Pyropia yezoensis genome reveals diverse mechanisms of carbon acquisition in the intertidal environment

Dongmei Wang, Xinzi Yu, Kuipeng Xu, Guiqi Bi, Min Cao, Ehud Zelzion, Chunxiang Fu, Peipei Sun, Yang Liu, Fanna Kong, Guoying Du, Xianghai Tang, Ruijuan Yang, Junhao Wang, Lei Tang, Lu Wang, Yingjun Zhao, Yuan Ge, Yunyun Zhuang, Zhaolan Mo, Yu Chen, Tian Gao, Xiaowei Guan, Rui Chen, Weihua Qu, Bin Sun, Debashish Bhattacharya & Yunxiang Mao

Changes in atmospheric CO2 concentration have played a central role in algal and plant adaptation and evolution. The commercially important red algal genus, Pyropia (Bangiales) appears to have responded to inorganic carbon (Ci) availability by evolving alternating heteromorphic generations that occupy distinct habitats. The leafy gametophyte inhabits the intertidal zone that undergoes frequent emersion, whereas the sporophyte conchocelis bores into mollusk shells. Here, we analyze a high-quality genome assembly of Pyropia yezoensis to elucidate the interplay between Ci availability and life cycle evolution. We find horizontal gene transfers from bacteria and expansion of gene families (e.g. carbonic anhydrase, antioxidative related genes), many of which show gametophyte-specific expression or significant up-regulation in gametophyte in response to dehydration. In conchocelis, the release of HCO3- from shell promoted by carbonic anhydrase provides a source of Ci. This hypothesis is supported by the incorporation of 13C isotope by conchocelis when co-cultured with 13C-labeled CaCO3.
The seaweed used most frequently to produce **nori** (sushi wrap), *Pyropia yezoensis*, has a high nutritional value, including up to 25–30% protein by dry weight, vitamins (in particular, B12), and oligosaccharides. The mariculture of *nori* in countries such as Korea, Japan, and China reaches an annual production of ca. 1.8 million tons\(^1\), making it, globally, one of the most economically important marine crops. *P. yezoensis* is a member of the order Bangiales (family Bangiophyceae) in the red algae (Rhodophyta). Molecular phylogenetic studies, which incorporate the fossil record, suggest that Bangiophyceae diverged from their sister group, the Florideophyceae, around 900 million years ago (Mya). More limited phylogenomic data indicate that the radiation of species within the Bangiophyceae occurred about ~250 Mya ago in the Permo-Carboniferous era, leading to formation of the order Bangiales\(^2,3\). Species in Bangiales, for instance, *Pyropia* spp. (e.g., *P. yezoensis*, *P. haitanensis*, *Porphyra umbilicalis*) and *Bangia* spp., typically have a life cycle characterized by the alternation of heteromorphic generations that live in distinctive habitats (Fig. 1). The edible, leafy thallus (gametophyte) inhabits the upper intertidal zone of exposed coasts. These seaweeds attach to rocky substrata using a rhizoidal holdfast (Fig. 1a, g) and are exposed to air during low tide. Many intertidal algae, including *P. yezoensis*, can utilize atmospheric CO\(_2\) as the carbon source for photosynthesis\(^4\)–\(^6\). However, frequent emersion also leads to fluctuating environmental stresses, such as extreme drought, high ultraviolet (UV) irradiance, and high temperature. For this reason, the *P. yezoensis* thallus, which is capable of tolerating broad and extreme environmental stresses, is of interest as a model for studying molecular mechanisms of stress resistance. Photosynthesis in the thallus can recover rapidly from >80% water loss that results in the death of unwanted epiphytes\(^7\). This robustness facilitates commercial nori cultivation in the open sea.

In contrast to *P. haitanensis* (another edible *nori* species), *P. yezoensis* is monecious. Mature thalli release sperm to fertilize carpospores (Fig. 1b, c). The resulting diploid zygotes germinate into the filamentous and branched conchocelis phase (the sporophyte phase). The conchocelis burrows into the inner layers of mollusk shells and lives within the calcareous matrix (Fig. 1d, e). The shell appears to provide protection from UV damage and wave exposure\(^8\). However, CO\(_2\) is scarce in the enclosed space that has reduced water exchange, which may limit growth of the conchocelis. Past work shows that increased atmospheric CO\(_2\) does not affect conchocelis photosynthetic performance, suggesting the existence of an alternative carbon source\(^9\). Conchosporangia develop in mature conchocelis and release conchospores that attach to rocky substrata or *P. yezoensis* nets in mariculture and develop into thalli, i.e., seedlings (Fig. 1f, g).

*Pyropia* species appear to have evolved two distinct strategies that are reflected in their life cycle: (1) the gametophyte thallus utilizes atmospheric CO\(_2\) while living in a harsh environment and (2) the sporophyte conchocelis is protected in the shell environment but faces C\(_i\) limitation. The genome-wide impact of this alternation of heteromorphic generations is poorly understood. The genomes of *P. umbilicalis* and *P. haitanensis* have recently been published\(^9,10\). Here we present the genome of *P. yezoensis* and analyze Bangiales genome and transcriptomic data to gain insights into the evolutionary mechanisms that allow these

![Fig. 1 The life cycle of *P. yezoensis* showing the alternation of heteromorphic generations inhabiting distinctive niches.](Image)

Ploidy was added as “1N” and “2N” in the upper left corner of each subgraph. a Gametophyte thalli adhered to rocks are exposed to air during low tide. b Carpospores and sperm form in mature thalli; c released zygotes, showing one with a germ tube. d Purple patches on oyster shell are boring sporophyte conchocelis. e Swelling filamentous conchocelis that are going to produce conchosporangia. f Released conchospores. g Thallus seedling with a holdfast formed at the base. h *P. yezoensis* nets in mariculture farms. Scale bars in b, c, and e–g 50 μm. Asexual reproduction is not shown here. Microscope observation in b, c, and e, g were repeated independently on multiple *Pyropia* samples (n > 10) showing similar morphology.
species to thrive in their stressful habitats vis-à-vis their dimorphic life cycle.

