Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 19 (2009) 1561–1568
Progress in Natural Science
www.elsevier.com/locate/pnsc

Vertical adaptive radiation in ocean *Prochlorococcus*: Evolutionary implications of the Chl *b*/*a* ratio from molecular evidence

Song Qin [a,*], Yinchu Wang [a,b]

[a] *Yantai Institute of Coastal Zone Research for Sustainable Development, Chinese Academy of Sciences, Yantai 264003, China*
[b] *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

Received 23 February 2009; received in revised form 7 May 2009; accepted 18 May 2009

Abstract

Prochlorococcus is a marine cyanobacterium of global significance. Two ecotypes are adapted to either high-light (HL) or low-light (LL) conditions. The ratio between chlorophyll (Chl) *a* and *b* is a distinguishing characteristic of these two ecotypes. However, how this ratio evolved in *Prochlorococcus* during this ecotype differentiation remains unclear. Our analyses reveal that the ancestor of *Prochlorococcus* was typically low-light adapted. The LL ecotype showed a stagnant evolution, and the HL ecotype was recently diverged. There was an adaptive radiation after directional evolution in the Chl *b*/*a* ratio regulation. Recombination in chlorophyllide *a* oxygenase (CAO) and positive selection on Clp protease contributed to the directional evolution of *Prochlorococcus*. The recombinant fragments of CAO were correlated with a large group of shared coevolving sites. Evidence of positive selection was found in both subunits of Clp. Chl *b*/*a* ratio evolution, as annotated by molecular evidence, appears to be among the crucial reasons that explain how *Prochlorococcus* has become the dominant photosynthetic organism in the ocean.

Keywords: *Prochlorococcus*; Chl *b*/*a* ratio; Adaptive evolution; Trait evolution; Ecotype differentiation

1. Introduction

Plants and algae can potentially use the reactions of the chlorophyll (Chl) cycle to adjust the Chl *b*/*a* ratio to their particular needs under various physiological and environmental conditions [1].

Prochlorococcus is a marine cyanobacterium of global significance and is the most abundant photosynthetic organism on earth. It inhabits the ocean within the 40° N to 40° S latitudinal bands, and its high density and phototropic metabolism make it accountable for an important part of the primary production of the world's oceans. Its vertical distribution spans from the surface to a depth of 200 m, where the light level can be as low as 0.1% of surface irradiance. *Prochlorococcus* possesses a remarkable pigment complement, which includes divinyl derivatives of Chl *a* and Chl *b* [2]. Its contribution as a major component of the carbon cycle is at least partly due to its unique pigment systems. Divinyl Chl *b* is especially important for enabling *Prochlorococcus* to live and thrive within its specialized niche [3].

Prochlorococcus exhibits unusually large changes in Chl *b*/*a* ratio with depth [4]. There are at least two ecotypes of *Prochlorococcus* that coexist in the oceans; these are distinguished not only by their photophysiology but also by their molecular phylogeny [5]. One ecotype (LL) is adapted to the low light of the deep euphotic zone, while the other (HL) is adapted to the high light at the surface. The two ecotypes are characterized *in vivo* by high Chl *b*/*a* ratios (LL ecotype) and low Chl *b*/*a* ratios (HL ecotype). The ratio of Chl *b*/*a* increases as irradiance decreases, but at

* Corresponding author. Tel./fax: +86 532 82898500.
E-mail address: sqin@yic.ac.cn (S. Qin).

1002-0071/$ - see front matter © 2009 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2009.05.005

any particular growth irradiance, the HL isolates will always have significantly lower Chl *b*/*a* ratios than the LL isolates ($P < 0.01$) [2,4].

The reduction of Chl *b* to Chl *a* (and its hydrolysis to chlorophyllide *a*) and the biosynthesis of Chl *b* from chlorophyllide *a* can be considered to operate as a cycle that starts and ends with chlorophyllide *a*. Biosynthesis of both Chls begins with chlorophyllide *a*. Completion of Chl *a* biosynthesis needs only the step of esterification catalyzed by Chl synthase, while Chl *b* biosynthesis includes a two-step reaction catalyzed by chlorophyllide *a* oxygenase (CAO) and a subsequent esterification step [1,6,7]. The pathway of *Prochlorococcus* chlorophyll synthesis is essentially the same as the known pathways in other algae and in higher plants. The pathway is illustrated in Fig. 1, which was revised from Kettler et al. [8] and generated by Pathway Tools [9]. It is reasonable to consider that the biosynthesis of chlorophyll *a* and *b* is tightly regulated [10].

Biochemical and genetic evidence has shown that CAO is the only enzyme necessary for the transition of *a*-type to *b*-type chlorophyll pigments [1]. The CAO gene is a unique gene in the synthesis of Chl *b* and has recently been identified in *Prochlorococcus* [3,11]. CAO protein levels correlate with changes in Chl *b* levels and with the Chl *b*/*a* ratio [12]. Regulation is thought to occur via interaction between Lhcb apoproteins and the CAO protein, as suggested by the enhanced Chl *b* synthesis that has been shown in cyanobacteria that coexpress both CAO and Lhcb [12,13].

