino acid composition in Galdieria**-type mitochondrial genes**

Enzymes from extremophilic organisms have high thermostability, more charged amino acid composition, and reduced hydrophobic surfaces to withstand extreme temperatures and pH [33, 34]. Furthermore, proteins of thermophilic species are shorter in length than those in their mesophilic counterparts [35]. To investigate protein characteristics, we compared protein charge, hydropathy, and stability in 16 genes (genes in Fig. 3C) that are retained in all 12 mitogenomes (i.e., *G*-type, *C*-type, non-cyanidiophycean red algae).

Amino acid composition of these 16 conserved mitochondrial genes shows that 13 out of 20 amino acids are significantly different (asterisks in Fig. 3A; Table S7) between *G*-type and *C*-type and this resulted in a difference in the structure of amino acids that effectively modified the protein properties of the genes. Whereas there is a similar amino acid composition between the non-cyanidiophycean red algae (outgroup) and C-type, G-type mitogenomes have a distinct amino acid composition (Fig. 3A). *G*-type genomes show a higher proportion of positively charged amino acids than those of *C*-type (*G*-type: 10.77–10.82%, *C*-type: 6.84–7.44%; see in Figure S6). A lower negative charge amino acid composition was found in *G*-type genomes when compared to *C*-type (*G*-type: 2.26–2.47%, *C*-type: 4.06–4.21%; see in Figure S6). Likewise, the influence of amino acid changes altered protein charge, hydrophilicity, and hydrophobicity (Fig. 3B,C).

Hydrophilicity, which was measured for 16 conserved mitochondrial proteins (Fig. 3B), representing the *G*-type species showed relatively higher hydrophilic amino acids in mitochondrial proteins (50.56–50.97%) than those in other red algal species (*C*-type and outgroup species: 41.64–43.49%). Because some genes showed dramatic difference in amino acid composition or in protein size, we examined individual gene hydropathy (a scale of hydrophobicity and hydrophilicity) to avoid a biased assessment [36,

37]. *G*-type proteins are clearly less hydropathic than other groups (Fig. 3C) and 11/16 mitochondrial genes in *G*-type tend to have reduced protein length (Figure S7A) when compared to other species. These dramatic differences in amino acids (e.g., charge, length) of mitochondrial genes are critical to protein structure that can affect solubility, stability, and their functions [38]. We applied *in silico* analysis (e.g., instability index, aliphatic index) to calculate the stability of conserved mitochondrial proteins and also found *G*-type and *C*-type have a few significant differences (Figure S7B, C) in their mitochondrial proteins. Conserved mitochondrial genes in Cyanidiophyceae are mostly membrane-bound proteins or in mitochondria they form protein complexes (e.g., mitochondrial respiratory complexes). In these cases, protein folding, which is a key factor to understand protein stability and their activity, can be highly dependent on lipid composition of the mitochondrial membrane or a protein-protein interaction with other supermatrix-forming proteins [39, 40]. Therefore, these protein interactions with mitochondrial membrane lipids need further studies.

Extreme GC-skew in Galdieria**-type mitogenomes and its associated characteristics**

On the basis of mitogenomes comparison, *G*-type mitogenomes have distinctive characteristics, such as high gene divergence and asymmetric nucleotide substitution. The difference in GC-contents between the *C*-type and *G*-type is pervasive across genomes (Figure S8) showing that *C*-type species have lower GC-contents (25.0-27.1%) than that of *G*-type species (41.4–44.0%) excluding mesophilic *Cd. chilense* (44.5%). However, GC-skew (G-C/G + C) and AT-skew (A-T/A + T) are clearly different in *C*-type and *G*-type (see Table 1): *C*-type composed symmetric AT and GC composition balances (AT-skew: 0.01–0.03; GC-skew: 0.01–0.06). In contrast, *G*-type showed unbalanced AT composition (AT-skew: 0.25–0.29) and extremely asymmetric composition of GC nucleotides (GC-skew: 0.66–0.74). While in a member *C*-type species, mesophilic *Cd. chilense*, the GC-content (44.5%) is close to those of *G*-type, but GC-skew (0.01) or AT-skew (0.03) of mesophilic *Cd. chilense* are more similar to other *C*-type mitogenomes. In other words, mesophilic *Cd. chilense* can be regarded as an intermediate state between *C*-type and *G*-type based on its genomic features and phylogenetic position.

All genes, including 17 CDSs, seven tRNAs, and two rRNAs, are located in a single strand of *G*-type *G. sulphuraria* 074W mitogenome except for the anticlockwise *cob* gene, whereas genes in *C*-type *Cz. merolae* 10D mitogenome are distributed in both strands as usual (Fig. 4A). According to the directional distribution of genes and extreme GC-skew, genes in *G*-type mitogenome appear to be substantially strand-biased. The *cob* gene, which is located in an antisense orientation of *G*-type species, has lower GC-skew than average GC-skew of *G*-type mitogenomes (*cob* gene region GC-skew: 0.43–0.48, mitogenome GC-skew: 0.66–0.74; see Table 1, Figure S9) and shows a higher TIGER value compared to other genes meaning that *cob* gene contains lower variable sites (TIGER value of *cob* gene: 0.766, average TIGER value: 0.630; see Figure S4). Based on these observations, we examined the potential impact of extreme GC-skew on *G*-type mitogenomes.

One of the key indicators to distinguish leading and lagging strands is the GC-skew [41]. In most cases, positive GC-skew reflects the leading strand, whereas negative GC-skew represents the lagging strand [42, 43]. GC-skew analysis of the *G*-type shows all positive values and its cumulative GC-skew is gradually increased without any decreasing points unlike other red algal species including *C*-type (Fig. 4A; Figure S9). Mitogenomes with a positive GC-skew in a single strand have been well studied for their replication system, particularly in human mitochondria. Although their precise replication mechanisms are still under investigation, it is accepted that they have a unique asymmetric replication process that contains one unidirectional leading strand and the other unidirectional lagging strand assisted by RNAs without any Okazaki fragments [44–46]. *G*-type species likely have an asymmetric replication mechanism: a guanine-rich leading strand (H-strand) and a cytosine-rich lagging strand (L-strand). After the separation of L-strand from H-strand, without a bidirectional replication fork, a daughter lagging strand is synthesized by a nascent leading strand and a nascent lagging strand synthesizes a daughter leading strand (Fig. 4B). The synthesis of lagging strand is considered to be more accurate than the leading strand due to the short Okazaki fragments [47], which have been exemplified in some mitochondria (e.g., yeast, fish, and bacterial species) with experimental verifications [48–50]. In addition, different gene substitution rates have been reported between two DNA strands of mitochondrial genome (e.g., higher in lagging strand of fish) [51]. Taken together, *G*-type species have higher rates of mutation that is likely to be accelerated by unidirectional replication (Fig. 4B).

