International Research Journal of Biological Sciences _____ ISSN 2278-3202
Vol. **2(12),** 90-101, December **(2013)** *Int. Res. J. Biological Sci.*

Genomic and Proteomic Properties of the Genes involved for Zinc Transportation in Firmicutes

Mondal Sunil Kanti* **and Chakraborty Papiya**
Department of Biotechnology, The University of Burdwan, Golapbag, Burdwan, 713104, West Bengal, INDIA

Available online at: www.isca.in, www.isca.me
Received 14th August 2013, revised 21st September 2013, accepted 16th October 2013

Abstract

Two types of zinc transport systems known as high affinity and low affinity on the basis of zinc availability in medium have been identified in bacteria. Genes responsible for the high affinity uptake system of zinc are ycdH, ycdI, yceA and for low affinity uptake system are yciA, yciB, yciC in firmicutes. From phylogenetic tree analysis of genes responsible for high affinity zinc uptake, Bacillus is the earliest to have evolved among the 4 genuses. Amino acids composition and the Gravy's score analysis shows that ycdH, ycdI, and yciC genes are hydrophilic amino acid rich, suggesting that they can either form a transport channel for zinc entry or can bind to zinc cation for transport, and yceA is hydrophobic amino acid rich showing its function in membrane composition. Also, all the three genes of both high affinity and low affinity uptake system are AT rich. The relationship between GC content and amino acid groups based on polarity and charge, depicts that the gene ycdH and ycdI shows similar trend but different from yceA, whereas yciA, yciB, and yciC shows similar trend in all the groups except acidic and basic polar and aromatics in yciC. For high affinity genes, a hierarchical clustering based on amino acid frequencies of the proteins encoded by the genes, the GC3 content and RSCU values of these genes, shows that all the organisms under a particular genus falls under same cluster, supporting their taxonomical lineage. ENc plot shows that all the genes involved in high affinity system for zinc uptake are under mutational bias except few ycdI and yceA from Listeria whereas the genes involved in low affinity zinc uptake system are under selectional bias except few yciA genes. Correspondence analysis shows that ycdH and ycdI follows similar pattern and yceA follows pattern which is opposite to both ycdH and ycdI whereas yciA and yciC follows similar pattern but yciB is different.CAI values predicts that the degree of expression of the genes for high affinity system from Bacillus and Paenibacillus to be high but from Enterococcus and Listeria to be low and expression of low affinity genes is high, except from Staphylococcus and few sub-species of Bacillus subtilis.

Keyword: RSCU: Relative Synonymous Codon Usage, GC3: GC% at third position of a codon; CAI: Codon Adaptation Index, Enc: Effective number of codon.

Introduction

Prokaryotic organisms are unicellular and simple in structure having no sub-cellular compartments. Since a single membrane separates the cell from its environment, cellular zinc concentration depends solely on zinc import, sequestration by metallo-chaperons and export[1-3]. Each metal ion has one or more high or low-affinity uptake systems for zinc uptake, selective for their target metal ions, which is tightly regulated according to metal ion requirements. This homeostasis is maintained in either metal limitation or excess. *Bacillus sp.* encodes three known zinc uptake systems: An ABC transporter encoded by the Zur-regulated ycdHI-yceA operon, a postulated low-affinity transport system encoded by the yciABC operon[5], and the P-type ATPase ZosA regulated by the PerR protein[6]. Here we have considered only the high and low affinity transport systems. Among the low affinity transport systems, yciA encodes GTP cyclohydrolase I (GCYH-I), the first enzyme of the de novo tetrahydrofolate biosynthetic pathway[7] which can substitute, under zinc limiting conditions for the Zn dependent FolE protein. yciB encodes a metal uptake system lipoprotein i.e., it is a transporter and its function is to uptake zinc[8]. And yciC encodes for a metallochaperon with NTPase activity and is postulated to allow metal insertion into the yciA protein[9]. Similarly, in high affinity systems, ycdH encodes a zinc ABC transporter.ABC transporters are energy (ATP) dependent active transporters which function to translocate substrate across membrane by driving conformational changes in the TMD (Transmembrane Domain) by using energy of ATP binding or hydrolysis. ycdI encode for a zinc transporter and yceA encodes for a Zinc ABC transporter permease. YciA, YciB, and YciC proteins may function as part of the same Zn(II) transport pathway and they are not related to any known transporter family, so they may define a new class of metal ion uptake system[9]. The aim of this study is to find out the following properties for the genes and proteins involved in zinc uptake by bacteria: Base composition of genes and amino acid composition of proteins, codon usage pattern and biasness, mutational pressure on genes, amino acid usage of proteins, hydrophobicity of proteins and their biological implications, comparison between genes and calculation of few more indices,

correspondence analysis, adaptability of the genes and gene expressibility.

Material and Methods

Collection of Data: Taxonomic information, nucleotide sequences and amino acid sequences of the six genes involved in the high affinity (table - 1a) and low affinity (table - 1b) system for zinc uptake and their 16S rRNA sequences were collected for 60 species (all Firmicutes) from KEGG database (www.genome.jp/kegg-bin/show_organism), and their cDNA sequences downloaded from PATRIC2 sequence (www.brcdownloads.vbi.vt.edu/patric2/genomes/).

Table-1a

Name and codes (HC*: Codes used in hierarchical cluster; PC#: triple/four letter code which has been used in phylogeny of the organisms from keg; T$: Codes used in place of lineages) of the organisms selected to study the high affinity system

HC*	PC#	T$	Name	HC*	PC#	T$	Name
1	bao	B : Bacillus	*Bacillus amyloliquefaciens* DSM 7	23	lms	L: Listeria	*Listeria monocytogenes* Finland 1998
2	baz		*Bacillus amyloliquefaciens* TA208	30	lmot		*Listeria monocytogenes* SLCC2540
3	bql		*Bacillus amyloliquefaciens* LL3	38	lmos		*Listeria monocytogenes* SLCC7179
4	bxh		*Bacillus amyloliquefaciens* XH7	47	lmoc		*Listeria monocytogenes* SLCC5850
6	bamp		*Bacillus amyloliquefaciens subsp. plantarum* AS43.3	48	lmt		*Listeria monocytogenes* 10403S
8	baq		*Bacillus amyloliquefaciens subsp. plantarum* CAU B946	41	lmz		*Listeria monocytogenes* serotype 7 SLCC2482
10	baml		*Bacillus amyloliquefaciens subsp. plantarum* UCMB5036	26	lmoa		*Listeria monocytogenes* ATCC 19117
7	bqy		*Bacillus amyloliquefaciens* Y2	27	lmog		*Listeria monocytogenes* serotype 4b LL195
9	bami		*Bacillus amyloliquefaciens* IT-45	28	lmol		*Listeria monocytogenes* L312
5	bay		*Bacillus amyloliquefaciens* FZB42	29	lmoo		*Listeria monocytogenes* SLCC2378
20	bae		*Bacillus atrophaeus*	31	lmp		*Listeria monocytogenes* 07PF0776
11	bsh		*Bacillus subtilis subsp. subtilis* 6051-HGW	33	lmoa		*Listeria monocytogenes* ATCC 19117
14	bsq		*Bacillus subtilis* QB928	25	lmf		*Listeria monocytogenes* F2365
15	bsu		*Bacillus subtilis subsp. subtilis* 168	32	lmg		*Listeria monocytogenes* FSL R2-561
17	bsn		*Bacillus subtilis* BSn5	35	lmoy		*Listeria monocytogenes* SLCC2479
19	bjs		*Bacillus sp.* JS	36	lmx		*Listeria monocytogenes* SLCC2372
12	bsl		*Bacillus subtilis subsp. subtilis* BSP1	58	efd		*Listeria monocytogenes* M7
16	bsy		*Bacillus subtilis subsp. subtilis* BAB-1	34	lmon		*Listeria monocytogenes* SLCC2376
18	bsr		*Bacillus subtilis subsp. subtilis* RO-NN-1	37	lmj		*Listeria monocytogenes* J0161
21	bss		*Bacillus subtilis subsp. spizizenii* W23	40	lmy		*Listeria monocytogenes* 08-5923
13	bso		*Bacillus subtilis subsp. natto* BEST195	42	lmh		*Listeria monocytogenes* HCC23
49	pjd	P@	*Paenibacillus sp.* JDR-2	61	efu	E: Enterococcus	*Enterococcus faecium* DO
44	ppm		*Paenibacillus polymyxa* SC2	59	ehr		*Enterococcus hirae*
45	ppo		*Paenibacillus polymyxa* M1	43	lmq		*Enterococcus faecium* NRRL B-2354
46	ppy		*Paenibacillus polymyxa* E681	60	efm		*Enterococcus faecalis* D32
51	cml		*Carnobacterium maltaromaticum*	54	efs		*Enterococcus faecalis* Symbioflor 1
50	lsg	L	*Listeria seeligeri*	55	ene		*Enterococcus sp.* 7L76
24	lwe		*Listeria welshimeri* SLCC5334	52	efa		*Enterococcus faecalis* V583
39	liv		*Listeria ivanovii*	53	efi		*Enterococcus faecalis* OG1RF
22	lin		*Listeria innocua*	56	efl		*Enterococcus faecalis* 62
				57	efc		*Enterococcus faecium* Aus0004

