International Journal of Pharma and Bio Sciences

EXPLORATION OF CODON USAGE PATTERNS IN SOME *BRUCELLA* GENOMES

ARVIND K GOYAL[1], ARNAB SEN*[1], SAUBASHYA SUR[1] AND ASIM K BOTHRA[2]

[1] NBU Bioinformatics Facility, University of North Bengal, Siliguri-734013, West Bengal, India.
[2] Bioinformatics Cheminformatics Laboratory, Department of Chemistry, Raiganj College, Raiganj-733134, West Bengal, India.

***Corresponding author** senarnab_nbu@hotmail.com

ABSTRACT

Comparative analysis of codon usage patterns in some *Brucella* strains were performed to predict expression levels for protein coding genes, find out horizontally transferred pathogenesis related genes, investigate patterns of pathogenesis related genes with respect to expression levels and monitor involvement of predicted highly expressed genes with lifestyle of *Brucella* strains. Selection for translational efficiency plays a major role in codon usage variation. Codon bias is also strongly influenced by GC3 compositional constraints. Thirty-five PHX (potentially highly expressed) genes related to pathogenicity have also been identified in the seven strains. High number of PHX genes associated with the metabolism COG group divulges that metabolic genes has an important part to play in effecting the survival of the bacteria against the action of host's resistance, antibiotics etc. thus establishing infection. Pathogenicity related homologs reveal that they help them to protect from the selective pressure of evolution.

KEY WORDS

Brucella, codon bias, COGs, correspondence analysis, pathogenicity, potentially highly expressed (PHX) genes.

ABBREVIATIONS

CAI, Codon Adaptation Index; COGs, Cluster of Orthologous Groups of genes; Fop, Frequency of Optimal codons; HGT, Horizontal Gene Transfer; Nc, Effective Number of codons; PHX, Potentially highly expressed genes.

INTRODUCTION

Brucella is a small gram negative, pathogenic bacteria measuring 0.6 to 1.5 µm by 0.5-0.7 µm belonging to the α-2 subdivision of proteobacteria causing brucellosis, a true zoonotic disease. They are non-sporing, having one polar flagellum with sheath[1]. *Brucella* mainly occurs as coccobaccilli, but coccal and bacillary forms are also found[2].

The genus is classified into ten species on the basis of its cultural, metabolic and

antigenic characteristics as: *B. melitensis*, *B. abortus*, *B. suis*, *B. ovis*, *B. canis* and *B. neotomae*[3]. *B. ceti, B. pinnnipedialis*[4], *B. microti, B. inopinata*[5] *B. melitensis*, *B. suis*, *B. abortus and B. canis* are recognized as human zoonoses causing undulant fever and a systemic, febrile illness in humans[6]. *B. melitensis* primarily affects sheep and goats, cattle is infected by *B. abortus*, *B. suis* infests pigs, *B. ovis* affects rams and ewes, male dogs and bitches are affected by *B. canis* while, *B. neotomae* infects desert wood rats[7]. Brucellosis is a true "zoonotic disease" endemic to many regions of the world, characterized by chronic infections in animals leading to abortion and infertility in wild and domestic ungulates[8]. Disease transmission is effected by consumption of unpasteurized milk and milk products and direct contact with infected animals or carcasses[9,10]. Veterinarians, slaughterhouse employees, dairy farmers and workers, livestock handlers, and laboratory personnel are more prone to brucellosis[11]. In humans brucellosis occurs by consumption of unpasteurized milk and milk products; breathing; penetration of oral or ocular mucosa; direct entry into the bloodstream through abrasions in the skin or vaccinations[12]. Virulent brucellae survive in polymorphonuclear and mononuclear phagocytes repressing chemotaxis and phagocytosis by polymorphonuclear leucocytes[13]. Non specific clinical symptoms include fever, chills, headache, pain, fatigue and arthritis. *B. melitensis* is highly infectious and is responsible for most of the cases followed by *B. suis* whose virulence rates are moderate to high. *B. abortus* is moderately virulent in humans while *B. canis* accounts for some[14]. Disease occurrence in humans increases, when brucellosis is frequent in sheep and goats[15]. Although treatment is available for brucellosis, prolonged antibiotic therapy harbors well for the patient. Early diagnosis is problematic but vaccines are available for animals[16]. Even though many countries have extensive eradication programs, brucellosis is still a serious disease challenging the veterinary and medical professions.

Complete genome sequences for *Brucella* provide opportunity for undertaking *in-silico* approaches focusing on synonymous codon usage analysis. A number of studies[17,18,19,20,21,22] exemplified the non-random species specific characteristics of synonymous codon usage in bacteria. Translation selection and/or mutational pressure play an important role in effecting codon usage bias[23]. It has been reported[24,25] that within a genome highly expressed genes are more biased than lowly expressed genes. Although gene adaptations are species-specific close similarities are detected amongst same genera of the organisms. Sharp and Li[26] postulated that translational selection governs the codon bias of highly expressed genes while mutational bias directs codon bias of lowly expressed genes. Codon usage patterns are known to be restrictive in some genes encoding abundant polypeptides[27]. A number of indices are available for measuring the extent of codon bias. These are GC (genomic GC content) content, GC3 (GC content at 3rd position of the codons) [19], Nc (Effective number of codons used in a gene) [28], Fop (frequency of optimal codons)[29], CAI (codon adaptation index)[26] etc. Although there are several other indices the aforesaid ones are preferred to extract fruitful information from the genomes.

The aim of the present work is to execute a comparative analysis of the codon usage patterns in seven strains of *Brucella*, predict expression levels of genes, detect horizontally transferred pathogenesis related genes, investigate the patterns of pathogenesis related genes to analyse their expression levels and examine association of the predicted highly expressed genes in COGs (Cluster of orthologous groups) [30] with the lifestyle of these bacteria.

