Molecular Phylogenetics and Evolution 62 (2012) 206–213

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Codon usage patterns and adaptive evolution of marine unicellular cyanobacteria *Synechococcus* and *Prochlorococcus*

Tonghai Yu [a,1], Jinsong Li [a,1], Yang Yang [b], Liu Qi [a], Biaobang Chen [a], Fangqing Zhao [c], Qiyu Bao [a,*], Jinyu Wu [a,*]

[a] Institute of Genomic Medicine/Zhejiang Provincial Key Laboratory of Medical Genetics, Wenzhou Medical College, Wenzhou 325000, China
[b] School of Life Sciences, Peking University, Beijing 100871, China
[c] Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100871, China

ARTICLE INFO

Article history:
Received 23 January 2011
Revised 3 September 2011
Accepted 23 September 2011
Available online 21 October 2011

Keywords:
Synechococcus
Prochlorococcus
Codon usage patterns
Adaptive evolution

ABSTRACT

Marine unicellular cyanobacteria, represented by *Synechococcus* and *Prochlorococcus,* dominate the total phytoplankton biomass and production in oligotrophic ocean. In this study, we employed comparative genomics approaches to extensively investigate synonymous codon usage bias and evolutionary rates in a large number of closely related species of marine unicellular cyanobacteria. Although these two groups of marine cyanobacteria have a close phylogenetic relationship, we find that they are highly divergent not only in codon usage patterns but also in the driving forces behind the diversification. It is revealed that in *Prochlorococcus*, mutation and genome compositional constraints are the main forces contributing to codon usage bias, whereas in *Synechococcus*, translational selection. In addition, nucleotide substitution rate analysis indicates that they are not evolving at a constant rate after the divergence and that the average d_N/d_S values of core genes in *Synechococcus* are significantly higher than those in *Prochlorococcus*. Our evolutionary genomic analysis provides the first insight into codon usage, evolutionary genetic mechanisms and environmental adaptation of *Synechococcus* and *Prochlorococcus* after divergence.

1. Introduction

Prochlorococcus (Chisholm et al., 1988) and *Synechococcus* (Johnson and Sieburth, 1979) are similar both genetically and physiologically, but distinct in terms of photosynthetic apparatus, with the former employing chlorophyll a/b and the latter utilizing phycobilisomes as the light-harvesting antenna (Partensky et al., 1999). *Prochlorococcus* lives in a wide range of environments from the bottom of the euphotic zones to the upper layer of oligotrophic zones. However, *Synechococcus*, composed of at least seven lineages classified according to pigment properties and 16s–23s ribosomal internal transcribed spacer sequences (Huang et al., 2009; Rocap et al., 2002), has even more remarkable genetic diversities and higher rates of evolution. Owing to the breakthrough of high-throughput genome sequencing technologies, currently 22 marine *Synechococcus* and *Prochlorococcus* (with different physiological features) genomes are available on the Integrated Microbial Genomes (IMG) database (http://img.jgi.doe.gov/). Hence, marine unicellular cyanobacteria along with their available data resources can be used as excellent examples for investigating evolutionary mechanisms of genetic diversity and genome evolution across different species and their functional implications.

Synonymous codon usage has been documented in a wide range of organisms from prokaryotes to unicellular and multicellular eukaryotes, with similar patterns in closely related species. Ingvarsson (2008) has revealed that the five species of populus with close phylogenetic relationship have significantly different synonymous codon bias, resulting from the evolutionary pressure of mutation, genetic drift and natural selection in the process of divergence of genome composition. Vicario et al. (2007) and Pouwels and Leunissen (1994) have found a similar pattern in *Drosophila* and *Lactobacillus*, respectively. No such investigation has been conducted on cyanobacteria. Fortunately, the availability of a large number of newly sequenced genomes of marine unicellular cyanobacteria provides a wonderful opportunity to unveil the synonymous codon usage bias in them.

Genome-wide synonymous mutations are responsible for nonrandom patterns of synonymous codon usage, which directly testify the divergence of species evolution from their latest common ancestor. The rate of nonsynonymous mutation and synonymous mutation determine the adaptive evolution. If nonsynonymous mutations are favored by positive selection, they will be fixed at a higher rate than synonymous mutations, which is thus the

* Corresponding authors.
E-mail addresses: baoqy@genomics.org.cn (Q. Bao), iamwujy@yahoo.com.cn (J. Wu).
[1] These authors contributed equally to this work.

1055-7903/$ - see front matter Crown Copyright © 2011 Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.09.013

evidence for adaptive protein evolution (Yang and Nielsen, 2002). Mes et al. (2006) examined the distribution of synonymous and nonsynonymous substitutions in 12 genes of natural populations of cyanobacteria to infer functional changes. Zhao and Qin (2006) found that positive selection may drive the diversification of phycobiliproteins in cyanobacteria and that two ecotypes of *Prochlo*rococcus may follow two distinct evolutionary patterns in phycoerythrin gene locus. These studies, however, were all based on a small number of genes from various cyanobacterial genomes, thus failed to provide a whole picture of genome evolution in closely related species.

In this study, using comparative genomics approaches, we extensively investigated the synonymous codon usage bias and evolutionary rates in a large number of closely related species of marine unicellular cyanobacteria. Our results show that the codon usage patterns of *Prochlorococcus* and *Synechococcus* are quite different despite their close phylogenetic relationship. In addition, we find that *Synechococcus* has a faster average evolutionary rate than *Prochlorococcus* and that a number of genes involved in metabolism, DNA repairing and translation are under positive selection.

2. Materials and methods

2.1. Data sources

The predicted genes and proteins of the available 22 *synechococcus* and *prochlorococcus* genomes were downloaded from the IMG database (http://img.jgi.doe.gov/). Orthologous groups were identified using the OrthoMCL program (Li et al., 2003) by default, which has been proved to be very powerful to infer orthologous families from multiple genomes and has been widely adopted in related studies. Among the identified families, only those with one-to-one orthologous (defined as the core-set genes) relationship from the 22 cyanobacterial genomes were included for further analyses. To minimize sampling errors, genes less than (or equal to) 100 codons or containing internal stop codons were excluded. Finally, the core set, comprising of 1115 genes for each species, was further analyzed.

2.2. Multiple sequence alignment and phylogenetic analysis

Amino acid alignment was carried out using the ClustalW program with default settings (Thompson et al., 1994), and the corresponding nucleotide sequences were then aligned following the same gap patterns using the tranalign program implemented in the EMBOSS package (Rice et al., 2000).

The phylogenetic tree was constructed based on the 16S rRNAs using the neighbor-joining (NJ) method, Maximum-likelihood (ML) and Bayesian methods implemented in the MEGA 4.0 (Tamura et al., 2007), PHYML v3.0 (Guindon and Gascuel, 2003) and MrBayes v3.1.2 (Huelsenbeck et al., 2001), respectively, For ML and Bayesian tree constructions, the optimal nucleotide substitution model was chosen using Akaike information criterion (AIC) implemented in ModelTest program (Posada, 2009). The reliability of the NJ tree and ML reconstructions were evaluated with 1000 replicates of bootstrapping test, and only high bootstrap values ($\geqslant$50%) were shown on the branches.