Results

Genome assembly. Assembly of the P. yezoensis genome using PacBio long-read sequencing data followed by removal of bacterial contaminant sequences resulted in 660 contigs with N50 = 340.8 kbp (Supplementary Tables 1 and 2; Supplementary Figs. 1 and 2). The genome size was estimated to be ca. 108 Mb, which is bigger than P. umbilicalis albeit comparable to the florideophyte Chondrus crispus. Hi-C and BioNano data linked the contigs into three scaffolds, S1 = 29.4 Mbp, S2 = 34.3 Mbp, and S3 = 43.6 Mbp, equivalent to the three chromosomes observed in haploid P. yezoensis (Supplementary Tables 1–3, Supplementary Figs. 3 and 4). These scaffolds included 95% of the genome data with the remaining 524 contigs comprising 5.5 Mbp. Nearly one-half of the P. yezoensis genome (48.0%) is predicted to be repeat elements, of which long terminal repeat (LTR) retrotransposons are predominant (Supplementary Note 1, Supplementary Table 4). The numbers of intact LTR is positively correlated with genome size in red algae (R² = 0.9437), explaining genome size growth in Pyropia (Supplementary Fig. 5) and Gracilaria (Supplementary Fig. 6).

Combining ab initio prediction with PacBio transcriptome data, 12,855 protein-coding genes were identified in the P. yezoensis genome (Supplementary Table 3). These contain 84% of the core eukaryotic single-copy genes in the BUSCO database, supporting the completeness of the assembly (Supplementary Fig. 6). To elucidate the evolution of genome structure in Bangiophyceae, we compared gene synteny in P. yezoensis, P. haitanensis, and P. umbilicalis. Long syntenic regions were identified in the two Pyropia genomes, whereas only a limited amount of synteny was found between P. yezoensis and P. umbilicalis (Supplementary Note 2, Supplementary Fig. 7). These results suggest that multiple genome rearrangements occurred after the divergence of these two genera.

Expanded orthogroups (OGs) and their phase-specific transcription. To gain insights into the origin of stress tolerance pathways in the P. yezoensis genome, we inferred orthologous relationships between the three available Bangiales genomes and 15 other genomes representing Florideophyceae, green algae, heterokonts, and plants. A total of 10,327 OGs were found, of which 424 single-copy OGs were used to build a concatenated gene phylogeny (Supplementary Fig. 8). These data were used to estimate the divergence time of Bangiales, which was found to be during the Perm–Carboniferous glaciation (300–225 Mya, Supplementary Note 3, Supplementary Fig. 8), consistent with previous studies. A total of 578 OGs were significantly expanded in Bangiales when compared to Florideophyceae and include the following Pfam annotations: antioxidative systems (e.g., thioredoxin (TRX), catalase (CAT)), Cu/Zn-superoxide dismutase (SOD), carboxymuconolactone decarboxylase, D-arabinono-1,4-lactone oxidase), stress resistance (e.g., lipoxygenase, HSP20, bacterial low temperature requirement A protein, subtilase, beta-lactamase), detoxification, carbohydrate metabolism (GH17, GH27, GH36, GH35, GH53, GH70, GH85, GH97, GT2, GT21, CMB, CMB53), transport, and tyrosine metabolism (Fig. 2a and Supplementary Fig. 9).

To determine whether a link exists between OG expansion and life history stage, we studied the differential expression of genes (DEGs) in P. yezoensis when comparing the gametophyte thallus (gam) and the sporophyte conchocelis (spo). Dually expressed genes with fragments per kilobase of transcript per million mapped reads (FPKM) < 5 in gam and > 25 in spo and unique genes with > 10-fold FPKM in spo were defined as spo-specifically transcribed genes. Gam-specifically transcribed genes were defined in the same way in the opposite direction. Genes with FPKM > 5 in both generations and fold change < 10 were considered to be constitutively expressed. Because desiccation is one of the most important environmental stresses faced by P. yezoensis thalli in the intertidal zone, we studied gene expression patterns during a time course of desiccation and rehydration. In this regard, SOD catalyzes the dismutation of two molecules of the superoxide radical into one molecule of oxygen and one molecule of hydrogen peroxide (H₂O₂), which can be degraded into water and oxygen in the presence of catalase. Thus the antioxidant activity of SODs is of particular importance to the amelioration of cell damage induced by reactive oxygen species (ROS) and maintenance of cell homeostasis. In plants, the SOD family consists of three classes of metalloenzymes, Fe-SOD, Mn-SOD, and Cu/Zn-SOD. In the P. yezoensis genome, we found 14 SOD genes, whereas this number is 7, 3, and 3 in C. crispus, Porphyridium purpureum, and Cyanidioschyzon merolae, respectively (Fig. 2a). DEG analysis showed that SODs in the conchocelis and thallus, as well as under dehydration-stressed condition, can be categorized into three groups: (1) spo-specifically transcribed SODs that include two genes exhibiting high relative transcriptional levels in sporophyte (py00127.11, py01435.11); (2) five gam-specifically transcribed SOD genes; (3) the remaining four genes show constitutive expression in the two stages (Fig. 2b). Interestingly, 4/5 gam-specifically transcribed SOD genes exhibit significant upregulation under dehydration, whereas the other genes show minimal expression variation (Fig. 2c). The correlation between SOD gene expansion, specific expression in the gametophyte thallus, and burst of transcriptional activity after water loss suggest that SODs played a significant role in the origin and maintenance of the heteromorphic life cycle of P. yezoensis as a response to life in the intertidal zone. In addition to SOD, several other oxidative stress-related OGs, including lipoxygenase (LOX), catalase, and tyrosinase (TYR) have expanded gene numbers in Pyropia species (Fig. 2a). Members of these gene families exhibit similar transcriptional patterns to SODs, with upregulation in a specific life history stage and high expression of gam-specifically transcribed genes under osmotic stress (Supplementary Note 4, Supplementary Figs. 10 and 11). The conserved response across these different gene families hints at the evolution of a regulatory mechanism in Bangiales for life in the intertidal zone.

