Codon usage analysis of photolyase encoding genes of cyanobacteria inhabiting diverse habitats

Rajneesh¹ · Jainendra Pathak¹ · Vinod K. Kannaujiya¹ · Shailendra P. Singh¹ · Rajeshwar P. Sinha¹

Received: 6 February 2017 / Accepted: 31 May 2017 / Published online: 29 June 2017
© Springer-Verlag GmbH Germany 2017

Abstract Nucleotide and amino acid compositions were studied to determine the genomic and structural relationship of photolyase gene in freshwater, marine and hot spring cyanobacteria. Among three habitats, photolyase encoding genes from hot spring cyanobacteria were found to have highest GC content. The genomic GC content was found to influence the codon usage and amino acid variability in photolyases. The third position of codon was found to have more effect on amino acid variability in photolyases than the first and second positions of codon. The variation of amino acids Ala, Asp, Glu, Gly, His, Leu, Pro, Gln, Arg and Val in photolyases of three different habitats was found to be controlled by first position of codon (G1C1). However, second position (G2C2) of codon regulates variation of Ala, Cys, Gly, Pro, Arg, Ser, Thr and Tyr contents in photolyases. Third position (G3C3) of codon controls incorporation of amino acids such as Ala, Phe, Gly, Leu, Gln, Pro, Arg, Ser, Thr and Tyr in photolyases from three habitats. Photolyase encoding genes of hot spring cyanobacteria have 85% codons with G or C at third position, whereas marine and freshwater cyanobacteria showed 82 and 60% codons, respectively, with G or C at third position. Principal component analysis (PCA) showed that GC content has a profound effect in separating the genes along the first major axis according to their RSCU (relative synonymous codon usage) values, and neutrality analysis indicated that mutational pressure has

resulted in codon bias in photolyase genes of cyanobacteria.

Keywords Codon usage bias · Cyanobacteria · Mutational pressure · Photolyase · Photoreactivation

Abbreviations

RSCU Relative synonymous codon usage
PCA Principal component analysis
CAI Codon adaptation index
Nc Effective numbers of codon

Introduction

There are a number of cyanobacterial strains whose genome has been fully sequenced. This availability of cyanobacterial genomic information has given a way to study the variability in nucleotides and corresponding amino acid codons of genes. It is well known that individual genes and the entire genome of an organism can vary significantly in nucleotide composition (Bernardi and Bernardi 1986; Muto and Osawa 1987). Genome of some organisms is disproportionately rich in guanine and cytosine (GC), while others have adenine and thymine (AT) rich genome. The variation in nucleotide composition is mostly found in the synonymous codon positions of genes and because of these variations, DNA content may have little effect on the amino acid content of the encoded proteins (Loomis and Smith 1990; Lockhart et al. 1992).

Generally, overall GC content ranges from 25 to 75% of the genome, however, GC content of genes may vary from 7 to 95% (Lobry and Sueoka 2002). In prokaryotic system, it

✉ Rajeshwar P. Sinha
r.p.sinha@gmx.net; rpsinhabhu@gmail.com

¹ Laboratory of Photobiology and Molecular Microbiology, Centre of Advanced Study in Botany, Institute of Science, Banaras Hindu University, Varanasi 221005, India

has been found that amino acid composition has a high correlation with the G + C composition (Sueoka 1961). In bacterial system, G + C variability results in variation in amino acid composition through codon redundancy, which is highest on third nucleotide of codon in contrast to the first and the second nucleotide position (D'Onofrio et al. 1991; Wada 1992; de Miranda et al. 2000; Singer and Hickey 2000; Knight et al. 2001; Harrison and Charlesworth 2011). Singer and Hickey (2000) has demonstrated correlation between the nucleotide composition and amino acid composition in 21 completely sequenced archaeal and eubacterial genomes.

It was hypothesized that GC content has an effect on codon usage and that the correlation between GC content and amino acid or codon usage is modulated by both mutation and selection pressure (Knight et al. 2001). The variation of G3C3 (content of G + C at third codon positions) along the coding sequences was initially determined by sliding window analyses of a group of phage and bacterial genes (Wada and Suyama 1985). Bharanidharan et al. (2004) noticed that proteins encoded by G + C rich genomes contain a greater proportion of the GARP amino acids and codons used are G + C rich. Similarly, the G + C poor genomes codes mostly FYMINK group of amino acids and possess A + T rich codons. There are plenty of reports available about the genome variability and its effect on codon usage and subsequent amino acid composition of protein in bacteria as well as mammals, however, similar studies are severely lacking for cyanobacteria (Kannaujiya et al. 2014). Cyanobacteria show variation in G + C constituents similar to other bacterial systems. The genetic codes are degenerative and except Met and Trp, all other amino acids have more than one codon. Synonymous codons are often used with different frequencies both within and among genomes (Grantham et al. 1980; Lloyd and Sharp 1992). Several studies have reported that codon usage in various organisms is influenced by factors such as compositional constraints and translational selection (Sharp and Li 1986a, b; Chen et al. 2013). Previous studies suggest that the nucleotide composition of the genes have an effect on protein evolution (Singer and Hickey 2000). Studies of synonymous codon usage are much needed for improving our knowledge of the mechanism(s) underlying behind synonymous codon usage bias (Powell and Moriyama 1997). Furthermore, information obtained from such studies can be useful in selection of an appropriate host for heterogenous protein expression study (Powell and Moriyama 1997; Zheng et al. 2007), in designing degenerate primers (Zhou et al. 2005), in prediction of genes from nucleotide sequences (Salamov and Solovyev 2000) and functional classification of proteins (Lin et al. 2002). The information obtained from codon usage study could provide inputs on level of genes expression, evolution of the genes and genomes and could be also utilized to enhance the

immunogenicity of vaccines (Crameri et al. 1996). Evolutionary studies conducted by Hooper and Berg (2000) confirms that the pattern of codon usage varies within a single gene, between the genes, as well as in genomes, which collectively suggests that there is a proper balance between biases generated by mutation, random genetic drift and natural selection (Lobo et al. 2006; Qi et al. 2015). Cyanobacteria are Gram-negative photoautotrophic organisms having wide range of proteins involved in repair mechanism(s) of damages caused by various environmental factors, including ultraviolet radiation (UVR; 280–400 nm) (Sinha et al. 1995; Sinha and Häder 2002; Richa et al. 2011).