P@: Paenibacillus

Table-1b
Name and codes (PC#: triple letter code of the organisms from kegg where the codes started by 'F' are user defined) which has been used in analysis of the organisms selected to study the low affinity system.

Gene	PC#	Name	PC#	Name	Gene
yciA	F01	*Exiguobacterium antarcticum* B7	F12	*Bacillus subtilis* XF-1	yciC
	F02	*Staphylococcus simulans* ACS-120-V-Sch1	bao	*Bacillus amyloliquefaciens* DSM 7	
	F03	*Staphylococcus lugdunensis* ACS-027-V-Sch2	F13	*Bacillus amyloliquefaciens subsp. plantarum* UCMB5036	
	F04	*Staphylococcus aureus subsp. aureus* MSHR1132	F14	*Bacillus amyloliquefaciens* IT-45	
	F05	*Staphylococcus aureus subsp. aureus* CIG1165	F15	*Bacillus amyloliquefaciens subsp. plantarum* M27	
	F06	*Staphylococcus aureus subsp. aureus* T0131	baq	*Bacillus amyloliquefaciens subsp. plantarum* CAU B946	
	F07	*Bacillus sp.* BT1B_CT2	F16	*Staphylococcus aureus subsp. aureus* C101	
	F08	*Selenomonas sp.* CM52	F17	*Staphylococcus aureus* A9765	
			F18	*Staphylococcus aureus subsp. aureus* 71193	
yciB	bsq	*Bacillus subtilis* QB928	F19	*Bacillus cereus* ATCC 14579	
	bsn	*Bacillus subtilis* BSn5	F20	*Macrococcus caseolyticus* JCSC5402	
	F09	*Bacillus subtilis subsp. subtilis* str. SC-8	bsq	*Bacillus subtilis* QB928	
	bjs	*Bacillus sp.* JS	bsn	*Bacillus subtilis* BSn5	
	bae	*Bacillus atrophaeus* 1942	bjs	*Bacillus sp.* JS	
	F10	*Bacillus subtilis subsp. inaquosorum* KCTC 13429	bae	*Bacillus atrophaeus* 1942	
	F11	*Bacillus subtilis subsp. spizizenii* ATCC 6633	bss	*Bacillus subtilis subsp. spizizenii* str. W23	
	bss	*Bacillus subtilis subsp. spizizenii str.* W23	F21	*Bacillus sp.* 5B6	

Evolutionary Analysis: For the common organism set of 60 species possessing the high affinity zinc uptake system the bootstrapped (1000 times) phylogenetic tree using 16S rRNA and encoding nucleotide sequence of ycdH, ycdI, yceA were generated through Clustalw (www.ebi.ac.uk/tools/msa/clustalw2) and using PHYLIP version 3.69 [11,12,13] and Tree view [14] software. The hierarchical clustering on the basis of their amino acid frequencies, GC3% and RSCU values was created using the programme DIANA within the package cluster of R statistical software[15]. But due to unavailability of significant number of organisms having common set of genes responsible for zinc uptake in low affinity system these analysis could not be performed.

Compositional Analysis: Parameters like amino acid frequencies, GC content, and RSCU (a measure of relative synonymous codon usage biasness) values were calculated using in house PERL script and considered for compositional analysis, Gravy's score (indicator of hydrophobicity / hydrophilicity of the protein) and Nc[16] i.e., effective number of codon which will provide useful information regarding existence of mutational pressures acting on the genes[17] were generated using CodonW (http://codonw.sourceforge.net/). The expected effective number of codon i.e., Enc were calculated following equation 1, where S denotes GC3s.

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

RSCU values close to one indicates lack of biasness whereas much higher and lower values indicate preference and avoidance of those particular codons, respectively. Using codonW, the correspondence analysis[13] has been performed to investigate major trend in RSCU variation among genes and

distribute the genes along continuous axes in accordance with these trends.

Relationship between amino acid frequency and GC content: Correlation coefficient and RSQ values between amino acid frequencies and GC% of all genes have been calculated using MS Excel 2007.

Expressional probability: The relative adaptiveness of each codon is the ratio of the usage of each codon, to that of the most abundant codon for the same amino acid. The geometric mean of these relative adaptiveness values is known as Codon Adaptatation Index (CAI) i.e. the measure of gene's probable expression. We have calculated it by following Sharp and Li method [18] and using in house PERL script and MS Excel 2007.

Results and Discussion

Evolutionary Analysis: From the phylogenetic tree[19-24] on the basis of 16S rRNA (figure-1a), it has been observed that *Enterococcus* and *Paenibacillus* are more closer to each other whereas *Enterococcus* and *Listeria* are closer to each other as observed in the phylogenetic trees (figure - 1b-d) on the basis of ycdH, ycdI and yceA. Bacillus tends to form a distant cluster in all except ycdH. Most importantly, in all the cases, *Bacillus* seems to be the earliest, whereas other genus evolved eventually under mutational circumstances.

Also the three genes ycdH, ycdI, and yceA were hierarchically clustered on the basis of their amino acid frequencies, GC3% content and their RSCU values as shown in figure - 2a-i, and we have observed that all the organisms under a particular genus falls under same cluster, supporting their taxonomical lineage.

Analysis of Average and Standard Deviations of amino acid frequencies of ycdH, ycdI, and yceA : There are 20 amino acids, but for better understanding of their effect on different genes and organisms, we have classified the 20 amino acids in five groups based on their polarity and charge, i.e., Acidic Polars (D,E), Basic Polars (H,K,R), Aromatics (F,Y,W), Neutral non-polars (P,C,M,G,A,V,I,L), and Neutral polars (Q,N,S,T).