C₄ assimilation. Beyond environmental stresses, pathways of C₄ acquisition are common targets for selection in algae and plants. Therefore, it is significant that one of the most prominent cases of gene expansion in P. yezoensis is for carbonic anhydrase (CA) encoding sequences. As a key component of the carbon-concentrating mechanism (CCM) in photosynthetic organisms, intracellular CA catalyzes the reversible conversion of carbon dioxide to bicarbonate, preventing the escape of CO₂ through cellular membranes. Bicarbonate is stored in the pyrenoid and subsequently converted to CO₂ by CA near the active site of RuBiSCO to allow efficient carbon fixation. CA genes are well characterized in Chlamydomonas reinhardtii, diatoms, and plants. Both internal and external CA are required for the uptake of HCO₃⁻ in Porphyra. There are six CA subfamilies (α, β, γ, δ, ε, η) that exhibit high protein sequence divergence and different subcellular localizations. We found 24, 22, and 16 putative CA homologs in P. yezoensis, P. haitanensis, and P. umbilicalis, respectively, which greatly exceeds the number in Florideophyceae, green algae, and algae with secondary plastids, such as diatoms. CA genes are distributed across the three P. yezoensis chromosomes. A total of 16 out of the 24 CAs formed...
conserved orthologous groups with counterparts in *P. haitanensis* in collinear blocks that are shared by these two species (Supplementary Note 2, Supplementary Fig. 12). Gene expansion in *P. yeoensis* is primarily explained by αCAs (15 copies) and βCAs (6 copies). The number of γCAs did not show significant variation among photosynthetic organisms (Fig. 3a).

The αCAs in *P. yeoensis* show relatively low protein identity (pairwise values ranging from 37 to 87%) and are distributed in seven subgroups based on protein sequence similarity. Eleven of these *P. yeoensis* genes are grouped with the corresponding orthologs from *P. umbilicalis* (Supplementary Fig. 13). The lack of *P. umbilicalis* orthologs of the other four αCAs in *P. yeoensis* may be attributed to gene loss or missing data. Most αCAs harbor one typical CA domain, however, in the longest protein αCA15 (and Pu-αCA11) that is of length 1337 amino acids, four CA domains are tandemly repeated (Supplementary Fig. 14). This multi-domain CA was limited to *P. yeoensis*. Although most αCAs are not considered to form multimers, several studies with fungi21, *C. reinhardtii*22, and animals23 report dimer organization of α-CA. Dimers, tetramers, and octamers are often found for βCAs that form multiple active sites24. Therefore, the four CA domains in αCA15 likely enhance enzyme stability and activity.

Based on signal peptide prediction using PredAlgo, nine, four, five, and six CA isoforms may be localized in plastid, mitochondria, extracellular matrix, and other regions in the cytoplasm, respectively (Fig. 3b, Supplementary Fig. 13). The extracellular localization of αCA5 and αCA9, as well as the cytoplasmic localization of βCA4, were confirmed using transient expression of green fluorescent protein (GFP) fusion proteins in tobacco (Supplementary Table 5, Supplementary Fig. 15). Although αCA14 was predicted to be a cytosolic protein, it was clearly localized to the extracellular matrix in tobacco. The following CAs αCA2, αCA4, αCA7, αCA13, αCA11, βCA1 and βCA3 were plastid-targeted CAs, although their suborganellar localization were unclear. These proteins might be involved in the conversion of CO₂ and HCO₃⁻ in the plastid stroma or the

---

**Fig. 2** The copy numbers of genes related to antioxidative functions and the transcriptional pattern of SOD genes. a Copy numbers of SOD, CAT, LOX, TYR, and TRX genes in algal genomes. The dendrogram was generated based on the phylogenetic tree constructed based on the 424 single-copy orthologous gene sets identified from *P. yeoensis* and the other 17 published algal genomes. The expanded orthologous related to antioxidative functions are shown in different colors. The size of the circle represents gene copy number. Percentages above the heatmap denote the degree of water loss and 80%-re and 50%-re denote the rehydrated samples after 80 and 50% of water loss, respectively. WL water loss.

---

**ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-17689-1**

4 NATURE COMMUNICATIONS | (2020) 11:4028 | https://doi.org/10.1038/s41467-020-17689-1 | www.nature.com/naturecommunications
their likely localization in diverse subcellular compartments, P. yezoensis of stages of desiccation (Fig. 3b). Overall, the enhanced expression of the individual CAs were markedly different. The extracellular in at least one stage of osmotic stress, yet the expression patterns were observed with the other extracellular followed by a decrease when rehydrated. Similar transcriptional transcription during dehydration from 20 to 80% water loss, Plastidal α the thallus than in the conchocelis (Supplementary Fig. 16). We CAs) and another three that exhibit higher expression in the lower level than in the sporophyte. Plastid-targeting CAs might function in either chloroplast stroma or thylakoid (green dashed square) where RuBisCO is localized. Heatmap beside or under these, transcriptionally active in the leafy thallus and conchocelis. Of α CA9 because they are exclusively transcribed in sporophyte (blue ellipses), except for α-CA11 which is transcribed in gametophytes at a much lower level than in the sporophyte. Interestingly, we also found three spo-specifically transcribed CAs and another three that exhibit higher expression in the conchocelis than in the thallus (Supplementary Fig. 16). Two of them (α-CA5 and α-CA9) were experimentally validated to be secreted proteins (Supplementary Fig. 13, Supplementary Fig. 15). The external CA enzymatic activity in free conchoelis was 7.86 Units per g fresh weight, slightly higher than the value previously reported in the Pyropia thallus25. Extracellular CAs in microbes can accelerate calcium carbonate dissolution in the presence of H⁺ that is produced during the hydration of CO₂26,27. Conchocelis live en masse in boreholes in the calcareous matrix of mollusk shells (Fig. 4a, b). The nature of evolution forces driving conchocelis burrowing into shell and how these cells acquire sufficient C₄ to support photosynthetic growth are open questions. Given that shells are composed primarily of calcium carbonate, we tested whether the conchocelis utilizes this substrate as a carbon source with the aid of extracellular CAs. To do this, we determined the Ca²⁺ concentration and pH near the shell in control medium (alga-free shell), growth medium (conchocelis-dwelling shell), and CA inhibitory medium (conchocelis-dwelling shell with the CA inhibitor acetazolamide). The Ca²⁺ concentration in the growth medium increased rapidly and continuously, until reaching a plateau on the 15th day when the conchocelis covered all of the shell. These values in the control and inhibitory media showed mild increases in the initial 3 days and went up gradually at a similar rate, possibly due to the natural release of Ca²⁺ from the shell matrix (Fig. 4c). The increasing rates over the first 3 days in the inhibitory and control medium were slightly different, possibly attributed to the incomplete inhibition of CAs because it may have taken some time for the acetazolamide to enter the shell boreholes. A similar result was found with respect to variation of pH values (Supplementary Fig. 18). Previous studies in eudendolithic cyanobacterium Mastigocoleus testarum28 showed that the released Ca²⁺ from the calcocite passively diffused into the apical cell, moved thereafter from cell to cell, and was finally pumped into the liquid medium by the distal cell through a plastidial α-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, and as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plastidal α-CA4 exhibited an opposite pattern, with low expression when rehydrated. Similar transcriptional patterns were observed with the other extracellular α-CA13, the cytoplasmic β-CA4, as well as the plastidal α-CA13 and β-CA3. Plasti...
P-type calcium ATPase. In *Pyropia*, two genes encoding P-type calcium ATPase were identified and both of them exhibited high expression in the conchocelis. We added vanadate, the specific inhibitor of P-type cation-transporting ATPase, into the growth medium and found that Ca$^{2+}$ extrusion was strongly inhibited, implying a pivotal role for the transporters in this process (Fig. 4c). These results suggest that the extracellular CAs produced by the conchocelis were capable of converting shell CaCO$_3$ into HCO$_3^-$, which was subsequently taken up by the alga to support photosynthetic growth.