UVR has the ability to induce inflammation, oxidative stress, free radical production, sunburn, DNA damage and skin cancer (Afaq et al. 2005; Bickers and Athar 2006; Halliday and Lyons 2008; Timares et al. 2008). DNA of all living organism residing on Earth is susceptible to UV-B (280–315 nm) due to its inductive effect on formation of pyrimidine dimers in genome. UV-induced DNA damage results in cis-syn cyclobutane pyrimidine dimers (CPDs), pyrimidine pyrimidone (6-4) photoproducts, and dewar isomers (Batista et al. 2009; Rastogi et al. 2010). Photoreactivation is one of the unique mechanisms found in cyanobacteria which reverse UV-induced photo-adducts into their normal state in presence of blue light. The process of photoreactivation is carried out by family of DNA repair enzyme called photolyase (Sinha and Häder 2002). Three types of photolyases have been identified in various organisms: CPD (cyclobutane pyrimidine dimers) photolyase repair cyclobutane pyrimidine dimers, (6-4) photolyases repair (6-4) pyrimidine pyrimidone, and cryptochrome-DASHs exhibit a variety of physiological functions including single-strand DNA photolyase activity (Selby and Sancar 2006; Essen and Klar 2006). Thus, photolyases are very critical enzymes for the survival of photoautotrophic organisms, including cyanobacteria.

In this study, we conducted codon usage bias analysis of photolyase encoding genes from cyanobacteria inhabiting different ecological habitats, and further elucidated if mutation or selection pressure has predominant effect on shaping the codon usage pattern in photolyase encoding genes. We also explored the photolyases of cyanobacteria by studying the amino acid variation within and among different species from three different habitats.

Materials and methods

Data collection and analysis

The nucleotide and amino acid sequence of photolyase from different cyanobacterial species were obtained from NCBI database. For this study, cyanobacterial species from three

different habitats, i.e., hot springs, fresh water and marine water, were selected. The information about locus region, accession number, total nucleotide and protein length has been given in Table 1. The percentage composition of amino acids was calculated by protein specific tools (http://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/). The percentage G + C composition was obtained using bioinformatics tools (http://www.genomicsplace.com/cgi-bin/gc_calculator.pl). CAIcal server (<http://genomes.urv.es/CAIcal/>) was used to calculate the uses of codon, relative synonymous codon usage (RSCU), codon adaptation index (CAI), effective numbers of codon (N_c) and G + C percentage.

Sequence analysis

The graph of the photolyase encoding gene was plotted on an average basis separately to determine the amino acid variability using Sigma Plot 11.0. The cause of respective amino acid variability was found out from nucleotide variability by checking the total G + C and A + T composition. The variability in G + C bias was determined from all synonymous codon substitutions as G1 + C1, G2 + C2 and G3 + C3 through correlation and regression analysis. Codon variability in G1 + C1, G2 + C2 and G3 + C3 bias was identified by online tools mentioned above. Termination codons were excluded for all analysis.

Principal component analysis (PCA)

Principal component analysis (PCA) was performed using the software XLSTAT. PCA is a multivariate statistical analysis for simplifying the multidimensional information of the data matrix into a two-dimensional map (Morrison 1990).

Statistical analysis

The variations in G + C content and amino acid composition of the photolyase encoding genes were tested statistically through correlation and regression analysis. This analytical test was done for each position of codons from G1 + C1, G2 + C2 and G3 + C3. A one-way analysis of variance was applied to confirm the significance of data according to Duncan's multiple comparison tests at $p < 0.05$. Microsoft excel 2007 and Sigma Plot 11.0 version were used for all statistical analysis.

Results

Amino acid variability in photolyase

The average composition of amino acids in photolyases from studied organisms has been shown in Fig. 1. Cyanobacterial

Table 1 The cyanobacterial species with their Locus/accession number and gene/protein length that were used in this study

Sp. ID	Primary locus (accession no.) gene/protein length
Fresh water cyanobacteria	
Sp.1	SYN7509_RS0210765/NZ_ALVU02000001.1 1430/476
Sp.2	SGL_RS08150/NC_000911.1 1439/479
Sp.3	SYN6312_RS01110/NC_019680.1 1442/480
Sp.4	M744_RS02410/NZ_CP006471.1 1452/484
Sp.5	NOS7107_RS01815/NC_019676.1 1437/479
Sp.6	ANA_RS21775/NC_019439.1 1434/478
Marine cyanobacteria	
Sp.7	Syncc8109_RS00925/NZ_CP006882.1 1431/477
Sp.8	TERY_RS18880/NC_008312.1 1428/476
Sp.9	KR49_RS11700/NZ_CP006270.1 1428/496
Sp.10	KR100_RS08130/NZ_CP006269.1 1485/495
Sp.11	SYNRCC307_RS05695/NC_009482.1 1524/508
Sp.12	WH8103_RS01135/NZ_LN847356.1 1431/477
Hot spring cyanobacteria	
Sp.13	II0425/NC_004113.1 1440/480
Sp.14	NK55_RS07070/NC_023033.1 1440/480
Sp.15	WP_011429671.1/NC_007775.1 1437/479
Sp.16	CYB_RS09280/NC_007776.1 1575/525
Sp.17	PLE7327_RS05170/NC_019689.1 1431/477
Sp.18	OSCCY1_RS01820/NZ_KL662191.1 1488/496

Sp1 Synechocystis sp. PCC 7509; *Sp2 Synechocystis* sp. PCC 6803; *Sp3 Synechococcus* sp. PCC 6312; *Sp4 Synechococcus* sp. UTEX 2973; *Sp5 Nostoc* sp. PCC 7107; *Sp6 Anabaena* sp. 90; *Sp7 Thermosynechococcus elongatus* BP1; *Sp8 Thermosynechococcus* sp. NK55; *Sp9 Synechococcus* sp. JA-3-3Ab; *Sp10 Synechococcus* sp. JA-2-3B⁺a(2-13); *Sp11 Pleurocapsa* sp. PCC 7327; *Sp12 Leptolyngbya* sp. JSC1; *Sp13 Synechococcus* sp. WH 8109; *Sp14 Trichodesmium erythraeum* IMS101; *Sp15 Synechococcus* sp. KORDI-49; *Sp16 Synechococcus* sp. KORDI-100; *Sp17 Synechococcus* sp. RCC307; *Sp18 Synechococcus* sp. WH 8103

