DNA Research, 2016, 23(5), 441–449
doi: 10.1093/dnares/dsw027
Advance Access Publication Date: 17 July 2016
Full Paper

OXFORD

Full Paper

Discrepancy among the synonymous codons with respect to their selection as optimal codon in bacteria

Siddhartha Sankar Satapathy[1], Bhesh Raj Powdel[2], Alak Kumar Buragohain[3,4], and Suvendra Kumar Ray[3,*]

[1]Department of Computer Science and Engineering, Tezpur University, Napaam, Tezpur 784028, Assam, India, [2]Department of Statistics, Darrang College, Tezpur 784001, Assam, India, [3]Department of Molecular Biology and Biotechnology, Tezpur University, Napaam, Tezpur 784028, Assam, India, and [4]Office of the Vice-Chancellor, Dibrugarh University, Dibrugarh 786004, Assam, India

*To whom correspondence should be addressed. Tel. (+91) 3712 275406. Fax. (+91) 3712 - 267005/267006. Email: suven@tezu.ernet.in

Edited by Prof. Kenta Nakai

Received 6 January 2016; Accepted 19 May 2016

Abstract

The different triplets encoding the same amino acid, termed as synonymous codons, are not equally abundant in a genome. Factors such as G + C% and tRNA are known to influence their abundance in a genome. However, the order of the nucleotide in each codon *per se* might also be another factor impacting on its abundance values. Of the synonymous codons for specific amino acids, some are preferentially used in the high expression genes that are referred to as the 'optimal codons' (OCs). In this study, we compared OCs of the 18 amino acids in 221 species of bacteria. It is observed that there is amino acid specific influence for the selection of OCs. There is also influence of phylogeny in the choice of OCs for some amino acids such as Glu, Gln, Lys and Leu. The phenomenon of codon bias is also supported by the comparative studies of the abundance values of the synonymous codons with same G + C. It is likely that the order of the nucleotides in the triplet codon is also perhaps involved in the phenomenon of codon usage bias in organisms.

Key words: codon usage bias; optimal codon; genome composition; high expression genes; bacteria

1. Introduction

In the genetic codes, although synonymous codons encode the same amino acid, these codons are not used with equal proportions. Certain synonymous codons, known as the optimal codons (OCs), are preferred to the other codons, as witnessed more prominently in the high expression genes (HEGs) in genomes.[1,2] The selection on OCs in the HEGs is reportedly a function of cellular adaptation for faster and/or more accurate translation.[3–7] Different factors such as genome G + C%,[8] anticodon modification and codon–anticodon interaction[4,9–12] and tRNA gene numbers[13–17] are believed to influence OC selection in the HEGs in bacteria. In addition to the above, selection on mRNA secondary structure and stability,[18–22] protein structure,[23] protein folding kinetics,[6,24–26] are known to influence codon usage bias in genes.

Whether the selection of OCs is dependent upon these extrinsic factors mentioned above or upon the triplet sequence itself, has not been explored in detail. One of the approaches to find out the influence of the nucleotide sequence itself on their selection as OCs for different amino acids is to study and compare the OCs for different

amino acids across different bacteria. A simple assumption is that the extrinsic factors influencing codon usage bias is likely to vary across bacteria but the intrinsic feature of a codon remains invariable. In this regard, there are several reports of analysis of the OCs in bacteria. Sharp et al. (2005)[27] observed that the C-ending codons of Phe, Tyr, Asn and Ile are always used as the OCs in preference to the U-ending codons in different species of bacteria. The findings of Sharp et al. (2005)[27] were in support of the earlier observation that growth rate influences the strength of translational selection in bacteria.[15] So, it is the C-ending codons in these amino acids that are selected as the OCs across bacteria.[27–29] Sharp et al. (2005)[27] analysed only the WWY codons of the four amino acids that are decoded by one anticodon each. Whether the observation is also valid for the other three amino acids such as His, Asp and Cys that are encoded only by the NNY codons was not explored in their study.

In case of amino acids encoded by six or four synonymous codons, there are isoacceptor tRNAs in an organism for decoding these codons. The variation of the copy number of the isoacceptor tRNAs, multiple codon–anticodon interactions increase the variability of the extrinsic mechanisms in the selection of the OC in a bacterium. Recently, we reported that GGU and CGU with respect to Gly and Arg, respectively are selected as the OCs in bacteria.[30] This finding is important because both Gly and Arg are encoded by the codons belonging to the 4- and 6-fold degenerate families, respectively. This study[30] indicates that the sequence of nucleotides in a triplet is important for OC selection.

Published articles on the roles of the extrinsic factors such as genome composition and tRNA gene number on the selection of OCs have not provided any conclusion in this regard.[8,31] To address the above issue, we analysed the codon usage bias in all the 18 amino acids (Met and Trp encoded by one codon each were not considered) in 221 different species of bacteria belonging to 15 different phylogenetic groups. Comparison of the codon usage bias between the HEGs and the whole genome *vis-à-vis* the genome composition suggested that there is amino acid specific influence with regard to OC selection across bacteria. What is interesting is that only some of synonymous codons are selected as the OCs across bacteria. It was also observed that OCs for some amino acids can be different among bacteria belonging to different phylogenetic groups. Observations in this study indicate that the intrinsic differences among the synonymous codons may contribute towards their abundance values in a genome and in the choice of specific synonymous codon as the OC with respect to particular amino acid.

2. Materials and methods

Gene sequences of the bacterial genomes considered in the present study were downloaded from the DDBJ site (www.gib.genes.nig.ac.jp) (the gene sequences were taken from the site in 2011). Several HEG sequences (ribosomal protein, translation factor, *rpoB* and *rpoC* gene sequences)[27] were also taken from the NCBI site (www.ncbi.nlm.nih.gov). The total tRNA gene numbers were collected from the Genomic tRNA Database (http://gtrnadb.ucsc.edu), which uses tRNAscan-SE to classify tRNA into different groups on the basis of their anticodon sequences.[32] In this study, we considered 221 unique species of bacteria whose tRNA gene numbers and gene annotations were available in the online databases. These bacteria were with genomic G + C% between 72.83 and 22.44 and belong to 15 different phylogenetic groups of bacteria (Table 1b).

In general, the codons present in greater abundance and therefore used more frequently in the HEGs in comparison to all the genes in the genome are considered as the OCs.[4,30] In our study here, to avoid the borderline cases, we considered those codons as the OCs, which are with abundance values of 5% or higher in the HEGs than the same with respect to all the genes of the genome. Codon abundance values, synonymous codon frequency ranks and the percentage of amino acid usage were calculated using a computer programme written in the C language. Figures were plotted using the Microsoft Excel. The relative synonymous codon usage values (RSCU) of codons were calculated according to the formula given by Sharp and Li (1986).[33] $RSCU_{diff}$ is defined as the subtraction value between the RSCU value of a codon in the HEGs and the RSCU value of a codon in the genome.

