Journal of Proteomics & Bioinformatics - Open Access

Research Article JPB/Vol.1/July 2008

Analysis of Synonymous Codon Usage Bias in *Pseudomonas Syringae* Phages: Implication in Phage Therapy for Halo Blight Disease

Shashank Sharma, Ambrish Sharan Vidyarthi and Raju Poddar*

Department of Biotechnology

Birla Institute of Technology, Mesra, Ranchi-835 215, India

*Corresponding author: Raju Poddar, Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi-835 215, India, Tel: +91-651-2276223; Fax: +91-651-2275401; E-mail: rpoddar@bitmesra.ac.in

Received April 22, 2008; Accepted July 16, 2008; Published July 17, 2008

Citation: Shashank S, Ambrish SV, Raju P (2008) Analysis of Synonymous Codon Usage Bias in *Pseudomonas Syringae* Phages: Implication in Phage Therapy for Halo Blight Disease. J Proteomics Bioinform 1: 206-218. doi:10.4172/jpb.1000025

Abstract

Halo blight disease caused due *Pseudomonas syringae* results extensive losses in dry Beans. Phage therapy can be tried as an alternative treatment. For evaluation of synonymous codon usage and codon variation, 20 phages of *Pseudomonas* has been taken for testing. The effect of GC concentration in the phages can be considered for analyzing virulence in these phages. Mutational biasing and translational selection are also the important factors for predicting the appropriate biasing, which can be analyzed through Nc plot and correspondence analysis. Our analysis indicating that out of 20 phages 3 phages namely phage D3, D3112 and 119X are extremely virulent as they have high translational efficiency. Based on our data, we conclude that the phage D3 will be best suited as a phage therapy for treatment against halo blight disease.

Keywords: Halo blight; Relative synonymous codon usage; Correspondence analysis; Translational selection; Multivariate statistical analysis; *Pseudomonas syringae*

Abbreviations: RSCU: Relative Synonymous Codon Usage; CAP: Correspondence Analysis; N_c; Effective number of codons; GC_{3s}: The frequency of (G + C) at synonymous third codon positions.

Introduction

Halo blight disease(Saettler, 1991) observed through world wide and it can cause extensive losses in dry beans. *Pseudomonas syringae* is a legume pathogen of worldwide importance and is mainly responsible for Halo blight in Beans (Burkholder, 1926; Burkholder, 1930). Phage therapy of this bacterial pathogen can be tried as an alternative treatment for protecting these legumes against such losses.

Many amino acids are coded by more than one codon and therefore the multiple codons for a given amino acid are synonymous (Ikemura, 1985). However many genes displays a non random usage of specific amino acid and the measure of the extent to this non randomness is given by Relative Synonymous Codon Usage (Sharp and Li, 1987). Some genes have extremely biased codon usage: these genes appear to be expressed at higher levels, and other genes (apparently those expressed at low levels) have relatively unbiased codon usage. The other factors which are responsible for variation in codon usage are mutational biasing (Levin and Whittome, 2000) and translational selection (Grantham et al., 1981).

The total of 1214 genes of twenty phages of Pseudomonas are considered which are classified in six families viz. podoviridae, myoviridae, siphoviridae, inoviridae, cystoviridae and leviviridae (Krieg et al., 1984). The genome of Pseudomonas phages shows a considerable variation in their genome size ranging from 2300 – 280000 bps. The Genome of *Pseudomonas* phage is rich in G-C content which an average accounts for nearly 55% of the total genome content. We found that the optimal codons exhibited variation (AT- or GC-ending codons) in different phages. We also found that genes that were specifically expressed had different patterns of codon usage and local genomic GC (GCg) content. Our efforts to work on this project will provide a path for treatment of halo blight using phage therapy of *Pseudomonas*. In this paper, we studied the synonymous codon usage bias in all the phages of *Pseudomonas* whose

genomic sequences are known.

Materials and Methods

The gene sequences of twenty different phages of *Pseudomonas* were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/). The gene sequences are retrieved in the FASTA format. The total of 1214 genes is considered for analysis. These genes are of twenty different phages of six families namely podoviridae, myoviridae, siphoviridae, inoviridae, cystoviridae and leviviridae, and were used to study codon bias in these phages. The basic nature and status of the above twenty phages are presented in Table 1.

The gene number for these phages varies from 4 to 301. All the above phage genomes were extracted from featurable table of genome according to gene bank records and all this gene sequences were used for the comparative analysis of the codon usage studies.

For multivariate analysis we performed Correspondence

Family	Phage name	Accession No.[a]	No. of genes per phage[b]
Podoviridae	119X	NC_007807	53
	F116	NC_006552	70
	gh1	NC_004665	42
	PaP2	NC_005884	58
	phiKMV	NC_005045	49
	Pap3	NC_004466	75
Myoviridae	phiCTX	NC_003278	47
	phiKZ	NC_004629	306
	EL	NC_007623	201
Siphoviridae	B3	NC_006548	59
	D3112	NC_005178	55
	DMS3	NC_008717	52
	D3	NC_002484	99
Inoviridae	Pf1	NC_001331	14
	Pf3	NC_001418	9
Cystoviridae	Phi8L	NC_003299	6
	Phi12L	NC_004173	6
	Phi13M	NC_004171	5
	Phi6L	NC_003715	4
Leviviridae	PP7	NC_001628	4

Table 1: Tabulated classification of phages from NCBI.

[a] Gene Bank Accession No. , The number assigned by database NCBI (USA).
[b] No. of genes per phage as downloaded from database.

Analysis (Greenacre, 1984) which is available on CodonW 1.3 (accessible at www.molbiol.ox.ac.uk/cu). The Relative Synonymous Codon Usage (Sharp et al., 1987) identifies when a codon is being used more frequently than expected and when it is being used less frequently than expected RSCU values are the number of times a particular codon is observed, relative to the number of times that the codon would be observed in the absence of any codon usage bias. Sometimes the observed frequency will be greater than the expected frequency if RSCU value is greater than 1.00, and sometimes it will be less when RSCU value is less than 1.00. RSCU values of each codon for the two groups of genes located at the extreme ends of the first major axis are determined by Correspondence analysis. Each group contains 10% of sequences located on the two extremes of the first major axis.

A_{3s}, T_{3s}, G_{3s}, C_{3s}, are the frequencies of the bases A (adenine), C (cytosine), G (guanine), and T (thiamine) occurring in codon position in the genome and GC_{3s} is the G+C distribution at the synonymous third positions of codons. This G+C content is considered to be the most effective cause of mutational pressure (Sueoka et al., 2000). N_c value is an effective number of codons measure that quantifies how far the codon usage of a gene deposits from equal usage of synonymous codons (Wright, 1990). N_c values range from 61 for a gene that tends to use all codons with equal frequency to 20 for a gene that is effectively using only a single codon for each amino acid.