Fig. 1 Average percentage composition of each amino acid in photolyase of cyanobacteria from diverse habitats. **a** Fresh water, **b** marine and **c** hot spring. Error bars represent standard deviation of the mean (mean $\pm$ SD). Significance of data was evaluated by Duncan's multiple range test in the range of $p < 0.05$

photolyase enzyme from three different habitats showed lower occurrence of Cys, His, Met, Lys, Asn and Tyr. Moreover, the low percentage is constant in all photolyases

as indicated by the low value of the standard deviation. Clearly, the scarcity of formation of disulphide bridges by Cys residue is associated with its low occurrence (Bharanidharan et al. 2004). Met is principally an initiation amino acid, which justifies its low percentage (Kreil and Ouzounis 2001). However, abundance of non-polar amino acids such as Ala, Leu, Gly, Pro and polar amino acids such as Arg, Gln and Glu was found in all studied cyanobacteria. Ala and Leu were found to be the most abundant amino acids in photolyases of cyanobacteria.

Nucleotide content analysis of photolyase encoding genes

To find out possible means of variability in amino acid constituents, differences in G + C percentage were analyzed in the genome of all studied cyanobacteria. The genomic GC content of fresh water cyanobacteria varies from 38 to 55.5%. In hot spring species, genomic GC content was found to be 45–60.20%. However, GC content varies from 34 to 61.4% in genome of marine cyanobacteria. The GC content of photolyase encoding genes in fresh water cyanobacteria varies from 36 to 58.60% (Fig. 2a) and in hot spring cyanobacteria GC content was 41–65% (Fig. 2b), while in marine cyanobacteria GC content was 30–64% (Fig. 2c). Overall, the significant variability in GC constituent was found in all photolyase encoding genes of studied cyanobacteria but maximum variation was found in marine cyanobacteria followed by hot spring and fresh water cyanobacteria.

Correlation and regression analysis of GpC codon level in photolyase encoding genes

All codon positions with G + C variability such as G1 + C1, G2 + C2 and G3 + C3 are shown in Fig. 3, and change in slopes and correlation coefficients are shown in Table 2. In photolyase encoding genes of fresh water cyanobacteria, the third nucleotide position bias (G3 + C3) was found to be higher (slope = 2.303) than the first (G1 + C1, slope = 0.935) and second position (G2 + C2, slope = 2.020). In hot spring cyanobacteria, the nucleotide bias at G3 + C3 (slope = 1.559) was also found to be higher than G2 + C2 position (slope = 0.438) but lower than G1 + C1 position (slope = 2.229). In marine cyanobacteria the nucleotide bias at G3 + C3 (slope = 1.395) was also found to be higher than G2 + C2 position (slope = 0.716) but lower than G1 + C1 position (slope = 1.560). Thus, collectively these results suggest that nucleotide bias is higher at G3 + C3 position of codons for cyanobacterial photolyases, irrespective of the ecological niches.

Fig. 2 Variation in average percentage of GC composition of photolyases in cyanobacteria from diverse habitats. **a** Fresh water, **b** marine and **c** hot spring

Neutrality plot analysis

Neutrality plot analysis of the GC12 and G3C3 was conducted to quantify the extent of natural selection and

Fig. 3 Correlation and regression analysis of scatter plot of the Gi + Ci levels ('i' denotes first, second and third position of the codon) between total genomic GC constituents and the relative variability at the first, second and third nucleotide positions of photolyase encoding genes. **a** Fresh water, **b** marine and **c** hot spring

mutation pressure in the codon usage pattern of photolyase encoding genes (Fig. 4). In neutrality plots, when a notable correlation between GC12 and G3C3 exists and the

Table 2 Slopes, sum of the square of the correlation coefficient and standard deviation of all three nucleotides position of G + C bias in photolyases encoding gene of all studied cyanobacteria

	Slope	R^2	p value	SD
GC Position (A)				
G1C1	0.935 ± 0.2289	0.847	0.009	6.3321
G2C2	2.020 ± 0.2064	0.654	0.051	2.5740
G3C3	2.303 ± 0.2106	0.881	<0.005	15.9685
GC Position (B)				
G1C1	1.560 ± 0.1415	0.916	0.003	17.2831
G2C2	0.716 ± 0.1186	0.902	0.004	7.9949
G3C3	1.395 ± 0.1511	0.948	<0.001	15.1901
GC Position (C)				
G1C1	2.229 ± 0.2064	0.869	0.007	12.9515
G2C2	0.438 ± 0.0892	0.367	0.203	3.9170
G3C3	1.559 ± 0.1208	0.928	<0.001	8.7646

A fresh water, B marine and C hot spring

slope of the regression line is near to 1, mutation bias is supposed to be the major force in shaping the codon usage. Conversely, if there is no correlation between GC12 and G3C3 indicates natural selection which results in a narrow distribution of GC content (Sueoka 1988). In our study, a significant positive correlation was found between GC12 and G3C3 ($r = 0.825$, $p < 0.05$) (Fig. 4a) in fresh water cyanobacteria. A strong correlation between GC12 and G3C3 was observed in marine cyanobacteria ($r = 0.959$, $p < 0.01$) and hot spring cyanobacteria ($r = 0.925$, $p < 0.01$) (Fig. 4b, c). This strong correlation suggests that mutational pressure has acted on codon usage bias in photolyase encoding genes of studied cyanobacteria. The G3C3 value of cyanobacteria inhabiting fresh water, marine and hot spring habitats was found to be 48–67, 36.69–76.99 and 52.72–79.12%, respectively, which further suggest that mutational pressure dominates over natural selection in codon usage pattern of cyanobacterial photolyases (Wei et al. 2014).

Gene expression level of photolyase encoding genes

Previous findings showed significant correlation between CAI and N_c in *Taenia saginata* (Yang et al. 2014), *Haemophilus influenzae* (Supriyo et al. 2014), Herbaceous Peony (Wu et al. 2015) and transcription factor gene GATA2 from mammals (Mazumder et al. 2016). These imply gene expression level plays an important role in shaping the codon usage bias or vice versa.