To corroborate this hypothesis, we grew free conchocelis with $^{13}$C-labeled CaCO$_3$ and analyzed the abundance of $^{13}$C in organic acids of the tricarboxylic acid (TCA) cycle. A total of 39% of total carbon was labeled with $^{13}$C isotopes in this experiment (LABEL), whereas the abundance in conchocelis cells co-cultured with $^{12}$C-CaCO$_3$ (CTRL) was 1%. Specifically, CTRL cells produced 0.4–1.7% of $^{13}$C isotopologs of succinic acid, fumaric acid, 2-ketoglutaric acid, oxaloacetic acid, malic acid, and citric acid. Between 2.1 and 4.9% of $^{13}$C was detected in the LABEL cells, which is far greater than the natural abundance of $^{13}$C (Fig. 4d). These results demonstrate that conchocelis cells assimilate available $^{13}$C-labeled CaCO$_3$ that may be made available by excreted CAs that transform carbonate into bicarbonate.

**Role of horizontal gene transfer (HGT) in *P. yezoensis* evolution.** To elucidate the potential impact of HGT on *P. yezoensis* gene origin and the evolution of its distinct life history, we did a comprehensive phylogenetic analysis of the predicted gene inventory. This approach identified 51 genes (ca. 0.04% of the inventory) that were acquired specifically by the Bangiales ancestor via HGT from prokaryotic donors (Dataset S2). These genes all formed well-supported monophyletic clades with various prokaryotic lineages. The most common roles played by HGT-derived genes include stress response (including antioxidant functions), antibiotic resistance, carbohydrate-related metabolism, and transport (Fig. 5a). Among them, seven genes formed expanded gene families, including SOD, CA, and N-acetyl-D-glucosamine 2-epimerase genes, glycine N-methyltransferase gene, carbohydrate-active enzyme, and peroxidexin encoding genes. Interestingly, 52.9% of HGTs exhibited specific transcription in either of the two generations (29.4% in gametophyte and 23.5% in sporophyte), whereas the percentage of genome-wide expression was 16.7% (Fig. 5b). The enrichment of phase-specifically transcribed gene expression among HGT-derived genes suggests that these foreign sequences provide adaptive functions to *Pyropia* species. This hypothesis is supported by the significant variation in transcriptional level under osmotic stresses resulting from dehydration or rehydration. In total, 13/15 gam-specifically transcribed HGT-derived genes exhibited more than twofold change in at least two time points under stress (Fig. 5c).

HGT is an important mechanism for gaining access to novel genes and functions. Among eukaryotes, this principle is exemplified by the red algal genus *Galdieria* that has about 1% of the gene inventory in species arising through HGT$^{29}$. Many of these novel genes confer an adaptive function in the hot spring...
habitats where these species live. The HGT origin of the SOD gene\textsuperscript{30} in Pyropia and the LOX gene in the kelp Saccharina japonica\textsuperscript{31} have previously been reported. Among the HGT candidates we found in P. yezoensis, eight genes are involved in stress responses. Combining the identification of HGT-derived CA gene for CCM, this indicates the important role of HGT in allowing Pyropia to thrive in its stressful habitat and overcome C\textsubscript{i} limitation. This hypothesis is supported by the thallus-specific gene expression patterns of HGT-derived genes.

High GC content in Bangiales genomes. An interesting feature of the P. yezoensis genome is the high overall GC content of 65%, which reaches 71% in coding regions. This attribute is shared with other Bangiophyceae including P. haitanensis\textsuperscript{10} (67.9%), Bangia fuscopurpurea\textsuperscript{32} (64%), and P. umbilicalis\textsuperscript{9} (66%), whereas the GC content ranges from 37 to 57% in other red algae (e.g., C. crispus\textsuperscript{33}, P. purpureum\textsuperscript{34}, Galdieria sulphuraria\textsuperscript{29}, and C. merolae\textsuperscript{35}; Supplementary Fig. 19a). No significant bias in GC content was observed in biological functions encoded by the P. yezoensis genome (Supplementary Fig. 19b). Comparison of the GC content between the two clades indicated an expansion of AT–GC transitions in the ancestor of Bangiophyceae. Among the 61 codons, 20 were enriched in Bangiales when compared to Florideophyceae (Supplementary Fig. 19c). Their GC content, on average was 80%; moreover, all of them had G or C at the third codon position.