From table – 2, in ycdH, acidic polars maintain an overall balance among the different genus. For example, D in *Enterococcus* is low which is compensated by high E. Of the basic polar groups, K is highest, may be due to its simple structure, and therefore less hindrance. In *Bacillus* it has high frequency and variation. In aromatics, Y being the polar one is highest among all (except Paenibacillus). Neutral non-polars are the most oftenly occurred ones with A being the highest in all the lineages, specially in *Paenibacillus.* A trend has been observed, that is, those having hydrocarbon side chain are most frequent, of these A has the simplest hydrocarbon chain so it appears most frequently. For neutral polars, in *Bacillus* S is the highest and also S and T dominates for most genus. For ycdI, acidic polars are high in frequency with less variation, being highest in *Enterococcus* and lowest in *Paenibacillus*. Overall E is highest in freq. In basic polar groups, here also K is highest among all in frequency and variation. In contrast to ycdH, the frequency of R is high here, with highest in *Listeria*. In aromatics, F is highest among all the genus with highest in *Listeria* then *Enterococcus*. Among neutral non polars, L is extremely high in all the lineages, with the highest in *Paenibacillus*. G also is present high proportion showing protein flexibility. High variation is seen in L, V, G. In neutral non-polars S is the highest for all genus, specially *Enterococcus*. After S, T has some highest numbers. In gene yceA, D and E has very low frequency (expect in *Paenibacillus*) and balance each other. Basic polar groups are also very less in frequency, except R. In aromatics, F shows less variation and high frequency with the highest in *Enterococcus* and *Bacillus*. Neutral non-polars are extremely high specially K (except in *Paenibacillus)*. V shows very high variation and frequency of V, I, and A are similar. Among neutral polars, S is highest in frequency followed by T, with *Bacillus* showing highest in both. Overall, in yceA, the hydrophobic amino acids are extremely high in frequency as compared to ycdH and ycdI. It proves that yceA must be embedded in the membrane which goes with the role played by yceA i.e. zinc ABC transporter permease. The negatively charged amino acids are higher in ycdH and positively charged amino acids are higher in ycdI and yceA, showing that ycdH must have a role in forming the transport channel that brings the zinc (cation) inside the cell. ycdI has high G content which justifies its flexibility.

Analysis of Average and Standard Deviations of amino acid frequencies of yciA, yciB, and yciC: Table - 2 shows that acidic polars are highest in yciC. But basic polars are lowest in yciC, so the negative charge must play a role in carrying the positively charged zinc cation. Basic polars are highest in yciB, but from its functional aspect it plays role in carrying the zinc, so the positively charged amino acids must be embedded inside a globular structure which it may be predicted to possess, from the high frequency of glycine. yciA has acidic polars a little higher than basic polars and has high frequency of hydrophobic amino acids which signifies its role as a transmembrane protein is transport channel formation.

Observation on the basis of correlation coefficient and RSQ values between amino acid frequencies and GC% of ycdH, ycdI, yceA, yciA, yciB, and yciC: Within the gene ycdH as shown in figure - 3a, for acidic polars, *Bacillus* shows negative trend. *Enterococcus*, *Listeria* and *Paenibacillus* is negative for D and positive for E. In basic polars, *Bacillus* and *Enterococcus* and *Listeria* shows positive trend in R, but negative in H. *Paenibacillus* shows just opposite variation. In aromatics, *Bacillus* and *Enterococcus* shows negative trend. *Listeria* shows positive trend in F and W and slight negative in Y. *Paenibacillus* shows very high variation in Y and W and very low in F. Among neutral non-polars, in *Bacillus* G and I are highly positive and V is highly negative. In neutral non-polars, *Enterococcus* shows high positive trend, except for M. *Listeria*

shows balanced trend whereas *Paenibacillus* is more positive except for A and P. Finally in neutral polars, *Bacillus* shows a negative trend, except for S. *Enterococcus* shows positive trend except, for N. *Listeria* shows positive trend and *Paenibacillus* shows balanced trend with slightly on higher side. For gene ycdI (figure - 3b), acidic polars, D shows highly negative trend for *Bacillus, Paenibacillus* and *Listeria* and positive for *Enterococcus.* E is negative for Enterococcus and *Paenibacillus*. Among basic polars, *Bacillus* shows positive trend in H and R but highly negative in K, *Enterococcus* shows negative and highly negative in H and K and highly positive in R. *Listeria* has very low frequency and *Paenibacillus* shows negative trend in all. In aromatics, *Bacillus* shows highly positive trend in Y, and slightly negative in F and W. *Enterococcus* shows highly negative in F and Y and highly positive in W. *Listeria* is almost balanced. *Paenibacillus* is highly negative in F then positive in Y and W. In neutral non-polars, *Bacillus* shows positive trend in M, G and A, and highly negative in L and I. *Enterococcus* shows positive trend in C, M, G and A, and negative in P and I, *Listeria* shows average increasing trend. *Paenibacillus* shows positive trend in P, A and negative in G and V. Finally among neutral polars, *Bacillus* shows very high trend in Q and N and very low in S. *Enterococcus* shows very high trend in Q and S, and very low in T. *Paenibacillus* shows very high in Q, N and S and very low in T. In yceA (figure - 3c), acidic polars balance each other but is present in extremely low frequency! Comparatively, *Paenibacillus* has higher frequency. Basic polar groups are also very low in frequency, but R has comparatively high frequency. In aromatics, less variation is seen. F shows very high frequency compared to other two, with the highest in *Enterococcus* and *Bacillus.* Neutral non-polars seems are most frequent, but extremely high especially K (except *Paenibacillus)*. V shows very high variation and frequency of V, L, I are approximately similar. Finally in neutral polars, S is highest in frequency followed by T, with *Bacillus* showing highest in both. Therefore, in gene ycdH and ycdI, acidic polar groups, basic polar groups and aromatics shows negative trend and neutral non-polar and polars shows positive trend, but in yceA it is different. Also from figure - 3d, yciA and yciB gene shows highly negative trend whereas yciC shows positive trend for acidic polar groups, which is just opposite for basic polar groups. For aromatics, yciA and yciB is highly negative but yciC slightly positive. In neutral non-polar, all shows positive trend whereas in neutral polar all shows negative trend.

Analysis on the basis of gene GC1, GC2, GC3, and total GC%: Figure - 4a shows that GC%[25] in first position dominates in all the three genes involved in high affinity system for zinc uptake from all the selected genus, with *Paenibacillus* being the highest. GC% at 3rd position is highest in *Bacillus* among the four genuses. And since the GC% is less than 50% in all the genes for all the groups, we can say that the genes are AT rich. On average, GC in 2nd position is the least in all the genes.

From (figure - 4b), genes yciA, yciB and yciC i.e., the genes involved in low affinity system for zinc uptake are AT rich. GC% at 1st position is highest in yciC gene. GC% in 2nd position is highest in yciB and GC% in 3rd position is also highest in yciC. Overall GC% is highest in yciC.

Codon preference check: The ycdH, ycdI, and yceA gene's RSCU values as shown in figure – 5a-c also lead to the establishment of a codon preference model, which thereby shows preference towards GC- ending codons by the neutral non-polars (C in ycdH and ycdI, V in ycdI and yceA), and S and Q among neutral polars. Preference of C3 over G3 is observed here, this preference is more pronounced in ycdI gene, in yceA gene, G3 and C3 preference is equal. D, F, G, H, I, N, and Y have preference for C3 in all the genes whereas E, K, L, M, P, Q, and W have preference for G3 in all. And for genes yciA, yciB, and yciC (figure – 5d), preference of G3 over C3 is prevalent in yciA, yciB, and also yciC. So we can conclude a difference between the high and low affinity zinc uptake genes, that at 3rd position, C dominates for ycdH, ycdI, and yceA, whereas G dominates in yciA, yciB, and yciC.