3. Results

3.1. Optimal codons for different amino acids do not follow a common pattern with regard to genome composition

The study was based on the codon usage bias in 221 bacteria (Supplementary Table 1). The bacteria considered in this study included single strain from different species and also contained bacteria from 15 different phylogenetic groups. The 221 bacteria were divided into two groups depending on the total number of tRNA genes present in their genomes: (i) the high tRNA number (HTN) group having a number of tRNA genes ≥ 50 and (ii) the low tRNA number (LTN) group having the number of tRNA genes < 50. Bacteria included in the HTN were presumed to have a strong translational selection in codon usage bias and those included in the LTN were presumed to have a weak translational selection as had been defined earlier.[15,31] Both HTN and LTN bacteria were further arranged in different groups according to their genome G + C%. As the maximum G + C% known in bacterial genomes is 75%, five different groups were made ranging from the very high (VH) G + C% to the very low (VL) G + C% genomes as follows: VH (G + C% ≥ 65), high (H; $55 \leq$ G + C% < 65), moderate (M; $45 \leq$ G + C% < 55), low (L; $35 \leq$ G + C% < 45) and VL (G + C% < 35). The HTN included total 119 bacteria whereas the LTN included 102 bacteria (Table 1a).

In several instances, bacteria belonging to different phylogenetic groups are also found with restricted genome G + C composition. For example, Actinobacteria and β-Proteobacteria are with high genome G + C composition, whereas Firmicutes and Tenericutes are with low genome G + C composition. In this study, the number of bacteria from the different phylogenetic groups considered for analysis of OC is variable. To find out if there is specific OC choice in bacteria according to their phylogenetic group, that might have been overlooked during analysis by dividing the bacteria according to genome G + C composition as described above, analysis of OCs was also performed in these bacteria by dividing them into different groups according to their phylogeny (Table 1b). Therefore, the OCs for the different amino acids were found out in the above bacteria on the basis of genome composition as well as phylogenetic affiliations (Table 1a and b; Fig. 1a and b). The OC compositions in different amino acids grouped under different degenerate codons were as follows.

3.2. Optimal codons for the amino acids encoded only by the Y-ending codons

Only the C-ending codons were found as the OCs for the five amino acids, *viz*, Asn, Asp, His, Phe and Tyr (Table 2) in all the bacteria irrespective of the genome G + C compositions. The observation was also consistent across the different phylogenetic groups (Table 2). Selection

of the C-ending codons in these amino acids can be explained on the basis of codon–anticodon interaction.[28] The only anticodon G_{34} (G at the first anticodon position) is used to decode both the synonymous codons of an amino acid. The C_3:G_{34} (C_3: C at the third codon position; G_{34}: G at the first anticodon position) pairing at the wobble position is preferred to the U_3:G_{34} pairing at the same position. But the same explanation is not tenable in case of the Cys codons. In the case of Cys, though the C-ending codon was found to be the OC in the high G + C% genomes as we observed for the other amino acids discussed above, the U-ending codon was the OC in the low G + C% genomes. Our observation here for the Cys codons is consistent with the observations made earlier by Iriarte et al. (2013)[34] in the family Enterobacteriaceae. Usually, the abundance of Cys is not high in the proteome of an organism. Whether this is the reason for the exceptional behaviour of the Cys codons in the low G + C% genomes is not known. In case of *Saccharomyces cerevisiae*, UGU was observed as the selected codon in the HEGs.[35,36] Wald et al. (2012)[29] gave an additional explanation of translational accuracy to explain the selection of C_3 to U_3 in amino acids encoded only by the Y-ending codons in the split codon boxes (the four codons in a box are not synonymous; His and Gln are encoded by split codon box). In a split codon box, usually the U_{34} anticodon decodes the NNR codons (the Gln codons, for example) while the G_{34} anticodon decodes NNY codons (the His codons, for example). While the G_{34} accurately decodes the NNY codons, the U_{34} apart from accurately decoding the NNR codons has a possibility of wrongly coding the U_3 codon of the split codon box. So to avoid the translational error in the HEGs due to mispairing of the U_{34} anticodon with the U_3 codon of a split codon box, the C_3 codons are selected more than the U_3 codons in these genes. As the UGG codon for Trp (Trp and Cys are part of the same spilt codon box) is decoded by the C_{34} anticodon, the greater abundance of the U_3 codon Cys is not likely to be a problem for the cell. If it is true, then the observations in the case of Cys should also have been extended to the Tyr (UAY) and Ile (AUY) codons as there is no U_{34} anticodon in these two split codon boxes. While in the case of Tyr, the selection of U-ending codon was not observed in general, in case of Ile, the selection of U-ending codon was observed in some bacterial groups. It is pertinent to note that there is a strong G:C base pairing at the second codon position in case of the Cys codons, and there is absence of any G:C base pairing at the first and the second codon positions in case of Ile and Tyr codons, which makes the C-ending codon more important for Ile and Tyr than in the case of Cys. The other possibility that selection of the U-ending codons in the HEGs might be an adaptation against the non-sense error,[37] i.e. the release factor has an erroneous preference to UGC than to UGU to cause premature translation termination. But this argument contradicts the selection of UGC in the high G + C genomes and also the selection of the C-ending codon in case of Tyr, which is encoded by the spilt codon box having stop codons. The other possibility of the selection of OCs in the direction of genome composition may be attributed to more stable mRNA secondary structure of HEGs.

Table 1a. Number of bacteria studied in the HTN and the LTN groups

G + C%	HTN	LTN
VH	21	13
H	30	8
M	24	10
L	32	35
VL	12	36
Total	119	102

VH (very high; G + C% ≥ 65), H (high; 55 ≤ G + C% < 65), M (moderate; 45 ≤ G + C% < 55), L (low; 35 ≤ G + C% < 45) and VL (very low; G + C%<35).

3.3. Optimal codons for the amino acids encoded only by the R-ending codons

In case of Glu, the A-ending codon was found as the OC across all G + C compositions. In case of Gln and Lys, the G-ending codons were found as the OC in G + C% high genomes while the A-ending codons were found as the OC in low G + C% genomes. Observations in the HTN and LTN were similar. The analysis in bacteria revealed that the selection of OCs was phylogeny specific in case of Glu. For example, in Actinobacteria (high G + C) and Spirochaetes (low G + C) the GAG was the OC whereas in β-Proteobacteria (high G + C) and Firmicutes (low G + C) the GAA was the OC. In case of y-Proteobacteria, GAA is selected as the OC, irrespective of genome composition. Glu is an example where the

Table 1b. Bacteria belonging to different phylogenetic groups and their genomic G + C% considered in the present study

S. No.	Group	No. of strains	Maximum G + C%	Minimum G + C%	Average G + C%
1	Actinobacteria	23	72.83	46.31	64.45
2	α-Proteobacteria	35	68.79	27.51	48.11
3	Aquificae	1	43.3	43.3	43.3
4	Bacteroidetes	8	66.13	22.44	40.26
5	β-Proteobacteria	17	68.49	50.72	63.47
6	Chlamydiae	6	41.31	39.19	40.10
7	Chlorobi	1	56.53	56.53	56.53
8	Chloroflexi	1	47.03	47.03	47.03
9	Cyanobacteria	8	62.00	31.32	47.09
10	Δ-Proteobacteria	6	71.38	50.65	62.27
11	ϵ-Proteobacteria	11	44.54	27.05	36.56
12	Firmicutes	38	56.98	28.21	38.44
13	γ-Proteobacteria	47	68.67	22.48	48.19
14	Spirochaetes	5	40.24	27.77	33.81
15	Tenericutes	14	40.01	23.77	28.58

Note: Aquificae, Chlorobi and Chloroflexi groups consisting of only one bacterium each are not considered in the result analysis shown in Fig. 1b.

