


RESEARCH ARTICLE

Open Access



The Antarctic sea ice alga *Chlamydomonas* sp. ICE-L provides insights into adaptive patterns of chloroplast evolution

Zhenhua Zhang¹, Meiling An², Jinlai Miao^{2,3}, Zhiqiang Gu¹, Chang Liu⁴ and Bojian Zhong^{1*} 

Abstract

Background: The ice alga *Chlamydomonas* sp. ICE-L is the main contributor to primary productivity in Antarctic sea ice ecosystems and is well adapted to the extremely harsh environment. However, the adaptive mechanism of *Chlamydomonas* sp. ICE-L to sea-ice environment remains unclear. To study the adaptive strategies in *Chlamydomonas* sp. ICE-L, we investigated the molecular evolution of chloroplast photosynthetic genes that are essential for the accumulation of carbohydrate and energy living in Antarctic sea ice.

Results: The 60 chloroplast protein-coding genes of *Chlamydomonas* sp. ICE-L were obtained, and the branch-site test detected significant signatures of positive selection on *atpB*, *psaB*, and *rbcl* genes in *Chlamydomonas* sp. ICE-L associated with the photosynthetic machinery. These positively selected genes were further identified as being under convergent evolution between *Chlamydomonas* sp. ICE-L and the halotolerant alga *Dunaliella salina*.

Conclusions: Our study provides evidence that the phototrophic component of *Chlamydomonas* sp. ICE-L exhibits adaptive evolution under extreme environment. The positive Darwinian selection operates on the chloroplast protein-coding genes of Antarctic ice algae adapted to extreme environment following functional-specific and lineages-specific patterns. In addition, three positively selected genes with convergent substitutions in *Chlamydomonas* sp. ICE-L were identified, and the adaptive modifications in these genes were in functionally important regions of the proteins. Our study provides a foundation for future experiments on the biochemical and physiological impacts of photosynthetic genes in green algae.

Keywords: Antarctic ice algae, Positive selection, Photosynthesis, Convergent evolution

Background

The Antarctic area is a particularly inclement environment with extremely low temperatures. Seasonal fluctuations in temperatures from winter to summer (from -80 °C to 5 °C), high UV-B radiation and alternating long periods of sunlight and darkness are major challenges for organisms to survive in Antarctica [1]. The Antarctic sea ice is an important constituent of the polar extreme environments, and it comprises a system of concentrated brine channels under low temperature, high salinity and low light conditions [2]. Survival in this harsh environment demands complex adaptations in physiology and metabolism. Antarctic ice algae thrive in the sea ice, and their prolific

growth makes them the main contributors to primary productivity in sea ice ecosystems [2, 3]. It has been reported that Antarctic ice algae form a series of physiological and biochemical mechanisms to adapt to the extreme environments [1, 4–7]. The rapid growth of molecular data provides us with a new approach to study adaptive mechanisms. Uncovering adaptive strategies at the molecular level can reveal evolutionary patterns and provide novel insights into adaptability of algae to extreme environments.

Green algae can convert inorganic carbon into organic compounds by photosynthesis in chloroplasts, and photosynthetic systems are sensitive to abiotic environmental stresses (low light, low temperature and high salinity). The chloroplast is the photosynthetic reaction center, and the chloroplast genomes are useful resources for evolutionary studies on photosynthesis [8, 9]. This maternally inherited,

* Correspondence: bjzhong@gmail.com

¹Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, China
Full list of author information is available at the end of the article



non-recombining chloroplast genome encodes many important subunits of photosynthetic systems (some subunits are encoded by the nuclear genes), and the chloroplast genome sizes vary among different green algae [10, 11]. Comparative genomic analyses of three model organisms (*Arabidopsis thaliana*, *Chlamydomonas reinhardtii* and *Thermosynechococcus elongatus*) indicate that 48 chloroplast genes fall into two broad groups: 29 genes for the photosynthetic apparatus; 19 genes for components of the chloroplast genetic system. These 29 chloroplast genes encode five photosynthetic complexes: photosystem I (PSI), photosystem II (PSII), Cytochrome *b₆f* (Cytb₆f), ATP synthesis and Rubisco. The protein complexes PSI, PSII and Cytochrome *b₆f* (Cytb₆f) are membrane-intrinsic. These complexes form the photosynthetic electron transport chain which can transfer electron from PSII to oxidized PSI reaction centers where NADP⁺ is reduced to NADPH. Both PSI and PSII are stress sensitive, and genes encoding the two complexes can respond to abiotic stress at transcriptional and post-transcriptional levels [12, 13]. During electron transfer, proton gradients are built and utilized by chloroplastic ATP synthase. The ATP is generated that can be used in the process of photosynthetic fixation of CO₂. Rubisco is extrinsic to the membrane and catalyzes the first step of the reductive pentose phosphate pathway of carbon dioxide assimilation. High concentration of NaCl and low temperature may limit the activity of Rubisco, which would inhibit the fixation of CO₂ [14, 15]. The *rbcL* gene encoding Rubisco in many plants adapted to extreme environments shows various amino acid substitutions in Rubisco that influence abiotic stress responses [14].

Antarctic plants have evolved a number of adaptive molecular mechanisms in five complexes of the photosynthetic system to survive in the harsh environment [16–18]. It is reported that *Colobanthus quitensis* (angiosperm) living in Antarctica has adapted to maintain faster repair rate of the photosynthetic system [19]. The Antarctic macroalgae (e.g. *Himantothallus grandifolius* and *Desmarestia anceps*) evolved high photosynthetic efficiency, low light requirements for photosynthesis and notable UV tolerance [20]. Interestingly, the Antarctic ice diatom *Amphiprora kufferathii* utilizes epiphytic bacteria to consume the reactive oxygen species (ROS) produced during photosynthesis [21]. The increasing molecular data of Antarctic plants will help to investigate the adaptation of photosynthesis to the extreme environment of Antarctica from an evolutionary perspective. It has been reported that positively selected and rapidly evolving genes in animals contributed to extreme environment adaptations [22–24]. In plants, the evidence of positive selection in the Rubisco gene (involved in carbon fixation) from mosses has been associated with its adaptation to the declining levels of atmospheric CO₂ since their origination in the Ordovician [25]. The genes related

to the photosynthetic machinery in an endolithic green alga *Ostreobium quekettii* show strong purifying selection due to its low light lifestyle [26]. The genomic analyses of salt-tolerant *Populus euphratica* demonstrate rapid evolution in genes encoding photosynthetic electron transport chain [27].

The green alga *Chlamydomonas* sp. ICE-L is isolated from the floating sea ice with an optimum growth temperature range of 4–10 °C. As the primary productivity of the Antarctic sea ice, *Chlamydomonas* sp. ICE-L evolved specific morphological characteristics as adaptation to the extreme environment, such as the changes in pigments, lipids and fatty acids content for maintaining the stability of the thylakoid membranes and the normal physiological function of the chloroplast [7]. The particularly harsh environment in Antarctica may leave footprints in the photosynthetic genes of *Chlamydomonas* sp. ICE-L by changing rates of molecular evolution. However, there is little genetic evidence on the adaptation to the extreme environment regarding photosynthesis of this alga.

There are other green algae that potentially adapt to similar environmental stress. *Coccomyxa subellipsoidea* C-169 is a small elongated non-motile unicellular green alga isolated in the polar summer of 1959/60 at Marble Point, where there were large moist areas covered with a thick mat of the thalloid *Nostoc commune* [28], and the major environmental factors of *Coccomyxa subellipsoidea* C-169 are low light and low temperature. *Chlorella* sp. ArM0029B is isolated from drift ice in the Arctic region [29], and low temperature is regarded as the main environmental stress. *Dunaliella salina* is a halotolerant alga thriving in extreme saline environments [30], and high salinity is the abiotic stress for *Dunaliella salina*. These green algae as well as *Chlamydomonas* sp. ICE-L living in extreme environments are ideal organisms to comparatively investigate adaptive strategies in different abiotic stresses.

