academicJournals

Vol. 13(2), pp. 194-201, 8 January, 2014
DOI: 10.5897/AJB2013.13230
ISSN 1684-5315 ©2014 Academic Journals
http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Evaluating codon bias perspective in barbiturase gene using multivariate analysis

Zarrin Basharat

Fatima Jinnah Women University, 46000, Pakistan.

Accepted 5 January, 2014

Barbiturases exist solely in bacteria and encompass an undistinguished protein family. s-Triazine compound introduction into the environment owing to recent industrial practices have revitalized barbiturases. Codon usage patterns were analysed in this study for 17 barbiturase genes encoded by 16 bacterial species. Multivariate and correspondence study of amino acid and codon usage was employed for detecting the cause of variation in the gene content. GC3 of synonymously variable codons, RSCU, NC and CAI were estimated with statistical softwares. Examination of DNA composition along with codon usage was done to reveal dynamics of gene evolution and expression of this enzyme.

Key words: Codon usage, barbiturase gene, multivariate statistical analysis.

INTRODUCTION

Codon penchant arises in the majority species across all kingdoms and shows a discrepancy in genes of same and dissimilar species (Duret, 2002). The possible reasons effecting codon usage pattern are mutations, genetic drift, translational selection and horizontal gene transfer (Ermolaeva, 2001). Comparative sequence analysis is a potent approach for detecting mutations, selection methods and finding rationale of biased and unbiased gene changes and hence, evolutionary processes. Characterization of gene and functional element features in bacteria due to selective pressure is difficult. In the recent era, plenty of algorithms have been developed for such analyses. Sequence information has also been used to infer gene location effect on codon usage for a few prokaryotic species (Sharp and Matassi, 1994). Comparative homologous gene sequence analysis within same and different microbial species facilitates the accomplishment of this task and assists in exposing footprints left by evolution. In case of amino acids encoded by synonymous codons, even if similar, some are used more often than others. Preference of codon usage and its quantification facilitates evolution comprehension of various microorganisms. ‘Codon bias’ statistic approach has been employed (Karlin et al., 1998) to measure the difference in codon usage of barbiturase gene in one species relative to another in this study.

Barbiturases are primarily zinc comprising amidohydralases and possess a key function in the oxidative metabolism of pyrimidines by ring-opening catalysis of barbituric acid to ureidomalonic acid (Soong et al., 2001). Barbiturases exist only in bacteria and share homology with cyanuric acid hydrolases. They comprise of an uncommon protein family without any identifiable linkage to other protein X-ray structural class (Seffernick et al., 2012). Barbiturase is believed to catalyze the intermediate step in oxidative pyrimidine catabolic pathway that is, hydrolytic opening of barbituric acid (Soong et al., 2001, 2002). Barbiturases yield ureidomalonic acid by severing the single amide bond of barbiturate six-membered ring. The barbiturase protein family has been expanded due to sequence annotation using bioinformatic based approaches from sequenced genomes but still rare as appear on average of one in every 200 sequenced genomes (Seffernick et al., 2012).

E-mail: zarrin.iiui@gmail.com.

Barbiturase genes have been investigated for the reason that the adaptation capabilities and adjustment tactics of these enzymes can add immensely to their prospective environmental clean-up use. Due to their sole existence in bacteria, they can shed light on evolution pattern among barbiturases in various species. Codon usage study for any enzyme or protein that exists solely in bacteria does not exist prior and this study is the first of its kind.

Codon usage patterns can be distinguished at three main levels: inter-genomic, intra-genomic and intra-genic (Ermolaeva, 2001; O'Neil et al., 2013). Codon usage was studied within genome of Nocardioides sp. JS614 as well as intergenomic pattern of codon usage was explored across several bacterial genomes. Relative synonymous codon usage, frequency of Guanine and Cytosine at synonymous 3rd position of codon (GC3s), effective number of codons used by gene (Nc) were computed and a multivariate approach that is, correspondence analysis (CA) was employed to assess and visualize systematic synonymous codon usage. CA is a sophisticated way to correlate and visualize complex datasets as it is a graphical approach and reduces multidimensional space into lower dimensional sub-space for best representation of variation among datapoints (Henry, 2007). Trend of variation between barbiturase genes of different species was observed which pointed to either a sign of mutational bias (Levin and Whittome, 2000), translational selection (Grantham et al., 1981), replication-transcriptional selection (McInerney, 1998), different amino acid composition (Wright, 1990) and tRNA abundance or scarcity (Ikemura, 1985; Kanaya et al., 1999). Genes expressed more show superior codon usage bias as compared to poorly expressed genes (Grantham et al., 1981; Kanaya et al., 1999).

Study of codon usage is valuable due to its importance in the perception of the molecular basics plus potential utilization as a molecular marker for typifying molecular evolution. Oligonucleotide probes and DNA primers can also be designed for catching genes of interest based on this knowledge as well (Prabha et al., 2012). Likely selection patterns due to varying environmental factors effect gene expressivity and codon usage trend among genes in different species of bacteria (McInerney, 1997). Codon usage pattern of barbiturases was studied in order to understand the genetic organization and factors shaping the codon usage trend. Our study endorsed this fact that expression intensity and gene length play a key role in codon usage preference.