MATERIALS AND METHODS

Finished sequences of seven *Brucella* strains viz. *Brucella abortus* bv. 1 9-941, *Brucella canis* ATCC 23365, *Brucella melitensis* 16M,

Brucella melitensis bv. *abortus* 2308, *Brucella ovis* ATCC 25840, *Brucella suis* 1330 and *Brucella suis* ATCC 234459 (hereafter will be referred to as BA, BC, BMM, BMA, BO, BS, BSA) were obtained from the Integrated Microbial Genomes website (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi)[31].
The protein coding genes, ribosomal protein genes and pathogenesis related genes were explored using Codon W software (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=codonw)[29] and CAI Calculator 2 (http://www.evolvingcode.net/codon/cai/cais.php) [32, 33]. The information about the pathogenesis related genes were obtained from the available literature.

The G or C at the third position of codons (GC3s), effective number of codons (Nc)[28] and frequency of optimal codon (Fop)[34] were analyzed using the Codon W software. The effective number of codons (Nc) is a simple estimate used to study the codon usage biases in genes and genomes[35] whose values lie between 20 and 61. Fop is the proportion of optimal codons to synonymous codons. Depending upon whether any optimal codon is present within a gene or not, its value varies from 0 to 1. CAI Calculator 2[32, 33] another widely used measure of codon bias is a web-based application which was employed to work out the CAI (codon adaptation index) [26] values taking the ribosomal genes as a reference set. The CAI value fluctuates in between 0 to 1.0. Higher CAI values usually advocate that the gene of concern have a codon usage pattern analogous to that in the reference genes[29].

To judge whether there is any discrepancy in the values of ribosomal protein genes and pathogenicity related genes with that of the protein coding genes Z test[36] was performed.

The information about the horizontally transferred genes of the studied strains was determined using the web server (http://cbcsrv.watson.ibm.com/HGT_SVM/Archaea+Bacteria/index.html)[37]. From the list, pathogenicity related genes were sorted out (G:\Brucella organism profile\VFDB - comparative pathogenomic composition of *Brucella*.htm). The sorted genes were subjected to IMG genome BLAST in the Integrated Microbial Genomes database to find out the sequence homologs against all the strains under study. Two criteria viz. minimum percent identity and the maximum E (expect) value were set at 90% and le-2 respectively.

In order to find the tendency of variation of codon and amino acid among the genes Correspondence analysis (COA) on codon count using Codon W software (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=codonw) was performed.

RESULTS

Codon Usage Patterns in seven Brucella strains genomes

The foremost aim of the current work on the codon usage patterns among *Brucella* genomes was to determine the intensity of heterogeneity in codon usage. Codon heterogeneity has been reported in majority of the bacteria with a considerable amount of AT/GC content in the genome[38]. It is known that gene expression level is related to codon variation. Accordingly, highly expressed genes are considered to contain a higher frequency of transitionally optimal codons[24]. To determine whether any codon heterogeneity exits among genes of the *Brucella* genomes, GC3 and Nc values for all genes in the genomes were calculated. Figure 1 shows the Nc/GC3 plot for the studied strains.

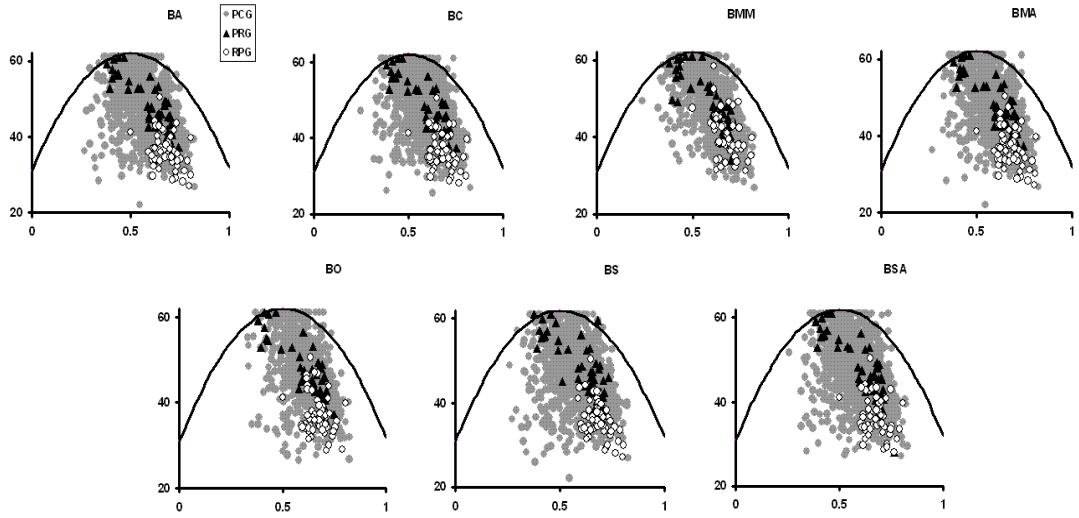

Figure 1

The effective number of codons used (Nc) (Y axis) plotted against the G+C content at the synonymous third position of codons (GC3) (X axis) for all seven Brucella strains. The curve in each plot indicates null hypothesis that the GC bias at the synonymous position is solely due to mutation but not selection. PCG= Protein Coding genes, PRG= Pathogenicity Related Genes and RPG= Ribosomal Protein Genes.

The mean values of different indices used in the aforesaid study are portrayed in Table 1. Analysis of the Fop values for the different gene sequences among different species of *Brucella* showed variation. Ribosomal protein genes had higher mean Fop values compared to protein coding genes and pathogenesis related genes. This result indicates that ribosomal protein genes have higher proportion of optimal codons. If these genes have had low Fop values, high degree of mutational bias would have been inferred.

Table 1

Mean values of Codon adaptation index (CAI), effective numbers of codons (Nc), guanine cytosine ratio at third position (GC3), guanine cytosine percentage (GC) and frequency of optimal codons (Fop) of the genes in seven Brucella strains.

