International Journal of Advanced Research in Biological Sciences
ISSN : 2348-8069
www.ijarbs.com

Research Article

Study of phoBR in *Escherichia coli* and phoPR in *Bacillus*: Computational Approach

Sunil Kanti Mondal [1,2*]

[1]Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta, 92, APC Road Kolkata-700009.
[2]Department of Biotechnology, The University of Burdwan, Rajbati, Burdwan-713104, West Bengal, India.
*Corresponding author

Abstract

One of the major essential macronutrients for biological growth and development is phosphorus which is required for energy metabolism, synthesis of important biological molecules such as phospholipids and nucleic acids. Phosphorus is present in environment in mostly insoluble forms. Uptake of phosphate by bacteria most commonly occurs via two systems, the low-affinity, constitutively expressed Pit system and the high-affinity, phosphate starvation induced Pst system. The genes encoding high-affinity Pi transport systems, such as the pst genes, or for transport systems for alternative phosphorus sources, such as the phn genes, are generally induced when the organism encounters Pi-limited conditions. This induction is mediated by the Pi-responsive two-component regulatory system, phoBR in *Escherichia coli* or phoPR in *Bacillus*. In both cases, phoR is a membrane-bound sensor and phoP or phoB are the response regulators. The compositional analysis shows significant difference between phoP from *Escherichia coli* and phoB from *Bacillus* and between phoR from both of them. Though the gene set phoBR is GC rich but I am getting some exceptional situations where the RSCU values of few /all synonymous codon for some amino acids are showing reverse result in reference to our expectation in *Escherichia coli*. In phoPR of *Bacillus* positive substitution takes place where as phoBR of *Escherichia coli* are static. Though the codon usage patterns of the genes are different but all are under mutational pressure. Similar and high values of CAI, reflects the co-expression of phoBR in *Escherichia coli* and phoPR in *Bacillus* and also their importance for survivability.

Keywords: pst: phosphate specific transporter; phn: Phosphonate transporter; phosphate inorganic transport; RSCU, CAI.

Introduction

Phosphorus is an essential nutrient for all cells and is required for energy metabolism and for the synthesis of important biological molecules such as phospholipids and nucleic acids. Living systems satisfy their need for phosphorus (P) primarily through Pi, in which P is in its highest (+5) oxidation state. The main source of phosphorus for bacteria is inorganic Phosphate.[1] Phosphorus is present at levels of 400–1200 mg/kg of soil [2, 3] while its concentration in soluble form is very low, normally at levels of 1 ppm or less then 1ppm. Phosphate solubilising bacteria (PSB) can convert insoluble phosphates into soluble forms and may accumulate excess P_i in the form of polyphosphate (polyP) which is a linear polymer of P_i with a chain length of up to 1000 residues or more. PolyP can serve as a P source for the biosynthesis of nucleic acids and phospholipids under P_i starvation conditions. PolyP is likely to function as a P_i reservoir with osmotic advantages [4, 5, 6].

Forms Of Phosphorus In Soil

A large proportion of phosphorus is present in insoluble forms. Organic matter is an important reservoir of immobilized phosphorus that accounts for 20–80% of soil P [7, 8]. Inositol phosphate (soil phytate)

is the largest and most stable form of organic phosphorus. It is synthesized by microorganisms and plants and accounts for up to 50% of the total organic phosphorus. Other organic phosphorus compounds in soil are phosphomonoesters, phosphodiesters (phospholipids, nucleic acids) and phosphotriesters [7, 9]. Mineral forms of phosphorus which are part of stratum rock present in soil are apatite, hydroxyapatite and oxyapatite. They are the biggest reservoirs of this element in soil and under appropriate conditions can be solubilised by bacteria.

Organic Phosphate Solubilization

The microbial mineralization of organic phosphorus is strongly influenced by the physicochemical and biochemical properties of molecules, environmental parameters and moderate alkalinity [10]. The variety of enzymes helping in this activity is Nonspecific phosphatases, Phytases, Phosphonatases and C–P Lyases.

Phosphate Transport Systems In Bacteria

Uptake of phosphate by bacteria most commonly occurs via two systems, the low-affinity, constitutively expressed Pit system and the high-affinity, phosphate starvation induced Pst system [11, 12]. Pit (phosphate inorganic transport) system consists of a single membrane protein, encoded by pitA or pitB, and is energized by the proton motive force. They recognize metal phosphate as their substrate [12, 13]. Pst (phosphate specific transport) system are multi subunit ABC transporters encoded by a four-gene operon pstSCAB. pstS is a substrate binding protein (pstS promoter is expressed at a low basal level when Pi is present in excess and it shows 100-fold derepression during Pi-limited growth). This shows that Pst system is induced under low Pi concentration (< 4 x 10^{-6} M). pstC and pstA are transmembrane proteins while pstB is ATPase. Pst recognizes phosphate as its substrate [11, 12, 14 - 17].

Bacteria are also able to utilize phosphonates or phosphites as alternative sources of phosphorus which are compounds that contain carbon phosphorus bond. These compounds are taken up into the cells by ABC type phn system. This system has three components which are phnC (ATPase), phnD (substrate-binding protein), and phnE (permease). Pi is a non-specific substrate for this system [18]. A third high affinity ABC type transport system is phoCDET which can uptake both Phosphate and phosphonates. phoC and phoD are homologous to phnC and phnD, respectively. phoE and phoT both encode transmembrane proteins homologous to phnE [9, 10]. In a study it has been found that, the direct oxidation causes phosphate solubilisation in Gram-negative bacteria via diffusion of the strong organic acids produced in the periplasm into the adjacent environment [19]. The genes encoding high-affinity Pi transport systems, such as the pst genes, or for transport systems for alternative phosphorus sources, such as the phn genes, are generally induced when the organism encounters Pi-limited conditions [15, 16, 20 - 22]. This induction is mediated by the Pi-responsive two-component regulatory system, phoBR in *Escherichia coli* or phoPR in *Bacillus*. In both cases, phoR is a membrane-bound sensor and phoP or phoB are the response regulators. Repression of the Pho response requires pstS [16, 20, 23 - 25]. phnF is the putative transcriptional regulator of phnDCE operon [18, 26]. Once inside the cell, Phosphate is captured in several metabolic reactions which require energy and produce organophosphate esters [27, 28]. The ugpBAEC transporter apparently has a role in the entry of Pi esters. glpT and uhpT transport glycerol 3-phosphate and hexose 6-phosphates respectively [28].

In this work I want to see the compositional similarity in the gene and amino acid level, codon usage pattern, expression probability of phoR from *Escherichia coli* and *Bacillus* as in both the gene is responsible for the same activity i.e, membrane-bound sensor/histidine kinase [29, 30] and the same analysis for phoB and phoP which are present in *Escherichia coli* and *Bacillus* respectively and acting as response regulators.

Materials and Methods

Collection of Data

A common set of 44 organisms of *Escherichia coli* for phoBR and 47 organisms of *Bacillus* were selected for our study. I have retrieved the specific gene and amino acid sequences from KEGG (http://www.genome.jp/dbget-bin/www_bget?-f+-n+a+eco:b0400) and all cDNA sequences for the selected organisms from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/) and PATRIC2 (ftp://ftp.patricbrc.org/patric2/genomes/).