MATERIALS AND METHODS

Sequence data

All 17 barbiturase gene sequences used in the study were retrieved from NCBI (www.ncbi.nlm.nih.gov) with particular accession numbers mentioned in Table 2.

The parameters relative synonymous codon usage (RCSU), frequency of G and C at synonymous 3rd position of codon (GC3S), frequency of G at synonymous 3rd position of codon (G3S) and frequency of C at synonymous 3rd position of codon (C3S) were determined using program codonW v1.4 (Peden, 1999). CodonO (Wan et al., 2006; Angellotti et al., 2007) was used for obtaining several parameters and comparing them with those obtained by CodonW for verification. Codon usage datasets were normalized using method of Sharp and Li (1987). CAI was also estimated for gene expressivity by CAIcal webserver (Puigbo et al., 2008). Nc codon usage statistics (Wright, 1990) for effective number of codons used were calculated by EMBOSS Chips package. Correspondence analysis (CA) was carried out to inspect and evaluate the difference of RSCU values among genes. CA plot of genes pertaining to their informative codon usage were visualized using PAST (Hammer et al., 2001). Phylogenetic and degeneracy analysis was done using MEGA (Tamura et al., 2011).

RESULTS AND DISCUSSION

GC content is recognized to be a leading driver of codon usage bias in bacteria (Chen et al., 2004). Overall coding GC content in the 17 genes was observed to be 64.95% which is moderately high and is consistent with the observation that these genes have been taken from bacterial genomes with high G+C% (Table 4). Owing to compositional restraint, it is anticipated that meagre biasness or none at all in distribution of A-, T-, G- and C-ending codons exists in the barbiturase genes. Among 19 principally used codons, ten were GC-ending (five C and five G-ending) whereas nine codons were AT-ending (six T- ending and three A-ending) as inferred from data in Table 1.

Synonymous codon usage variation in barbiturase

Nc defines the effective quantity of codons usage by a gene and is usually used for measuring the synonymous codon bias. Nc is not dependent on the composition of amino acid or codon number (Wright, 1990). Nc values vary in range from 20 (in case of single codon usage for every amino acid) to 61 (codon usage with equal chance). Exceedingly influenced genes are usually more expressed. Nc values show general non-uniformity of synonymous codon usage among genes of some species. Nc values of barbiturases ranged from 27.18 to 49.17 with an overall mean value of 35.81. Nc values never approached 61 which is an indicator of non-random or systematic use of sense codons. Nc values were lower indicating stronger bias, but the values also never approached extreme value of 20 showing that only one synonym was not being used for each amino acid. Barbiturase genes varied in their GC content from 0.40 to 0.72 (Table 2), and this biased base composition maybe the cause of values of Nc being lower than 61 in the absence of any selective use of codons but the exact interplay is not fully understood. Nc values are also affected by genomic G+C% leading to high gene expression which in turn leads to exhibition of non-

Table 1. Relative synonymous codon usage of barbiturase gene in 17 species for an average number of 355 codons.

Codon	*f*	RSCU	Codon	*f*	RSCU	Codon	*f*	RSCU	Codon	*f*	RSCU
UUU(F)	1.9	**1.21**	UCU(S)	2.5	0.63	**UAU(Y)**	2.1	**1.21**	UGU(C)	4.4	0.78
UUC(F)	1.2	0.79	**UCC(S)**	4.5	**1.16**	UAC(Y)	1.4	0.79	**UGC(C)**	6.8	**1.22**
UUA(L)	0.8	0.29	UCA(S)	3.2	0.81	UAA(*)	1.7	0.75	**UGA(*)**	4.2	1.84
UUG(L)	1.5	0.57	***UCG(S)***	6.2	**1.58**	UAG(*)	0.9	0.41	UGG(W)	6.7	1
CUU(L)	3.6	**1.36**	CCU(P)	6.3	0.72	**CAU(H)**	8.3	**1.18**	CGU(R)	13.2	0.95
CUC(L)	4.9	**1.82**	CCC(P)	6.6	0.76	CAC(H)	5.7	0.82	***CGC(R)***	17.4	**1.25**
CUA(L)	1.9	0.7	CCA(P)	7.6	0.87	***CAA(Q)***	8.2	**1.3**	***CGA(R)***	22	**1.59**
CUG(L)	3.4	**1.25**	***CCG(P)***	14.5	**1.65**	CAG(Q)	4.4	0.7	***CGG(R)***	20.4	**1.47**
AUU(I)	1.6	1.05	ACU(T)	2.8	0.71	AAU(N)	2.4	0.98	AGU(S)	3.1	0.8
AUC(I)	2.4	**1.5**	ACC(T)	4	1	AAC(N)	2.5	1.02	AGC(S)	4.1	1.03
AUA(I)	0.7	0.45	ACA(T)	3.2	0.79	**AAA(K)**	2.5	**1.18**	AGA(R)	5.8	0.42
AUG(M)	2	1	ACG(T)	6	1.5	AAG(K)	1.8	0.82	AGG(R)	4.4	0.32
GUU(V)	2.6	0.84	GCU(A)	9.2	1.11	***GAU(D)***	9.8	**1.26**	**GGU(G)**	11.2	**1.13**
GUC(V)	5.1	1.61	GCC(A)	6.5	0.78	GAC(D)	5.8	0.74	GGC(G)	11.5	1.16
GUA(V)	1.2	0.37	GCA(A)	7.2	0.87	GAA(E)	4.1	1.22	GGA(G)	10.4	1.05
GUG(V)	3.7	1.18	**GCG(A)**	10.2	**1.23**	GAG(E)	2.6	0.78	GGG(G)	6.6	0.66