In this study, we sequenced 60 chloroplast protein-coding genes of *Chlamydomonas* sp. ICE-L, and used other available chloroplast genomes of aforementioned green algae living in similar extreme environments (low light, low temperature and high salinity) to investigate the potential genetic basis of adaptation to abiotic stress in the photosynthetic machinery of *Chlamydomonas* sp. ICE-L. Adaptive evolution appears to target the protein-coding components of the chloroplast in a function (photosynthetic components) and lineage specific manner, and the adaptive modifications in these positively selected genes were in functionally important regions. Our analyses revealed signatures of positive selection and convergent evolution among the chloroplast protein-coding genes of Antarctic ice algae, supporting the notion that the adaptation to extreme environments in algae are associated with the altered patterns of selection on chloroplast proteins.

Methods

Algal samples and DNA sequencing

The strain of *Chlamydomonas* sp. ICE-L was obtained from the floating sea ice near the Zhongshan Station of Antarctica (China) during the Chinese 18th Antarctic Science Exploration from 2001 to 2002. The *Chlamydomonas* sp. ICE-L was isolated from Antarctic sea ice and monoclonal cultured in the lab. Cells were grown in Provasoli seawater medium [31] under a photon flux density of 40 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$, 14 L: 10D cycle, and temperature of 8 °C. Total genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN Bio-Tec), and sequenced using Illumina HiSeq™ 2000 platform.

Chloroplast genes assembly and annotation

A total of 6130 Mb clean reads (125-bp paired-end) were generated by removing adapters, poly-N and low quality reads using in-house perl scripts. All downstream analyses were based on the clean reads. We performed de novo assembly using SPAdes software [32] with the parameter “-plasmid”, “-careful” and other default settings. We created a local database of published chloroplast genomes of green algae, and performed similarity searches to identify the scaffolds from the chloroplast genome using BLAST with a cut-off E-value of 1×10^{-10} [33]. We further used Geneious (version 9.0.4, <http://www.geneious.com/>) to annotate the chloroplast genes, and confirmed the gene annotations using the DOGMA [34] online tools (<http://dogma.ccbb.utexas.edu/>).

DNA alignment and phylogenetic inference

We generated multiple sequence alignments using two different programs: MUSCLE [35] and PRANK [36]. We aligned 60 chloroplast protein-coding genes using six Chlorophyceae and seven Trebouxiophyceae green algae (Table 1).

Nucleotide sequences were aligned at the codon level with the option “-codon” using MUSCLE and at the protein level with the option “-translate” using PRANK. Stop codons were removed from the sequences prior to alignment. Each aligned nucleotide sequence was trimmed to exclude poorly aligned positions using Gblocks0.91b with default parameter [37], and concatenated into a single data set using Geneious. Two alignments were almost identical except for a few sites. We used the aligned data produced by PRANK in the subsequent analyses as PRANK has lower false-positive rates [38]. The *rpoB1*, *rpoB2* and *rpoC* genes were excluded due to alignment ambiguities according to visual assessment.

The phylogenetic trees were reconstructed based on the nucleotide data with GTRGAMMA model and amino acid data with PROTGAMMAAUTO model using RAXML v8.2 [39] with a rapid bootstrap search of 1000 replicates. The 3rd codon positions are problematic to phylogenetic inference because of high saturation [40], then only 1st and 2nd codon positions of nucleotide data are used for phylogenetic analyses to minimize negative effects of saturation [41].

Identification of genes under positive selection

Branch-specific estimates of rates of synonymous (dS) and non-synonymous substitutions (dN) were calculated under the basic model (model = 0, Nsites = 0, which assumes no site-wise or branch-wise dN/dS variation). We discarded 11 genes with high dS value to avoid misestimation of dN/dS. The dS values of remaining 46 genes were below 1 for nearly all branches, ensuring sufficient phylogenetic resolution to accurately estimate dN/dS across the phylogeny [42] (Additional file 1: Table S1). Alignment gaps and uncertainties were deleted to avoid false positives [43]. The codon frequencies were determined by the

Table 1 The 13 green algae used in this study

Order	Taxa	Accession number
Chlorophyceae	<i>Dunaliella salina</i>	[GenBank: NC_016732]
	<i>Chlamydomonas applanata</i>	[GenBank: KT625417]
	<i>Microglena monadina</i>	[GenBank: KT624717 - KT624805]
	<i>Chlamydomonas</i> sp. ICE-L	[GenBank: MF280291 - MF280350]
	<i>Chlamydomonas reinhardtii</i>	[GenBank: NC_005353]
	<i>Volvox carteri</i> f. nagariensis	[GenBank: GU084820]
Trebouxiophyceae	<i>Paradoxia multiseta</i>	[GenBank: KM462879]
	<i>Choricystis parasitica</i>	[GenBank: KM462878]
	<i>Coccomyxa subellipsoidea</i>	[GenBank: NC_015084]
	<i>Dicloster acuatus</i>	[GenBank: NC_025546]
	<i>Parachlorella kessleri</i>	[GenBank: NC_012978]
	<i>Chlorella</i> sp. ArM0029B	[GenBank: KF554427]
	<i>Chlorella vulgaris</i>	[GenBank: NC_001865]

F3 × 4 model. The remaining 46 genes were divided into two functional classifications (photosynthesis and genetic system), and were separately concatenated. Three different codon-based likelihood models (branch-specific, site-specific and branch-site models) were used to explore the selection patterns and identify positive selection on each of the 46 protein-coding genes and two concatenated data. The branch and branch-site model tests need to assign foreground branch based on a priori knowledge, and incorrect assignments may compromise the power of the test. The goal of our study is to explore the role of positive selection in the adaptive patterns of Antarctic sea ice algae, thus four selected algae adapted to extreme environments are used to perform the selection analyses.

To detect adaptive evolving genes in *Chlamydomonas* sp. ICE-L, we first conducted the branch model analyses using Codeml in PAML. The non-synonymous to synonymous rate ratio ω (dN/dS) among branches indicates changes in patterns of natural selection, where $\omega = 1$, $\omega < 1$ and $\omega > 1$ correspond to neutral evolution, purifying and positive selection, respectively [44]. We used three starting ω values (0.5, 1 and 2) to avoid potential local optima. The null model specified one ω for the entire phylogenetic tree, whereas the alternative model allowed two different ω values for *Chlamydomonas* sp. ICE-L and background lineages. Furthermore, the other three green algae (*Coccomyxa subellipsoidea* C-169, *Chlorella* sp. ArM0029B and *Dunaliella salina*) with extreme living conditions were specified as foreground and repeated branch-model test. The likelihood ratio test (LRT) with a χ^2 distribution was used to determine which models were statistically different from the null model at a threshold of $P < 0.05$. We applied the FDR correction to the P values for the multiple tests performed with a significance level of 0.05 [45].

The site-specific model assumes that selection pattern varies among sites in the alignment but not among branches in the phylogeny. We used a pair of site model comparisons to test for positive selection. The alternative models allow for site-specific positive selection, but the null models do not (M8 vs. null M8a) [44]. LRT statistics were compared to a χ^2 null distribution with the corresponding degrees of freedom, and the FDR correction was applied to the P values calculated above to account for the multiple tests performed with a significance level of 0.05 [45].

To detect evidence of positive selection on specific sites along specific lineage, the improved branch-site model (model = 2, Nsites = 2, fixed omega = 0, omega = 2) implemented in the Codeml program [44] were compared with the null model (model = 2, Nsites = 2, fixed omega = 1, omega = 1). The alternative model assumes four categories of sites: the first two assume sites under purifying selection ($0 \leq \omega_0 \leq 1$) or neutrality ($\omega_1 = 1$) across the phylogeny, whereas the remaining two assume sites under either purifying selection or neutrality along the background but with

positive selection ($\omega_2 > 1$) in foreground. The null model fixed the $\omega_2 = 1$ in foreground. Because branch-site models do not allow for multiple foreground branches, the branch leading to *Chlamydomonas* sp. ICE-L, *Coccomyxa subellipsoidea* C-169, *Chlorella* sp. ArM0029B and *Dunaliella salina* was separately chosen as foreground branch in each analysis. The likelihood ratio test (LRT) with a χ^2 distribution was used to determine which models were statistically different from the null model at a threshold of $P < 0.05$. We applied the FDR correction to the P values calculated above to account for the multiple tests performed with a significance level of 0.05 [45]. Bayes empirical Bayes (BEB) method was used to statistically identify sites under positive selection with posterior probabilities ≥ 0.95 [46].

