------ Raf. J. Sci., Vol. 23, No.4 pp 10-22, 2012------

Codon Usage and Sequence Analysis of *Shigella flexneri* Genes

Mohammed Z. Saeed
Department of Biology
College of Education
University of Mosul
mzalkhat@yahoo.com
(Received 20 / 2 / 2012 ; Accepted 11 / 6 / 2012)

ABSTRACT

The nucleotide sequences of fifty genes of *Shigella flexneri* 5 strain 8401 were analyzed. The very lowly expressed genes (nine genes) have P_2 value range of (0.29-0.42) while moderately low expressed genes (thirteen genes) have P_2 range of (0.42-0.5). Highly expressed genes (sixteen genes) range between(0.50-0.63) and very highly expressed genes (twelve genes) range was (0.63-0.85). GC% content fluctuated from 44%-55% with a mean value of 52% indicating the nucleotides G and C slightly higher than A and T and it is a GC rich microorganism. The points of the ENc plot against GC_3 composition are quite spreaded which suggest that there are other contributors to the codon usage pattern besides the genomic composition. CAI values clearly parallel levels of gene expressions so highly expressed genes have higher CAI values.

Keywords: Effective number of codons (ENc), Codon adaptation index(CAI), Codon usage, P_2 index.

تحليل التسلسلات والشفرات الوراثية المستخدمة لمورثات جرثومة *Shigella flexneri*

الملخص

تم تحليل تسلسلات خمسين جينا من جينات جرثومة *Shigella flexneri* وبلغت مديات قيم دليل P_2 للجينات الواطئة جدا للتعبير الجيني (تسعة جينات) بين 0.29-0.42 وللواطئة التعبير (ثلاثة عشر جينا) بين 0.42-0.5 فيما كانت القيم للجينات عالية التعبير (ستة عشر جينا) بين 0.5-0.63 وللعالية جدا للتعبير الجيني (اثنا عشر جينا) بين 0.63-0.85. تراوح محتوى الكوانين والسايتوسين بين 44%-55% وبوسيط مقداره 52% والذي يشير الى محتوى عالٍ من هذه القواعد في هذا الكائن الحي. كانت قيم أعداد الشفرات المؤثرة في منحني الشفرات المؤثرة مع نكليوتيدات الكوانين والسايتوسين الكائنة في الموقع الثالث في الشفرات منتشرة وتؤكد وجود عوامل اخرى مساهمة في تحديد نمط الشفرات المستخدمة بالإضافة إلى المحتوى الوراثي للكائن وان قيم دليل تكيف الشفرات يوازي مستويات التعبير الجيني ولذا الجينات العالية التعبير لها قيم عالية منه.

الكلمات الدالة : العدد المؤثر للشفرات، دليل تكيف الشفرات، الشفرات المستخدمة، دليل P_2.

INTRODUCTION

Shigella species are gram negative facultative anaerobes and which recognized as causative agent of Bacillary dysentery in 1980s and subgrouped into 4 species : *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei*. Worldwide annual episodes due to *Shigella* infections are 160 millions of which 1.1 million deaths occur in children below 5 years of age (Ranade *et al*., 2009). The genes status and species classification appear no longer valid, as compelling evidence indicates that *Shigella* as well as enteroinvasive *Escherchia coli*, are derived from multiple origins of *E. coli* and forms a single pathovar.The *Shigella* chromosomes shares most of their genes with that of *E. coli* (Yang *et al*., 2005). Wei *et al*. (2003) determined the complete genome sequence of *Shigella flexneri* serotype 2a strain 2457T. The genome exhibits the backbone and island mosaic structure of *E. coli* pathogens, albeit with much less horizontally transferred DNA and lacking 357 genes present in *E. coli*. The strain is distinctive in its large complement of insertion sequences, cryptic prophages, 372 pseudogenes and 195 *S. flexneri*–specific genes. The 2457T genome was compared with that of *S. flexneri* 2a strain 301, and the data is consistent with *Shigella* being phylogenetically indistinguishable from *E. coli*.

Analysis of codon usage data has both practical and theoretical importance in understanding the basics of molecular biology and evolution. It is well known that synonymous codon usage bias is non random and species specific. Moreover, codon usage pattern differ significantly among different genes within the same taxa (Gupta *et al*., 2004). Different factors have been proposed to explain the preferential usage of a subset of a synonymous codons, including biased mutation pressure toward G/C or T/A and natural selection for optimizing translational selection (Sharp *et al*., 2005). Although the genomic-wide mutational bias should act on the entire genome, the extent is stronger for the third positions of codons since the first two positions of codons are constrained by protein-coding requirements (Muto *et al*.,1987) thus, the mutational bias could be the cause of the preferential usage of G/C or A/T ending codons. The translation selection should act mainly as highly expressed genes and should be the cause of preferentially optimal codons, which are best recognized by the most abundant tRNA species in the cell (Ikemura, 1985 ; Kanaya *et al*.,1999).