f refers to codon frequencies and are count of frequency averages over taxa. RSCU>1.2 means codon usage with high frequency. Data in bold, codons with high frequency. Data in italics, codons with extremely high frequency (RSCU>2.0).

Table 2. Summary of GC content, GC3s, CAI, NC (effective number of codons), number of codons for each gene and gene length of the 17 nucleotide sequences dataset of barbiturase genes. gi|119714272(2) refers to second barbiturase gene found in same bacterial specie.

Serial number	Accession number	Bacterial species name	GC	GC3s	CAI	Nc	Number of codons	Gene length
1	gi|256389232	*Catenulispora acidiphila* DSM 44928	0.720867	0.95664	0.769	29.547	369	1107
2	gi|312193897	*Frankia* sp. EuI1c	0.71526	0.913043	0.745	31.223	391	1173
3	gi|284988629	*Geodermatophilus obscurus* DSM 43160	0.711957	0.975543	0.767	28.299	368	1104
4	gi|257054089	*Saccharomonospora viridis* DSM 43017	0.676367	0.899471	0.781	33.389	378	1134
5	gi|271961609	*Streptosporangium roseum* DSM 43021	0.711538	0.972527	0.823	28.114	364	1092
6	gi|379005799	*Sulfobacillus acidophilus* DSM 10332	0.590909	0.659091	0.675	49.171	352	1056
7	gi|317126741	*Bacillus cellulosilyticus* DSM 2522	0.40634	0.270893	0.608	47.264	347	1041
8	gi|226359415	*Rhodococcus opacus* B4	0.667568	0.867568	0.733	36.617	370	1110
9	gi|226303489	*Rhodococcus erythropolis* PR4	0.633333	0.764865	0.711	40.457	370	1110
10	gi|119714272	*Nocardioides* sp. JS614	0.706093	0.981183	0.787	27.744	372	2232
11	gi|119714272(2)	*Nocardioides* sp. JS614	0.709677	0.983871	0.806	27.145	372	1116
12	gi|83588874	*Moorella thermoacetica* ATCC 39073	0.479167	0.399457	0.645	57.361	368	1104
13	gi|389861738	*Modestobacter marinus*	0.708333	0.98913	0.809	27.467	368	1104
14	gi|317123177	*Intrasporangium calvum* DSM 43043	0.693931	0.939314	0.775	31.675	379	1137
15	gi|434390831	*Gloeocapsa* sp. PCC 7428	0.468416	0.411079	0.601	48.501	343	1029
16	gi|379733551	*Blastococcus saxobsidens* DD2	0.703804	0.967391	0.760	28.463	368	1104
17	gi|148251626	*Bradyrhizobium* sp. BTAi1	0.701519	0.855228	0.721	36.427	373	1119

random codon usage because the codons corresponding to most abundant tRNA species are most preferentially used. Optimal codon usage of studied bacteria seems to be mostly due to translational selective pressure. The G+C base composition of 'silent' third positions expressed as the GC3s value was calculated (Table 2). Base composition and GC3s is generally described to have high authority on codon usage.

Isochoric presence has been inferred due to GC3s disparity among prokaryotic genes (D'onofrio and Bernardi, 1992; Sueoka, 1992). Nc and GC3s was used for further analysis of assorted codon usage extent. Genes with weak codon bias display GC3s variation which indicates relation to chromosomal position, with a lower GC3s near the terminus of replication (Deschavanne and Filipski, 1995). GC3s is presumed to be variable among species by a factor of ten and such preference is inclined towards mutational bias. Value of GC3s ranged from 0.27 to 0.97 with a mean value of 0.81. This reflected that there is a mutational bias existing among species and a noticeable manifest of codon usage variation in the genes among few bacterial species while most species showed similar pattern of codon usage.