Convergent and parallel evolution analyses

To determine whether similar patterns of adaptive evolution occurred in distant lineages habitually exposed to similar environmental factors, we performed convergent evolution analyses. Convergent sites include both “parallel” and “convergent” sites as defined by J. Zhang and S. Kumar [47]. We reestimated the branch length and reconstructed the ancestral amino acid sequences using the Codeml program in PAML packages [44]. We identified convergent and parallel amino acid substitutions between each pair of species with the following criteria: (1) amino acid residues of both extant species in one pair were identical; (2) amino acid change occurred between the extant lineages and their most recent common ancestor. We then performed a Poisson test to verify whether the observed number of convergent sites of each gene was significantly more than the expected number caused by random substitution under the gcpREV- f_{gene} amino acid substitution models [48, 49]. It has been reported that gcpREV- f_{gene} model performs well using small number of sequences for convergent analyses, although it could cause false positive detection of excessive amount of convergence [49]. To alleviate the possible false positive detection, we used a stringent criterion that the positively selected gene with nonrandom convergent amino acid substitutions is a conservative signature of adaptive evolution [50, 51]. We also applied the FDR correction to the P values calculated above to account for the multiple tests performed with a significance level of 0.05 [45].

Structural analyses of photosynthetic genes under adaptive evolution

In order to gain more insights into the influence of non-random convergent substitutions and positively selected sites on the structure and function of the chloroplast proteins, we performed structural analyses using available X-ray crystal structure data. The secondary structures of chloroplast genes with conservative signatures

of adaptive evolution in the *Chlamydomonas* sp. ICE-L were predicted using Phyre2 server. Phyre2 is an upgrade to the original Protein Homology/analogy Recognition Engine [52]. The three-dimensional (3D) structures were predicted using the homology modeling software provided by the I-TASSER server [53]. The functional information were derived from the Uniprot (<http://www.uniprot.org/>). We aligned genes with its homologous protein in *Chlamydomonas reinhardtii* and *Arabidopsis thaliana*. The non-random convergent substitutions and positively selected sites were mapped on the predicted secondary and 3D structures to visualize the functional locations using ePlant Web server [54].

Results

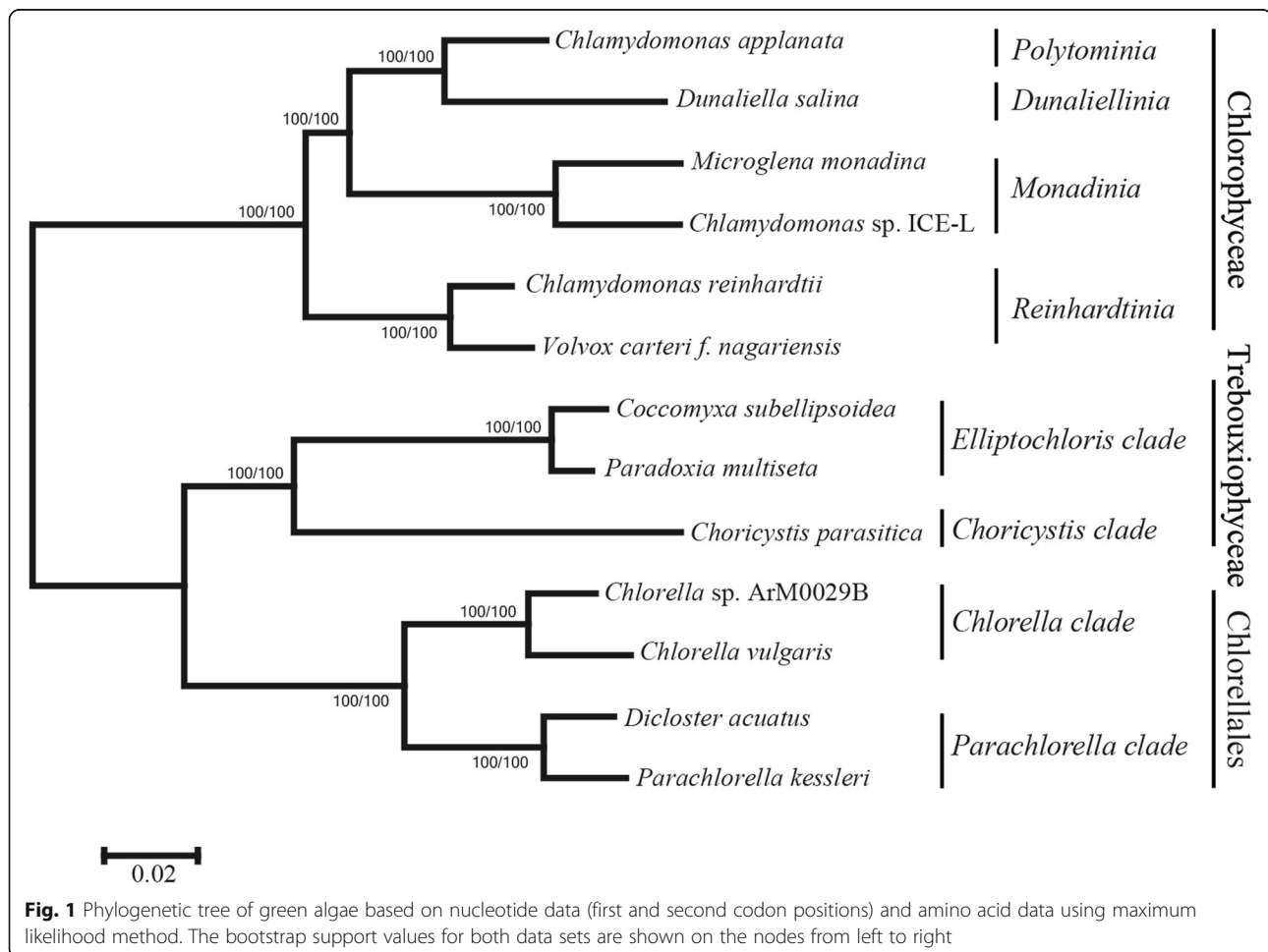
Positive selection on photosynthetic chloroplast genes

The maximum-likelihood trees confirm that Chlamydomonadales (Chlorophyceae) comprise multiple clades [55, 56], and *Chlamydomonas* sp. ICE-L is a sister group to *Microglena monadina* with high bootstrap support

(100%) based on 1st + 2nd codon positions and amino acid data (Fig. 1). The divergence time analyses demonstrate that *Chlamydomonas* sp. ICE-L originated in Permian period (240.7Mya; 95% confidence interval: 138.7~379.9Mya) (Additional file 1: Figure S1).

In order to test whether the evolutionary patterns of chloroplast protein-coding genes in *Chlamydomonas* sp. ICE-L are related to extreme environmental adaptation, four algae living in extreme environments were selected for comparative evolutionary analyses. We first tested the basic model (invariant ω across both sites and branches), and found that gene-specific estimates of ω range from 0.00096 to 0.057. All 46 gene-specific ω values were less than 1 (Additional file 1: Table S1). The average of all 46 gene-specific ω values was 0.02999. Only 9 out of 28 photosynthetic genes and 12 out of 18 genetic system genes have larger ω values than the average, respectively.

