ELSEVIER

Gene 273 (2001) 63–70

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

# Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*

S.K. Gupta, T.C. Ghosh*

*Distributed Information Centre, Bose Institute, P 1/12, C.I.T. Scheme, VII M, Calcutta 700 054, India*

Received 31 January 2001; received in revised form 4 June 2001; accepted 14 June 2001
Received by T. Gojobori

**Abstract**

Codon usage biases of all DNA sequences (length greater than or equal to 300 bp) from the complete genome of *Pseudomonas aeruginosa* have been analyzed. As *P. aeruginosa* is a GC-rich organism, G and/or C are expected to predominate in their codons. Overall codon usage data analysis indicates that indeed codons ending in G and/or C are predominant in this organism. But multivariate statistical analysis indicates that there is a single major trend in the codon usage variation among the genes in this organism, which has a strong negative correlation with the expressivities of the genes. The majority of the lowly expressed genes are scattered towards the positive end of the major axis whereas the highly expressed genes are clustered towards the negative end. This is the first report where the prokaryotic organism having highly skewed base composition is dictated mainly by translational selection, though some other factors such as the lengths of the genes as well as the hydrophobicity of genes also influence the codon usage variation among the genes in this organism in a minor way. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords*: Synonymous codon usage; Correspondence analysis; Translational selection; *Pseudomonas aeruginosa*; Highly expressed genes; Lowly expressed genes

## 1. Introduction

It is well established that the codon usage pattern is non-random and species-specific. It has also been reported that there is significant variation of codon usage bias among the different genes within the same organism (Grantham et al., 1981; Wada et al., 1990). Biased usage of synonymous codons may result from various factors. In organisms having highly skewed base composition, it was observed that compositional constraints are the main factors in determining the codon usage variation among the genes (Ohkubo et al., 1987; Ohama et al., 1990; Andersson and Sharp, 1996b; Andersson et al., 1998; Musto et al., 1998). It was suggested that translational selection manifests the codon usage biases of highly expressed genes and subsequently it was reported that preferred codons in highly expressed genes are recognized by most abundant tRNAs (Ikemura, 1981a,b, 1982; Bennetzen and Hall, 1982). In some unicellular organisms it has been reported that both translational selection and compositional constraints operate in dictating the codon usage variation among genes (Ikemura, 1981a,b; Gouy and Gautier, 1982; Andersson and Sharp, 1996a; Romero et al., 2000b; Ghosh et al., 2000). In recent years, codon usage analyses on several complete genomes has indicated that the physical location of each gene on the chromosome determines codon usage patterns (Kerr et al., 1997). It has also been reported that in different organisms, replication-translational selection is the main source of variation of codon usage biases among the genes (McInerney, 1998; Lafay et al., 1999; Romero et al., 2000a). Recently, it has been reported that the hydrophobicity of each gene is one of the main factors in determining the codon usage variation among the genes in *Mycobacteria* (de Miranda et al., 2000).

*Pseudomonas aeruginosa* is a GC-rich Gram-negative bacteria. It has been considered as one of the major opportunistic human pathogens during the past century (Stover et al., 2000). *Pseudomonas aeruginosa* is a well-studied microorganism and is the largest prokaryote whose

Abbreviations: $A_{3s}$, $T_{3s}$, $G_{3s}$, $C_{3s}$ are the distributions of adenine, thymine, guanine and cytosine at the synonymous third positions of codons, respectively; CAI, codon adaptation index; GC, molar fraction of guanine + cytosine in DNA; $GC_{3s}$, frequency of guanine + cytosine at the synonymous third positions of codons; $N_c$, effective number of codons used by a gene; RSCU, relative synonymous codon usage

* Corresponding author. Fax: +91-33-334-3886.
*E-mail address:* tapash@boseinst.ernet.in (T.C. Ghosh).

0378-1119/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0378-1119(01)00576-5

complete genome has been published (Stover et al., 2000). Thus, extensive possibilities, earlier unavailable, have opened up to gain insight into the biology of this microorganism. In the present study, we have used all coding sequences (greater than or equal to 300 bp) and analyzed the codon usage data with a view to understanding the genetic organization of the *P. aeruginosa* genome. Our results demonstrate that translational selection is the major source of codon usage variation among the genes in this highly skewed compositional genome.

## 2. Materials and methods

The complete genome of *P. aeruginosa* has been downloaded from www.ncbi.nlm.nih.gov/genbank/genomes. To minimize sampling errors we have taken only those sequences which are greater than or equal to 300 bp and have the correct initial and termination codons. Finally, 5239 sequences were selected for data analysis. The coding sequences from the complete genome were retrieved using a program in C, developed by us.

Relative synonymous codon usage (RSCU) was used to study the overall codon usage variation among the genes. RSCU is defined as the ratio of the observed frequency of codons to the expected frequency if all the synonymous codons for those amino acids are used equally (Sharp and Li, 1986). RSCU values greater than 1.0 indicate that the corresponding codon is more frequently used than expected, whereas the reverse is true for RSCU values less than 1.0.