The expression level of photolyase encoding genes assessed by CAI. For fresh water cyanobacteria it ranges from 0.53 to 0.59 with mean value 0.55 and standard deviation of 0.021. In marine habitat cyanobacteria value ranges from 0.55 to 0.73 with mean 0.61 and standard deviation of

Fig. 4 Neutrality plot of individual photolyase encoding genes plotted as the average of GC content in the first and second codon position versus the GC content of the third codon position (G3C3). **a** Fresh water, **b** marine and **c** hot spring

0.07, while in hot spring it ranges from 0.49 to 0.71 with mean 0.62 and standard deviation of 0.07. Further, we conducted correlation analysis between N_c and CAI (Table 3). No correlation was found between two factors in fresh water and marine cyanobacteria, however, strong negative

Table 3 Effective number of codon (Nc) and Codon adaptation index (CAI) of cyanobacterial species

Species	CAI	Nc
Fresh water		
Sp.1	0.531	49.4
Sp.2	0.535	54.8
Sp.3	0.568	48.8
Sp.4	0.591	44.5
Sp.5	0.546	41.9
Sp.6	0.537	43
Marine		
Sp.7	0.551	46.7
Sp.8	0.551	49.4
Sp.9	0.732	38.8
Sp.10	0.674	43
Sp.11	0.543	46
Sp.12	0.602	53.1
Hot spring		
Sp.13	0.707	39.4
Sp.14	0.495	50.2
Sp.15	0.659	45
Sp.16	0.575	47.8
Sp.17	0.650	42.5
Sp.18	0.677	41.9

correlation was observed in hot spring cyanobacteria ($r = -0.952$, $p < 0.01$). These findings suggest that the expression of photolyase encoding genes in cyanobacteria inhabiting fresh water and marine habitats is not affected by the codon bias pattern. Similar findings have been reported for enterovirus (Ma et al. 2014) and for cytochrome P450 genes which are associated with coronary artery disease in human (Malakar et al. 2016). However, the expressions of photolyase encoding genes in hot spring cyanobacteria were found to be affected by pattern of codon bias.

Correlation between G1 + C1 codon and their respective amino acids

The guanine and cytosine at the first position (G1 + C1) encode 10 amino acids using thirty-two associated codons. Ala is encoded by all four codons GCT, GCC, GCA and GCG. Codons GCT, GCC and GCG showed high codon usage in photolyase encoding genes of fresh water cyanobacteria while GCA was least used codon (Fig. 5). However, in hot spring cyanobacteria, GCC and GCG codons showed higher usage in comparison to GCT and GCA codons. In marine habitat except GCA, codons GCT, GCC, and GCG were found to be highly utilized. In case of Leu, the CTG codon is mainly used, while CTT and CTC codons showed equal usage contrary to CTA, which is least used codon in photolyase encoding genes

Fig. 5 G1 + C1 associated codon variation with respect of their frequency per unit for 10 amino acids (abbreviated as A, D, E, G, H, L, P, Q, R and V) in photolyase of cyanobacteria

of all cyanobacteria. CTG and CTC were found to be most utilized codons, whereas CTT is equally utilized and CTA is least utilized codon in members of marine and hot spring cyanobacteria. For Gln, codon CAG was more preferred than CAA in hot spring and marine cyanobacteria, whereas CAA was more preferred codon in fresh water cyanobacteria. In case of Arg, codons CGC and CGG showed higher usage than CGT and CGA in all studied cyanobacteria. GGT, GGC and GGG codons for Gly showed higher utilization in comparison to GGA in fresh water habitat. However, higher usage of GGT and GGC codons, and GGC and GGG codons for Gly were found to be in marine and hot spring cyanobacteria, respectively. In fresh water cyanobacteria, all codons for Val are utilized equally while in hot spring and marine habitats codon usage of GTC and GTG codons is higher than GTT and GTA. In fresh water cyanobacteria, codon CCA is more frequently utilized than other codons for Pro, while in marine habitat codons CCG and CCC are frequently utilized. In hot spring cyanobacteria CCC and CCA codons are highly utilized for Pro amino acid in photolyase. The codon GAC for Asp is predominantly utilized by fresh water cyanobacteria while codons GAT and GAC are equally utilized in marine and hot spring cyanobacteria. In marine and hot spring cyanobacteria, codons CAT and CAC are equally utilized while in fresh water cyanobacteria codon CAT is preferentially utilized over CAC. The codon usage for Glu was more or less similar in three habitats. Thus, a higher degree of variation in codon usage was observed in photolyases of cyanobacteria from different habitats, and G1 + C1 provided the first stage of codon variability in terms of their amino acid composition.

Correlation between G2 + C2 codon and respective amino acids

The second position of guanine and cytosine (G2 + C2) associated with 8 amino acids is translated from 31 codons

Fig. 6 G2 + C2 associated codon variation with respect to their frequency per unit for 8 amino acids (abbreviated as A, C, G, P, R, S, T, W) in photolyase of cyanobacteria

(Fig. 6). Contrary to G1 + C1, the G2 + C2 codons were found to be slightly rich in cytosine than guanine residue in all studied cyanobacteria. For Ser, codon TCA was least utilized, however, codons TCT, TCG and TCC were equally utilized in three habitats. For Gly, all codons (GGT, GGG and GGC) except GGA showed higher usage in fresh water cyanobacteria. However, in hot spring cyanobacteria, codons GGC and GGG were highly utilized and codons GGC and GGT were frequently utilized in marine cyanobacteria. The usage of codons CGC and CGG for Arg was high in fresh water cyanobacteria while other codons were equally utilized. The usage of CGC and CGG was higher in hot springs and marine cyanobacteria than fresh water cyanobacteria. In Ala, all codons except GCA were equally utilized in all habitats. ACT and ACC show higher usage in the fresh water cyanobacteria for amino acid Thr, while ACC is frequently used in marine and hot spring cyanobacteria. Fresh water and hot spring cyanobacteria frequently utilize CCC and CCA codons for Pro, whereas marine cyanobacteria showed higher usage of CCA and CCG codons for the same amino acid. Codons TGT and TGC showed equal usage for Cys in fresh water cyanobacteria, while in marine and hot spring cyanobacteria TGC is frequently utilized.