There is a relationship between N_c and the base composition of a gene with genes that have more biased base compositions being expected to have lower N_c values. Usually, lower N_c values might be dictated by the base composition of the gene. This might be taken as evidence that there is some kind of selective pressure on the gene to use a smaller subset of codons. This selective pressure could be translational selection for 'optimal' codons. Optimal codons are those that correspond to the major abundance tRNA for that amino acid. In such circumstances, there could be a selective pressure to use a particular codon that corresponds to this tRNA (Dong et al., 1996).

The copy number of tRNA species in *Pseudomonas* phage D3 strain and corresponding anticodon sequences were determined by the program tRNA scan-SE (available at http://selab.janelina.org/tRNA scan-SE/). It has been shown that the pattern of codon usage in the highly expressed genes of *Escherichia coli* and *Saccharomyces cerevisiae* correlates very strongly with the known abundances of the iso-accepting transfer-RNAs (tRNAs)) (Ikemura, 1981; Bennetzen and Hall, 1982; Ikemura, 1982; Sharp and Cowe, 1991). The advantage of this system (translational selection) is self-evident- using a codon for which there is an abundant cognate tRNA can speed up the process of mRNA translation.

The graph was plotted between various attributes of codon usage like graph of N_c-GC_{3s}, Axis 1 - Axis 2, AXIS 2 - N_c were plotted by the SIGMA PLOT 9.0 . The Correlations coefficient between the positions of genes along the first two major axis with different parameters for codon usage was calculated through SYSTAT 11.0. The alignment among the 20 phages for generating an aligned sequence for dendogram is produced by Clustalw 1.83. A dendrogram representing the extent of divergence in synonymous codon usage among the total phages of pseudomonas was constructed by the DS GENE-SCAN. The tRNA counts present in the genome of phage are calculated by tRNA SCAN SE Server.

Results and Discussion

Variations in Synonymous Codon Usage

The values of RSCU had been determined in 1214 genes of all the 20 phages of *Pseudomonas.* It has been observed that all the phages carry GC rich genome. G and C ending codon are predominant in all phages. The concentration of GC present in overall genome of the phage ranges from 36% (phage phiKZ) to 64% (phage DMS3 and phage D3112). To detect codon usage variation if present in any gene of the above-mentioned phages, effective number of codons used by a gene (N_c) and the (G + C) percentage of the synonymous third positions of the codons (GC_{3s}) were determined.

The value of GC3s ranges from 0.144 to 0.912 with an average of 0.549 whereas the value of Nc ranges from 22.74 to 61 with an average of 48.16. The marked intragenomic variation in GC3s (standard variation > 7%) and in N_c values (standard deviation > 4.4% except for phage PP7). These observations indicate that there is a significant heterogeneity in composition within the phage genome of *Pseudomonas*. The average codon usage bias and the base composition of 20 *pseudomonas* phage are mentioned in Table 2.

The Effect of Mutational Bias on Codon Usage Variation

A study of correlations between introns and coding region base composition shows that variation in mutation pattern also contributes to codon bias variation (Kliman and Hey, 2003). The strength of base composition correlations between introns and codon third positions is greater for genes with low codon bias than for genes with high codon bias. One direct effect of mutation bias on genome evolution is to influence genome composition, which can be measured by G+C content. For analyzing the determinants of codon usage bias in the phages of *Pseudomonas*, N_c plots (a plot of N_c versus GC_{3s}) and the correspondence analysis (CA) are used widely.

The N_c plot drawn for the genes of 20 Pseudomonas phages are displayed in Figure 1. Some of the points especially of phage phiKZ, phage119X and phage B3 lie on the expected curve towards GC-poor regions (GC value 0.144 to 0.2) which certainly originates from extremely mutational bias. It is evident from the figure that a considerable number of genes lie well below the expected curve, indicating that codon usage bias of these genes are influenced by the forces other than genomic GC composition. Points demon-

Family	Phage name	Accession No.[a]	GC%	T_{3s}	C_{3s}	A_{3s}	G_{3s}	N_c	GC_{3s}	Gravy
Podoviridae	119X	**NC_007807**	44	0.37	0.26	0.33	0.29	49	0.43	-0.22
	F116	**NC_006552**	63	0.22	0.35	0.27	0.32	48	0.58	-0.85
	gh1	**NC_004665**	57	0.19	0.33	0.22	0.38	46	0.62	-0.43
	PaP2	**NC_005884**	45	0.27	0.28	0.34	0.32	48	0.49	-0.15
	phiKMV	**NC_005045**	62	0.25	0.30	0.31	0.29	49	0.52	-0.88
	Pap3	**NC_004466**	52	0.26	0.41	0.22	0.35	46	0.60	-0.30
Myoviridae	phiCTX	**NC_003278**	62	0.24	0.32	0.27	0.33	48	0.56	-0.89
	phiKZ	**NC_004629**	36	0.37	0.21	0.41	0.28	48	0.37	0.17
	EL	**NC_007623**	49	0.28	0.31	0.28	0.35	51	0.53	-0.33
Siphoviridae	B3	**NC_006548**	63	0.16	0.44	0.15	0.41	42	0.73	-0.39
	D3112	**NC_005178**	64	0.17	0.48	0.18	0.37	43	0.70	-0.52
	DMS3	**NC_008717**	64	0.15	0.40	0.18	0.41	44	0.70	-0.63
	D3	**NC_002484**	57	0.20	0.38	0.23	0.39	49	0.63	-0.55
Inoviridae	Pf1	**NC_001331**	61	0.24	0.33	0.24	0.34	47	0.57	-0.62
	Pf3	**NC_001418**	45	0.40	0.31	0.27	0.23	52	0.45	0.30
Cystoviridae	Phi8L	**NC_003299**	54	0.27	0.38	0.22	0.32	48	0.59	-0.32
	Phi12L	**NC_004173**	54	0.32	0.30	0.33	0.24	48	0.45	-0.83
	Phi13M	**NC_004171**	57	0.35	0.32	0.28	0.20	49	0.45	-0.52
	Phi6L	**NC_003715**	55	0.29	0.33	0.28	0.28	52	0.52	-0.49
Leviviridae	PP7	**NC_001628**	54	0.28	0.28	0.29	0.31	56	0.50	-0.53

Table 2: Various attributes of twenty phages of *Pseudomonas*.

a. Gene Bank Accession No. ,The no. assigned by database NCBI (USA), NC stands for NCBI code number .
b. % GC, GC content.
c. No. of genes per Phage as downloaded from database.

strated by phage D3 and phage D3112 lie away from the expected curve in comparison with the rest of the phage genes which indicate that the effect of mutational bias on codon usage variation in the former three phage genes is very weak. This phenomenon was further verified with other statistical analysis like correspondence analysis.