Code	Name of the Organism (E.coli) selected for phoBR	Code	Name of the Organism (*Bacillus*) selected for phoPR
eck	*Escherichia coli 55989 (EAEC)*	bao	*Bacillus amyloliquefaciens DSM 7*
eab	*Escherichia coli ABU 83972*	bay	*Bacillus amyloliquefaciens FZB42*
ecv	*Escherichia coli APEC O1*	bql	*Bacillus amyloliquefaciens LL3*
ebr	*Escherichia coli B REL606*	baq	*Bacillus amyloliquefaciens subsp. plantarum CAU B946*
ebe	*Escherichia coli BL21(DE3)*	bamb	*Bacillus amyloliquefaciens subsp. plantarum NAU-B3*
ebl	*Escherichia coli BL21(DE3)*	bama	*Bacillus amyloliquefaciens subsp. plantarum UCMB5033*
ebw	*Escherichia coli BW2952*	baml	*Bacillus amyloliquefaciens subsp. plantarum UCMB5036*
elc	*Escherichia coli clone D i14*	bamn	*Bacillus amyloliquefaciens subsp. plantarum UCMB5113*
eld	*Escherichia coli clone D i2*	bya	*Bacillus amyloliquefaciens subsp. plantarum YAU B9601-Y2*
edj	*Escherichia coli DH1*	baz	*Bacillus amyloliquefaciens TA208*
ecw	*Escherichia coli E24377A (ETEC)*	bxh	*Bacillus amyloliquefaciens XH7*
eih	*Escherichia coli IHE3034*	bqy	*Bacillus amyloliquefaciens Y2*
ecd	*Escherichia coli K-12 DH10B*	bai	*Bacillus anthracis A0248*
ecok	*Escherichia coli K-12 MDS42*	ban	*Bacillus anthracis Ames*
eco	*Escherichia coli K-12 MG16955*	bar	*Bacillus anthracis Ames 0581 (Ames Ancestor)*
ecj	*Escherichia coli K-12 W3110*	bah	*Bacillus anthracis CDC 684*
ekf	*Escherichia coli KO11FL*	banv	*Bacillus anthracis Vollum*
elf	*Escherichia coli LF82*	bcx	*Bacillus cereus 03BB102*
ena	*Escherichia coli NA114*	bcr	*Bacillus cereus AH187*
eoh	*Escherichia coli O103:H2 12009 (EHEC)*	bcu	*Bacillus cereus AH820*
eoi	*Escherichia coli O111:H- 11128 (EHEC)*	bca	*Bacillus cereus ATCC 10987*
ecg	*Escherichia coli O127:H6 E2348/69 (EPEC)*	bcb	*Bacillus cereus B4264*
ecoo	*Escherichia coli O145:H28 RM13514 (EHEC)*	bcz	*Bacillus cereus E33L (zebra killer)*
ecoh	*Escherichia coli O145:H28 RM13516 (EHEC)*	bcg	*Bacillus cereus G9842*
ecf	*Escherichia coli O157:H7 EC4115 (EHEC)*	bcq	*Bacillus cereus Q1*
ece	*Escherichia coli O157:H7 EDL933 (EHEC)*	bcl	*Bacillus clausii KSM-K16*
etw	*Escherichia coli O157:H7 TW14359 (EHEC)*	bha	*Bacillus halodurans C-125*
eum	*Escherichia coli O17:K52:H18 UMN026*	blh	*Bacillus licheniformis 9945A*
eci	*Escherichia coli O18:K1:H7 UTI89 (UPEC)*	bli	*Bacillus licheniformis ATCC 14580*
eoj	*Escherichia coli O26:H11 11368 (EHEC)*	bmd	*Bacillus megaterium DSM 319*
ecz	*Escherichia coli O45:K1:H7 S88 (ExPEC)*	bmq	*Bacillus megaterium QM B1551*
eok	*Escherichia coli O55:H7 CB9615*	bmh	*Bacillus megaterium WSH-002*
ecc	*Escherichia coli O6:K2:H1 CFT073*	bpf	*Bacillus pseudofirmus OF4*
eoc	*Escherichia coli O7:K1 CE10*	bpu	*Bacillus pumilus SAFR-032*
ect	*Escherichia coli O7:K1 IAI39 (ExPEC)*	bsq	*Bacillus subtilis QB928*
ecr	*Escherichia coli O8 IAI1*	bso	*Bacillus subtilis subsp. natto BEST195*
ecq	*Escherichia coli O81 D1a*	bst	*Bacillus subtilis subsp. spizizenii TU-B-10*
ecx	*Escherichia coli O9 HS*	bss	*Bacillus subtilis subsp. spizizenii W23*
elp	*Escherichia coli P12b*	bsu	*Bacillus subtilis subsp. subtilis 168*
ecoi	*Escherichia coli PMV-1*	bsh	*Bacillus subtilis subsp. subtilis 6051-HGW*
ecm	*Escherichia coli SMS-3-5*	bsul	*Bacillus subtilis subsp. subtilis JH642 AG174*
eun	*Escherichia coli UMNK88*	bsr	*Bacillus subtilis subsp. subtilis RO-NN-1*
ell	*Escherichia coli W*	bsx	*Bacillus subtilis XF-1*
elw	*Escherichia coli W*	btk	*Bacillus thuringiensis 97-27 (serovar konkukian)*
		btl	*Bacillus thuringiensis Al Hakam*
Triple/four letter code is the organism code from KEGG.		btb	*Bacillus thuringiensis BMB171*
		btc	*Bacillus thuringiensis serovar chinensis CT-43*

Table1: List of organisms selected to study phoR, phoB and phoP.

Compositional Analysis

The parameters like GC content, amino acid frequencies and RSCU (Relative Synonymous Codon Usage) were used for compositional analysis. I have generated GC1%, GC2%, GC3%, GC%, A3%, U3%, Nc [31] and gravy score of phoR, phoB and phoP using codonW (http://codonw.sourceforge.net/) and in house PERL script. These provided useful information regarding existence of mutational pressures acting on the genes [32]. ENc, the expected effective number of codon were calculated from GC3s under H0 (Null hypothesis, i.e., no selection) according to the given equation, where S denotes GC3.

$$Enc=2+S+ \{29/[S^2 + (1-S)^2]\} \quad ...(1)$$

I have grouped the amino acids into five (5) classes based on their physic chemical properties as follows: i. Negatively charged R group (asp, glu); ii. positively charged R group (his, lys, arg); iii. Polar uncharged R group (ser, thr, cys, asn, gln); iv. Aromatic group (phe, tyr, trp); and v. Non polar aliphatic R group (gly, ala, pro, val, leu, ile, met) containing amino acids. The GC1%, GC2%, GC3% of the organisms was calculated using in house perl script and considering all cDNA sequences. I have also calculated amino acid frequency of the protein encoded by phoR, phoB and phoP, number of times a particular codon used and RSCU values of all codons for those genes using in house or home based PERL programme.

RSCU, the relative synonymous codon usage values close to one indicates lack of biasness where as much higher and lower values indicate preference and avoidance of those particular codons, respectively. Using codonW, the correspondence analysis [33] has been performed to investigate major trend in RSCU variation among genes and distribute the genes along continuous axes in accordance with these trends also I have calculated gravy score to know about the hydrophobicity / hydrophilicity of the protein.

Expressional Probability

The geometric mean of the weight associated to each codon over the length of the gene sequence (measured in codons) is known as Codon Adaptatation Index [34] (CAI) i.e. the measure of gene's probable expression. I have calculated it by following Sharp and Li method [27] and using in house PERL script and MS Excel 2007.

Statistical Analysis

Statistical significance (z) test was performed based on GC1%, GC2%, GC3%, GC%, G3%, C3%, A3%, U3%, CAI, rscu of all codons, frequency of all amino acids and their groups based on physic chemical properties of phoR, phoB and phoP and the protein sequences encoded by them for different situations: i. between *Escherichia coli* and *Bacillus* for the gene phoR and ii. between phoB of *Escherichia coli* and phoP of *Bacillus*.

$$Z_i = (AVERAGE_a - AVERAGE_b)/ \sqrt{(STDEV_a^2/N_a + STDEV_b^2/N_b)} \quad ...(2)$$

Here, i denotes the parameter, a and b denotes the genus and N indicates the sample size.

Motif Analysis

Types of motifs in gene sequence are calculated using the online tool MOTIFSCAN (http://hits.isb-sib.ch/cgi-bin/PFSCAN) and are aligned determining each genes conservation among groups using clustalw2 (http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalw2-E20150224-055314-0919-45389162-oy&analysis=summary) and MultAlin (http://multalin.toulouse.inra.fr/multalin/cgi-bin/multalin.pl) [35]. Also I have checked the similarity of the consensus sequences part responsible for the domains only using bl2seq of BLAST from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with default parameter values (Expect threshold: 10; Word size: 3; Matrix: BLOSUM62; Gap costs: Existence 11 & Extension 1).

.

Results and Discussion

Compositional Variability

Variation in phoR gene based on GC content, Amino Acid Frequencies and Relative Synonymous Codon Usage values-

In table 2, I am observing that the compositional parameters i.e, GC1, GC2, GC3 and even G3, C3 are significantly higher in phoB and phoR from *Escherichia coli* than phoP and phoR from *Bacillus* respectively.