To examine whether selective pressure is driving GC evolution in Bangiophyceae, we calculated the evolutionary rate of GC content change based on the 424 single-copy orthologous gene families identified by Orthofinder in Bangiales genomes and the other 15 genomes of photosynthetic eukaryotes. We classified amino acids into three groups based on GC content at the first two sites of a codon as high GC, low GC, or neutral. The substitution rate within groups (d\textsubscript{within}) and between groups (d\textsubscript{between}) in the two clades was compared. In both clades, mean d\textsubscript{within} was greater than mean d\textsubscript{between}, suggesting that amino acid changes trended toward a similar GC content. However, the mean ratio d\textsubscript{between}/d\textsubscript{within} was significantly greater in the comparisons involving the Bangiales clade (0.789 ± 0.003) than those in Florideophyceae clades (0.646 ± 0.002; paired t test; P < 0.01). The different pattern of codon replacement for Bangiales suggested an increased rate of substitution toward GC-rich codons in this clade. For each single copy orthologous gene family in Bangiophyceae and Florideophyceae, the ancestral sequence was constructed through an optimized non-homogeneous model. Their average GC content was estimated as 53.85% (Supplementary Fig. 19a). A significantly greater frequency of A/T to G/C changes was found in Bangiales (126,126) compared with the Florideophyceae clade (27,585; paired t test: P < 0.05) while the frequency of G/C to A/T changes (42,760) was nearly equal in the comparisons involving the Bangiales clade (0.789 ± 0.003) than those in Florideophyceae clades (0.646 ± 0.002; paired t test; P < 0.01).

Discussion

A variety of evidence from physiology, evolutionary biology, paleontology, geosciences, and molecular genetics provides insights into the co-evolution of the Earth’s atmosphere, geosphere, and biology. The advent of oxygenic photosynthesis around 2.4 Ga increased atmospheric and surface ocean O\textsubscript{2} levels.
and reduced atmospheric CO₂36,37. Atmospheric CO₂ concentrations reached a maximum of 0.5% around 500 Mya and then experienced a long downward trend, reaching a level that is unable to sustain Rubisco about ~286 Mya. This long-term decline in CO₂ concentration has shaped the evolutionary history of photosynthetic eukaryotes, resulting in several fundamental adaptations that impact kinetic properties of Rubisco and the CCM38–40. The timescale of Bangiales evolution suggests that its ancestor experienced long-term CO₂ limitation, which may have been a major driving force in this lineage. Bangiales displayed an alternation of two heteromorphic generations in distinctive habitats. The morphological differences and the resulting ecological advantages in their alternative life cycles might be a secondary adaptation to environmental stresses in the evolutionary history of photosynthetic organisms41.

Analyses of the P. yezoensis genome reveal the important role of carbon limitation in Bangiales evolution. The Pyropia lineage has adapted to C₄ limitation through the large-scale duplication of CA genes and enhancement of its antioxidant system with increased gene copy of SOD, LOX, TRX, and TYR. HGTs from bacteria and local gene duplication (indicated by θCA8 and θCA9) partially explain gene expansion in Bangiales. These gene families tend to show phase-specific expression, implying that they may be an adaptive response to the different growth conditions in the heteromorphic life history. For the thallus phase, inhabiting the upper intertidal zone, fluctuating stresses during emersion lead to intracellular over-production of ROS, which would then damage cellular structures and macromolecules42. The expanded gene repertoire of the antioxidative system and their upregulation in the thallus phase strongly support the essential roles of these genes in dealing with environmental stresses during emersion.

Based on the gene expression patterns of CA isoforms and related bicarbonate transporters in the two life history phases, as well as their distinct subcellular localizations, we propose a model for the CCM in Pyropia. Multiple gam-specifically transcribed CAs and related transporters were likely to contribute to carbon acquisition in the thallus. Their potential functions in response to CO₂ fluctuation during emersion in the upper intertidal zone is supported by elevated transcription during dehydration. Moreover, the transcription of CAs is apparently induced at different time points of dehydration as well as rehydration, indicating diverse roles of individual CAs. Genetic manipulation is needed for each CA isoform to determine their physiological function, as well as greater knowledge about their subcellular localization, to confirm these observations.

The conchocelis utilizes the carbon available in mollusk shell by promoting dissolution of CaCO₃. This phenomenon of mineral-sourced autotrophy has been previously documented in the cyanobacterium Mastigocoleus testarum43. In Mastigocoleus, the dissolution of CaCO₃ starts from a passive diffusion of Ca²⁺ into the apical cell. The lowered of Ca²⁺ concentration in the interstitial microenvironment shifts the ion equilibrium toward dissolution. Our observation of increased Ca²⁺ concentration and pH in the external medium in cultures of Pyropia conchocelis, as well as the central role of P-type cation transporters in transporting Ca²⁺ into distal cells, support this model. Moreover, we found that the extracellular CAs, in particular the two specifically transcribed in the conchocelis, play essential roles in this process. Therefore, we speculate that, in addition to the spontaneous Ca²⁺ ionization as proposed in the Mastigocoleus model, CA excreted by the conchocelis cell may promote the dissolution of Ca²⁺ through generating protons that attack CaCO₃ (Fig. 4e), as proposed in some limestone-dissolving bacteria26,27. Two bicarbonates are produced in this reaction, one from CO₂ and one from CaCO₃. These might immediately enter the cell through bicarbonate transporters and their rapid elimination in the microenvironment would push the chemical reaction in this direction. The import of Ca²⁺ is coupled with the excretion of two protons to maintain charge. One proton would combine with ionized CO₂⁻ to form bicarbonate. The fate of bicarbonate could either be entering the cell and being converted to CO₂ by intracellular CA or being converted directly into CO₂ with another excreted proton by the way of extracellular CA. The dissolution of CaCO₃ might also play a role in conchocelis penetration into the shell matrix. Based on our results, the utilization of shell CaCO₃ as a carbon source, with the participation of extracellular CA, P-type ATPase, and the subsequent intracellular CCM components expressed in the conchocelis, is strongly supported. However, the contributions of CaCO₃ to the total C uptake in the conchocelis, the physiological function of CAs, and calcium transporters remain to be determined.

In marine bacteria with divergent GC content, carbon limitation may be an important factor driving GC content evolution because the GC pair requires one less carbon atom than AT33–45. In the evolutionary history of the Bangiales, their common ancestor has experienced long-term carbon limitation resulting from the sharp decline in atmospheric CO₂ in the Permo-Carboniferous era. Therefore, we speculate that, as in marine bacteria, the preference to use carbon-saving GC bases might have been an adaptive strategy responding to ancient carbon limitation. Carbon limitation also has left imprints on amino acid usage in Bangiales. We calculated the average content of each amino acid in the single copy orthologous genes of Bangiales and Florideophyceae genomes. Among the 19 amino acids with carbon in side chains, Gly and Ala containing the least carbon atoms (2 and 3, respectively) were enriched in Bangiales when compared with Florideophyceae (paired t test; P < 0.01), whereas the content of amino acids with 5, 6, and even more carbons in the two clades were equivalent (Supplementary Table 6). The strong bias toward a reduced frequency of carbon-containing amino acids in Bangiales genomes is consistent with (but does not prove) the hypothesis of carbon limitation driving the evolution of Bangiales. It is clear that other factors may impact GC content such as genome size increase driven by transposon expansion (e.g., in grasses46) and high levels of intra-genic recombination (i.e., GC-based gene conversion) that can act independently from selection to create codon usage bias47.