Mutational bias effect on codon usage variation of barbiturse

To uncover out the grounds of codon usage variation, Nc plots (a plot of Nc vs. GC3s) and correspondence analysis method was used. The Nc plot of the genes suggests that maximum points lie away from the expected curve with three bacterial species (*Bacillus cellulosilyticus* DSM 2522, Streptosporangium roseum DSM 43021 and *Moorella Thermoacetica* ATCC 39073) towards GC poor region while the rest of the species form a cluster towards GC rich region (Figure 1). Almost all genes showed GC skew. GC skew refers to compositional constraint towards or away from G+C nucleotide pairs due to mutational biases (Wolfe et al., 1989) or natural selection (Bernardi et al., 1993). Most of the genes lying away from the curve of the predicted values as shown in Figure 1 seemed to have codon selection constrained by a G+C mutation bias. This can be used to infer that impact of mutational bias is robust on codon usage difference in barbiturases of twelve species, but it is also an indicator that GC mutational bias is not the only factor shaping the codon usage because genes having codon choice constrained merely by a G+C mutation bias, will lie on or just underneath the predicted values curve (Wright, 1990) and as apparent from Figure 1, most of the species lie far away from the predicted curve line. It is suggested that GC mutational bias maybe the sole cause of codon usage variation in only *Bacillus cellulosilyticus* DSM 2522 and *Rhodococcus erythropolis* PR4.

Observing position of demonstrating points that lie away from the expected line, it is also suggested that mutational bias effect on codon use discrepancy of these bacteria is not uniform in *Bacillus cellulosilyticus* DSM 2522, *Rhodococcus erythropolis* PR4, *Gloeocapsa* sp. PCC 7428, *Sulfobacillus acidophilus* DSM 10332, *Moorella thermoacetica* ATCC 39073, *Bradyrhizobium* sp. BTAi1, *Saccharomonospora viridis* DSM 43017 and *Frankia* sp. EuI1c, whereas it is near uniform in *Nocardioides* sp. JS614, *Modestobacter marinus*, *Rhodococcus opacus* B4 and *Bacillus cellulosilyticus* DSM 2522. This helps us infer that there is a variation between barbiturase genes in most species as apparent from the discrepancy between positions of demonstrating points (Figure 2).

Relative synonymous codon usage is calculated as the ratio of observed codon frequency to the frequency expected if codon usage was uniform. RSCU values close to 1.0 indicated that this codon usage occurred at expected frequency. Higher values indicated more frequent codon usage than the expectation. The correspondence analysis of RSCU values of 17 Barbiturase coding genes (of the above 16 bacterial species) states that mutational bias is effecting a great deal on the codon usage as most of the demonstrating points lie away from the expected curve line. Correspondence analysis (CA) is a multivariate statistical analysis method employed for studying codon usage differences between genes (Wright, 1990).

To investigate whether amino acid composition has an influence on synonymous codon usage, correspondence analysis on relative synonymous codon usage values was performed. Correspondence analysis is a refined technique with codon usage data plotted in a 59 axes area (Met, Trp and stop codons are excluded) and an axis is identified which is a representation of the major factors causing gene variance. Gene positions of two major axes are shown in Figure 3. The quantity as well as occurrence of all codons and their RSCU values for 17 barbiturase genes with maximum codon biases are shown as Table 1. Gene length was found to be maximum for *Nocardioides* sp. JS614 with 2232 amino acids and least for *Gloeocapsa* sp. PCC 7428 with 1029 amino acids. Observing the length and location of other genes on the graph, it is suggested that the variations in codon usage of barbiturases are not correlated with nucleotide composition of the genes. However, the position of two genes of the same bacterial species that is, *Nocardioides* sp. JS614 near axis2 (Figure 2) shows that they are clustered.

Translational selection effect on codon usage variation in barbiturase

Translational selection influence on codon usage variation in barbiturase coding genes was investigated by scatter plot. A scatter plot of the gene demonstrating