Enc plot analysis: For genes responsible for zinc uptake by high affinity system, the ENc plot analysis (figure – 6a) was used to investigate patterns of synonymous codon usage, which shows that all the organisms lie below the expected curve thus are under mutational bias but some of the ycdI and yceA genes of *Listerias* lie above the curve showing that they are under selectional bias. Similarly, for genes responsible for zinc uptake by low affinity system, ENc plot (figure - 6b) shows that most of the organisms lie above the expected curve showing that they are under selectional bias, except yciA is equally mutationally biased.

Correspondence analysis: To determine the codon usage of ycdH, ycdI, and yceA among the genuses, correspondence analysis on the genes RSCU values was carried out by a standard procedure[26]. The distribution of the three genes from the four genuses on the first two major axes of the correspondence analysis shows that genes are recognized based on their genuses. In gene ycdH and ycdI (figure - 7a), *Listeria* and *Bacillus* are separated by the 2nd major axis having *Listeria* in the right hand side and *Bacillus* in the left hand side of the 2nd major axis which shows that they follow similar pattern, whereas, in gene yceA (figure - 7a), *Bacillus* lies on the right hand side of the 2nd major axis showing that it follows a trend against both ycdH and ycdI. Whereas, codon usage pattern of yciA, yciB and yciC of the low affinity system (figure - 7b) shows that, yciA and yciC are distributed along the major axis 1 and yciB is distributed along major axis 2. Also, neither of the genes seem to cluster at a particular point.

Codon adaptation index (CAI): For ycdH, ycdI and yceA (figure - 8a), CAI predicting the degree of expression of the three genes involved in zinc transport by high affinity zinc uptake system shows a high rate of expression of 0.75 to 0.85 in *Bacillus* expect in bae and *Paenibacillus*. Low in *Listeria* (especially yceA is very low), and moderate towards low in

Enterococcus with very high variation. For yciA, yciB and yciC (figure - 8b)**,** CAI predicts a high rate of expression of 0.72 to 0.76, indicating high expression except the species of *Staphylococcus* in yciA, 2 subspecies of *Bacillus subtilis* in yciB and 2 species of *Bacillus*, 1 species of *Macrococcus* and all *Staphylococcus* in yciC.

Gravy's Score: Finally from the gravy's score as shown in figure – 9a, we get a clear indication that yceA is hydrophobic since it has positive gravy's score for all the organisms. And ycdH and ycdI are hydrophilic since they have negative gravy's score for all the organisms. Whereas yciA, yciB, and yciC, all have negative gravy's score as shown in figure – 9b therefore all are hydrophilic.

Table–2
Average and standard deviation (within bracket) of the amino acids of the proteins encoded by ycdH, ycdI, yceA, yciA, yciB and yciC

Gene/Amino acid 1lt code		D	E	H	K	R	F	Y	W	P	C
ycdH	B	6.8(0.3)	9.1(0.6)	3.6(0.3)	11.7(0.6)	0.6(0.3)	3(0.2)	4(0.2)	1.3(0.1)	3.9(0.2)	0.7(0.1)
	E	5.3(1.6)	11(0.9)	2.7(0.4)	9.3(0.9)	1.6(0.3)	3.2(0.5)	4.4(0.8)	0.7(0.3)	4.1(0.5)	0.4(0.1)
	L	7.3(0.3)	10.2(0.3)	3.6(0.2)	10.5(0.5)	0.7(0.2)	3.2(0.2)	3.9(0.1)	1(0.1)	3.6(0.1)	0.4(0.1)
	P	7.1(0.3)	10(0.8)	2.7(0.6)	9(1.2)	0.7(0.2)	3.3(0.4)	2.5(0.1)	1.3(0.1)	4.3(0.1)	0.4(0.1)
ycdI	B	4.6(0.4)	7.7(0.4)	3.2(0.2)	8.1(0.7)	4.9(0.3)	3.9(0.1)	2.4(0.3)	2.2(0.1)	3.1(0.1)	1.3(0.1)
	E	6.3(0.4)	8.1(0.4)	4.3(0.3)	7.2(0.9)	6.8(1.4)	4.3(0.2)	3.8(0.2)	1.4(0.1)	3.7(0.4)	0.5(0.1)
	L	6.1(0.5)	7.1(0.4)	3.3(0.3)	9.3(0.3)	5.6(0.3)	5.1(0.2)	2.2(0.2)	1.3(0.1)	3.6(0.3)	1.3(0.1)
	P	4.8(0.3)	7(0.3)	3(0.6)	4.9(0.3)	6.1(0.3)	4.1(0.5)	2.1(0.1)	1.7(0.1)	2.9(0.4)	2(0.2)
yceA	B	1.5(0.1)	2.1(0.3)	0.4(0.2)	2.8(0.3)	4.5(0.3)	6.4(0.3)	3.3(0.1)	0.6(0.2)	2.6(0.1)	0(0)
	E	1.9(0.2)	1.8(0.2)	1.1(0.1)	1.8(0.3)	3.6(0.5)	6.7(0.8)	3(0.4)	0.8(0.4)	2.9(0.1)	0.2(0.2)
	L	1.1(0.2)	1.9(0.1)	0.4(0.1)	1.9(0.2)	3(0.1)	6(0.5)	2.3(0.1)	0(0)	2.7(0.1)	0.4(0.1)
	P	1.2(0.2)	2.5(0.3)	1(0.4)	2.8(0.3)	4.3(0.6)	5.5(0.3)	3.2(0.1)	0.5(0.2)	3(0.2)	0.3(0.2)
yciA		6.6(1.4)	9.6(0.8)	2.8(1.1)	7.1(1.2)	4.9(0.5)	4.5(0.5)	3.6(0.9)	1.1(0.2)	3.5(1.1)	1.5(0.8)
yciB		5.4(0.6)	8.5(1.5)	4.2(1)	11.6(1.2)	2(1.1)	2.5(1.4)	3.3(1)	2.7(0.9)	4.1(0.4)	1.8(1.4)
yciC		7.5(0.9)	10.2(0.7)	1.8(0.4)	4.8(0.8)	4.6(0.6)	3.8(0.5)	1.9(0.4)	2(0.2)	3.5(0.5)	1.9(0.4)
Gene/Amino acid 1 lt code		M	G	A	V	I	L	Q	N	S	T
ycdH	B	2.2(0.2)	5.9(0.7)	8.8(0.4)	6.3(0.3)	5.8(0.9)	8.3(0.2)	3.3(0.5)	3(0.4)	8(0.6)	4.8(0.3)
	E	1.9(0.4)	4.8(0.8)	9.6(0.9)	7(0.6)	5(0.5)	8.9(0.7)	4.1(1.2)	4.2(0.9)	6.1(0.5)	6.7(1)
	L	2(0.1)	4.6(0.3)	9.9(0.7)	7.6(0.4)	4.7(0.3)	8(0.2)	3.9(0.2)	4.2(0.5)	5.3(0.3)	6.6(0.6)
	P	1.8(0.4)	6(0.2)	12.7(0.6)	7.8(0.4)	3.7(0.6)	8.7(0.3)	3.8(0.9)	5.3(1)	4.7(1)	5.3(1.3)
ycdI	B	4.2(0.3)	8.5(0.3)	3.3(0.7)	8.1(0.4)	4.6(0.3)	9.8(0.7)	4.2(0.7)	4.7(0.4)	6.3(0.3)	6.1(0.2)
	E	2.8(0.3)	6.5(0.6)	5.2(0.6)	6(0.5)	6.5(0.2)	8.4(0.2)	3.9(0.5)	2.6(0.2)	7.5(0.1)	5.2(1.3)
	L	3.1(0.1)	6.8(0.2)	6.4(0.3)	5.3(0.5)	7.6(0.7)	9.1(0.3)	4.1(0.5)	2.9(0.5)	6.5(0.1)	4.2(0.3)
	P	4.1(0.5)	9.8(1.3)	5.5(1.3)	7.4(0.8)	4.3(0.4)	10.4(0.3)	5.8(0.4)	2.4(0.7)	7(1.6)	5.7(0.7)
yceA	B	4.1(0.4)	7.6(0.8)	9.6(0.2)	7.3(1.1)	12.7(0.4)	14.9(0.3)	2.7(0.3)	2.2(0.1)	9(0.5)	6.4(0.6)
	E	6.7(1.2)	6.9(0.5)	9.8(0.7)	9.9(0.3)	11.5(0.1)	14.7(0.5)	2(0.8)	3.3(0.2)	7(0.3)	5.4(0.5)
	L	5(0.4)	8.3(0.1)	10.2(0.4)	12.7(0.6)	12.6(0.5)	15.3(0.2)	3.4(0.2)	0.8(0.2)	7.9(0.3)	5(0.2)
	P	5.1(0.5)	8.7(0.2)	12.5(0.2)	10(1)	10.3(0.8)	12.3(1)	2.5(1.1)	2.3(0.4)	8.3(1.6)	5(0.2)
yciA		2.3(0.7)	4.3(0.9)	7.3(2.4)	6.4(1.3)	6.2(0.8)	7.8(1)	3.4(0.9)	4.4(1.4)	6.2(1)	7.5(1.9)
yciB		2.1(0.6)	6(1.1)	7.7(3.2)	6.5(0.9)	8.1(1.1)	5.4(2.1)	2.8(0.8)	2.4(0.8)	7.2(1.3)	6.6(2.2)
yciC		2.2(0.5)	5.7(0.6)	5.9(1.1)	6.7(0.7)	7.4(1.1)	10.8(0.6)	4.4(0.8)	4(1.4)	6.2(0.5)	5.6(0.5)

