Occurrence of all nucleotide combinations at the third and the first positions of two adjacent codons in open reading frames of bacteria

The occurrence of an amino acid in a polypeptide chain is independent of the ones preceding and succeeding it in the chain. During the translation process, two tRNAs pair with two adjacent codons on the mRNA occupy the P and A sites within the ribosome. Catalysis of the peptide bond is stimulated by the correct pairing of codon and anticodon at the 'A' site, following accommodation of the aminoacyl-tRNA[1]. The location of the anticodon is in the anticodon stem loop (ASL) region of tRNA that consists of a RNA duplex whose width is ~ 20 Å. Each codon–anticodon pairing forms a helix of three base pairs with a length of ~ 10 Å (3 × 3.4 Å). Hence two adjacent codon–anticodon pairs occupy a length of ~ 20 Å (2 × 10 Å). This is paradoxical because two ASLs whose combined dimension is 40 Å, are to be accommodated within a space of about ~ 20 Å (Figure 1 *a*). To resolve the paradox, it was suggested that mRNA bends between the A and P sites, allowing both tRNAs to pair with their cognate anticodons simultaneously[2] (Figure 1 *b*). Recent X-ray crystallography studies have demonstrated the presence of a kink of 45° in the mRNA between the P and A site codons[3,4]. According to our estimation, this bending angle is sufficient to keep the two ASLs separated from each other while interacting with their cognate codons (see Figure 1 legend). However, the relationship of bending in mRNA during translation with the accommodation of two tRNAs is yet to be demonstrated.

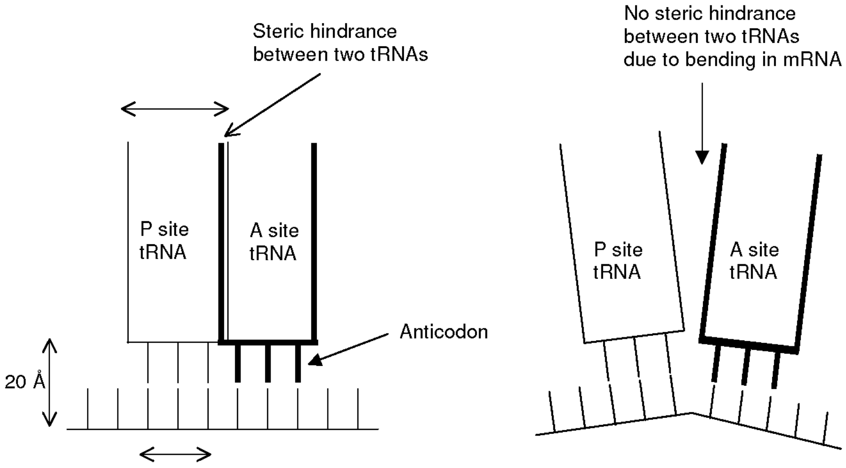

Figure 1. Schematic representation of mRNA bending and steric hindrance between two tRNAs binding to two adjacent codons on the mRNA at the P and A sites. The anticodon of a tRNA is present at the anticodon stem loop region (ASL) which is 20 Å in width. Anticodon–codon pairing forms a three-base pair helix that is 10 Å (3.4 × 3) × 20 Å (RNA duplex) in dimension. Binding of two tRNAs simultaneously to two adjacent codons causes steric hindrance (***a***). The P site tRNA has been shown in thin line and the A site tRNA has been shown in thick line to highlight overlapping between the two tRNAs. A bending between the two codons on mRNA would facilitate simultaneous binding of the two tRNAs to the codons without steric hindrance (***b***). Considering the dimensions of an anticodon–codon pairing (10 Å × 20 Å) and the overlapping region between the two tRNA ASLs around 10 Å (width of a RNA duplex; shown in (***a***)), the bending angle in mRNA can be calculated to be 28° (maximum), which is sufficient for overcoming steric hindrance. (Figures have not been drawn to scale.)