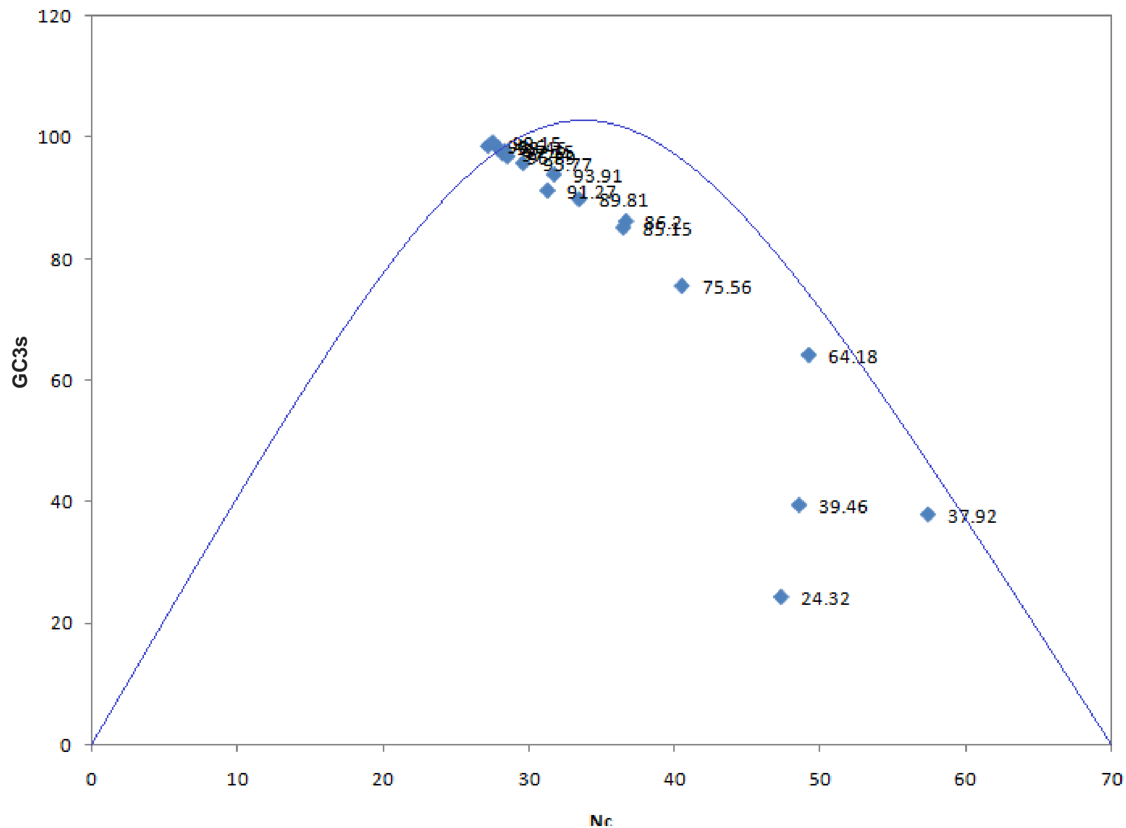

Figure 1. Nc versus GC3s plot (Nc denotes the effective number of codons of each gene, and GC3s denotes the G+C content on the third synonymous codon position of each gene) of *barbiturase* genes. The solid line indicates the expected ENc value if the codon bias is only due to GC3s. Genes shown by respective serial numbers allotted in Table 2.

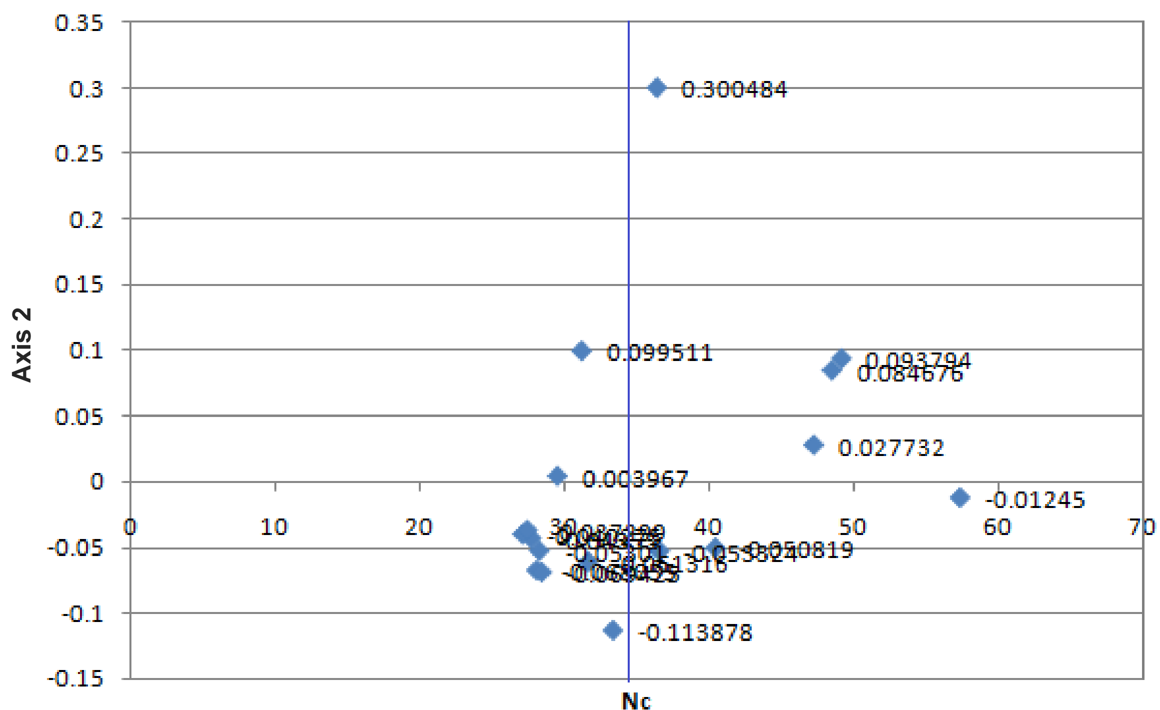

Figure 2. Nc versus Axis 2. Genes position shown by respective serial numbers allotted in Table 2.