MATERIALS AND METHODS

The nucleotide sequences of fifty genes of *Shigella flexneri* 5 strain 8401 were downloaded from Gene bank. Gene length, Relative synonymous codon usage (RSCU), Base composition, Codon adaptation index (CAI), Effective number of codons (ENc), GC_3 content were calculated for each gene using codon W program available at :
http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::codonw
and CAI calculator available at http://genomes.urv.cat/CAIcal/
and genomatrix software suite available at :
http://www.genomatix.de/cgi-bin/tools/tools.pl

The relative synonymous codon usage (RSCU) is defined as the ratio of the observed frequency of codons to the expected frequency if all synonymous codons were used equally for a given amino acid so detect codon usage variation:

$$RSCUij= (obsij/\sum^{ni} j=1\ obsij)/(1/ni)$$

Where obsij is the observed number o codon j for i amino acid, which is encoded by ni synonymous codons in specific gene sample (Gu *et al*., 2004).

The codon adapation index (CAI) is calculated according to Sharp and Li (1987). It quantifies the relative adaptiveness of a gene's codon usage that is the quantity of usage of each codon to that most plentiful codon within synonymous family. The CAI value varies from 0.1-1. The effective number of codons (ENc) is used to measure codon bias (measure the real number of used codon), so ENc values ranges from 20 (when only one codon is used per amino acid to 61 (when all synonymous codons used equally for each amino acid. The expected ENc value under random codon usage can be calculated for any values of GC_3 as $ENc=2+S+29/(S^2+(1-S)^2)$ where S represent the given GC_3 value (Zhang *et al*., 2011). P_2 index (use of intermediate energy codon)was calculated according to Grosjean and Fiers (1982):

$$P_2 =(WWC+SSU)/(WWY+SSY),$$

where W=A or U, S=G or C, Y=C or U and for example WWC is the observed number of codons of that description. P_2 value of 0.5 indicate no bias. Lowly expressed genes have values less than 0.5. P_2 value gives a measure of translational pressure and correlated to high gene expression.

RESULTS AND DISCUSSION

For the fifty genes (Table 1), the lowly expressed genes had P_2 values less than 0.50, ranged between 0.29-0.49 and the very lowly expressed genes had a mean P_2 value of 0.37, ranged between 0.29-0.42 while moderately low expressed genes had a mean P_2 value of 0.46, ranged between 0.42-0.5. Highly expressed genes on the other hand had values greater than 0.50, ranged between 0.50-0.85 so highly expressed genes had a mean value of 0.55, ranged between 0.50-0.63 and very highly expressed genes of a mean P_2 value of 0.68, ranged between 0.63-0.85. Accordingly, the fifty genes of *S. flexneri* can be categorized to: Very low expressed genes (nine genes): *cheY, ispF, apbA, fucA, dsbE, rfbB, panD, coaD, fimH* : moderately lowly expressed genes (thirteen genes): *trpR, ndx, rfbA, ubiC, arcB, holB, asd, rpiA, pdxH, mepA, ccd, serA, aphA* :Highly expressed genes (sixteen genes):*dsbA, cybC, adk, purR, zipA, gyrB, dnaJ, mltB, gadB, rseB, rfbD, gnd, galF, narH, pldA , glmS,*Very highly expressed (twelve genes) *fba, fabI, ompF, rplY, groEL, rpsL, surA, rpmh, rplx, rpsF, rplN, rpmA.*

Seven of these highly expressed genes were ribosomal proteins encoding genes reflecting the high demand for their products for the microorganism to build cell components and metabolic activity. Sharp and Li (1986) studied 165 *E. coli* genes which were categorized as twenty seven very highly expressed genes, fifteen highly expressed genes, fifty seven moderately low expressed and fifty eight very lowly expressed genes and the remaining eight

regulatory/repressor genes which were reported to have an extraordinary high frequency of rare codons. They identified a clear and general trend in codon usage bias, from the very high bias seen in very highly expressed genes attributed to selection to a rather low bias in other genes which seems to be influenced by mutation rather than selection. Values of P_2 index for the highly expressed genes ranged between 0.52 and 0.82 while lowly expressed genes ranged between 0.3-0.76.