2.3. Codon usage analysis

In order to elucidate codon usage bias in different cynaobacterial genomes, a number of indices, including RSCU, ENc, CAI, CBI, F_{op} and GC_{3s}, were measured using the codonW 1.42 program (http://codonw.sourceforge.net/). Among them, RSCU (relative synonymous codon usage), defined as the ratio of the observed frequency of codons to the expected frequency with the support of all the synonymous codons that are used equally (Sharp and Li, 1986), was calculated for all the protein coding sequences. RSCU values more than 1.0 indicate that the corresponding codon is used more frequently than expected, whereas less than 1.0 means the reverse (Sau and Deb, 2009). ENc (effective number of codons) was used to measure the magnitude or strength of codon bias for an individual gene, yielding values ranging from 20 for a gene with an extreme bias using only one codon per amino acid to 61 for a gene with no bias in using synonymous codons (Wright, 1990). CAI (codon adaptation index) (Sharp and Li, 1987), CBI (codon bias index) and F_{op} (frequency of optional codons) are three indices of directional codon usage relative to a subset of pre-defined reference optimal codons for a species. The identification of the reference set for each species in this study was done by examining the highly expressed genes (low ENc), such as ribosomal proteins. Thus, these three indices measure deviation from the optimum codon usage pattern defined 0, the furthest from the optimal set, meaning no optimal codons are used and less bias, and 1 indicating only the use of optimal codons and therefore a stronger codon usage bias, while GC_{3s} is the frequency of G+C at the third synonymously variable coding position, excluding Met, Trp and termination codons.

All statistical analyses were performed with the Matlab 2008b package. The correlation between GC_{12s} (represents the average frequency of the nucleotide G+C at the one and two synonymous codon positions) and GC_{3s} among genes was analyzed using a nonparametric Spearman's rank correlation analysis.

2.4. Substitution rates calculation

For each orthologous gene set, the pairwise d_N (the number of nonsynonymous substitutions per nonsynonymous site) and d_S (the number of synonymous substitutions per synonymous site) were estimated using the yn00 program in PAML (Yang, 1997). To identify specific genes subjected to positive selection, the maximum-likelihood method of Nielsen and Yang implemented in the codeml program was applied (Yang, 1997). Then, site-specific models M7 and M8 were used to compare the fitness of two nested models to the data. In brief, Model M7 assumes a beta distribution over the interval (0,1) and therefore does not allow for sites with $\omega > 1$, providing a flexible null hypothesis for testing positive selection. Model M8 adds an extra class of sites to M7 and is used to estimate relative rates ($\omega = d_N/d_S$). An LRT (likelihood ratio test) analysis is conducted to compare M7 with M8, and the level of significance is calculated as twice the difference of the likelihood scores (2Δ lnL). The functional category of each positively selected gene was obtained at the COG database (Zhaxybayeva et al., 2009) by BLASTing with an *E*-value of 10^{-5}.

3. Results and discussion

3.1. Uneven codon usage patterns in Synechococcus and Prochlorococcus

The pattern of codon usage bias is an effective indication of species environmental adaptation at the molecular level. It is species-specific to enhance the translation speed and accuracy, for instance, to minimize mismatches or frame shifting errors during translation (Huang et al., 2009). *Synechococcus* and *Prochlorococcus* have made great contributions to earth's photosynthetic biomass. In the present study, we investigated for the first time all of the available marine *Synechococcus* and *Prochlorococcus* genomes to reveal the patterns of their codon usage and to shed more lights on the evolution of the genetic codes in cyanobacteria.

Surprisingly, although the two marine cyanobacteria *Synec hococcus* and *Prochlorococcus* are phylogenetically close (Zhaxybayeva et al., 2009; Zwirglmaier et al., 2008), they distinct dramatically in both genomic characteristics and codon usage patterns (Fig. 1). The genome size, GC content, GC_{3s}, CAI, CBI, F_{op} and PC-GC_{3s} (the frequency of G + C at the third position in optimal codons) of *Synechococcus* are all remarkably higher (*t* test, $p < 0.001$) than those of *Prochlorococcus*. All these observations indicate that *Synechococcus* and *Prochlorococcus* have undergone great changes after their divergence.

Concretely, by investigating the changes of GC content, GC_{3s} and PC-GC_{3s} in the 22 genomes, we found that *Prochlorococcus* has a genome-wide biased mutational orientation, i.e., from G/C to A/T, whereas *Synechococcus* has an opposite mutational orientation, i.e., from A/T to G/C. Previous study indicates that the genome reduction of *Prochlorococcus* has affected its GC content and protein evolution (Dufresne et al., 2005). Here, in the view of codon usage, it is indicated that the divergence of genomic features of *Prochlorococcus* and *Synechococcus* results in their dissimilar G/C content in codons and has a direct impact on codon usage. Based on ENc observation, we found that different cyanobacteria species have different ENc values, ranging from the lowest 38.601 in *Synechococcus* sp. RS 9917 to the highest 51.837 in *Prochlorococcus marinus* MIT 9303. Such observation also strongly suggests that the overall intensity of cyanobacteria codon usage bias differs sharply. However, ENc is nondirectional, which is used only to estimate the overall intensity of codon usage bias for an individual gene and thus cannot reveal which codons are preferred. Therefore, there is a possibility that genes may have different codon usage bias patterns even if with equal ENc values. In this study, three directional indices, CAI, CBI and F_{op}, were measured to complement ENc. We found that all these three directional indices of *Synechococcus* are significantly higher than those of *Prochlorococcus* (*t* test, $p < 0.001$), suggesting that *Synechococcus* prefers optimal codons and has a stronger bias than *Prochlorococcus*. However, the CBI values for *Prochlorococcus* genes (except P9313 and P9303) are all less than 0, indicating that the majority of genes in those species of *Prochlorococcus* prefer usage of more kinds of codons to usage of optimal codons.

Prochlorococcus sp. P9313 and P9303 are reported to use the same photosynthetic apparatus with other species of *Prochlorococcus*. In this study, interestingly, we observed that while their genomic features and patterns of codon usage bias, including genome size, GC_{3s} and CAI, are more similar to those of *Synechococcus*, other indices such as F_{op}, PC-GC_{3s} and CBI are somewhere between those of *Prochlorococcus* and *Synechococcus*. Such incongruence implied that the genomes of P9313 and P9303 may undergo homologous recombination or horizontal gene transfer with *Synechococcus* (Kettler et al., 2007). Given that *Prochlorococcus* and marine *Synechococcus* diverged over a very short period, another possibility of the incongruence is the rapid radiation of *Prochlorococcus* spp. followed by an incomplete lineage sorting of ancestral polymorphisms. Additionally, genome scale synteny analysis also revealed that P9313 and P9303 share much more conserved gene orders with *Synechococcus* spp. than with other *Prochlorococcus* spp. (data not shown). Thus, it is also indicated that P9313 and P9303 have undergone frequent introgression and then resulted its genome becoming more "*Synechococcus*-like" but still maintain the genes for its ecological niches (Zhaxybayeva et al., 2009).