Parameter / gene_organism code[#]	B_E	P_B	R_B	R_E	Z value of	
					B_E & P_B	R_E & R_B
GC1	64.54(0.26)	53.83(2.36)	51.8(2.75)	60.87(0.25)	30.95	22.59
GC2	37.81(0.1)	29.33(1.35)	32.16(2.93)	39.29(0.16)	43.29	16.7
GC3	60.36(0.96)	38.54(11.26)	41.62(12.28)	59.11(0.67)	13.25	9.76
T3S	0.28(0.02)	0.4(0.08)	0.4(0.08)	0.3(0.01)	-11.2	-9.05
C3S	0.34(0.02)	0.19(0.09)	0.24(0.1)	0.28(0.01)	12.54	3.1
A3S	0.25(0.01)	0.42(0.09)	0.37(0.1)	0.23(0.01)	-14.87	-9.61
G3S	0.4(0.01)	0.31(0.09)	0.3(0.07)	0.45(0.01)	7.77	14.85
CAI	0.7(0.02)	0.75(0.04)	0.74(0.05)	0.71(0.02)	-8.35	-4.17

Table2: Average (Standard deviation) and Z values of different positional nucleotide %'s (GC1, GC2, GC3 are within 100 & T3, A3, C3, G3 are within 1) (individual/combination) and CAI.

in place of name of the gene and organism this code may have been used. phoB_*E.coli* (B_E); phoP_*Bacillus* (P_B); phoR_*Bacillus* (R_B); phoR_*E.coli* (R_E)

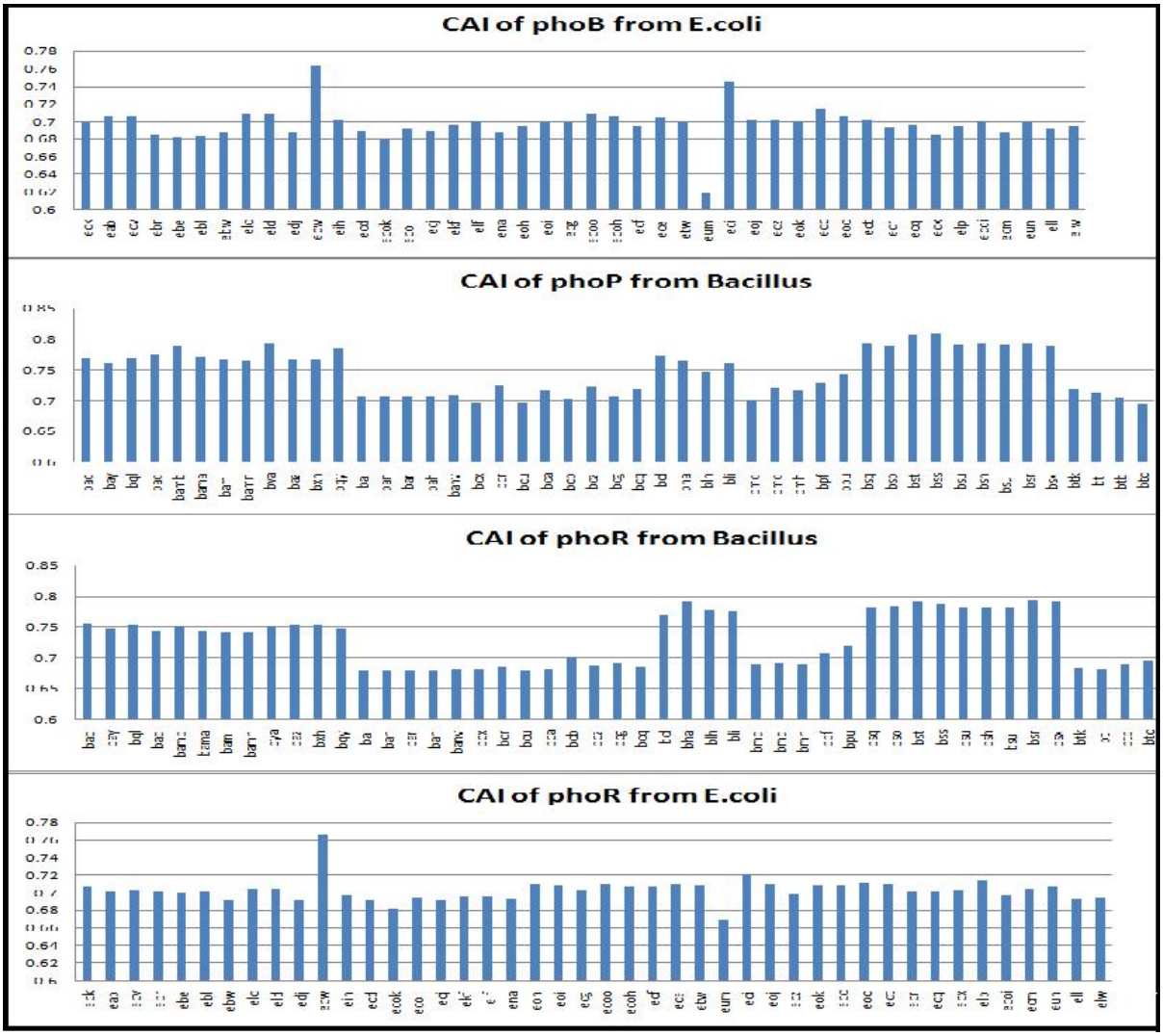

Figure 1: Codon Adaptation Index (CAI) of phoB & phoR from *E.coli* and phoP & phoR from *Bacillus*.

From figure 1 it is clear that the values of Codon Adaptation Index (CAI) are similar for phoB and phoR in *Escherichia coli* and also for phoP and phoR in *Bacillus*. This similarity in expression indicates their co expression. The CAI values are varying between 0.61 to 0.81 which also showing the importance of the gene set for their survivability. But the z values for CAI in table 2 depicting that the levels of expression are significantly more for the gene set from *Bacillus* than from *Escherichia coli*.

In table 3&4 under the column 'Z values of', values which are >+1.96 and <-1.96 indicates significant difference between a. phoB from *Escherichia coli* & phoP from *Bacillus* (under sub column 'B_E & P_B') and b. phoR from *Bacillus* and *Escherichia coli* (under sub column 'R_E & R_B') for corresponding parameter (in table3: RSCU value of the codon and in table 4: individual Amino acid frequency or the group). The positive values indicate that the corresponding parameter value is higher in *Escherichia coli* than *Bacillus*.

Amino Acids	codons	B_E	P_B	R_B	R_E	Z values of	
						B_E & P_B	R_E & R_B
A	gca	0.44(0.17)	1.07(0.35)	1.01(0.4)	0.76(0.1)	-11.22	-4.2
	gcc	0.37(0.01)	0.48(0.39)	0.83(0.56)	0.91(0.05)	-2.04	0.95
	gcg	2.52(0.12)	1.44(0.4)	1.34(0.39)	1.93(0.12)	18.04	10.04
	gcu	0.68(0.13)	1.03(0.42)	0.83(0.34)	0.42(0.08)	-5.46	-8.2
C	ugc	2(0)	0.6(0.93)	0.86(0.5)	1.5(0.39)	10.42	7.05
	ugu	0(0)	1.41(0.93)	1.07(0.52)	0.51(0.39)	-10.42	-5.99
D	gac	0.83(0.09)	0.41(0.3)	0.63(0.3)	0.17(0.03)	9.36	-10.64
	gau	1.18(0.09)	1.6(0.3)	1.38(0.3)	1.84(0.03)	-9.36	10.64
E	gaa	1.29(0.05)	1.38(0.19)	1.41(0.16)	1.11(0.04)	-3.12	-12.81
	gag	0.72(0.05)	0.63(0.19)	0.6(0.16)	0.9(0.04)	3.12	12.81
F	uuc	0.45(0.01)	0.6(0.35)	0.46(0.16)	0.53(0.07)	-3	2.8
	uuu	1.56(0.01)	1.41(0.35)	1.55(0.16)	1.48(0.07)	3	-2.8
G	gga	0.61(0.24)	1.28(0.48)	1.28(0.47)	0.66(0.1)	-8.74	-8.9
	ggc	1.77(0.15)	0.9(0.71)	0.96(0.66)	1.34(0.1)	8.22	3.98
	ggg	1(0.24)	0.86(0.29)	0.92(0.22)	0.78(0.06)	2.7	-4.31
	ggu	0.64(0.15)	0.98(0.79)	0.86(0.28)	1.24(0.08)	-2.9	9.09
H	cac	1.43(0.01)	0.25(0.33)	0.5(0.21)	0.67(0.04)	25.34	5.79
	cau	0.58(0.01)	1.76(0.33)	1.51(0.21)	1.34(0.04)	-25.34	-5.79
I	aua	0.02(0.06)	0.33(0.31)	0.34(0.14)	0.32(0.08)	-7.08	-0.72
	auc	1.2(0.22)	0.95(0.54)	0.92(0.45)	1.06(0.05)	2.88	2.07
	auu	1.8(0.2)	1.73(0.4)	1.76(0.46)	1.64(0.08)	1.16	-1.77
K	aaa	1.15(0.01)	1.5(0.13)	1.42(0.17)	1.67(0.05)	-19.13	10.01
	aag	0.86(0.01)	0.51(0.13)	0.59(0.17)	0.34(0.05)	19.13	-10.01
L	uua	1.06(0.16)	2.04(1.27)	1.79(1.15)	0.69(0.04)	-5.26	-6.58
	uug	0.32(0.07)	0.73(0.2)	0.65(0.22)	0.86(0.1)	-13.59	6.11
	cua	0.18(0.16)	0.27(0.32)	0.27(0.23)	0.43(0.07)	-1.8	4.6
	cuc	1.16(0.22)	0.39(0.38)	0.78(0.5)	0.48(0.05)	12.2	-4.16
	cug	2.75(0.21)	1.44(1.32)	1.07(0.73)	3.09(0.1)	6.74	18.84
	cuu	0.56(0.12)	1.16(0.38)	1.48(0.35)	0.49(0.05)	-10.63	-19.43
M	aug	1(0)	1(0)	1(0)	1(0)	#DIV/0!	#DIV/0!