Methods

Genome sequencing and assembly. To avoid contamination of surface bacteria and organelle DNA, we first isolated nuclei from de-contaminated homogenous Pyropia thalli and then extracted genomic DNA to construct sequencing libraries (Supplementary Methods 1–3, Supplementary Figs. 20 and 21). Long read sequencing data from Pac-Bio RSII platform were collected and used in initial assembly via RS_HGAP_Assembly 3 protocol. Mate pair Illumina reads with insert size at 500 bp were also used to generate scaffolds using SPADES48. Gap closing through PBJelly249 and genome polishing through Quiver algorithm from SMRT Analysis v2.3.050 were then done (Supplementary Method 4). Scaffolds in the primary assembly were then subject to a five-step de-contamination pipeline to detect potential bacterial sequences (Supplementary Method 5). Twenty scaffolds were filtered out in this way. To improve the assembly, optical maps of the Blob-Nano system were generated and used for scaffolding, followed by a proximity-guided assembly based on Hi-C data. Bacterial sequences were manually investigated and removed from the final assembly.

For transcriptome sequencing, full-length cDNA sequencing was first done with the Pac-Bio SMART sequencing platform using an RNA mixture from multiple development stages and stress treatments to cover as many transcripts as possible (Supplementary Method 6). Differential gene expression sequencing for gametophyte thalli and sporophyte conchocelis, as well as thalli, under desiccated and re-hydrated osmotic stresses were performed on the Illumina GAIIx platform (Supplementary Methods 7 and 8). The relative transcriptional abundances of each gene model under various conditions were calculated as FPKM values and normalized using the cufflink program51.
Genome annotation. Repeat elements were identified using RepeatModeler and Tandem Repeats Finder. De novo prediction via AUGUSTUS, homology searches, and RNA aiding prediction via PASA followed by EVM integration was used to predict protein-coding gene models from the repeat-masked genome. The gene models were then functionally annotated using the NR, InterPro, GO, KOG, KEGG, CAzyM, and CDD databases (Supplementary Methods 9–11). Subcellular localization of Pyropia proteins were predicted with PredAlgo, which is a multi-subcellular localization prediction tool designed for algae. The trans-membrane helices were predicted using TMHMM (v2.0) at http://www.cbs.dtu.dk/services/TMHMM/.

Evolutionary analysis of nucleotide composition in red algae. We implemented the strategy described in Luo et al. to infer the evolution of nucleotide composition in several red algae genomes. Based on the phylogenetic tree built from single-copy orthologous genes, we first estimated substitution parameters of the alignment sequence with the Bio++ maximum-likelihood algorithm using the non-homogeneous model as implemented in the BppML programs from Bio++ suite to infer ancestral states of a continuous character. Then the ancestral sequences for each node of the tree were computed using the BppAncestor program. Conservative and radical amino acid replacements were defined as amino acid substitutions within and between groups, respectively. Rates of conservation (dc) and radical (dr) replacements, that is, the number of conservative nonsynonymous nucleotide substitutions per conservative nonsynonymous site and the number of radical nonsynonymous nucleotide-substitution per radical nonsynonymous site, between two orthologous genes were calculated using the HON-new software. The dc and dr substitution of each node were computed by classifying amino acids into three groups according to GC content at the first two sites of a codon (most codons encoding one amino acid may differ in the third site): high G+C, low G+C, and neutral.

Gene expanding and contraction analysis. To examine genome divergence among red algae, we carried out a phylogenetic analysis using the P. yezeensis and five other red algal genomes as well as model algal genomes to identify orthologous genes using OrthoFinder. Single-copy orthologous groups were identified and used to construct a phylogenetic tree with MrBayes3.1.2.1. Divergence time was calculated using the Bayesian relaxed molecular clock approach with the divergence time of the Bangiaceae and Florideophyceae set as the standard.2,2,6,2,6 The number of genes in each OG were statistically compared among red algal genomes using the CAFE program (Supplementary Method 12). HGT candidates were inferred following phylogenetic analysis described for the genome of Paullinia chromatophora, with detailed information included in Supplementary Method 13.

Detection of the variation of Ca2+ concentration and pH value. Three conchocelis-dwelling shells with similar sizes were cultured in a sterilized culture medium to inhibit the activity of external CA (99% atom%, Sigma-Aldrich), in 20 ± 1 °C under 50% relative humidity and with a light cycle of 12:12 light:dark cycle (12:12 light:dark cycle) with three rounds of gentle shaking every day. Control samples of equal weight were co-cultured with 12C-CaCO3. Three biological replicates were done for each group. Provasoli’s enrichment medium was refreshed every 5 days. For Ca2+ and pH detection, 30–50 ml medium in the vicinity of shells was aspirated using a syringe with minimal disturbance to the medium on the 0th, 3rd, 6th, 9th, 12th, and 15th day. Ca2+ concentration was tested by an atomic absorption spectrophotometer. pH values were measured at the same time points by a pH meter.