Sharp and Shields (1987) studied fifty six genes of *B. subtilis* genes and showed that the synonymous codon usage is less biased than those of *E. coli*. In *E. coli*, highly expressed genes had P_2 values of 0.7-0.9 indicating a strong preference, while other genes had values close to 0.5 indicating little preference. In *B. subtilis* few genes had P_2 values greater than 0.5 but in a comparison of the highly expressed genes, P_2 values from *B. subtilis* were lower than those from *E. coli*.

Table 1: Genes of *S. flexneri* .

Gene symbol	Gene nomenclature	length	P_2 index
dsbE	Disulfide oxidoreductase	186	0.29
ispF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	160	0.34
apbA	2-dehydropantoate 2-reductase	304	0.34
fimH	Fimh protein	301	0.34
panD	Aspartate alpha-decaroxylase	127	0.37
cheY	Chemotaxis requlatory protein	130	0.40
fucA	L-fuculose phosphate aldolase	216	0.40
rfbB	dTDP-glucose 4,6-dehydratase	362	0.41
coaD	Phosopantetheine adenylytransferase	160	0.41
pdxH	Pyridoxamine 5-phosphate oxidase	219	0.42
arcB	Aerobic respiration control sensor protein ArcB	779	0.44
rpiA	Ribose-5-phosphate isomeraseA	220	0.44
aphA	Acid phosphatase/phosotransferase	318	0.45
asd	Aspirate-semialdehyde dehdrogenase	368	0.45
trpR	Trp operon repressor	109	0.46
ndk	Nucleoside diphosphate kinase	144	0.46
serA	D-3-phosphglycerate dehydrogenase	411	0.47
ubiC	Chorismate pyruvate lyase	203	0.47
ccd	Cytidine deminase	295	0.48
holB	DNA polymerase III subunit delta	335	0.48
rfbA	Glucose-1-phosphate thymidyly transferase	293	0.48
mepA	Penicillin-insensitive murein endopeptidase	275	0.49

Table 1: Continued

Gene symbol	Gene nomenclature	length	P_2 index
galF	UTP-glucose-1-phosphate uridylyltransferase subunit galF	298	0.5
mltB	Murein hydrolaseB	382	0.51
rseB	Periplasmic negative regulator of sigmaE	319	0.52
pldA	phosolipaseA	290	0.52
purR	DNA-binding transcriptional repressor	342	0.53
dsbA	Periplasmic protein disulfide isomease	209	0.53
dnaJ	Heat shock protein	377	0.55
gadB	Glutamate decarboxylase isozyme	467	0.55
rfbD	dTDP-4-dehydrorhamnose reductase	300	0.55
cybC	Cytochrome b (562)	129	0.58
zipA	Cell division protein	342	0.58
gyrB	DNA gyrase subunit B	805	0.58
gnd	6-phosogluconate dehydrogenase	469	0.58
narH	Nitrate reductase 1 subunit beta	513	0.59
glmS	Glucosamine-fructose-6-phosphate minotransferase	610	0.6
adk	Adenylate kinase	215	0.62
surA	Peptidyl-prolyl cis-transisomerase SurA	429	0.63
fabI	Enoyl-(acyl carrier protein) reductase	263	0.64
rplX	50S ribosomal protein L24	105	0.66
rplN	50S ribosomal protein L14	124	0.68
rpmH	50S ribosomal protein L34	86	0.68
rpsF	30S ribosomal protein 56	131	0.73
rpmA	50S ribosomal protein L27	47	0.74
ompF	Outer membrane proteinF	363	0.77
groEL	Chaperonin GroEL	549	0.77
rpsL	30S ribosomal protein S12	125	0.81
rplY	50S ribosomal protein L25	95	0.83
fba	Fructose-bisphosohate aldolase	360	0.85

The codons in the fifty genes were analyzed using codon W and Table 2 shows the results. The codons whom their frequencies lesser than 10% are (CUA, AUA, CCC, UCA, AGU, ACA, AGG, AGA, CGA, CGG, UGU, UGC, GGA, UUA, CCU, CCA, ACG, GGG) which represent rare codons. There is a supporting evidence that the higher frequency of rare codons reflects mutation biases rather than positive selection for rare codons (Sharp and Cowe,1991).

Luo *et al*.(2011) analyzed the synonymous codon usage in *S. flexneri* 2a strain 301(sf301) and performed a comparative analysis of synonymous codon usage patterns in sf301 and other strains of *Shigella* and *E. coli*. Although the a significant variation in codon usage bias among different sf301 genes, there was a slight but observable codon usage bias that could primarily be attributable to mutational pressure and translational selection. By comparing the relative synonymous codon usage values across different *Shigella* and *E. coli* strains, they suggested that synonymous codon usage pattern in *Shigella* genomes was strain specific. Synonymous codons are used with different frequencies both among species and among genes within the same genome. Highly expressed genes such as those encoding translation elongation factors and ribosomal proteins tend to use optimal (preferred) codons and exhibit very high levels of bias. These patterns have been interpreted as natural selection for more efficient and accurate translation in contrast to some other studies that have demonstrated that the first factor shaping codon usage is the nucleotide composition (GC content) (Rao *et al*., 2011).