$GC_{3s}$ is the frequency of (G + C) and $A_{3s}$, $T_{3s}$, $G_{3s}$, and $C_{3s}$ are the distributions of A, T, G and C at the synonymous third positions of codons. $N_c$ is the effective number of codons used by a gene, generally used to measure the bias of synonymous codons and independent of amino acid compositions and codon numbers (Wright, 1990). The values of $N_c$ range from 20 (when one codon is used per amino acid) to 61 (when all the codons are used with equal probability). The expected value of $N_c$ under random codon usage can be expressed as:

$$N_c = 2 + s + \{29/[s + (1 - s)^2]\}$$

where $s = GC_{3s}$.

Gene expressivities were measured by calculating the parameter codon adaptation index (CAI) as defined by Sharp and Li (1987).

All the above-mentioned parameters were calculated by using the program CodonW 1.3 (available at www.molbiol.ox.ac.uk/cu). Correspondence analysis (Greenacre, 1984) available in CodonW was used to investigate the major trend in codon usage variation among genes.

GC skew, defined as the ratio of (G − C) to (G + C) along the DNA sequences, was calculated using a sliding window of 60 kb and a step size of 6 kb.

## 3. Results and discussion

### 3.1. Overall codon usage analysis

A total of 5239 genes were used in this study. Since *P. aeruginosa* is a GC-rich genome, it is expected that G and/or C containing codons will predominate in the coding regions. Table 1 shows the overall RSCU values of this organism. It is evident that codons ending in G and/or C are predominant in the entire coding region. However, the overall codon usage values may obscure some heterogeneity of codon usage bias among the genes that might be superimposed on the extreme genomic composition of this organism.

### 3.2. Heterogeneity of codon usage

Two indices, viz. effective number of codons used by a gene ($N_c$) and (G + C) percentage at the third synonymous codon positions ($GC_{3s}$) are generally used to study the codon usage variation among the genes in any organism. Fig. 1

Table 1
Overall codon usage data of *P. aeruginosa* genes[a]

| AA | Codon | *N* | RSCU | AA | Codon | *N* | RSCU |
|---|---|---|---|---|---|---|---|
| Phe | UUU | 3036 | 0.10 | Ser | UCU | 1451 | 0.09 |
| | UUC | 59277 | 1.90 | | UCC | 21085 | 1.31 |
| Leu | UUA | 502 | 0.01 | | UCA | 1017 | 0.06 |
| | UUG | 15434 | 0.42 | | UCG | 22987 | 1.42 |
| Tyr | UAU | 9276 | 0.42 | Cys | UGU | 1716 | 0.20 |
| | UAC | 35292 | 1.58 | | UGC | 15560 | 1.80 |
| ter | UAA | 445 | 0 | ter | UGA | 3977 | 0 |
| ter | UAG | 578 | 0 | Trp | UGG | 26052 | 1.00 |
| Leu | CUU | 5406 | 0.15 | Pro | CCU | 3712 | 0.17 |
| | CUC | 48876 | 1.34 | | CCC | 22897 | 1.03 |
| | CUA | 2478 | 0.07 | | CCA | 3815 | 0.17 |
| | CUG | 145724 | 4.00 | | CCG | 58605 | 2.63 |
| His | CAU | 10932 | 0.58 | Arg | CGU | 13859 | 0.62 |
| | CAC | 26925 | 1.42 | | CGC | 86650 | 3.88 |
| Gln | CAA | 10931 | 0.29 | | CGA | 4148 | 0.19 |
| | CAG | 63685 | 1.71 | | CGG | 24843 | 1.11 |
| Ile | AUU | 4986 | 0.21 | Thr | ACU | 2874 | 0.16 |
| | AUC | 66177 | 2.73 | | ACC | 57531 | 3.15 |
| | AUA | 1662 | 0.07 | | ACA | 1404 | 0.08 |
| Met | AUG | 35200 | 1.00 | | ACG | 11218 | 0.61 |
| Asn | AAU | 6558 | 0.28 | Ser | AGU | 4646 | 0.29 |
| | AAC | 39863 | 1.72 | | AGC | 45731 | 2.83 |
| Lys | AAA | 6134 | 0.25 | Arg | AGA | 834 | 0.04 |
| | AAG | 43605 | 1.75 | | AGG | 3540 | 0.16 |
| Val | GUU | 4717 | 0.16 | Ala | GCU | 8338 | 0.16 |
| | GUC | 50504 | 1.67 | | GCC | 119213 | 2.33 |
| | GUA | 6958 | 0.23 | | GCA | 8491 | 0.17 |
| | GUG | 58674 | 1.94 | | GCG | 68493 | 1.34 |
| Asp | GAU | 18396 | 0.39 | Gly | GGU | 14471 | 0.39 |
| | GAC | 74944 | 1.61 | | GGC | 108992 | 2.94 |
| Glu | GAA | 40792 | 0.77 | | GGA | 7329 | 0.20 |
| | GAG | 65665 | 1.23 | | GGG | 17489 | 0.47 |

[a] AA, amino acids; *N*, number of codons; RSCU, cumulative relative synonymous codon usage of 5239 genes.