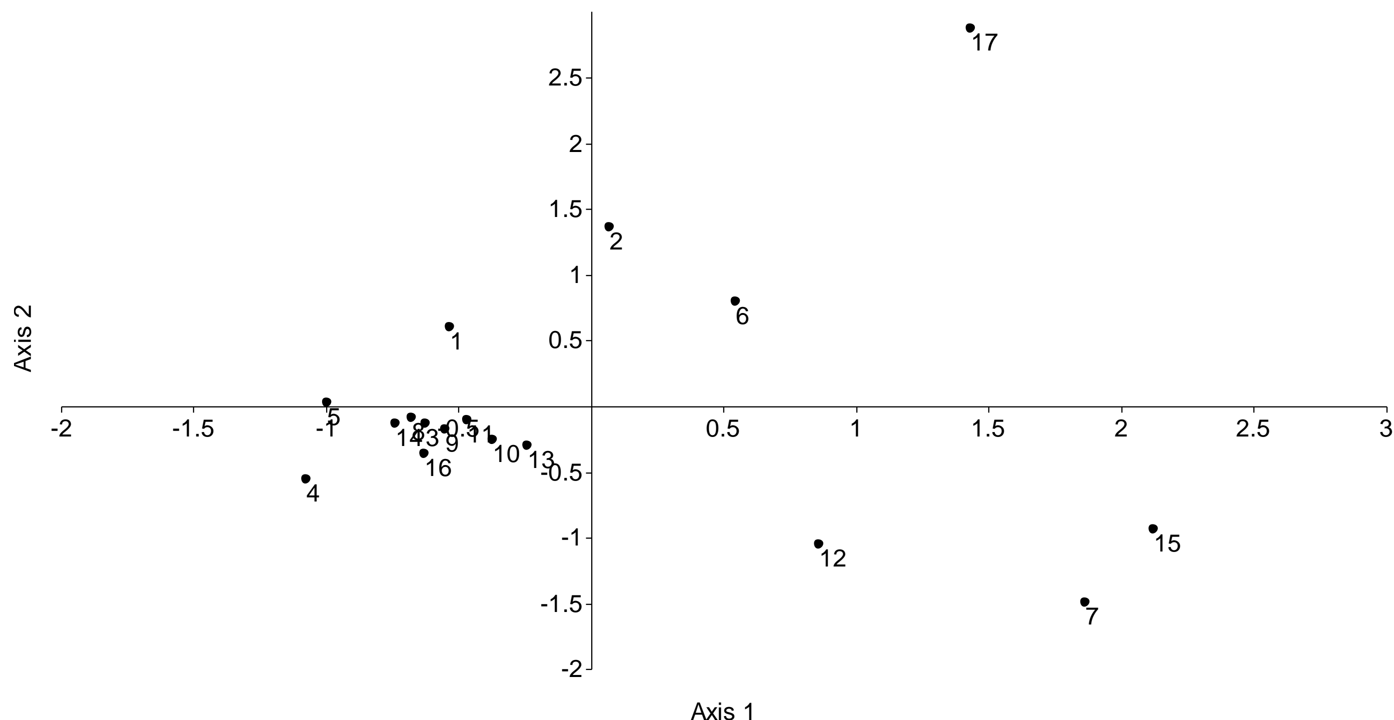

Figure 3. Position of barbiturase genes along the major axes (1 and 2) of variation in the correspondence analysis of relative synonymous codon usage. Genes shown by respective serial numbers allotted in Table 2.

points along the 2nd major axis and Nc values is shown (Figure 2). It can be seen that barbiturase gene from *Bradyrhizobium* sp. BTAi1 has maximum Nc value in the entire model system and the rest with lower Nc values show less bias as compared to *Bradyrhizobium* sp. BTAi1 barbiturase.

Evolutionary analyses

Evolutionary analyses was conducted in MEGA5 and maximum composite likelihood estimate of the pattern of nucleotide substitution of 17 nucleotide sequences was estimated to be 0.08, 0.30, 0.65, 1.21 and 2.77 substitutions per site. The nucleotide frequencies of A = 18.29%, T/U = 16.64%, C = 31.01% and G = 34.07% were found. Average codon usage bias was found to be 44%. The number of base substitutions per site from averaging over all sequence pairs was calculated to be 0.598. The transition/transversion rate ratios are k_1 = 1.233 (purines) and k_2 = 1.424 (pyrimidines). The overall transition/transversion bias is R = 0.602, where $R = [A*G*k_1 + T*C*k_2]/[(A+G)*(T+C)]$. Each entry shows the probability of substitution (r) from one base (row) to another base (column) in Table 3. For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in *italics*. By estimating maximum likelihood values, a tree topology was automatically computed. The maximum log likelihood for this computation was -11559.606 involving 17 nucleotide sequences.

Table 3. Maximum composite likelihood estimate of the pattern of nucleotide substitution.

	A	T	C	G
A	-	5.01	9.33	**12.63**
T	5.5	-	**13.28**	10.25
C	5.5	**7.13**	-	10.25
G	**6.78**	5.01	9.33	-

Codon positions included in the analysis were 1st, 2nd, 3rd and non-coding with a total of 911 positions in the ultimate dataset with gap-containing and missing data elimination. Evolutionary record was deduced by means of the maximum likelihood method utilizing Tamura-Nei model (Tamura and Nei, 1993). The tree with the highest log likelihood (-11949.0384) is shown (Figure 4).