A second key enzyme that responds to Chl *b*/*a* ratio regulation is Clp protease that consists of two subunits, ClpC1 and ClpP. This enzyme regulates the level of CAO through destabilization of the CAO protein in response to the accumulation of Chl *b* [10]. Clp is widely distributed in cyanobacteria and chloroplasts. The CAO in *Prochlorococcus* contains AB domains that are Clp signal sequences and that lead to CAO degradation by Clp [3,14,15].

The level of transcripts encoding Clp proteins and the chlorophyll fluorescence per cell both vary with the light intensity in *Prochlorococcus* [16] in conjunction with the changes in Chl *b*/*a* ratio [4]. These same phenomena are seen in *Arabidopsis thaliana* [10]. Clp protease has been shown to control chlorophyll *b* synthesis by regulating the level of CAO in *Arabidopsis thaliana* [10].

The reverse pathway, the transition of the *b*-type to *a*-type Chl pigments, seems to play a role in the degradation of Chl *b* and operates during acclimation of plants to different light regimes. However, this transition has so far been characterized only by activity tests, which point to two different enzymes for this transition, and this type of enzyme activity has not been detected in algae [1]. Therefore, chlorophyll *b* reductase was not examined in the present study.

Other pigments, such as phycoerythrin (PE) [17,18], also show different patterns of expression in HL and LL *Prochlorococcus*. In addition, the phycoerythrin content of *Prochlorococcus* is extremely low (PE/DV-Chl *b* ratio of about 1:330 while the typical phycobilin/Chl ratio is about 1:2 for a cyanobacterium). Thus, the capacity of

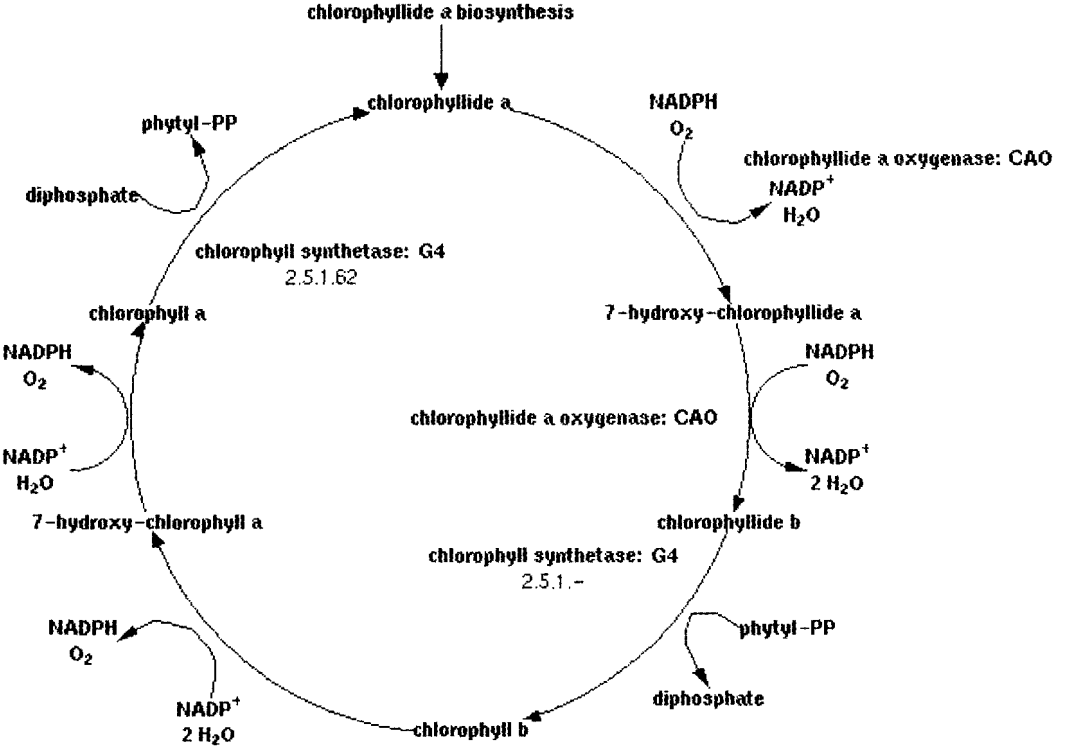

Fig. 1. The chlorophyll cycle of *Prochlorococcus*. If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.

phycoerythrin to function primarily as a photosynthetic light-harvesting pigment in *Prochlorococcus* is also extremely low (less than 1.8%) [19].

Taken together, the available information on the photophysiology of *Prochlorococcus* suggests that this is quite unique. Of equal interest, it also appears to be different in the HL and LL ecotypes. Therefore, in this study, we further examined the ecological significance of the Chl *b*/*a* ratio and how it is maintained in the two ecotypes of this species, by exploring the evolutionary history of *Prochlorococcus* at the molecular level.