Previous study suggested that synonymous codons usage bias is idiosyncratical among organisms, but optimal codon shift in close species is quite steady during the process of evolution. Through a straightforward and effective method, we identify optimal codons according to the values of RSCU among the 22 marine unicellular cyanobacteria. We found that *Prochlorococcus* and *Synechococcus* have employed different optimal codons to adapt to their environments despite their close relationships (Fig. 2). For example, in *Prochlorococcus,* the amino acids of arginine, leucine, valine, threonine, serine, praline, isoleucine, glycine and glutamine are

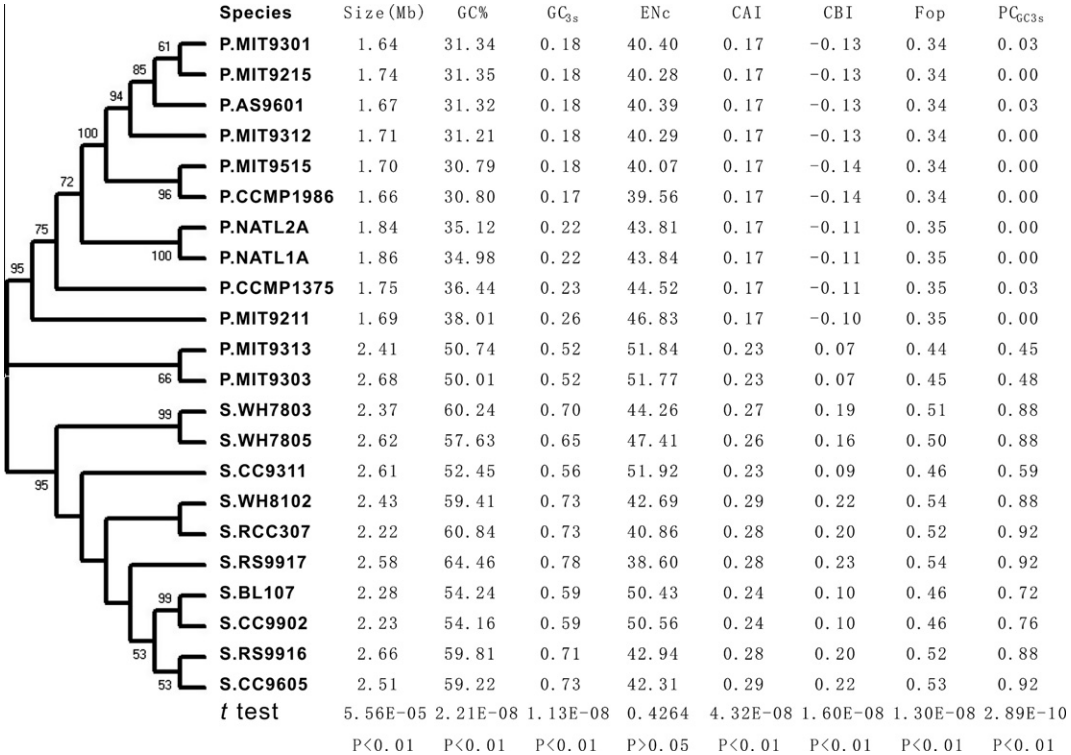
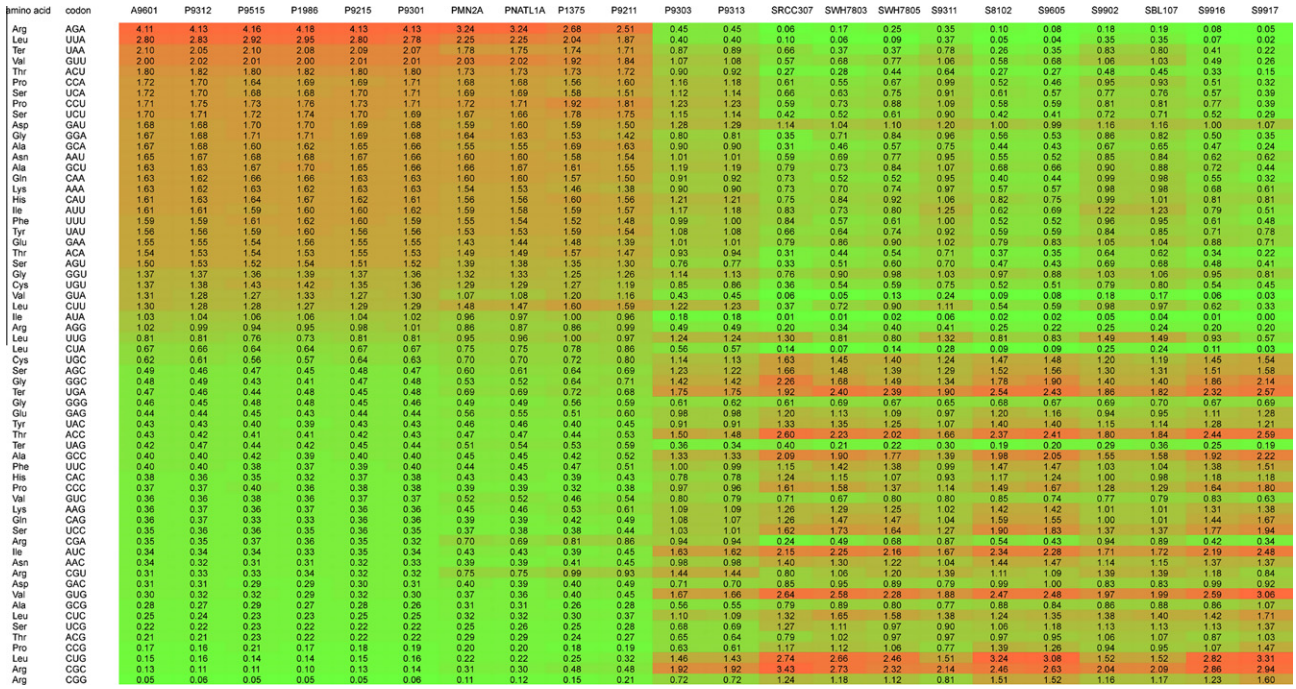


Fig. 1. Genomic characteristics and codon usage patterns in marine *Synechococcus* and *Prochlorococcus*. The organism tree was constructed based on the 16S RNA sequences of 22 *Synechococcus* and *Prochlorococcus* organisms. We choose the representative tree constructed by NJ methods. Bootstrap support values above 50% are shown on the nodes. The organism tree based on the 16S RNA with branch length, as well as the phylogenetic tree constructed by ML and Bayesian method, is shown in Fig. S1.