N	aac	0.74(0.15)	0.59(0.28)	0.65(0.2)	1.21(0.07)	3.27	18.69
	aau	1.27(0.15)	1.42(0.28)	1.36(0.2)	0.8(0.07)	-3.27	-18.69
P	cca	0.92(0.12)	1.63(1.07)	0.93(0.9)	0.44(0.1)	-4.55	-3.78
	ccc	0.54(0.13)	0.28(0.3)	0.22(0.22)	0.37(0.05)	5.6	4.72
	ccg	1.99(0.07)	1.51(0.78)	2.11(0.66)	3.01(0.11)	4.26	9.27
	ccu	0.56(0.07)	0.6(0.61)	0.76(0.41)	0.2(0.03)	-0.39	-9.39
Q	caa	0.62(0.11)	1.03(0.64)	1.05(0.68)	0.67(0.06)	-4.39	-3.85
	cag	1.39(0.11)	0.98(0.64)	0.96(0.68)	1.34(0.06)	4.39	3.85
R	aga	0.55(0.01)	1.9(1.12)	1.26(0.27)	0(0)	-8.34	-32.28
	agg	0(0)	0.75(0.57)	0.7(0.46)	0.16(0.03)	-8.96	-8.02
	cga	0.41(0.14)	0.48(0.62)	0.43(0.34)	0.37(0.08)	-0.78	-1.19
	cgc	3.13(0.14)	1.05(0.65)	1.22(0.41)	2.58(0.16)	21.58	21.68
	cgg	0.41(0.14)	0.72(0.75)	0.76(0.59)	0.81(0.02)	-2.74	0.66
	cgu	1.53(0.14)	1.14(1.1)	1.66(0.53)	2.09(0.12)	2.41	5.46
S	agc	0.58(0.28)	1.86(0.95)	1.03(0.36)	1.64(0.17)	-8.83	10.69
	agu	1.43(0.28)	1.52(0.95)	1.03(0.6)	1.82(0.14)	-0.63	8.9
	uca	0.67(0.01)	1.1(0.9)	0.82(0.32)	0.37(0.15)	-3.31	-8.85
	ucc	0.69(0.11)	0.31(0.47)	0.98(0.57)	0.29(0.02)	5.34	-8.33
	ucg	1.82(0.31)	0.56(0.59)	0.83(0.29)	1.42(0.06)	13.08	13.92
	ucu	0.84(0.33)	0.68(0.57)	1.35(0.48)	0.49(0.14)	1.65	-11.86
T	aca	1(0.12)	1.53(0.41)	1.29(0.52)	0.59(0.08)	-8.76	-9.28
	acc	2.15(0.17)	0.64(0.55)	0.63(0.4)	1.76(0.03)	18.13	19.71
	acg	0.66(0.08)	1.33(0.53)	1.63(0.41)	1.45(0.12)	-8.83	-2.99
	acu	0.22(0.18)	0.52(0.41)	0.47(0.3)	0.22(0.11)	-4.6	-5.4
V	gua	0.43(0.09)	0.95(0.45)	1.06(0.31)	0.56(0.04)	-7.84	-11.09
	guc	1.07(0.07)	0.65(0.55)	0.78(0.37)	0.46(0.08)	5.24	-5.86
	gug	2.52(0.08)	1.07(0.29)	1.08(0.21)	2.21(0.05)	33.77	36.66
	guu	0.01(0.03)	1.36(0.56)	1.1(0.27)	0.79(0.08)	-16.75	-7.93
W	ugg	1(0)	1(0)	1(0)	1(0)	#DIV/0!	#DIV/0!
Y	uac	0.4(0)	0.54(0.23)	0.79(0.41)	0.99(0.06)	-4.09	3.34
	uau	1.6(0)	1.47(0.23)	1.22(0.41)	1.02(0.06)	4.09	-3.34

Table3: Average (Standard Deviation) of RSCU values of the codons and z score. Z Values which are non significant at 5% significance level are highlighted with grey colour. Exceptional values are highlighted by Green and Red colour.

From table 2 its clear to us that the gene set from *Escherichia coli* is GC rich than the homologous gene set from *Bacillus*. So it is expected that the GC rich codons must appear more frequently than the AT/U rich codons in *Escherichia coli* than *Bacillus*. But in table 3, I am getting some exceptional situations where the RSCU values of few /all synonymous codon for some amino acids are showing reverse result in reference to our expectation for phoB (Y; S:agc; R:cgg; L:uug; F; A:gcc)/phoR (V:guc; S:ucc; L:cuc; K:aag; G:ggu & ggg; D) (highlighted by red colour) or for both (R:agg & cgu) (highlighted by green colour) in *Escherichia coli*.

Amino Acids		B_E	P_B	R_B	R_E	Z values of	
Triple letter code	Single letter code					B_E & P_B	R_E & R_B
asp	D	57(0)	75.45(7.34)	52.52(6.07)	27.71(0.71)	-17.24	-27.85
glu	E	105(0)	103.56(7.31)	87.43(10.25)	74.12(1.11)	1.36	-8.86
his	H	31(0)	13.69(4.86)	26.05(4.81)	34.39(1.39)	24.44	11.43
lys	K	31(0)	85.15(10.1)	75.15(8.92)	39.16(0.75)	-36.78	-27.56
arg	R	95.91(0.61)	54.66(4.72)	49.54(9.89)	86.07(0.46)	59.47	25.32
phe	F	39(0)	28.07(7.68)	42.26(3.99)	39.23(1.47)	9.78	-4.88
tyr	Y	22(0)	34.52(4.71)	29.47(5.54)	19(0)	-18.22	-12.97
trp	W	13(0)	4.39(1.84)	3.56(1.81)	25.66(2.27)	32.18	51.34
pro	P	61(0)	39.49(4.16)	19.9(2.25)	48.91(0.71)	35.51	84.13
cys	C	4(0)	4.39(1.84)	3.88(1.78)	6.91(0.61)	-1.43	11.06
met	M	52(0)	33.92(4.39)	27.22(3.64)	28.1(0.43)	28.29	1.65
gly	G	66(0)	51.13(8.22)	67.58(4.78)	70.07(0.46)	12.41	3.57
ala	A	47.91(0.61)	52.39(8.35)	55.41(10.1)	50.57(1.32)	-3.67	-3.26
val	V	96.1(0.61)	67.96(10.94)	74.73(9.63)	65.46(1.49)	17.61	-6.52
ile	I	48(0)	68.52(7.47)	80.54(6.68)	39.19(0.7)	-18.84	-42.21
leu	L	87(0)	130.6(7.65)	105.47(6.84)	143.44(3.85)	-39.08	32.93
gln	Q	31.1(0.61)	34.35(9.06)	38.26(8.77)	46.6(1.55)	-2.46	6.42
asn	N	22(0)	29.75(5.73)	38.41(3.17)	49.39(1.46)	-9.28	21.49
ser	S	39(0)	35.24(11.6)	63.54(8.56)	48.98(0.55)	2.23	-11.64
thr	T	52(0)	52.81(12.13)	58.15(3.76)	58.12(1.11)	-0.46	-0.07
negatively charged R gr		162(0)	179(11.48)	139.94(6.17)	101.82(1.21)	-10.16	-41.55
positively charged R gr		157.91(0.61)	153.49(10.48)	150.73(4.95)	159.62(2.12)	2.89	11.28
aromatic gr		74(0)	66.96(4.15)	75.28(7.49)	83.89(3.41)	11.64	7.14
non polar aliphatic R gr		458(0)	443.98(13.96)	430.81(6.34)	445.71(2.12)	6.89	15.24
polar uncharged R gr		148.1(0.61)	156.52(15.99)	202.22(11.94)	209.98(2.61)	-3.61	4.36

Table4: Average (Standard Deviation) of amino acids frequencies and z score.

Though phoB of *Escherichia coli* and phoP of *Bacillus* encoding the homologous protein which are acting as response regulator, in table 4 the Z score showing that almost there is no similarity in the frequency level of amino acids except few like: glu, cys & thr in their sequences. Individual frequencies of D, K, Y, A, I, L, Q, N and as group negatively charged and polar uncharged R group containing amino acids have significant higher frequencies in phoP from *Bacillus* than phoB from *Escherichia coli*. Similarly except the amino acids cys and met, the frequency of the amino acids are totally different in phoR sequence selected from *Escherichia coli* and *Bacillus*. In phoR sequence individual frequencies of D, E, K, F, Y, A, V, I, S and as group the negatively charged R group containing amino acids have the higher use in *Bacillus* than *Escherichia coli*.