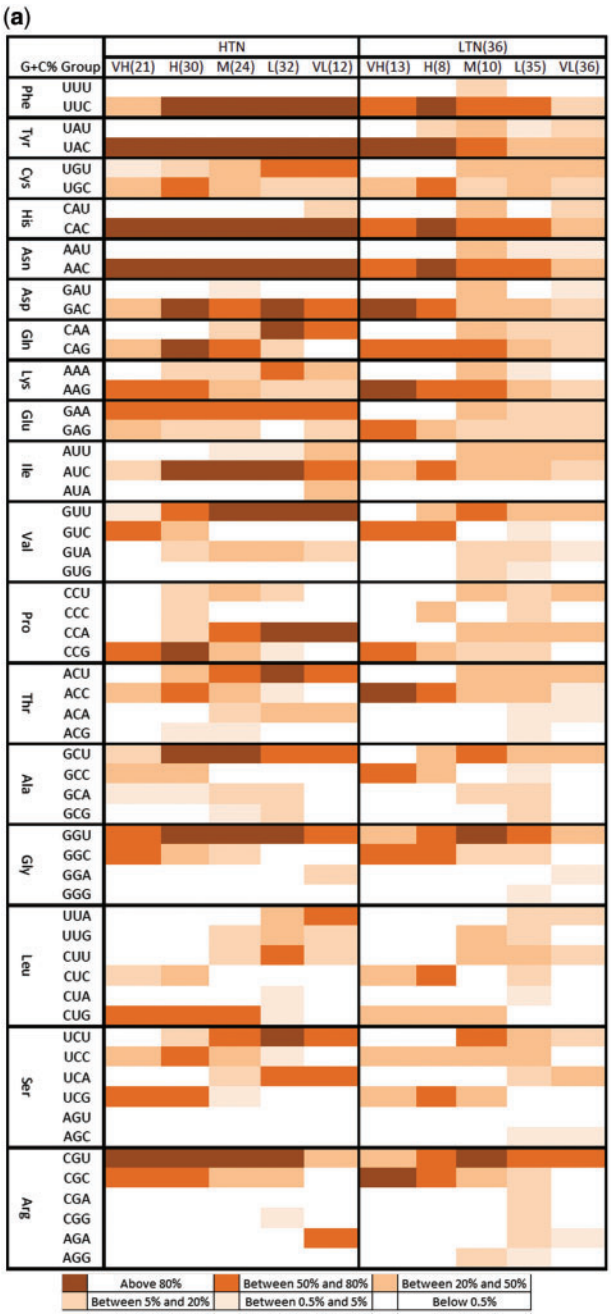

Figure 1a. Optimal codons of 18 amino acids in the HTN and in the LTN groups of bacteria within different genomic G + C% groups. A codon is considered as the OCs if it is used more frequently in the set of HEGs in comparison to all the genes in a genome. In our study here, to avoid any borderline cases, we considered those codons as the OCs, which are with abundance values of 5% or higher in the HEGs than the same with respect to all the genes of the genome. OCs were found out in 221 bacteria (Supplementary Table 1). The bacteria were considered in two groups: (A) HTN and (B) LTN. Further, in each of the two groups, bacteria were grouped into five subgroups according to their genome G + C% (i) VH, 65.0 ≤ G + C%; (ii) H, 55.0 ≤ G + C% < 65.0; (iii) M, 45.0 ≤ G + C% < 55.0; (iv) L, 35.0 ≤ G + C% < 45.0 and (v) VL, G + C% < 35.0. Numbers of bacteria considered in each subgroup were given in Table 1. The colour code is used to represent % of bacteria in the G + C% group where the codon was observed more frequent in the HEGs than the whole genome. For the raw data, please refer to Supplementary Table 2.

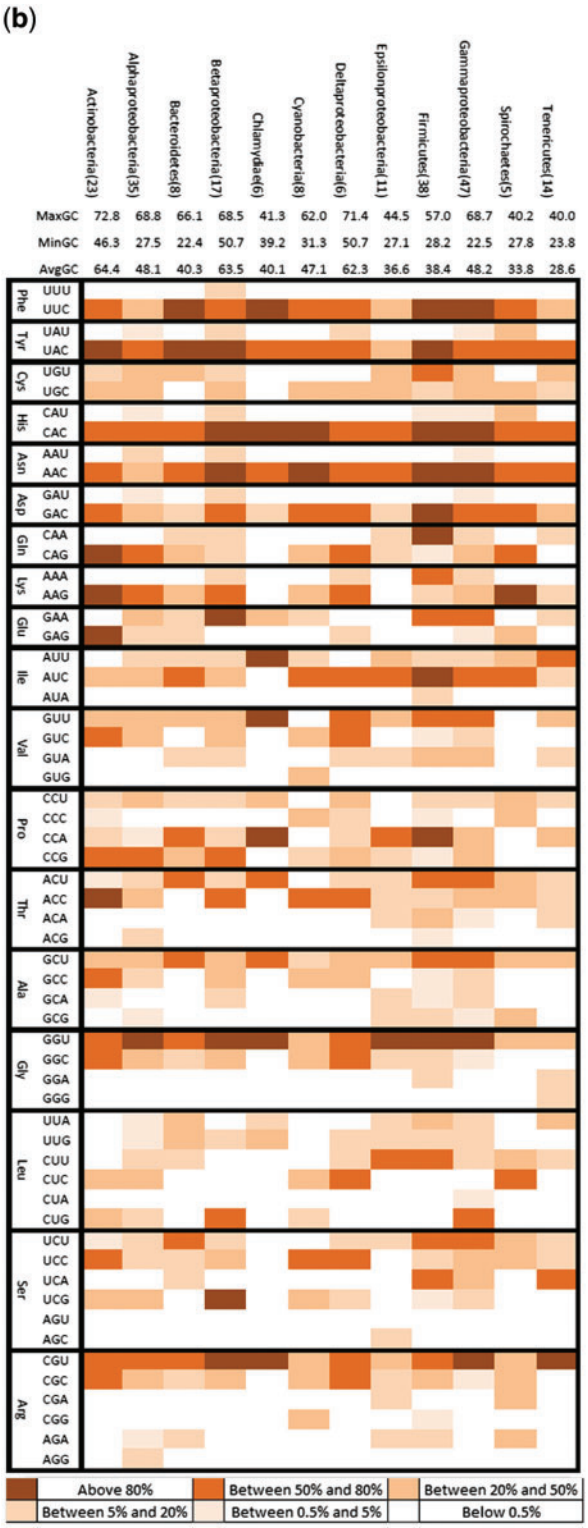

Figure 1b. Optimal codons of 18 amino acids in bacteria within different phylogenetic groups. As described earlier for Fig.1a, the OCs found out in 221 bacteria and were considered in 15 groups (Table 1b) according to the phylogeny. Out of the 15 groups, three groups having only one bacterium each were not considered for further analysis. The colour code is used to represent % of bacteria in the phylogenetic group where the codon was observed more frequent in the HEGs than the whole genome. For the raw data, please refer to Supplementary Supplementary Table 3.

