NBU Journal of
Plant Sciences
Copyright © 2010, University of North Bengal. All Rights Reserved

Vol. 4, March 2010 p. 65-71
ISSN No. 0974 6927

Bioinformatics of codon usage pattern in pathogenic proteobacteria *Burkholderia mallei*

Manash K Ghosh[1], Asim K Bothra[2], and Arnab Sen[1]*
[1]Department of Botany, University of North Bengal, Siliguri 734013, India.
[2]Department of Chemistry, Raiganj University College, Raiganj, India.

Abstract

Burkholderia mallei are pathogenic *gram* negative β Proteobacteria those are available in Africa, Asia, Middle East, Central and South America and abundantly known as the causal organism of glanders.In this study, the synonymous patterns of four *Burkholderia mallei* (*Burkholderia mallei* ATCC 23344, *Burkholderia mallei* NCTC 10229, *Burkholderia mallei* NCTC 10247, *Burkholderia malle* SAVP1) genome were compared and analyzed to each other. It was observed that *Burkholderia mallei* have high G+C content and moderately biased. Using cogon adaptation index (CAI) as a numerical estimator of gene expression level where ribosomal protein coding genes were considered as a reference of highly expressed genes. Here, we also studied the functional analysis of the PHX genes, gene expression level, correspondence analysis and horizontally transferred pathogenicity related genes activity. COGs are also associated with metabolism especially those linked to carbohydrate metabolism and amino acid transport.

Keywords: *Burkholderia mallei*, Codon bias,Correspondence analysis, PHX, COG.

Abbreviation: BMAL-44-*Burkholderia mallei* ATCC 23344, **BMAL-29**- *Burkholderia mallei* NCTC 10229, **BMAL-47**-*Burkholderia mallei* NCTC 10247, **BMAL-VP1**-*Burkholderia malle* SAVP1, **PCG**-Protein coding gene, **RPG**-Ribosomal protein genes, **PRG**-Pathogenecity related genes, **HTG**- Horizontally transferred genes, **PHX**-Predicted highly expressed genes, **COG**-Clusters of orthologous groups of protein, **CAI**-Codon adaptation index

Horizontal gene transfer refers to the transfer of genes or genetic material directly from one individual to another by processes similar to infection. It is distinct from the normal process of *vertical* gene transfer - from parents to offspring - that occurs in reproduction. Genetic engineering bypasses reproduction altogether by exploiting horizontal gene transfer, so genes can be transferred between distant species that would never interbreed in nature. For example, human genes are transferred into pig, sheep, fish and bacteria. Toad genes are transferred into potatoes. Completely new, exotic genes can therefore be introduced into food crops.

Horizontal gene transfer is the lateral movement of genes between organisms. HGT may be designated as infectious transfer. This is a very common mechanism among bacteria even they are distantly related. By this way one bacterium become drug resistance. Transformation, transduction and conjugation are three classical mechanisms of HGT. Most of the organisms are affected by HGT.

The purpose behind this study was to implement a comparative analysis of the synonymous codon usage patterns, predicted expression levels for the protein coding genes in these pathogenic bacterial strains with special reference to pathogenicity related genes, study of horizontally transferred pathogenicity related genes to detect their presence in the strains and scrutinize the nature of highly expressed genes upon their lifestyles. We consider that the result of this study would be helpful in further work on these bacteria.

*Corresponding author:
E-mail: senarnab_nbu@hotmail.com

Materials & Methods:

The complete genome sequences for four strains of *Burkholderia mallei* (*Burkholderia mallei* ATCC 23344, *Burkholderia mallei* NCTC 10229, *Burkholderia mallei* NCTC 10247, *Burkholderia malle* SAVP1, henceforth referred to as BMAL-44, BMAL-29, BMAL-47and BMAL-VP1 respectively) bearing Gene Bank accession numbers NC_006348, NC_006349,NC_008836, NC_008835,NC_009080, NC_009079, NC_008785 and NC_008784 were obtained from the IMG website (www.img.jgi.doe.gov). The gene sequences are retrieved in the FASTA format. All the protein coding genes, horizontally transferred genes, pathogenicity related genes and ribosomal protein genes were examined using Codon W software (http://bioweb2.pasteur.fr) (Peden, 1999) and CAI Calculator2 (http://www.evolvingcode.net/codon/CalculateCAIs.php) (Wu *et al.*, 2005) Codon W (Peden, 1999) was used to determine the GC3 content, effective number of codons (Nc) (Wright,1990) and the frequency of optimal codons (Fop) (Peden, 1999). The effective number of codons (Nc) is a straightforward measure of codon bias (Wu *et al.*, 2005). It generally ranges between 20 and 61. Fop (Peden, 1999) calculates the section of synonymous codons that are optimal. Its value varies from 0 to 1.0. All negative Fop values were adjusted to zero. The 'codon adaptation index' (CAI)