The correspondence analysis of RSCU values of 1214 genes of the 20 *Pseudomonas* phages confirms that mutational bias and other factors are also responsible for codon usage variation. The main objective to plot genes in axis 1 and axis 2 space is finding of optimal codons. Optimal codons are defined as those codons that occur significantly more often in highly expressed genes relative to their frequency in lowly expressed genes. Significance is assessed by a two-way chi square contingency test with the criterion of $p < 0.01$. The advantage of using a test of significance to identify optimal codons is that variation in codon usage between highly and lowly expressed genes, that is due to random noise is suppressed. Correspondence analysis is a multivariate statistical analysis technique to study codon usage variation among genes (Wright, 1990). In this analysis, the data are plotted in a multidimensional space of 59 axes (excluding Met, Trp and stop codons), then the most prominent axes are determined that contribute to the codon usage variation among the genes.

The positions of genes along the first as well as the second major axis (generated by CA) are analyzed with the nucleotide composition at the third codon shows that the first major axis is positively correlated with G3 (r, correlation coefficient = 0.067), C3 (r = 0.261 with $P < 0.05$) and negatively correlated with A3 (r, correlation coefficient = -0.043) and T3 (r correlation coefficient = -0.104). In con-

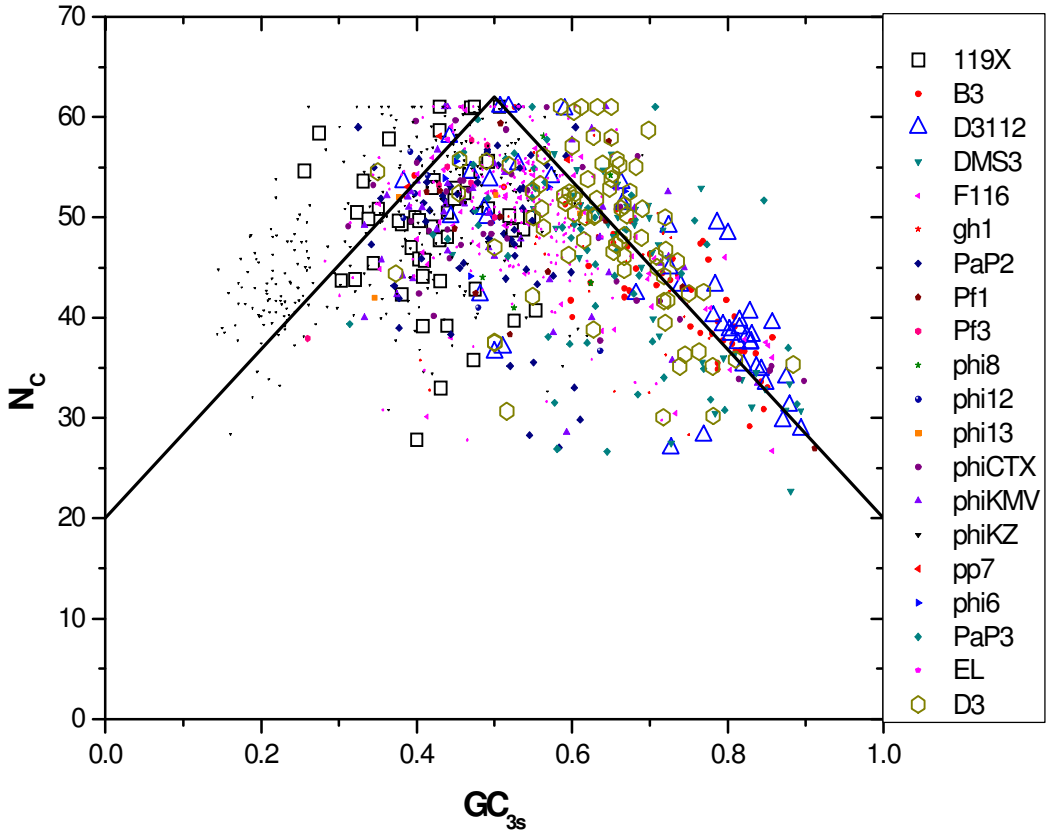

Figure 1: N_c plot of genes from 20 *Pseudomonas* phages.

trast the reverse is true for the second major axis.

The correlation coefficient between the second axis and GC_{3s} is relatively small as compared to that between the axis1 and GC_{3s} (Table 2). But it is worth mentioning that the axis2 exhibits strong negative correlation with G_{3s} and positive correlation with C_{3s} (Table 3). These observations indicate that G_{3s} and C_{3s} interact synergistically in the first principal axis resulting in the increase of GC_{3s} content, but antagonistically in the second principal axis so that increase in the frequency of C_{3s} is accompanied by a decrease in G_{3s} and vice-versa.

The position of genes of the first two major axes (Figure 2) clearly shows that the majority of genes of phage D3, phage D3112 and phage 119X are not clustered with genes of the other phages. To investigate the difference between these two clusters of genes, the codon usage of 10% of the genes located at extreme right side of axis 1 was compared with 10% of the genes located at the extreme left side of axis 1. To access the variation in codon usage between these two genes, chi-square tests were performed taking $P < 0.01$ as the significant criterion. The number and occurrence of each codon and its RSCU values for the two groups of genes are displayed in Table 3.

Out of 21 predominant codons there are 11 C ending codons and 8 G ending codons which actually represent 90.47% of total G and C ending codons. This result suggests that genomic GC composition has a profound effect on in separating the genes along the first major axis according to their RSCU values. It has been reported that RNY codons are more advantageous for translation (Shepherd, 1981). It was also demonstrated that in highly expressed genes of *Escherichia coli*, C is the prominent base at the third codon position (Gutiérrez et al., 1996). The high occurrence of C ending codons in highly expressed genes demonstrates that compositional constraints are not the only factor in determining the codon usage variation in this organism. If compositional constraints are the only dictator in codon usage variation in this organism, the base composition in the third codon position among these optimal codons should have also A or T ending codons as observed in the overall RSCU values of this organism. A similar type of observation was also reported for *Plasmodium falciparum* (Musto et al., 1999).