Correlation between G3 + C3 codon and respective amino acids

The third nucleotide position of codon was found to play a crucial role in the variation of total amino acids with specific codons. G3 + C3 position is associated with all 20 amino acid translated from 31 codons (Fig. 7). The codons TTC (Phe), ATC (Ile), AGC (Ser), GCC (Ala), GGC (Gly), TGC (Cys) and TTG (Leu) showed high usage, while codons TCC and TCG (Ser), CAC (His), GTG (Val), GAC

Fig. 7 G3 + C3 associated codon variation with respect to their frequency per unit for 20 amino acids (abbreviated as A, C, D, E, F, G, H, I, L, K, M, N, P, Q, R, S, V, T, Y and W) in photolyase of cyanobacteria

(Asp), AAG (Lys), ACG (Thr) and AGG (Arg) were least utilized. However, CTC (Lys), GTC (Val), GCG (Ala), AAC (Asn), and CAG (Gln) were moderately utilized codons in photolyase encoding gene of fresh water cyanobacteria. In marine habitat, codons CAG (Gln), CTG (Leu), GAG (Glu), GCC (Ala), GGC (Gly), CGC (Arg) and CCG (Pro) have higher usage, while CCC (Pro), CGC (Ala), GGG (Gly), TTG (Leu) and AAG (Lys) were moderately utilized codons. AGG (Lys), TCG (Ser), TAC (Tyr) are least utilized codons in marine habitat. In hot spring cyanobacteria, CTG (Lys), GGC (Gly), CAG (Gln), GCC (Ala) and GAG (Glu) codons showed higher usage, while ATG (Met), TAC (Tyr) and AAC (Asn) were moderately utilized codons. AGG (Arg), ACG (Thr), TCG (Ser) and AAG (Lys) codons showed least usage in hot spring cyanobacteria.

Preferential codon usage

The RSCU values of all 59 sense codons in cyanobacteria from different habitats are shown in Tables 4, 5 and 6. Twenty-eight codons including TTG, CCA, AGC and CGC were found to be highly utilized in photolyase encoding genes of fresh water cyanobacteria. More than half of frequently used codons (17/28) ended with G or C at their third position. Twenty-three codons, including CTG, AGC, ACC and CGC were frequently used codons in marine cyanobacteria. More than 80% (19/23) of the frequently used codons ended with G or C at their third position. In hot spring cyanobacteria, 26 codons including AGC, CGC, CTG, GGC, TTG and ACC were found to be highly utilized and more than three-fourth (22/26) of frequently used codons ended with G or C. These results suggest that nucleotides composition of an organism could also play an important role in determining the codon usage pattern, i.e.,

Table 4 Relative synonymous codon usage (RSCU) of fresh water cyanobacteria

Codon	AA	RSCU	Codon	AA	RSCU
TTT	F	0.9402	GCC	A	1.5488
TTC		1.0598	GCA		0.2787
TTA	L	0.3057	GCG		1.2803
TTG		2.8540	TAT	Y	0.4335
CTT		0.4983	TAC		1.2332
CTC		0.6773	CAT	H	1.4185
CTA		0.1167	CAC		0.5815
CTG		1.5482	CAA	Q	1.2755
ATT	I	1.0768	CAG		0.7245
ATC		1.2115	AAT	N	1.4530
ATA		0.7117	AAC		0.5470
GTT	V	1.0822	AAA	K	1.6078
GTC		1.1438	AAG		0.3922
GTA		0.7168	GAT	D	1.6057
GTG		1.0565	GAC		0.3943
TCT	S	0.6562	GAA	E	0.8272
TCC		0.8173	GAG		1.1728
TCA		0.0977	TGT	C	0.7152
TCG		0.6280	TGC		1.2848
AGT		1.0898	CGT	R	0.7622
AGC		2.7112	CGC		1.7628
CCT	P	0.2115	CGA		0.478
CCC		0.9025	CGG		0.966
CCA		2.1568	AGA		1.0065
CCG		0.7292	AGG		1.0243
ACT	T	1.4242	GGT	G	0.8857
ACC		1.4323	GGC		1.5378
ACA		0.4442	GGA		0.2000
ACG		0.6993	GGG		1.3763
GCT	A	0.8920			

The codons displayed in bold are preferred
AA amino acid

GC rich organisms could prefer codons having G or C at third position.

Principal component analysis

To study the relative contribution of two major factors, i.e., natural selection and mutational pressure on codon usage, we performed PCA analysis taking RSCU scores to find out major trends of codon usage in photolyase encoding genes of cyanobacteria. A plot of F1 and F2 showed important features of the pattern of codon usage in photolyase encoding genes of cyanobacteria from different habitats (Fig. 8). From this analysis major trends in codon usages were detected in which axis (F1) accounted for 36.28%, whereas axis (F2) accounted

Table 5 Relative synonymous codon usage (RSCU) of marine cyanobacteria

Codon	AA	RSCU	Codon	AA	RSCU
TTT	F	0.6607	GCC	A	1.8237
TTC		1.3393	GCA		0.5060
TTA	L	0.1027	GCG		0.8465
TTG		1.0598	TAT	Y	0.7817
CTT		0.6658	TAC		0.8850
CTC		1.5378	CAT	H	1.1128
CTA		0.0947	CAC		0.8872
CTG		2.5388	CAA	Q	0.4715
ATT	I	0.8455	CAG		1.5285
ATC		1.8852	AAT	N	0.6247
ATA		0.2693	AAC		1.3753
GTT	V	0.7057	AAA	K	0.8702
GTC		0.8945	AAG		1.1298
GTA		0.4072	GAT	D	1.2098
GTG		1.9930	GAC		0.7902
TCT	S	0.6053	GAA	E	0.7603
TCC		1.0428	GAG		1.2398
TCA		0.1955	TGT	C	0.7998
TCG		0.892	TGC		1.2002
AGT		0.7068	CGT	R	0.6977
AGC		2.5577	CGC		2.4140
CCT	P	0.5328	CGA		0.4497
CCC		1.2630	CGG		1.3772
CCA		0.9648	AGA		0.6773
CCG		1.2393	AGG		0.3843
ACT	T	0.6788	GGT	G	1.0838
ACC		2.0373	GGC		1.6590
ACA		0.3117	GGA		0.4632
ACG		0.9722	GGG		0.7943
GCT	A	0.8240			