Table 2. A general pattern indicating amino acid specific determination of OCs in different bacterial groups

Degeneracy	Amino acids	Selected OCs in the high expression genes*
Two (NNY)	Phe, Tyr, Asn,His and Asp	C-ending codons in all genomes
	Cys	C-ending codon in G + C high genomes (e.g. Actinobacteria, β-Proteobacteria) and U-ending codon in G + C low genomes (e.g. Firmicutes)
Two (NNR)	Gln and Lys	The G-ending codons in the G + C high (e.g. Actinobacteria, β-Proteobacteria) and the A-ending codons in the G + C low genomes (Firmicutes). Exception is Spirochaetes (G + C low) where the G-ending codon
	Glu	The G-ending codons in G + C high/low genomes (e.g. Actinobacteria and Spirochaetes) and the A-ending in G + C high/low genomes (e.g. β-Proteobacteria and Firmicutes)
Three	Ile	C-ending codons in G + C high and U/C-ending in G + C low genomes
Four	Val, Thr and Ala	Y-ending codons (C-ending codons in G + C high and U-ending in G + C low genomes)
	Pro	R-ending codons (G-ending codons in G + C high and A-ending in G + C low genomes)
	Gly	Y-ending codons (U/C-ending codons in G + C high and U-ending in G + C low genomes)
Six	Leu and Ser	Family box codons (G/C-ending codons in G + C high and A/U-ending in G + C low genomes)
	Arg	Y-ending codon (U/C-ending codons in G + C high and U-ending codon in G + C low genomes)

*All are statistically significant (*chi*-square significance test P value < 0.01).

OC is selected according to the phylogenetic affiliation of the bacteria but not by genome G+C%. Earlier, it was demonstrated that GAA was a favoured codon over GAG in the HEGs in *Escherichia coli* and the speed of translation was higher in case of GAA than that of GAG.[38,39] Unlike the amino acids encoded by the Y-ending codons described before, for the amino acids encoded by the NNR codons, bacteria uses either U_{34} or both U_{34} and C_{34} anticodons: the U_{34} can decode both the A_3 and the G_3 codons, while the C_{34}, decodes only the G_3 codons. While the occurrence of the U_{34} is there in all bacteria, the C_{34} is generally found in bacteria with high G+C genomes.[40] The availability of two anticodons increases the effectiveness/probability of codon–anticodon pairing for an amino acid. Though the C_{34} anticodon usually occurs in the high G+C% genomes, in case of β-Proteobacteria, in spite of being high G+C genomes, C_{34}-Glu anticodon was not observed.[40] This might be due to the fact that GAA is selected as the OC in this group of bacteria. In the case of Actinobacteria, which is also high G+C genome, where the GAG codon was selected as the OC, the C_{34}-Glu was present in all the bacteria in this group.[40] However, in case of Spirochaetes, C_{34}-Glu was not found in all the bacteria, though GAG was selected as the OC. This indicates that the occurrence of C_{34} is not essential for the GAG selection as the OC, rather some phylogeny specific feature may contribute for its selection.

The U_{34} modification is same in case of tRNA of Gln (tQ), Glu (tE) and Lys(tK).[10] However, the choice of OC in case Glu was not found to be the same as for Lys and Gln in all the bacteria under consideration. A recent study with anticodon modifying enzyme in eukaryotes have revealed that the hypomodification of U_{34} affects translation of CAA and AAA codons of Gln and Lys, but not GAA,[12] the codon of Glu. This indicates that choice of OC in case of these three amino acids may not be always dependent upon the anticodon modification. In the Tenericutes, Chlamydiae and ϵ-Proteobacteria with low G+C genomes, no OC was evident as far as these three amino acid codons are concerned, though the selection of OC for the other amino acids was observed in these bacteria. The reason for the lack of specific choice of any synonymous codons as an OC in these groups needs further investigation. In case of Lys and Gln, the G-ending codon selection in the case of low G+C Spirochaetes was found as an exception also.

Krüger et al. (1998)[41] reported that the different modifications of the U_{34} nucleotide in the anticodon of tE had different effects on the decoding rates of the GAA and GAG codons. Anticodon modifying enzymes has an impact on the selection of OC in bacteria.[4] It has been demonstrated now that this modification is important for the rate of decoding of the cognate codons in tune with the kinetics of translation and co-translational protein folding.[12] Therefore, a detailed study of the effect of U_{34} modification enzymes on the anticodon of tRNA in decoding the codons of split boxes will give insight into the role of these enzymes and the choice of the OC in bacteria belonging to different phylogenetic groups.

3.4. Optimal codon for Ile

The C-ending codon was the OC in most of the bacteria while the U-ending codon was the OC in some low G+C genomes such as Tenericutes and Chlamydiae. Observations with respect to the HTN and the LTN were similar (Table 2). As has been discussed above, the selection of AUU as the OC in some low G+C bacteria in case of Ile is likely to have similarity with Cys due to the absence of the U_{34} anticodon for the split codon box. Except in some bacteria belonging to Firmicutes, AUA was generally not observed as the OC in any other group of bacteria. AUA is a rare codon in bacteria, which is decoded by the anticodon CAU. The C_{34} is modified to $k2C_{34}$, to make correct pairing with AUA and to avoid incorrect pairing with the AUG codon.[42] There are also instances where AUA is decoded by U_{34} in some specific groups of bacteria.[43] The intrinsic feature such as the UA dinucleotide, which is generally avoided in the HEGs in bacteria,[44] might be a reason for AUA not being the OC in bacteria.