Stable isotope-labeling experiment. For the 13C-labeling experiment, 0.3 g free conchocelis was cultured in 200 ml artificial sea water with 0.3 g 13C-labeled CaCO3 (99 atom%, Sigma-Aldrich), in 20 ± 1 °C under 50–60 μmol photons m−2 s−1 (12:12 light:dark cycle) with three rounds of gentle shaking every day. Control samples of equal weight were co-cultured with 13C-CaCO3. Three biological replicates were done with these experiments. After 20 days of cultivation, the abundance of 13C atom in the organic acids related to the TCA cycle were quantitatively analyzed using gas chromatography–mass spectrometry based on the retention time, electron ionization mass spectrum of corresponding standards, and matching to the NIST library, as described in Mulat et al. Information about the methodology used in transient expression of CA::GFP fusion proteins in tobacco and enzymatic activity assay of extracellular CA in conchocelis are included in Supplementary Methods 14 and 15, respectively.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The datasets generated and analyzed during the current study are available from the corresponding author upon request. Whole-genome sequencing data of P. yezeensis were deposited in NCBI under the BioProject PRJNA589917. Genomic sequencing raw data were deposited under SRR10487098 (Illumina platform), SRR10484745 (Hi-C platform), and SRR10489066–SRR10489010 (Paclío platform). The assembly final is available at DBJ/ENA/GenBank under the accession number WMLA0000000. The version described in this paper is version WMLA010001. The data were deposited under SRR10502194–SRR10502233 (osmotic stresses), SRR10527930–SRR10527973 (two life cycle stages), and SRR10502264–SRR10502266 (full-length cDNA sequencing using the Paclío platform). The source data underlying Fig. 4, d as well as Supplementary Fig. 18 are provided as a Source data file.

Received: 18 December 2019; Accepted: 9 July 2020
Published online: 12 August 2020

References. 1. FAO. FAO Yearbook of Fishery and Aquaculture Statistics. http://www.fao.org/fishery/static/Yearbook/1YB2017_USocard/index.htm (2017). 2. Yang, E. C. et al. Divergence time estimates and the evolution of major lineages in the florideophyte red algae. Sci. Rep. 6, 21361 (2016). 3. Xu, K. P. et al. Divergence time, historical biogeography and evolutionary rate estimation of the order Bangiales (Rhodophyta) inferred from multilocus data. J. Oceanol. Limnol. 36, 870–881 (2018). 4. Huan, L., Wu, G., Gao, S. & Wang, G. Photosynthetic activity and proteinaceous analysis highlights the utilization of atmospheric CO2 by Ulva prolifera. Plant Pho. 592, 1103–1113 (2016). 5. Zou, D. H. & Gao, K. S. Effects of desiccation and CO2 concentrations on emersed photosynthesis in Porphyra haitanensis (Bangiales, Rhodophyta), a species farmed in China. Eur. J. Phyc. 37, 578–592 (2002). 6. Huan, L. et al. Preliminary comparison of atmospheric CO2 enhancement to photosynthesis of Pyropia yezeensis (Bangiales, Rhodophyta) leafy thalli and filamentous thalli. Physol. Pho. 66, 117–126 (2018). 7. Blouin, N. A., Brodie, J. A., Grossman, A. C., Xu, P. & Brawley, S. H. Porphyra: a marine crop shaped by stress. Trends Plant Sci. 16, 29–37 (2011). 8. Jiang, H., Gao, K. & Walter Helbling, E. The conchocelis of Porphyra haitanensis (Rhodophyta) is protected from harmful UV radiation by the covering calcareous matrix. J. Phyc. 45, 1270–1277 (2009). 9. Brawley, S. H. et al. Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). Proc. Natl Acad. Sci. USA 114, E6361–E6370 (2017). 10. Cao, M. et al. A chromosome-level genome assembly of Pyropia haitanensis (Bangiales, Rhodophyta). Mol. Ecol. Resour. 20, 216–227 (2019). 11. Yabu, H. Observation on chromosomes in some species of Prophrya. Bull. Fish. Sci. Hokkaido Univ. 19, 239–249 (1969). 12. Lee, J. et al. Analysis of the draft genome of the red seaweed Gracillariopsis chorda provides insights into genome size evolution in Rhodophyta. Mol. Biol. Evol. 35, 1869–1886 (2018). 13. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriştenváňa, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212 (2015). 14. Gee, C. W. & Niyogi, K. K. The carbonic anhydrate CAH1 is an essential component of the carbon-concentrating mechanism in Nannochloropsis oceanica. Proc. Natl Acad. Sci. USA 114, 4537–4542 (2017). 15. DiMario, R. J., Machingura, M. C., Waldrop, G. L. & Moroney, J. V. The many types of carbonic anhydrases in photosynthetic organisms. Plant Sci. 268, 11–17 (2018). 16. Hopkinson, B. M., Dupont, C. L. & Matsuda, Y. The physiology and genetics of CO2 concentrating mechanisms in model diatoms. Curr. Opin. Plant Biol. 35, 51–57 (2016). 17. Tomar, V., Sidhu, G. K., Nogia, P., Mehrotra, R. & Mehrotra, S. Regulatory components of carbon concentrating mechanisms in aquatic unicellular photosynthetic organisms. Plant Cell Rep. 36, 1671–1688 (2017). 18. Mackinder, L. C. M. et al. A spatial interactome reveals the protein organization of the algal CO2-concentrating mechanism. Cell 171, 133–147 (2017). 19. Mercado, J. M., Niell, F. X. & Figueroa, F. L. Regulation of the mechanism for HCO3- use by the inorganic carbon level in Porphyra leucosticta Thur. in Le Jolis (Rhodophyta). Planta 201, 319–325 (1997). 20. Li, X. S., Xu, J. T. & He, P. M. Comparative research on inorganic carbon acquisition by the macroalgae Ulva prolifera (Chlorophyta) and Pyropia yezeensis (Rhodophyta). J. Appl. Phyc. 28, 491–497 (2016). 21. Cuesta-Seijo, J. A. et al. Structure of a dimeric fungal alpha-type carbonic anhydrase. FEBS Lett. 585, 1042–1048 (2011).
22. Benloch, R. et al. Crystal structure and functional characterization of photosystem II-associated carboxylic anhydride CAH3 in Chlamydomonas reinhardtii. Plant Physiol. 167, 950–962 (2015).

23. De Simone, G. & Supuran, C. T. Carboxylic anhydride IX: biochemical and crystallographic characterization of a novel antitumor target. Biochim. Biophys. Acta 1804, 404–409 (2010).

24. Aggarwal, M., Chua, T. K., Pinard, M. A., Szepenyi, D. M. & McKenna, R. Carbon dioxide “trapped” in a β-carboxylic anhydride. Biochemistry 54, 6631–6638 (2015).

25. Giordano, M. & Maberly, S. C. Distribution of carbonic anhydrase in British marine macroalgae. Oecologia 81, 534–539 (1989).