The codons displayed in bold are preferred
AA amino acid

for 27.47% of total variation in fresh water cyanobacteria (Fig. 8a). In marine cyanobacteria axis (F1) accounted for 47.45% and axis (F2) accounted for 26.15% (Fig. 8b), whereas in hot spring cyanobacteria axis (F1) accounted for 39.95% and axis (F2) accounted for 23.99% (Fig. 8c). It was clearly seen that in fresh water cyanobacteria CGC, CCC, CTG, CAA, CGG, AAG, GGA and TTT were oriented most distantly on the positive axis ranging from 0.945 to 0.832. However, CCA, GCG, AAA, TTC, CCT, GAT and TCT ranges from −0.928 to −0.800 on negative axis. Other codons such as GCA, TGC, CTG, AAT, AAC, CGT AAG and ATA were positioned far away from the origin. In marine cyanobacteria, codons AGG, ATA, CAT, AAA, CCA, ACT, TCT, GGA, GGC, GAA, GTA and GGG were most distantly

Table 6 Relative synonymous codon usage (RSCU) of hot spring cyanobacteria

Codon	AA	RSCU	Codon	AA	RSCU
TTT	F	0.9368	GCC	A	1.5475
TTC		1.0632	GCA		0.4080
TTA	L	0.1498	GCG		1.3638
TTG		1.7247	TAT	Y	0.9985
CTT		0.5073	TAC		1.0015
CTC		1.190	CAT	H	0.9553
CTA		0.222	CAC		1.0447
CTG		2.2065	CAA	Q	0.7092
ATT	I	0.9343	CAG		1.2908
ATC		1.4805	AAT	N	1.0277
ATA		0.5852	AAC		0.9723
GTT	V	0.7165	AAA	K	1.3608
GTC		1.3718	AAG		0.6392
GTA		0.54	GAT	D	1.0393
GTG		1.3717	GAC		0.9607
TCT	S	0.6788	GAA	E	0.8552
TCC		1.1127	GAG		1.1448
TCA		0.1595	TGT	C	0.6065
TCG		0.8982	TGC		1.3935
AGT		0.979	CGT	R	0.5678
AGC		2.3538	CGC		2.2160
CCT	P	0.3642	CGA		0.5120
CCC		1.4873	CGG		1.3005
CCA		1.3177	AGA		0.4103
CCG		0.831	AGG		0.9938
ACT	T	0.8188	GGT	G	0.5638
ACC		2.0377	GGC		1.8668
ACA		0.4493	GGA		0.4838
ACG		0.6940	GGG		1.0857
GCT	A	0.6808			

The codons displayed in bold are preferred

AA amino acid

placed on the positive axis ranging from 0.995 to 0.808, whereas CAC, AAG, GTG, CGC, GAG, CCC, ATC, CTG and CCG were placed most distantly on the negative axis ranging from -0.978 to -0.783 . Codons such as CCG, AGC, GTT, TTG and GCC were placed far away from the origin. In hot spring cyanobacteria, codons CGT, ACA, ATA, AAT, CCA, AAA, TAT and GAC were placed most distantly (0.945–0.831) on positive axis. On negative axis, AGG, CTA, AAG, TAC, CGC, CGG, GAT, and GCC (0.960 to -0.822) were distantly placed while codons such as CTG, ATC, GTC, GTG, TCA, CCC, ACT, ACC and GGA were also placed distantly from origin. These findings suggest that genomic GC content has a profound effect in separating the genes along the first main axis according to their RSCU values.

Discussion

The relationship between variability of amino acid composition and nucleotide bias is a new area of research in cyanobacterial genomics. In this study, we analyzed first, second and third position of codon, and differences in codon usage for amino acids in photolyase encoding genes of cyanobacteria inhabiting three different habitats. There was abundance of codons enriched with G or C at third position which showed higher usage in comparison to codons with G or C at first and second position. Codon usage analysis confirmed that codons with G or C at third position are preferred in photolyase encoding genes of cyanobacteria. To explore the level of variation in GC content and amino acids along with codon usage pattern, the codons of all 20 standard amino acids were determined. Results showed that G3 + C3 codes 31 codons and 20 amino acids while G2 + C2 and G1 + C1 were found to code for 10 and 8 amino acids, respectively. The third position of a codon is considered as the most likely position to reflect the genome base composition which is different in different organisms (Chen et al. 2013). Bharanidharan et al. (2004) observed that amino acid composition is decided by the nucleotide frequency to a great extent; however, the amino acid composition of the proteome is not decided entirely by nucleotide frequency due to the effect of selection pressure. Therefore, changes in compositions may be one facet of the wider degeneracy seen in biological systems (Edelman and Gally 2001).

Previous studies performed on large number of divergent species showed that there are several factors influencing the patterns of synonymous codon usage bias in various organisms (Ingvarsson 2008; Pouwels and Leunissen 1994; Sharp and Li 1986a, b). The balance of genomic compositional mutation, genetic drift and natural selection are the major contributors for variation in codon usage that affect gene translation. However, analysis of nine cyanobacteria showed that genetic drift is a weak evolutionary force in these organisms (Yu et al. 2012). Genome-wide directional mutation, especially change in GC content, has profound impact on codon usage in addition to the selection pressure, acting on highly expressed genes (Sueoka 1988). Codon usage analysis in *Schistosoma mansoni* suggested that codon usage bias is dependent on the base composition of gene (Ellis et al. 1995; Milhon and Tracy 1995; Musto et al. 1999). In bacteria, high levels of GC content were found to be positively correlated with amino acids (GARP), while negatively correlated with amino acids (FYMINK) observed in low GC levels (Banerjee et al. 2005). In this study, the abundance of GARP was found. Recently, Malakar et al. (2016) showed that mutational pressure and natural

Fig. 8 Principal component analysis depicting the variation among the RSCU values of codons in the photolyase gene. **a** Fresh water, **b** marine and **c** hot spring

selection could be the main factors accounting for codon usage bias in cytochrome P450 (CYP) genes. In *Actinobacteria*, mutation is a main driving force than natural selection for extreme GC content and codon bias (Lal et al. 2016). In this study, mutational pressure was found to be

the main driver of codon bias in photolyase encoding genes of cyanobacteria. However, only a few reports are available on effect of GC variation on codon usage in cyanobacteria. GC correlation and regression analysis was performed at G1 + C1, G2 + C2 and G3 + C3 levels. In photolyase

4.1.99.3) to traditional sunscreen was found to significantly reduce the UV-induced damage of human skin by significantly reducing the pyrimidine dimers (An et al. 2013; Stege et al. 2000). Thus, information obtained from this study could be further utilized for commercial/laboratory scale production of cyanobacterial photolyases for their structural studies and/or for the development of new class of sunscreens.