(Peden, 1999) values were computed using 'The CAI Calculator 2 (http://www.evolvingcode.net/codon/CalculateCAIs.php) (Wu *et al.*, 2005) taking the ribosomal protein genes as a reference. The CAI value varies from 0 to 1.0 with higher CAI values suggestive of the fact that the gene of study has a codon usage pattern like that of the reference genes (Sharp and Li, 1987).

An analysis of the horizontally transferred pathogenicity related genes among the *Burkholderia* strains were carried out to recognize whether they were located in all the strains or indigenous to a specific strain.

Among different types of genes, the pathogenicity related horizontal gene transfer mechanisms in the studied strains were sorted out acquires genes. Using the Integrated Microbial Genomes database (www.img.jgi.doe.gov) the sorted pathogenicity related genes for each strain were subjected to IMG Genome BLAST against the studied strains to find out the sequence homologs. The minimum percent identity was set at 90% and the maximum E value 1e-2.

Correspondence analysis (COA) is a sensitive method identifying non-random usage of synonymous codons in many organisms. Correspondence analysis of codon count was performed with Codon W (http://bioweb2.pasteur.fr) (Peden, 1999). This method investigates the key trends in codon and amino acid disparity among the genes.

Results and discussion

Codon usage patterns in four strains of *Burkholderia mallei* genome:

It was proved that codon usage pattern is more advanced for the prokaryotes than for the eukaryotes; much of this information is based on the relatively few species that have been subjected to a concerted molecular genetic analysis. It is truth that codon usage among the Gram-negative proteobacteria is much more advanced than in any other group of species (Grantham *et al.*, 1980; Gouy and Gautier, 1982). Generally most bacterial genome with a balanced AT/GC genome content had high degree of heterogeneity in codon usage pattern. Codon heterogeneity was closely associated with gene expression level. Thus highly expressed genes contain a high frequency of codons i.e. translationally optimal (Ikemura, 1981; Lafay *et al.*, 2000; Ikemura, 1985). The studied strains of *Burkholderia mallei* genomes have moderately high G+C content, GC3 and Nc values that help to determine the existence of codon heterogeneity among them. The results from Nc versus GC3 plots, suggested the codon usage variations among genes of the same genome showed that the Nc values of the genes ranging from 21-61 for the four strains of *Burkholderia mallei* genomes suggesting the considerable heterogeneity is present in these GC-rich genomes. The RPGs, which were expected to be expressed at high levels during rapid cell growth are recognized and represented in Nc vs GC3 plot. The clustering of most RPGs in *Streptomyces* (Wu *et al.*, 2005) *Xanthomonas Frankia, Azotobacter* (Sen *et al.*, 2007; 2008) genome remain at low end of the Nc vs GC3 curve that signify stronger codon bias in the RPGs. The RPGs and PRGs, in all the studied bacteria formed similar clusters at the low end of the Nc versus GC3 plot that indicate stronger codon bias in RPGs and PRGs of *Burkholderia mallei*.

Codon usage is similar in closely related species but difference depends on phylogenetic distance (Peden, 1999). Genes having < 40 Nc value tend to have stronger codon bias manipulated by mutational pressure (Cameron and Aguade, 1998). Table 1 show that the mean Nc values of protein coding genes is 34 for all the studied strains with mean standard deviation value is around 6 where as Nc value of HTGs are high ranging from 38-40. The GC3 values for the four strains of *Burkholderia mallei* were quite high ranging from 84-85 in PCGs but the GC3 values of PRGs are higher than that. PCGs, PRGs and RPGs have higher Fop values in compared to HTGs. GC value of the studied genomes is very similar in the respective indices which are also shown in the Table 1.