The branch model test was used to detect positive selection using gene-specific data in *Chlamydomonas* sp. ICE-L. There was no evidence of positive selection in branch model analyses (corrected P -value =1 and $\omega < 1$;



Additional file 1: Table S2), indicating purifying selection is dominant in chloroplast protein-coding genes. Similarly, there was no evidence of positive selection considering the other three green algae (*Coccomyxa subellipsoidea* C-169, *Chlorella* sp. ArM0029B and *Dunaliella salina*) living in extreme environments as foreground (Additional file 1: Table S2). When using two concatenated data based on two functional classifications (photosynthesis and genetic system), in no case did the test reveal evidence for positive selection in any of the four algae (*Chlamydomonas* sp. ICE-L, *Coccomyxa subellipsoidea* C-169, *Chlorella* sp. ArM0029B and *Dunaliella salina*) (Additional file 1: Table S2).

The random-site model, which ignores ω variation among lineages, was used to identify whether some sites were the targets of positive selection. To test site-specific positive selection in each gene, we used a pair of site models (M8a vs. M8) and FDR correction to the multiple LRT comparisons (corrected P -value < 0.05). There was no specific site identified under positive selection using

the Bayes empirical Bayes (BEB) method, confirming that purifying selection is predominant force shaping the evolution of chloroplast genes of green algae (Additional file 1: Table S3). Moreover, the examination of site-specific posterior mean ω across two concatenated data showed that ω values were nearly zero for almost all sites. There are a few of sites with ω value more than 0.5 (Additional file 1: Figure S2) implying possible weak purifying selection or episodic positive selection operating at these sites [42].

Branch-site analyses of gene-specific data identified the evidence of positive selection along the lineages leading to *Chlamydomonas* sp. ICE-L, *Dunaliella salina* and *Chlorella* sp. ArM0029B with posterior probabilities ≥ 0.95 using a Bayes empirical Bayes (BEB) method (Table 2) [46]. BEB analyses identified positively selected genes: three (*atpB*, *psaB*, and *rbcL*) in *Chlamydomonas* sp. ICE-L, three (*atpA*, *atpB*, and *psbC*) in *Dunaliella salina* and two (*atpE* and *petA*) in *Chlorella* sp. ArM0029B. The LRT

Table 2 Positively selected sites in corresponding genes of *Chlamydomonas* sp. ICE-L, *Dunaliella salina* and *Chlorella* sp. ArM0029B

Gene	Branch-site model	-lnL	$2\delta(\lnL)$	Q Value	ω Values	Positively selected sites
<i>atpA</i>	Branch (<i>Dunaliella salina</i>)					
	Null	9513.725			$\omega_0 = 0.026, \omega_1 = 1, \omega_2 = 1$	
	Alternative	9503.900	19.650	4.28E-04	$\omega_0 = 0.026, \omega_1 = 1, \omega_2 = 49.199$	259–0.994, 269–0.951, 373–0.977, 382–0.996, 426–0.969, 429–0.978, 432–0.992, 434–0.998, 450–0.952, 455–0.978
<i>atpB</i>	Branch (<i>Chlamydomonas</i> sp. ICE-L)					
	Null	8590.140			$\omega_0 = 0.023, \omega_1 = 1, \omega_2 = 1$	
	Alternative	8567.141	45.998	5.43E-10	$\omega_0 = 0.022, \omega_1 = 1, \omega_2 = 999$	24–1.000, 107–0.978, 109–0.999, 375–0.998,
	Branch (<i>Dunaliella salina</i>)					
	Null	8566.141			$\omega_0 = 0.021, \omega_1 = 1, \omega_2 = 1$	
	Alternative	8558.602	15.079	2.37E-03	$\omega_0 = 0.021, \omega_1 = 1, \omega_2 = 57.824$	112–0.975, 297–0.970, 305–0.970, 331–0.966, 371–0.977
<i>atpE</i>	Branch (<i>Chlorella</i> sp. ArM0029B)					
	Null	2572.548			$\omega_0 = 0.030, \omega_1 = 1, \omega_2 = 1$	
	Alternative	2568.079	8.936	4.29E-02	$\omega_0 = 0.030, \omega_1 = 1, \omega_2 = 694.685$	100–0.987
<i>petA</i>	Branch (<i>Chlorella</i> sp. ArM0029B)					
	Null	5724.705			$\omega_0 = 0.042, \omega_1 = 1, \omega_2 = 1$	
	Alternative	5720.106	9.198	4.29E-02	$\omega_0 = 0.042, \omega_1 = 1, \omega_2 = 999$	216–0.984, 243–0.986
<i>psaB</i>	Branch (<i>Chlamydomonas</i> sp. ICE-L)					
	Null	12,670.404			$\omega_0 = 0.015, \omega_1 = 1, \omega_2 = 1$	
	Alternative	12,665.032	10.746	1.60E-02	$\omega_0 = 0.016, \omega_1 = 1, \omega_2 = 998.999$	220–0.972, 356–0.985, 646–0.990
<i>psbC</i>	Branch (<i>Dunaliella salina</i>)					
	Null	7158.342			$\omega_0 = 0.019, \omega_1 = 1, \omega_2 = 1$	
	Alternative	7151.779	13.126	4.47E-03	$\omega_0 = 0.019, \omega_1 = 1, \omega_2 = 999$	207–0.958, 355–0.988, 360–0.980, 410–0.992
<i>rbcL</i>	Branch (<i>Chlamydomonas</i> sp. ICE-L)					
	Null	6852.947			$\omega_0 = 0.016, \omega_1 = 1, \omega_2 = 1$	
	Alternative	6843.536	18.822	3.31E-04	$\omega_0 = 0.017, \omega_1 = 1, \omega_2 = 365.261$	372–0.969, 450–0.999, 457–0.994

The number for amino acid residues identified by Bayes empirical bayes (BEB) analyses corresponds to their alignment positions. Number behind hyphen is the posterior probability (PP) under BEB analysis

of branch-site model in three genes (*atpB*, *psbC* and *psbD*) in *Coccomyxa subellipsoidea* C-169 were statistically significant after correction, but there was no positively selected site identified in BEB analyses ($PP > 0.95$). The same result was obtained in *atpB* gene in *Chlorella* sp. ArM0029B. A total of 10 and 19 positively selected sites were separately identified in *Chlamydomonas* sp. ICE-L and *Dunaliella salina*, whereas only 3 positively selected sites in *Chlorella* sp. ArM0029B. More importantly, all positively selected genes were photosynthetic and encoded vital photosynthetic proteins which are sensitive to abiotic stresses.

The branch-site analyses of two concatenated data identified signatures of positive selection in seven genes (*atpB*, *psaA*, *psaB*, *psbA*, *psbD*, *rbcl* and *rpl2*; 10 sites in total) in *Chlamydomonas* sp. ICE-L, 12 genes (*atpA*, *atpB*, *petA*, *psaA*, *psbC*, *psbD*, *ccsA*, *chlN*, *rpl5*, *rpl23*, *rps4* and *ycf12*; 30 sites in total) in *Dunaliella salina* (Additional file 1: Table S4). Although the amounts of positively selected genes based on concatenated data are different from that of gene-specific data, it consistently showed that most positively selected sites (9/10 and 23/30) were within photosynthetic genes in *Chlamydomonas* sp. ICE-L and *Dunaliella salina*, respectively. There was no evidence of positive selection for *Coccomyxa subellipsoidea* C-169 and *Chlorella* sp. ArM0029B. It is presumable that these discrepancies based on gene-specific and concatenated data reflect a trade-off between the benefit of increased data and the cost of shared parameter estimates across genes in the large concatenated data compared to the small gene-specific data. It is undeniable that the concatenated data analyses confirmed the results from gene-specific analyses.

In summary, positive selection analyses using both gene-specific and concatenated data revealed putatively adaptive evolutionary patterns of chloroplast protein-coding genes in green algae living in extreme environments: they show in a functional-specific and lineage-specific manner that positive selection mainly target the photosynthetic genes in *Chlamydomonas* sp. ICE-L and *Dunaliella salina*. These genes under positive selection were used in subsequent

convergent analyses to discover conservative signatures of adaptive evolution.