STRAINS	GENES	CAI	Nc	GC3%	GC%	Fop
BA	PCG	0.530 ± 0.118	44.44 ± 6.25	64.20 ± 0.077	57.42 ± 0.404	0.331 ± 0.047
	RPG	0.550 ± 0.164	35.99 ± 4.52	68.27 ± 0.061	58.73 ± 0.025	0.423 ± 0.055
	PRG	0.514 ± 0.123	47.67 ± 6.73	60.56 ± 0.106	55.32 ± 0.052	0.344 ± 0.046
BC	PCG	0.469 ± 0.113	44.31 ± 6.13	64.39 ± 0.076	57.39 ± 0.041	0.331 ± 0.047
	RPG	0.651 ± 0.081	36.42 ± 4.62	67.99 ± 0.060	58.76 ± 0.025	0.419 ± 0.058
	PRG	0.415 ± 0.119	47.68 ± 6.90	60.31 ± 0.108	55.30 ± 0.053	0.344 ± 0.042
BMM	PCG	0.568 ± 0.102	44.49 ± 6.00	65.02 ± 0.069	57.79 ± 0.036	0.330 ± 0.044
	RPG	0.734 ± 0.078	37.15 ± 5.00	67.62 ± 0.106	58.64 ± 0.025	0.416 ± 0.057
	PRG	0.509 ± 0.125	47.97 ± 6.52	60.35 ± 0.069	55.23 ± 0.052	0.344 ± 0.046
BMA	PCG	0.569 ± 0.113	44.26 ± 7.11	64.52 ± 0.074	57.47 ± 0.040	0.332 ± 0.047
	RPG	0.741 ± 0.076	36.46 ± 4.80	68.08 ± 0.058	58.87 ± 0.253	0.416 ± 0.061
	PRG	0.516 ± 0.119	47.54 ± 6.75	60.82 ± 0.106	55.41 ± 0.052	0.345 ± 0.046
BO	PCG	0.563 ± 0.102	44.22 ± 5.94	65.05 ± 0.065	57.71 ± 0.035	0.329 ± 0.046
	RPG	0.736 ± 0.078	37.12 ± 4.79	67.32 ± 0.055	58.54 ± 0.025	0.424 ± 0.049
	PRG	0.512 ± 0.097	47.66 ± 6.47	60.92 ± 0.104	55.22 ± 0.049	0.346 ± 0.043
BS	PCG	0.526 ± 0.121	44.21 ± 6.24	64.20 ± 0.079	57.27 ± 0.042	0.332 ± 0.048
	RPG	0.742 ± 0.074	36.51 ± 4.61	67.65 ± 0.057	58.70 ± 0.026	0.421 ± 0.058
	PRG	0.480 ± 0.112	47.75 ± 6.87	60.89 ± 0.107	55.42 ± 0.052	0.347 ± 0.047
BSA	PCG	0.540 ± 0.115	44.43 ± 6.12	64.12 ± 0.076	57.25 ± 0.040	0.331 ± 0.047
	RPG	0.741 ± 0.076	35.97 ± 4.45	68.28 ± 0.061	58.75 ± 0.025	0.423 ± 0.054
	PRG	0.488 ± 0.124	47.63 ± 7.19	60.70 ± 0.109	55.40 ± 0.054	0.346 ± 0.045

BA= *Brucella abortus* bv. 1 9-941, BC= *Brucella canis* ATCC 23365, BMM= *Brucella melitensis* 16M, BMA= *Brucella melitensis* bv. *abortus* 2308, BO= *Brucella ovis* ATCC 25840, BS= *Brucella suis* 1330 and BSA= *Brucella suis* ATCC 234459.

Z test, which was performed for all the *Brucella* strains for different indices viz. CAI, Nc, GC and GC3 to find if any difference in the values of ribosomal protein genes, pathogenicity related genes with that of the protein coding genes exists. Z test revealed significant differences in values obtained for GC and GC3 as depicted in table 2a and 2b but no significant differences were seen for CAI and Nc.

Table 2a

Z test value for GC (guanine cytosine percentage) in protein coding genes (PCG), ribosomal protein genes (RPG) and pathogenicity related genes (PRG) of seven Brucella strains.

Strains	PCG	RPG	PRG
BA	0.0693	1.48	-0.058
BC	23.08	-0.8	-3.151
BMM	-11.722	-0.76	4.52
BMA	3.9	-0.356	1.288
BO	0.386	-8.84	1.265
BS	-1.929	-10.077	-1.135
BSA	-3.8	0.84	2.185

Table 2b

Z test value for GC3 (guanine cytosine ratio at third position) in protein coding genes (PCG), ribosomal protein genes (RPG) and pathogenicity related genes (PRG) of seven Brucella strains.

Strains	PCG	RPG	PRG
BA	14.052	3.115	-0.208
BC	23.08	-7.933	-3.037
BMM	-18.478	-3.43	-0.696
BMA	-0.013	-1.81	0.106
BO	7.185	-2.109	0.192
BS	3.835	1.86	-1.82
BSA	-5.355	7.25	0.188

BA= *Brucella abortus* bv. 1 9-941, BC= *Brucella canis* ATCC 23365, BMM= *Brucella melitensis* 16M, BMA= *Brucella melitensis* bv. *abortus* 2308, BO= *Brucella ovis* ATCC 25840, BS= *Brucella suis* 1330 and BSA= *Brucella suis* ATCC 234459.

The number of horizontally transferred genes for the four strains of *Brucella* was 442, 401, 388 and 426 for BA, BMM, BMA and BS respectively. Out of which BA, BC, BMM, BMA, BO, BS and BSA had 19, 18, 19, 20, 16, 20 and 18 pathogenesis related genes correspondingly (G:\Brucella organism profile\VFDB - comparative pathogenomic composition of Brucella.htm).