Table-1: Mean Values of Nc, GC%, GC3%, Fop, CAI of the studied genes of *Burkholderia mallei* (Mean ± Standard deviation

Strain	Gene	Mean Nc	MeanGC (%)	MeanGC3 (%)	Mean Fop	Mean CAI
BMAL-44	PCG	34.661±6.22	67.918±0.048	84.5±0.09	0.576±0.05	0.600±0.14
	RPG	33.471±4.83	61.747±0.029	80.0±0.04	0.607±0.05	0.700±0.07
	PRG	34.713±5.25	68.702±0.046	86.2±0.07	0.584±0.04	0.612±0.09
	HTG	40.31±7.65	63.609±0.064	76.0±0.10	0.539±0.073	0.512±0.151
BMAL-29	PCG	34.104±5.64	68.368±0.049	85.3±0.08	0.579±0.057	0.612±0.12
	RPG	33.384±4.607	61.716±0.032	80.0±0.051	0.609±0.053	0.697±0.076
	PRG	33.106±4.351	68.789±0.035	87.8±0.054	0.587±0.043	0.645±0.091
	HTG	38.248±6.800	64.268±0.066	78.3±0.100	0.550±0.187	0.540±0.137
BMAL-47	PCG	34.419±5.96	68.352±0.049	84.6±0.090	0.576±0.059	0.600±0.136
	RPG	33.219±4.65	61.908±0.034	80.0±0.052	0.607±0.053	0.700±0.083
	PRG	33.249±4.14	68.969±0.036	87.7±0.057	0.587±0.044	0.639±0.089
	HTG	39.013±7.03	63.939±0.063	76.7±0.103	0.545±0.068	0.519±0.145
BMAL-VP1	PCG	34.297±5.822	68.238±0.050	84.7±0.089	0.577±0.058	0.617±0.129
	RPG	33.636±4.891	61.964±0.033	79.9±0.052	0.606±0.049	0.703±0.089
	PRG	33.164±3.916	68.677±0.033	87.4±0.055	0.586±0.039	0.652±0.068
	HTG	38.443±6.732	63.916±0.062	77.1±0.100	0.546±0.071	0.543±0.137

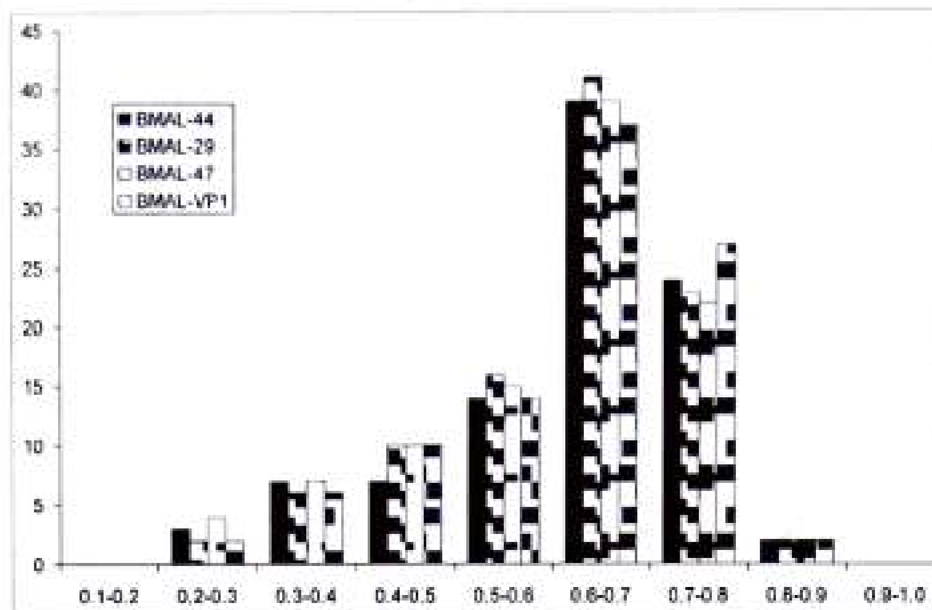

Figure 1: Frequency distribution of CAI values for all coding genes in the four *Burkholderia mallei* genome

Analysis of horizontally transferred pathogenicity related genes:

Gene transfer is generally associated with transposon-like elements or insertion sequences (Groisman *et al.*, 1992; Groisman *et al.*, 1993; Simon, *et al.*, 1980). Horizontally gene transferred mechanism is not easy process and there are some transfer barriers that prevent the delivery of genetic information from a donor cell that block inheritance of newly acquired genes (Matic *et al.*, 1995). But genes acquired by horizontal transfer often have atypical G+C content, codon bias and repetitive elements (Medigue *et al.*, 1991) and only approach the characteristic codon usage and G+C content of their host after millions of years (Groisman *et al.*, 1993). The studied strains of *Burkholderia mallei* contained 391, 428, 513 and 468 HTGs for BMAL-44, BMAL-29, BMAL-47and BMAL-VP1 respectively. Amongst them, the number of PRGs was 47,6,12 and 10 for BMAL-44, BMAL-29, BMAL-47and BMAL-VP1 respectively.