Convergent evolution of *Chlamydomonas* sp. ICE-L in high salinity condition

Adaptive convergence means two or more distant lineages evolve similar traits independently due to similar environmental factors [57]. We identified nonrandom convergent and parallel amino acid substitutions between two halotolerant algae (*Chlamydomonas* sp. ICE-L and *Dunaliella salina*). A total of 19 parallel substitutions from five genes (*petA*, *psbF*, *psbZ*, *rbcl*, *tufA*) and four convergent substitutions from four genes (*atpB*, *atpI*, *petA*, *psaB*) showed significant convergence between *Chlamydomonas* sp. ICE-L and *Dunaliella salina* (Additional file 1: Table S5). Seven convergence genes were photosynthetic, and one gene (*tufA*) encoding elongation factor Tu was responsible for catalyzing the binding of an aminoacyl-tRNA (aa-tRNA) to the ribosome. Consistent with the result of positive selection, it showed same functional specificity that convergent evolution mainly targets photosynthetic genes. To unravel more conservative signatures of adaptive evolution, we focused on positively selected genes with non-random convergent amino acid substitutions. As a result, three genes (*atpB*, *psaB*, and *rbcl*) with both signatures of positive selection and convergent evolution were identified although positively selected sites were not overlapped with those that undergo convergent evolution (Table 3). Considering the similar high-salinity environments of *Chlamydomonas* sp. ICE-L and *Dunaliella salina*, we speculate that the conservative signature of adaptive evolution on these genes is associated with acclimation of salt stress.

Structural and functional analyses

Both positively selected sites and nonrandom convergent substitutions were mapped onto the three-dimensional (3D) crystal structure. Most identified positively selected sites and nonrandom convergent substitutions in *Chlamydomonas* sp. ICE-L were localized in or close to critical residues and functional regions, indicating that many sites were functionally relevant. In *rbcl* gene, the positively

Table 3 Convergent amino acid substitutions between *Chlamydomonas* sp. ICE-L and *Dunaliella salina*

Gene	Function	P value ^a	Q-value	Position(s)	AA Changes
<i>atpB</i>	ATP synthase subunit beta, chloroplastic	3.18E-03	0.0365	2	Val-Ile
<i>psaB</i>	Photosystem I P700 chlorophyll a apoprotein A2	1.27E-04	0.0058	116	Ser-Ala
<i>rbcl</i>	Ribulose biphosphate carboxylase large chain	1.95E-04	0.0019	32	Arg-Ser
				116	Met-Leu
				141	Pro-Ser

The numbering of amino acid residues identified by convergent evolution analyses corresponds to their alignment positions. ^a: A statistical test is conducted under the assumption that the number of convergent (or parallel) sites follows a Poisson distribution with the mean equal to the expected number. When the observed number is smaller than the expected, the lower tail probability is given; when the observed number is larger than the expected, the upper tail probability is given

selected sites SER₄₅₀ and CYS₄₅₇ were close to the heterodimer interface (TRP₄₅₁ and PRO₄₅₃) at the carboxy-terminus of the large subunit, and nonrandom convergent amino acid substitutions LEU₁₁₆ and SER₁₄₁ were one and nine amino acids from the homodimer interface (ASN₁₁₅ and ALA₁₃₂) (Fig. 2a, b). The positively selected site ASP₁₂₅ in *atpB* gene was five amino acids from the alpha subunit interaction interface position ASN₁₂₀ (Fig. 2c). As for *psaB* gene, the positively selected site PHE₆₄₆ localized in the transmembrane domain (position 644–666 amino acid), and it is eight amino acids from cofactors Mg²⁺ binding site HIS₆₅₅ (position 655 amino acid). In addition, the positively selected site ALA₂₂₀ of *psaB* gene were 10 amino acids from the predicted carbohydrate binding sites ASN₂₁₀ (Fig. 2d).

Discussion

Photosynthesis is a vital energy accumulation process for algae, and different abiotic stresses may drive adaptive modifications of photosynthetic proteins. The chloroplast protein-coding genes of algae living in extreme environments were used to conduct comparative analyses, aiming to uncover their relationship to the adaptability of extreme environments.

The branch model tests of both gene-specific and concatenated data showed no significant evidence of positive selection targeting the chloroplast protein-coding genes, which indicated strong purifying selection ($\omega < 1$). In general, the functional genes are highly conservative and positive selection may only target specific sites on

specific branches, thus positive selection is rare in the branch model test. The random-site model tests based on gene-specific and concatenated data showed that no positively selected site was identified by BEB analyses, implying that purifying selection is the predominant force shaping the evolution of algae chloroplast genes. Two possible reasons were deduced. One reason is that the chloroplast protein-coding genes are conserved to maintain photosynthetic function, and another reason is that the positive selection may have operated on some sites in specific lineages intermittently.

As expected, the notion that positive selection may target specific sites in specific lineages was confirmed through the branch-site model test. In our analyses, only 3 sites were identified under positive selection on *Chlorella* sp. ArM0029B and no site on *Coccomyxa subellipsoidea* C-169, whereas a large number of sites were identified as such along *Chlamydomonas* sp. ICE-L and *Dunaliella salina* in gene-specific and concatenated data analyses. Our results showed that positive selection exclusively targeted chloroplast photosynthetic protein-coding genes along *Chlamydomonas* sp. ICE-L and *Dunaliella salina* lineages.

Considering the observed similar physiological characteristics (salt-tolerant) in *Chlamydomonas* sp. ICE-L and *Dunaliella salina*, elevated salinity is likely the motivation of the adaptive molecular evolution of chloroplast protein-coding genes. Our convergent analyses demonstrated that seven out of eight convergent genes between *Chlamydomonas* sp. ICE-L and *Dunaliella salina* were photosynthetic showing their functional specificity. Why

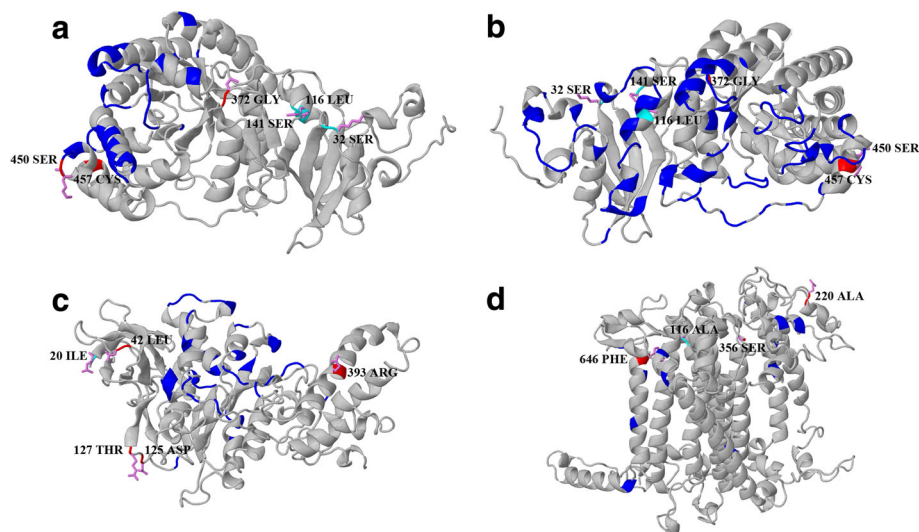


Fig. 2 Spatial distribution of positively selected sites and nonrandom convergent substitutions identified on *Chlamydomonas* sp. ICE-L. Positively selected sites (red) and nonrandom convergent substitutions (cyan) were close to functional domains in *Chlamydomonas* sp. ICE-L. The labels of positively selected sites and nonrandom convergent substitutions correspond to the positions in homologous proteins of *Arabidopsis thaliana*. **a** Nonrandom convergent substitutions in *rbcl* gene near the heterodimer interface (in blue). **b** Positively selected sites in *rbcl* gene close to the homodimer interface (in blue). **c** Positively selected site in *atpB* gene near the alpha subunit interaction interface (in blue). **d** Positively selected sites in *psaB* gene close to the predicted carbohydrate binding sites (in blue)