3.5. Optimal codon for the amino acids encoded only by the 4-fold degenerate family box codons

Based upon the choice of the OCs in bacteria, in general, the five amino acids, *viz.*, Ala, Gly, Pro, Thr and Val, can be divided into three groups—first, in case of Ala, Thr and Val, the C-ending codons were found as the OC in the high G+C% genomes while the U-ending codons were found as the OC in the low G+C% genomes; second, in the case of Gly, both the U-ending and the C-ending codons

Table 3. Comparison of the abundance values of the synonymous codons in bacteria with same G + C composition within family boxes

AA		VH	H	M
Val	HTN	$G_3 > C_3$*	$G_3 > C_3$**	$G_3 > C_3$*
	LTN	$G_3 \sim C_3$	$G_3 > C_3$*	$G_3 > C_3$**
Pro	HTN	$G_3 > C_3$*	$G_3 > C_3$*	$G_3 > C_3$*
	LTN	$G_3 \sim C_3$	$G_3 \sim C_3$	$G_3 \sim C_3$
Thr	HTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 > G_3$*
	LTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 > G_3$*
Ala	HTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 \sim G_3$
	LTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 > G_3$*
Gly	HTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 > G_3$*
	LTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 > G_3$*
Leu	HTN	$G_3 > C_3$*	$G_3 > C_3$*	$G_3 > C_3$*
	LTN	$G_3 > C_3$*	$G_3 > C_3$*	$G_3 > C_3$*
Ser	HTN	$G_3 > C_3$*	$G_3 > C_3$*	$G_3 \sim C_3$
	LTN	$G_3 > C_3$*	$G_3 > C_3$*	$G_3 \sim C_3$
Arg	HTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 > G_3$*
	LTN	$C_3 > G_3$*	$C_3 > G_3$*	$C_3 > G_3$*

*P value <0.05.

**P value <0.01 (*chi*-square significant test; H_0: G and C equal preference; H_A: unequal preference between G and C).

were found as the OC in the high G + C% genomes, while only the U-ending codon was found as the OC in the low G + C% genomes; third, in case of Pro, the G-ending codon was found as the OC in the G + C% high genomes while the A-ending codon was found as the OC in the G + C% low genomes (Table 2). Therefore, in these five amino acids either the Y-ending codons or the R-ending codons were observed as the OCs in bacteria. The HTN and the LTN groups were similar for all the amino acids with respect to the OC.

The observation of more than one OCs for these amino acids in a genome is possible due to higher degeneracy. We analysed the OCs for these five amino acids in different bacteria according to their phylogenetic affiliations (Table 2). Bacteria use different isoacceptor tRNAs to decode the four different synonymous codons in a family box, which is described under various sparing strategies.[10] Choice of OCs in family box codons is influenced by anticodon modification and tRNA gene copy number in bacteria.[4] The modification of U_{34} in tRNAs across different family box codons is similar and this modification is different from that of the U_{34} in tRNAs of split codon boxes.[10] In case of Gly, the Y-ending codons were selected as the OC across different groups of bacteria. It has been reported earlier that decoding of the GGG codon is prone to −1 frameshift during translation.[45] This might be a reason for the less preference of this codon in the HEGs. In case of Pro, the CCC codon was not selected as the OC in bacteria which might be due to the fact that this codon is prone to +1 frameshift during translation.[46] Apart from this, translation speed is hindered by the Pro codons both at the 'P' and the 'A' sites of a ribosome due to the difference in the chemical structure of this amino acid from the other amino acids.[47] In a case similar to Gly, the G-ending codons were not selected as the OC in case of Val, Ala and Thr, across different groups of bacteria. We believe that the intrinsic nature of these codons of different amino acids is likely to contribute towards their choice of OCs in bacteria.

The choice for the OCs is likely to be influenced by the tRNA gene copy number in bacteria. The gene copy number of G_{34} anticodon, which is used to decode the Y-ending codons, is higher for Gly and is lower for Pro than the G_{34} anticodons for the other amino acids.[40] This is in concordance with the choice of GGU/GGC as the OC(s) and avoidance of CCC as an OC in bacteria. Transfer RNA with either U_{34} or G_{34} at the wobble position can pair with either the R-ending or the Y-ending codons, respectively. Therefore, if the GGG triplet is not favoured as the OC, there will be no selection for the higher copy number of the tRNA with U_{34} anticodon. Low U_{34} tRNA will automatically not prefer the GGA triplet as the OC. Therefore, in general, either the Y-ending or the R-ending codons are being selected as the OC for these amino acids in bacteria. As the G-ending codon in Val and Ala are not preferred as the OC, the C_{34} is also a less preferred anticodon for these amino acids.[40]

3.6. Optimal codon for the amino acids encoded by 6-fold degenerate codons

Depending upon the choice of the OCs across bacteria, the three amino acids Arg, Leu and Ser can be divided into two groups—first, both U-ending and C-ending codons were found as the OC in the high G + C% genomes, while only the U-ending codon was found as the OC in the low G + C% genomes, in the case of Arg; second, in the case of Ser and Leu, the C-ending and the G-ending codons were found as the OC in the high G + C% genomes, while the U-ending and the A-ending codons were found as the OC in the low G + C% genomes (Table 2). In the three amino acids, the OCs were observed mostly from the family box codons and were less preferred from the split codon boxes. The observation with respect to the HTN and the LTN were similar for all the amino acids.

We studied the OCs of these three amino acids in bacteria that were divided according to their phylogenetic groupings (Table 2). In case of Arg, the CGU and the CGC codons were largely selected as the OC across different bacteria. The other four synonymous codons were generally not observed as the OC; however, in some bacterial groups such as Spirochaetes, Firmicutes and ϵ-Proteobacteria with low genome G + C%, AGA was also observed as the OC to some extent. It is pertinent to note that anticodons and its modifying enzymes used in bacteria to decode the Arg family box codons are different from the family box codons of the other amino acids.[10] I_{34} in tR is used to decode the U/C/A-ending codons whereas C_{34} is used to decode the G-ending codon. It is important to note that though there is similarity between Arg and Gly with respect to the choice of the OC (Fig. 1b), there is a difference regarding modification in anticodons of tG and tR.

In case of Leu, the choice of the OC in bacteria was found to be influenced by their phylogenetic affiliations. In case of γ-Proteobacteria and β-Proteobacteria which are more closely related phylogenetically in comparison to the Proteobacteria,[48] CUG was selected as the OC, which is different from δ-Proteobacteria and ϵ-Proteobacteria. In case of Firmicutes, UUA, UUG and CUU were selected as the OCs. CUA was avoided as the OC in all the bacterial group. In case of γ-Proteobacteria and β-Proteobacteria, the tRNA gene copy number with C_{34} anticodon, which decodes the CUG codon, was higher than that in the other group of bacteria.[40]

In case of Ser, the choice of OCs was found only from the four family box codons. OCs were found to be selected according to genome composition. For example, UCA and UCU were selected as OC in Firmicutes, Tenericutes (low G + C% genomes) whereas UCC and UCG were selected as OC in Actinobacteria, β-Proteobacteria (high G + C% genomes). In case of ϵ-Proteobacteria and Chlamydiae, low selection of OC was observed.

Study of codon usage bias in different amino acids exhibited that there is amino acid specific influence for the selection of OCs in

bacteria. For example, out of the six codons of Ser, the AGY codons were never selected as the OCs in any group of bacteria. Though the OCs were selected within the family box codons, the principle for the Leu and Ser that are encoded by six synonymous codons is different from the amino acids encoded by the four synonymous codons. This relationship between degeneracy of the codons and the principle of selection for the OCs in amino acids is surprising and interesting considering the fact that the set of anticodons used to decode the different family box codons as well as the anticodon modifying enzymes for family box tRNA is similar in a bacterium. Therefore, it is less likely that the above observation was a consequence of the anticodons.