26. Li, W., Yu, L. J., Wu, Y., Jia, L. P. & Yuan, D. X. Enhancement of Ca2+ release from limestone by microbial extracellular carboxylic anhydride. Bioresource. Technol. 98, 1953–1958 (2007).

27. Li, W. et al. Limestone dissolution induced by fungal mycelia, acidic materials, and carboxylic anhydride from fungi. Mycopathologia 167, 37–46 (2009).

28. Guida, B. S., Bose, M. & Garcia-Pichel, F. Carbon fixation from mineral carbonates. Nat. Commun. 8, 1025 (2017).

29. Schonknecht, G. et al. Gene transfer from bacteria and archaea facilitated evolution of an extremophlic eukaryote. Science 339, 1207–1210 (2013).

30. Bhattacharya, D. et al. When less is more: red algae as models for studying gene loss and genome evolution in eukaryotes. Crit. Rev. Plant Sci. 37, 81–99 (2018).

31. Teng, L. H. et al. Evolution and expansion of the prokaryote-like lipoxygenase family in the brown alga Saccarina japonica. Front. Plant. Sci. 8, 2018 (2017).

32. Cao, M. et al. Integrating transcriptomics and metabolomics to characterize homeostasis in Chlamydomonas reinhardtii. ISME J. 12, 1180–1187 (2018).

33. English, A. C. et al. Mind the gap: upgrading genomes with pacemaker transgenes in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 493–507 (2012).

34. Young, J. N., Rickaby, R. E. M., Kapralov, M. V. & Filatov, D. A. Adaptive Young, J. N. & Hopkinson, B. M. The potential for co-evolution of CO2-assoc- ionases and carbon oxidation cycles. Proc. Natl Acad. Sci. USA 110, 5247–5252 (2013).

35. Bhattacharya, D. et al. The genome of red algae Porphyridium purpureum. Nat. Commun. 4, 1941 (2013).

36. Matsuoki, M. et al. Genome sequence of the ultrasmall unicellular red alga Cyanidioidochytrium marsei 103D. Nature 428, 653–657 (2004).

37. Rye, R., Kuo, P. H. & Holland, H. D. Atmospheric carbon-dioxide concentrations before 2.2-billion years ago. Nature 378, 603–605 (1995).

38. Watanuki, K. Evolution of the Earth’s atmosphere; variation of CO2 concentration and climatic changes. Energy 22, 109–114 (1997).

39. Raven, J. A., Giordano, M., Beardall, J. & Maberly, S. C. Algal evolution in relation to atmospheric CO2: carbonylases, carbon-concentrating mechanisms and carbon oxidation cycles. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 493–507 (2012).

40. Young, J. N., Rickaby, R. E. M., Kapralov, M. V. & Filatov, D. A. Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 483–492 (2012).

41. Young, J. N. & Hopkinson, B. M. The potential for co-evolution of CO2-concentrating mechanisms and Rubisco in diatoms. J. Exp. Bot. 68, 3751–3762 (2017).

42. Couceiro, L. et al. Evolution and maintenance of haploid-diploid life cycles in natural populations: the case of the marine brown alga Ectocarpus. Evolution 69, 1808–1822 (2015).

43. Wang, W. L. et al. Regulatory mechanisms underlying the maintenance of homeostasis in Pyropia haitanensis under hypersaline stress conditions. Sci. Total Environ. 662, 168–179 (2019).

44. Luo, H. W., Swan, B. K., Stepanauskas, R., Hughes, A. L. & Moran, M. A. Evolutionary analysis of a streamlined line of surface ocean Rosebacter. ISME J. 8, 1428–1439 (2014).

45. Luo, H. W., Thompson, L. R., Stingl, U. & Hughes, A. L. Selection maintains low genomic GC content in marine SAR11 lineages. Mol. Biol. Evol. 32, 2738–2748 (2015).

46. Hellweger, F. L., Huang, Y. J. & Luo, H. W. Carbon limitation drives GC content evolution of a marine bacterium in an individual-based genome-scale conversion hypothesis expands. ISME J. 12, 1180–1187 (2018).

47. Smarda, P. et al. Ecological and evolutionary significance of genomic GC content diversity in monocytes. Proc. Natl Acad. Sci. USA 111, E4096–E4102 (2014).

48. Lassalle, F. et al. GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. PLoS Genet. 11, e1004941 (2015).

49. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SPSPACE. Bioinformatics 27, 578–579 (2011).

50. English, A. C. et al. Mind the gap: upgrading genomes with pacemaker transgenes in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, e47788 (2012).

51. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515 (2010).

52. Hubley, R. & Smit, A. RepeatModeler Open-10. http://www.repeatmasker.org (2008).

53. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580 (1999).

54. Stankiewicz, P. et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34, W435–W439 (2006).

55. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9, R7 (2008).

56. Tardif, M. et al. PredAlgo: a new subcellular localization prediction tool Technical 98, 953–958 (2007).

57. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. P. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 587–590 (2001).

58. Gueguen, L. et al. Bio plus plus: efficient extensible libraries and tools for computational molecular evolution. Mol. Biol. Evol. 30, 1745–1750 (2013).

59. Zhang, J. Z. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. 50, 56–68 (2000).

60. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157 (2015).

61. Makarova, K. S. et al. MvpaSearch v2.0, efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542 (2012).

62. Herron, M. D., Hackett, J. D., Aylward, F. O. & Michod, R. E. Triassic origin and early radiation of multicellular volvocine algae. Proc. Natl Acad. Sci. USA 106, 3254–3258 (2009).

63. Tamura, K. et al. Estimating divergence times in large molecular phylogenies. Proc. Natl Acad. Sci. USA 109, 1180–1183 (2012).

64. De Bie, T., Cristiani, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22, 1297–1297 (2006).

65. Nowack, E. C. 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 (2016).

66. Mulat, D. G. & Feilberg, A. GC/MS method for determining carbon isotope enrichment and concentration of underivatized short-chain fatty acids by direct aqueous solution injection of biogas digestor samples. Talanta 143, 56–63 (2015).
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-17689-1.

Correspondence and requests for materials should be addressed to D.B. or Y.M.

Peer review information Nature Communications thanks Mathieu Pernice, John Raven, and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

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) 2020