Table 2: Codon usage of the *S. flexneri* Genes

	N RSCU		N RSCU		N RSCU		N RSCU
Phe	U UU 223 0.90	Ser	UCU 180 1.40	Tyr	UAU 229 1.01	Cys	UGU 61 0.84
	UUC 270 1.10		UCC 146 1.14		UAC 226 0.99	TER	UGC 85 1.16
Leu	U UA 136 0.59		UCA 74 0.58	TER	UAA 33 1.98	Trp	UGA 14 0.84
	UUG 150 0.65		UCG 100 0.78		UAG 3 0.18	Arg	UGG 173 1.00
	CUU 152 0.66	Pro	CCU 102 0.61	His	CAU 167 1.01		CGU 411 3.13
	C UC 111 0.48		CCC 45 0.27		CAC 164 0.99		CGC 305 2.32
	CUA 36 0.16		CCA 109 0.65	Gln	CAA 174 0.55		CGA 31 0.24
	CUG 790 3.45		CCG 411 2.46		CAG 463 1.45	Ser	CGG 23 0.18
Ile	AUU 371 1.35	Thr	ACU 162 0.86	Asn	AAU 214 0.69		AGU 73 0.57
	AUC 429 1.56		ACC 370 1.97		AAC 410 1.31	Arg	AGC 197 1.54
	AUA 27 0.10		ACA 76 0.40	Lys	AAA 620 1.53		AGA 12 0.09
Met	AUG 396 1.00		ACG 144 0.77		AAG 189 0.47	Gly	AGG 6 0.05
Val	GUU 347 1.33	Ala	GCU 314 0.90	Asp	GAU 485 1.17		GGU 487 1.70
	GUC 183 0.70		GCC 287 0.82		GAC 346 0.83		GGC 478 1.67
	GUA 175 0.67		GCA 303 0.87	Glu	GAA 685 1.45		GGA 69 0.24
	GUG 337 1.29		GCG 495 1.42		GAG 262 0.55		GGG 114 0.40

*N is the number of the codons
**RSCU is the relative synonymous codon usage

The values of nucleotide contents in the fifty genes were analyzed (Table 3), evidently GC% content fluctuated from 44%-55% with a mean value of 52% and standard of deviation (SD) of 2.81 indicating the nucleotides G and C slightly higher than A and T and it is a GC rich microorganism. The values of A,G,C,T and GC were compared with A_3, G_3, C_3, T_3 (contents of third position of codon) and GC_3 and Table 4 shows the correlation among them. There is a strong correlation between GC and G_3,C_3 and GC_3 because the organism is GC rich. Zhang *et al*. (2011a) suggested that the nucleotide constraint could possibly influence synonymous codon usage. The base composition is the most frequently reported DNA feature and is one of the most pervasive influences on codon usage. Peden (1999) indicated that a base composition is a balance between mutational pressure towards or away from GC nucleotides, either these compositional constraints are the result of mutational biases or natural selection and plays a major role to preferential fixation of non-random dinucleotides and base frequencies.

GC_3 was calculated for the genes with an average of 53, ranged between 35%-64%, and a standard deviation (SD) of 6.133 while ENc was with an average 43.4 the range was between 28-57 and SD of 7.4. It was reported that a plot of ENc against GC_3 can be effectively used to explore this heterogeneity. The ENc plots of the genes ,whom codon choice is constrained only by GC_3 composition, will lie on or just below the curve of the predicted values , i.e. is principally influenced by the mutational bias only (Zhang *et al*., 2011b). Fig.1 shows the points of the ENc plot against GC_3 composition, the points are quite spread and the genes appear to follow a higher slope than that of the theoretical curve, and suggest that there are other contributors to the codon usage pattern besides the genomic composition. The genome-wide codon usage of each organism is set primarily by mutational forces (point mutation) which create a point about which the codon bias of individual genes is additionally perturbed from the genome-wide average codon bias by selective or other mutational forces acting during translation but this effect is relatively much smaller (Chen *et al*., 2004).