amino acid	codon	A9601	P9312	P9515	P1986	P9215	P9301	PMN2A	PNATL1A	P1375	P9211	P9303	P9313	SRCC307	SWH7803	SWH7805	S9311	S8102	S9605	S9902	SBL107	S9916	S9917
Arg	AGA	4.11	4.13	4.16	4.18	4.13	4.13	3.24	3.24	2.68	2.51	0.45	0.45	0.06	0.17	0.25	0.35	0.10	0.08	0.18	0.19	0.08	0.05
Leu	UUA	2.80	2.83	2.92	2.95	2.80	2.78	2.25	2.25	2.04	1.87	0.40	0.40	0.10	0.06	0.09	0.37	0.05	0.04	0.35	0.35	0.07	0.02
Ter	UAA	2.10	2.05	2.10	2.08	2.09	2.07	1.78	1.75	1.74	1.71	0.87	0.89	0.66	0.37	0.37	0.78	0.26	0.35	0.83	0.80	0.41	0.22
Val	GUU	2.00	2.02	2.01	2.00	2.01	2.01	2.03	2.02	1.92	1.84	1.07	1.08	0.57	0.68	0.77	1.06	0.58	0.68	1.06	1.03	0.49	0.26
Thr	ACU	1.80	1.82	1.80	1.82	1.80	1.80	1.73	1.73	1.73	1.72	0.90	0.92	0.27	0.28	0.44	0.64	0.27	0.27	0.48	0.45	0.33	0.15
Pro	CCA	1.72	1.70	1.64	1.69	1.69	1.71	1.68	1.68	1.56	1.60	1.16	1.18	0.61	0.55	0.67	0.99	0.52	0.46	0.95	0.93	0.51	0.32
Ser	UCA	1.72	1.70	1.68	1.68	1.70	1.71	1.69	1.69	1.58	1.51	1.12	1.14	0.66	0.63	0.75	0.91	0.61	0.57	0.77	0.76	0.57	0.39
Pro	CCU	1.71	1.75	1.73	1.76	1.73	1.71	1.72	1.71	1.92	1.81	1.23	1.23	0.59	0.73	0.88	1.09	0.58	0.59	0.81	0.81	0.77	0.39
Ser	UCU	1.70	1.71	1.72	1.74	1.70	1.69	1.67	1.66	1.78	1.75	1.15	1.14	0.42	0.52	0.61	0.90	0.42	0.41	0.72	0.71	0.52	0.29
Asp	GAU	1.68	1.68	1.70	1.70	1.69	1.68	1.59	1.60	1.59	1.50	1.28	1.29	1.14	1.04	1.10	1.20	1.00	0.99	1.16	1.16	1.00	1.07
Gly	GGA	1.67	1.68	1.71	1.71	1.69	1.68	1.64	1.63	1.53	1.42	0.80	0.81	0.35	0.71	0.84	0.96	0.56	0.53	0.86	0.82	0.50	0.35
Ala	GCA	1.67	1.68	1.60	1.62	1.65	1.66	1.55	1.55	1.69	1.63	0.90	0.90	0.31	0.46	0.57	0.75	0.44	0.43	0.67	0.65	0.47	0.24
Asn	AAU	1.65	1.67	1.68	1.68	1.67	1.66	1.60	1.60	1.58	1.54	1.01	1.01	0.59	0.69	0.77	0.95	0.55	0.52	0.85	0.84	0.62	0.62
Ala	GCU	1.63	1.63	1.67	1.70	1.65	1.66	1.66	1.67	1.61	1.55	1.19	1.19	0.79	0.73	0.84	1.07	0.68	0.66	0.90	0.88	0.72	0.44
Gln	CAA	1.63	1.62	1.66	1.66	1.63	1.63	1.60	1.60	1.57	1.50	0.91	0.92	0.73	0.52	0.52	0.95	0.40	0.44	0.99	0.98	0.55	0.32
Lys	AAA	1.63	1.62	1.63	1.62	1.63	1.63	1.54	1.53	1.46	1.38	0.90	0.90	0.73	0.70	0.74	0.97	0.57	0.57	0.98	0.98	0.68	0.61
His	CAU	1.61	1.63	1.64	1.67	1.62	1.61	1.56	1.56	1.60	1.56	1.21	1.21	0.75	0.84	0.92	1.06	0.82	0.75	0.99	1.01	0.81	0.81
Ile	AUU	1.61	1.61	1.59	1.60	1.60	1.62	1.59	1.58	1.59	1.57	1.17	1.18	0.83	0.73	0.80	1.25	0.62	0.69	1.22	1.23	0.79	0.51
Phe	UUU	1.59	1.59	1.61	1.62	1.60	1.59	1.55	1.54	1.52	1.48	0.99	1.00	0.84	0.57	0.61	1.00	0.52	0.52	0.96	0.95	0.61	0.48
Tyr	UAU	1.56	1.56	1.59	1.60	1.56	1.56	1.53	1.53	1.59	1.54	1.08	1.08	0.66	0.64	0.74	0.92	0.59	0.59	0.84	0.85	0.71	0.78
Glu	GAA	1.55	1.55	1.54	1.56	1.55	1.55	1.43	1.44	1.48	1.39	1.01	1.01	0.79	0.86	0.90	1.02	0.79	0.83	1.05	1.04	0.88	0.71
Thr	ACA	1.54	1.53	1.54	1.53	1.55	1.53	1.49	1.49	1.57	1.47	0.93	0.94	0.31	0.44	0.54	0.71	0.37	0.35	0.64	0.62	0.34	0.22
Ser	AGU	1.50	1.53	1.52	1.54	1.51	1.52	1.39	1.38	1.35	1.30	0.76	0.77	0.33	0.51	0.60	0.70	0.47	0.43	0.69	0.68	0.48	0.41
Gly	GGU	1.37	1.37	1.36	1.39	1.37	1.36	1.32	1.33	1.25	1.26	1.14	1.13	0.76	0.90	0.98	1.03	0.97	0.88	1.03	1.06	0.95	0.81
Cys	UGU	1.37	1.38	1.43	1.42	1.35	1.36	1.29	1.29	1.27	1.19	0.85	0.86	0.36	0.54	0.59	0.75	0.52	0.51	0.79	0.80	0.54	0.45
Val	GUA	1.31	1.28	1.27	1.33	1.27	1.30	1.07	1.08	1.20	1.16	0.43	0.45	0.06	0.05	0.13	0.24	0.09	0.08	0.18	0.17	0.06	0.03
Leu	CUU	1.30	1.28	1.28	1.27	1.29	1.29	1.48	1.47	1.60	1.59	1.22	1.23	0.37	0.72	0.90	1.11	0.54	0.59	0.98	0.97	0.62	0.33
Ile	AUA	1.03	1.04	1.06	1.06	1.04	1.02	0.96	0.97	1.00	0.96	0.18	0.18	0.01	0.01	0.02	0.06	0.02	0.02	0.05	0.04	0.01	0.00
Arg	AGG	1.02	0.99	0.94	0.95	0.98	1.01	0.86	0.87	0.86	0.99	0.49	0.49	0.20	0.34	0.40	0.41	0.25	0.22	0.25	0.24	0.20	0.20
Leu	UUG	0.81	0.81	0.76	0.73	0.81	0.81	0.95	0.96	1.00	0.97	1.24	1.24	1.30	0.81	0.80	1.32	0.81	0.83	1.49	1.49	0.93	0.57
Leu	CUA	0.67	0.66	0.64	0.64	0.67	0.67	0.75	0.75	0.78	0.86	0.56	0.57	0.14	0.07	0.14	0.28	0.09	0.09	0.25	0.24	0.11	0.03
Cys	UGC	0.62	0.61	0.56	0.57	0.64	0.63	0.70	0.70	0.72	0.80	1.14	1.13	1.63	1.45	1.40	1.24	1.47	1.48	1.20	1.19	1.45	1.54
Ser	AGC	0.49	0.46	0.47	0.45	0.48	0.47	0.60	0.61	0.64	0.69	1.23	1.22	1.66	1.48	1.39	1.29	1.52	1.56	1.30	1.31	1.51	1.58
Gly	GGC	0.48	0.49	0.43	0.41	0.47	0.48	0.53	0.52	0.64	0.71	1.42	1.42	2.26	1.68	1.49	1.34	1.78	1.90	1.40	1.40	1.86	2.14
Ter	UGA	0.47	0.46	0.44	0.48	0.45	0.48	0.69	0.69	0.72	0.68	1.75	1.75	1.92	2.40	2.39	1.90	2.54	2.43	1.86	1.82	2.32	2.57
Gly	GGG	0.46	0.45	0.48	0.48	0.45	0.46	0.49	0.49	0.56	0.59	0.61	0.62	0.61	0.69	0.67	0.65	0.68	0.67	0.69	0.70	0.67	0.69
Glu	GAG	0.44	0.44	0.45	0.43	0.44	0.44	0.56	0.55	0.51	0.60	0.98	0.98	1.20	1.13	1.09	0.97	1.20	1.16	0.94	0.95	1.11	1.28
Tyr	UAC	0.43	0.43	0.40	0.39	0.43	0.43	0.46	0.46	0.40	0.45	0.91	0.91	1.33	1.35	1.25	1.07	1.40	1.40	1.15	1.14	1.28	1.21
Thr	ACC	0.43	0.42	0.41	0.41	0.42	0.43	0.47	0.47	0.44	0.53	1.50	1.48	2.60	2.23	2.02	1.66	2.37	2.41	1.80	1.84	2.44	2.59
Ter	UAG	0.42	0.47	0.44	0.42	0.45	0.44	0.51	0.54	0.53	0.59	0.36	0.34	0.40	0.21	0.22	0.30	0.19	0.20	0.29	0.36	0.25	0.19
Ala	GCC	0.40	0.40	0.42	0.39	0.40	0.40	0.45	0.45	0.42	0.52	1.33	1.33	2.09	1.90	1.77	1.39	1.98	2.05	1.55	1.58	1.92	2.22
Phe	UUC	0.40	0.40	0.38	0.37	0.39	0.40	0.44	0.45	0.47	0.51	1.00	0.99	1.15	1.42	1.38	0.99	1.47	1.47	1.03	1.04	1.38	1.51
His	CAC	0.38	0.36	0.35	0.32	0.37	0.38	0.43	0.43	0.39	0.43	0.78	0.78	1.24	1.15	1.07	0.93	1.17	1.24	1.00	0.98	1.18	1.18
Pro	CCC	0.37	0.37	0.40	0.36	0.38	0.38	0.39	0.39	0.32	0.38	0.97	0.96	1.61	1.58	1.37	1.14	1.49	1.67	1.28	1.29	1.64	1.80
Val	GUC	0.36	0.36	0.38	0.36	0.37	0.37	0.52	0.52	0.46	0.54	0.80	0.79	0.71	0.67	0.80	0.80	0.85	0.74	0.77	0.79	0.83	0.63
Lys	AAG	0.36	0.37	0.36	0.37	0.36	0.36	0.45	0.46	0.53	0.61	1.09	1.09	1.26	1.29	1.25	1.02	1.42	1.42	1.01	1.01	1.31	1.38
Gln	CAG	0.36	0.37	0.33	0.33	0.36	0.36	0.39	0.39	0.42	0.49	1.08	1.07	1.26	1.47	1.47	1.04	1.59	1.55	1.00	1.01	1.44	1.67
Ser	UCC	0.35	0.36	0.36	0.35	0.36	0.35	0.37	0.38	0.38	0.44	1.03	1.01	1.62	1.73	1.64	1.27	1.90	1.83	1.37	1.37	1.77	1.94
Arg	CGA	0.35	0.35	0.37	0.36	0.35	0.32	0.70	0.69	0.81	0.86	0.94	0.94	0.24	0.49	0.68	0.87	0.54	0.43	0.94	0.89	0.42	0.34
Ile	AUC	0.34	0.34	0.34	0.33	0.35	0.34	0.43	0.43	0.39	0.45	1.63	1.62	2.15	2.25	2.16	1.67	2.34	2.28	1.71	1.72	2.19	2.48
Asn	AAC	0.34	0.32	0.31	0.31	0.32	0.33	0.39	0.39	0.41	0.45	0.98	0.98	1.40	1.30	1.22	1.04	1.44	1.47	1.14	1.15	1.37	1.37
Arg	CGU	0.31	0.33	0.33	0.34	0.32	0.32	0.75	0.75	0.99	0.93	1.44	1.44	0.80	1.06	1.20	1.39	1.11	1.09	1.39	1.39	1.18	0.84
Asp	GAC	0.31	0.31	0.29	0.29	0.30	0.31	0.40	0.39	0.40	0.49	0.71	0.70	0.85	0.95	0.89	0.79	0.99	1.00	0.83	0.83	0.99	0.92
Val	GUG	0.30	0.32	0.32	0.29	0.32	0.30	0.37	0.36	0.40	0.45	1.67	1.66	2.64	2.59	2.28	1.88	2.47	2.48	1.97	1.99	2.59	3.06
Ala	GCG	0.28	0.27	0.29	0.27	0.28	0.26	0.31	0.31	0.26	0.28	0.56	0.55	0.79	0.89	0.80	0.77	0.88	0.84	0.86	0.88	0.86	1.07
Leu	CUC	0.25	0.24	0.23	0.23	0.25	0.25	0.32	0.32	0.30	0.37	1.10	1.09	1.32	1.65	1.58	1.38	1.24	1.35	1.38	1.40	1.42	1.71
Ser	UCG	0.22	0.22	0.23	0.22	0.22	0.22	0.25	0.26	0.25	0.28	0.68	0.69	1.27	1.11	0.97	0.90	1.06	1.18	1.13	1.13	1.13	1.37
Thr	ACG	0.21	0.21	0.23	0.22	0.22	0.22	0.29	0.29	0.24	0.27	0.65	0.64	0.79	1.02	0.97	0.97	0.97	0.95	1.06	1.07	0.87	1.03
Pro	CCG	0.17	0.16	0.21	0.17	0.18	0.19	0.20	0.20	0.18	0.19	0.63	0.61	1.12	1.12	1.06	0.77	1.39	1.26	0.94	0.95	1.07	1.47
Leu	CUG	0.15	0.16	0.14	0.14	0.15	0.16	0.22	0.22	0.25	0.32	1.46	1.43	2.74	2.66	2.46	1.51	3.24	3.08	1.52	1.52	2.82	3.31
Arg	CGC	0.13	0.11	0.11	0.10	0.13	0.14	0.31	0.30	0.48	0.48	1.92	1.92	3.43	2.73	2.32	2.14	2.46	2.63	2.04	2.09	2.86	2.94
Arg	CGG	0.05	0.06	0.05	0.05	0.05	0.06	0.11	0.12	0.15	0.21	0.72	0.72	1.24	1.18	1.12	0.81	1.51	1.52	1.16	1.17	1.23	1.60