Conclusion

An inquiry on synonymous codon bias of barbiturase

Table 4. Genomic G+C% of bacterial genomes studied for barbiturase gene codon usage. Two barbiturase genes encoded by Nocardioides sp. JS614 have been used in the study.

Serial number	EMBL genome databank accession number	Bacterial species name	Genome length (bp)	Genomic G+C%
1	CP001700	*Catenulispora acidiphila* DSM 44928	10,467,782	69.8
2	CP002299	*Frankia* sp. EuI1c	8,815,781	72.3
3	CP001867	*Geodermatophilus obscurus* DSM 43160	5,322,497	74.0
4	CP001683	*Saccharomonospora viridis* DSM 43017	4,308,349	67.3
5	CP001814	*Streptosporangium roseum* DSM 43021	10,341,314	70.9
6	CP003179	*Sulfobacillus acidophilus* DSM 10332	3,472,898	56.8
7	CP002394	*Bacillus cellulosilyticus* DSM 2522	4,681,672	36.5
8	AP011115	*Rhodococcus opacus* B4	7,913,450	67.6
9	AP008957	*Rhodococcus erythropolis* PR4	6,516,310	62.31
10	CP000509	*Nocardioides* sp. JS614	4,985,871	71.65
11	CP000232	*Moorella thermoacetica* ATCC 39073	2,628,784	55.8
12	FO203431	*Modestobacter marinus*	5,575,517	74.13
13	CP002343	*Intrasporangium calvum* DSM 43043	4,024,382	70.7
14	CP003646	*Gloeocapsa* sp. PCC 7428	5,431,448	71.16
15	FO117623	*Blastococcus saxobsidens* DD2	4,875,340	72.95
16	CP000494	*Bradyrhizobium* sp. BTAi1	8,264,687	66

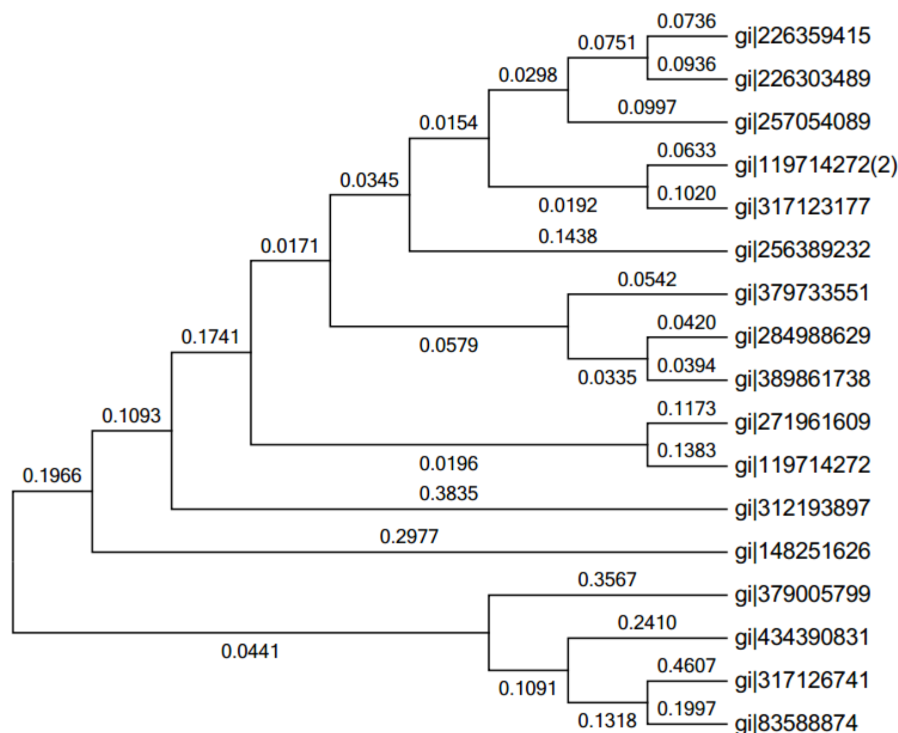

Figure 4. Phylogenetic tree. Genes shown by respective gene accession numbers shown in rectangular boxes with species names in Table 2. The second gene from Nocardioides sp. JS614 is shown as gi|119714272(2).

gene was carried out in the present work. Codon usage of 17 barbiturase genes from 16 different bacterial species were explored and a relative examination of codon usage along with RSCU values of barbiturase coding gene was done. All of the genes were found to yield vivid expression as inferred from codon adaptation index (CAI) used extensively as a gauge to study gene expression level of organisms. Mutational bias effect study on variation in codon usage of the barbiturase gene suggested overall strong biasing in all the genes studied. Mutations effects all organisms and contour gene codon usage in the absence of selection (Nichols et al., 1980; Sueoka, 1988). Relation between gene expression level and optimal codons can partially explain common features of codon use. Based on the analysis, it can be concluded that translational selection has also influenced codon usage variation in case of gene expression in barbiturases. The calculated values suggest that barbiturases are an ancient set of genes with a strong mutational bias along with translational selection mode playing a role in their continued existence. This can be used to postulate that there is a co-adaptation of synonymous codon usage to optimize translation efficiency in barbiturases.