IMG BLAST results revealed homologs having sequence identity with a number of similar horizontally transferred pathogenicity related gene coding proteins in the four studied strains of *Burkholderia mallei*. Among them, BMAL-44 was the most pathogenic strain containing virulence factors essential for pathogenicity. These are extra cellular capsule, *Salmonella typhimurium*-like type III secretion system and type VI secretion system (Schell *et al.*, 2008). It was observed that out of 47 horizontally transferred pathogenicity related genes in BMAL-44, 33,36 and 32 horizontally transferred homologs are identical (percent identity ranging from 90-100) in BMAL-29, BMAL-47and BMAL-VP1 respectively.

Correspondence Analysis:

Correspondence analysis (CA) is a multivariate method and, as such, its aim is to summarize data structures in high-dimension space by projection onto low-dimension subspaces, while loosing as little information as possible. The principal factors are therefore along the directions of maximum variability in the dataset. Correspondence Analysis (COA) of synonymous codon usage of all protein coding genes in *Burkholderia mallei* strains were performed on simple codon count. Correspondence analysis of PCGs, RPGs and PRGs revealed the position of first and second major axis showed the position of above-mentioned genes. In BMAL-44, the Protein Coding Genes scattered as two indistinguishable clouds where as the PRGs and PCGs clustered from -0.5 to +0.5 on X-axis and Y-axis respectively. In BMAL-29, the PCGs distributed on the both axes but PRGs were not overlapping on RPGs that means the RPGs are not highly expressed. In BMAL-47, the scattering of studied genes were almost like BMAL-44and in BMAL-VP1, all these genes formed a clumped horizontal line where PRGs and RPGs were overlapped with each other. However, correlation with first major axis and GC3, GC, Nc, CAI values revealed some interesting results. The position of the genes on the first major axis of variation showed strong positive correlation with GC3 ($r = 0.941, r = 0.922$; $p<0.001$) for BMAL-44 and BMAL-29 and a moderate positive correlation with GC3 ($r = 0.602, r = 0.444$; $p<0.001$) for BMAL-47 and BMAL-VP1. The position of the genes on the first major axis of variation showed positive correlation with Nc ($r = 0.795, r = 0.705, r = 0.447$; $p<0.001$) for BMAL-44, BMAL-29, BMAL-47 and negative correlation with Nc ($r = -0.335$; $p<0.001$) for BMAL-VP1.On the other hand, the genes of the first major axis showed strong negative correlation with CAI ($r = -0.951, r = -0.945, r = -0.617$; $p<0.001$) for BMAL-44, BMAL-29, BMAL-47 and positive correlation with CAI ($r = 0.413$; $p<0.001$) for BMAL-VP1. From these results, we can predict that the genes virulent strain has significant negative correlation with first major axis and CAI whereas avirulent has simple positive correlation.

Identification of predicted highly expressed genes:

The effective measures of synonymous codon usage bias is the codon adaptation index (CAI).Gene expression was measured by calculating CAI value in the four strains of *Burkholderia mallei*. Here, RPGs were taken as reference set of highly expressed genes (Sharp and Wen-Hsiung, 1987). The distribution of CAI