3.7. The abundance of most of the OCs is higher than the synonymous codons with same G + C composition in a genome

It is known that genome G + C% strongly influences the abundance value of a synonymous codon in genomes.[49–51] Two synonymous codons with the same G + C composition were available in case of amino acids encoded by 4- or 6-fold degeneracy codons. In the above section, we observed that out of the two synonymous codons with same G + C%, one of the codons was selected as the OC in many bacteria. This prompted us to compare the abundance values between two synonymous codons having the same G + C% and their selection as the OCs in genomes. Their abundance comparison in amino acids encoded by 4-fold degeneracy and 6-fold degeneracy codons was presented in Table 3. The results of the comparison between the C-ending and the G-ending codons in VH, H and M groups in the eight family boxes were presented.

The C-ending codons were more abundant than the G-ending codons in case of Ala, Gly and Thr. In fact, the C-ending codons were the OCs in these amino acids. In case of Pro, the G-ending codon was more abundant than the C-ending codon, which was in concordance with the selection as the OC. However, in case of Val, the G-ending codon was more abundant than the C-ending codon, though the latter was selected as the OC. In the HTN and the LTN, the abundance patterns of the different amino acid codons were similar.

In case of Arg, the abundance value of CGC was more than that of CGG. In fact, CGC was selected as the OC. In case of Leu, CUG abundance was higher than CUC. CUG was selected as the OC more often than CUC. In case of Ser, abundance values of the synonymous codons with the same G + C% were similar.

It is often observed that the abundance values of the OCs are higher than the abundance values of the synonymous codons with the same G + C%. In some cases, variations were observed. The differential abundance values between two synonymous codons with the same G + C composition for an amino acid were following a uniform pattern across different bacteria. Why in some amino acids such as Val and Ser, abundance values of OCs is not higher than the other synonymous codons with same G + C composition is an interesting question.

4. Discussion

The earlier notion was that the extrinsic factors such as genome G + C%, tRNA gene numbers, etc., cause the difference in abundance values among the synonymous codons. The interesting point emerged in this study is the amino acid specific choice of OC in bacteria depending upon their phylogenetic affiliations, which is independent of their genome composition and tRNA gene number. For example, the observation of GAA as OC in β-Proteobacteria and GAG as OC in the Spirochaetes in this study are the vindication of the influence of some bacterial phylogeny specific mechanisms influencing OC choice for Glu in these bacteria. Similarly, the observation of CUG as OCs in β-Proteobacteria and γ-Proteobacteria, that are phylogenetically closed than any other Proteobacteria also indicates a role of phylogeny specific mechanisms for OC choice in case of Leu in bacteria. The OCs for the three amino acids Gln, Glu and Lys, encoded only by the R-ending codons in split codon boxes were observed to be influenced maximally by the phylogeny of bacteria. The U_{34} of these tRNA undergoes different modifications to avoid frameshift mutation,[52] correct decoding,[53,54] to avoid miscoding of the non-synonymous codons of the amino acids of the split codon boxes,[54,55] effective speed of decoding,[12] during the translation. Our future research is aimed at doing a comparative study of the anticodon modifying enzymes across different phylogenetic groups of bacteria.

The observation of the OCs limiting to only some synonymous codons for different amino acids across bacteria is interesting in this study. Why is this discrepancy among the codons? There are some known mechanisms such as frameshift mutation in case of GGG and CCC codon for which these codons are avoided as the OC in bacteria. But the reason is not known for some other non-OCs observed in this study such as CUA, CGR, G-ending codons of Ala, Val and Thr, and AGY. The findings in this study clearly indicate that the potential of all synonymous codons with regard to their selection as OCs in bacteria is not same even though they are encoding the same amino acid.

There may be an influence of the preference and avoidance of dinucleotides during OC selection. The UA dinucleotide is not preferred in genomes,[44,56] which might influence UUA, CUA, AUA and GUA for not being selected as the OC in any bacteria. Similarly for selection of CGU, GGU and UGU as the OCs, it may be argued that GU is the selected dinucleotide at the second codon position. Interestingly, the reverse complement of GU, AC are also observed to be preferred at the second position, as UAC, CAC are selected as the OCs in bacteria. Our finding on OCs is in concordance with the dinucleotide constraint on codon usage bias already reported in bacteria.[44]

According to our present understanding, codon usage bias in organisms is a cumulative effect of the extrinsic and intrinsic factors. In the selection of the OCs in bacteria whether it is the extrinsic or the intrinsic factor that plays dominant role is not clear. We believe that the amino acid specific role for the selection of the OC in bacteria is decided first by the codon–anticodon interaction. The selection of the C-ending codons in case of amino acids encoded by the NNY codons is a suitable example. In these amino acids, only one anticodon is available to decode both the codons. So out of the two codon–anticodon interactions, the more stable one was selected in all the bacteria irrespective of genome composition. It is also not dependent upon the tRNA gene numbers.[28] In case of the amino acids encoded by higher degeneracy codons, more than one codon can function as the OCs due to multiple codon–anticodon interactions. In this context, the increase in the copy number of the cognate tRNA complements the efficient translation of the OCs, which is required for bacterial adaptation to growth rate.[15,16] It is not the tRNA gene copy number *per se* which may be essential for selection of the OC for an amino acid.[28,57] Recently, it has been demonstrated that tRNA genes rapidly change in evolution to meet novel translational demands.[58] It has also been observed that the copy number of only specific isoacceptor tRNA genes increased across bacteria.[40] The role of anticodon modification enzymes and its effect on the selection of codons in

different major taxa of life[11,59] also supports our assumption that codon–anticodon interactions are important determinant factors for the selection of the OCs in bacteria. More research on codon–anticodon pairing during translation is likely to provide empirical evidence in favour of the role of codon–anticodon interaction on the selection of OCs. It is usually believed that mutation pressure maintains the abundance values of two synonymous codons with same G + C% similar in a genome. The consistently higher abundance value of one synonymous codon over the other synonymous codons with same G + C composition across bacteria further supports the contribution from the inherent nucleotide sequence of the triplet codon to its abundance in genomes.