Fig. 2. Individual codon usage bias in marine *Synechococcus* and *Prochlorococcus*. The colors in the figure indicate the gradient of codon usage bias, from the strongest bias (red) to the weakest bias (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

preferred to be encoded by AGA, UUA, GUU, ACU, CCA, UCA, UCU, AUU, GGU and CAA, respectively. However, in *Synechococcus,* they are prone to use CGC, CUG, GUG, ACC, CCC, AGC, AUA, GGC and CAG, respectively. In addition, the values of PC-GC_{3s} are also diverging acutely. The mean of PC-GC_{3s} in *Synechococcus* is 86.859%, whereas it is 8.9484% in *Prochlorococcus,* with PC-GC_{3s} being 50% and 46.667% in P9313 and P9303, respectively.

3.2. Different factors driving the bias of codon usage

Previous studies performed on a large number of divergent species show that there are several factors influencing the patterns of synonymous codon usage bias in a variety of organisms (Ingvarsson, 2008; Pouwels and Leunissen, 1994; Sharp and Li, 1986). Whereas among these numerous factors, including mutational bias, replicational and transcriptional selection and tRNA numbers (dos Reis et al., 2003), the balance of genomic compositional mutation, genetic drift and natural selection that affect gene translation is the major contribution, which has been widely accepted to account for codon usage variation. However, analysis of nine cyanobacteria shows that genetic drift is a weak evolutionary force. Huang et al. found that the evolutionary forces of *Prochlorococcus* result mainly from elevated mutation rates rather than genetic drift (Huang et al., 2009). While genome-wide directional mutation, especially the change of GC content, has a profound impact on the pattern of codon usage, selection forces across different steps of protein expression shape the usage of codons as well, especially for highly expressed genes (Sueoka, 1988).