Correspondence Analysis And GC3 Vs. Enc/Nc Plot:

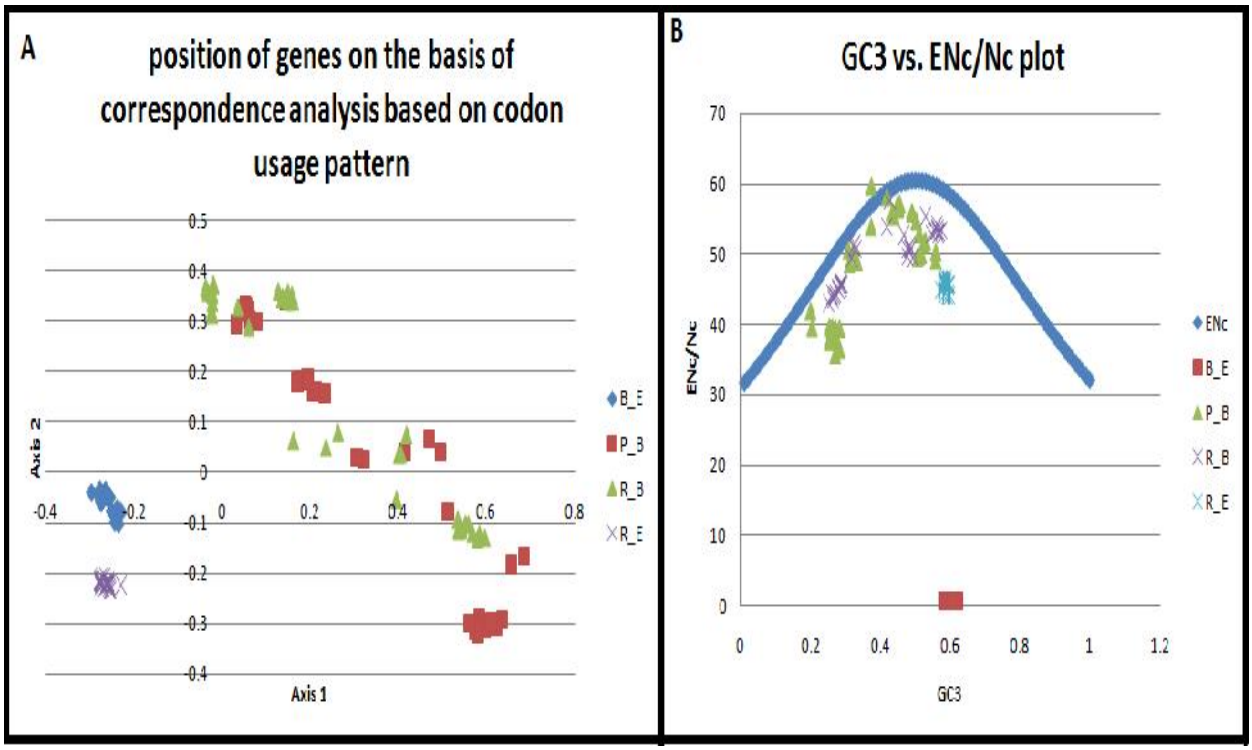

Figure 2: A. Correspondence analysis of the genes based on their RSCU values. B. $GC_{3\ vs.}$ ENc/Nc plot of the four genes.

Correspondence analysis on the basis of codon usage pattern as in figure 2A which showing that the gene phoB from *Escherichia coli* and its homolog phoP from *Bacillus* are not occupying same position in graph and similar observation for phoR from both. From GC3 Vs. ENC/Nc plot i.e, the figure 2B, I am observing that all the genes are under mutational pressure except phoP of *Bacillus halodurans C-125*(bha).

Motif Analysis

The entire consensus sequence of phoP from *Bacillus* and phoB from *Escherichia coli* are as follows where the red coloured portion (length ~ 117 aa) is responsible for response regulation:

>P_B_ consensus

MSGKMEAYIVNMNKRILVVDDEESIATLIQYNLERAGYEVITASDGEEALQKAKEEKPDLIILDVMLPKMDG
FEVCKQLRQQKVMVPILMLTAKDDEFDKVLGLELGADDYMTKPFSPREVTARVKAILRRVEVASESS-
EQNEEDEKAGTITIGELKINPEHYEVYFKGELLELTPKEFELLLYLANHKGRVLTRDQLLNAVWNYDFAGDT
RIVDVHISHLRDKIEPNTKKPVYIKTIRGLGYKLEEPK

>B_E_ consensus

MARRILVVEDEAPIREMVCFVLEQNGFQPVEAEDYDSAVNQLNEPWPDLILLDWMLPGGSGIQFIKHLKRES
MTRDIPVVMLTARGEEEDRVRGLETGADDYITKPFSPKELVARIKAVMRRISPMAVEEVIEMQGLSLDPTSH
RVMAGEEPLEMGPTEFKLLHFFMTHPERVYSREQLLNHVWGTNVYVEDRTVDVHIRRLRKALEPGGHDRM
VQTVRGTGYRFSTRF

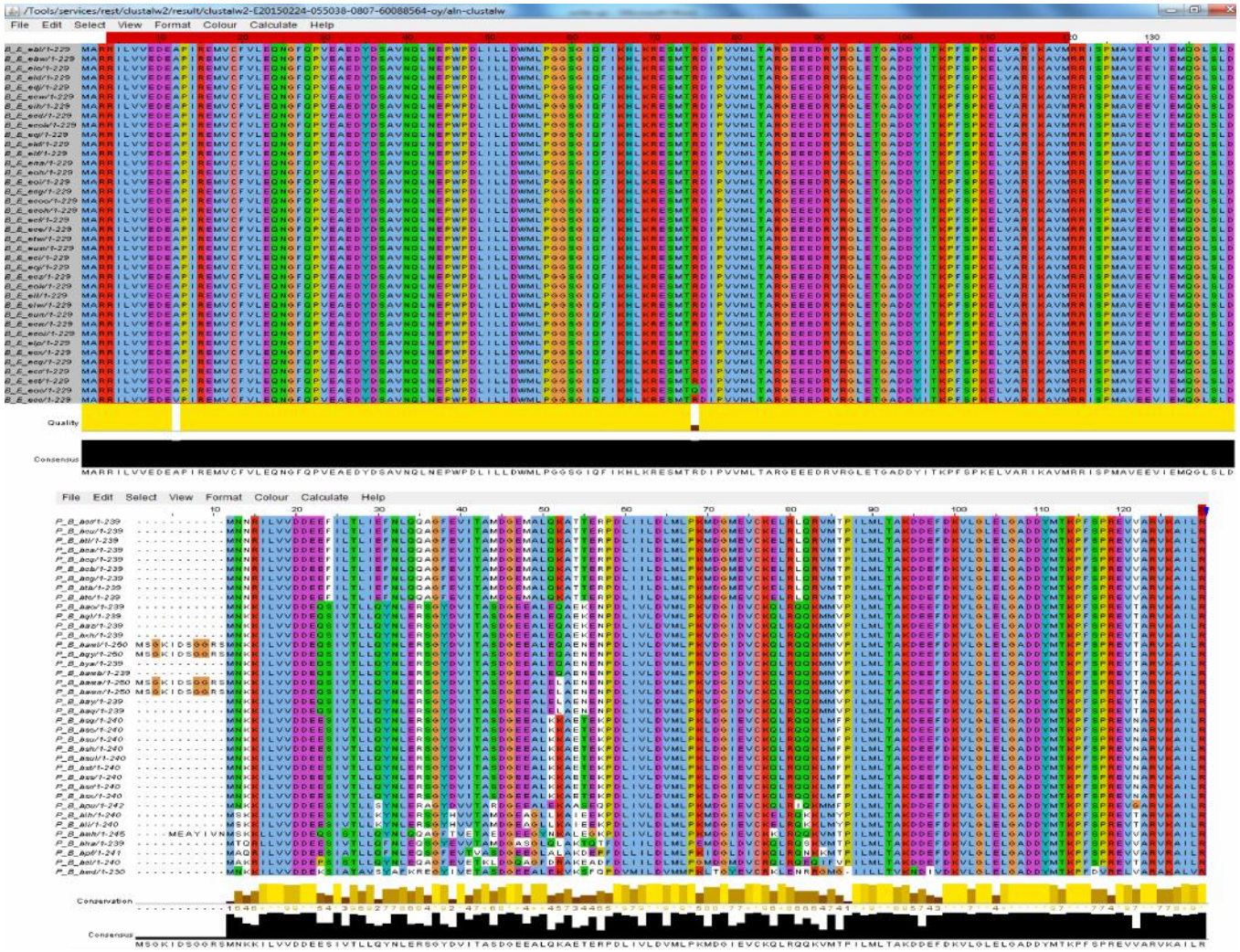

Figure 3: Jalview of multiple sequence alignment of phoB from *Escherichia coli* (upper part) and phoP from *Bacillus* (lower part).