What is the evolutionary advantage or the selective force for having the amino acid specific principle for OC choice across bacteria? It is likely that the principle has enabled microorganisms to receive and express genes from the other microorganisms by lateral gene transfer. For example, both the U- and the C-ending codons of Gly are efficiently translated by the same anticodon.[10] The translation rate of the gene inside the receptive host is less likely to be affected by the difference in codon usage bias. Now the well-known phenomenon of co-translational protein folding[60] and the kinetics of translation is important for the structure and function of a protein.[26] Thus, lateral gene transfer may also get inhibited due to the differential translational kinetics of a gene across different bacteria.[61,62] A recent study has revealed that OC choice between the two archaea groups, Crenarchaea and Euryarchaea, is different for many amino acids,[63] unlike what we observed for bacteria in this study. As bacteria are evolutionarily different from archaea, the selection force on codon usage bias between the two groups of archaea may not be the same. Similarly in human, which is phylogenetically very different from bacteria and archaea, there is no difference between the high and the low expression genes with respect to codon usage bias.[64]

A recent publication by Babbitt et al. (2014)[65] discusses intrinsic DNA flexibility and codon usage bias across different bacteria, which is also in support of the analysis discussed in this paper. In conclusion, our study indicates a vital role of the nucleotide sequence of the triplet in selection as OCs in bacteria that was proposed three decades ago by Thomas et al. (1988).[66]

Acknowledgements

We are very much thankful to the two anonymous reviewers for their comments which helped us to improve the manuscript in different ways and to analyse the data differents. We also thank Ms Ruksana Aziz, Ms Ishani Goswami, Department of Molecular Biology and Biotechnology, Tezpur University for comments on the manuscript. S.K.R. and S.S.S. are thankful to DBT, Govt. of India for the twinning project under Bioinformatics.

Conflict of Interest

Authors do not have any conflict of interest.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

References

1. Sharp, P.M. and Li, W.H. 1987, The codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res*., **15**, 1281–95.
2. Sharp, P.M., Emery, L.R. and Zeng, K. 2010, Forces that influence the evolution of codon bias, *Philos. Trans. R. Soc. Lond. B. Biol. Sci*., **365**, 1203–12.
3. Bulmer, M. 1991, The selection-mutation-drift theory of synonymous codon usage, *Genetics*, **129**, 897–907.
4. Ran, W. and Higgs, P.G. 2010, The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria, *Mol. Biol. Evol*., **27**, 2129–40.
5. Quax, T.E., Claassens, N.J., Söll, D. and van der Oost, J. 2015, Codon bias as a means to fine-tune gene expression, *Mol. Cell*, **59**, 149–61.
6. Yu, C.H., Dang, Y., Zhou, Z., et al. 2015, Codon usage influences the local rate of translation elongation to regulate co-translational protein folding, *Mol. Cell*, **59**, 1–11.
7. Ray, S.K. and Goswami, I. 2016, Synonymous codons are not same with respect to the speed of translation elongation, *Curr. Sci.*, **110**, 1612–4.
8. Hershberg, R. and Petrov, D.A. 2009, General rules for optimal codon choice, *PLoS Genet*., **5**, e1000556.
9. Grosjean, H. and Fiers, W. 1982, Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes, *Gene*, **18**, 199–209.
10. Grosjean, H., de Crécy-Lagard, V. and Marck, C. 2010, Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes, *FEBS Lett*., **584**, 252–64.
11. Novoa, E.M., Pavon-Eternod, M., Pan, T. and Ribas de Pouplana, L. 2012, A role for tRNA modifications in genome structure and codon usage, *Cell*, **149**, 202–13.
12. Nedialkova, D.D. and Leidel, S.A. 2015, Optimization of codon translation rates via tRNA modifications maintains proteome integrity, *Cell*, **161**, 1606–18.
13. Ikemura, T. 1981, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *J. Mol. Biol*., **151**, 389–409.
14. Ikemura, T. 1985, Codon usage and tRNA content in unicellular and multicellular organisms, *Mol. Biol. Evol*., **2**, 13–34.
15. Rocha, E.P. 2004, Codon usage bias from tRNA's point of view, redundancy, specialization, and efficient decoding for translation optimization, *Genome Res*., **14**, 2279–86.
16. Higgs, P.G. and Ran, W. 2008, Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage, *Mol. Biol. Evol*., **25**, 2279–91.
17. Shah, P. and Gilchrist, M.A. 2011, Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift, *Proc. Natl. Acad. Sci. U. S. A*., **108**, 10231–6.
18. Katz, L. and Burge, C.B. 2003, Wide spread selection for local RNA secondary structure in coding regions of bacterial genes, *Genome Res*., **13**, 2042–51.
19. Gu, W., Zhou, T. and Wilke, C.O. 2010, A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes, *PLoS Comput. Biol*., **6**, e1000664.
20. Zur, H. and Tuller, T. 2012, Strong association between mRNA folding strength and protein abundance in *S. cerevisiae*, *EMBO Rep*., **13**, 272–7.
21. Park, C., Chen, X., Yang, J.R. and Zhang, J. 2013, Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly, *Proc. Natl. Acad. Sci. U. S. A*., **110**, E678–86.
22. Presnyak, V., Alhusaini, N., Chen, Y.H., et al. 2015, Codon optimality is a major determinant of mRNA stability, *Cell*, **160**, 1111–24.
23. Zhou, M., Wang, T., Fu, J., Xiao, G. and Liu, Y. 2015, Nonoptimal codon usage influences protein structure in intrinsically disordered regions, *Mol. Microbiol*., **97**, 974–87.
24. Ray, S.K., Baruah, V.J., Satapathy, S.S. and Banerjee, R. 2014, Co-translational protein folding is revealing the selective use of synonymous codons along the coding sequence of a low expression gene, *J. Genet*., **93**, 613–7.
25. Nissley, D.A., Sharma, A.K., Ahmed, N., et al. 2015, Accurate prediction of cellular co-translational folding indicates proteins can switch from post- to co-translational folding, *Nat. Commun*., **7**, 10341.