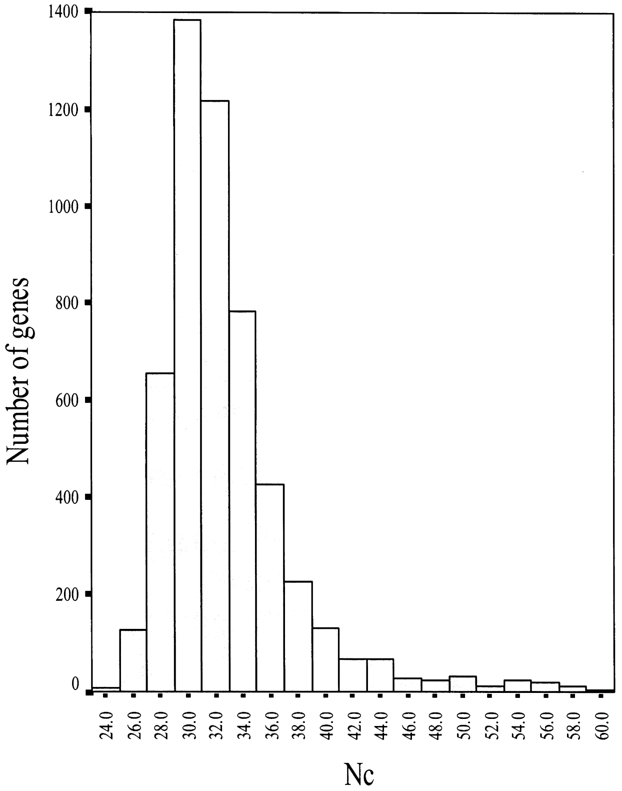

Fig. 1. $N_c$ distribution of *P. aeruginosa* genes.

shows the $N_c$ distribution of different genes in *P. aeruginosa*. The $N_c$ values range from 23.43 to 60.86 (with a mean of 32.72 and standard deviation of 4.77), indicating that there is a wide variation of codon usage bias among the genes. The heterogeneity of codon usage biases among the genes is further confirmed from the distributions of (G + C) at the third synonymous codon positions, shown in Fig. 2. From this figure it is obvious that (G + C) at the synonymous third position of codons varies from 23 to 97% with a mean of 86% and standard deviation of 6%. These results indicate that apart from compositional constraints, other trends might influence the overall codon usage variation among the genes in *P. aeruginosa.*

### *3.3. Exploring different factors in determining the codon usage variation*

#### *3.3.1. The $N_c$ plot*

Wright (1990) suggested that a plot of $N_c$ against $GC_{3s}$ could be effectively used to explore the codon usage variation among the genes. It was demonstrated by Wright (1990) that the comparison of the actual distribution of genes with the expected distribution under no selection could be indicative, if the codon usage bias of the genes had some influence other than the genomic GC composition. In other words, if $GC_{3s}$ is the only determinant of the codon usage variation among the genes then the values of $N_c$ would fall on the continuous curve (the $N_c$–$GC_{3s}$ plot, see Fig. 3), representing random codon usage. From Fig. 3 it is evident that only a small number of points lie on the expected curve, mostly in the GC-rich regions. This indicates that compositional constraints play a role in defining the codon usage variation among those genes. There are also a large number of points which lie well below the expected curve (having low $N_c$ values), which indicate that these genes have an additional codon usage bias, other than genomic GC composition.

#### *3.3.2. Multivariate statistical analysis*

To explore the other possibilities in shaping the codon usage variation among the genes in *P. aeruginosa* we have subjected the data to multivariate statistical analysis, a method that has been successfully used to study the codon usage variation among genes in different organisms (Romero et al., 2000a,b; Ghosh et al., 2000; Wright and Bibb, 1992; Stenico et al., 1994; Musto et al., 1999). Correspondence analysis is one of the multivariate statistical analyses in which the data are plotted in a multidimensional space of 59 axes (excluding Met, Trp and stop codons) and then the most prominent axes contributing to the codon usage variation among the genes are determined. In the present work, correspondence analysis has been performed on RSCU values to minimize the effects of amino acid composition. Fig. 4 shows the positions of genes along the first two major axes. The first major axis accounted for 18.5% of the total variation while none of the other axes accounted for more than 5%. It must be remembered that

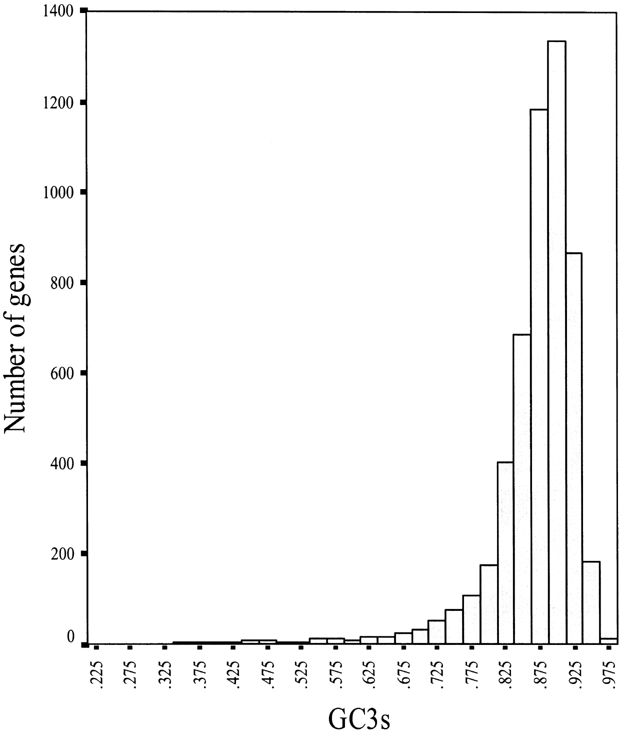

Fig. 2. Compositional distribution of (G + C) at the synonymous third positions of codons ($GC_{3s}$) of *P. aeruginosa* genes.