2. Materials and methods

2.1. Sequence collection and alignment

Our exploration was based on the work of Moore and Chisholm [4] that focused on the photophysiology of ten different *Prochlorococcus* isolates from diverse oceanographic locations. Satoh and Tanaka [3], who identified CAO in *Prochlorococcus,* also pointed out that CAO is the only non-heme oxygenase in *Prochlorococcus* genomes. Therefore, we retrieved gene and protein sequences of the CAO cluster (CLS1023049) from the NCBI database (http://www.ncbi.nlm.nih.gov/). This cluster includes the Rieske iron–sulfur protein 2Fe–2S subunit sequences (namely CAO) of ten *Prochlorococcus* strains that belong to two ecotypes (HL: AS9601, MIT 9301, MIT 9312, MIT 9515, MED4; LL: MIT 9303, MIT 9313, SS120, NATL1A and NATL2A). We also retrieved gene and protein sequences of ClpC1 and ClpP in protein clusters (CLS1105073 and PRK00277) of the same 10 isolates. These strains have been isolated both at the surface and at depth from various sites in oceans around the world [2,4], which makes the sample sufficiently representative and randomized for molecular population analyses.

Amino acid alignments were carried out using CLUSTAL W software with the default settings [20]. The nucleotide sequences were then aligned following the same gap patterns using Tranalign online (http://embossgui.sourceforge.net/demo/tranalign.html).

2.2. Analyses of evolutionary traits

Recent progress on methods for studying trait evolution has inspired new insights, built on a solid statistical foundation. BayesTraits [21] is a computer package for performing analyses of trait evolution among groups of species for which a phylogeny or sample of phylogenies is available, including MultiState, MultiAns, Discrete and Continuous. It uses Markov chain Monte Carlo (MCMC) methods to derive posterior distributions and maximum likelihood (ML) methods to derive point estimates, log likelihoods, the parameters of statistical models and the values of traits at ancestral nodes of phylogenies. In this study, BayesMultiState was used to reconstruct how the ecotypical traits of *Prochlorococcus* may have evolved on phylogenetic trees. This is useful for reconstructing ancestral states and for testing models of trait evolution [22]. Many phylogenetic trees (proteins of CpeB, ClpC1, ClpP, CAO, 16S RNA, etc.) obtained by three methods (neighbor-joining, maximum parsimony and maximum likelihood) were used, and the two traits were encoded. Hypotheses were tested by using BayesContinuous [23,24] about models of evolution, ancestral states and correlations among pairs of traits. Four likelihood ratio tests were taken: (1) Influence of Phylogeny; (2) Drift versus Directional Models of Evolution; (3) Punctional and Gradual Traits Evolution; and (4) Adaptive Radiation versus Species Adaptation. All data for the computation are derived from Moore and Chisholm [4] and are listed in Table 1.

2.3. Neutrality tests

To collect any evidence in response to Chl *b*/*a* ratio evolution at the molecular level and investigate how natural selection plays its role, the following analyses were carried out.

To expediently compare with previous work regarding positive selection constraints on phycoerythrin in *Prochlorococcus* [18], we adopted the same neutrality tests, parameters and software. Two measures of diversity were computed for the population samples: (1) π, the direct estimate of per-site heterozygosity derived from the average number of pairwise sequence differences in the sample [25]; (2) Watterson's θ_W, based on the number of segregating sites in the sample, which is an estimate of the expected per-site nucleotide heterozygosity, theoretically equal to the neutral mutation parameter: $4Ne\mu$ [26]. To test whether the frequency spectrum of mutations conformed to the expectations of the standard neutral model, we calculated the values of the following statistics using the total number of segregating sites: Tajima's D statistic [27], Fu and Li's D^* and F^* statistics [28]. Significance values for each of the above three test statistics were estimated by means of 10,000 coalescent simulations of a Wright–Fisher equilibrium model that was conditioned on the sample size and

Table 1
Traits and states of 10 *Prochlorococcus* isolates.

Isolate	State	$I(k, g)$	$R - I(k, g)$	$R - I(30)$
MED4	H	45	0.14201	0.16864
MIT 9201	H	40	0.23077	0.23077
MIT 9215	H	59	0.26627	0.27515
MIT 9202	H	59	0.28402	0.30178
MIT 9302	H	56	0.38166	0.44379
MIT 9312	H	36	0.44379	0.46154
MIT 9313	L	25	0.84320	0.79882
MIT 9303	L	22	1.03846	0.95858
MIT 9211	L	23	1.42899	1.34911
SS120	L	27	1.43787	1.38462

The states H and L stand for the ecotype High-light adaptive and Low-light adaptive, respectively; $I(k, g)$ is irradiance at which growth is saturated, in units of μmol Q m^{-2} s^{-1}. $R - I(k, g)$ and $R - I(30)$ are the Chl *b*/*a* ratios in response to irradiances of $I(k, g)$ and 30 μmol Q m^{-2} s^{-1}.

level of polymorphism of the observed data, with no recombination [29], using DnaSP version 4.0 software [30]. Statistical significance for all of the above tests was established using 1000 permutations.

In addition, other neutrality tests were carried out using NeutralityTest [31]. Zeng–Fu–Shi–Wu's E test, a statistical test for detecting positive selection by utilizing high-frequency variants, showed the most power in detecting the recovery phase after the loss of genetic diversity, which includes the postselective sweep phase [32]. Afterwards, we used Fay–Wu's H test to detect whether there was any hitchhiking under positive Darwinian selection by measuring an excess of high compared to intermediate frequency variants [33]. A McDonald and Kreitman test of neutrality [34] was employed by comparing the distribution of synonymous and non-synonymous (replacement) variation within and between species. These tests were performed using the default parameters set of NeutralityTest with the purpose of testing for an excess of recent mutations or rare alleles (Left-side test). Critical points of the tests were based on 5000 simulated samples.