Table 3: Base content (%), Effective number of codons (ENc) and codon adaptation index (CAI)

Gene	A	G	C	T	A_3	G_3	C_3	T_3	GC	GC_3	ENc	CAI
fba	26	25	26	23	17	21	35	27	51	56	30	0.86
fabI	22	26	27	25	16	16	38	30	53	54	34	0.76
dsbA	26	28	20	26	19	31	19	31	48	50	37	0.74
cybC	31	26	24	19	25	27	26	22	50	53	40	0.58
ompF	27	25	23	26	15	13	36	36	48	49	37	0.77
rplY	31	25	25	20	22	20	35	23	50	55	29	0.78
cheY	27	30	19	24	22	31	22	25	49	53	53	0.48
ispF	23	30	24	23	19	23	27	31	54	50	49	0.49
trpR	25	30	24	21	21	35	23	21	54	58	49	0.43
groEL	26	30	24	21	23	18	32	27	53	50	28	0.87
adk	26	28	24	22	19	23	29	29	52	52	33	0.79
ndk	23	26	27	25	17	17	34	32	53	51	33	0.79
purR	24	29	25	22	19	26	31	26	54	57	44	0.63
apbA	26	26	26	25	21	23	30	26	51	53	47	0.45
fucA	26	26	23	25	19	27	20	34	49	47	57	0.44
zipA	24	27	28	21	19	31	21	29	54	52	49	0.53
gyrB	25	28	26	22	14	29	35	22	55	64	35	0.73
rpsL	24	26	27	20	18	20	29	33	53	49	37	0.74
dnaJ	24	31	25	22	17	24	32	27	55	56	42	0.66
mltB	23	28	27	23	15	31	32	22	55	63	48	0.54
gadB	23	27	27	23	15	23	37	25	54	60	41	0.67
dsbE	23	30	25	28	14	33	27	26	54	60	42	0.52
rfbA	28	26	18	26	23	22	13	42	44	35	43	0.5
rseB	24	25	26	24	19	22	30	29	51	52	54	0.43
rfbD	27	25	25	28	18	26	25	31	49	51	54	0.46
rfbB	28	25	19	25	21	20	17	42	44	37	50	0.48
gnd	25	26	24	20	15	23	30	32	51	53	41	0.68
surA	26	28	26	27	15	28	33	24	54	61	42	0.67
ubaC	23	26	24	27	28	29	20	23	50	49	51	0.33
galB	26	27	25	23	20	26	29	25	52	55	50	0.54

Table 3:Continued

Gene	A	G	C	T	A_3	G_3	C_3	T_3	GC	GC_3	ENc	CAI
arcB	26	28	23	23	19	31	25	25	51	56	45	0.53
rpmA	28	27	23	23	14	16	20	50	50	36	39	0.78
panD	24	26	27	23	15	14	41	30	53	55	48	0.62
narH	25	28	25	21	15	29	32	24	54	61	35	0.71
pldA	25	27	24	24	15	27	32	26	51	59	48	0.54
holB	22	27	27	24	22	26	24	28	54	50	54	0.42
rplX	31	25	21	23	22	18	32	28	46	50	42	0.7
coaD	23	29	24	24	14	37	21	28	53	58	39	0.56
rpsF	26	27	26	21	15	19	35	31	53	54	32	0.79
asd	22	28	26	24	15	24	32	29	54	56	49	0.52
glmS	23	28	26	24	15	28	28	29	54	56	37	0.71
rpiA	23	29	24	23	19	20	33	28	53	53	42	0.62
pdxH	23	26	25	25	17	23	29	31	52	52	43	0.53
mepA	23	27	28	22	18	28	28	26	55	56	50	0.47
rplN	25	28	23	24	15	23	31	31	51	54	48	0.63
fimH	23	27	23	27	17	25	24	34	50	49	52	0.44
rpmH	21	24	28	27	24	10	33	33	52	43	43	0.75
ccd	22	28	26	24	15	30	25	30	55	55	47	0.54
serA	25	28	24	23	17	30	27	26	53	57	43	0.61
aphA	29	21	23	26	23	18	26	33	45	44	55	0.43

	A_3	G_3	C_3	T_3	GC_3
A	0.345821	-.14709	-0.16904	0.124667	-0.30301
G	-0.27718	0.541457	-0.05876	-0.33919	0.460695
C	-0.30135	-0.07672	0.583724	-0.36439	0.487515
T	0.020063	-0.04615	-0.23481	0.287127	-0.27138
GC	-0.6787	0.565805	0.789445	-0.32679	0.710292