Homologs having similar sequences in other strains were obtained by the IMG genome BLAST. Different pathogenicity related genes of BA such as glycosyl transferase, mannose-6-phosphate isomerase, mannose-1-phosphate guanylyltransferase, phosphormanno mutase, hypothetical mannosyl transferase, GDP-mannose 4,6-dehydratase, perosamine synthase, rfbD, rfbE, wbkB, formyl transferase, glcosyl transferase, virB1, virB4, virB5, virB6, virB8, virB9, virB10 were found to have 18, 19, 20, 16, 20 and 18 horizontally transferred homologs for BC, BMM, BMA, BO, BS and BSA respectively (percentage identity ranging from 99-100)

To study the dissimilarity in the codon usage among the genes in the different organisms, multivariate statistical analysis is extensively used[35]. Correspondence analysis of codon count of all the *Brucella* strains for all the protein coding genes of the genome, ribosomal protein genes and pathogenesis related genes was performed and plotted. Figure 2 reveals the positions of the codon count for protein coding genes, ribosomal protein genes and pathogenesis related genes generated by correspondence analysis on the

plane defined by the first and second axes. The principal axis i.e., Axis1 was correlated with Nc, GC3, GC and CAI values for all the strains. Table 3 depicts the correlation among the different indices viz. GC3 and Nc, GC3 and GC, GC3 and CAI, CAI and Nc, Axis1 and Nc, Axis1 and GC3, Axis1 and GC and Axis1 and CAI for all the strains. From the table it is seen that there exists strong correlation between the different indices.

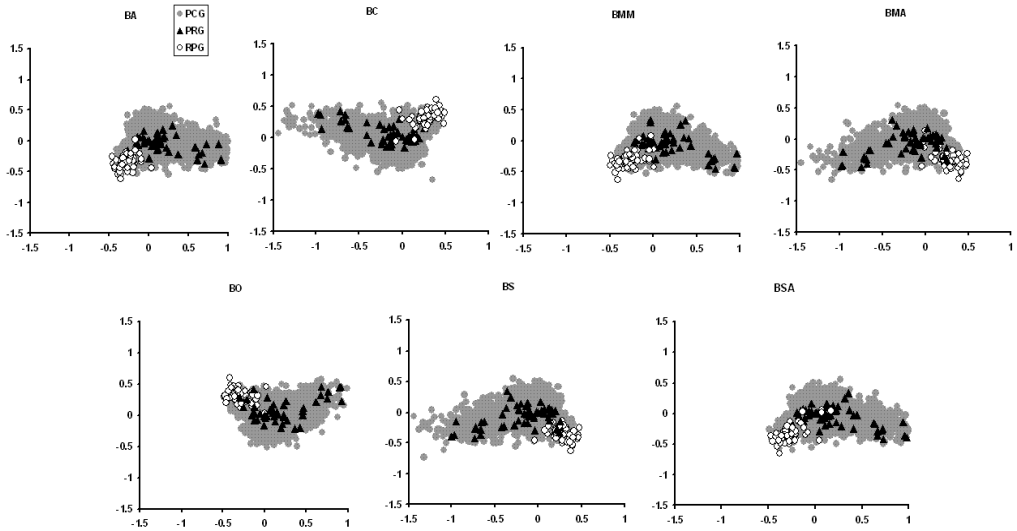

Figure 2

Correspondence analysis of codon usage patterns on codon count for all the seven Brucella strains. X and Y axes in all the seven plots corresponds to axes 1 and 2 of the analysis. PCG= Protein Coding genes, PRG= Pathogenicity Related Genes and RPG= Ribosomal Protein Genes.

Table 3

Correlations among the different indices of the seven Brucella strains.

Correlations	BA	BC	BMM	BMA	BO	BS	BSA
GC3 and GC	0.756	0.770	0.725	0.764	0.720	0.784	0.774
GC3 and CAI	0.996	0.964	0.987	0.989	0.980	0.971	0.972
CAI and Nc	0.939	0.973	0.960	0.953	0.976	0.964	0.969
Axis1 and Nc	0.981	0.903	0.983	0.899	0.992	0.897	0.982
Axis1 and GC3	0.843	0.997	0.849	0.999	0.884	0.996	0.823
Axis1 and GC	0.859	0.989	0.849	0.997	0.878	0.987	0.808
Axis1 and CAI	0.875	0.972	0.919	0.987	0.954	0.978	0.924

BA= *Brucella abortus* bv. 1 9-941, BC= *Brucella canis* ATCC 23365, BMM= *Brucella melitensis* 16M, BMA= *Brucella melitensis* bv. *abortus* 2308, BO= *Brucella ovis* ATCC 25840, BS= *Brucella suis* 1330 and BSA= *Brucella suis* ATCC 234459.

Codon adaptation index is an approximation of directional codon bias. CAI measurement summarizes the codon usage of a gene relative to the codon usage of a reference set of genes particularly highly expressed genes[21]. To recognize the potentially highly expressed genes, the CAI values for these *Brucella* genomes were checked. The CAI values for all the genes in different *Brucella* strains were computed and finally plotted. Ribosomal protein genes were found to have higher CAI value in comparison to other genes, indicating high levels of gene expression. The average CAI values for different gene groups associated with diverse functions also varied. The CAI value ranged from 0.139 to 0.874, 0.122 to 0.834, 0.175 to 0.885, 0.165 to 0.887, 0.205 to 0.880, 0.144 to 0.886 and 0.165 to 0.890 for BA, BC, BMM, BMA, BO, BS and BSA respectively.

As per Wu *et al.* 2005a[32], the top 10% of genes in terms of CAI values were classified as predicted Highly Expressed (PHX) genes. The CAI cut off value corresponded to 0.665, 0.607, 0.683, 0.696, 0.682, 0.664 and 0.673 for BA, BC, BMM, BMA, BO, BS and BSA respectively. Among the seven strains of *Brucella* about 35 genes associated with pathogenicity related genes has been found to be associated within the highly expressed categories. BA, BC, BMM, BMA, BO, BS and BSA have 2, 3, 3, 4, 3, 4 and 16 genes respectively. The pathogenesis related genes within the PHX category are shown in Table 4.