Acknowledgements Rajneesh and J. Pathak are thankful to Department of Biotechnology (DBT-JRF/13/AL/143/2158) and Council of Scientific and Industrial Research (09/013/0515/2013-EMR-I), New Delhi, India, respectively, for the financial support in the form of fellowships.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- Afaq F, Adhami VM, Mukhtar H (2005) Photochemoprevention of ultraviolet B signaling and photocarcinogenesis. *Mutat Res* 571:153–173
- An M, Mou S, Zhang X, Ye N, Zheng Z, Cao S, Xu D, Fan X, Wang Y, Miao J (2013) Temperature regulates fatty acid desaturases at a transcriptional level and modulates the fatty acid profile in the Antarctic microalga *Chlamydomonas* sp. ICE-L. *Bioresour Technol* 134:151–157
- Banerjee T, Gupta SK, Ghosh TC (2005) Role of mutational bias and natural selection on genome-wide nucleotide bias in prokaryotic organisms. *BioSystems* 81:11–18
- Batista LF, Kaina B, Meneghini R, Menck CF (2009) How DNA lesions are turned into powerful killing structures: insights from UV-induced apoptosis. *Mutat Res* 681:197–208
- Bernardi G, Bernardi G (1986) Compositional constraints and genome evolution. *J Mol Evol* 24:1–11
- Bharanidharan D, Bhargavi GR, Uthamallian K, Gautham N (2004) Correlations between nucleotide frequencies and amino acid composition in 115 bacterial species. *Biochem Biophys Res Commun* 315:1097–1103
- Bickers DR, Athar M (2006) Oxidative stress in the pathogenesis of skin disease. *J Invest Dermatol* 126:2565–2575
- Chen L, Liu T, Yang D, Nong X, Xie Y, Fu Y, Wu X, Huang X, Gu X, Wang S, Peng X (2013) Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochem Biophys Res Commun* 430:1344–1348
- Cramer A, Whitehorn EA, Tate E, Stemmer WP (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* 14:315–319
- D'Onofrio G, Mouchiroud D, Aissani B, Gautier C, Bernardi G (1991) Correlations between the compositional properties of human genes, codon usage, and amino acid composition of proteins. *J Mol Evol* 32:504–510
- Dass JFP, Sudandiradoss C (2012) Insight into pattern of codon biasness and nucleotide base usage in serotonin receptor gene family from different mammalian species. *Gene* 503:92–100
- de Miranda AB, Alvarez-Valin F, Jabbari K, Degraeve WM, Bernardi G (2000) Gene expression, amino acid conservation, and hydrophobicity are the main factors shaping codon preferences in *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *J Mol Evol* 50:45–55
- Duan X, Yi S, Guo X, Wang W (2015) A comprehensive analysis of codon usage patterns in blunt snout bream (*Megalobrama amblycephala*) based on RNA-Seq data. *Int J Mol Sci* 16:11996–12013
- Edelman GM, Gally J (2001) Degeneracy and complexity in biological systems. *Proc Natl Acad Sci USA* 98:13763–13768
- Ellis J, Morrison DA, Kalinna B (1995) Comparison of the patterns of codon usage and bias between *Brugia*, *Echinococcus*, *Onchocerca* and *Schistosoma* species. *Parasitol Res* 81:388–393
- Essen LO, Klar T (2006) Light-driven DNA repair by photolyases. *Cellular and Molecular Life Sciences* 63:1266–1277
- Grantham R, Gautier C, Gouy M, Mercier R, Pave A (1980) Codon catalog usage and the genome hypothesis. *Nucleic Acids Res* 8:R49–R62
- Halliday GM, Lyons JG (2008) inflammatory doses of UV may not be necessary for skin carcinogenesis. *Photochem Photobiol* 84:272–283
- Harrison RJ, Charlesworth B (2011) Biased gene conversion affects patterns of codon usage and amino acid usage in the *Saccharomyces sensu stricto* group of yeasts. *Mol Biol Evol* 28:117–129
- Hooper SD, Berg OG (2000) Gradients in nucleotide and codon usage along *Escherichia coli* genes. *Nucleic Acids Res* 28:3517–3523
- Ingvarsson PK (2008) Molecular evolution of synonymous codon usage in *Populus*. *BMC Evol Biol* 8:307
- Kannaujiya VK, Rastogi RP, Sinha RP (2014) GC constituents and relative codon expressed amino acid composition in cyanobacterial phycobiliproteins. *Gene* 546:162–171
- Knight RD, Freeland SJ, Landweber LF (2001) A simple model based on mutation and selection explains trends in codon and amino acid usage and GC composition within and across genomes. *Genome Biol* 2:1
- Kreil DP, Ouzounis CA (2001) Identification of thermophilic species by the amino acid compositions deduced from their genomes. *Nucleic Acids Res* 29:1608–1615
- Lal D, Verma M, Behura SK, Lal R (2016) Codon usage bias in phylum Actinobacteria: relevance to environmental adaptation and host pathogenicity. *Res Microbiol* 167:669–677
- Lin K, Kuang Y, Joseph JS, Kolatkar PR (2002) Conserved codon composition of ribosomal protein coding genes in *Escherichia coli*, *Mycobacterium tuberculosis* and *Saccharomyces cerevisiae*: lessons from supervised machine learning in functional genomics. *Nucleic Acids Res* 30:2599–2607
- Lloyd AT, Sharp PM (1992) Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Res* 20:5289–5295
- Lobo NF, Behura SK, Aggarwal R, Chen M-S, Collins FH, Stuart JJ (2006) Genomic analysis of a 1 Mb region near the telomere of Hessian fly chromosome X2 and avirulence gene vH13. *BMC Genomics* 7:7
- Lobry JR, Sueoka N (2002) Asymmetric directional mutation pressures in bacteria. *Genome Biol* 3:1
- Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD (1992) Substitutional bias confounds inference of cyanelle origins from sequence data. *J Mol Evol* 34:153–162
- Loomis WF, Smith DW (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. *Proc Natl Acad Sci USA* 87:9093–9097
- Ma M, Hui L, Wang M, Tang Y, Chang Y, Jia Q, Wang X, Yan W, Ha X (2014) Overall codon usage pattern of enterovirus 71. *Genet Mol Res* 13:336–343
- Malakar AK, Halder B, Paul P, Chakraborty S (2016) Cytochrome P450 genes in coronary artery diseases: codon usage analysis reveals genomic GC adaptation. *Gene* 590:35–43
- Mazumder TH, Uddin A, Chakraborty S (2016) Transcription factor gene GATA2: association of leukemia and nonsynonymous to