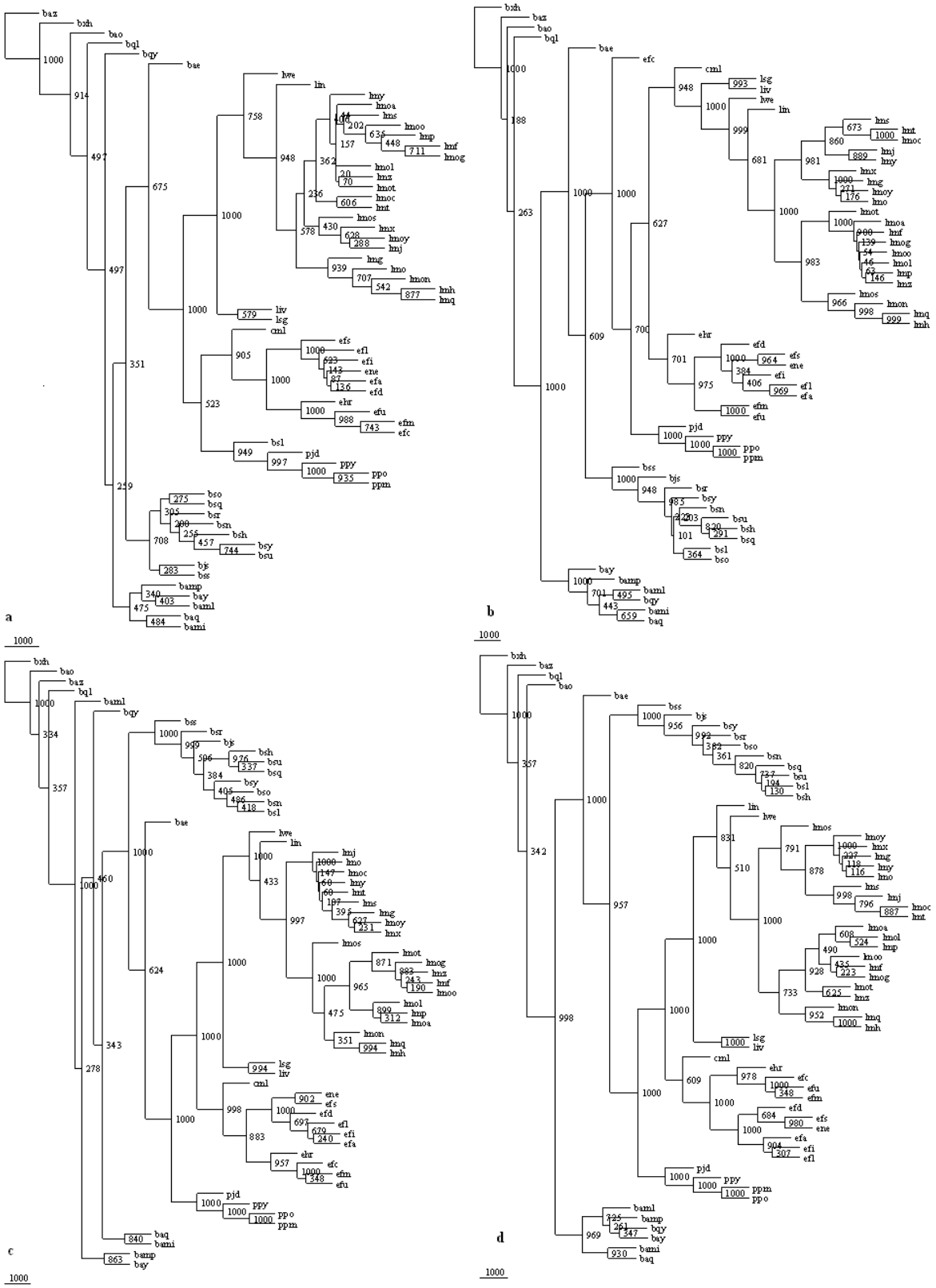


Figure-1

1000 times bootstrapped phylogenetic tree based on gene sequences of a. 16S rRNA, b. ycdH, c. ycdI and d. yceA.