On the basis of these findings, we propose that unidirectional replication in *G*-type mitochondria led to higher divergence than in the *C*-type and other non-cyanidiophycean red algae, which have bidirectional mitogenome replication. Because of mitochondrial

replication system divergence, *G*-type mitogenomes may have different uses of the DNA polymerases described in Jain et al. (2015), resulting in sequence variation. Such accelerated mutation rates of the two DNA strands may potentially contribute to *G*-type mitogenomes having a higher fitness in rapidly changing environments and play a role in adaptation [52]. We speculate that extreme GC-skew in the *G*-type resulted in the use of unidirectional replication that led to changes in protein properties, affecting the fitness of mitochondrial proteins in harsh environments.

Conclusions

The rapid evolution we report here in *Galdieria*-type mitogenomes (i.e., long branches of *G*-type in the molecular phylogenies) may be explained by the fact that Cyanidiophyceae species inhabit extreme conditions unlike other mesophilic red algae. Among Cyanidiophyceae, *Cyanidium*-type (*C*-type) species inhabit ecologically more protected niches (e.g., aquatic habitats such as hot springs and ditches) within extreme environments than do *G*-type species. Therefore, they are likely to be less prone to environmental pressures such as temperature or pH fluctuation than *Galdieria* species [4, 10]. Reactive oxygen species in the internal environment of mitochondria, mitochondria replication, and absence or degeneration of mitochondrial DNA repair systems can elevate mitochondrial mutation [53]. In cyanidiophycean mitochondria, high temperature and an acidic environment are potential oxidative stress inducers, which is detrimental to cellular components like DNA or their repair system. Along with these common factors, another mutagenic factor could be the unidirectional replication mode of *G*-type mitogenomes. Strand-based compositional asymmetry of the two separate replicated strands can be introduced through replication-associated mutational stress [54]. However, in comparison to other mitogenomes that also have unidirectional replication, *G*-type species have been exposed to extreme habitats and diverged more than 800 Mya [13]. Since then, unidirectional mitochondrial DNA replication system in *G*-type species has facilitated mutagenesis in stressful habitats over a long evolutionary time span, resulting in exceptionally divergent *G*-type mitochondria that is not observed in unidirectional replication systems of other red algae or even in nuclear and plastid of *G*-type species. Specifically, we assume that unidirectional replication was established in the ancestral population of *G*-type mitogenomes. Unlike the nuclear genome, mitochondria are subject to rapid genetic drift due to predominantly uniparental inheritance and lack of sexual recombination [55]. The reduced effective population size of these organelles, that weakens the power of natural selection, may have allowed the initial survival of mutations that proved adaptive in the longer run, such as those that confer thermostability of DNA and RNA (Supplementary Information 4).

However, because of their long evolutionary history, it is difficult, if not impossible, to assign a cause-and-effect relationship between a large mutagenesis event and unique trait evolution (e.g., protein properties). Interestingly, EGT-derived genes that were moved to the nucleus in the early stages of eukaryote evolution accumulated fewer mutations than mitochondrial-encoded genes (Supplementary Information 5). In addition, mutations of *G*-type mitochondrial genes are found to be under purifying selection, similar to mammalian mitochondrial genes [56], based on analysis of the nonsynonymous and synonymous mutation (K_a/K_s) ratio (Figure S11). This suggests that *G*-type mitogenomes acquired mutations faster than other *C*-type mitogenomes or non-cyanidiophycean red algal mitogenomes during the same time period. The reason why these changes are restricted to the *G*-type mitochondrial genome and not evident in plastid or nuclear DNA remains unclear. In summary, we speculate that a combination of factors (e.g., unidirectional replication system, polyextreme habitats, heterotrophic metabolism, reduced effective population size of mitochondria) have driven *G*-type mitogenome evolution. These results lay the foundation for future studies to better understand the biology of the intriguing Cyanidiophyceae.

Methods

Sample preparation

Cyanidium caldarium samples were obtained from the predecessor strain (*Cd. caldarium* ACUF 019; Siena, Italy) established by Claudia Ciniglia. Due to strain contamination issues in some isolated strains, single cells from two established strains of *Cyanidiococcus yangmingshanensis* (*Cc. yangmingshanensis* 8.1.23; Kula Manisa, Turkey) and *Galdieria sulphuraria* (*G. sulphuraria* SAG 108.79; Yellow Stone National Park, USA) were isolated using fluorescence-activated cell sorting (FACS) to establish single cell derived culture strains. The cells were spread on a sucrose-agar medium after initial cultivation in a liquid medium (5x Allen medium), and an individual colony from each species (SAG 108.79: isolate E11, 8.1.23: isolate F7) was

transferred to liquid media to promote growth. Blue-green biofilms were obtained from the shaded side of the Sybil Cave tuff wall (Cume, Italy) to collect mesophilic *Cyanidium chilense* (*Cd. chilense*). Collected samples were mixed with 15 mL phosphate-buffered saline (PBS) and centrifuged briefly at 550 rpm, repeatedly, to harvest cells from environmental samples.

Transmission electron microscopy

Samples of *G. sulphuraria* 108.79 E11 and *Cc. yangmingshanensis* 8.1.23 F7 were initially fixed in a solution of 1.5% glutaraldehyde + 8% sucrose in 0.1M phosphate buffer (pH 7.4) and incubated overnight at 4 °C. Fixed samples were rinsed for 10 min and centrifuged at 12,000 rpm for 5 min. After repeating the rinse three times, 1% osmium tetroxide in 0.1M phosphate buffer was treated to the samples for 1.5 h. We repeated this step three times by rinsing the samples for 10 min and centrifuging at 12,000 rpm for 5 min. Pellets of fixed cyanidiophycean cells were dislodged and solidified in 1% agarose to divide them into the appropriate size for the section. The dehydration step in ethanol was done after the sectioning step and the pellet segments were embedded in Spurr's resin. Samples were cut into 70 nm thickness and stained with uranyl acetate and lead citrate. The samples were observed using a Bio-HVEM System (JEM-1400 Plus at 100 kV and JEM-1000BEF at 1,000 kV [JEOL, Japan]) at the Korea Basic Science Institute (Ochang, Korea).