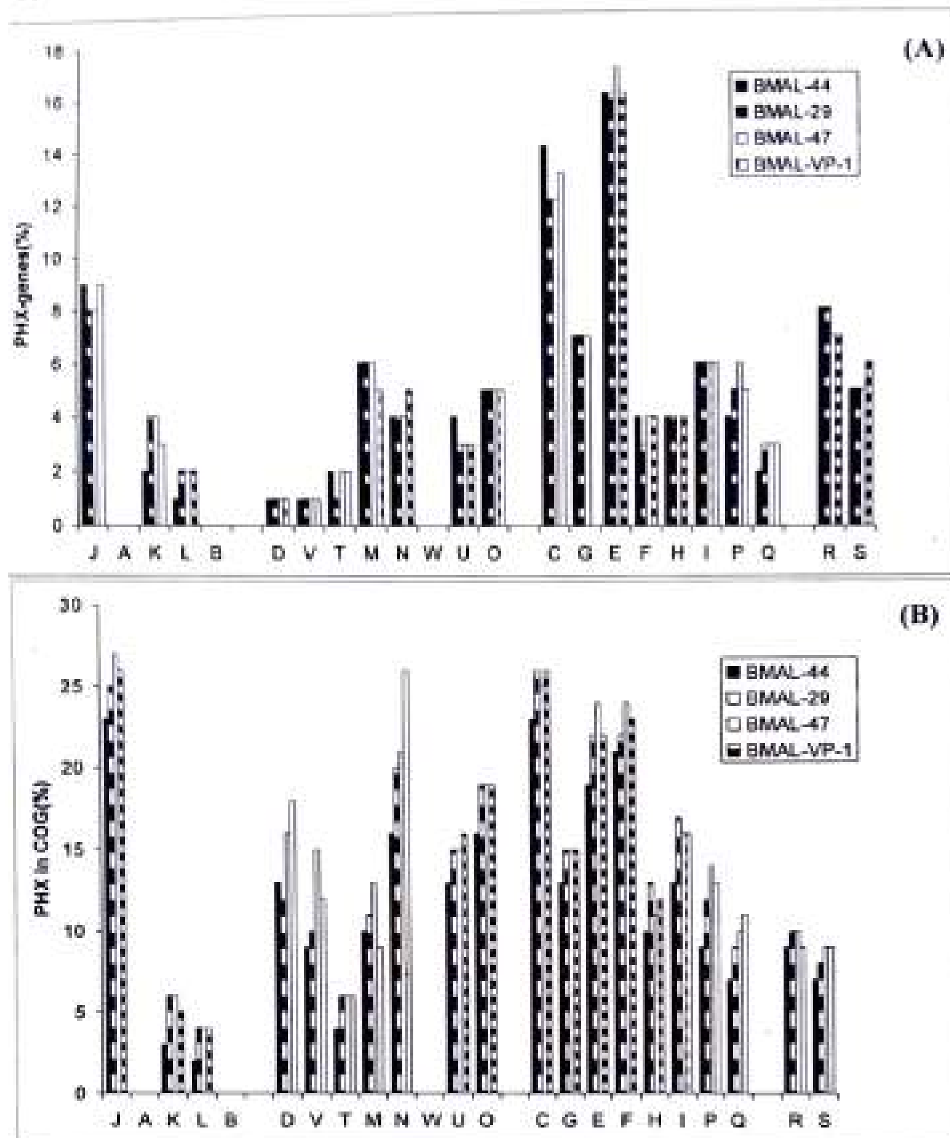

Figure 4 (A-B): Distribution of *Burkholderia mallei* predicted highly expressed genes within functional COG groups (as in text)

value of different genomes are shown in the Figure 1 .The CAI value ranges from 0.09-0.86, 0.12-0.86,0.12-0.86 and 0.11-0.87 for BMAL-44, BMAL-29, BMAL-47 and BMAL-VP1 respectively. The mean CAI value of PCG, PRG and HTG were from 60-61, 61-65 and 51-54 where the mean CAI value of highly expressed RPG was quiet high (Table1).

According to descending order of CAI value, the top 10% of total CAI value were classified as the predicted highly expressed genes (PHX), and corresponded to CAI cutoff values are 0.751, 0.74, 0.739 and 0.748 for BMAL-44, BMAL-29, BMAL-47and BMAL-VP1 respectively. BMAL-44 had 449 PHX genes, including 15 RPG, 22 PRG and 11 HTG. BMAL-29 had 554 PHX genes, with 16 RPG, 16 PRG and 19 HTG. BMAL-47 had 589 PHX genes, including 18 RPG, 24 PRG and 23 HTG. BMAL-VP1 had 521 PHX genes, including 17 RPG, 23 PRG and 20 HTG.