Table 4: Correlation among the bases

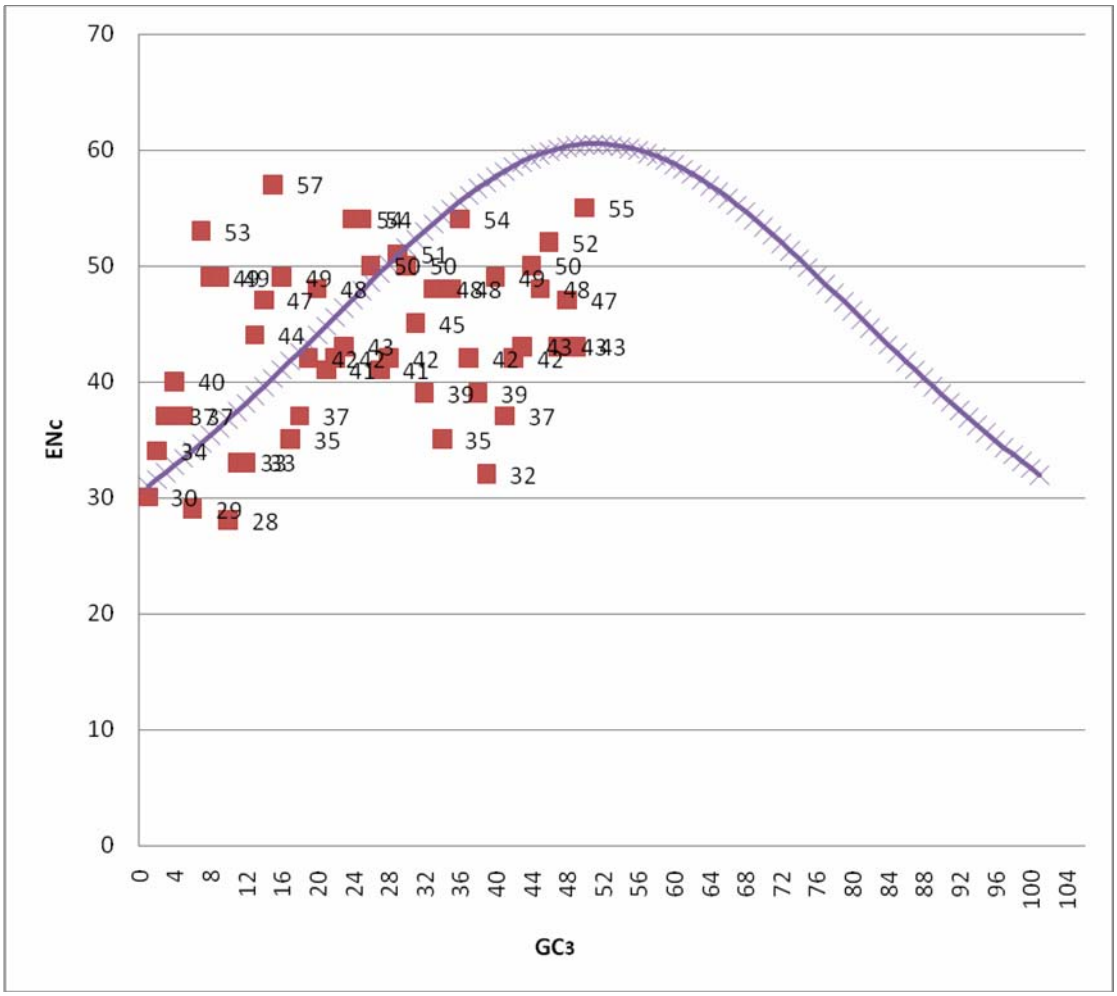

Fig. 1: ENc plot against GC_3

A simple effective measure of synonymous codon usage bias ,is the codon adaptation index which uses a reference set of highly expressed genes from a species to assess the relative merits of each codon ,and a score for a gene is calculated from the frequency of use of all codons in that gene. This index assesses the extent to which selection has been effective in molding the pattern of codon usage. It is useful for predicting the level of expression of a gene and CAI values clearly parallel levels of gene expressions (Sharp and Li, 1987) so highly and very highly expressed genes (Table 3) have higher CAI values and measuring correlation between P_2 and CAI yield 0.757309 indicating the positive relation between gene expression and this index (Fig. 2,3). Codon adaptation index value has been proved to be the best gene expression theoretical value and been extensively used as a measure of gene expression level. All the data suggests that genes with the higher expression level, exhibiting greater degree of codon usage bias, higher P_2 and CAI values and the later has now been considered as an accepted measure of gene expression (Liu *et al*., 2010). It was reported the correlations of codon usage bias with gene expression level and GC content bias are not ubiquitous , thus

codon usage diversity within any genome could be the result of a balance among different evolutionary forces and their relative contributions vary among different genomes (Suzuki *et al*., 2009).