DNA extraction and rbcL sequencing

The genomic DNA of three culture strains (*Cd. caldarium* ACUF 019, *Cc. yangmingshanensis* 8.1.23 F7, *G. sulphuraria* 108.79 E11) was extracted by DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and purified using the LaboPass™ DNA Isolation Kit (Cosmo Genetech, Seoul, Korea). Genomic DNA of environmental samples from Sybil Cave was extracted using the FastDNA® SPIN Kit for Soil (MP Bio, Santa Ana, USA). To identify species, *rbc*L gene was amplified using DNA KOD FX Neo (Toyobo, Osaka, Japan) polymerase from each strain. A different combination of *rbc*L primer sets were used for PCR; *rbc*L_rc_214F: 5'-GTTGTWTGGACWGATTTATTAAC-3' (23 mers), *rbc*L_rc_1234R: 5'-GCTTGWATWCCATCTGGATC-3' (20 mers), *rbc*L_90F: 5'-CCATATGCYAAAATGGGATATTGG-3' (24 mers), *rbc*L_R: 5'-ACATTTGCTGTTGGAGTCTC-3' (20 mers) [57]. Amplified DNAs were purified by LaboPass™ PCR Purification Kit (COSMO Genetech, Seoul, Korea) and products were sent to a sequencing company (Macrogen, Seoul, Korea). We manually removed low quality sequences and merged forward and reverse strand sequences into a single aligned sequence.

Whole genome sequencing and mitogenome analysis

The raw data of *Galdieria phlegrea* DBV 009 (Naples, Italy), sequenced in a previous study [22], was used here to complete the mitogenome. Other whole genome data were generated using the Illumina HiSeq 2500 (Illumina, San Diego, USA) with 2 × 100 bp paired-end sequencing at the DNA-Link (Seoul, Korea). Reads were assembled into contigs using SPAdes assembler 3.10.1 [58] and mitogenomes were constructed following the established procedure described in Cho et al., 2018 [59] using two published mitogenomes (NC_000887, NC_024666) as reference. Using the mitochondrial proteins of *Cz. merolae* 10D [26], assembled contigs that include mitochondrial genes were selected by tBLASTn. Next, the sorted contigs were re-assembled to get a scaffold sequence of mitogenomes using Geneious Prime 2019.0.3 (Biomatters, Auckland, New Zealand). After the re-assembly step, a complete circular mitogenome or partial linear mitogenome contigs were obtained from each species. The completeness of these mitogenome sequences were tested by LASTZ [60] and by comparison with two published mitogenomes (NC_000887: *Cz. merolae* 10D, NC_024666: *G. sulphuraria* 074W). The genome sequences were mapped with raw reads to correct sequencing errors by mapping coverage or filling gaps for a linear mitogenome.

Protein coding genes (CDS) were manually annotated based on a homologous region of existing mitogenomes (NC_000887, NC_024666, KJ569774). The "standard genetic code 1" has been generally used in cyanidiophycean mitogenomes [25, 26] without a clear evidence, although there is an experimental report of genetic code in Gigartinales among the red algae that has been using use the "protozoan mitochondrial genetic code 4" [61]. Therefore, both genetic codes, "standard genetic code 1" and "standard genetic code 1", were tested in our study and verified the "standard genetic code 1" were more acceptable without having CDS extension (Figure S1). The ribosomal RNA sequences were acquired by BLASTn searching against other mitochondrial rDNA sequences of Cyanidiophyceae. tRNAs, and other small RNAs were searched by using tRNAscan-SE 2.0 [62] and ARAGORN [63]. For comparative analysis using mitochondrial genes, several genes and ambiguous sequences from the published mitogenomes were re-annotated (Table S1) in this study [64]. GC-content was calculated with a sliding window scale of 48 bp, and the GC-skew

of each species was calculated (window size: 1 kbp, step: 20 bp) using a Python script. Potential G-quadruplex forming sequences were predicted by G4Hunter using a default option [65].

Phylogenetic analyses

The *rbc*L gene was chosen to identify the phylogenetic position of the newly sequenced strains. Cyanidiophyceae *rbc*L sequences were collected from the NCBI nucleotide database (ntDB) and aligned with *rbc*L sequences of mitogenome-sequenced strains in this study. A total of 269 *rbc*L sequences were used for phylogenetic analysis including 10 other red algal species (i.e., Rhodophytina) as outgroups. For phylogenetic analysis of mitochondrial genes, eight Cyanidiophyceae taxa were selected, along with four other red algae as an outgroup. The outgroup taxa included two species from Florideophyceae (*Chondrus crispus* and *Hildenbrandia rubra*), one species from Bangiophyceae (*Porphyra purpurea*), and one species from Compsopogonophyceae (*Compsopogon caeruleus*). Gene alignments were produced using MAFFT v7 [66] with the default options. Maximum likelihood (ML) phylogenetic analysis with automated model selection was done using IQ-TREE [67] with 1,000 ultrafast bootstrap replications. A total of 32 mitochondrial proteins present in at least five species were concatenated for the ML analysis. The model test option was applied to each protein in this alignment.

Comparison of mitochondrial genes

Because most of datasets used for statistical tests did not show a normal distribution, the 'Independent-Samples Kruskal-Wallis Test' was used to evaluate hypotheses. The null hypothesis was that the distribution of factor (e.g., amino acid composition) is identical among taxa from two ingroup clades (i.e., *Cyanidium*-type and *Galdieria*-type) and outgroup taxa. If the null hypothesis is rejected, pairwise comparisons among three groups were applied. Sixteen conserved mitochondrial genes (*sdh*C, *atp*4, *atp*6, *atp*8, *atp*9, *cob*, *cox*1, *cox*2, *cox*3, *nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6) that all encoded in mitogenomes of 13 species were chosen to test amino acid properties.

The alignment of each protein and their ML trees were used for pairwise comparisons and all results were summarized in Table S2. Only conserved sites excluding any gaps were selected from the protein alignment of conserved genes for analysis (3,845/4,825 sites). Average proportions of amino acids were recorded in each group and the values were visualized with standard errors. To compare protein properties for 16 conserved mitochondrial genes, GRAVY (grand average of hydropathy), aliphatic index, and instability index were calculated using ProtParam in ExPASy [37], and the results were summarized in Table S3. The mean values of each feature were illustrated with bar plots with standard error bars.

For the EGT-derived gene detection and transit peptide prediction, nuclear genomes of *G. sulphuraria* 074W and *Cz. merolae* 10D were used as references to find mitochondrion-derived nuclear-encoded genes [23, 29, 68]. EGT genes were further verified based on homologous search by MMSeqs2 [69]. Transit peptide sequences in EGT candidates were predicted by TargetP [70].