Functional analysis of the PHX genes:

Orthologs are evolutionary significant as they are evolved from common ancestor and opposed to paralog. Basically, orthologous proteins have the same structural domain and the same function. In order to realize the functional distribution of the PHX genes among the four *Burkholderia mallei* genomes the Cluster of Orthologous Groups (COG) of Proteins were studied. For these *Burkholderia mallei*, genomes, 23 COG categories were analyzed.Fig.-2 (A and B) showed the portion of the PHX into each COG category based on the total PHX genes and the title genes within the COG groups and expressed as percentage. To support the analysis, each of the COG categories were classified into 4 COG groups: *Information and storage processing* comprising to COG connected to J-Translation; A-RNA processing; K-Transcription; DNA replication ,recombination and repair; B-Chromatin structure and dynamics,(COG-1);*Cellular process and signaling* comprising of COGs related to D-Cell divisions-Defense mechanism-Signal transduction-Cell envelope biogenesis; N-Cell motility and secretion; W-Extra cellular structure; U-Intracellular trafficking; O-

Table-2: Pathogenecity related PHX genes of BMAL-44, BMAL-29, BMAL-47, BMALVP1 and their CAI value

Locus tag of PRG	Description	CAI value
	BMAL-44	
BMAA0440	hypothetical protein	0.793
BMA0504	unknown function	0.786
BMAA0455	hypothetical protein	0.786
BMAA0755	putative outer membrane nitrite reductase	0.785
BMA_IS407A-37	endoribonuclease, L-PSP family	0.776
BMAA1542	type III secretion system protein BsaQ	0.771
BMAA1633	HrpB2-like protein	0.769
BMAA1107	alcohol dehydrogenase, iron-containing	0.767
BMAA1457	putative lipoprotein	0.766
BMAA1534	type III secretion system protein BsaY	0.764
BMA0994	copper ABC transporter, periplasmic copper-binding protein, putative	0.763
BMA0995	nitrous-oxide reductase precurser	0.762
BMAA1786	alkyl hydroperoxide reductase, subunit c	0.761
BMAA1449	putative syringomycin synthesis regulator SyrP	0.761
BMAA1628	type III secretion inner membrane protein SctR	0.759
BMAA1812	putative sugar ABC transporter, periplasmic sugar-binding protein	0.758
BMA3346	unknown function	0.757
BMAA1531	BipB protein	0.755
BMA0991	protein disulfide isomerase NosL, putative	0.755
BMAA1610	putative type IV pilus biogenesis protein PilN	0.754
BMAA0405	hypothetical protein	0.754
BMAA1536	type III secretion system protein BsaW	0.751
	BMAL-29	
BMA10299_A3397	putative lipoprotein	0.773
BMA10299_A3203	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9) (IMGterm)	0.744
BMA10299_A3014	putative outer membrane porin	0.766
BMA10299_A2984	ompA family protein	0.78
BMA10299_A2858	phospholipase C	0.77
BMA10299_A2852	radical SAM domain protein	0.785
BMA10299_A2805	acyl carrier protein	0.835
BMA10299_A2742	chaperonin, 60 kDa	0.821
BMA10299_A2542	endoribonuclease, L-PSP family	0.774
BMA10299_A2476	putative transcriptional regulator	0.748
BMA10299_A2336	thioesterase domain protein	0.758
BMA10299_A2129	putative threonine efflux protein	0.776
BMA10299_A2029	sodium/bile acid symporter family protein	0.763
BMA10299_A1467	putative ABC transporter, permease protein	0.765
BMA10299_A1465	putative syringomycin synthesis regulator SyrP	0.768
BMA10299_A0751	ketol-acid reductoisomerase (EC 1.1.1.86) (IMGterm)	0.844
	BMAL-47	
BMA10247_0398	ketol-acid reductoisomerase (EC 1.1.1.86) (IMGterm)	0.845
BMA10247_2943	ABC transporter, ATP-binding protein	0.802
BMA10247_2950	LysE family protein	0.8
BMA10247_3221	putative outer membrane porin	0.789
BMA10247_1750	radical SAM domain protein	0.786
BMA10247_A1658	putative outer membrane nitrite reductase	0.782
BMA10247_3517	putative lipoprotein	0.781
BMA10247_3387	putative threonine efflux protein	0.774
BMA10247_A0750	type III secretion system protein BsaY	0.772
BMA10247_A0742	type III secretion system protein BsaQ	0.769
BMA10247_2893	putative syringomycin synthesis regulator SyrP	0.768
BMA10247_A1276	alcohol dehydrogenase, iron-containing	0.767
BMA10247_2255	sodium/bile acid symporter family protein	0.764
BMA10247_2860	SCO1/SenC family protein	0.762
BMA10247_A2046	alkyl hydroperoxide reductase, subunit C	0.76
BMA10247_A0637	type III secretion inner membrane protein SctR	0.76
BMA10247_3050	aldehyde dehydrogenase family protein	0.759
BMA10247_A0753	BipB protein	0.754