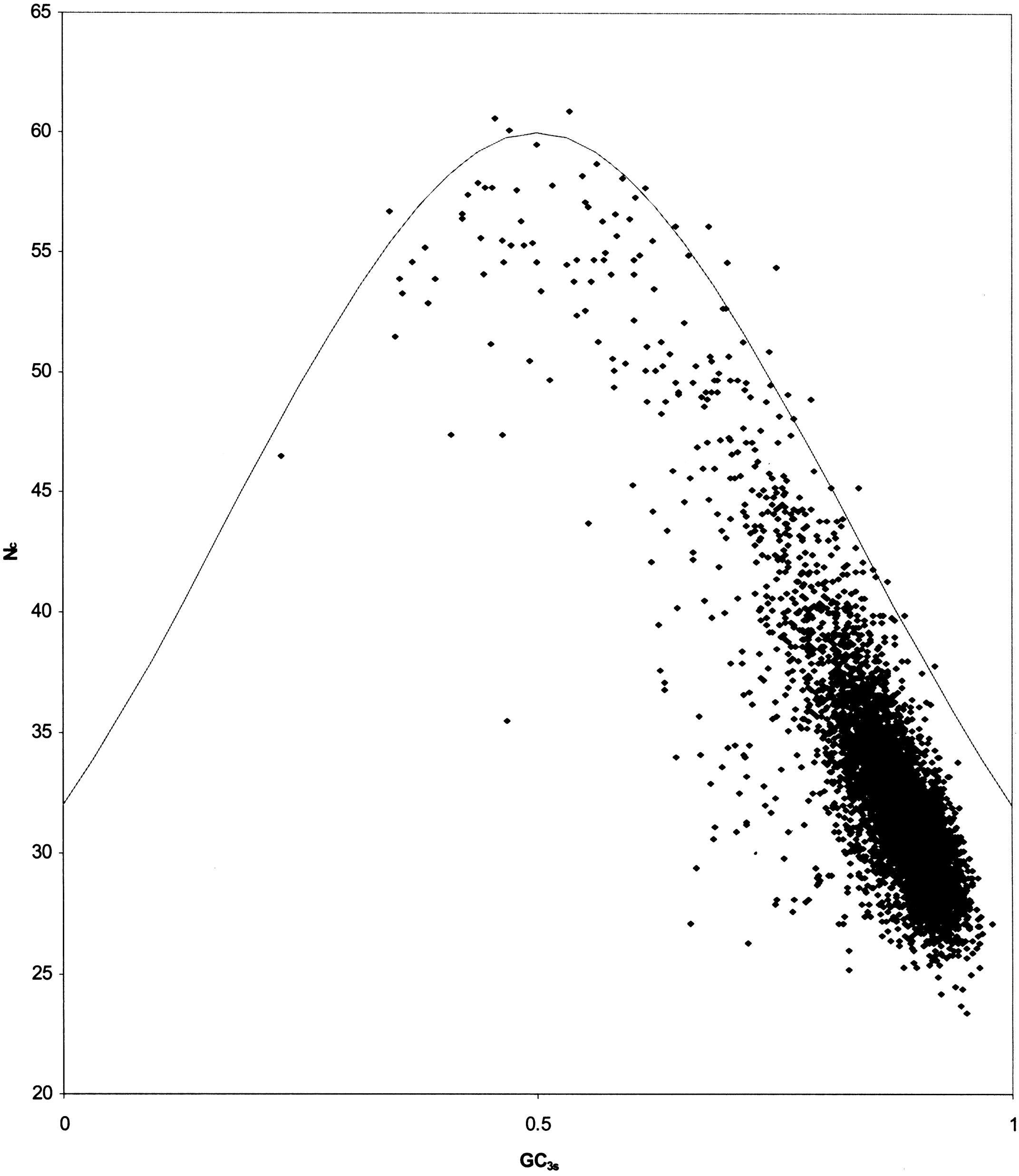

Fig. 3. $N_c$ plot of *P. aeruginosa* genes. The continuous curve represents the expected curve between $GC_{3s}$ and $N_c$ under random codon usage.

although the first principal axis explains a substantial amount of variation of codon usage among the genes in this microorganism, its value is still lower than found in other organisms studied earlier (Alvarez et al., 1994; Musto et al., 1998). The low value might be due to the extreme genomic composition of this genome. It is also obvious from Fig. 4 that the majority of the points are clustered in a spherical shape around the origin of axes, having an average $N_c$ value of 32.72 and standard deviation of 4.77. This indicates that these genes have more or less similar codon usage biases. There are very few points that are widely scattered along the positive side of the first major axis with an average $N_c$ value of 45.04 and standard deviation of 6.42, indicating that codon usage biases of these genes are not homogeneous. From these results it can be stated that there is a single major trend in codon usage variation among the genes in *P. aeruginosa*.