2.4. Recombination and coevolution analyses

The evolutionary pattern of CAO was further analyzed with possible recombination taken into account. The nucleotide substitution bias model was selected by referring to the result of HyPhy package [35]. GARD [36] implemented within the HyPhy software package was used to screen the sequences for recombination breakpoints, to identify non-recombinant regions and to allow each to have its own phylogenetic tree. PARRIS [37], a robust inference of positive selection from recombining coding sequences, was then used to deal with recombinant data. The significance level of the p value was limited to 0.1. CAPS [38] was used to measure the intramolecular coevolution between amino acid sites belonging to different recombinant segments (Significance test: random sampling = 10,000, threshold *alpha* value = 0.001; weight correlation by the divergence time between sequences using synonymous distances by Li 1993; minimum $R = 0.1$ GrSize = 5). The sample size of 10 was sufficient for accurate results by CAPS analysis [38].

3. Results

3.1. Analyses of trait evolution

The BayesMultiState result suggested that the ancestor of *Prochlorococcus* is likely the LL ecotype, and the likelihood value reached its minimum (−0.966807) using the phylogenetic tree based on CpeB by the Neighbor-Joining method (Table 2). BayesContinuous also indicated the ancestor to be a typical LL ecotype and estimated the ancestral states $I(k, g)$ and $I(30)$ to be 1.13587 and 1.12172, respectively (Fig. 2). The covariance between these two traits was 0.119913. In the Influence of Phylogeny analysis, the parameter λ, which assesses the contribution of phylogeny, was equal to 1. This indicated that the phylogeny fit the default hypothesis that any sort of phylogenetic correction was unnecessary, and the phylogenetic history had no minimal effect. By comparing Model A to Model B using the LR test, the model of directional evolution had the most significant likelihood (Lh) value (random walk: 22.829971; directional: 23.182243; regression: 19.309152. All of these allowed covariance to be non-zero). This means that *Prochlorococcus* underwent a directional evolution rather than a random walk or regression.

Table 2
BayesMultiState result estimating ancestral state.

Tree no.	Lh	qHL	qLH	Root P(H)	Root P(L)
1	−0.966807	0	1.662426	0	1
2	−0.967289	0	1.645328	0	1
3	−1.532260	0.839656	0.959389	0.472147	0.527853
4	−1.532260	0.839656	0.959389	0.472147	0.527853
5	−1.532260	0.839656	0.959388	0.472147	0.527853
6	−2.069781	7.337113	2.05701	0.677228	0.322772
7	−2.073936	7.191661	2.111109	0.665518	0.334482
8	−2.176884	5.201918	3.431442	0.386856	0.613144
9	−2.400973	0	0.959375	0	1
10	−2.413757	0	0.957907	0	1
11	−2.579437	0	0.949009	0	1

Lh are the likelihood values of 11 tested phylogenetic trees; qHL and qLH are the transition rates between the state HL and LL adaption. Root P(H) and Root P(L) are the values of the reconstructed probabilities of the two states at the root of the tree.

In the model of Punctional versus Gradual Trait Evolution, the former was comparatively significant, with a likelihood value of 22.409401 to 21.715087. The parameter κ was estimated to be 0.39447. This implied some form of gradualism and stasis in the longer branches, namely the LL ecotype clade. Moreover, in the model of Adaptive Radiation versus Species Adaptation, the former was comparatively significant, with the likelihood value of 21.42714–21.40913. The parameter δ was estimated to be 0.813327, which suggests that adaptive radiation has been dominant in *Prochlorococcus*.

3.2. Neutrality tests

Results of nucleotide diversity and neutrality tests are listed in Table 3. Overall, estimates of nucleotide diversity (π) and the numbers of segregating sites (θ_W) for the HL population were higher than those for the LL population, but they were lower than those for the collectivity, especially for CAO. This is because rare variants in HL contributed more to π than to θ_W. From another angle, this means that CAO has more significant π and θ_W than do the two genes encoding the Clp protease.

Coincidentally, Fay and Wu's H and revised H indicated that there might be hitchhiking in all three of these genes. McDonald and Kreitman's G tests, which proposed a two-by-two contingency test using the numbers of non-synonymous and synonymous polymorphisms within species and the numbers of non-synonymous and synonymous dif-