Table-2: *Continued*

Locus tag of PRG	Description	CAI value
BMA10247_A0658	type IVB pilus formation outer membrane protein, R64 PilN family	0.754
BMA10247_A0748	type III secretion system protein BsaW	0.752
BMA10247_0916	arginine/ornithine antiporter	0.748
BMA10247_1384	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9) (IMGterm)	0.746
BMA10247_A0371	GatB/Yqey family protein	0.743
BMA10247_A1232	putative sugar ABC transporter, periplasmic sugar-binding protein	0.741
	BMAL-VP1	
BMASAVP1_A1115	ketol-acid reductoisomerase (EC 1.1.1.86) (IMGterm)	0.849
BMASAVP1_A0789	thioredoxin-disulfide reductase	0.83
BMASAVP1_A0165	cytochrome c oxidase, subunit III	0.801
BMASAVP1_A3005	flagellar hook-associated protein 3	0.799
BMASAVP1_A3526	lipoprotein, putative	0.792
BMASAVP1_0593	putative outer membrane nitrite reductase	0.791
BMASAVP1_A1655	FimA	0.788
BMASAVP1_A2959	putative threonine efflux protein	0.782
BMASAVP1_A1654	fimbrial biogenesis outer membrane usher protein	0.778
BMASAVP1_A0708	endoribonuclease, L-PSP family	0.775
BMASAVP1_A0123	putative syringomycin synthesis regulator SyrP	0.775
BMASAVP1_0107	alcohol dehydrogenase, iron-containing	0.775
BMASAVP1_A3064	sodium/bile acid symporter family protein	0.772
BMASAVP1_A3320	aldehyde.dehydrogenase family protein	0.77
BMASAVP1_A2581	OmpA family protein	0.767
BMASAVP1_0779	alkyl hydroperoxide reductase, subunit C	0.765
BMASAVP1_A3080	aminotransferase (EC 2.6.1.-) (IMGterm)	0.763
BMASAVP1_A1549	butyryl-CoA:acetate CoA transferase (EC 2.8.3.8) (IMGterm)	0.762
BMASAVP1_A1622	TonB-dependent siderophore receptor	0.761
BMASAVP1_1518	GatB/Yqey family protein	0.758
BMASAVP1_0065	putative sugar ABC transporter, periplasmic sugar-binding protein	0.757
BMASAVP1_A2111	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9) (IMGterm)	0.754
BMASAVP1_A1282	putative homoserine/threonine efflux protein	0.752

Post transnational modification, protein turnover and chaperons,(COG-2); *Metabolism* consisting of COGs related to C-Energy production and conversion-Carbohydrate transport and metabolism-Amino acid transport and metabolism's-Nucleotide transport and metabolism-Coenzyme transport and metabolism; I-Lipid transport and metabolism-Inorganic ion transport and metabolism-Secondary metabolites biosynthesis and transport, (COG-3);*General function prediction and unknown function* comprising of COGs related to R-General function prediction; S-Unknown function, (COG -3). Some pathogenicity related genes found in PHX category which are shown in the Table-2. There were 22, 16, 24 and 23 PR-PHX genes in BMAL-44, BMAL-29, BMAL-47and BMAL-VP1 respectively. Another study revealed that the most pathogenic strain of *Burkholderia mallei* i.e. BMAL-44 has 17 known PR-PHX genes, those are iron-containing alcohol dehydrogenase, alkyl hydroperoxide reductase, BipB protein, copper ABC transporter- periplasmic copper-binding protein, endoribonuclease, L-PSP family, HrpB2-like protein, nitrous-oxide reductase precursor, putative outer membrane nitrite reductase, putative lipoprotein, putative sugar ABC transporter, periplasmic sugar-binding protein, putative syringomycin synthesis regulator SyrP, protein disulfide isomerase NosL, putative, putative type IV pilus biogenesis protein PilN, type III secretion system protein BsaY,type III secretion system protein BsaQ,type III secretion inner membrane protein SctR and type III secretion system protein BsaW with mean CAI-value is 0.76±0.008 that is quiet high.but the CAI-value of rest three strains are 0.70±0.108, 0.70±0.103 and 0.68±0.122 for BMAL-29, BMAL-47and BMAL-VP1 respectively. It indicates that in BMAL-44, the virulence gene expression level was comparatively high.