Cluster analysis has been successfully used to study the frequency of codon usage divergence among the genes of an organism and also among the organisms (Sharp and Li,

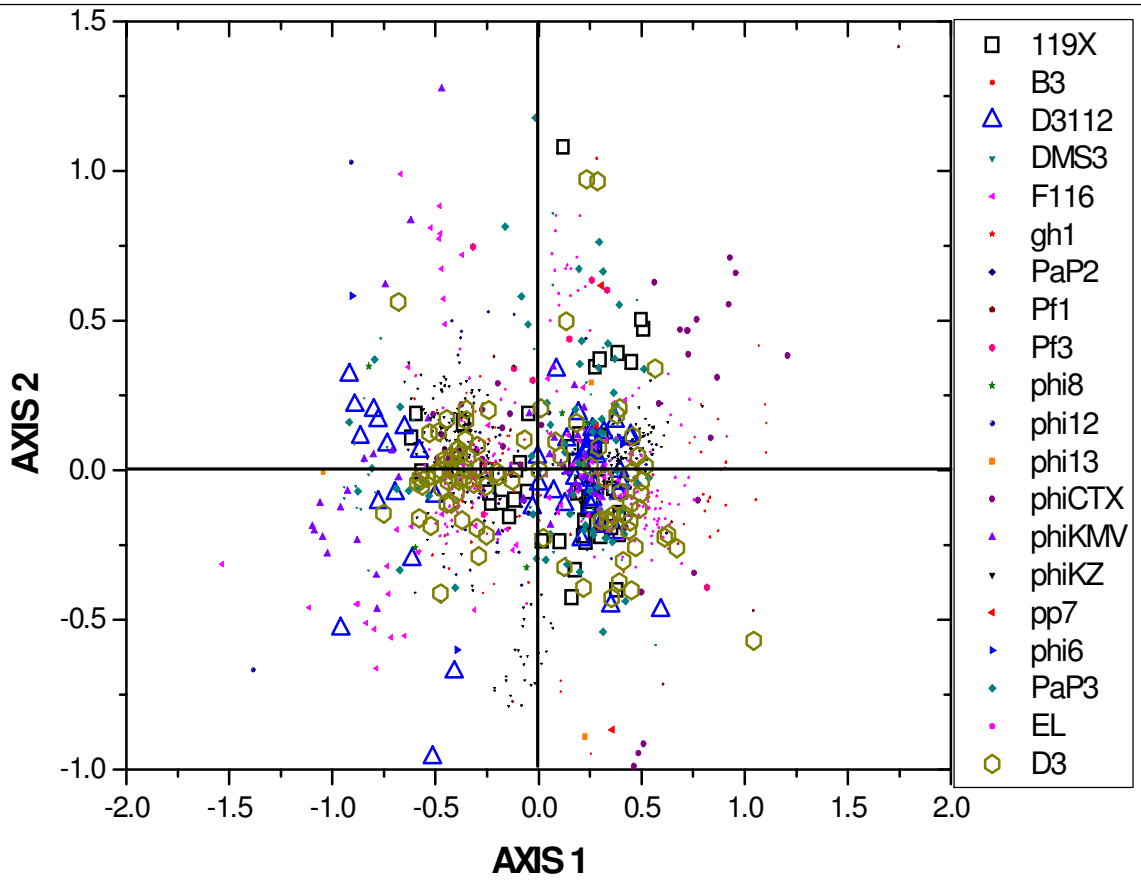

Figure 2: Position of genes from 20 *Pseudomonas* phages along the two major axes of variation in the correspondence analysis on RSCU values.

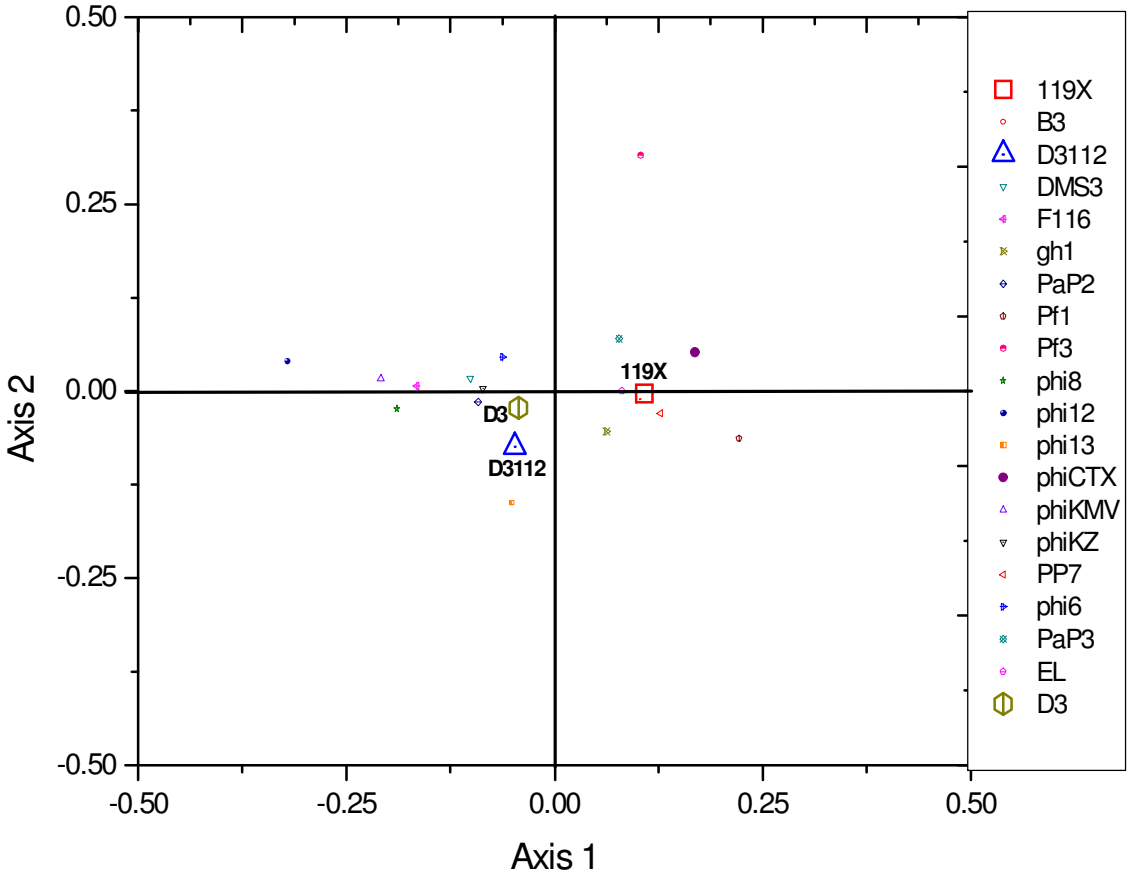

Figure 3: Plot between the average value of Axis 1 and Axis 2 of 20 phages of *Pseudomonas syringae*.