To address whether the codon usage bias of *Prochlorococcus* and *Synechococcus* is caused by genome composition difference, GC_{3s} vs. ENc is calculated to investigate the determinants of codon usage variations among genes in two marine picocyanobacteria (Wright, 1990). Interestingly, the two groups of cyanobacteria form two independent clusters, indicating that the codon usages of the two groups are influenced differently by genomic mutation. By comparing the *Synechococcus* and two exceptional members of *Prochlorococcus*, we observed that the distribution of ENc plots of *Prochlorococcus* is more similar to the expected curve (Fig. 3A). Such observation suggests that codon usage bias in *Prochlorococcus* is impacted mainly by mutation and much influenced by genome compositional constraints than *Synechococcus*. However, it should be noted that only a small proportion of plots of *Synechococcus* and two exceptional *Prochlorococcus* lie on the expected curve, suggesting that, apart from the effect of genome compositional constraints, there might be some additional factors that influence the codon usage in these species.

The heterogeneity of codon usage among species was shaped by the equilibrium between neutral mutation pressure and selective constraints (Sueoka, 1988). To further investigate the effect caused by translational selection on the pattern of codon usage bias in both groups of cyanobacteria, neutrality plots (GC_{12s} vs. GC_{3s}) were analyzed for the 22 marine cyanobacteria genomes. Because of the discrepancy of GC content distribution in those species, it can be seen that the 10 *Synechococcus* fall on the upper right corner (Fig. 3A), whereas the *Prochlorococcus* fall on the lower left corner. In addition, the two exceptional *Prochlorococcus* (P9303 and P9313) genomes show similar GC content distribution with *Synechococcus*. However, what we concern about are the correlations between GC_{3s} and GC_{12s} for each gene. Sueoka suggested that if there are significant correlations and the slope of the regression line is close to 1, it indicates that those genes are undergoing the similar impact at the different positions of codons and the codon usage pattern is mainly shaped by mutation pressure (Sueoka, 1988). In contrast, if translational selection is the dominant factor, then selection would go against mutational bias, counteracting the correlation between GC_{12s} and GC_{3s}, and further forcing the slope of the regression line close to 0 (Kawabe and Miyashita, 2003). We can observe that genes of ten *Prochlorococcus* genomes have significant correlations between the different positions of codons (Fig. 3B), indicating there is a strong correlation between GC mutation bias in the 10 *Prochlorococcus* species and GC content in each of three codon positions. In other words, the pattern of codon usage bias of the *Prochlorococcus* results mainly from genomic compositional mutation pressure, which is consistent with

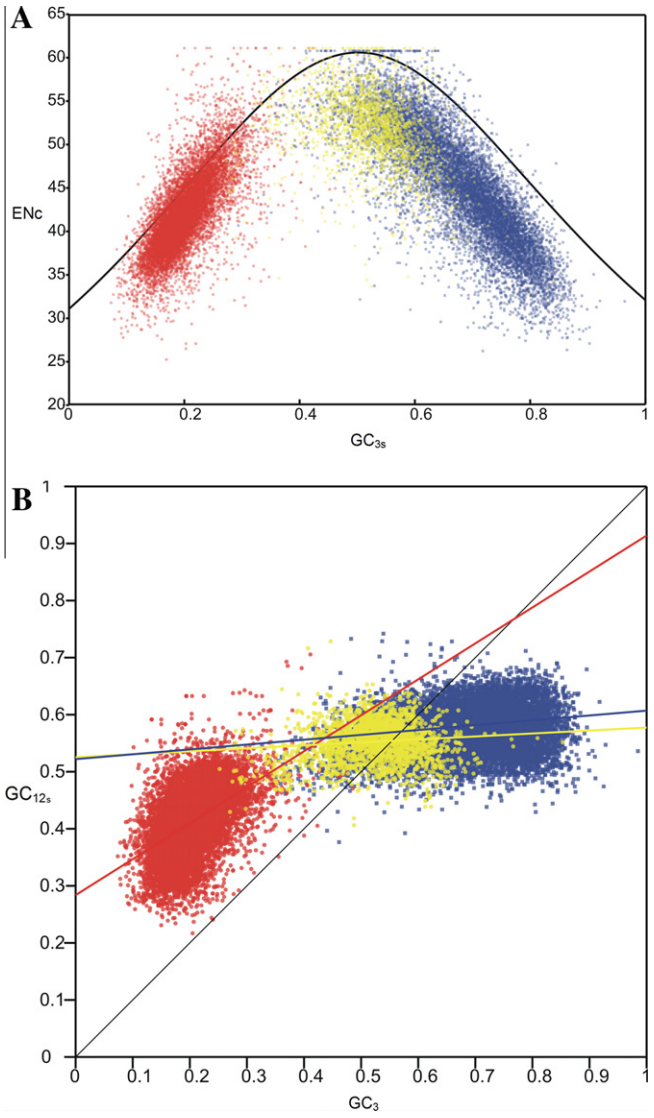


Fig. 3. (A) Nc-plot (ENc vs. GC_{3s}) of the two marine picocyanobacteria. The continuous curve (the expected curve) represents codon usage bias of genes, which is determined by GC_{3s} content alone. (B) Neutrality plot (GC_{12s} vs. GC_{3s}). The regression line of *Prochlorococcus* (red) is $y = 0.63017x + 0.28378$, $r = 0.4945$, $p < 0.01$, the regression line of P9313 and P9303 (yellow) is $y = 0.051971x + 0.52545$, $r = 0.0425$, $p < 0.05$, and the regression line of *Synechococcus* (blue) is $y = 0.085005x + 0.52209$, $r = 0.1740$, $p < 0.01$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

previous findings. The potential reasons for genomic compositional mutation may be due to its adapted habit that protects *Prochlorococcus* from some extrinsic mutagens like UV-B and low growth rate of *Prochlorococcus* (Partensky et al., 1999). While the correlation between GC_{12s} and GC_{3s} of *Synechococcus* and those of P9313 and P9303 is unobvious, there is low mutation bias or high conservation level of GC contents throughout the genome. Therefore, it can be concluded that the codon usage of *Synechococcus* and two exceptional members of *Prochlorococcus* (P9303 and P9313) is dominated by high-level translational selection, which counteracts the effect caused by nucleotide. Although *Synechococcus* and *Prochlorococcus* are coexists in the environment, *Synechococcus* is considered to be more of a generalist than *Prochlorococcus* with the ability to grow over a broader range of nutrient concentrations and temperatures (Moore et al., 1998). Such observation indicated that translational selection severing as a major driving force behind the codon usage of *Synechococcus* will make it more adaptive to environment. Indeed, it is indicated that the GC-rich *Synechococcus* can utilize the full repertoire of 40 tRNAs at most, while *Prochlorococcus* can use only a subset of them (Limor-Waisberg et al., 2011).

3.3. Nucleotide substitution and adaptive evolution

Marine *Synechococcus* and *Prochlorococcus* are indicated to evolve from the same ancestor through a near-simultaneous diversification (Urbach et al., 1998) and share many conserved core-set genes in comparison with other cyanobacteria, such as *Synechocystis* PCC 6803 and *Anabaena* sp. PCC 7120. However, both genome phylogeny and nucleotide substitution rate indicate that these two groups of marine cyanobacteria are not evolving at a constant rate after the divergence (Huang et al., 2009). In this study, we found that the nucleotide substitution rates for the two lineages of cyanobacteria have a large excess of synonymous over nonsynonymous substitutions in the core-set genes of *Synechococcus* and *Prochlorococcus* (Fig. 4). Plot of nucleotide substitutions indicated a significant difference in the frequency spectra and the slope of the linear relationship between d_N and d_S. The distribution of d_S in *Synechococcus* had a peak around 1.8 and then a long tail, with the mean synonymous substitution rate of 2.133 ± 0.882, which is significantly lower than that of *Prochlorococcus* (2.978 ± 1.303; t test, $p < 0.001$). It is suggested that the high rate of synonymous substitution in *Prochlorococcus* may result from recent relaxation of selective constraints on codon usage pattern (Morton, 1997). Similar pattern is also found in d_N. The nonsynonymous substitutions of *Synechococcus* exhibited a much smooth distribution from 0.1 to 0.4, with an average rate of 0.202 ± 0.148, indicating the similar heterogeneity level of all the sampled *Synechococcus* strains.