### 3.3.3. *Effect of gene expressivities on codon usage*

For a long time it has been noted that in organisms with a highly skewed base composition, mutational bias is the main factor in shaping the codon usage variation among the genes whereas translational selection plays a minor role. Overall RSCU values (shown in Table 1) and $N_c$ plot (shown in Fig. 3) provide definite indications that mutational bias is acting in this organism in dictating the codon usage variation among the genes. However, correspondence analysis indicates that there is a single major

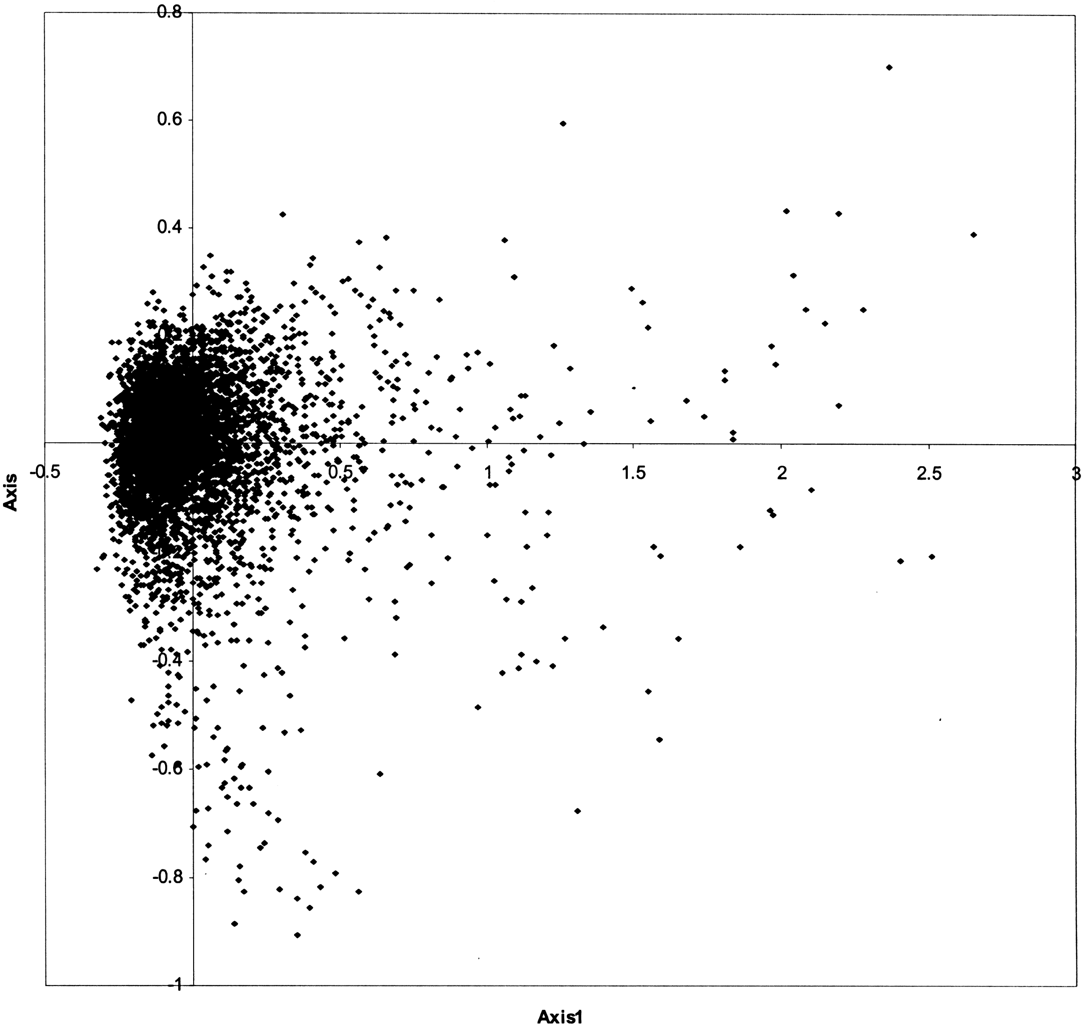

Fig. 4. Correspondence analysis on the RSCU values of *P. aeruginosa* genes. Each point on the plot corresponds to the coordinates on the first and second principal axes produced by the correspondence analysis.

trend in the codon usage among the genes in this bacterium. To assess the effect of expressivities of genes on codon usage biases, we have calculated the expressivity for each gene of this organism. The CAI has been widely used to estimate the expressivities of genes by different workers (Gutierrez et al., 1996; Nakamura and Tabata, 1997; Pan et al., 1998; Tiller and Collins, 2000) and is now considered a well-accepted measure of gene expressivities. In this study we have used the CAI to measure the expression level of a gene. A scatter diagram of the positions of genes on the first major axis was plotted against their corresponding CAI values, as shown in Fig. 5. The correlation coefficient between the positions of genes along the first major axis against their corresponding CAI values as estimated from Fig. 5 is $-0.920$ ($P < 0.01$). This is a clear indication that gene expression is the main cause for the codon usage variation among the genes in this organism. The positions of the genes along the first major axis are also significantly negatively correlated with $GC_{3s}$ ($r = -0.896$, $P < 0.01$) and with the length of the gene ($r = -0.119$, $P < 0.01$). We also obtained a significant negative correlation between the positions of the genes in the first major axis and the hydropathy of each protein ($r = -0.060$, $P < 0.01$) and a positive correlation with the third major axis ($r = 0.060$, $P < 0.01$). The positions of the genes along the second major axis are strongly positively correlated with $GC_{3s}$ ($r = 0.207$, $P < 0.01$). These results suggest that highly expressed genes have lower (G + C) content at their synonymous third codon positions than the lowly expressed genes. It also indicates that highly expressed genes are longer. Earlier it was observed that in case of *Escherichia coli* there was a strong positive correlation between the synonymous codon usage and gene length and it was proposed that longer genes should impose stronger constraints on the codon usage bias (Moriyama and Powell, 1998). The positive correlation between the synonymous codon usage and gene length in this organism is not surprising since it was observed (by comparing gene sequences of *P. aeruginosa*) that *E. coli* is the closest relative of *P. aeruginosa* (Stover et al., 2000).