1986). The codon frequency of 64 codons for each organism was compared with the codon frequency with all other organisms. Figure 4 shows the clustering produced by UPGMA (unweighted pair group method using arithmetic averages) method (Sneath, 1973). From the figure 4 it is evident that there are two distinct branches for the twenty *Pseudomonas* phages. Second branch out of the two branches comprises of PP7, PaP3, phiKMV, gh-1, phi-12L, phi-6l and phiEL and the rest are in first branch. Phage D3, D3112 and 119X are of same branch and it suggest that they have nearly similar synonymous codon usage pattern, provided 119x is of family podoviradae, while the other two phage D3 and D3112 are from Siphoviridae.

The above cluster analysis has not only supported our correspondence as well as variance analyses (mentioned above) but also assisted in understanding the intra- and inter-genomic diversities of the *Pseudomonas* bacteriophages in a much better way.

AXIS	A_3	T_3	G_3	C_3	GC_{3s}[a]	N_c[b]	GRAVY
First	-0.043	-0.104	0.067	0.261*	0.502**	-0.134	0.139
Second	0.163	0.018	-0.290**	-0.498	-0.384**	0.014	0.255

Table 3: Correlations coefficient between the positions of genes along the first two major axis with different parameters.

** Correlation is significant at the P-value < 0.01 level.
* Correlation is significant at the P-value < 0.05 level.
[a] GC_{3s}, is the frequency of (G + C) content in the phage gene.
[b] N_c, The effective number of codons used by a gene.

The Influence of Translational Selection Over Codon Usage Variation in Pseudomonas Syringae

In *Caenorhabditis elegans* and *Drosophila melanogaster,* which are characterized by extensive variation in codon usage, the factors governing the choices have been attributed to equilibrium between mutational biases and translational selction (Shields et al., 1988, Sharp and Li, 1989; Moriyama and Gojobori, 1992; Carulli et al., 1993; Akashi, 1993; Akashi, 1997; Stenico et al., 1994; Moriyama and Powell, 1997; Powell et al., 1997). In many organisms, selection acts on synonymous codons to improve translation. The selection on synonymous codon use in *E. coli* is largely due to selection for translation accuracy (Gouy and Gautier, 1982). The plot between second major Axis and N_c values suggest that a substantial number of phage genes, particularly to the phages of D3, D3112 and 119X have lower N_c values as compare to other phage genes (Figure 5).

The first major axis is negatively correlated with N_c, whereas the second major axis is positively correlated with N_c (Table 3). This suggests that considerable number of phage genes carrying GC-rich codons have low N_c values. On the basis of these results, we urge that a balance between mutation and selection due to translational efficiency is strongly operating in selecting the codon usage variation among the genes of *Pseudomonas* phages. Besides phages D3, D3112, 119X mostly carry highly expressed genes.

Various reports suggests that the synonymous codon choices appear to be positively correlated with the relative abundance of tRNAs, with the correlation being very strong for highly expressed genes (Ikemura, 1981; Bennetzen et al., 1982; Ikemura, 1982; Ikemura, 1981; Bennetzen et.al., 1982; Gouy and Gautier, 1982; Sharp et al., 1986; Bulmer, 1988; Bulmer, 1991; Kanaya et al., 1991) In several organisms, cellular tRNA abundance was shown to be directly proportional to their tRNA copy number (Kanaya et al., 1991 Kanaya et al., 2001).

To point any positive correlation between host tRNA abundance and synonymous codon usage of phage D3 and phage D3112 a comparative analysis was made between the copy numbers of *Pseudomonas syringae* specific tRNA and over represented synonymous codons of phage D3 and phage D3112 separately (Table 5).

It was found that out of total 23 over-represented codons of D3, 20 are recognized by those *Pseudomonas syringae* specific tRNAs that have at least one copy in the cell. In contrast, 2 less over-represented codons of D3112 are recognized by abundant tRNAs *Pseudomonas syringae* specific tRNAs. The data indicates that most of the genes of D3 and D3112 have high translation efficiency.

AA	Codon	RSCU[a]	N[a]	RSCU[b]	N[b]	AA	Codon	RSCU[a]	N[a]	RSCU[b]	N[b]
Phe	UUU	0.00	0	1.11	5	Ser	UCU	0.57	2	0.94	20
	UUC*	2.00	11	0.89	4		UCC*	2.29	8	0.70	15
Leu	UUA	0.00	0	0.67	3		UCA	0.00	0	1.03	22
	UUG*	0.63	6	1.78	8		UCG*	0.57	2	1.55	33
	CUU	0.53	5	0.67	3	Pro	CCU	0.75	3	1.03	16
	CUC	0.74	7	0.67	3		CCC*	1.00	4	0.71	11
	CUA	0.32	3	0.44	2		CCA	0.75	3	1.16	18
	CUG*	3.79	36	1.78	8		CCG*	1.50	6	1.10	17
Ile	AUU	0.52	7	0.33	1	Thr	ACU	0.17	1	0.64	13
	AUC*	2.40	32	2.33	7		ACC*	2.26	13	1.14	23
	AUA	0.08	1	0.33	1		ACA	0.35	2	0.79	16
Met	AUG	1.00	16	1.00	16		ACG	1.22	7	1.43	29
Val	GUU	0.57	3	0.29	1	Ala	GCU	0.68	20	0.57	6
	GUC	1.52	8	2.00	7		GCC*	2.15	63	0.48	5
	GUA	0.76	4	0.57	2		GCA	0.79	23	1.71	18
	GUG*	1.14	6	1.14	4		GCG	0.38	11	1.24	13
Tyr	UAU	0.44	2	1.20	3	Cys	UGU	0.00	0	0.67	7
	UAC	1.56	7	0.80	2		UGC	2.00	4	1.33	14
TER	UAA	0.75	1	0.10	1	TER	UGA	2.25	3	2.32	24
	UAG	0.00	0	0.58	6	Trp	UGG	1.00	11	1.00	31
His	CAU	0.25	1	1.25	5	Arg	CGU*	0.69	7	0.20	4
	CAC*	1.75	7	0.75	3		CGC*	3.93	40	0.46	9
Gln	CAA	0.48	7	0.67	4		CGA	0.49	5	1.12	22
	CAG*	1.52	22	1.33	8		CGG	0.69	7	1.07	21
Asn	AAU	0.30	3	0.40	1	Ser	AGU	0.00	0	0.84	18
	AAC*	1.70	17	1.60	4		AGC*	2.57	9	0.94	20
Lys	AAA	0.47	9	0.80	4	Arg	AGA	0.10	1	1.53	30
	AAG*	1.53	29	1.20	6		AGG*	0.10	1	1.63	32
Asp	GAU	0.82	18	1.00	3	Gly	GGU	0.56	5	0.11	1
	GAC*	1.18	26	1.00	3		GGC*	2.56	23	1.00	9
Glu	GAA*	1.01	41	1.00	4		GGA	0.22	2	1.67	15
	GAG	0.99	40	1.00	4		GGG	0.67	6	1.22	11

Table 4: RSCU values of each codon for the two groups of genes located at the extreme ends of the first major axis as determined by CA, Each group contains 10% of sequences located on the two extremes of the first major axis.