Figure 3, multiple sequence alignment showing that in phoP there are more variations of character/amino acids in the columns where as in phoB the columns are static. The regular expression for the response regularity part of phoP from *Bacillus* is as follows:

MNK[KR]ILVVDDE[EQ][FS]I[VL]TLL[EQ][FY]NL[QE][RQ][SA]G[FY][DE]VITA[SM]DGE[EM]AL[QEK][KQ]A[ET][TN]E[RNT]PDLI[VI]LD[VL]MLPK[MLV]DG[IM][ED]VCK[QE]LR[QL]Q[KR][VML]M[TVF]PILMLTAKDDEFDKVLGLELGADDYMTKPFSPREV[VTN]ARVKAILR

Consensus sequence of phoR from *Escherichia coli* and *Bacillus* obtained through multiple sequence alignment are as follows:

>R_B_ consensus

MNKFRTRLFFALIVLIILVFVGLGLFLGQLFENYYEDHLSERMKKEAEYVASLVDEDGIPNSKQNQQIIEEAG
RELDVRVSIIDADGKVLYHSGGDPEEMENHSDVMEISLSEKDKQVLKMRFSLTVTEESLYYAVPIQTEQGEQ
LGYVLISSSVEPLQDINQEIWGMLAVSFTTAFIVIVFLGMRITSQYVRPIESATKVAVELAKGNYKARTYEDYS

DETGMLGRSMNALAYNLQEMTRTHEMQQDRLNTLIENMGSGLILIDNRGYINLVNRSYREQFHVNPDEWL
HRLYHEVFEHEEIIQLVEEIFMTETKKRKLLRLPINIERRYFEVDGAPIMGTNDEWKGIVLVFHDKRKLEFAIT
ELKKLEQMRKDFVANVSHELKTPITSIKGFTETLLDGAMDDPQLMEEFLSIILKESERLQSLIQDLLDLSKIEQ
QNFKLNIEQVDLKKILEDIZELLKNKAEEKGISLHLNVPEKPAYVWGDPERLKQIFINLVNNAITYTPEGGKVS
VSLKEQENSVVIKVSDTGIGIQKEEIPRIFERFYRVDKDRSRNSGGTGLGLAIVKHLVEAHEGKIEVESEEGKG
TTFTVTFPKKTEKKQ

>R_E_consensus

MLERLSWKRLVLELLLCCLPAFILGAFFGYLPWFLLASVTGLLIWHFWNLLRLSWWLWVDRSMTPPPGRGS
WEPLLYGLHQMQLRNKKRRRELGNLIKRFRSGAESLPDAVVLTTEEGGIFWCNGLAQQILGLRWPEDNGQN
ILNLLRYPEFTQYLKTRDFSRPLNLVLNTGRHLEIRVMPYTHKQLLMVARDVTQMHQLEGARRNFFANVSH
ELRTPLTVLQGYLEMMDEQPLEGAVREKALHTMREQTQRMEGLVKQLLTLSKIEAAPTHLLNEKVDVPMM
LRVVEREAQTLSQKKQTFTFEIDNGLKVSGNEDQLRSAISNLVYNAVNHTPEGTHITVRWQRVPHGAEFSVE
DNGPGIAPEHIPRLTERFYRVDKARSRQTGGSGLGLAIVKHAVNHHESRLNIESTVGKGTRFSFVIPERLIAKN
SD

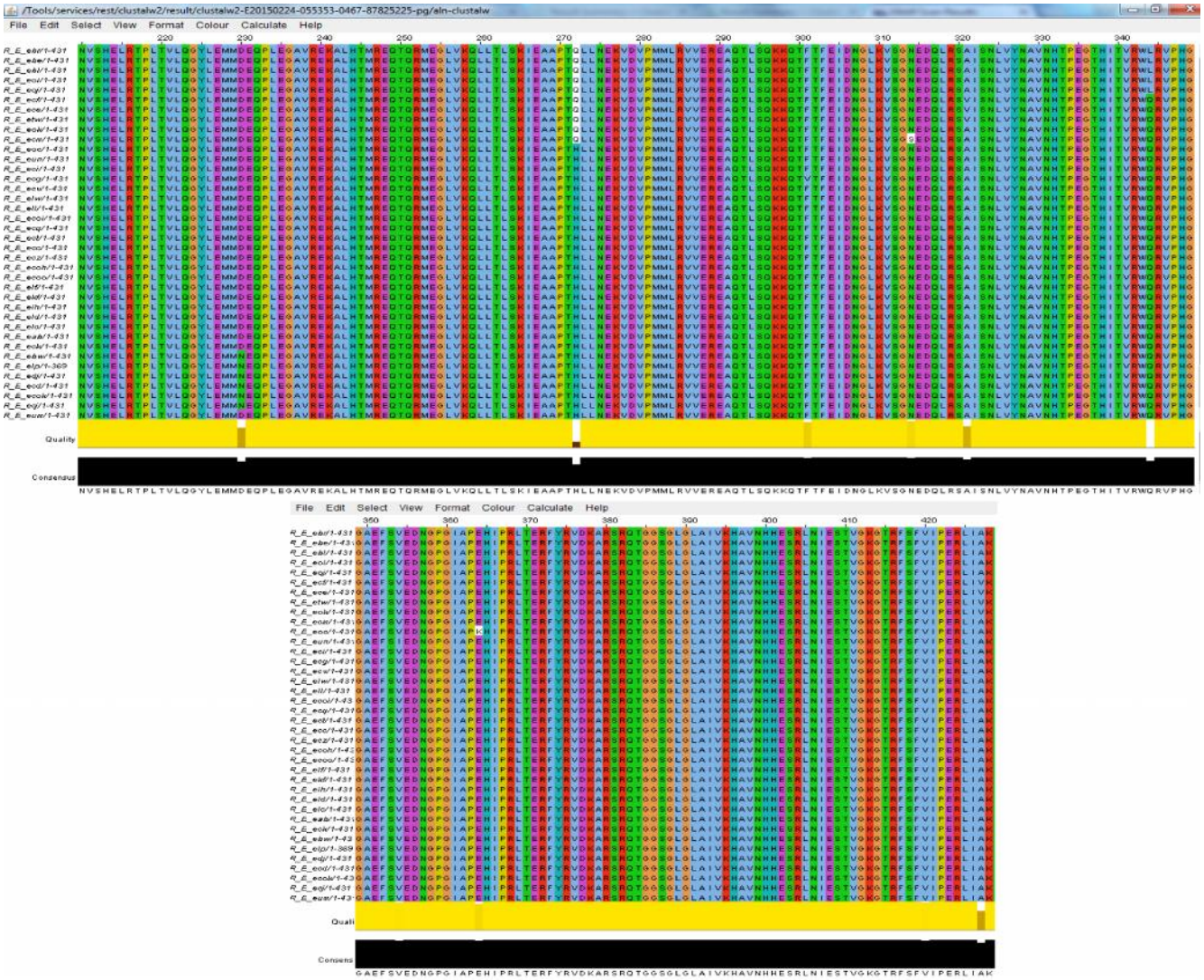

Figure 4 A: Jalview of multiple sequence alignment of phoR from *Escherichia coli* (membrane bound sensor/histidine kinase motif part only belongs from 210 to 425 within the sequence).