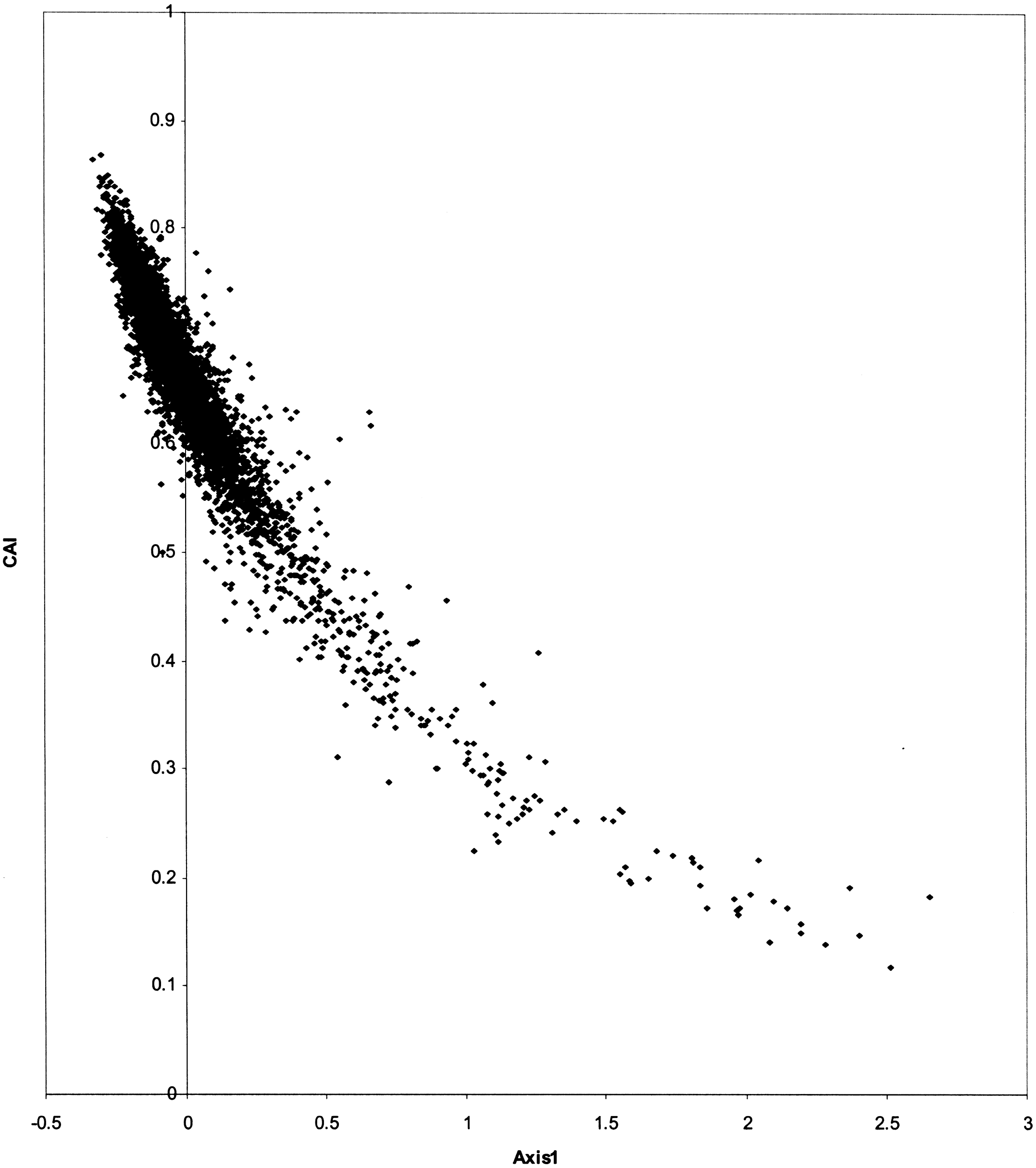

Fig. 5. The scatter diagram of the positions of *P. aeruginosa* genes on the first major axis generated by correspondence analysis against their CAI values.