The * indicates the codons whose occurrences are significantly ($P < 0.01$) higher in the group of genes on left side than the genes on the right side of first major axis.

[a] RSCU and *N* (the number of codons) values of genes of left group.

[b] RSCU and *N* (the number of codons) values of genes of right group.

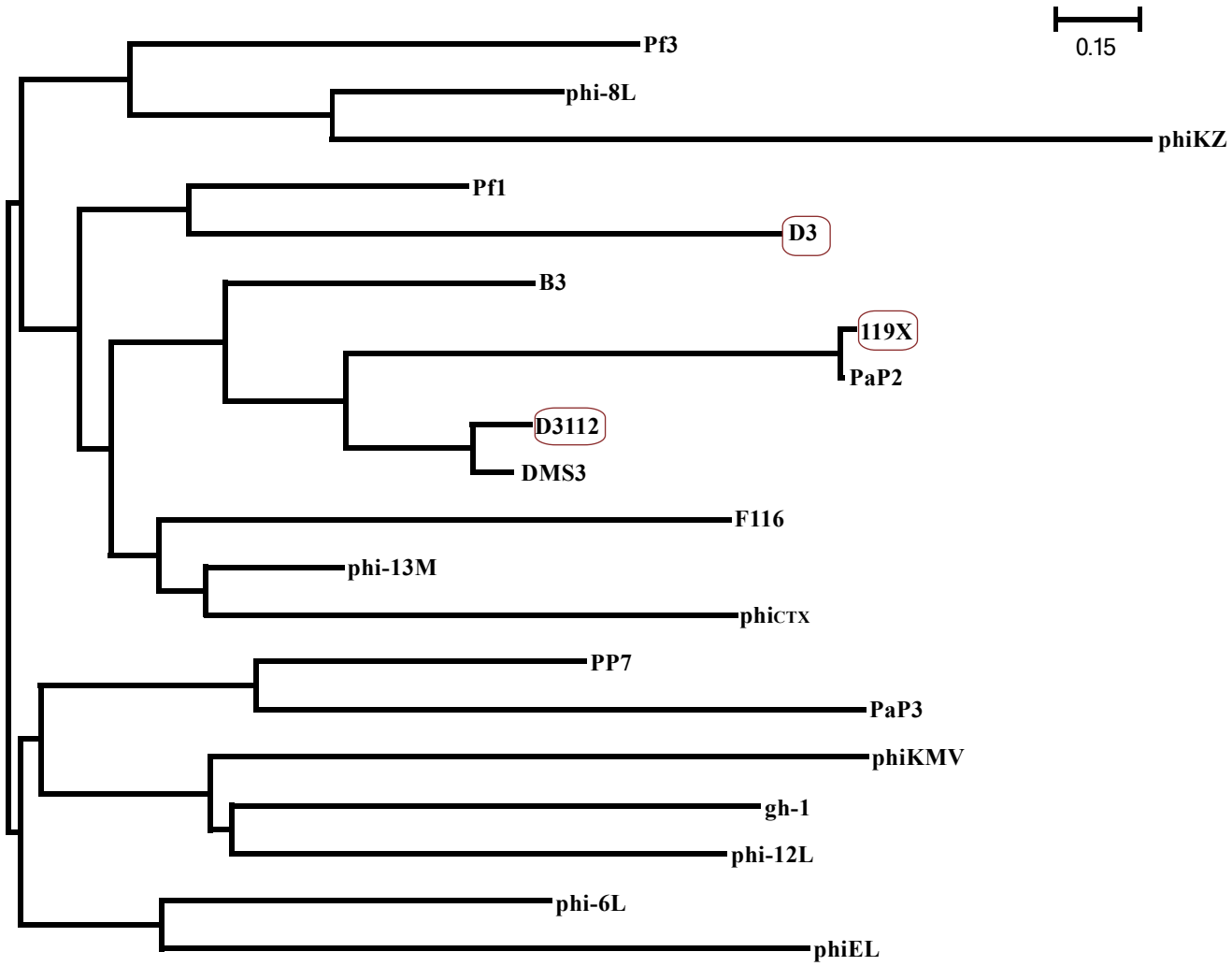

Figure 4: A phylogram representing the extent of divergence in synonymous codon usage of 20 phages of *Pseudomonas syringae*.

Conclusion

The study of Synonymous codon usage of 20 phages of *Pseudomonas syringae* has been done in our work. Out of these 20 phages, phage D3, D3112 and 119X codon usage, their correlation with expression levels and their translational selection clearly suggest that these may be useful for curing *Pseudomonas* infections.

Further, a comparative analysis of codon usage and RSCU value of the protein coding genes of these 20 phages was done. The cluster analysis (Figure 2 and Table 3) also indicates the similarity in the synonymous codon usage and the divergence among the phages D3, D3112 and 119X. The RSCU value suggest that D3 and D3112 have high GC content and the phage 119X though have high AT rich regions, but the codon variation is uniform in phage 119X (shown in Figure 1). It is observed that D3 carries most highly expressed genes compared with the other phages. Also the over-represented codons of D3 are preferentially reorganized compared with other phages. Based on this data (Figure 5 and Table 5), we suggest that the genes of D3 are expressed rapidly by host's translation machinery. Several lytic phages had been used successfully to cure infected