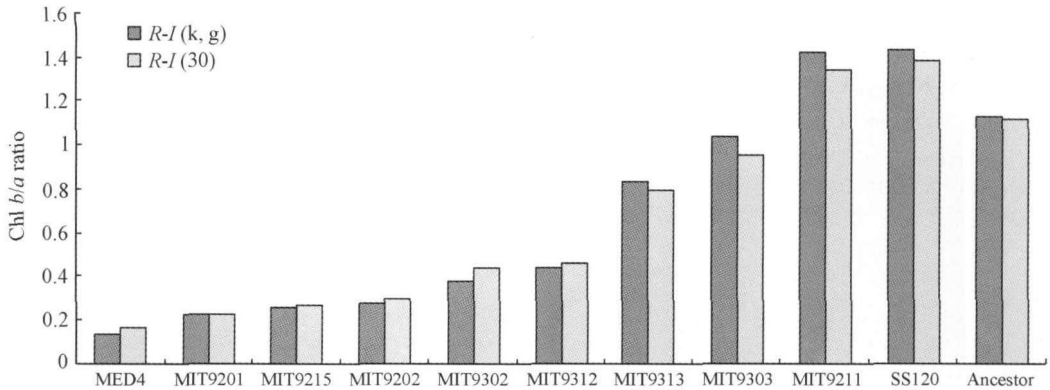

Fig. 2. Chl *b/a* ratios of 10 *Prochlorococcus* isolates and their inferred ancestor. $R - I(k, g)$ and $R - I(30)$ are the Chl *b/a* ratios in response to irradiances of $I(k, g)$ and 30 μmol Q m^{-2} s^{-1}.

Table 3
Results of nucleotide diversity and neutrality tests.

Gene	Ecotype	π	θ_W	D	D^*	F^*	G	H	Revised H	E
CAO										
1404 bp	HL ecotype	0.18740	0.16255	1.16800	0.18500	0.55248	**50.05200**	**−139.500**	**−1.297**	−1.539
	LL ecotype	0.32082	0.26344	1.66558	0.17804	0.51879		**−215.800**	**−1.264**	−1.326
	Collectivity	0.33737	0.23566	**2.16502**	0.55516	1.42504	–	**−279.022**	**−1.194**	−1.319
clpC1										
609 bp	HL ecotype	0.12786	0.11224	1.05996	−0.23405	0.02695	**12.83600**	**−47.000**	**−1.326**	**−1.771**
	LL ecotype	0.20267	0.16160	**1.93825**	0.52232	0.87753		**−59.300**	**−1.188**	−0.912
	Collectivity	0.20633	0.14257	**2.23404**	0.44049	1.00646	–	**−82.667**	**−1.231**	−1.147
clpP										
2580 bp	HL ecotype	0.15600	0.13660	1.08629	−0.17395	0.08520	**114.13500**	**−235.300**	**−1.331**	**−1.731**
	LL ecotype	0.21472	0.17349	1.81745	0.45634	0.79495		**−276.900**	**−1.221**	−1.059
	Collectivity	0.22664	0.15444	**2.34605**	0.25817	0.88434	–	**−368.089**	**−1.187**	−1.138

π is the mean number of pairwise differences per site. θ_W indicates Watterson's estimator. D, D^*, F^*, G, H and E indicate the tests of Tajima's D, McDonald and Kreitman's G, Fay and Wu's H and Zeng–Fu–Shi–Wu's E, respectively. Significant results at the 5% level or less are in bold font.

ferences between species, indicated that all three genes were experiencing selective constraints between the HL and LL populations. Tajima's D tests, which evaluate the normalized difference between π and θ_W of the three genes in the *Prochlorococcus* strains as a whole, showed significant positive values. This indicated that there were balancing selections in the three genes as a complete unit of *Prochlorococcus*. In addition, Tajima's D value of LL-*clpC1* was significant as well, because π and θ_W of the LL were higher than those of the HL population, and the two genes are much shorter than the others. Among Tajima's D and Fu–Li's tests, the D^* of *clpC1* and *clpP* in the HL population had conspicuous negative values, although these were not sufficiently significant.

Fu and Li's D test shares much information with Tajima's D test. However, for certain population genetic scenarios, particularly selective sweeps, this test is more sensitive than Tajima's D test [28]. Due to the selective sweeps of LL-*clpC1* and LL-*clpP* detected by McDonald and Kreitman's G tests, we accepted the results of Fu and Li's D^*, which then indicated potential positive selection (selective sweep) or an expanding population. To confirm the selection on the two Clp genes and to avoid the disturbance of a selective sweep, we *a priori* used Zeng–Fu–Shi–Wu's E Test, a recently developed statistical test for detecting positive selection by utilizing high-frequency variants. This test is most powerful in detecting the recovery phase after loss, which relies on the difference between θ_L and Watterson's θ_W genetic diversity, and which includes the postselective sweep phase [32]. Zeng–Fu–Shi–Wu's E tests were carried out with the purpose of testing for an excess of recent mutations or rare alleles (Left-side test), because the HL was the later diverged population. These tests verified the logical reasoning and showed that these two HL-Clp genes had significant values, which indicated recent positive selection not only on *clpC1* ($p = 0.033$, *alpha* $0.05 = -1.662$) but also on *clpP* ($p = 0.042$, *alpha* $0.05 = -1.653$). Nevertheless, the other genes, including LL-Clp genes, showed no significant E values because they had a longer evolutionary history, while E tests lack power under a balancing selection condition [32].