- the synonymous substitution rate across five mammals. *Genomics* 107:155–161
- Milhon JL, Tracy JW (1995) Updated codon usage in *Schistosoma*. *Exp Parasitol* 80:353–356
- Musto H, Romero H, Zavala A, Jabbari K, Bernardi G (1999) Synonymous codon choices in the extremely GC-poor genome of *Plasmodium falciparum*: compositional constraints and translational selection. *J Mol Evol* 49:27–35
- Morrison DF (1990) *Multivariate statistical methods*. McGraw-Hill Inc., New York
- Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci USA* 84:166–169
- Nair RR, Nandhini MB, Sethuraman T, Dos G (2013) Mutational pressure dictates synonymous codon usage in freshwater unicellular α -cyanobacterial descendant *Paulinella chromatophora* and β -cyanobacterium *Synechococcus elongatus* PCC6301. *Springer plus* 2:492
- Naya H, Romero H, Carels N, Zavala A, Musto H (2001) Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. *FEBS Lett* 501:127–130
- Pouwels PH, Leunissen JA (1994) Divergence in codon usage of *Lactobacillus* species. *Nucl Aci Res* 22:929–936
- Powell JR, Moriyama EN (1997) Evolution of codon usage bias in *Drosophila*. *Proc Natl Acad Sci USA* 94:7784–7790
- Qi Y, Xu W, Xing T, Zhao M, Li N, Yan L, Xia G, Wang M (2015) Synonymous codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity. *Evol Bioinform Online* 11:65
- Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP (2010) Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J Nucleic Acids* 592980:32. doi:10.4061/2010/592980
- Richa, Rastogi RP, Kumari S, Singh KL, Kannaujiya VK, Singh G, Keshri M, Sinha RP (2011) Biotechnological potential of mycosporine-like amino acids and phycobiliproteins of cyanobacterial origin. *Biotechnol Bioinform Bioeng* 1:159–171
- Salamov AA, Solovyev VV (2000) Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res* 10:516–522
- Selby CP, Sancar A (2006) A cryptochrome/photolyase class of enzymes with single-stranded DNA specific photolyase activity. *Proc Nat Acad Sci USA* 103:17696–17700
- Sharp PM, Li WH (1986a) An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol* 24:28–38
- Sharp PM, Li WH (1986b) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res* 14:7737–7749
- Singer GAC, Hickey DA (2000) Nucleotide bias causes a genome wide bias in the amino acid composition of proteins. *Mol Biol Evol* 17:1581–1588
- Sinha RP, Häder DP (2002) UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* 1:225–236
- Sinha RP, Lebert M, Kumar A, Kumar HD, Häder D-P (1995) Spectroscopic and biochemical analyses of UV effect on phycobiliprotein of *Anabaena* sp. and *Nostoc carmum*. *Bot Acta* 108:87–92
- Stege H, Roza L, Vink AA, Grewe M, Ruzicka T, Grether-Beck S, Krutmann J (2000) Enzyme plus light therapy to repair DNA damage in ultraviolet-B-irradiated human skin. *Proc Natl Acad Sci USA* 97:1790–1795
- Sueoka N (1961) Correlation between base composition of deoxyribonucleic acid and amino acid composition and protein. *Proc Natl Acad Sci USA* 47:1141–1149
- Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci USA* 85:2653–2657
- Supriyo C, Prosenjit P, Mazumder TH (2014) Codon usage bias prefers AT bases in coding sequences among the essential genes of *Haemophilus influenzae*. *Notulae Sci Biol* 6:417
- Timares L, Katiyar SK, Elmets CA (2008) DNA damage, apoptosis and langerhans cells—Activators of UV-induced immune tolerance. *Photochem Photobiol* 84:422–436
- Wada A (1992) Compliance of genetic code with base-composition deflecting pressure. *Adv Biophys* 28:135–158
- Wada A, Suyama A (1985) Third letters in codons counterbalance the (G+C)-content of their first and second letters. *FEBS Lett* 188:291–294
- Wan X-F, Xu D, Kleinhofs A, Zhou J (2004) Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol* 4:19
- Wei L, He J, Jia X, Qi Q, Liang Z, Zheng H, Ping Y, Liu S, Sun J (2014) Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evol Biol* 14:1
- Wu Y, Zhao D, Tao J (2015) Analysis of codon usage patterns in herbaceous Peony (*Paeonia lactiflora* Pall.) based on transcriptome data. *Gene* 6:1125–1139
- Yang X, Luo X, Cai X (2014) Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset. *Parasit Vectors* 7:527
- Yu T, Li J, Yang Y, Qi L, Chen B, Zhao F, Bao Q, Wu J (2012) Codon usage patterns and adaptive evolution of marine unicellular cyanobacteria *Synechococcus* and *Prochlorococcus*. *Mol Phylogenet Evol* 62:206–213
- Zheng Y, Zhao WM, Wang H, Zhou YB, Luan Y, Qi M, Cheng YZ, Tang W, Liu J, Yu H, Yu XP, Fan YJ, Yang X (2007) Codon usage bias in *Chlamydia trachomatis* and the effect of codon modification in the MOMP gene on immune responses to vaccination. *Biochem Cell Biol* 85:218–226
- Zhou T, Gu W, Ma J, Sun X, Lu Z (2005) Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. *Biosystems* 81:77–86