### 3.3.4. *Influence of replication-transcriptional selection on codon usage*

In recent years it was reported that in several bacteria, codon usage bias is mainly dictated by transcriptional-translational selection (McInerney, 1998; Lafay et al., 1999; Romero et al., 2000a). We explored this possibility by studying codon usage bias in the leading and lagging strands of *P. aeruginosa*. For locating the leading and lagging strands we have calculated GC skew along the DNA of *P. aeruginosa* by taking a 60 kb window and a step size of 6 kb. GC skew is generally used to locate the leading and lagging strands of a prokaryotic organism and it has been reported that in most prokaryotes there is a change in sign of GC skew near the origin of replication (Lobry, 1996; McLean et al., 1998). Fig. 6 shows the GC skew of this organism. There is a change in sign around 240 kb and the major part of the genome lies on the negative side of the GC skew curve. It has also been reported that in the leading strand there is an excess of G over C and most of the genes lie on the leading strand (McInerney, 1998; McLean et al., 1998). But in the GC skew positive side, we have obtained only 2038 genes and the rest (3201 genes) are found to lie in the lagging strand. Earlier it was reported that in most prokaryotic genomes a larger number of genes are located in the leading strand than in the lagging strand (McLean et al., 1998). In this respect, the sequence of *P. aeruginosa* is exceptional in having more genes in the lagging strand. Interestingly, we have also observed that of the genes clustered near the

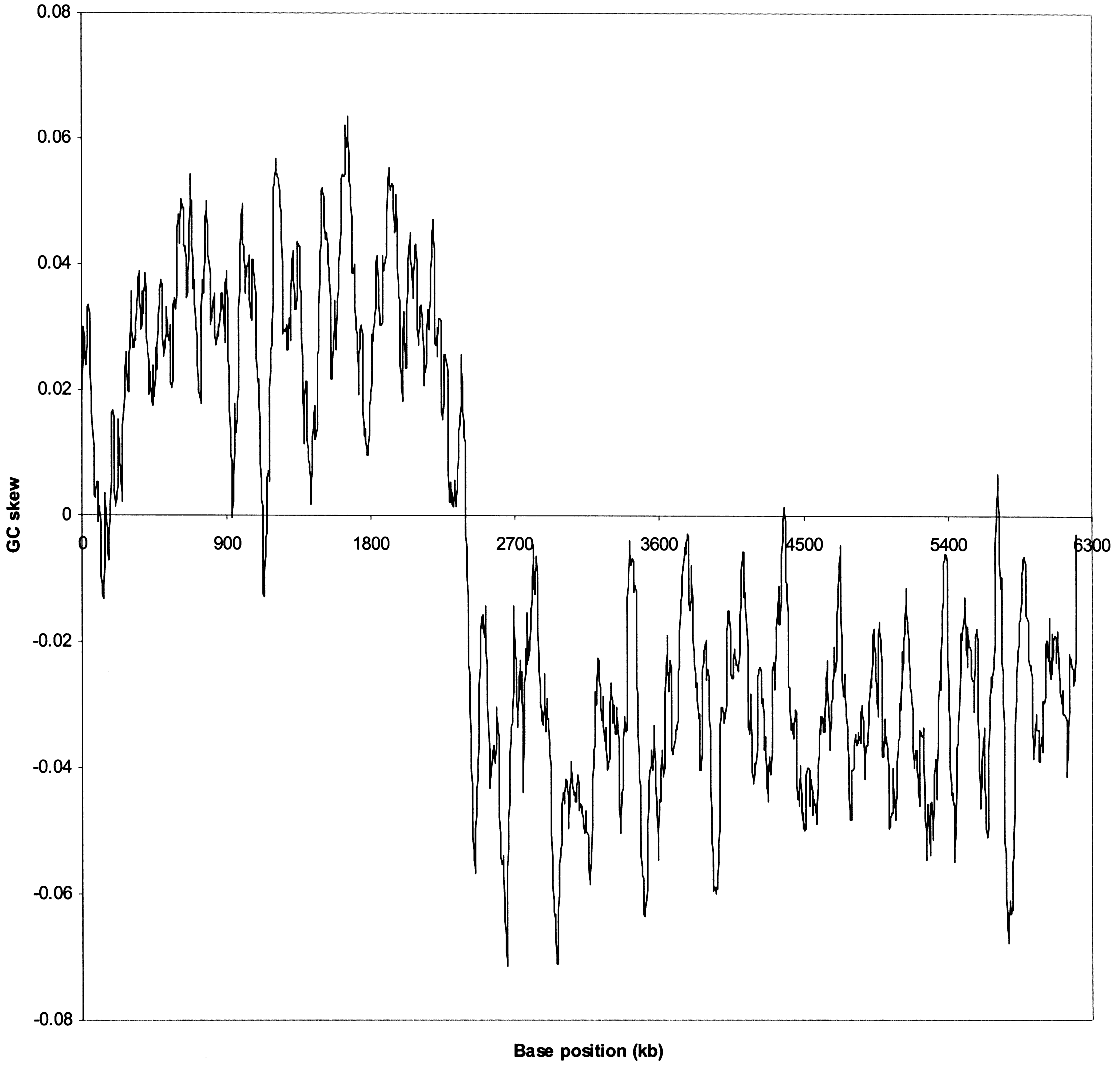

Fig. 6. GC skew (the ratio of (G − C) and (G + C)) of the *P. aeruginosa* genome. The GC skew was calculated by taking a window of size 60 kb with a step of 6 kb.

origin of the first two major axes generated by correspondence analysis, 91% are located in the positive side of the GC skew curve and 94% are located in the negative. From this it is clear that replication-transcriptional bias has no effect in shaping the codon usage variation among the genes in this organism. In conclusion it can be asserted that codon usage bias among the genes in *P. aeruginosa* is mainly dictated by translation selection, though some other factors also have an influence in a minor way. This is obviously a new paradigm on the codon usage biases in this highly skewed base compositional genome.

## Acknowledgements

The authors are thankful to the Department of Biotechnology, Government of India, for financial help, and to Dr P.K. Parrack and Professor M.K. Pal for improving the manuscript.