Little evidence was found to support the hypothesis of positive selection on CAO. Both HL- and LL-CAO show a neutral evolution, but CAO seems to be a balancing selection when viewed from the angle of genus level. We assumed that recombination might have promoted the

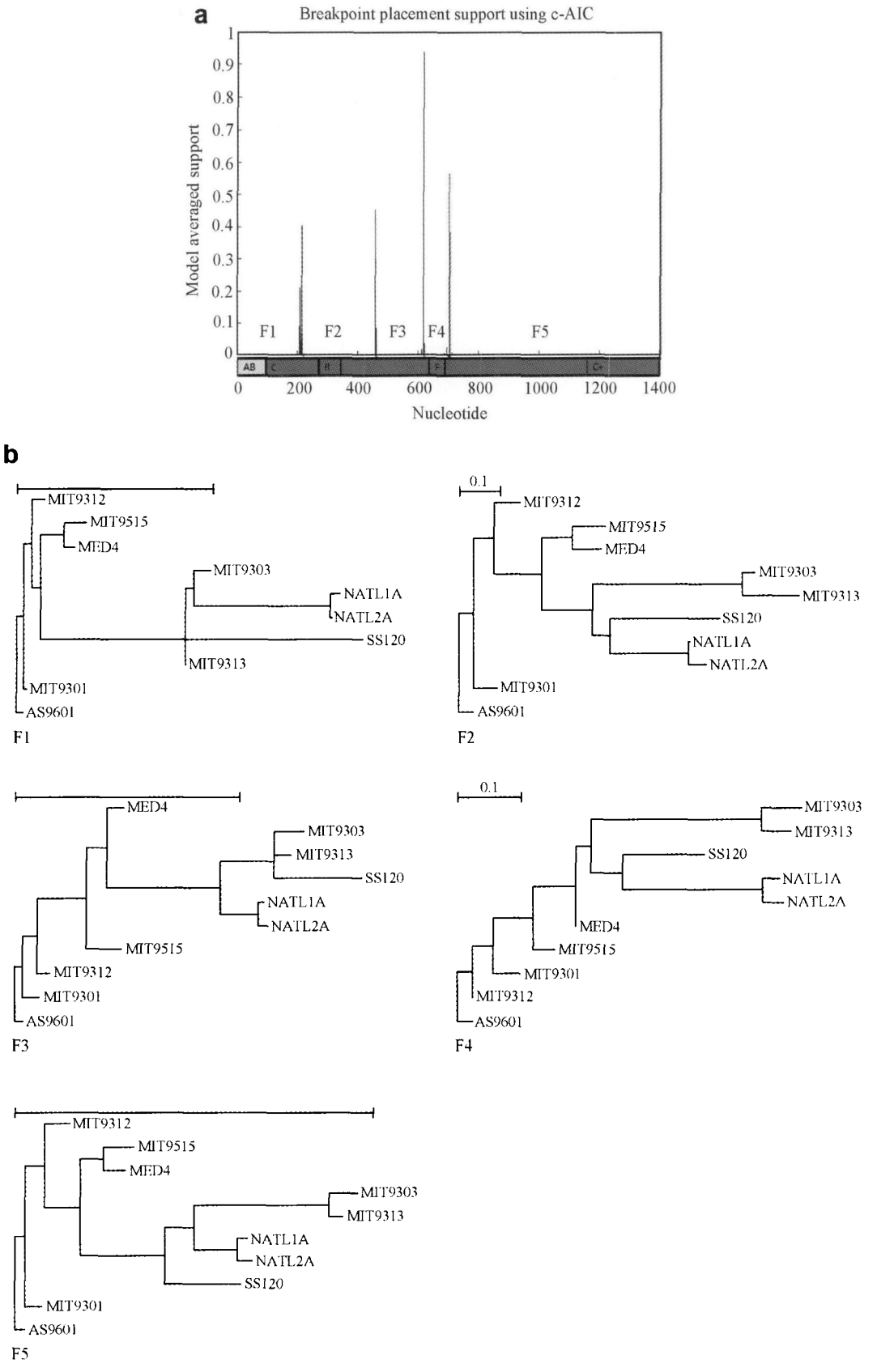

Fig. 3. Schematic structure of *Prochlorococcus* chlorophyll *a* oxygenase with inferred segment topologies. F1–F5 stand for the five recombinant fragments. (a) The position of breakpoints on the *Prochlorococcus* chlorophyll *a* oxygenase gene. The letters AB, C, R, F and C+ with ground color of different gray scales stand for the AB domains, C domain and C domain extension, respectively. The letters R and F stand for the Rieske center, the non-heme iron-binding motif region respectively. Breakpoint placement support using c-AIC is shown with the same length measurement. (b) Inferred topologies of the five recombinant fragments.

divergence of HL- and LL-CAO. This hypothesis was verified as described later.

3.3. Recombination and coevolution tests

Adopting the optimal model (001232) with AIC of 18458.8, GARD found evidence of four breakpoints on CAO with c-AIC = 18415.1. All of these breakpoints are located in the C domain. The Rieske center is comprised of a fragment (F2) and the non-heme iron-binding motif region is another individual fragment (F4). Inferred segment topologies showed that the LL strains always formed a clade that separated them from the HL strains, although the phylogenies and evolutionary rates of these fragments were different from each other (Fig. 3). These phylogenetic topologies are in accordance with previous research [39]. However, there was no evidence of positive selection in these segments at $p < 0.1$ (Null model, $\log(L) = -8398.99$; Alternative model, $\log(L) = -8398.99$; LRT = 0, p value = 1). Therefore, it can be concluded that recombination contributed to the adaptive evolution of CAO. Coevolution analysis showed a large group including a great number of coevolving sites on CAO. These sites were interspersed among the five fragments. Accordingly, these fragments were intensively correlated to each other during recombination (Fig. 4).