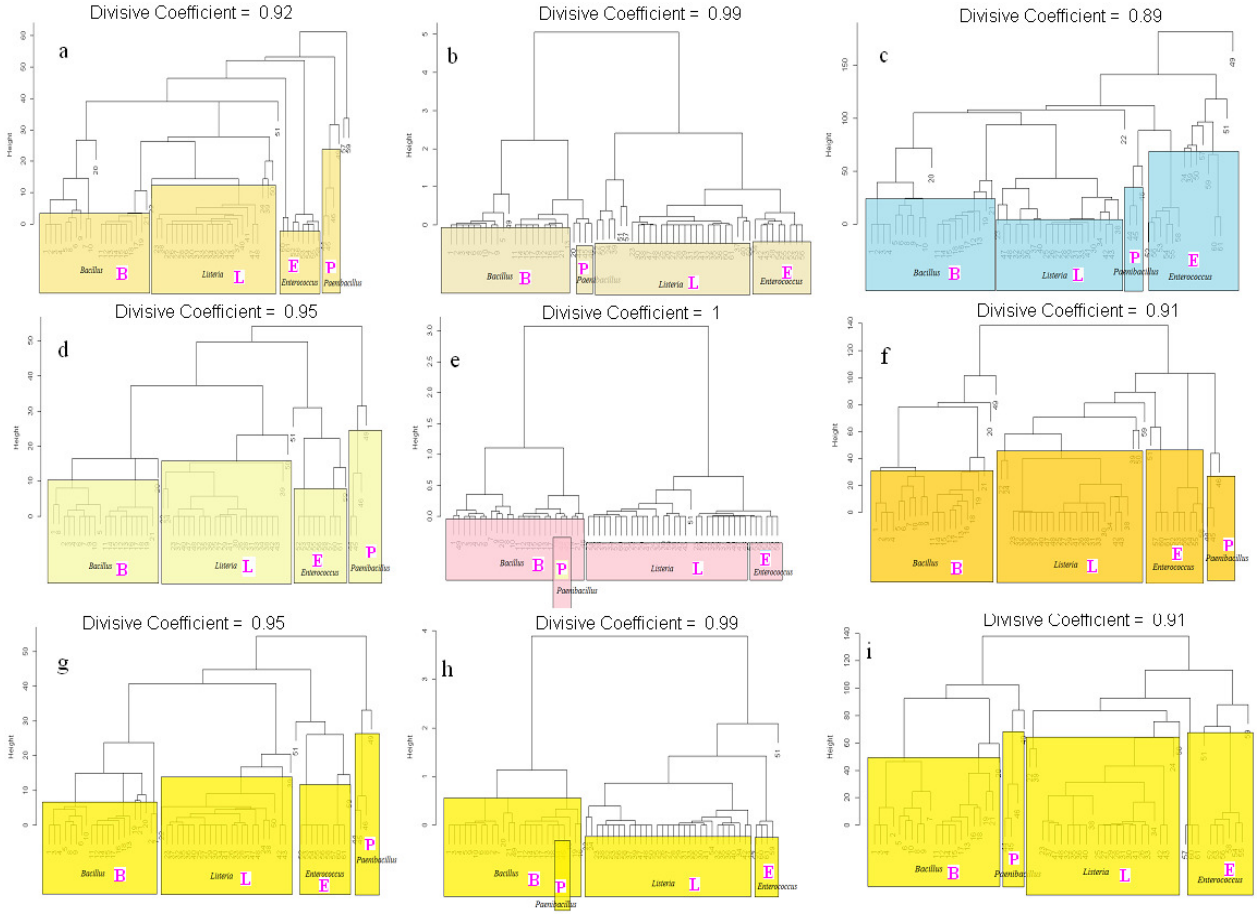


Figure-2

Dendograms based on a. amino acid frequency, b. gene GC3%, c RSCU of the codons of ycdH, d. amino acid frequency, e. GC3%, f. RSCU of the codons of ycdI, g. amino acid frequency, h. GC3%, i. RSCU of the codons of yceA

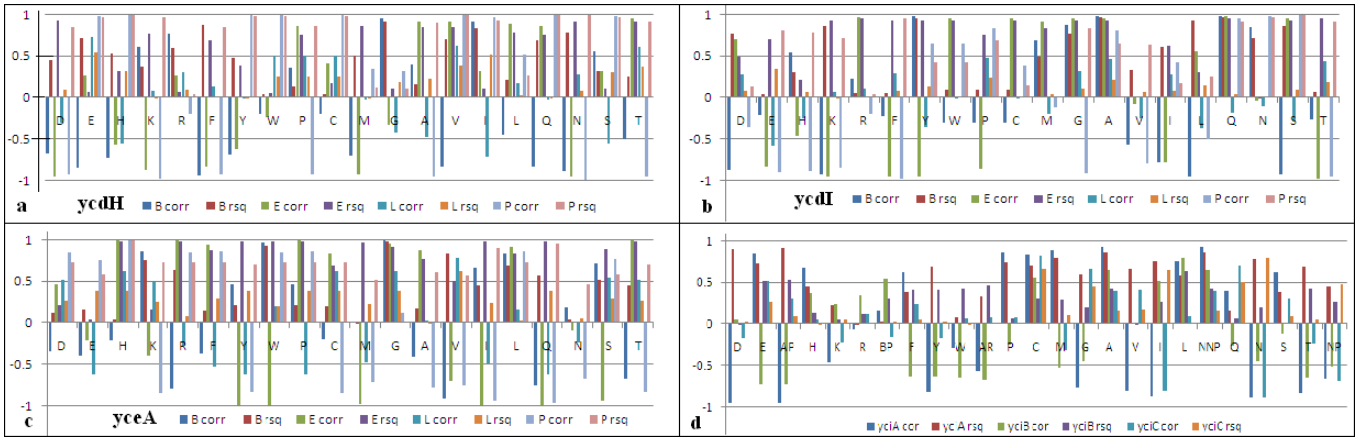


Figure-3

Correlation Coefficient and RSQ values between amino acid frequency and GC% of a. ycdH, b. ycdI, c. yceA and d. yciA, yciB and yciC

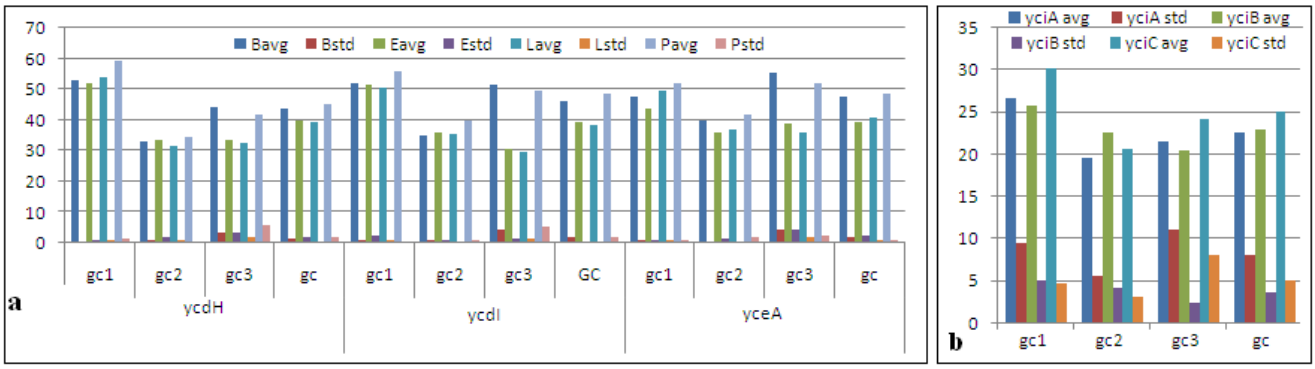
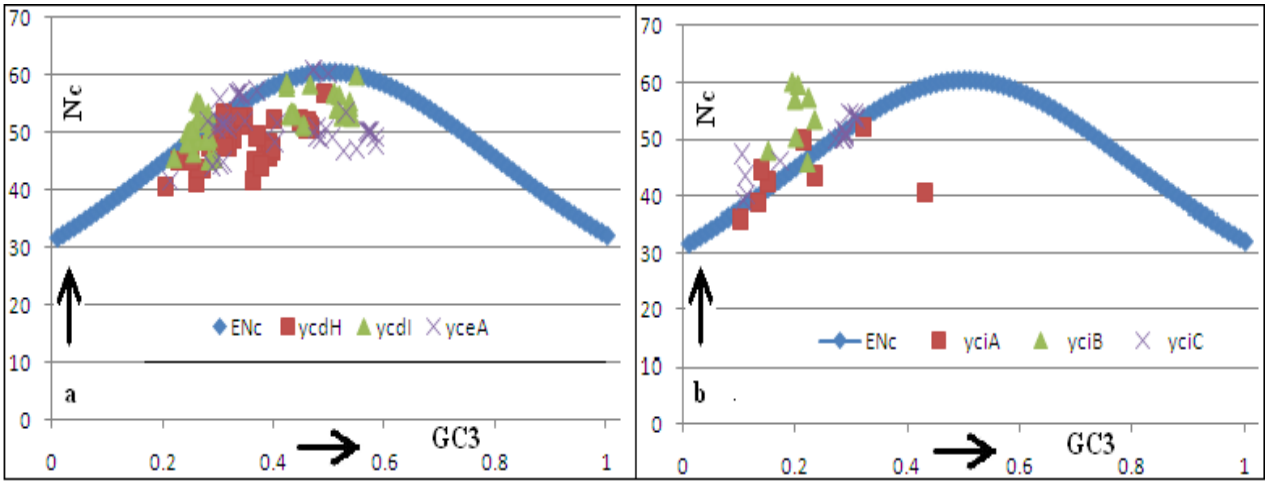