Estimation of evolutionary rate and selection

Concatenated datasets of mitochondrial proteins were divided into 64 separate partitions ('-b 100': allocate the sites in 1 to 100 based on the evolutionary rates) based on TIGER values (Tables S4, S5), which calculate the evolutionary rates for each site on the basis of tree-independent approaches to differentially weighted characters [71]. Non-synonymous substitutions per non-synonymous sites (K_a) and synonymous substitutions per synonymous sites (K_s) have been widely used to determine evolutionary selection on genes [72]. A total of 16 conserved mitochondrial genes were selected for alignments to calculate ratios of non-synonymous substitutions per non-synonymous sites and synonymous substitutions per synonymous sites (K_a/K_s), which have been widely used to determine the evolutionary selection on genes [72]. ParaAT [73] was used with MAFFT version 7 [66] and KaKs_Calculator [74] to do this analysis.

Abbreviations

C-type
Cyanidium-type, *G*-type:*Galdieria*-type CZME:*Cyanidioschyzon merolae*), CYSP:Cyanidiophyceae sp., CCYA:*Cyanidiococcus yangmingshanensis*, CDCA:*Cyanidium caldarium*, CDCH:*Cyanidium chilensis*, GAPH:*Galdieria phlegrea*, GASU:*Galdieria sulphuraria*

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Cho CH, Park SI, Ciniglia C, Yang EC, Graf L Bhattacharya D, Yoon HS (2020) Data from: Potential causes and consequences of rapid mitochondrial genome evolution in thermoacidophilic *Galdieria* (Rhodophyta). Dryad Digital Repository. https://doi.org/10.5061/dryad.nvx0k6dp7

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Collaborative Genome Program of the KIMST funded by the MOF [20180430], the NRF of Korea [NRF-2017R1A2B3001923], the Next-generation BioGreen21 Program [PJ01389003] from the RDA, Korea. D.B. was supported by a NIFA-USDA Hatch grant [NJ01170].

Authors' contributions

CHC, SIP, ECY, LG analyzed the data. CC provided materials. HSY and DB supervised the research.

Acknowledgements

The authors thank for the assistance and constructive comments from Soyeong Park, Dongseok Kim, Seung Su Han, and Sunghoon Jang. We would also like to thank Woongghi Shin and Minseok Jeong for taking the microscopic images of cyanidiophycean cells, and the Korea Basic Science Institute (KBSI) for providing electron microscopy facilities.

References

1. Reeb V, Bhattacharya D: **The thermo-acidophilic cyanidiophyceae** (**Cyanidiales**). In: *Red Algae in the Genomic Age.* Edited by Seckbach J, Chapman D: Springer; 2010: 409–426.
2. Hough DW, Danson MJ. Extremozymes. Curr Opin Chem Biol. 1999;3(1):39–46.
3. Rothschild LJ, Mancinelli RL. Life in extreme environments. Nature. 2001;409(6823):1092.
4. Ciniglia C, Yoon HS, Pollio A, Pinto G, Bhattacharya D. Hidden biodiversity of the extremophilic Cyanidiales red algae. Mol Ecol. 2004;13(7):1827–38.
5. Yoon HS, Müller KM, Sheath RG, Ott FD, Bhattacharya D. Defining the major lineages of red algae (Rhodophyta). J Phycol. 2006;42(2):482–92.
6. Seckbach J. **Systematic problems with** *Cyanidium caldarium* **and** *Galdieria sulphuraria* **and their implications for molecular studies**. J Phycol. 1991;27(6):794–6.
7. Ott FD, Seckbach J: **New classification for the genus** Cyanidium **Geitler 1933**. In: *Evolutionary pathways and enigmatic algae: Cyanidium caldarium (Rhodophyta) and related cells.* Springer; 1994: 145–152.
8. Pinto G, Albertano P, Ciniglia C, Cozzolino S, Pollio A, Yoon HS, Bhattacharya D. **Comparative approaches to the taxonomy of the genus** *Galdieria* **Merola (Cyanidiales, Rhodophyta)**. Cryptogamie Algologie. 2003;24(1):13–32.
9. Schwabe G. Uber einige Blaualgen aus dem mittleren und sudlichen Chile. Verh Dt Wiss Verein Santiago NF. 1936;3:113–74.