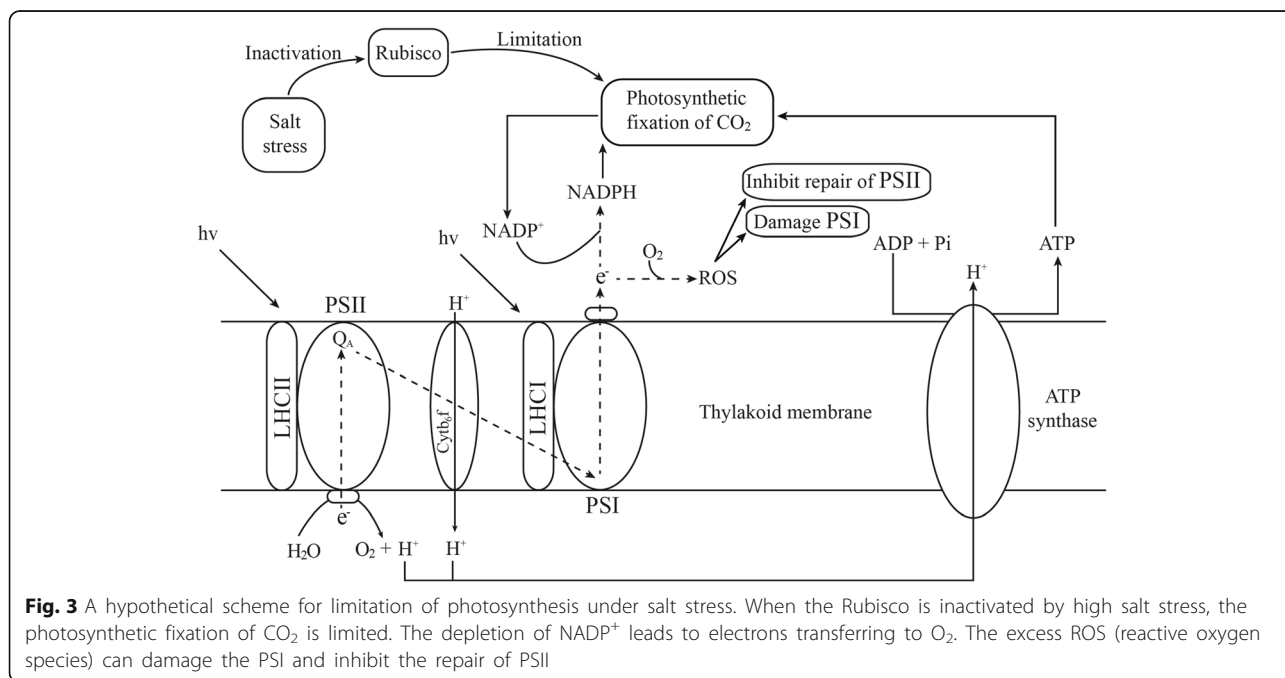
do different evolutionary patterns appear in four algae adapted to extreme environments? One possibility is that adaptive evolution of photosynthetic system may involve in nuclear-encoded subunits. We note that the chloroplast genome is not functionally independent of the nuclear genome [9]. Antenna complexes (nuclear-encoded) experienced adaptive modifications reducing light absorption efficiency and further decreasing the probability of damage to photosystem [58]. Another possible explanation is that adaptive modifications on other abiotic stresses targeting genes were sufficient to maintain the homeostasis for photosynthesis without the need of adaptive evolution of chloroplast-encoded photosynthetic subunits [59, 60].

The three-dimensional (3D) crystal structure analyses revealed more insights into how the adaptive sites may influence the function and structure of the chloroplast photosynthetic proteins. Here we focus on the positively selected genes (*rbcl*, *psaB* and *atpB*) with convergent amino acid substitutions in *Chlamydomonas* sp. ICE-L. The *rbcl* gene encoding Rubisco was identified with conservative signatures of adaptive evolution, and numerous adaptive modifications detected in our analyses may be responsible for high salinity stress. The Rubisco is one of the essential enzymes for inorganic carbon fixation in photosynthesis and the primary target of salt stress. High concentration of NaCl may limit the activity of Rubisco, which would inhibit the fixation of CO₂ (Fig. 3) [15, 61, 62]. Limitation of the photosynthetic fixation of CO₂ can induce the production of ROS [63]. ROS is mainly generated in plant chloroplast, which could be enhanced

under salt stress. It is reported that ROS produced under salt stress decreases photosynthetic rate and carboxylation efficiency resulting in the limitation of photosynthesis [64, 65]. The nonrandom convergent substitutions (SER₃₂, LEU₁₁₆ and SER₁₄₁) and positively selected sites (SER₄₅₀ and CYS₄₅₇) in *rbcl* gene were near the homodimer/heterodimer interface, which might contribute to the stability and activity of the enzyme [66]. It is likely that the Rubisco in *Chlamydomonas* sp. ICE-L evolved to restore activity and adapt to abiotic stresses (e.g. high salinity).

Moreover, the ROS is produced at the PSI reaction center, which in turn damage the PSI [67]. Protecting PSI from photodamage in excess ROS generated under salt stress is essential [68]. It is reasonable to assume that *Chlamydomonas* sp. ICE-L have some adaptive modifications to increase resistance to photodamage. For *psaB* gene, adaptive positively selected site PHE₆₄₇ is close to the cofactors Mg²⁺ binding site HIS₆₅₅ possibly contributing to maintain the PSI intactness for normal energy transfer efficiency. The PSI reaction center polypeptides PsaA and PsaB bind many electron transport components [69–71]. Our results showed that *psaB* genes involved in phototrophic electron transport chain experienced adaptive evolution which may help to maintain the stabilization of PSI in *Chlamydomonas* sp. ICE-L.

We also found that *atpB* gene associated with ATP syntheses in chloroplast were under adaptive evolution (Table 3). The beta subunits of ATP synthase encoded by *atpB* gene is responsible for catalysis of ATP synthesis. The positively selected site ASP₁₂₅ in *atpB* gene was near the alpha/beta subunits interaction interface ASN₁₂₀.



According to available functional information, this adaptive modification may contribute to the activity of ATP synthase, and the increased ATP synthase activity in *Chlamydomonas* sp. ICE-L could meet the need of CO₂ fixation. More structural and functional information are essential to gain more insights into the adaptive modifications in *atpB* gene. Together, these confirmative molecular evidences demonstrated that *Chlamydomonas* sp. ICE-L might evolve to maintain high photosynthetic rate and carboxylation efficiency under high salinity environment, and the adaptive modifications in photosynthetic genes of *Chlamydomonas* sp. ICE-L respond to selection pressure from salt stress.

It is noted that both positive selection and convergence are detected at different sites in genes that have undergone adaptive evolution. There are multiple adaptive strategies that influence the adaptive substitutions, and it is reasonable that putatively adaptive substitutions in different green algae living in extreme environments is not convergent in the current study. Similarly, a recent work involving convergent evolution of marine mammals revealed no significant positive correlation between convergence and positive selection in marine mammals [72]. In addition, there are different types of convergent evolution, such as sequence, functional, mechanistic and structural convergence [73]. A previous study has shown strong pattern of convergence emerged at the level of genes, operons, and functional complexes in *Escherichia coli* [74]. We believe that further functional and structural analyses of these adaptive substitutions in *Chlamydomonas* sp. ICE-L will complement our findings.

Conclusions

We investigate adaptive mechanisms of chloroplast photosynthetic genes from the Antarctic green alga *Chlamydomonas* sp. ICE-L. Our study provides strong evidences that positive Darwinian selection operates on the chloroplast protein-coding genes of *Chlamydomonas* sp. ICE-L adapted to extreme environment following functional-specific and lineages-specific patterns. We identified three genes with conservative signatures of adaptive evolution in *Chlamydomonas* sp. ICE-L, and the adaptive modifications in these genes were in functionally important regions of the proteins. Our results reveal the phototrophic component of *Chlamydomonas* sp. ICE-L exhibits adaptive evolution under extreme environments, and high salinity is likely an important trigger of the adaptations at the chloroplast level. It should be noted that some subunits of photosynthetic systems are nuclear genome-encoded, and it would be more beneficial to understand the adaptive strategies by analyzing all photosynthetic genes in both chloroplast and nuclear genomes. Further experimental and structural analyses will provide confirmative evidences uncovering the adaptive evolution of photosynthetic systems in *Chlamydomonas* sp. ICE-L.