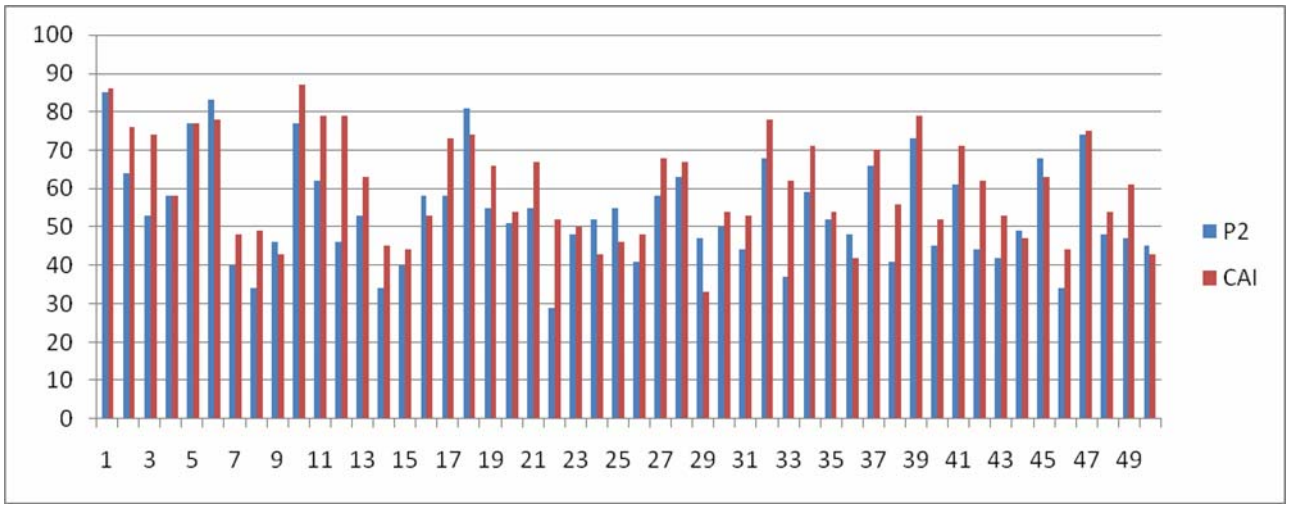

Fig. 2: A histogram showing P_2 index values with CAI.

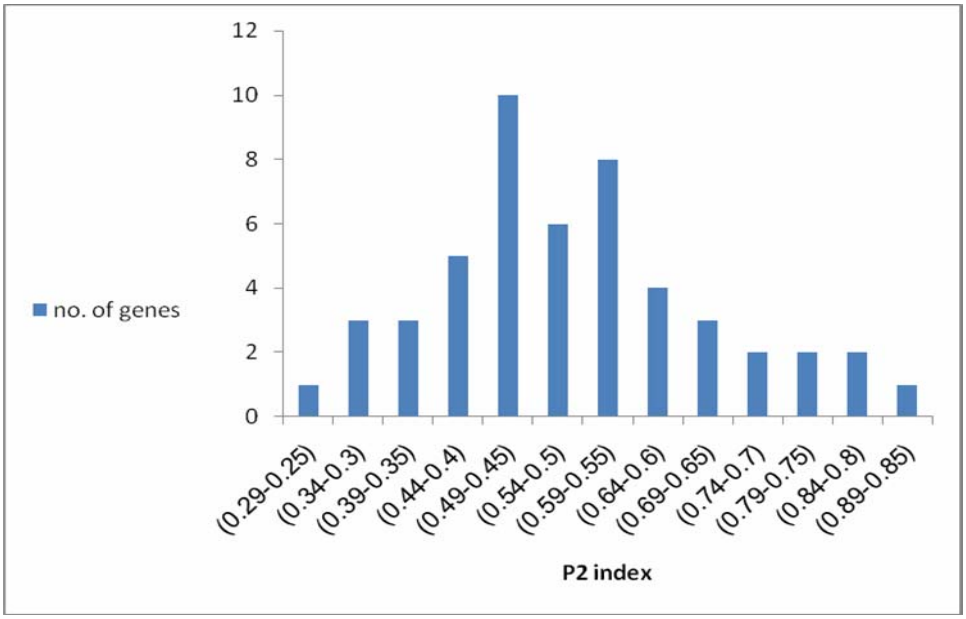

Fig. 3: A histogram showing P_2 index of the genes

REFERENCES

Grosjean, H.; Fiers,W.(1982).Preferential codon usage in prokaryotic genes ,the optimal codon–anticodon interaction energy and selective codon usage in effectively expressed genes.*Gene*,**18**:199-209.

Gu, W.; Zhou, T.; Ma, J.; Sun, X.; Lu, Z.(2004).The relationship between synonymous codon usage and protein structure in *E. coli* and *Homo Sapiens*. *Biosystems*, **37**, 89-97.