10. Yoon HS, Ciniglia C, Wu M, Comeron JM, Pinto G, Pollio A, Bhattacharya D. Establishment of endolithic populations of extremophilic Cyanidiales (Rhodophyta). BMC Evol Biol. 2006;6(1):78.
11. Iovinella M, Eren A, Pinto G, Pollio A, Davis SJ, Cennamo P, Ciniglia C: **Cryptic dispersal in non-acidic environments from Turkey of Cyanidiophytina (Rhodophyta)**. *Extremophiles* 2018.
12. Ciniglia C, Cennamo P, De Natale A, De Stefano M, Sirakov M, Iovinella M, Yoon HS, Pollio A. *Cyanidium chilense* **(Cyanidiophyceae, Rhodophyta) from tuff rocks of the archeological site of Cuma, Italy**. Phycological Research. 2019;67(4):311–9.
13. Yang EC, Boo SM, Bhattacharya D, Saunders GW, Knoll AH, Fredericq S, Graf L, Yoon HS. Divergence time estimates and the evolution of major lineages in the florideophyte red algae. Sci Rep. 2016;6:21361.
14. Hsieh CJ, Zhan SH, Lin Y, Tang SL, Liu SL. **Analysis of *rbc*L sequences reveals the global biodiversity, community structure, and biogeographical pattern of thermoacidophilic red algae (Cyanidiales)**. J Phycol. 2015;51(4):682–94.
15. Yoon HS, Nelson W, Lindstrom SC, Boo SM, Pueschel C, Qiu H, Bhattacharya D: **Rhodophyta**. In: *Handbook of the Protists.* Edited by Archibald JM, Simpson AGB, Slamovits CH. Cham: Springer International Publishing; 2017: 89–133.
16. Toplin JA, Norris TB, Lehr CR, McDermott TR, Castenholz RW. Biogeographic and phylogenetic diversity of thermoacidophilic Cyanidiales in Yellowstone National Park, Japan, and New Zealand. Appl Environ Microbiol. 2008;74(9):2822–33.
17. Skorupa DJ, Reeb V, Castenholz RW, Bhattacharya D, McDermott TR. Cyanidiales diversity in Yellowstone National Park. Lett Appl Microbiol. 2013;57(5):459–66.
18. Ciniglia C, Yang EC, Pollio A, Pinto G, Iovinella M, Vitale L, Yoon HS. **Cyanidiophyceae in Iceland: plastid *rbc*L gene elucidates origin and dispersal of extremophilic *Galdieria sulphuraria* and *G. maxima* (Galdieriaceae, Rhodophyta)**. Phycologia. 2014;53(6):542–51.
19. Gross W, Oesterhelt C, Tischendorf G, Lederer F. **Characterization of a non-thermophilic strain of the red algal genus *Galdieria* isolated from Soos (Czech Republic)**. Eur J Phycol. 2002;37(3):477–83.
20. Qiu H, Rossoni AW, Weber APM, Yoon HS, Bhattacharya D. **Unexpected conservation of the RNA splicing apparatus in the highly streamlined genome of *Galdieria sulphuraria***. BMC Evol Biol. 2018;18(1):41.
21. Gross W, Schnarrenberger C. **Heterotrophic growth of two strains of the acido-thermophilic red alga *Galdieria sulphuraria***. Plant Cell Physiol. 1995;36(4):633–8.
22. Qiu H, Price DC, Weber APM, Reeb V, Chan Yang E, Lee JM, Kim SY, Yoon HS, Bhattacharya D. **Adaptation through horizontal gene transfer in the cryptoendolithic red alga *Galdieria phleagrea***. Curr Biol. 2013;23(19):R865–6.
23. Schönknecht G, Chen W-H, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Bräutigam A, Baker BJ, Banfield JF, Garavito RM. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. Science. 2013;339(6124):1207–10.
24. Rossoni A, Price D, Seger M, Lyska D, Lammers P, Bhattacharya D, Weber A: **The genomes of polyextremophilic Cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions**. *eLife* 2019, **8**.
25. Jain K, Krause K, Grewe F, Nelson GF, Weber APM, Christensen AC, Mower JP. **Extreme features of the *Galdieria sulphuraria* organellar genomes: a consequence of polyextremophily?**. Genome Biol Evol. 2015;7(1):367–80.
26. Ohta N, Sato N, Kuroiwa T. **Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence**. Nucleic Acids Res. 1998;26(22):5190–8.
27. Liu S-L, Chiang Y-R, Yoon HS, Fu H-Y. **Comparative genome analysis reveals *Cyanidiococcus* gen. nov., a new extremophilic red algal genus sister to *Cyanidioschyzon* (Cyanidioschyzonaceae, Rhodophyta)**. *Journal of Phycology* 2020 (accepted).
28. Suzuki K, Kawano S, Kuroiwa T. **Single mitochondrion in acidic hot-spring alga: behaviour of mitochondria in *Cyanidium caldarium* and *Galdieria sulphuraria* (Rhodophyta, Cyanidiophyceae)**. Phycologia. 1994;33(4):298–300.
29. Matsuzaki M, Misumi O, Shin-i T, Maruyama S, Takahara M, Miyagishima S-y, Mori T, Nishida K, Yagisawa F, Nishida K *et al*: **Genome sequence of the ultrasmall unicellular red alga** Cyanidioschyzon merolae **10D**. *Nature* 2004, **428**:653.
30. Adams KL, Daley DO, Whelan J, Palmer JD. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. Plant Cell. 2002;14(4):931–43.

31. Smits P, Smeitink JAM, van den Heuvel LP, Huynen MA, Ettema TJG. Reconstructing the evolution of the mitochondrial ribosomal proteome. Nucleic Acids Res. 2007;35(14):4686–703.
32. Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S. On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. Res Microbiol. 2011;162(1):53–70.
33. Elleuche S, Schröder C, Sahm K, Antranikian G. Extremozymes—biocatalysts with unique properties from extremophilic microorganisms. Curr Opin Biotechnol. 2014;29:116–23.
34. Jaenicke R. Stability and stabilization of globular proteins in solution. J Biotechnol. 2000;79(3):193–203.
35. Das R, Gerstein M. The stability of thermophilic proteins: a study based on comprehensive genome comparison. Funct Integr Genomics. 2000;1(1):76–88.
36. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32.
37. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook. Edited by Walker: Springer; 2005. pp. 571–607.
38. DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet. 2005;6(9):678–87.
39. Hong H, Joh NH, Bowie JU, Tamm LK: **Methods for Measuring the Thermodynamic Stability of Membrane Proteins**. In: *Methods in Enzymology.* Edited by Johnson ML, Holt JM, Ackers GK, vol. 455: Academic Press; 2009: 213–236.
40. Tamm LK, Hong H, Liang B. Folding and assembly of β-barrel membrane proteins. Biochimica et Biophysica Acta - Biomembranes. 2004;1666(1–2):250–63.
41. Necşulea A, Lobry JR. A new method for assessing the effect of replication on DNA base composition asymmetry. Mol Biol Evol. 2007;24(10):2169–79.
42. Grigoriev A. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. 1998;26(10):2286–90.
43. Agier N, Fischer G. The mutational profile of the yeast genome Is shaped by replication. Mol Biol Evol. 2011;29(3):905–13.
44. Yasukawa T, Reyes A, Cluett TJ, Yang M-Y, Bowmaker M, Jacobs HT, Holt IJ. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. The EMBO Journal. 2006;25(22):5358–71.
45. Pohjoismäki JL, Goffart S. Of circles, forks and humanity: Topological organisation and replication of mammalian mitochondrial DNA. Bioessays. 2011;33(4):290–9.
46. Holt IJ, Reyes A. Human mitochondrial DNA replication. Cold Spring Harb Perspect Biol. 2012;4(12):a012971.
47. Lujan SA, Williams JS, Pursell ZF, Abdulovic-Cui AA, Clark AB, McElhinny SAN, Kunkel TA. Mismatch repair balances leading and lagging strand DNA replication fidelity. PLoS Genet. 2012;8(10):e1003016.
48. Fijalkowska IJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM. **Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome**. Proc Natl Acad Sci U S A. 1998;95(17):10020.
49. Pavlov YI, Mian IM, Kunkel TA. Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. Curr Biol. 2003;13(9):744–8.
50. Kunkel TA, Burgers PM. Dividing the workload at a eukaryotic replication fork. Trends Cell Biol. 2008;18(11):521–7.
51. Bielawski JP, Gold JR. Mutation patterns of mitochondrial H- and L-strand DNA in closely related Cyprinid fishes. Genetics. 2002;161(4):1589–97.
52. Szczepanik D, Mackiewicz P, Kowalczuk M, Gierlik A, Nowicka A, Dudek MR, Cebrat S. Evolution rates of genes on leading and lagging DNA strands. J Mol Evol. 2001;52(5):426–33.
53. Lynch M, Koskella B, Schaack S. Mutation pressure and the evolution of organelle genomic architecture. Science. 2006;311(5768):1727.
54. Lin Q, Cui P, Ding F, Hu S, Yu J. Replication-associated mutational pressure (RMP) governs strand-biased compositional asymmetry (SCA) and gene organization in animal mitochondrial genomes. Curr Genomics. 2012;13(1):28–36.
55. Bergstrom CT, Pritchard J. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics. 1998;149(4):2135–46.

56. Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, Trifunovic A, Larsson N-G. Strong purifying selection in transmission of mammalian mitochondrial DNA. PLoS Biol. 2008;6(1):e10.
57. Yoon HS, Hackett JD, Bhattacharya D. A single origin of the peridinin-and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. Proc Natl Acad Sci U S A. 2002;99(18):11724–9.
58. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
59. Cho CH, Choi JW, Lam DW, Kim KM, Yoon HS. Plastid genome analysis of three Nemaliophycidae red algal species suggests environmental adaptation for iron limited habitats. PLoS One. 2018;13(5):e0196995.
60. Harris RS: **Improved pairwise alignmnet of genomic DNA**. *Dissertation*. Pennsylvania State University; 2007.
61. Boyen C, Leblanc C, Bonnard G, Grienenberger JM, Kloareg B. **Nucleotide sequence of the *cox*3 gene from *Chondrus crispus*: evidence that UGA encodes tryptophan and evolutionary implications**. Nucleic Acids Res. 1994;22(8):1400–3.
62. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44(W1):W54–7.
63. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11–6.
64. Salomaki ED, Lane CE. Red algal mitochondrial genomes are more complete than previously reported. Genome Biol Evol. 2017;9(1):48–63.
65. Bedrat A, Lacroix L, Mergny J-L. Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. 2016;44(4):1746–59.
66. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
67. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2014;32(1):268–74.
68. Nozaki H, Takano H, Misumi O, Terasawa K, Matsuzaki M, Maruyama S, Nishida K, Yagisawa F, Yoshida Y, Fujiwara T, et al. **A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga** *Cyanidioschyzon merolae*. BMC Biol. 2007;5(1):28.
69. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026.
70. Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2(4):953.
71. Cummins CA, McInerney JO. A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. Syst Biol. 2011;60(6):833–44.
72. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. PLoS Genet. 2008;4(12):e1000304.
73. Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. Biochem Biophys Res Commun. 2012;419(4):779–81.
74. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics. 2006;4(4):259–63.

Figures

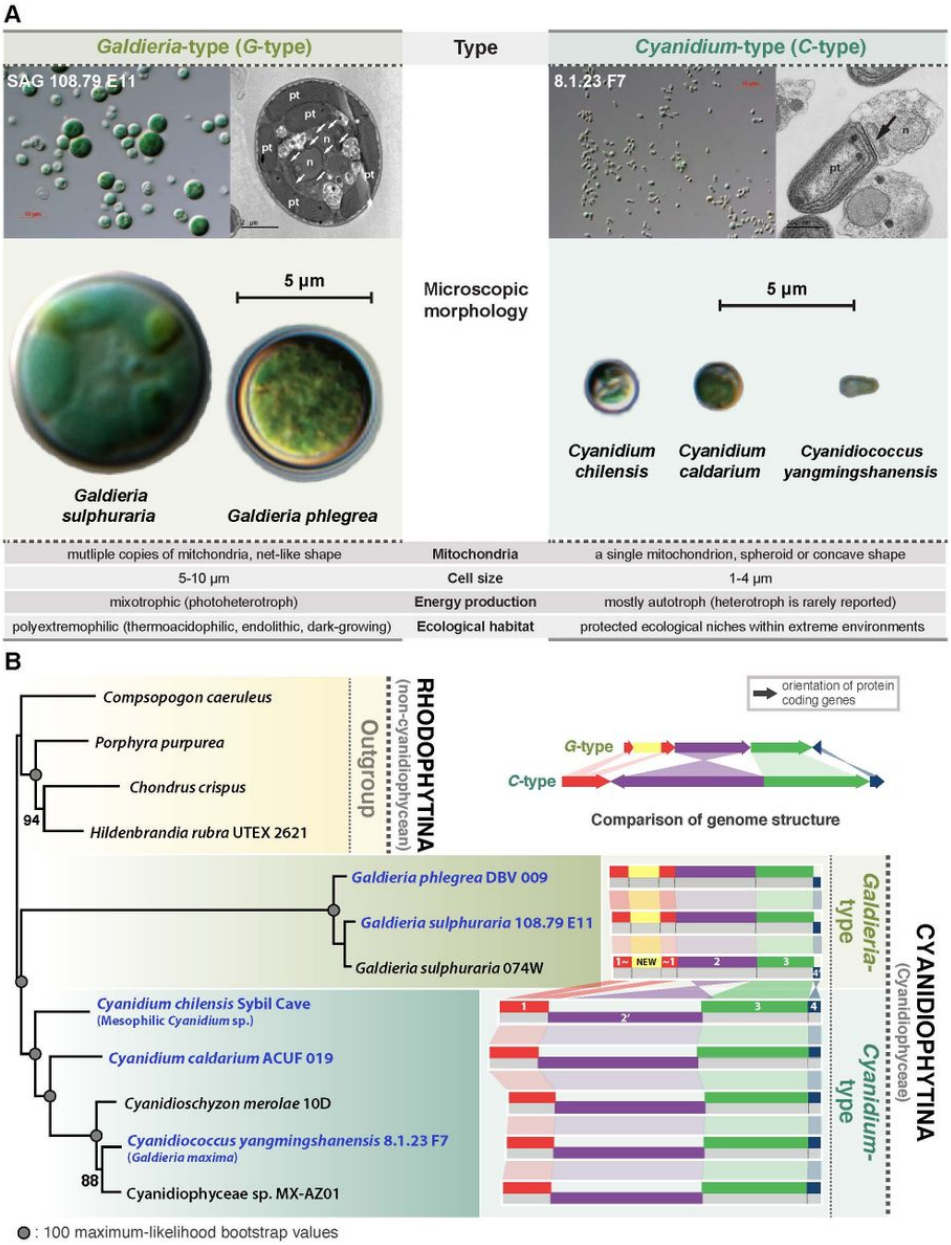

Figure 1

Overview of the major characteristics of Cyanidiophyceae and its phylogeny. (A) Comparison of key characteristics of the Cyanidium-type and Galdieria-type species providing morphological observations from two different types of cyanidiophycean cells. n: nucleus, pt: plastid, arrow: mitochondria. (B) Maximum-likelihood phylogeny using a protein alignment of 12 mitochondrial genomes. This tree was constructed based on the alignment of 32 mitochondrial proteins. Four non-cyanidiophycean species were chosen as the outgroup. The simplified genome structure of cyanidiophycean mitochondria are illustrated next to the phylogenetic tree.