Additional file

Additional file 1: Table S1. Results of the basic (M0) model test of individual genes and the concatenated data. **Table S2.** Results of the correct *P*-value of LRT in branch-model test. **Table S3.** Results of site-model test to each gene-specific and the concatenated data. **Table S4.** Results of branch-site model analyses using the concatenated data. **Table S5.** Results of convergence test between *Chlamydomonas* sp. ICE-L and *Dunaliella salina*. **Figure S1.** Results of divergence time analyses. The estimated divergence times were showed with 95% confidence intervals. **Figure S2.** The dN/dS estimates for the concatenated data using the M8 random-site model. The y-axis shows the BEB posterior mean estimate of dN/dS for each site. (DOCX 341 kb)

Abbreviations

BEB: Bayes empirical Bayes; BS: bootstrap support; LHCS: Light-harvesting complexes; LRT: likelihood ratio test; Phyre: Protein Homology/analogy Recognition Engine; PP: posterior probability; ROS: Reactive oxygen species; Rubisco: Ribulose biphosphate carboxylase/oxygenase

Acknowledgments

We are grateful to Zhengting Zou and Jianzhi Zhang for help with convergent evolution analyses, and Harold Zakon for critical reading and English editing. We thank two anonymous reviewers for their valuable comments on this manuscript.

Funding

This study was financially supported by the National Natural Science Foundation of China (31570219), the Natural Science Foundation of Jiangsu Province (BK20150971), the Jiangsu Province Key Project for Scientific Research (16KJA180002), the Six Talent Peaks Project of Jiangsu Province (16KJA180002), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and Qing Lan Project.

Availability of data and materials

Chloroplast protein-coding sequences of *Chlamydomonas* sp. ICE-L are available at GenBank, and accession numbers are MF280291 - MF280350. Multiple sequence alignments and phylogenetic trees are available in the FigShare repository. <https://figshare.com/s/8efff1838bb4a25852eb> [75].

Authors' contributions

BZ conceived the project and designed the experiments. ZZ, MA, JM and ZG performed the experiments, and ZZ analyzed the data. ZZ and BZ drafted the manuscript. All authors revised the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, China. ²Medical College, Qingdao University, Qingdao, China. ³Key Laboratory of Marine Bioactive Substance, The First Institute of Oceanography, State Oceanic Administration, Qingdao, China. ⁴State Key Laboratory of Natural Medicines, School of Life Science and Technology, China Pharmaceutical University, Nanjing, China.

Received: 5 December 2017 Accepted: 27 March 2018

Published online: 03 April 2018

References

- Blanc G, Agarkova I, Grimwood J, Kuo A, Brueggeman A, Dunigan DD, Gurnon J, Ladunga I, Lindquist E, Lucas S. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol.* 2012;13(5):1–12.
- Thomas D, Dieckmann G. Antarctic Sea ice: a habitat for extremophiles. *Science.* 2002;295(5555):641–4.
- Arrigo K, Worthen D, Lizotte M, Dixon P, Dieckmann G. Primary production in Antarctic Sea ice. *Science.* 1997;276(5311):394–7.
- An M, Mou S, Zhang X, Zheng Z, Ye N, Wang D, Zhang W, Miao J. Expression of fatty acid desaturase genes and fatty acid accumulation in *Chlamydomonas* sp. ICE-L under salt stress. *Bioresour Technol.* 2013;149:77–83.
- Liu C, Wang X, Wang X, Sun C. Acclimation of Antarctic *Chlamydomonas* to the sea-ice environment: a transcriptomic analysis. *Extremophiles.* 2016; 20(4):437–50.
- Michel C, Legendre L, Theriault J-C, Demers S. Photosynthetic responses of Arctic Sea-ice microalgae to short-term temperature acclimation. *Polar Biol.* 1989;9(7):437–42.
- Wang Y B, Liu F, Zhang X, Zhang A, Wang B, Zheng Z, Sun C, Miao J. Composition and regulation of thylakoid membrane of Antarctic ICE microalgae *Chlamydomonas* sp. ICE-L in response to low-temperature environment stress. *J Mar Biol Assoc UK.* 2017;97(6):1241–49.
- Wicke S, Schneeweiss GM. Next-generation organellar genomics: potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. In: Hörandl E, Appelhans MS, editors. *Next-generation sequencing in plant systematics.* Vienna: Koeltz Scientific Books; 2015.
- Allen JF, De Paula WBM, Puthiyaveetil S, Nield J. A structural phylogenetic map for chloroplast photosynthesis. *Trends Plant Sci.* 2011;16(12):645–55.
- de Koning AP, Keeling PJ. The complete plastid genome sequence of the parasitic green alga *Helicospiridium* sp. is highly reduced and structured. *BMC Biol.* 2006;4(1):12.
- Maul JE, Lilly JW, Cui L, Miller W, Harris EH, Stern DB. The *Chlamydomonas reinhardtii* plastid chromosome islands of genes in a sea of repeats. *Plant Cell.* 2002;14(11):2659–79.
- Mou S, Zhang X, Ye N, Dong M, Liang C, Liang Q, Miao J, Xu D, Zheng Z. Cloning and expression analysis of two different LhcSR genes involved in stress adaptation in an Antarctic microalga, *Chlamydomonas* sp. ICE-L. *Extremophiles.* 2012;16(2):193–203.
- Tang GQ, Novitzky WP, Carol Griffin H, Huber SC, Dewey RE. Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. *Plant J.* 2005;44(3):433–46.
- Devos N, Ingouff M, Loppes R, Matagne RF. RUBISCO adaptation to low temperatures: a comparative study in psychrophilic and mesophilic unicellular algae. *J Phycol.* 1998;34(4):655–60.
- Ziska LH, Seemann JR, DeJong TM. Salinity induced limitations on photosynthesis in *Prunus salicina*, a deciduous tree species. *Plant Physiol.* 1990;93(3):864–70.
- Zacher K. UV effects on photosynthesis and DNA in propagules of three different Antarctic macroalgae species (*Adenocystis utricularis*, *Monostroma hartioides* and *Porphyra endiviifolium*). *Planta.* 2007;225(6):1505–16.
- Roleda MY, Zacher K, Wulff A, Hanelt D, Wiencke C. Photosynthetic performance, DNA damage and repair in gametes of the endemic Antarctic brown alga *Ascoseira mirabilis* exposed to ultraviolet radiation. *Austral Ecol.* 2007;32(8):917–26.
- Cvetkovska M, Hüner NP, Smith DR. Chilling out: the evolution and diversification of psychrophilic algae with a focus on *Chlamydomonadales*. *Polar Biol.* 2017;40(6):1169–84.
- Bascuñán-Godoy L, Sanhueza C, Cuba M, Zuñiga GE, Corcuera LJ, Bravo LA. Cold-acclimation limits low temperature induced photoinhibition by promoting a higher photochemical quantum yield and a more effective PSII restoration in darkness in the Antarctic rather than the Andean ecotype of *Colobanthus quitensis* Kunt Bartl (Cariophyllaceae). *BMC Plant Biol.* 2012;12(1):1–15.
- Huovinen P, Gómez I. Photosynthetic characteristics and UV stress tolerance of Antarctic seaweeds along the depth gradient. *Polar Biol.* 2013;36(9):1319–32.
- Hünken M, Harder J, Kirst G. Epiphytic bacteria on the Antarctic ice diatom *Amphiprora kufferathii* Manguin cleave hydrogen peroxide produced during algal photosynthesis. *Plant Biol.* 2008;10(4):519–26.
- Qu Y, Zhao H, Han N, Zhou G, Song G, Gao B, Tian S, Zhang J, Zhang R, Meng X. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun.* 2013;4:2701.
- Tian R, Wang Z, Niu X, Zhou K, Xu S, Yang G. Evolutionary genetics of hypoxia tolerance in cetaceans during diving. *Genome Biol Evol.* 2016; 8(3):827–39.
- Wang Y, Yang L, Zhou K, Zhang Y, Song Z, He S. Evidence for adaptation to the Tibetan plateau inferred from Tibetan loach transcriptomes. *Genome Biol Evol.* 2015;7(11):2970–82.
- Raven JA, Colmer TD. Life at the boundary: photosynthesis at the soil–fluid interface. A synthesis focusing on mosses. *J Exp Bot.* 2016;67(6):1613–23.
- R Marcelino V, Cremen MCM, Jackson CJ, Larkum AA, Verbruggen H: Evolutionary dynamics of chloroplast genomes in low light: a case study of the endolithic green alga *Ostreobium quekettii*. *Genome Biol Evol.* 2016, 8(9):2939–2951.
- Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, Liu B, Qiu Q, Wang Z, Zhang J. Genomic insights into salt adaptation in a desert poplar. *Nat Commun.* 2013;4:2797.
- Holm-Hansen O. Isolation and culture of terrestrial and fresh-water algae of Antarctica 1. *Phycologia.* 1964;4(1):43–51.
- Ahn J-W, Hwangbo K, Lee SY, Choi H-G, Park Y-I, Liu JR, Jeong W-J. A new Arctic *Chlorella* species for biodiesel production. *Bioresour Technol.* 2012;125:340–3.
- Polle JF W, Tran D, Ben-Amotz A. History, distribution, and habitats of algae of the genus *Dunaliella* Teodoresco (Chlorophyceae)[J]. In: *The Alga Dunaliella*: Science publishers; 2009. p. 1–13.
- Provasoli L. Media and prospects for the cultivation of marine algae. In: Watanabe A, Hattori R, editors. *Culture and collections of algae.* Proc U.S.-Japan Conference Hakone. 12-15 September 1966. p. 63-75.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402.
- Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 2004;20(17):3252–5.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
- Löytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A.* 2005;102(30):10557–62.
- Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 2000;17(4):540–52.
- Redelings B. Erasing errors due to alignment ambiguity when estimating positive selection. *Mol Biol Evol.* 2014;31(8):1979–93.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.
- Breinholt JW, Kawahara AY. Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biol Evol.* 2013;5(11):2082–92.
- Nei M, Kumar S. *Molecular evolution and phylogenetics.* New York: Oxford university press; 2000.
- Escalona T, Weadick CJ, Antunes A. Adaptive patterns of mitogenome evolution are associated with the loss of shell scutes in turtles. *Mol Biol Evol.* 2017;34(10):2522–36.
- Fletcher W, Yang Z. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol Biol Evol.* 2010;27(10):2257–67.
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003;100(16):9440–5.
- Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol.Biol.Evol.* 2005;22(12):2472–9.
- Zhang J, Kumar S. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol.Biol.Evol.* 1997;14(5):527–36.
- Cox CJ, Foster PG. A 20-state empirical amino-acid substitution model for green plant chloroplasts. *Mol Phylogenet Evol.* 2013;68(2):218–20.
- Zou Z, Zhang J. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol.Biol.Evol.* 2015;32(8):2085–96.

50. Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S, Van Elk C, Hunter ME, Joshi V. Convergent evolution of the genomes of marine mammals. *Nat Genet.* 2015;47(3):272–5.
51. Thomas GWC, Hahn MW. Determining the null model for detecting adaptive convergence from genomic data: a case study using Echolocating mammals. *Mol.Biol.Evol.* 2015;32(5):1232–6.
52. Kelley LA, Sternberg MJ. Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc.* 2009;4(3):363–71.
53. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008;9(1):1–8.
54. Fucile G, Di Biase D, Nahal H, La G, Khodabandeh S, Chen Y, Easley K, Christendat D, Kelley L, Provart NJ. ePlant and the 3D data display initiative: integrative systems biology on the world wide web. *PLoS One.* 2011;6(1):e15237.
55. Lemieux C, Vincent AT, Labarre A, Otis C, Turmel M. Chloroplast phylogenomic analysis of chlorophyte green algae identifies a novel lineage sister to the Sphaeropleales (Chlorophyceae). *BMC Evol Biol.* 2015;15(1):264.
56. Nozaki H, Nakada T, Watanabe S. Evolutionary origin of Gloeomonas (Volvocales, Chlorophyceae) based on ultrastructure of chloroplasts and molecular phylogeny. *J Phycol.* 2010;46(1):195–201.
57. Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature.* 2013;502(7470):228–31.
58. Alboresi A, Ballottari M, Hienerwadel R, Giacometti GM, Morosinotto T. Antenna complexes protect photosystem I from Photoinhibition. *BMC Plant Biol.* 2009;9(1):1–14.
59. Dolhi JM, Maxwell DP, Morgan-Kiss RM. The Antarctic *Chlamydomonas raudensis*: an emerging model for cold adaptation of photosynthesis. *Extremophiles.* 2013;17(5):711–22.
60. Hirooka S, Hirose Y, Kanesaki Y, Higuchi S, Fujiwara T, Onuma R, Era A, Ohbayashi R, Uzuka A, Nozaki H. Acidophilic green algal genome provides insights into adaptation to an acidic environment. *Proc Natl Acad Sci U S A.* 2017;114(39):E8304–13.
61. Parry MA, Andralojc P, Khan S, Lea PJ, Keys AJ. Rubisco activity: effects of drought stress. *AnnBot.* 2002;89(7):833–9.
62. Lu K, Cao B, Feng X, He Y, Jiang D. Photosynthetic response of salt-tolerant and sensitive soybean varieties. *Photosynthetica.* 2009;47(3):381–7.
63. Asada K. The water-water cycle in chloroplasts: scavenging of active oxygens and dissipation of excess photons. *Annu Rev Plant Biol.* 1999;50(1):601–39.
64. Neale PJ, Melis A. Salinity-stress enhances Photoinhibition of photosynthesis in *Chlamydomonas reinhardtii*. *J Plant Physiol.* 1989;134(5):619–22.
65. Murata N, Takahashi S, Nishiyama Y, Allakhverdiev SI. Photoinhibition of photosystem II under environmental stress. *BBA-Bioenergetics.* 2007; 1767(6):414–21.
66. Hong S, Spreitzer RJ. Complementing substitutions at the bottom of the barrel influence catalysis and stability of ribulose-bisphosphate carboxylase/oxygenase. *J Biol Chem.* 1997;272(17):11114–7.
67. Sonoike K. Degradation of psaB gene product, the reaction center subunit of photosystem I, is caused during photoinhibition of photosystem I: possible involvement of active oxygen species. *Plant Sci.* 1996;115(2):157–64.
68. Sonoike K. Photoinhibition of photosystem I. *Physiol Plantarum.* 2011;142(1):56–64.
69. Rüdiger W. Biosynthesis of chlorophyll b and the chlorophyll cycle. *Photosynth Res.* 2002;74(2):187–93.
70. Luciński R, Jackowski G. The structure, functions and degradation of pigment-binding proteins of photosystem II. *Acta Biochim Pol.* 2006;53(4):693–708.
71. Green B, Durnford D. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol.* 1996;47(1):685–714.
72. Zhou X, Seim I, Gladyshev VN. Convergent evolution of marine mammals is associated with distinct substitutions in common genes. *Sci Rep.* 2015;5:16550.
73. Doolittle RF. Convergent evolution: the need to be explicit. *Trends Biochem Sci.* 1994;19(1):15–8.
74. Tenailon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. The molecular diversity of adaptive convergence. *Science.* 2012;335(6067):457–61.
75. Zhang Z, An M, Gu Z, Miao J, Zhong B. The Antarctic sea ice alga *Chlamydomonas* sp. ICE-L provides insights into adaptive patterns of chloroplast evolution. *FigShare.* 2017. <https://figshare.com/s/8efff1838bb4a25852eb>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