26. Buhr, F., Jha, S., Thommen, M., et al. 2016, Synonymous codons direct cotranslational folding toward different protein conformations, *Mol. Cell*, **61**, 341–51.
27. Sharp, P.M., Bailes, E., Grocock, R.J., Peden, J.F. and Sockett, R.E. 2005, Variation in the strength of selected codon usage bias among bacteria, *Nucleic Acids Res*., **33**, 1141–53.
28. Satapathy, S.S., Dutta, M., Buragohain, A.K. and Ray, S.K. 2012, Transfer RNA gene numbers may not be completely responsible for the codon usage bias in asparagine, isoleucine, phenylalanine and tyrosine in the high expression genes in bacteria, *J. Mol. Evol*., **75**, 34–42.
29. Wald, N., Alroy, M., Botzman, M. and Margalit, H. 2012, Codon usage bias in prokaryotic pyrimidine-ending codons is associated with the degeneracy of the encoded amino acids, *Nucleic Acids Res*., **40**, 7074–83.
30. Satapathy, S.S., Powdel, B.R., Dutta, M., Buragohain, A.K. and Ray, S.K. 2014, Selection on GGU and CGU codons in the high expression genes in bacteria, *J. Mol. Evol*., **78**, 13–23.
31. Wang, B., Shao, Z.-Q., Xu, Y., et al. 2011, Optimal codon identities in bacteria: implications from the conflicting results of two different methods, *PLoS One*, **6**, e22714.
32. Lowe, T.M. and Eddy, S.R. 1997, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res*., **25**, 955–64.
33. Sharp, P.M. and Li, W.-H. 1986, An evolutionary perspective on synonymous codon usage in unicellular organisms, *J. Mol. Evol*., **24**, 28–38.
34. Iriarte, A., Baraibar, J.D., Romero, H., Castro-Sowinski, S. and Musto, H. 2013, Evolution of optimal codon choices in the family Enterobacteriaceae, *Microbiology*, **159**, 555–64.
35. Akashi, H. 2002, Translational selection and yeast proteome evolution, *Genetics*, **164**, 1291–303.
36. Shah, P. and Gilchrist, M.A. 2010, Effect of correlated tRAN abundances on translation errors and evolution of codon usage bias, *PLoS Genet*., **6**, e1001128.
37. Gilchrist, M.A., Shah, P. and Zaretzki, R. 2009, Measuring and detecting molecular adaptation in codon usage against non-sense errors during protein translation, *Genetics*, **183**, 1493–505.
38. Sørensen, M.A., Kurland, C.G. and Pedersen, S. 1989, Codon usage determines the translation rate in *Escherichia coli*, *J. Mol. Biol*., **207**, 365–77.
39. Sørensen, M.A. and Pedersen, S. 1991, Absolute *in vivo* translation rates of individual codons in *Escherichia coli* the two glutamic acid codons, GAA and GAG are translated with threefold differences in rate, *J. Mol. Biol*., **222**, 265–80.
40. Prajapati, V.K., Satapathy, S.S., Satish Kumar, M.V., Buragohain, A.K. and Ray, S.K. 2015, Evidences indicating the involvement of selection mechanisms for the occurrence of C34 anticodon in bacteria, *J. Cell Sci. Mol. Biol*., **2**, 112.
41. Krüger, M.K., Pedersen, S., Hagervall, T.G. and Sørensen, M.A. 1998, The modification of the wobble base of $tRNA^{Glu}$ modulates the translation rate of glutamic acid codon *in vivo*, *J. Mol. Biol*., **284**, 621–31.
42. Gustilo, E.M., Vendeix, F.A.P. and Agris, P.F. 2008, tRNA's modifications bring order to gene expression, *Curr. Opin. Microbiol*., **11**, 134–40.
43. Suzuki, T. and Numata, T. 2014, Convergent evolution of AUA decoding in bacteria and archaea, *RNA Biol*., **11**, 1586–96.
44. Satapathy, S.S., Powdel, B.R., Dutta, M., Buragohain, A.K. and Ray, S.K. 2014, Constraint on dinucleotides by codon usage bias in bacterial genomes, *Gene*, **536**, 18–28.
45. O'Çonnor, M. 1998, tRNA imbalance promotes -1 frameshifting via near-cognate decoding, *J. Mol. Biol*., **279**, 727–36.
46. O'Çonnor, M. 2002, Imbalance of $tRNA^{Pro}$ isoacceptors induces +1 frameshifting at near-cognate codons, *Nucleic Acids Res*., **30**, 759–65.
47. Tarrant, D. and von der Haar, T. 2014, Synonymous codons, ribosome speed, and eukaryotic gene expression regulation, *Cell Mol. Life Sci*., **71**, 4195–206.
48. Rocha, E.P. 2004, The replication-related organization of bacterial genomes, *Microbiology*, **150**, 1609–27.
49. Muto, A. and Osawa, S. 1987, The guanine and cytosine content of genomic DNA and bacterial evolution, *Proc. Natl Acad. Sci. U. S. A*., **84**, 166–9.
50. Chen, S.L., Lee, W., Hotts, A.K., Shapiro, L. and McAdams, H.H. 2004, Codon usage between genomes is constrained by genome wide mutational processes, *Proc. Natl. Acad. Sci. U. S. A*., **101**, 3480–5.
51. Palidwor, G.A., Perkins, T.J. and Xia, X. 2010, A general model of codon bias due to GC mutational bias, *PLoS One*, **5**, e13431.
52. Brégeon, D., Colot, V., Radman, M. and Taddei, F. 2001. Translational misreading: a tRNA modification counteracts a +2 ribosomal frameshift, *Genes Dev*., **15**, 2295–306.
53. Agris, P.F. 2008, Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications, *EMBO Rep*., **9**, 629–35.
54. Armengod, M.E., Meseguer, S., Villarroya, M., et al. 2014, Modification of the wobble uridine in bacterial and mitochondrial tRNAs reading NNA/NNG triplets of 2-codon boxes, *RNA Biol*., **11**, 1495–507.
55. Hou, Y.-M., Gamper, H. and Yang, W. 2015, Post-transcriptional modifications to tRNA – a response to the genetic code degeneracy, *RNA*, **21**, 642–44.
56. Karlin, S., Campbell, A.M. and Mrázek, J. 1998, Comparative DNA analysis across diverse genomes, *Annu. Rev. Genet*., **32**, 185–225.
57. Supek, F., Škunca, N., Repar, J., Vlahoviček, K. and Šmuc, T. 2010, Translational selection is ubiquitous in prokaryotes, *PLoS Genet*., **6**, e1001004.
58. Yona, A.H., Bloom-Ackermann, Z., Frumkin, I., et al. 2013, tRNA genes rapidly change in evolution to meet novel translational demands, *eLife*, **2**, e01339.
59. Endres, L., Dedon, P.C. and Begley, T.J. 2015, Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses, *RNA Biol*., **12**, 603–14.
60. Holtkamp, W., Kokic, G., Jäger, M., Mittelstät, J., Komar, A. and Rodnina, M.V. 2015, Cotranslational protein folding on the ribosome monitored in real time, *Science*, **350**, 1104–7.
61. Angov, E., Hiller, C.J., Kincaid, R.L. and Lyon, J.A. 2008, Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression hosts, *PLoS One*, **3**, e2189.
62. Angov, E. 2011, Codon usage: nature's roadmap to expression and folding of proteins, *Biotechnol. J*., **6**, 650–9.
63. Baruah, V.J., Satapathy, S.S., Powdel, B.R., Konwarh, R., Buragohain, A.K. and Ray, S.K. 2015, Comparative analysis of codon usage bias in Crenarchaea and Euryarchaea genome reveals differential preference of synonymous codons to encode highly expressed ribosomal and RNA polymerase proteins, *J. Genet*. (In press).
64. Satapathy, S.S., Ray, S.K., Sahoo, A.K., Begum, T. and Ghosh, T.C. 2015, Codon usage bias is not significantly different between the high and the low expression genes in human, *Int. J. Mol. Genet. Gene Ther*., **1**, 1.
65. Babbitt, G.A., Alawad, M.A., Schulze, K.V. and Hudson, A.O. 2014, Synonymous codon bias and functional constraint on GC3-related DNA backbone dynamics in the prokaryotic nucleoid, *Nucleic Acids Res*., **42**, 10915–26.
66. Thomas, L.K., Dix, D.B. and Thompson, R.C. 1988, Codon choice and gene expression: synonymous codons differ in their ability to direct aminoacylated-transfer RNA binding to ribosomes *in vitro*, *Proc. Natl. Acad. Sci. U. S. A*., **85**, 4242–6.