The d_N/d_S ratio is often mostly used to compare the strength of positive selection on different genes. We found the average d_N/d_S values among orthologous for the core-set genes in *Synechococcus* (0.104 ± 0.084) are significantly higher than those of *Prochlorococcus* (0.080 ± 0.059; t test, $p < 0.001$). Such evidence strongly indicates that *Synechococcus* has undergone the episodic accelerated evolution and afterward is still under strong selective constraints. If nonsynonymous mutations are favored by positive selection, they will be fixed at a higher rate than synonymous mutations, which is the evidence for adaptive protein evolution. Based on the maximum-likelihood method of Yang and Nielsen (2002), we found that there are 129 core-set genes that may have undergone positive selection in *Prochlorococcus* and *Synechococcus* (Fig. 5). Functional classification of accelerated evolving genes based on the Cluster of Orthologous Groups (COGs) database shows that they are present in all primary functions, and most of them are related to metabolism, DNA repairing and translation. For example, we found a signature of adaptive evolution in the aminoacyl-tRNA synthetase, which is a key component of the protein translation machinery that catalyzes the esterification of a specific amino acid to its compatible cognate tRNA to form an aminoacyl-tRNA. It contains a conserved core domain involved in ATP binding and hydrolysis and combines with additional domains determining the specificity of interactions with the cognate amino acid and tRNA (Cusack, 1997). It has been proved that site-directed mutation of certain amino acids in this domain can result in the loss (or the decrease) of capacity to efficiently recognize and aminoacylate tRNA (Brevet et al., 2003; Feng et al., 2005; Kettler et al., 2007). In addition, we also observed strong signals of positive selection of glutathione transferase, which is involved in cellular defense against toxic electrophiles of both exogenous and endogenous origins. Previous report has shown that glutathione transferase was subject to rapid adaptive evolution in human through which elevated nonsynonymous substitutions were capable of driving

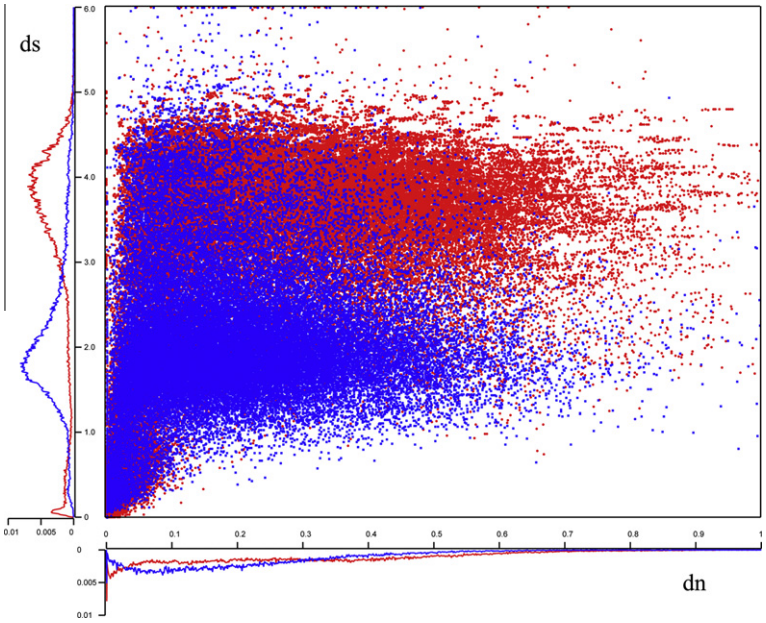


Fig. 4. Comparison of nucleotide substitution rates in the core-set genes of marine cyanobacteria. Plot of the rate of nonsynonymous substitutions (d_N) against the rate of synonymous substitutions (d_S) in *Prochlorococcus* (red) and *Synechococcus* (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

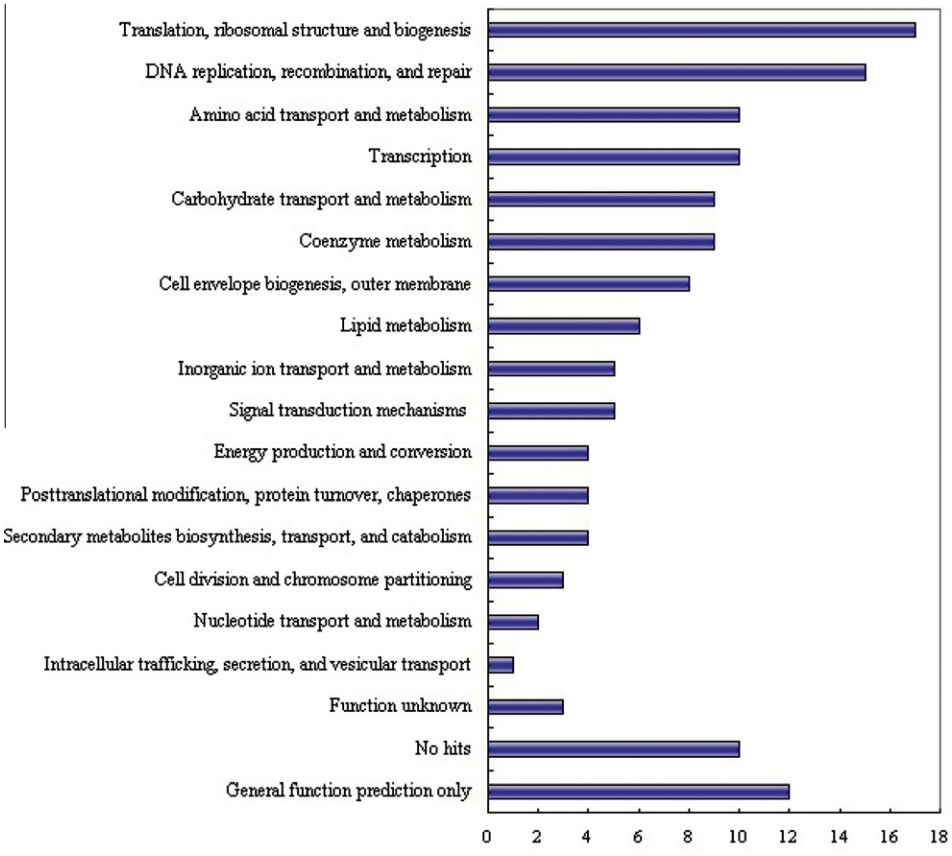


Fig. 5. Functional classification of positive selected genes in marine cyanobacteria.

functional diversification in substrate specificities (Ivarsson et al., 2003).