Figure-4

Average and standard deviations of GC1%, GC2%, GC3% and GC% of a. ycdH, ycdI and yceA, b. yciA, yciB and yciC

Figure–5

Average and standard deviation of the codons of a. ycdH, b. ycdI, c. yceA, d. yciA, yciB and yciC



Figure–6
ENc plot (Nc vs. GC3%) of a. ycdH, ycdI and yceA,, b. yciA, yciB, and yciC

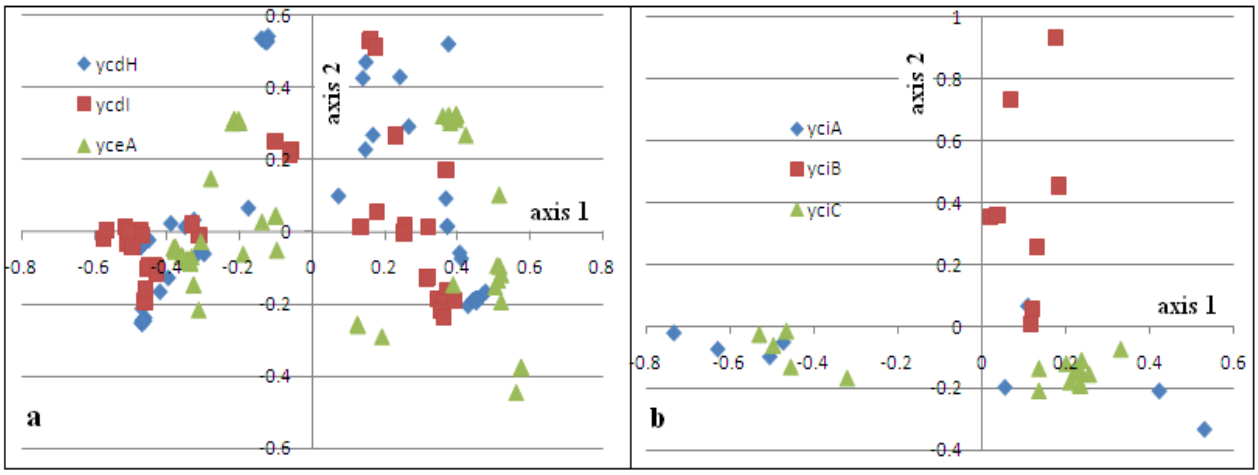


Figure-7
Correspondence analysis on the basis of the codon usage of a. ycdH, ycdI and yceA,, b. yciA, yciB, and yciC

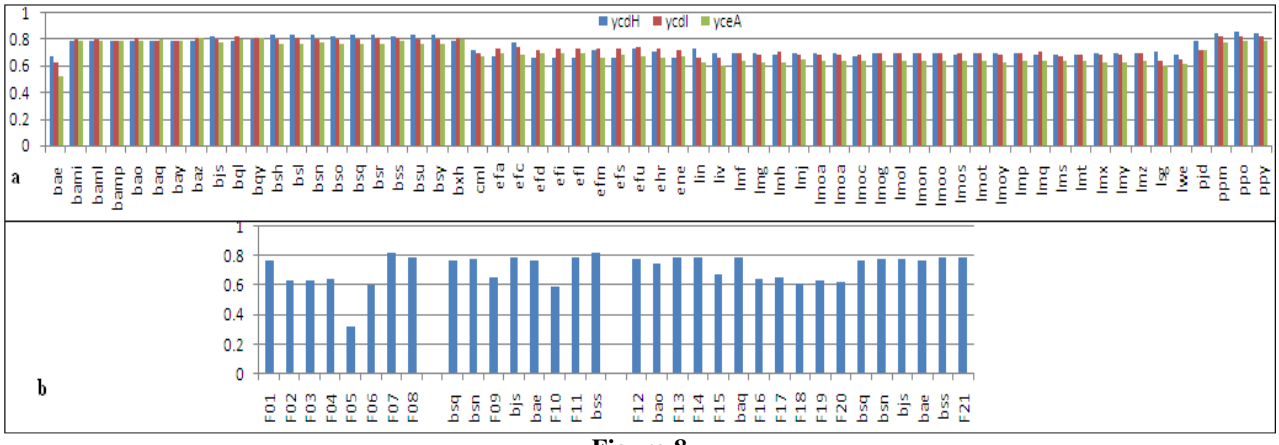
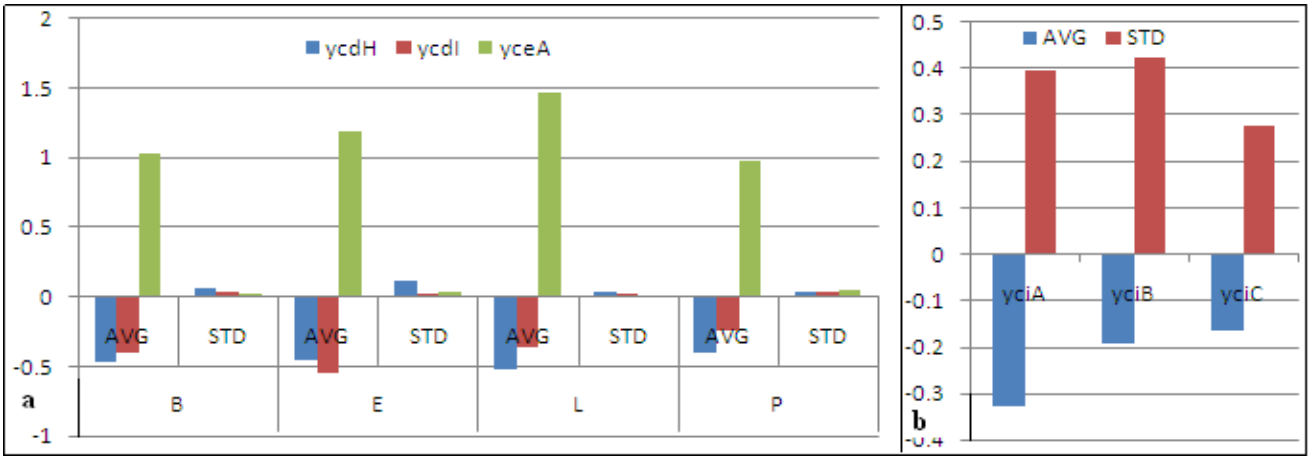