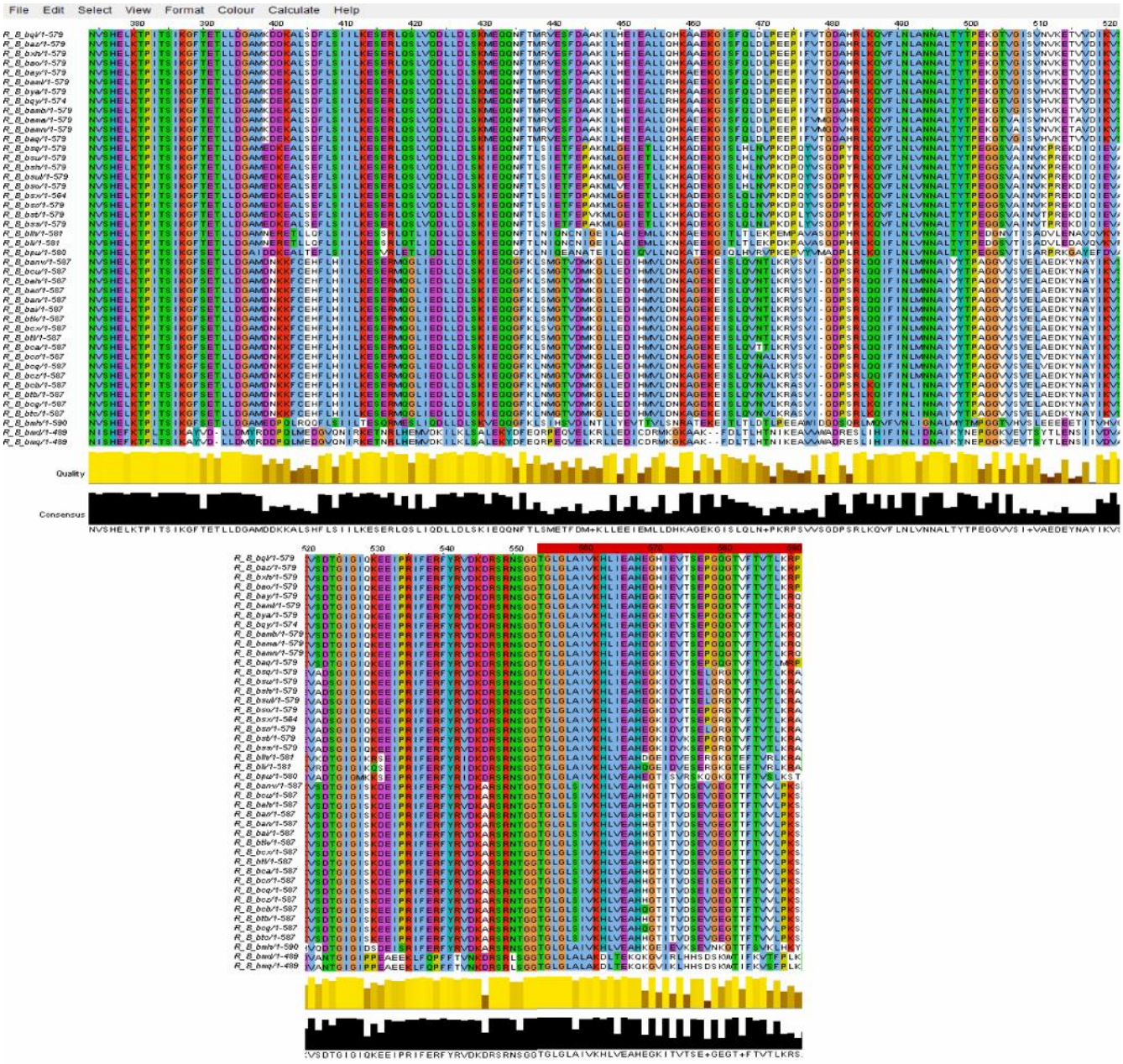

Figure 4 B: Jalview of multiple sequence alignment of phoR from *Bacillus* (membrane bound sensor/histidine kinase motif part only belongs from 359 to 577 within the sequence).

Figures 4A & 4B of multiple sequence alignment showing that in phoR from *Bacillus* there are more variations of character/amino acids in the columns where as static in *Escherichia coli*. The regular expression for the response regularity part of phoR from *Bacillus* is as follows:

NVSHELKTPITSIKGF[TS]ETLLDGAM[DEK][DN][DKE][KE][AF][LC][SE][DEH]FL[SH]IILKESER[LM]Q[SG]L[IV][DE]DLLDLSK[IM]EQQ[NG]F[TK][LM][SRN][MIV][EG][TS][FV][DE][MAP][KA][KG][LIM]L[EHG][DE]I[EH][MTA][LV]L[DQK][HN]KA[GDEA]EK[GE]IS[LF]Q[LV][ND][LVT][PL][KE][DER][PV][SQI][FTV][TS]GD[PA][SYH]RL[KQ]Q[VI]F[LI]NL[VIMA]NNA[LI][TV]YTP[EA][GK]G[VTS]V[SGA][IV][SNE][VL][NKAH][VEP][KRD][EK][TKY][VDN][VIA][DQY]I[KE]V[SA]D[TS]GIGI[QS]K[ED]EIPRIFERFYRVDK[DA]RSRN[ST]GGTGLGL[AS]IVKHL[VI]EAH[EH]G[KHT]I[TDE]V[TD]SE[PLV]G[EQR]GT[TV]FTV[TV]L[KP][RK][SPQA]

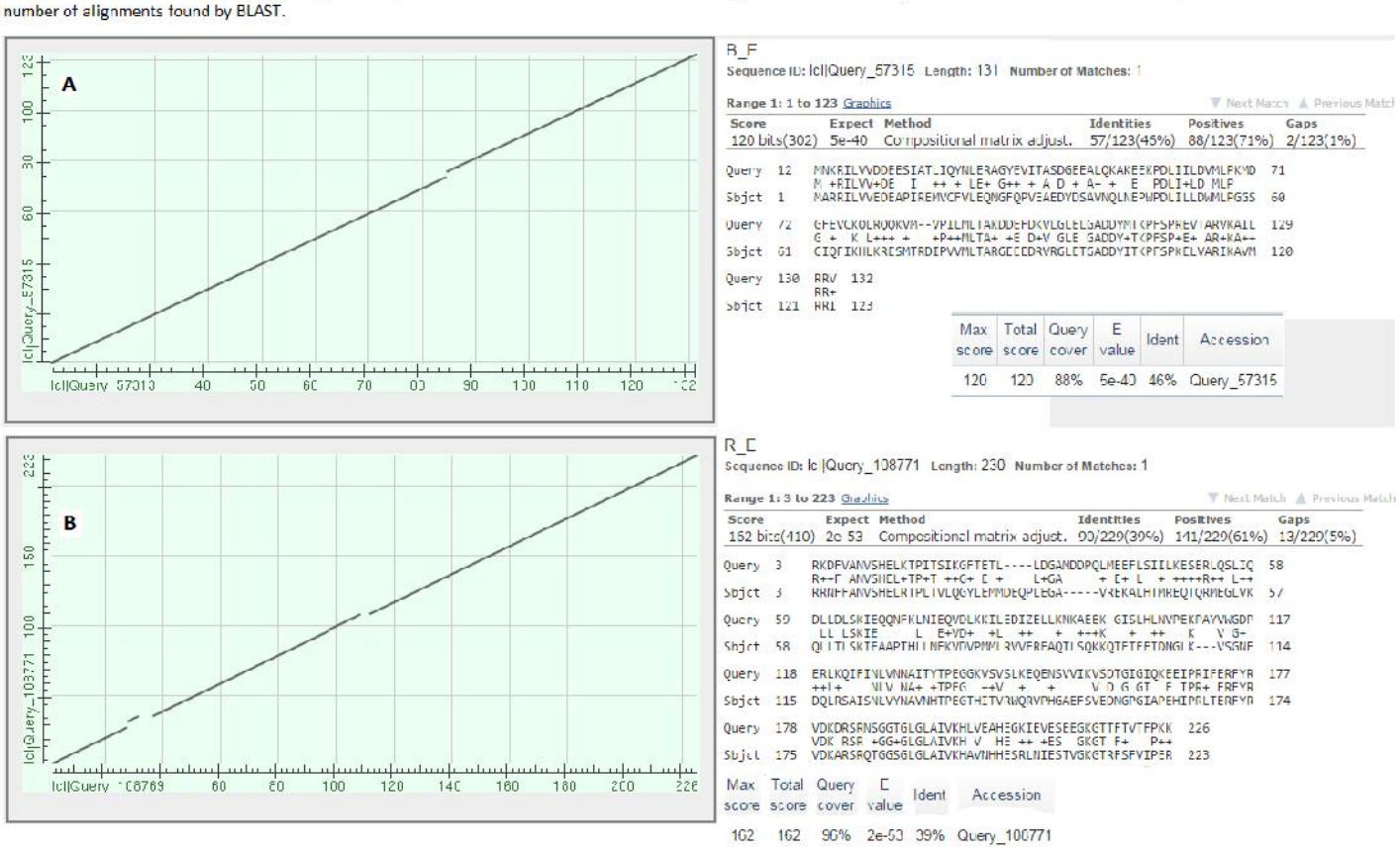

Figure 5: Pairwise alignment of the concensus sequence part responsible for A. response regulator from B_E & P_B and B. membrane-bound sensor from R_E & R_B.

Pair wise alignment with graphical representation as in figure 5 between consensus sequences for the response regularity motifs from *Escherichia coli* and *Bacillus* showing ~46% identities with ~71% positivity and for the membrane bound sensor/histidine kinase motifs from *Escherichia coli* and *Bacillus* showing ~39% identities with ~61% positivity which indicates the positive/acceptable substitution of the amino acids within the sequence to restore their activity.