Gupta, S.K.; Bhatta, Ch.; Ghosh, T.C.(2004).Synonymous codon usage in Lactococcus lactis: mutational bias versus translational selection. *J. Biomolecular Str. Dynamics*.**21**(4),1-9.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol*., **2**(1), 13-34.

Kanaya, S. ;Yamada, Y.; Kudo, Y.; Ikemura, T. (1999). Studies of codon usage and tRNA genes of 15 unicellular organisms and quantification of *Bacillus subtilis* tRNA: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene*, **238**(1), 143-155.

Liu, H.; He, R.; Zhang, H.; Huang, Y.; Tian, M.; Zhang, J. (2010). Analysis of Synonymous codon usage in *Zea mays*. *Mol. Biol. Rep*., **37**,677-684.

Luo, L.; X.; Xu, J.G.; Ye, Ch., Y. (2011). Analysis of Synonymous codon usage in *Shigella flexneri* 2a strain and other *Shigella* and *E. coli* strains. *Candian J. Microbiol.*, **57**(11), 1016-1023.

Muto, A.; Osawa, S. (1987).The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl. Acad. Sci.USA*, **84** (1), 166-169.

Peden, J.F.(1999). Analysis of codon usage. Ph.D thesis, Department of Genetics ,University of Nottingham , Nottingham ,UK, pp. 11-12.

Ranade, S.H.; Hossani, A .; Deobagkar, O.N.; Deobagkar, D.D.; Chopade, B.A. (2009). The nucleotide sequences of *Shigella flexneri* 1A : A common Indian isolate. *Indian J. Clin. Biochem*., **24**(2),142-149.

Rao, Y.; Wu, G.; Wang, Z.; Ch., X.; Nie, Q.; Zhang, X. (2011). Mutation bias is the driving force of codon usage in the *Gallus gallus* genome. *DNA Res*., **18**, 499-512.

Sharp, P.M.; Bailes, E., Grocock, R.J.; Peden, J.F.; Sockett, R.E. (2005). Variation in the strength of selected codon bias among bacteria. *Nucleic Acids Res*., **33**(4), 1141-1153.

Sharp, P.M.; Cowe, E. (1991). Synonymous codon usage in *Saccharomyces cervisiae*. *Yeast*, **7**, 657-678.

Sharp, P.M.; Li, W-H. (1986). Codon usage in regulatory genes in *E. coli* does not reflect selection for rare codons. *Nucleic Acids Res*., **14**(19), 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**(3), 1281-1295.

Sharp, P.M.; Shields, D.C. (1987). Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Res*., **15**(19), 8023-8040.

Suzuki, H.; Saito, R.; Tomita, M. (2009). Measure of synonymous codon usage diversity among genes in bacteria. *BMC Bioinformatics*, **10**, 167.

Wei, J.; Goldberg, M. B.; Burland, V.; Venkoatesan, M. M.; Deng, W.; Fournier, G; Mayhew, F.; PlunkeltIII, G.; Rose, D. J.; Darling, A.; Mau, B.; Perna, N. T.; Payne, S. M. ; Runyen-Janecky, L.J.; Zhou, S. ;Schwartz, D.C. ; Blattner, F.R. (2003). Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 245T . *Infect. Immun*.,**71**(5), 2775-2786.

Yang, F.; Yang, J.; Zhang, X. ;Chen, l.; Jiang, Y.; Yan, Y.; Tang, X.; Wang, J.; Xiong, Zh., Dong, J.; Xue, Y. ; Zhu, Y.; Xu, X.; Sun, L.; Chen, S.; Peng, J.; Xu, J., Wang, Y.; Yuan, Zh.; Wen, Y.; Yao, Zh.; Qiang, B.; HOu, Y.; Yu, J.; Jin, Qi. (2005). Genome dynamics and diversity of Shigella species: The etiologic agents of bacillary dysentery. *Nucleic acids Res*., **33**(19), 6445-6458.

Zhang, J.; Wang, M.; Liu, W.; Zhou, J.; Chen, H.; Ma, L.; Ding, Y.; Gu, Y.; Liu, Y. (2011a). Analysis of codon usage and nucleotide composition bias in polioviruses. *Virology J*., **8**, 146.

Zhang, Y.; Liu, Y; Liu, W.; Zhou, J.; Ch., H; Wang, Y.; Ma, L. (2011b). Analysis of Synonymous codon usage in hepatitis A virus. *Virology J*., **8**, 174-182.

Zhou, T.; Sun, X.; Lu, Z.(2006). Synonymous codon usage in environmental chlamydia UWE25 reflects on evolutional divergence from pathogenic chlomydiae, *Gena*, 368,117-125.