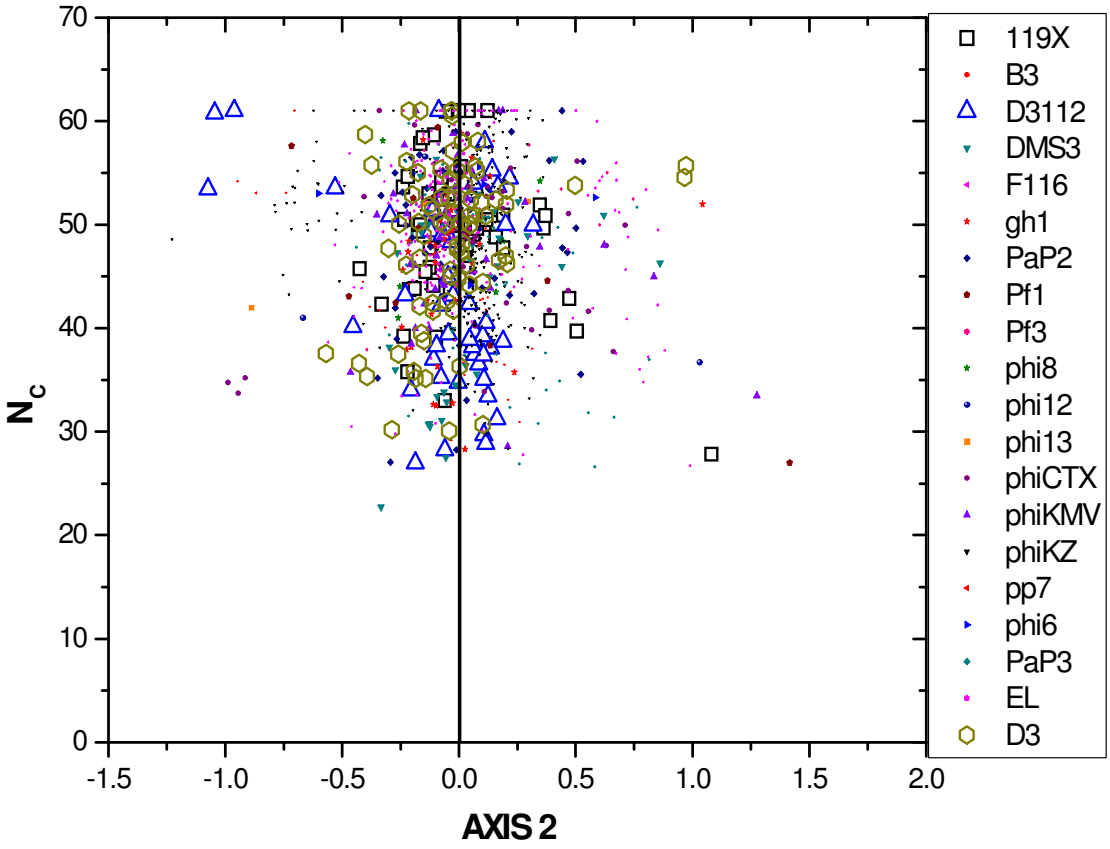

Figure 5: Plot between the genes of 20 *Pseudomonas* phages and N_c values.

patients (Sulakvelidze et al., 2001; Markoishvili et al., 2002; Jikia et al., 2005). It is speculated whether to use phages as a mixture of relatively weak and virulent phages or as one of a kind. A comparative analysis on codon usage of *E. coli* phages T1, T3, T4, T5 and T7 indicates that T4 carries the highest percentage of highly expressed genes among 5 phages and T4-like phages showed immense potential in therapy (Chibani-Chennoufi et al., 2004)

Our results interprets that the codon usage of different phages of *Pseudomonas* vary significantly which can be

AA	Codon	RSCU		tRNA Copy No. of *P.syringae*	AA	Codon	RSCU		tRNA Copy No. of *P.syringae*
		D3	D3112				D3	D3112	
Phe	UUU	0.00	0.00		Ser	UCU	0.57	0.86	
	UUC*	2.00	2.00	1		UCC*	2.29	0.43	1
Leu	UUA	0.00	0.41	1		UCA	0.00	2.57	1
	UUG	0.63	0.56	1		UCG*	0.57	0.86	1
	CUU	0.53	0.00		Pro	CCU	0.75	0.80	
	CUC	0.74	0.34	1		CCC*	1.00	0.20	1
	CUA	0.32	0.67	1		CCA	0.75	1.80	2
	CUG*	3.79	6.00	2		CCG	1.50	1.20	
Ile	AUU	0.52	0.75		Thr	ACU	0.17	1.50	
	AUC*	2.40	2.25	5		ACC*	2.26	1.50	1
	AUA	0.08	0.00			ACA	0.35	0.50	1
Met	AUG*	1.00	1.00	5		ACG*	1.22	0.50	
Val	GUU	0.57	0.00		Ala	GCU	0.68	1.85	
	GUC*	1.52	2.00	2		GCC	2.15	0.62	2
	GUA	0.76	1.23	2		GCA*	0.79	0.62	5
	GUG	1.14	2.00			GCG	0.38	0.92	
Tyr	UAU	0.44	0.00		Cys	UGU	0.00	1.20	
	UAC*	1.56	2.00	1		UGC*	2.00	0.80	1
TER	UAA	0.75	3.00		TER	UGA	2.25	1.50	
	UAG	0.00	0.00		Trp	UGG	1.00	1.00	1
His	CAU	0.25	0.00		Arg	CGU*	0.69	0.41	2
	CAC*	1.75	2.00	1		CGC	3.93	1.03	
Gln	CAA	0.48	0.84	1		CGA	0.49	2.07	
	CAG*	1.52	2.00			CGG	0.69	1.03	2

Table 5: Correlation between RSCU values phage D3 and phage D3112 and *P.syringae* specific tRNA copy number anticodons.

(*) represents putative optimal codons of phage D3.

suggested due to the occurrence of several factors such as mutational biasing, translational selection etc.

On the basis of our data we suggest that out of twenty phages of *Pseudomonas* phage D3 will be best suited for the treatment of *Pseudomonas* infections. Thus the phage D3 may be recommended to the phage therapy for the treatment of halo blight disease of *Pseudomonas syringae*.

Acknowledgement

The authors are thankful to the Sub-Distributed Information Center (BTISnet SubDIC), Department of Biotechnology (No. BT/BI/04/065/04), New Delhi, India for financial support.