Conclusion

The compositional analysis of phoBR and phoPR which are encoding the Pi-responsive two-component regulatory system in *Escherichia coli* and *Bacillus* respectively depicting that in genomic and proteomic level there is no similarity except few parameters and positive/acceptable substitution has takes place in phoPR of *Bacillus* where as phoBR of *Escherichia coli* is static. Variation in phoPR insisted us to propose the regular expression for the motifs. Though the gene set phoBR is GC rich but I am getting some exceptional situations where the RSCU values of few /all synonymous codon for some amino acids are showing reverse result in reference to our expectation in *Escherichia coli*. Correspondence analysis shows that the codon usage patterns of the genes are different but all are under mutational pressure except phoP of *Bacillus halodurans C-125*(bha). Those differences in composition have no reflection in the level of expression of phoB and phoP and varying from 0.61 to 0.81and and for phoR from 0.67 to 0.79 and also CAI values reflecting the co expression of phoBR in *Escherichia coli* and phoPR in *Bacillus*.

Acknowledgments

I am very much thankful to Prof. Sudip Kundu, Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta for his valuable suggestion about the calculation of the parameters used within this manuscript.

References

1. Metcalf, W. W. and Wolfe, R. S. (1998) Molecular genetic analysis of phosphate and hypophosphite oxidation by *Pseudomonas stutzeri* WM88. *J. Bacteriol*, **180**, 5547–5558.
2. Ehrlich, H. L. (1990) Geomicrobial processes. A physical and biochemical overview. In Geomicrobiology, 2nd Ed. Marcel Dekker.
3. Fernández, C., Novo, R., Microbiana, V.and Suelo, E. (1988) II. La Habana: Editorial Puebloy Educación fibroblasts. *Life Sci* **52**, 1909–1915.
4. Kulaev, I. S. and Vagabov, V. M. (1983) Polyphosphate metabolism in micro-organisms. *Adv. Microbiol. Physiol.*, **24**, 83-171.
5. Harold, F. M. (1966) Inorganic polyphosphates in biology: structure, metabolism, and function. *Bacteriol. Rev.*, **30**, 772-794.
6. Kornberg, A. (1995) Inorganic polyphosphate: toward making a forgotten polymer unforgettable. *J. Bacteriol.*, **177**, 491-496.
7. Paul, E. A. and Clark, F. E. (1988) Soil Microbiology and Biochemistry. San Diego, *CA: Academic Press*.
8. Dalal, R. C. (1977) Soil organic phosphorus. *Adv. Agron.* **29**, 83–117.
9. Richardson, A. E. (1994) Soil microorganisms and phosphorous availability. In Soil Biota: Management in Sustainable Farming Systems. Eds. CE Pankhurst, BM Doube and VVSR Gupta. CSIRO, Victoria, Australia.
10. Richardson, A.E., Hadobas, P. A., Hayes, J. E., O'Hara, C. P. and Simpson, R. J. (2001) Utilization of phosphorus by pasture plants supplied with myo-inositol hexaphosphate is enhanced by the presence of soil microorganisms. *Plant Soil*, **229,** 47-56.
11. Wanner, B. L. (1996) Phosphorus Assimilation and Control of the Phosphate Regulon. In *Escherichia coli and Salmonella: cellular and molecular biology Volume 1*. 2nd edition. Edited by: Neidhardt, F. C., Curtiss, R. III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M.and Umbarger, H. E. Washington, DC: ASM Press:1357-1381.
12. van Veen, H. W. (1997) Phosphate transport in prokaryotes: molecules, mediators and mechanisms. *Antonie Van Leeuwenhoek*, **72(4)**, 299-315.
13. van Veen, H. W., Abee, T., Kortstee, G. J., Konings, W. N. and Zehnder, A.J. (1993) Mechanism and energetics of the secondary phosphate transport system of *Acinetobacter johnsonii* 210A. *J Biol Chem*, **268(26),**19377-19383.
14. Allenby, N. E., O'Connor, N., Pragai, Z., Carter, N. M., Miethke, M., Engelmann, S., Wipat, A., Ward, A. C. and Harwood, C. R. (2004) Posttranscriptionalregulation of the *Bacillus* subtilis pst operon encoding a phosphate-specific ABC transporter. *Microbiology*, **150**, 2619–2628.
15. Qi, Y., Kobayashi, Y. and Hulett, F. M. (1997) The pst operon of *Bacillus* subtilis has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the Pho regulon. *J Bacteriol*, **179**, 2534–2539.
16. Wanner, B. L. (1996) Phosphorus assimilation and control of the phosphate regulon. In *Escherichia coli* and Salmonella: Cellular and Molecular Biology, Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology. 1357–1381.
17. Yuan, Z. C., Zaheer, R. and Finan, T. M. (2006) Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of Sinorhizobium meliloti. *J Bacteriol*, **188**, 1089–1102.
18. Metcalf, W. W. and Wanner, B. L. (1991) Involvement of the *Escherichia coli* phn (psiD) gene cluster in assimilation of phosphorus in the form of phosphonates, phosphite, Pi esters, and Pi. *J Bacteriol*, **173**, 587–600.
19. Bardin, S., Dan, S., Osteras, M. and Finan, T. M. (1996) A phosphate transport system is required for symbiotic nitrogen fixation by *Rhizobium meliloti*. *J Bacteriol*, **178**, 4540–4547.

20. Voegele, R. T., Bardin, S. and Finan, T. M. (1997) Characterization of the Rhizobium (Sinorhizobium) meliloti high- and low-affinity phosphate uptake systems. *J Bacteriol*, **179**, 7226–7232.
21. Bardin, S. D. and Finan, T. M. (1998) Regulation of phosphate assimilation in *Rhizobium (Sinorhizobium) meliloti. Genetics,* **148**, 1689–1700.
22. Sola-Landa, A., Rodriguez-Garcia, A., Franco-Dominguez, E. and Martin, J. F. (2005) Binding of phoP to promoters of phosphateregulated genes in *Streptomyces coelicolor*: identification of PHO boxes. *Mol Microbiol*, **56**, 1373–1385.
23. White, A. K. and Metcalf, W. W. (2004) The htx and ptx operons of Pseudomonas stutzeri WM88 are new members of the Pho regulon. *J Bacteriol,* **186**, 5876–5882.
24. von Kruger, W. M., Humphreys, S. and Ketley, J. M. (1999) A role for the phoBR regulatory system homologue in the *Vibrio cholerae* phosphatelimitation response and intestinal colonization. *Microbiology,* **145**, 2463–2475.
25. Wanner, B. L. and Chang, B. D. (1987) The phoBR operon in *Escherichia coli* K-12. *J Bacteriol*, **169**, 5569–5574.
26. Kriakov, J., Lee, S. and Jacobs, W. R., (2003) Identification of a regulated alkaline phosphatase, a cell surface-associated lipoprotein, in *Mycobacterium smegmatis*. *J Bacteriol*, **185**, 4983–4991.
27. Tran, S. L., Rao, M., Simmers, C., Gebhard, S., Olsson, K. and Cook, G. M. (2005) Mutants of *Mycobacterium smegmatis* unable to grow at acidic pH in the presence of the protonophore carbonyl cyanide m-chlorophenylhydrazone. *Microbiology,* **151**, 665–672.
28. Wanner, B. L. (1987) Phosphate regulation of gene expression in *Escherichia coli*. In: Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. and Umbarger, H. E. (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology,. American Society for Microbiology, Washington, D.C. **2**, 1326-1333.
29. Scholten, M., and Tommassen, J. (1993) Topology of the phoR protein of *Escherichia coli* and functional analysis of internal deletion mutants. *Mol. Microbiol.*, **8,** 269-275.
30. Shi, L., and Hulett, F. M. (1999) The cytoplasmic kinase domain of phoR is sufficient for the low phosphate-inducible expression of Pho regulon genes in *Bacillus subtilis*. *Mol. Microbiol*, **31,** 211-222.
31. Fu, C., Xiong, J. and Miao, W. (2009) Genome-wide identification and characterization of cytochrome P450 monooxygenase genes in the ciliate *Tetrahymena thermophila. BMC Genomics,* **10**, 208. doi:10.1186/1471-2164-10-208.
32. Meng, Z., Wei, L. and Xia, L. (2008) Analysis of synonymous codon usage in chloroplast genome of *Populus alba. Journal of Forestry Research,* **19**, 293−297.
33. Sharp, P. M., Tuohy, T. M. F. and Mosurski, K.R. (1986) Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research,* **14**, 5125-5143.
34. Sharp, P. M. and Li, W. H. (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*., **15**, 1281– 95.
35. Mondal, S. K., Shit, S. and Kundu, S., (2013) A comparative computational study of the '*rbc*L' gene in plants and in the three prokaryotic families—Archaea, cyanobacteria and proteobacteria. *IJBT,* **12**, 58-66.