## References

Alvarez, F., Robello, C., Vignali, M., 1994. Evolution of codon usage and base contents in kinetoplastid protozoan. Mol. Biol. Evol. 11, 790–802.

Andersson, S.G.E., Sharp, P.M., 1996a. Codon usage in the *Mycobacterium tuberculosis* complex. Microbiology 142, 915–925.

Andersson, S., Sharp, P., 1996b. Codon usage and base composition in *Rickettsia prowazekii*. J. Mol. Evol. 42, 525–536.

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Almark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H., Kurland, C.G., 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396, 133–140.

Bennetzen, J.L., Hall, B.D., 1982. Codon selection in yeast. J. Biol. Chem. 257, 3026–3031.

de Miranda, A.B., Alvarez-Valin, F., Jabbari, K., Degrave, W.M., Bernardi, G., 2000. Gene expression, amino acid conservation, and hydrophobicity are the main factors shaping codon preferences in *Mycobacterium tuberculosis* and *Mycobacterium leprae*. J. Mol. Evol. 50, 45–55.

Ghosh, T.C., Gupta, S.K., Majumdar, S., 2000. Studies on codon usage in *Entamoeba histolytica*. Int. J. Parasitol. 30, 715–722.

Gouy, M., Gautier, C., 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10, 7055–7074.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., Mercier, R., 1981. Codon catalog usage in genome strategy modulated for genes expressivity. Nucleic Acids Res. 9, r43–r74.

Greenacre, M.J., 1984. Theory and Applications of Correspondence Analysis. Academic Press, London.

Gutierrez, G., Marquez, L., Mann, A., 1996. Preference for guanine at first codon position in highly expressed *Escherichia coli* genes. A relationship with translational efficiency. Nucleic Acids Res. 24, 2525–2527.

Ikemura, T., 1981a. 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.

Ikemura, T., 1981b. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. 146, 1–21.

Ikemura, T., 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in proteins genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. 158, 573–587.

Kerr, A.R.W., Peden, J.F., Sharp, P.M., 1997. Systematic base composition variation around the genome of *Mycoplasma genitalium*, but not *Mycoplasma pneumoniae*. Mol. Microbiol. 25, 1177–1179.

Lafay, B., Lloyd, A.T., McLean, M.J., Devine, K.M., Sharp, P.M., Wolfe, K.H., 1999. Proteome composition and codon usage in *spirochaetes*: species-specific and DNA strand-specific mutational bias. Nucleic Acids Res. 27, 1642–1649.

Lobry, J.R., 1996. Origin of replication of *Mycoplasma genitalium*. Science 272, 745–746.

McInerney, J.O., 1998. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. Proc. Natl. Acad. Sci. USA 95, 10698–10703.

McLean, M.J., Wolfe, K.H., Devine, K.M., 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J. Mol. Evol. 47, 691–696.

Moriyama, E.N., Powell, J.R., 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res. 26, 3188–3193.

Musto, H., Romero, H., Maseda, H.R., 1998. Heterogeneity in codon usage in the flatworm *Schistosoma mansoni*. J. Mol. Evol. 46, 159–167.

Musto, H., Romero, H., Zavala, A., Jabbari, K., Bernardi, G., 1999. Synonymous codon choices in the extremely GC-poor genome of *Plasmodium falciparum*: compositional constraints and translational selection. J. Mol. Evol. 49, 27–35.

Nakamura, Y., Tabata, S., 1997. Codon-anticodon assignment and detection of codon usage trends in seven microbial genomes. Microbiol. Comp. Genomics 2, 299–312.

Ohama, T., Muto, A., Osawa, S., 1990. Role of GC-biased mutation pressure on synonymous codon choice in *Micrococcus luteus*, a bacterium with a high genomic GC-content. Nucleic Acids Res. 18, 1565–1569.

Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F., Osawa, S., 1987. The ribosomal protein gene cluster of *Mycoplasma capricolum*. Mol. Gen. Genet. 210, 314–322.

Pan, A., Dutta, C., Das, J., 1998. Codon usage in highly expressed genes of *Haemophillus influenzae* and *Mycobacterium tuberculosis*: translational selection versus mutational bias. Gene 215, 405–413.

Romero, H., Zavala, A., Musto, H., 2000a. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Res. 28, 2084–2090.

Romero, H., Zavala, A., Musto, H., 2000b. Compositional pressure and translational selection determine codon usage in the extremely GC-poor unicellular eukaryote *Entamoeba histolytica*. Gene 242, 307–311.

Sharp, P.M., Li, W.-H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res. 14, 7737–7749.

Sharp, P.M., Li, W.-H., 1987. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15, 1281–1295.

Stenico, M., Lloyd, A.T., Sharp, P.M., 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. Nucleic Acids Res. 22, 2437–2446.

Stover, C.K., et al., 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature 406, 959–964.

Tiller, E.R., Collins, R.A., 2000. The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. J. Mol. Evol. 50, 249–257.

Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori, T., Ikemura, T., 1990. Codon usage tabulated from GenBank genetic sequence data. Nucleic Acids Res. 18 (Suppl.), 2367–2411.

Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87, 23–29.

Wright, F., Bibb, M.J., 1992. Codon usage in the G + C rich *Streptomyces* genome. Gene 113, 55–65.