Cluster of Orthologous Genes (COGs) of protein among seven *Brucella* genomes were considered in order to understand the functional distribution of the PHX genes. The COG categories were divided into 4 COG groups to aid in investigation[30]. Group 1 (information storage and processing) consisting of COGs linked to transcription, translation, RNA processing, DNA replication, replication recombination and repair. Group 2 (cellular processes) included cell division and cell cycle control, nuclear structure, defense mechanism, signal transduction, cell envelop biogenesis, cell motility, intracellular structures and post translational modification. Group 3 (metabolism) incorporated COGs related to energy production and conversion, carbohydrate transport, amino acid transport, lipid transport and metabolism, nucleotide transport, coenzyme metabolism, inorganic ion transport and secondary metabolism biosynthesis. General function and functions unknown genes were included in group 4. Figure 4 illustrates the allotment of PHX into each COG category based on total PHX genes and the total genes within that COG category. The distribution of PHX in the COG functional groups of seven *Brucella* genomes were found to be as: BA 18.72, 17.81, 29.22 and 29.68%; BC 23.57, 12.10, 55.41 and 8.92%; BMM 21.43, 14.29, 52.68 and 11.61%; BMA 23.73, 13.61, 53.80 and 8.86%; BO 23.05, 15.26, 52.27 and 9.42%; BS 23.85, 13.79, 52.30 and 10.06 and BSA 23.05, 13.17, 52.69 and 11.08% for the COG functional groups (1-4) respectively.

Table 4

PHX genes in seven Brucella strains associated with pathogenicity.

Strains	GenBank accession	Description	Gene Length (bp)	CAI value
BA	YP_222879	Type IV secretion system protein VirB3	351	0.672
	YP_222880	Type IV secretion system protein VirB2	318	0.672
BC	YP_001591986	Glycosyltrnaferase 36	8604	0.742
	YP_001592317	Acyl carrier protein	237	0.698
	YP_001593891	Transcriptional regulatory protein chvI	720	0.620
BMM	NP_540028	Acyl carrier protein	282	0.728
	NP_540392	Acyl carrier protein	237	0.711
	NP_540754	Cellobiose phosphorylase	8604	0.702
BMA	YP_413604	Cyclic beta 1-2 glucan synthetase	8199	0.815
	YP_413944	Acyl carrier protein	237	0.780
	YP_414302	Acyl carrier protein	282	0.753
	YP_415418	Transcriptional regulatory protein	720	0.709
BO	YP_001258150	Cyclic beta 1-2 glucan synthetase	8595	0.747
	YP_001258477	Acyl carrier protein	237	0.728
	YP_001258830	Putative acyl carrier protein	282	0.711
BS	NP_697152	Cyclic beta 1-2 glucan synthetase	8199	0.807
	NP_697487	Acyl carrier protein	237	0.764
	NP_697869	Acyl carrier protein	282	0.734
	NP_699065	DNA-binding response regulator BvrR	720	0.676
BSA	YP_001621866	Hypothetical protein	741	0.827
	YP_001621867	Hypothetical protein	777	0.817
	YP_001621868	2,3-dihydroxy benzoate-AMP ligase	1620	0.812
	YP_001621869	Isochorismate synthase Hypothetical protein	1176	0.809
	YP_001621870	P-type DNA transfer	1392	0.807
	YP_001621912	ATPase VirB11 Hypothetical protein	1089	0.763
	YP_001621913	P-type DNA transfer	1176	0.762
	YP_001621914	ATPase VirB9 Hypothetical protein	870	0.761
	YP_001621915	Hypothetical protein	720	0.759
	YP_001621916	Hypothetical protein	174	0.759
	YP_001621917	P-type DNA transfer	1044	0.757
	YP_001621918	protein VirB5	717	0.756
	YP_001621919	Type IV secretion/ conjugal transfer ATPase, VirB4 family Hypothetical protein Hypothetical protein	2496	0.755
	YP_001621920	Hypothetical protein	351	0.754
	YP_001621921		318	0.753
	YP_001621922		717	0.753

BA= *Brucella abortus* bv. 1 9-941, BC= *Brucella canis* ATCC 23365, BMM= *Brucella melitensis* 16M, BMA= *Brucella melitensis* bv. *abortus* 2308, BO= *Brucella ovis* ATCC 25840, BS= *Brucella suis* 1330 and BSA= *Brucella suis* ATCC 234459.

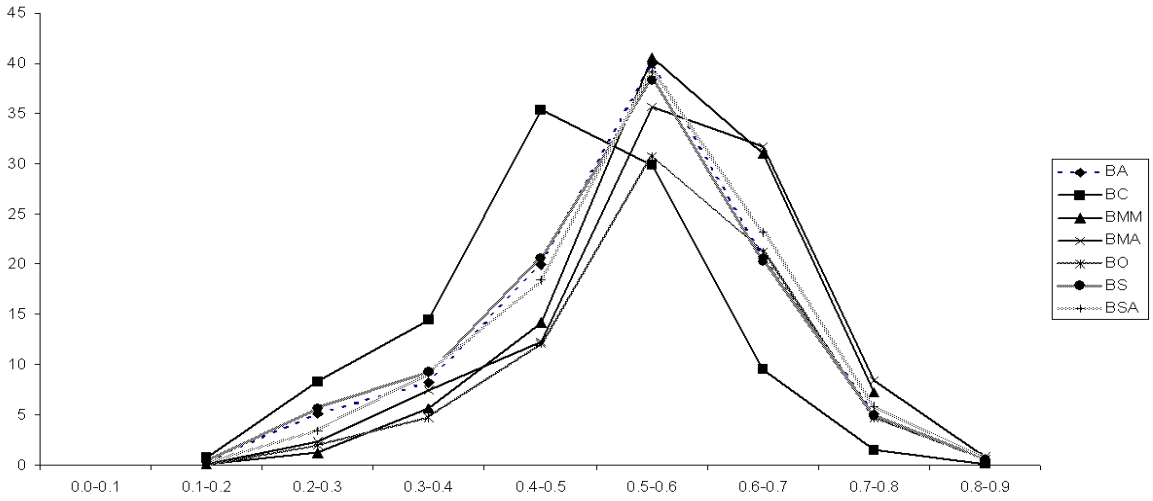

Figure 3

Frequency distribution of the CAI (Codon Adaptation Index) for all coding genes in all the seven Brucella genomes.