4. Discussion

The evolutionary history of *Prochlorococcus*, based on analyses of the Chl *b/a* ratio evolution, can be outlined here. The ancestor of *Prochlorococcus* was typically low-light adapted. Afterwards, *Prochlorococcus* gradually evolved and diverged. Some of the forms continued to accommodate to the environment of low light and these appeared to be in evolutionary stasis. On the other hand, the others began to carve out the competitive high-light environment. In this process, *Prochlorococcus* adapted to different irradiance conditions and formed a directional evolution. As a result, *Prochlorococcus* presents adaptive radiation along the vertical water columns of gradations in irradiance and formed two main ecotypes.

CAO had a more significant π and θ_W than the two genes encoding Clp protease. This may possibly be attributable to different evolution rates under different selective constraints during a similar period among the three genes. Therefore, Clp might have an excess of recent mutations or rare alleles. These results also suggest that the geographical differentiations within both lineages, if they exist, are much weaker than those between them. In addition, this would be in accordance with the evolutionary history and phylogenetic analyses speculated earlier, which suggests that the HL population diverged later.

Most substitutions that have occurred in *Prochlorococcus* have to be selectively neutral, as the large size of populations imposes low genetic drift and strong purifying selection [40]. The *Prochlorococcus* group is not currently experiencing higher levels of genetic drift [41]. As a result, CAO evolution is neutral and Clp protease is in the recovery phase after the loss of genetic diversity for either ecotype. However, there is remarkable recent evidence for positive selection of the Clp genes in HL *Prochlorococcus*.

Recombination in chlorophyllide *a* oxygenase (CAO) and positive selection for Clp protease contributed to the directional evolution of *Prochlorococcus*. The recombinant fragments of CAO are correlated with a large group of shared coevolving sites. CAO consists of A, B and C domains. The C domain alone has the catalytic function, and all of the breakpoints are located in the C domain. Evidence for positive selection is found in both subunits of Clp in high-light-adapted strains.

It would be expected that HL- and LL-phycoerythrin would be under different selective patterns [17,18], but phycoerythrin had lower genetic polymorphism in comparison with CAO and Clp protease. We assumed that CAO changed its catalytic function by recombination during

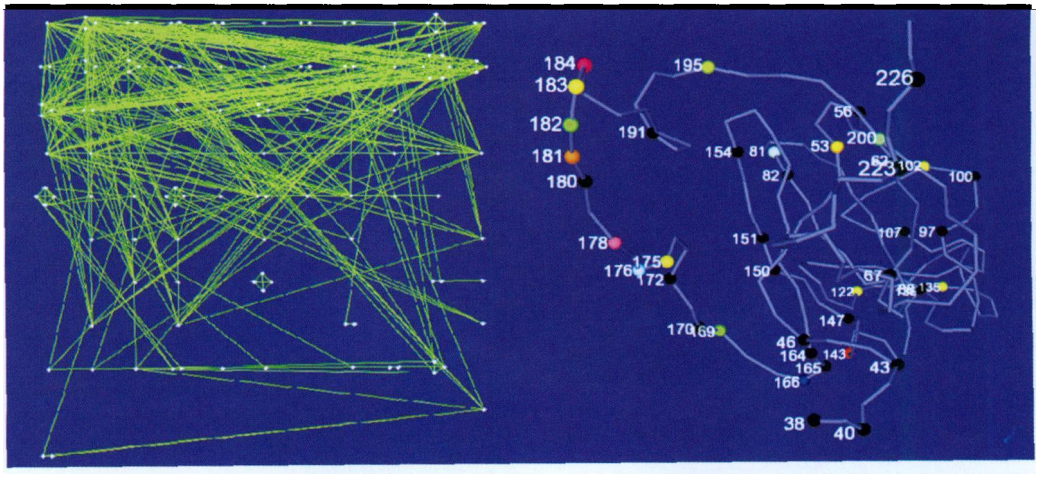

Fig. 4. Intramolecular analyses of *Prochlorococcus* CAO. (left) The relationship network of coevolving pairs of sites; (right) 3D display of coevolutionary sites on the CAO peptides, balls of the same color stand for coevolving pairs of sites in the same group.

directional evolution and that Clp protease evolved along with the structural changes in CAO. Due to the significance of Chl *b* to *Prochlorococcus* and of the Chl *b/a* ratio to its vertical ecological distribution in the euphotic layer [2,3], the molecular evidence for Chl *b/a* ratio evolution shows that this was one of the crucial reasons that explains how *Prochlorococcus* became the dominant photosynthetic organism in the ocean.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 30670165), the Key Innovative Project of the Chinese Academy of Sciences (Grant Nos. KZCX2-YW-209 and KZCX2-YW-216), and the CAS/SAFEA International Partnership Program for Creative Research Teams.