References

1. Akashi H (1994) Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. Genetics 136: 927-935. »CrossRef »Pubmed »Google Scholar
2. Akashi H (1997) Codon bias evolution in Drosophila Population genetics of mutation-selection drift. Gene 205: 269-278. »CrossRef »Pubmed »Google Scholar
3. Bennetzen JL, Hall BD (1982) Codon selection in yeast J Biol Chem 257: 3026-3031. »CrossRef »Pubmed »Google Scholar
4. Bulmer M (1988) Codon usage and intragenic position. J Theor Biol 133: 67-71. »Pubmed »Google Scholar
5. Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage Genetics 129: 897-907. »CrossRef »Pubmed »Google Scholar
6. Burkholder WH (1926) A new bacterial disease of the bean. Phytopathology 16: 915-928.
7. Burkholder WH (1930) The bacterial diseases of the bean: a comparative study. Memoirs, Cornell University Agricultural Experiment Station 127: 1-88.
8. Carulli JP, Krane DE, Hartl DL, Ochman H (1993) Compositional heterogeneity and patterns of molecular evolution in the Drosophila genome. Genetics 134: 837-845. »CrossRef »Pubmed »Google Scholar
9. Chibani CS, Sidoti J, Bruttin A, Kutter E, Sarker S, et al. (2004) In vitro and in vivo bacteriolytic activities of *E. coli* phages: implications for phage therapy. Antimicrob. Agents Chemother 48: 2558-2569. »CrossRef »Pubmed »Google Scholar
10. Dong H, Nilsson L, Kurland CG (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J Mol Biol 260: 649-663. »CrossRef »Pubmed »Google Scholar
11. Gouy M, Gautier C (1982) Codon usage in bacteria: correlation with gene expressivity Nucleic Acids Res 10: 7055-7074. »CrossRef »Pubmed »Google Scholar
12. Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res 9: 43-74. »CrossRef »Pubmed »Google Scholar
13. Greenacre MJ (1984) Theory and applications of correspondence analysis. Academic Press London.
14. Gutiérrez G, Márquez L, Marín A (1996) Preference for guanosine at first codon position in highly expressed *Escherichia coli* genes: A relationship with translational efficiency. Nucleic Acids Res 24: 2525-2527. »CrossRef »Pubmed »Google Scholar
15. 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. »CrossRef »Pubmed »Google Scholar
16. Ikemura T (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein 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: 573-597. »CrossRef »Pubmed
17. Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2: 13-34. »CrossRef »Pubmed »Google Scholar
18. Jikia D, Chkhaidze N, Imedashvili E, Mgaloblishvili I, Tsitlandze G, et al. (2005) The use of novel biodegradable preparation capable of the sustained release of bacteriophages and ciprofloxacin, in the complex treatment of multidrug-re-sistant *Staphylococcus aureus*-infected local radiation injuries caused by exposure to Sr90. Clin Exp Dermatol 30: 23-26. »CrossRef »Pubmed »Google Scholar
19. Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T (2001) Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG dinucleotide usages assessed by multivariate analysis. J Mol Evol 53: 290-298. »CrossRef »Pubmed »Google Scholar
20. Kanaya S, Yamada Y, Kudo Y, Ikemura T (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of

codon usage based on multivariate analysis. Gene 238: 143-155. »CrossRef » Pubmed » Google Scholar

21. Kliman RM, Hey J (2003) Hill-Robertson interference in Drosophila melanogaster, Genet Res 81: 89-90. »CrossRef » Pubmed » Google Scholar
22. Krieg, Noel R, Bergey's (1984) Manual of Systematic Bacteriology, Volume 1. Baltimore: Williams & Wilkins. ISBN 0683041088.
23. Levin DB, Whittome B (2000) Codon usage in nucleopolyhedroviruses. J Gen Virol 81: 2313-2325. »CrossRef » Pubmed » Google Scholar
24. Markoishvili K, Tsitlanadze G, Katsarava R, Morris JG Jr, Sulakvelidze A (2002) A novel sustained-released matrix based on biodegradable polyester amides and impregnated with bacteriophages and an antibiotic shows promise in management of infected venous stasis ulcers and other poorly healing wounds. Int J Dermatol 41: 453-458.» CrossRef » Pubmed » Google Scholar
25. Moriyama EN, Gojobori T (1992) Rates of synonymous substitution and base composition of nuclear genes in Drosophila. Genetics 130: 855-864.» CrossRef » Pubmed » Google Scholar
26. Moriyama EN, Powell JR (1997) Synonymous substitution rates in Drosophila: mitochondrial versus nuclear genes. J Mol Evol 45: 378-391. »CrossRef » Pubmed » Google Scholar
27. 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.» CrossRef » Pubmed » Google Scholar
28. Powell JR, Moriyama EN (1997) Evolution of codon usage bias in Drosophila Proc. Natl Acad Sci USA 94: 7784-7790. »CrossRef » Pubmed » Google Scholar
29. Saettler AW (1991) Halo Blight P30 in Compendium of Bean Diseases R Hall ed APS Press St Paul.
30. Sharp PM, Cowe E (1991) Synonymous codon usage in Saccharomyces cerevisiae. Yeast 7: 657-678. »CrossRef » Pubmed » Google Scholar
31. Sharp PM, Li WH (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 24: 28-38. »CrossRef » Pubmed » Google Scholar
32. Sharp PM, Li WH (1987) The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research 15**:** 1281-1295. »CrossRef » Pubmed » Google Scholar
33. Sharp PM, Li WH (1989) On the rate of DNA sequence evolution in Drosophila J Mol Evol 28: 398-402. »CrossRef » Pubmed » Google Scholar
34. Sharp PM, Tuohy TMF, Mosurski KR (1986) Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Research 14: 5125-5143.» CrossRef » Pubmed » Google Scholar
35. Shepherd JCW (1981) Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc Natl Acad Sci USA 78: 1596-1600.» CrossRef » Pubmed » Google Scholar
36. Shields DC, Sharp PM, Higgins DG, Wright F (1988) 'Silent' sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol 5: 704-716. »CrossRef » Pubmed » Google Scholar
37. Sneath P (1973) Sokal Numerical Taxonomy W.H. Freeman and Company. San Francisco pp230-234.
38. Stenico M, Lloyd AT, Sharp PM (1994) Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. Nucleic Acids Res 22: 2437-2446.» CrossRef » Pubmed » Google Scholar
39. Stoletzki, Nina, Eyre W, Adam (2007) Synonymous Codon Usage in Escherichia coli: Selection for Translational Accuracy. Molecular Biology and Evolution 24: 374-381.» CrossRef » Pubmed » Google Scholar
40. Sueoka N, Kawanishi Y (2000) DNA, G+C content of the third codon position and codon usage biases of human genes. Gene 261: 53-62.» CrossRef » Pubmed » Google Scholar
41. Sulakvelidze A, Alavidze Z, Morris JG Jr (2001) Bacteriophage therapy. Antimicrob Agents Chemother 45: 649-659. »CrossRef » Pubmed » Google Scholar
42. Wright F (1990) The effective number of codons used in a gene. Gene 87: 23-29.» CrossRef » Pubmed » Google Scholar