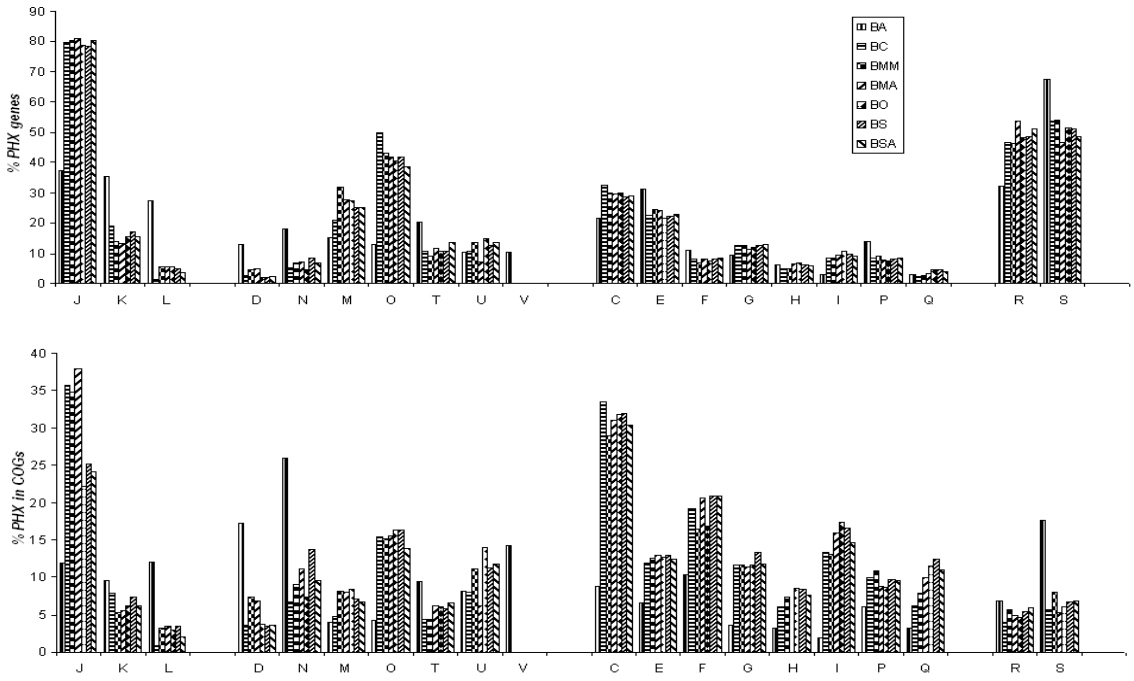

Figure 4

Distribution of predicted highly expressed (PHX) genes within functional groups of Brucella genomes. COG functional groups are as follows: (1) Information and storage processing: J- Translation, ribosomal structure and biogenesis; K- Transcription; L- DNA replication, recombination and repair, (2) Cellular processes: D- cell division and cell cycle control; M- cell envelop biogenesis; N- cell motility and secretion; O- Post translational modification; T- Signal transduction; U- intracellular trafficking; V- Defense mechanisms, (3) Metabolism: C- Energy production and conversion; E- Amino acid transport and metabolism; F- Nucleotide transport and metabolism; G- Carbohydrate transport and metabolism; H- Coenzyme metabolism and metabolism; I- Lipid transport and metabolism; P- Inorganic ion transport and metabolism; Q- Secondary metabolism biosynthesis, transport and catabolism, (4) Poorly characterized: R- General function; S- Functions unknown.

DISCUSSION AND CONCLUSION

The Nc/GC3 plot also highlights the ribosomal protein genes that are anticipated to be highly expressed during rapid cell growth. Nearly all of the ribosomal protein genes of all *Brucella* strains genomes are observed to strongly cluster at the lower end of the plot and implying a significant strong codon bias in these genes resulting out of selection for translational efficiency[39]. This is similar to results obtained for *Streptomyces*[32], *Xanthomonas*[20], *Frankia*[19], *Azotobacter*[21] and *Chlorobium*[22]. Apart from this, the pathogenesis related genes are also shown in the plot. The continuous curve indicates the factor influencing codon usage bias. If GC3s utterly controlled codon bias, Nc values ought to fall beneath the expected curve of the Nc/GC3 plot. Yet, we found that apart from a small number of genes, values for majority of the genes were well below the expected curve. On an average, the elevated Nc values of both the protein coding genes and pathogenesis related genes infer that they are lowly biased. It can be concluded from Table 1 that Nc is inversely proportional to GC3. The lower Nc value signifies high degree of codon bias. As evident from the table that ribosomal protein gene had a lower Nc values compared to the mean values obtained for all the protein coding genes for all the genome, while the pathogenesis related genes had Nc values higher than both the ribosomal protein gene and protein coding genes. It can be inferred from Table 1 that ribosomal protein genes are more highly biased compared to those associated with pathogenesis.