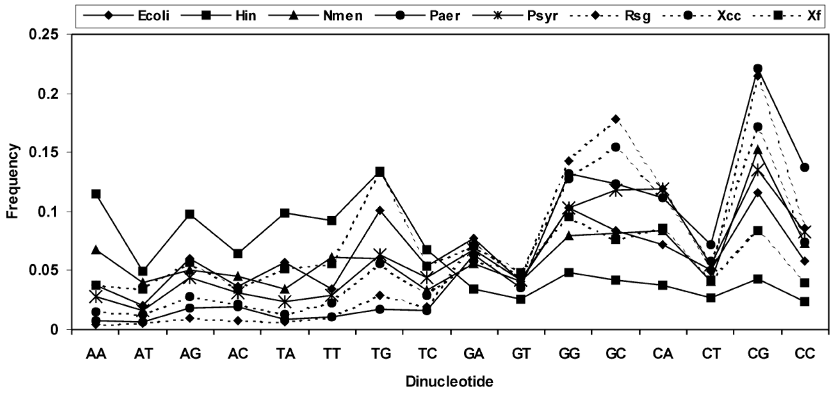

Figure 2. Dinucleotide frequency in the third and first positions respectively, of two adjacent codons in bacterial ORFs. In the *X*-axis all sixteen dinucleotides and in the *Y*-axis the average frequency of dinucleotides are given. Abundance of a dinucleotide occurrence in the third and first positions respectively, of two adjacent codons within an ORF was calculated by using the program 'dinsearch'. Frequency of a dinucleotide within the ORF was calculated by dividing the abundance value of the dinucleotide with the total abundance value of all sixteen dinucleotides. The average frequency of a dinucleotide occurring within the ORFs of a genome was calculated from the frequency values obtained from ten different ORFs randomly chosen from each genome sequence (Table 1). The frequency graph suggests a similarity in patterns in all these organisms. The standard deviation values are not shown in the graph for the sake of clarity.

Studies on codon context in *Escherichia coli* have suggested that there is a preference for the occurrence of certain synonymous codons in an open reading frame (ORF) depending upon the nucleotide succeeding immediately in the ORF. For example, lysine is preferentially encoded by AAA if guanosine (G) is at the adjacent 3′ position, but by AAG if cytidine (C) is at this position instead[5]. Both AAA and AAG codons are decoded by the same tRNA molecule due to the wobble base pairing at the third nucleotide position of codons[6]. In this case it is reasonable to argue that presence of wobble pairing and codon degeneracy both result in independent occurrence of amino acids in a polypeptide. Context-dependent nonsense suppression by tRNA has been reported in *E. coli*, which suggests the immediate 3′ nucleotide is responsible for it[7,8]. A molecular explanation regarding codon restriction is yet to be elucidated. Similarly, the exact reason for wobble pairing is yet to be found. It has been speculated that the third nucleotide wobbling is due to the fact that the anticodon lies in the ASL, prohibiting a perfectly linear alignment with the corresponding mRNA codon[9].

Looking at the wobble pairing at the third nucleotide position of a codon and a probable steric hindrance between the two tRNAs occupying the P and A sites simultaneously within the ribosome, we have addressed in this study, the question whether there is any restriction on the occurrence of a nucleotide at the third nucleotide position of a codon in relation to the first nucleotide of the adjacent succeeding codon within an ORF. We have analysed the presence of all sixteen possible dinucleotides at the third and first positions respectively, of two adjacent codons in an ORF. Our observation suggests the presence of all sixteen dinucleotides at the third and the first positions respectively, of two adjacent codons in several bacterial ORFs (Figure 2). From this it may be construed that all possible wobble pairing can occur at the third nucleotide position of codons in relation to base pairing at the first nucleotide position of the adjacent succeeding codon.

A total of eighty open reading frames (each one is more than 1 kb) from eight bacterial genome sequences (ten from each genome; Table 1) were randomly taken for analysing the occurrence of dinucleotides at the third and first nucleotide positions respectively, of two adjacent codons using a computer program 'dinsearch' (developed by the authors for this study). The program was developed to determine the abundance of a dinucleotide at the first and second positions, second and third positions of codons, and third and first positions respectively of two adjacent codons within an ORF. Of these, dinucleotide abundance involving the third and the first positions of two adjacent codons has been considered for this study. The first nucleotide of these dinucleotides represents the third nucleotide of the codons, so their abundance has a direct correlation with the genome GC% of the organism[10]. The second nucleotide of these dinucleotides represents the first nucleotide of a codon, which has correlation with the amino acid compositions of the polypeptide encoded by the ORF. In organisms having higher GC% in the genome (Table 1), the dinucleotide beginning with G/C occurs more frequently within the ORFs in relation to dinucleotides beginning with A/T (Figure 2). This is because a higher percentage of synonymous codons ending with G/C are used over the synonymous codons ending with A/T within the ORFs in these organisms (*R. solanacearum*, *X. campestris* pv. *campestris*, *P. aeruginosa*, and *P. syringae*; Table 1)[10]. Similarly, in case of *H. influenzae* ORFs, dinucleotides beginning with A/T occur more frequently than those beginning with G/C, as its genome is AT-rich. In *E. coli*, *N. meningitidis* and *X. fastidiosa*, no significant distribution bias was observed beteen codons beginning with either A/T or G/C, as the genome GC% of these organisms is close to 50. However, it is interesting to note that the correlation of a dinucleotide frequency in relation to the adjacent dinucleotides (Figure 2) exhibits similar pattern in all the organisms, e.g. AT is lower than AA and AG, whereas AG is higher than AT and AC; TA and TT values are similar, whereas TG value is higher than TT and TC; GT is lower than GA and GG, GG is higher than GT and GC; the frequency CT is lower than CA and CG, and the frequency of CG is higher than CT and CC. The similar pattern among these different organisms is due to the similarity in the pattern of relative amino acid compositions in their proteomes (unpublished result from our laboratory).