Several genes involved in ion transport, including peptide/nickel transporter, iron transporter and functionally unknown transporter have also been targets of positive selection in *Prochlorococcus* and *Synechococcus*. In addition, there is another selected gene called fur (ferric-uptake regulator), which can control the intracellular iron concentration in many bacteria (Crosa, 1997)

and can also regulate a variety of iron-dependent cellular processes, such as the oxidative-stress response (Escolar et al., 1999). Iron is considered as a limiting factor of primary productivity in open oceans (Behrenfeld and Kolber, 1999) and its vertical concentration distribution shows that it increases with depth. The adaptive evolution of ferric-uptake regulator and iron complex transporter in *Prochlorococcus* and *Synechococcus* waters may be an adaptation to iron-depletion environment.

4. Conclusion

Although the two groups of marine cyanobacteria *Prochlorococcus* and *Synechococcus* have shown to present a close phylogenetic relationship based on the analysis of their 16S RNA, they are different in many ways, such as the photosynthetic apparatus, genome size and the ability to grow in oligotrophic waters (Zhaxybayeva et al., 2009; Zwirglmaier et al., 2008). In this study, for the first time, we have employed comparative genomics approaches to extensively investigate synonymous codon usage bias and evolutionary rates of these two groups of marine cyanobacteria. As a result, in the view of codon usage, we found that *Prochlorococcus* and *Synechococcus* are highly divergent not only in codon usage patterns but also in the driving forces behind the diversification. It is revealed that in *Prochlorococcus*, mutation and genome compositional constraints are the main forces contributing to codon usage bias, whereas in *Synechococcus*, translational selection. In addition, nucleotide substitution rate analysis indicates that they are not evolving at a constant rate after the divergence and that the average d_N/d_S values of core genes in *Synechococcus* are significantly higher than those in *Prochlorococcus*. In conclusion, our evolutionary genomic analysis provides the first insight into codon usage, evolutionary genetic mechanisms and environmental adaptation of *Synechococcus* and *Prochlorococcus* after divergence.

Acknowledgments

This work is supported by the Science and Technology Foundation of Zhejiang Province, China (2009C33040); the National Natural Science Foundation of Zhejiang Province, China (Z307471) and Wenzhou Science and Technology Development Program, China (Y20080105 and S20090002).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.09.013.

Reference

Behrenfeld, M.J., Kolber, Z.S., 1999. Widespread iron limitation of phytoplankton in the south pacific ocean. Science (New York, N.Y) 283, 840–843.
Brevet, A., Chen, J., Commans, S., Lazennec, C., Blanquet, S., Plateau, P., 2003. Anticodon recognition in evolution: switching tRNA specificity of an aminoacyl-tRNA synthetase by site-directed peptide transplantation. The Journal of Biological Chemistry 278, 30927–30935.
Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., Welschmeyer, N.A., 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature 334, 340–343.
Crosa, J.H., 1997. Signal transduction and transcriptional and posttranscriptional control of iron-regulated genes in bacteria. Microbiology and Molecular Biology Reviews 61, 319–336.
Cusack, S., 1997. Aminoacyl-tRNA synthetases. Current Opinion in Structural Biology 7, 881–889.
dos Reis, M., Wernisch, L., Savva, R., 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. Nucleic Acids Research 31, 6976–6985.
Dufresne, A., Garczarek, L., Partensky, F., 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biology 6, R14.
Escolar, L., Perez-Martin, J., de Lorenzo, V., 1999. Opening the iron box: transcriptional metalloregulation by the Fur protein. Journal of Bacteriology 181, 6223–6229.
Feng, L., Yuan, J., Toogood, H., Tumbula-Hansen, D., Soll, D., 2005. Aspartyl-tRNA synthetase requires a conserved proline in the anticodon-binding loop for tRNA(Asn) recognition in vivo. The Journal of Biological Chemistry 280, 20638–20641.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52, 696–704.
Huang, Y., Eugene, V.K., David, J.L., Teresa, M.P., 2009. Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage. Nucleic Acids Research 37, 6799–6810.
Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science (New York, NY) 294, 2310–2314.
Ingvarsson, P.K., 2008. Molecular evolution of synonymous codon usage in Populus. BMC Evolutionary Biology 8, 307.
Ivarsson, Y., Mackey, A.J., Edalat, M., Pearson, W.R., Mannervik, B., 2003. Identification of residues in glutathione transferase capable of driving functional diversification in evolution. A novel approach to protein redesign. The Journal of Biological Chemistry 278, 8733–8738.
Johnson, P.W., Sieburth, J.M., 1979. Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototrophic biomass. Limnology and Oceanography 24, 928–935.
Kawabe, A., Miyashita, N.T., 2003. Patterns of codon usage bias in three dicot and four monocot plant species. Genes & Genetic Systems 78, 343–352.
Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G.M., Richardson, P., Chisholm, S.W., 2007. Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. PLoS Genetics 3, e231.
Li, L., Stoeckert Jr., C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13, 2178–2189.
Limor-Waisberg, K., Carmi, A., Scherz, A., Pilpel, Y., Furman, I., 2011. Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies. Nucleic Acids Research 39, 6016–6028.
Mes, T.H., Doeleman, M., Lodders, N., Nubel, U., Stal, L.J., 2006. Selection on protein-coding genes of natural cyanobacterial populations. Environmental Microbiology 8, 1534–1543.
Moore, L.R., Rocap, G., Chisholm, S.W., 1998. Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. Nature 393, 464–467.
Morton, B.R., 1997. Rate of synonymous substitution do not indicate selective constraints on the codon usage of the plant psb A gene. Molecular Biology Evolution 14, 412–419.
Partensky, F., Hess, W.R., Vaulot, D., 1999. Prochlorococcus, a marine photosynthetic prokaryote of global significance. Microbiology and Molecular Biology Review 63, 106–127.
Posada, D., 2009. Selection of models of DNA evolution with jModelTest. Methods of Molecular Biology 537, 93–112.
Pouwels, P.H., Leunissen, J.A., 1994. Divergence in codon usage of Lactobacillus species. Nucleic Acids Research 22, 929–936.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genetics 16, 276–277.
Rocap, G., Distel, D.L., Waterbury, J.B., Chisholm, S.W., 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. Applied Environmental Microbiology 68, 1180–1191.
Sau, K., Deb, A., 2009. Temperature influences synonymous codon and amino acid usage biases in the phages infecting extremely thermophilic prokaryotes. In Silico Biology 9, 1–9.
Sharp, P.M., Li, W.H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Research 14, 7737–7749.
Sharp, P.M., Li, W.H., 1987. The codon adaptation index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research 15, 1281–1295.
Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. Proceedings of the National Academy of Sciences of the United States of America 85, 2653–2657.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology Evolutionary 24, 1596–1599.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.
Urbach, E., Scanlan, D.J., Distel, D.L., Waterbury, J.B., Chisholm, S.W., 1998. Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of Prochlorococcus and Synechococcus (Cyanobacteria). Journal of Molecular Evolution 46, 188–201.
Vicario, S., Moriyama, E.N., Powell, J.R., 2007. Codon usage in twelve species of Drosophila. BMC Evolutionary Biology 7, 226.
Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87, 23–29.
Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13, 555–556.
Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular Biology and Evolution 19, 908–917.

Zhao, F., Qin, S., 2006. Evolutionary analysis of phycobiliproteins: implications for their structural and functional relationships. Journal of Molecular Evolution 63, 330–340.

Zhaxybayeva, O., Doolittle, W.F., Papke, R.T., Gogarten, J.P., 2009. Intertwined evolutionary histories of marine *Synechococcus* and *Prochlorococcus marinus*. Genome Biology Evolutionary 1, 325–339.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., Not, F., Massana, R., Ulloa, O., Scanlan, D.J., 2008. Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. Environmental Microbiology 10, 147–161.