Strong positive correlations of GC3 values with GC content indicate that GC3 increases with the increase of GC. Quite fascinatingly strong positive correlation of GC3 values with CAI values imply that gene expression plays a significant role in influencing synonymous codon bias in these organisms. This is further strengthened by the strong positive correlation observed between CAI and Nc values indicating that codon bias is strongly influenced by highly expressed genes. High positive correlations of the principal axis of correspondence analysis with the effective number of codons (Nc) indicates the increase of codon bias among the genes lying towards the right side of Axis 1 and the establishment of Nc as an important parameter in effecting synonymous codon usage. Strong correlation of the Axis 1 with GC and GC3 values portray the role played by compositional constraints in influencing codon bias. Similarly the positive correlation of CAI values of Axis 1 reaffirm the role of highly expressed genes.

Figure 2 implies that like *E. coli*[40], the scattered plot of BA, BMM, BMA, BS and BSA revealed a small core region with two ascending horns, on the other hand BC and BO had small core region with two descending horns. Most of the pathogenesis related genes of BA, BMM, BO and BS are located on the positive side of the Axis 1, whereas for others they are on the negative side. The ribosomal protein genes that are thought to be highly expressed are clustered together in the right horn of BC, BMA, BS and BSA and that of BA, BMM and BO in the left horn.

The study revealed that not all the pathogenesis related genes were acquired by horizontal gene transfer mechanisms. The pathogenicity related genes that are acquired by horizontal gene transfer include glycosyl transferase, mannose-6-phosphate isomerase, mannose-1-phosphate guanylyltransferase, phosphormanno mutase, hypothetical mannosyl transferase, GDP-mannose 4,6-dehydratase, perosamine synthase, rfbD, rfbE, wbkB, formyl transferase, glycosyl transferase, virB1, virB4, virB5, virB6, virB8, virB9, virB10. Though the percentage identities for rest of the pathogenicity related homologs in all the strains ranged from 99 to 100, they were not horizontally transferred. Thus it can be inferred that these genes are indigenous to these pathogenic bacteria that protect them from the selective pressure of evolution. On the other hand, the homologs that are horizontally transferred are known to acquire from other organisms. These genes are transportable within the genus because of high level of percentage identity within the strains.

From Table 4 it is seen that BSA houses the maximum number of pathogenesis related genes that are PHX. Yet, the presence of some

pathogenicity related genes in the PHX category entail that high expression levels of these genes play a significant role in influencing pathogenesis in these bacteria by taking control over the host's defense systems. High number of PHX genes associated with the metabolism COG group divulges that metabolic genes has an important part to play in effecting the survival of the bacteria against the action of host's resistance, antibiotics etc. thus establishing infection.

In conclusion it can be said that selection for translational efficiency favours codon usage bias in *Brucella*. Although protein coding genes and pathogenicity related genes show low bias ribosomal protein genes are highly biased. Besides translational efficiency, GC3 compositional constraints, effective number of codons and highly expressed genes strongly influence codon bias. Correspondence analysis reveals that ribosomal genes are strongly clustered along the horns. Homologs related to pathogenicity genes showing high identity levels indicated that they are indigenous to these pathogenic bacteria protecting them from selective pressures of evolution. Homologs horizontally transferred are known to be acquired from other organisms. The role played by metabolic genes reveals the way of survival of the bacterium in the environment.

ACKNOWLEDGEMENT

The authors are grateful to the Department of Biotechnology, Government of India, for providing financial support in setting up Bioinformatics Facility at the Department of Botany, University of North Bengal.

REFERENCES

1. Fretin D, Fauconnier A, Kohler S, Halling S, Leonard S, Nijskens C, Ferooz J, Lestrate P, Delrue RM, Danese I, Vandenhaute J, Tibor A, DeBolle X, and Letesson JJ, The sheathed flagellum of *Brucella melitensis* is involved in persistence in a murine model of infection. Cell Microbiol, 7:687-98, (2005).
2. Leslie C, Balows A, and Sussman M, Microbiology and microbial infections. System Bacteriol, 2:829-830, (1998).
3. M.J. Corbel, and W.J. Brinley-Morgan. Genus *Brucella*. In: N.R. Krieg, and J.G. Holt (eds.), *Bergey's Manual of Ssystematic Bacteriology*, vol. 1, The Williams & Wilkins Co., Baltimore, Maryland, USA,1984,pp. 377–388.
4. Foster G, Osterman BS, Godfroid J, Jacques I, and Cloeckaert A, *Brucella ceti* sp. nov. and *Brucella pinnipedialis* sp. nov. for *Brucella* strains with cetaceans and seals as their preferred hosts. Int J Syst Evol Microbiol, 57:2688–93, (2007).
5. Scholz HC, Nockler K, Gollner C, Bahn P, Vergnaud G, Tomaso H, Al-Dahouk S, Kampfer P, Cloeckaert A, Maquart M, Zygmunt MS, Whatmore AM, Pfeffer M, Huber B, Busse HJ, and De BK, *Brucella inopinata* sp. nov., isolated from a breast implant infection. Int J Syst Evol Microbiol, DOI 10.1099/ijs.0.011148-0, (2009).
6. Corbel MJ, Brucellosis: an overview. Emerg Infect Dis, 3:213-221, (1997).
7. Corbel MJ, and Macmillan AP, Brucellosis. Bacterial Infect, 3:819-820, (1998b).
8. Paulsen IT, Seshadri R, Nelson KE, Eisen JA, Heidelberg JF, Read TD, Dodson RJ, Umayam L, Brinkac LM, Beanan MJ, Daugherty SC, Deboy RT, Durkin AS, Kolonay JF, Madupu R, Nelson WC, Ayodeji B, Kraul M, Shetty J, Malek J, Aken SEV, Riedmuller S, Tettelin H, Gill SR, White O, Salzberg SL, Hoover DL, Lindler LE, Halling SM, Boyle SM, Fraser CM, The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc Natl Acad Sci USA, 13148-13153, (2002).