Wobble base pairing at the third nucleotide position of a codon might have evolved to enhance the efficiency of translation. During the synthesis of a polypeptide, amino acids are not directly diffused to the site of synthesis, rather aminoacyl tRNAs

Table 1. Open reading frames of different bacteria considered for analysis

E. coli	*Hin*	*Nmen*	*Paer*	*Psyr*	*Rsc*	*Xcc*	*Xf*
(50.78)[#]	(38.15)	(51.4)	(66.55)	(58.39)	(67.03)	(65.06)	(52.7)
ThrA	HI0015	NMB0049	NT03PA0006	PSPTO0006	RSc0004	XC0003	NT01XF0033
(2463)*	(1050)	(3165)	(2301)	(1308)	(1662)	(2445)	(3693)
DnaK	HI1356	NMB0075	NT03PA0047	PSPTO0037	RSc0005	XC0078	NT01XF0099
(1917)	(2100)	(2274)	(10608)	(4911)	(1446)	(2259)	(2706)
CarB	HI0057	NMB0082	NT03PA0080	PSPTO3446	RSc0014	XC1320	NT01XF0111
(3222)	(1830)	(2115)	(2091)	(1239)	(1446)	(2151)	(1938)
AceE	HI0070	NMB0249	NT03PA3797	PSPTO3482	RSc0062	XC1414	NT01XF1473
(2664)	(1677)	(2262)	(1287)	(2028)	(2406)	(2610)	(2085)
AcnB	HI0078	NMB0405	NT03PA3824	PSPTO3532	RSc0077	XC2004	NT01XF1811
(2598)	(1380)	(1497)	(1779)	(1992)	(2436)	(3153)	(1281)
IleS	HI0087	NMB0435	NT03PA3868	PSPTO3886	RSc0121	XC2089	NT01XF2409
(2817)	(1278)	(1200)	(1572)	(1773)	(1674)	(2526)	(1095)
CarA	HI0066	NMB0728	NT03PA4668	PSPTO3900	RSc0237	XC3905	NT01XF2657
(1149)	(1299)	(2364)	(1902)	(3474)	(3300)	(3009)	(1533)
YaaU	HI0038	NMB1140	NT03PA4785	PSPTO4812	RSc0400	XC3927	NT01ZF2708
(1332)	(1056)	(1284)	(1935)	(2607)	(1989)	(3291)	(6195)
KefC	HI0264	NMB1299	NT03PA4788	PSPTO4842	RSc2835	XC4298	NT01XF2710
(1863)	(2718)	(1365)	(2160)	(1914)	(1605)	(1800)	(3567)
NhaA	HI0334	NMB1820	NT03PA4764	PSPTO4845	RSc2316	XC4379	NT01XF2950
(1167)	(2232)	(1242)	(1101)	(4950)	(1536)	(1947)	(4167)

E. coli, *Escherichia coli*; *Hin*, *Haemophilus influenzae*; *Nmen*, *Neisseria meningitidis*; *Paer*, *Pseudomonas aeruginosa*; *Psyr*, *Pseudomonas syringae*; *Rsc*, *Ralstonia solanacearum* (chromosome); *Xcc*, *Xanthomonas campestris* pv. *campestris*; *Xf*, *Xyllela fastidiosa.*

[#] Numbers in parentheses indicate genome GC%.

*Numbers in parentheses indicate ORF length in bp.

are brought to the site of synthesis by the elongation factor (EFTu). Had there been no wobble pairing for all 61 codons that encode one amino acid each, there would have been a requirement of as many different types of t-RNAs. Wobble pairing has reduced this requirement. Lesser the variety of t-RNAs, faster the cognate tRNA appearing at the A site in the ribosome during translation. This also explains why synonymous codons are not randomly present in the genetic code (synonymous codons have same nucleotides at the first and second positions except few codons for serine, arginine and leucine). It seems, wobble pairing and non-randomness of synonymous codons are the consequences of a coevolutionary process leading to higher translation rate.

The program 'dinsearch' may have use in finding alien DNA sequences in genomes which have been acquired recently by horizontal gene transfer. Alien DNA sequences have been suggested to have different GC%, altered codon usage and encode polypepides having different amino acid compositions in relation to the host genome[11]. Genome GC percentage is often correlated with the occurrence of the nucleotides at the third nucleotide position of synonymous codons. The first and/or the second nucleoide(s) of codons give information about amino acid composition of the polypeptide encoded by an ORF. Both GC% and amino acid composition are signatures of a genome. So the dinucleotide abundance at the third and first codon positions respectively, of adjacent codons in an ORF might be considered as a signature for the organism. Comparison of dinucleotide abundance of an alien sequence present in the genome with the reference genome signature will give better resolution than the one obtained on the basis of either the GC% value or the amino acid composition value of the alien sequence.