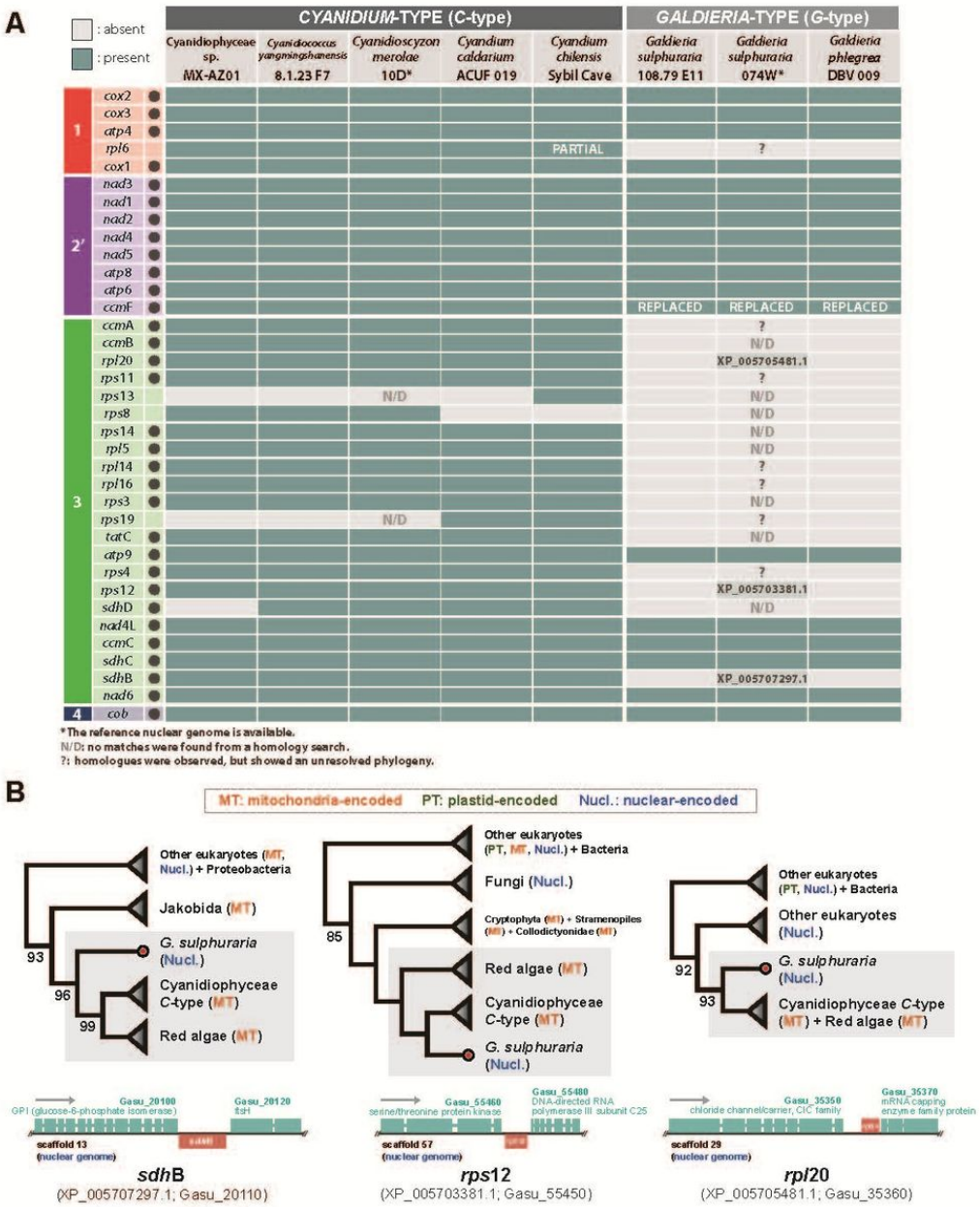

Figure 2

Mitochondrial gene content in cyanidiophycean mitogenomes and the phylogeny of three EGT-derived genes. (A) The presence and absence of 36 mitochondrial genes in Cyanidiophyceae is shown. Black dots indicate genes used for concatenated dataset phylogenetic analysis. Each number in colored box represents different gene synteny and reversed orientation is indicated by the prime mark (′). (B) The phylogeny of three EGT-derived genes and their location in the Galdieria sulphuraria 074W genome. Bootstrap values >90% support merged clades (triangles) and bootstrap support values <50% are now shown.