References

[1] Rüdiger W. Biosynthesis of chlorophyll *b* and the chlorophyll cycle. Photosynth Res 2002;74:187–93.
[2] Partensky F, Hess WR. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiol Mol Biol Rev 1999;63:106–27.
[3] Satoh S, Tanaka A. Identification of chlorophyllide *a* oxygenase in the *Prochlorococcus* genome by a comparative genomic approach. Plant Cell Physiol 2006;47:1622–9.
[4] Moore LR, Chisholm SW. Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. Limnol Oceanogr 1999;443:628–38.
[5] Scanlan D. Cyanobacteria: ecology, niche adaptation and genomics. Microbiol Today 2001;28:128–30.
[6] Tanaka A, Ito H, Tanaka R, et al. Chlorophyll *a* oxygenase CAO is involved in chlorophyll *b* formation from chlorophyll *a*. Proc Natl Acad Sci USA 1998;95:12719–23.
[7] Beale SI. Enzymes of chlorophyll biosynthesis. Photosynth Res 1999;60:43–73.
[8] Kettler GC, Martiny AC, Huang K, et al. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet 2007;3:e231.
[9] Karp PD, Paley S, Romero P. The Pathway Tools software. Bioinformatics 2002;18:225–32.
[10] Nakagawara E, Sakuraba Y, Yamasato A, et al. Clp protease controls chlorophyll *b* synthesis by regulating the level of chlorophyllide *a* oxygenase. Plant J 2007;49:800–9.
[11] Oster U, Tanaka R, Tanaka A, et al. Cloning and functional expression of the gene encoding the key enzyme for chlorophyll *b* biosynthesis CAO from *Arabidopsis thaliana*. Plant 2000;21:305–10.
[12] Harper AL, Gesjen SE, Linford AS, et al. Chlorophyllide *a* oxygenase mRNA and protein levels correlate with the chlorophyll *a/b* ratio in *Arabidopsis thaliana*. Photosynth Res 2004;79:149–59.
[13] Xu H, Vavilin D, Vermaas W. Chlorophyll *b* can serve as the major pigment in functional photosystem II complexes of cyanobacteria. Proc Natl Acad Sci USA 2001;98:14168–73.
[14] Nagata N, Satoh S, Tanaka R, et al. Domain structures of chlorophyllide *a* oxygenase of green plants and *Prochlorothrix hollandica* in relation to catalytic functions. Planta 2004;2186:1019–25.
[15] Flynn JM, Neher SB, Kim YI, et al. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. Mol Cell 2003;11:671–83.
[16] Mary I, Tu CJ, Grossman A, et al. Effects of high light on transcripts of stress-associated genes for the cyanobacteria *Synechocystis* sp. PCC 6803 and *Prochlorococcus* MED4 and MIT 9313. Microbiology 2004;150:1271–81.
[17] Qin S, Zhao F, Tseng C. Evidence for positive selection in phycoerythrin genes of red algae and cyanobacteria *Prochlorococcus* and *Synechococcus*. Photosynthetica 2005;43(1):141–6.
[18] Zhao FQ, Qin S. Comparative molecular population genetics of phycoerythrin locus in *Prochlorococcus*. Genetica 2007;129:291–9.
[19] Steglich C, Mullineaux CW, Teuchner K, et al. Photophysical properties of *Prochlorococcus* marinus SS120 divinyl chlorophylls and phycoerythrin *in vitro* and *in vivo*. FEBS Lett 2003;553:79–84.
[20] Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–8.
[21] Pagel M, Meade A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am Nat 2006;167:808–25.
[22] Pagel M, Meade A, Barker D. Bayesian estimation of ancestral character states on phylogenies. Syst Biol 2004;53:673–84.
[23] Pagel M. Inferring evolutionary processes from phylogenies. Zool Scr 1997;26:331–48.
[24] Pagel M. Inferring the historical patterns of biological evolution. Nature 1999;401:877–84.
[25] Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 1979;76:5269–73.
[26] Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol 1975;7:256–76.
[27] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989;123:585–95.
[28] Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics 1993;133:693–709.
[29] Hudson RR. Gene genealogies and the coalescent process. Oxf Surv Evol Biol 1990;7:1–44.
[30] Rozas J, Sánchez-DelBarrio JC, Messeguer X, et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 2003;19:2496–7.
[31] Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 1997;147:915–25.
[32] Zeng K, Fu YX, Shi S, et al. Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics 2006;174:1431–9.
[33] Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics 2000;155:1405–13.
[34] McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 1991;351:652–4.
[35] Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics 2005;21(5):676–9.
[36] Pond SLK, Posada D, Gravenor MB, et al. GARD: a genetic algorithm for recombination detection. Bioinformatics 2006;22(24):3096–8.
[37] Scheffler K, Martin DP, Seoighe C. Robust inference of positive selection from recombining coding sequences. Bioinformatics 2006;22:2493–9.
[38] Fares MA, David M. CAPS: coevolution analysis using protein sequences. Bioinformatics 2006;22(22):2821–2.
[39] Luo H, Shi J, Arndt W, et al. Gene order phylogeny of the genus *Prochlorococcus*. PLoS ONE 2008;3(12):e3837.
[40] Dufresne A, Garczarek L, Partensky F. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol 2005;6(2):R14.
[41] Hu J, Blanchard JL. Environmental sequence data from the Sargasso sea reveal that the characteristics of genome reduction in *Prochlorococcus* are not a harbinger for an escalation in genetic drift. Mol Biol Evol 2009;26:5–13.