1. Ogle, J. M., Carter, A. P. and Ramakrishnan, V., *Trends Biochem. Sci.*, 2003, **28**, 259–266.
2. Rich, A., *Annu. Rev. Biochem.*, 2004, **73**, 1–37.
3. Yusupov, M. M. *et al.*, *Science*, 2001, **292**, 883–896.
4. Yusupova, G. Z., Yusupov, M. M., Cate, J. H. D. and Noller, H. F., *Cell*, 2001, **106**, 233–241.
5. Shpaer, E. G., *J. Mol. Biol.*, 1986, **188**, 555–564.
6. Osawa, S., Jukes, T. H., Watanabe, K. and Muto, A., *Microbiol. Rev.*, 1992, **56**, 229–264.
7. Bossi, L. and Roth, J. R., *Nature*, 1980, **286**, 123–127.
8. Bossi, L., *J. Mol. Biol.*, 1983, **164**, 73–87.
9. Brown, T. A., In *Genomes 2*, John Wiley, New York, 2002, pp. 318–319.
10. Ishikawa, J. and Hotta, K., *FEMS Microbiol. Lett.*, 1999, **174**, 251–253.
11. Karlin, S., *Trends Microbiol.*, 2001, **9**, 335–343.

ACKNOWLEDGEMENTS. We thank Prof. Bolin K. Konwar, Head, Department of Molecular Biology and Biotechnology, Tezpur University for support. Thanks are due to Mr Suvam Sen for helping with the mathematical derivations on mRNA bending. We thank Mr Manabendra Mandal and Mr Sujit Kumar Verma for useful discussions. We also thank Dr Anurag Mishra and Mr Prabhu B. Patil for comments on the manuscript.

Received 3 June 2005; revised accepted 16 October 2005

DEBOJYOTI DAS[1]
SIDDHARATHA SANKAR SATAPATHY[2]
ALAK KUMAR BURAGOHAIN[1]
SUVENDRA KUMAR RAY[1,*]

[1]*Department of Molecular Biology and Biotechnology,*
Tezpur University
[2]*Department of Computer Science and Information Technology,*
Tezpur University,
Tezpur 784 028, India
**For correspondence.*
e-mail: suven@tezu.ernet.in

Mapping QTLs underlying seedling vigour traits in rice (*Oryza sativa* L.)

Cultivars having high seedling vigour are desirable for crop establishment in the direct-seeded rice system and in temperate rice-growing areas[1]. High seedling vigour helps the genotypes to suppress the weeds, which is a serious problem in large rainfed and upland areas in the tropics where dry seeding is practised.

The purpose of this study was to tag quantitative trait loci (QTLs) underlying seedling vigour-related traits using a DH mapping population derived from a cross between a high vigour *japonica* cultivar CT9993 and a low vigour *indica* cultivar IR62266. The linkage map of this population comprised 145 RFLPs, 153 AFLPs and 17 microsatellite markers covering 1788 cM in length with an average distance of 5.7 cM between adjacent markers[2]. In this experiment, seeds of 125 DH lines along with parents and one check, Azucena were sown in black cylindrical pipes measuring 30 cm in length and 10 cm in diameter, which were filled with a mixture of FYM, coir pith and soil in 1 : 1 : 1 proportion. The experiment was laid out in completely randomized design with three replications. After germination, one seedling was allowed to grow in each pipe. Plants were watered daily throughout the experiment to maintain moisture field capacity. After 21 days, the pipes were submerged in water for one hour to loosen the soil and avoid any damage to the seedling. Later, the seedling was carefully removed from the pipe and washed with water to remove any adhering material without damaging any part of seedling. The observations for vigour-related traits were recorded (Table 1). Analysis of variance was done to partition the variance. Interval analysis was performed to detect QTLs using MAPMAKER/QTL[3]. A locus with LOD > 3.00 was declared a putative QTL.

Analysis of variance of all the traits showed significant line differences, revealing desirable variation in the population. The ranges of the mean values of all traits extend beyond that of the parents, exhibiting transgressive segregation. Parent, CT9993, performed better for almost all the traits except for leaf number and root to shoot fresh weight ratio, when compared to parent IR62266.

A total of 29 QTLs for 14 morphological and growth-related traits were tagged to molecular markers (Table 2). The variance explained by each QTL ranged from 10.7 to 38.8. Significant QTLs were located on chromosomes 1 and 3. Four QTLs for total length were identified on chromosomes 3, 5, 10 and 12, which ex-