REFERENCES

Angellotti MC, Bhuiyan SB, Chen G, Wan X (2007). CodonO: codon usage bias analysis within and across genomes. Nucleic Acids Res. 35:W132-W136.

Bernardi G, Mouchiroud D, Gautier C (1993). Silent substitutions in mammalian genomes and their evolutionary implications. J. Mol. Evol. 37:583-589.

Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH (2004). Codon usage between genomes is constrained by genome-wide mutational processes. Proc. Natl. Acad. Sci. USA 101:3480-3485.

Deschavanne I, Filipski J (1995). Correlation of GC content with replication timing and repair mechanisms in weakly expressed E. co/i genes. Nucleic Acids Res. 23:1350-1353.

D'Onofrio G, Bernardi G (1992). A universal compositional correlation among codon positions. Gene 110(1):81-8.

Duret L (2002).Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. 12:640-649.

Ermolaeva MD (2001). Synonymous codon usage in bacteria. Curr. Iss. Mol. Biol. 3:91-97.

Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. 9:43-74.

Hammer Ø, Harper DAT, Ryan PD (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4(1):9.

Henry I (2007). Evolution of Codon Usage in Bacteria. Ph.D thesis. University of Nottingham.U.K.

Ikemura T (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2:13-34.

Kanaya S, Yamada Y, Kudo Y, Ikemura T (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene 238:143-155.

Karlin S, Campbell AM, Mrázek J (1998). Comparative DNA analysis across diverse genomes. Annu. Rev. Genet. 32:185-225.

Levin DB, Whittome B (2000). Codon usage in nucleopolyhedroviruses. J. Gen. Virol. 81:2313-2325.

McInerney JO (1997). Prokaryotic genome evolution as assessed by multivariate analysis of codon usage patterns. Microb. Compar. Genomics 2(1):1-10.

McInerney JO (1998). Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. Proc. Natl. Acad. Sci. USA 95: 10698-10703

Nichols BP, Miozzari GF, Cleemput MV, Bennet GN, Yanofsky C (1980). Nucleotide sequences of the trpG regions of *Escherichia coli, Shigella dysenteriae, Salmonella typhimurium* and *Serratia marcescens*. J. Mol. Biol. 142:503-517.

O'Neil PK, Or M, Erill I (2013). scnRCA: A Novel Method to Detect Consistent Patterns of Translational Selection in Mutationally-Biased Genomes. PLoS ONE 8(10):e76177.

Peden JF (1999). Analysis of Codon Usage, Ph.D. Thesis, University of Nottingham.

Prabha R, Singh DP, Gupta SK, Farooqi S, Rai A (2012). Synonymous codon usage in *Thermosynechococcus elongatus* (cyanobacteria) identifies the factors shaping codon usage variation. Bioinformation 8(13):622-628.

Puigbo P, Bravo IG, Garcia-Vallve S (2008) CAIcal: a combined set of tools to assess codon usage adaptation. Biol. Direct 3(1):38.

Seffernick JL, Erickson JS, Cameron SM, Cho S, Dodge AG, Richman JE, Sadowsky MJ, Wackett LP (2012). Defining Sequence Space and Reaction Products within the Cyanuric Acid Hydrolase (AtzD)/Barbiturase Protein Family. J. Bacterial. 194(17):4579 - 4588.

Sharp PM, Li WH (1987). The codon adaptation index—a measure of directional synonymous codon usage bias and its potential applications. Nucleic Acids Res. 15:1281-1295.

Sharp PM, Matassi G (1994). Codon usage and genome evolution. Curr. Opin. Genet. Dev. 4 (6):851-860.

Soong CL, Ogawa J, Sakuradani E, Shimizu S (2002). Barbiturase, a novel zinc-containing amidohydrolase involved in oxidative pyrimidine metabolism. J. Biol. Chem. 277:7051-7058.

Soong CL, Ogawa J, Shimizu S (2001). Novel amidohydrolytic reactions in oxidative pyrimidine metabolism: analysis of the barbiturase reaction and discovery of a novel enzyme, ureidomalonase. Biochem. Biophys. Res. Commun. 286:222-226

Sueoka N (1988). Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA 85(8): 2653-2657.

Sueoka N (1992). Directional mutation pressure, selective constraints and genetic equilibria. J. Mol. Evol. 34:95-114

Tamura K, Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512-526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28:2731-2739.

Wan XF, Zhou J, Xu D (2006). CodonO: a new informatics method measuring synonymous codon usage bias. Int. J. Gen.Sys. 35:109-125.

Wolfe KH, Sharp PM, Li WH (1989). Mutation rates differ among regions of the mammalian genome. Nature 337:283-285.

Wright F (1990).The effective number of codons used in a gene. Gene 87(1):23-29