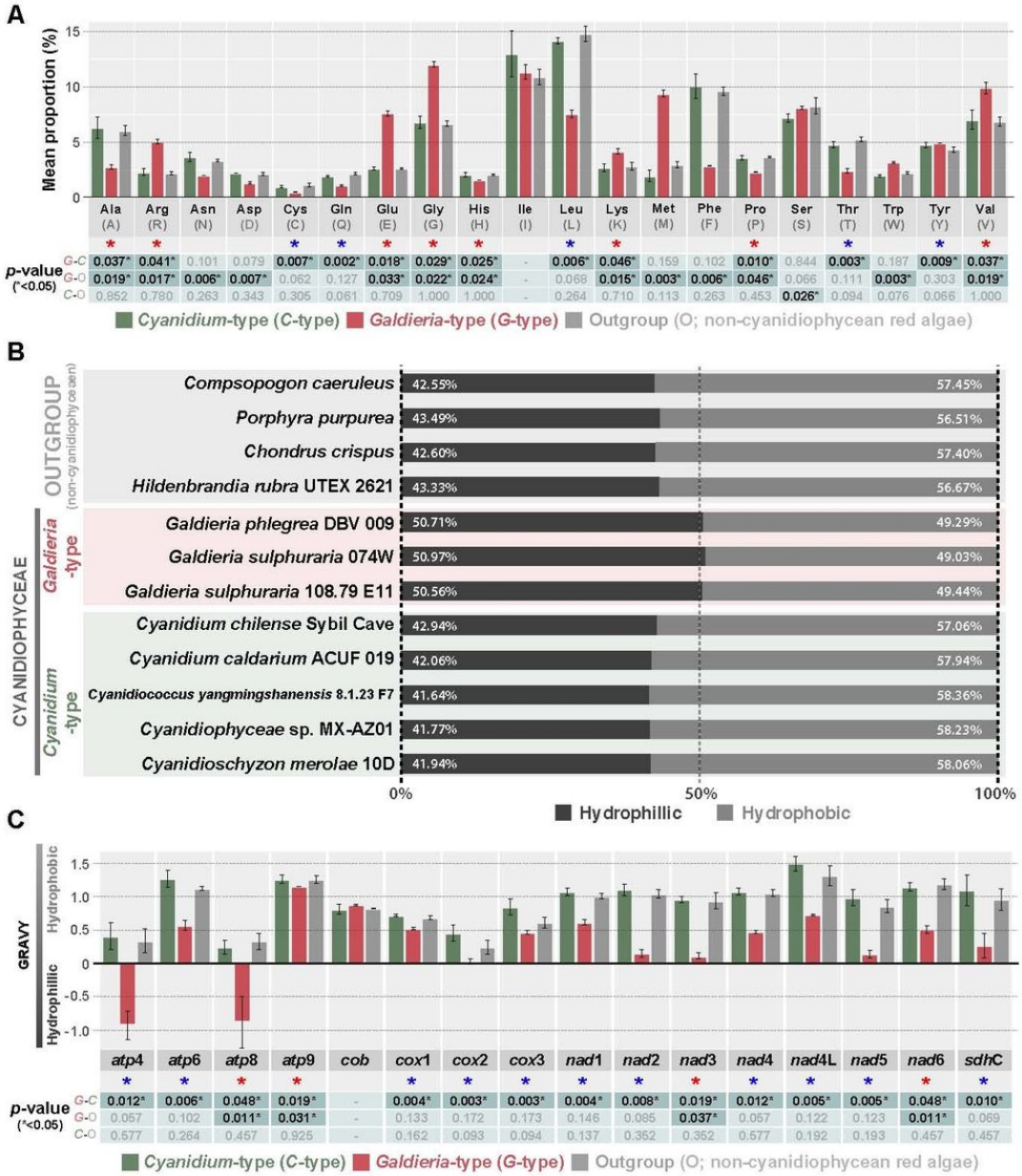

Figure 3

Amino acid composition of conserved sites and their properties. A total of 16 mitochondrial genes present 13 species was chosen for this analysis. Blue asterisks indicate a statistically significant difference between G-type and C-type. A significant difference between both G-type and C-type and G-type and non-cyanidiophycean red algae is indicated by a red asterisk (see Tables S6, S7). (A) Relative amino acid compositions of conserved mitochondrial genes. (B) Relative hydrophilic and hydrophobic composition of conserved mitochondrial genes. (C) Comparison of hydropathy in 16 conserved mitochondrial proteins.

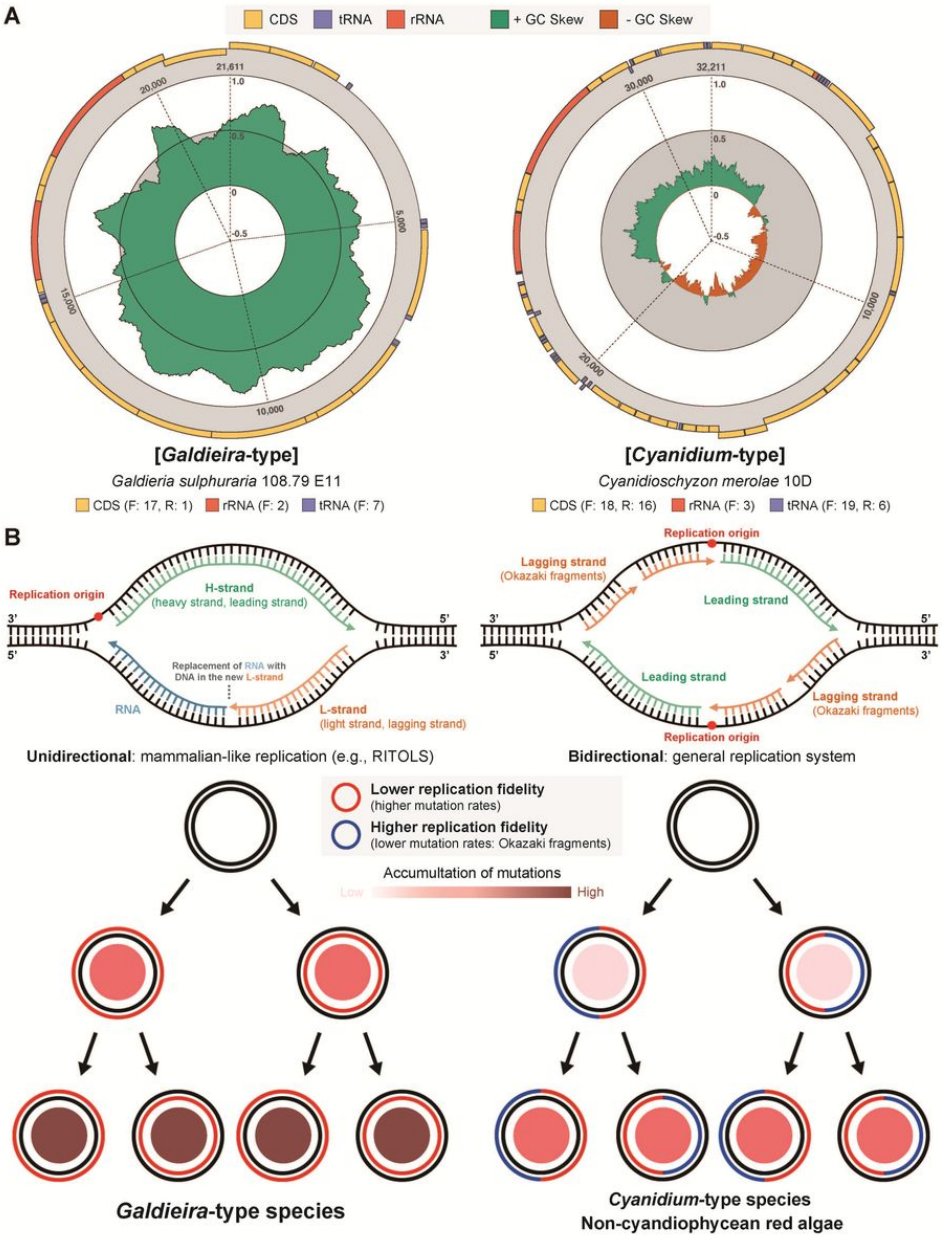

Figure 4

Two different models for mitogenome replication in Cyanidiophyceae. Unidirectional and conservative replication (separate leading and lagging strands for each daughter strand) in Galdieria-type and bidirectional and semiconservative replication (mixed leading and lagging strand for each daughter strand) in Cyanidium-type. (A) GC-skew of representative structure comparison. F: forward, R: reverse. (B) Hypothetical replication system and mitogenome inheritance model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesS1S8200527.xlsx
- SupplementaryMaterials20200618BMCEvolBiol.pdf