Figure-8
Representation of the CAI values for the a. ycdH, ycdI, and yceA genes from four genus, b. yciA, yciB and yciC genes from the genus firmicutes



Figure–9
Average and standard deviation of Gravy score of a. ycdH, ycdI and yceA,, b. yciA, yciB, and yciC

Conclusion

Current research paper highlights the differences and similarities between the genes responsible for high affinity and low affinity zinc uptake at the structural, compositional, and functional level justifying their biological significance. ycdH, ycdI, and yciC genes are hydrophilic amino acid rich, showing that they can either form a transport channel for zinc entry or can bind to zinc cation for transport, and yceA is hydrophobic amino acid rich showing its function in membrane composition. Also, all the genes are AT rich. The relationship between GC content and amino acid groups depicts that the gene ycdH and ycdI shows similar trend but different from yceA, whereas yciA, yciB, and yciC shows similar trend in all the groups except for the acidic and basic polars and aromatics in yciC. The taxonomical grouping is supported by the hierarchical clustering on the basis of different compositional parameters like: amino acid frequency, GC content and RSCU values. Genes involved in high affinity system for zinc uptake are under mutational bias except few ycdI and yceA genes of *Listerias* but the genes which are involved in low affinity system for zinc uptake, are under selectional bias except few yciA genes which are under mutational bias. Correspondence analysis shows that ycdH and ycdI follows similar codon usage pattern and different from yceA whereas yciA and yciC follows similar pattern and different from yciB. Degree of expression of the genes involved in high affinity system from *Bacillus* and *Paenibacillus* is high but from *Enterococcus* and *Listeria* is low and expression of low affinity genes is high, except from *Staphylococcus* and few sub-species of *Bacillus subtilis*.

References

1. Mishra P.C., Dash A.K., and Pradhan K., Metals in Environmental segments at Hirakud of Odisha, India, *ISCA Journal of Biological Sciences,* **1(1)**, 7-23 **(2012)**
2. Mohammed M.B., Mohammed S.S. and Adewumi A.A.J., Assessment of Zn Bioavailability in Dumpsites of Kaduna Metropolis,Nigeria, *Research Journal of Recent Sciences,* **1(12)**, 21-24, December **(2012)**
3. Francis A.R., and Masilamai D., Removal of Zinc (II) by Non Living Biomass of Agaricus Bisporus, *Research Journal of Recent Sciences* **1(9)**, 13-17, September **(2012)**
4. Blencowe DK, and Morby AP Zn(II) metabolism in prokaryotes, *FEMS Microbiol Rev,* **27**, 291–311**, (2003)**
5. Gaballa A, and Helmann J D., Identification of a zinc-specific metalloregulatory protein, Zur, controlling zinc transport operons in Bacillus subtilis, *J. Bacteriol*, **180**, 5815-5821, **(1998)**
6. Gaballa A, and Helmann J D, A peroxide-induced zinc uptake system plays an important role in protection against oxidative stress in Bacillus subtilis, *Mol. Microbiol*, **45**, 997-1005, **(2002)**
7. El Yacoubi B, Bonnett S, Anderson JN, Swairjo MA, Iwata-Reuyl D, and de Crécy-Lagard V, Discovery of a new prokaryotic type I GTP cyclohydrolase family, *J Biol Chem*, **281(49)**, 37586-93, **(2006)**
8. http://subtiwiki.uni-goettingen.de/wiki/index.php/YciB **(2013)**
9. Gaballa A, Wang T, Rick W. Ye, and Helmann J D, Functional Analysis of the Bacillus subtilis Zur Regulon, *J Bacteriol* , **184(23),** 6508–6514 **(2002)**
10. Hollenstein K, Dawson RJ, and Locher K.P., Structure and mechanism of ABC transporter proteins, *Curr. Opin. Struct. Biol* , **17 (4)**, 412–8 **(2007)**
11. Felsenstein J., PHYLIP: Phylogeny interference package (version 3.69) Department of Genome Sciences and Department of Biology, University of Washington, Washington, USA), 164-166 **(1989)**

12. Swofford D. L., Olsen G. J., Waddell P. J. and Hillis D. M., Phylogenetic inference, In D M Hillis, C Moritz & B K Mable (Eds.), *Molecular systematics*, Sunderland, USA: Sinauer Associates, Inc., Publishers, 2nd edn, 407-514 **(1996)**

13. Mondal S. K., Shit S. and Kundu S., 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 **(2013)**

14. Saldanha A. J., Java Treeview-extensible visualization of microarray data. *BIOINFORMATICS APPLICATIONS NOTE*, **20(17)**, 3246–3248 **(2004)** doi:10.1093/bioinformatics/bth349.

15. Kaufman L & Rousseeuw P J, Finding groups in data : An introduction to cluster analysis, (*John Wiley and Sons, Inc., New Jersey, USA*), **(1990)**

16. Fu C, Xiong J and Miao W, Genome-wide identification and characterization of cytochrome P450 monooxygenase genes in the ciliate Tetrahymena thermophila, *BMC Genomics,* **10,** 208, **(2009)**

17. Meng Z, Wei L and Xia L, Analysis of synonymous codon usage in chloroplast genome of Populus alba, *J for Res* **19,** 293-297, **(2008)**

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

19. Kumar S, Lingaiah K, Ramachandra N.B., and Nair M V., Genetic variations among Ecologically diverse species of Anurans at the level of Genus based on ISSR Marker, *International Research Journal of Biological Sciences* **1(7)**, 11-19, November **(2012)**

20. Dwivedi VD, Sharma T, Mishra S.K., and Pandey A.K., Insights to Sequence Information of Lactoylglutathione Lyase Enzyme from Different Source Organisms, *International Research Journal of Biological Sciences,* **1(6)**, 38-42, October **(2012)**

21. Sharma A., and Sharma P., Genetic and Phytochemical analysis of Cluster bean Cyamopsis tetragonaloba (L.) Taub) by RAPD and HPLC, *Research Journal of Recent Sciences,* **2(2)**, 1-9, February **(2013)**

22. Maithri S.K., Ramesh K.V., and Muntanga D, Theoretical structure prediction of TcaA from Photorhabdus luminescens and aminopeptidase receptor from Helicoverpa armigera, *Research Journal of Recent Sciences,* **2(2)**, 40-49, February **(2013)**

23. Bhattacharya A, Power J.B., and Davey M. R., Genetic Manipulation of Gibberellin (GA) Oxidase Genes in Nicotiana sylvestris using constitutive promoter to modify Plant Architecture, *Research Journal of Recent Sciences,* **1(5)**, 1-7, May **(2012)**

24. Maithri S.K., Ramesh K.V., Dieudonné M, and Deshmukh S., Molecular Modeling and Docking Studies of PirB Fusion Protein from Photorhabdus Luminescens, *International Research Journal of Biological Sciences,* **1(8)**, 7-18, December **(2012)**

25. Kamaraj M., Jansi L., Sivaraj R., Sama K., Salam H.A. and Rajiv P., Gas Chromatographic and UV-VIS spectrometric analysis of Bisphenol-A degradation in garden soil collected from Coimbatore district,Tamil Nadu, India, *International Research Journal of Biological Sciences,* **1(8)**, 54-60, December **(2012)**

26. Sharp P.M., Tuohy T.M.F. and Mosurski K.R., Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes, *Nucleic Acids Res,* **14,** 5125-5143, **(1986)**