9. Young EJ, Human brucellosis. Rev Infect Dis, 5:321-342, (1983).
10. Young EJ, An overview of human brucellosis. Clin Infect Dis, 21:283-289, (1995).
11. Kozukeev TB, Ajeilat S, Maes E, and Favorov M, Risk Factors for Brucellosis. Morb Mortal Weekly Report, 55:31-34, (2003).
12. B.A. Forbes, D.F. Sahm, and A.S. Weissfeld. *Staphylococcus*, M*icrococcus*, and similar organisms. In: K. Fabiano (eds), Bailey and Scott's diagnostic microbiology,11th Edition, Missouri, Andrew Allen, 2002, pp. 284-287.
13. Ocon P, and Reguera JM, Phagocytic cell function in active brucellosis. Infect Immun, 62:910-914, (1994).
14. Acha PN, and Szyfres B, Zoonoses and communicable diseases common to man and animals. Pan American Health Organization, Washington, DC, 28-45, (1980).
15. Shrestha JM, Zoonotic Diseases, Zoonoses Control sub-division, Epidemiology and Disease Control Division, Department of Health Services, Ministry of Health, (2004).
16. Schurig GG, Sriranganathan N, and Corbel MJ, Brucellosis vaccines: past, present and future. Vet Microbiol, 90:479-96, (2002).
17. Banerjee T, Basak S, Gupta SK, and Ghosh TC, Evolutionary forces in shaping the codon and amino acid usages in *Blochmannia floridanus*. J Biomol Struct Dyn, 22:13-24, (2004).
18. Basak S, Banerjee T, Gupta SK, and Ghosh TC, Investigation on the causes of codon and amino acid usages variation between thermophilic *Aqiflex aeolicus* and mesophilic *Bacillus subtilis.* J Biomol Struct Dyn, 22:205-214, (2004).
19. Sen A, Sur S, Bothra AK, Benson D, Normand P, and Tisa LS, The implication of lifestyle on codon usage patterns and predictably highly expressed genes for three *Frankia* genomes. Anton Van Leewen, 93:335-346, (2008).
20. Sen G, Sur S, Bose D, Mondal U, Furnholm T, Bothra A, Tisa L, Sen A, Analysis of codon usage patterns and predicted highly expressed gene for six phyopathogenic *Xanthomonas* genomes shows high degree of conservation. In Silico Biol, 7:547-558, (2007).
21. Sur S, Bhattacharya M, Bothra AK, Tisa LS, and Sen A, Bioinformatics analysis of codon usage patterns in a free-living diazotroph, *Azotobacter vinelandii*. Biotechnol, 7:242-249, (2008).
22. Sur S, Bothra AK, Bajwa M, Tisa LS, and Sen A, *In silico* analysis of *Chlorobium* genomes divulge insights into the lifestyle of the bacteria. Res J Micrbiol, 3:600-613, (2008).
23. Knight RD, Freeland SJ, and Landweber LF, A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol, 2:0010.1-0010.13, (2001).
24. Lafay B, Atherton JC, and Sharp PM, Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori*. Microbiol, 146:851-860, (2000).
25. Dos Reis M, Wernisch L, and Savva R, Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. Nucleic Acids Res, 31:6976-6985, (2003).
26. Sharp PM, and Li WH, The Codon Adaptation Index- a measure of directional synonymous codon usage bias and its potential applications. Nucleic Acids Res, 15:1281-1295, (1987).
27. Martin-Galiano AJ, Wells JM, and Campa de la AG, Relationship between codon biased genes, microarray expression values and physiological characteristics of *Sterptococcus pneumoniae*. Microbiol, 150:2313-2325, (2004).
28. Wright F, The "effective number of codons" used in a gene. Gene, 87:23-29, (1990).
29. Peden J, Analysis of codon usage. PhD Thesis. The University of Nottingham, United Kingdom, (1999).
30. Tatusov RL, Federova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin J

J, and Natale DA, The COG database: An updated version includes eukaryotes. BMC Bioinformatics, 4:41, (2003).
31. Markowitz VM, Ivanova N, Palaniappan K, Szeto E, and Korzeniewski F, An experimental metagenome data management and analysis system. Bioinformatics, 22:359-367, (2006).
32. Wu G, Culley DE, and Zhang W, Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. Microbiol, 151:2175-2187, (2005a).
33. Wu G, Nie L, and Zhang W, Predicted highly expressed genes in *Nocardia farcinica* and its implication for its primary metabolism and nocardial virulence. Anton Van Leewen, 89:135-146, (2005b).
34. Sur S, Sen A, and Bothra AK, Codon usage profiling and analysis of intergeneric association of *Frankia* EuIK1 *nif* genes. Ind J Microbiol, 46:363-369, (2006).
35. Ghosh TC, Gupta SK, and Majumdar S, Studies on codon usage in *Entamoeba histolytica*. Int J Parasitol, 30:715-722, (2000).
36. Walpole RE, Myers RH, Myers SL, and Ye K, Probability and Statistics for Engineers and Scientists, Pearson Education, Singapore, Pte. Ltd., Indian Branch, 482 F.I.E. Patparganj, Delhi-110092, India, (2004).
37. Tsirigos A and Rigoutsos I, A sensitive, support-vector-machine method for the detection of horizontal gene transfers in viral, archaeal and bacterial genomes. Nucleic Acids Res, 33:3699-3707, (2005).
38. Ikemura T, Correlation between abundance of *Escherichia coli* tRNAs and their occurrence of the respective codons in protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* system. J Mol Biol, 146:1-21, (1981).
39. Cutter AD, Payseur BA, Salcedo T, Estes AM, Good JM, Wood E, Hartl T, Maughan H, Strempel J, Wang B, Bryan AC, and Dellos M, Molecular correlates of genes exhibiting RNAi phenotypes in *Caenorhabditis elegans*. Genome Res, 13:2651-2657, (2003).
40. Medique C, Rouxel T, Vigier P, Henaut A, and Dancin A, Evidence for horizontal gene transfer in *Escherichia coli* speciation. J Mol Biol, 222:851-856, (1991).