Conclusion:

Moderate codon bias has been observed in the *Burkholderia* strains. Codon heterogeneity is associated with the gene expression levels. A number of pathogenesis related genes in the studied *Burkholderia* strains were horizontally transferred and had somewhat different pattern of codon usage compared to other genes. Correspondence analysis revealed variability among different sets of genes. A number of pathogenicity related genes directly associated with the virulence and toxicity in *Burkholderia* belonged to the PHX category. COGs associated with metabolism especially those linked to carbohydrate metabolism and amino acid transport and metabolism had the lion's share of PHX genes implying their significance in influencing the lifestyle of the *Burkholderia* strains as pathogens.

References

Akashi H, Eyre-Walker AC.1998. Translational selection and molecular evolution. *Curr Opin Genet Dev.* 8. 688–693.

Bulmer M.1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics*. 129: 897–907.

Cameron J, Aguade M.1998. An evaluation of measures of synonymous codon usage bias. *J Mol Evol.* 47:268-274.

Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proc Natl Acad Sci USA.* 96:4482–4487.

Godoy D, Randle G, Simpson AJ, Aanensen DM, Pitt TL, Kinoshita R, Spratt BG. 2003.Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J Clin Microbiol.* 41(5): 2068 - 2079.

Gouy M, Gautier C. 1982. Codon usage in bacteria correlation with gene expressivity. *Nucl Acid Res.* 10: 7055-7074.

Grantham R, Gautier C, Gouy M.1980. Codon frequencies in 119 genes confirm consistent choices of degenerate base according to genome type. *Nucl Acid Res.* 8: 1892-1912.

Groisman E A ,Sturmoski M A, Solomon F R ,Lin R ,Ochmann H. 1993. Molecular,functional and evolutionary analysis of sequence specific to *Salmonella*. *Proc Natl Acad Sci USA.* 90: 1033-1037.

Groisman E A, Saier M H,Ouchmann H, 1992. Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome. *EMBO Journal.* 11: 1309-1316.

Matic I, Rayssiguier C, Radman M. 1995. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell.* 80: 507-515.

Medigue C, Rouxel T,Vigier P,Henaut A, Danchin A, 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol.* 222: 851-856.

Miller WR,Pannell L,Cravitz L, Tanner WA, Rosebury T.1948. Studies on certain biological characteristics of *Malleomyces mallei* and *Malleomyces pseudo mallei*.II. Virulence and infectivity for animals.*J.bacterial.*55:127-135.

Nierman WC, DeShazer D, Kim HS, Tettelin H, Nelson KE, Feldblyum T, Ulrich RL, Ronning CM, Brinkac LM, Daugherty SC, Davidsen TD, Deboy RT, Dimitrov G, Dodson RJ, Durkin AS, Gwinn ML, Haft DH, Khouri H, Kolonay JF, Madupu R, Mohammoud Y, Nelson WC, Radune D, Romero CM, Sarria S, Selengut J, Shamblin C, Sullivan SA, White O, Yu Y, Zafar N, Zhou L, Fraser CM. 2004. Structural flexibility in the *Burkholderia mallei* genome. *Proc Natl Acad Sci USA.*.101(39): 14246-14251.

Peden J.1999.Analysis of codon usage. PhD thesis, The University of Nottingham, UK.

Schell MA, Lipscomband L, DeShazer D.2008. Comparative Genomics and an Insect Model Rapidly Identify Novel Virulence Genes of *Burkholderia mallei*. *J Bacteriol.* 2306 -2313.

Schell MA, Ricky L, Ulrich.2007. Type VI secretion is a major virulence determinant in Burkholderia mallei. Mol Microbiol .64(6): 1466–1485.

Sen A,Sur S,Bothra AK,Benson DR,Normand P,Tisa LS.2008.The implication of life style on codon usage patters and predicted highly expressed genes for three *Frankia* genomes. *Anton Leeuwenhoek Int J Gen.*93:335-346.

Sen G,Sur S, Bose D, Mondal U, Furnholm T, Bothra A, Tisa L ,Sen A.2007.Analysis of codon usage patterns and predicted highly expressed genes for six phytopathogenic *Xanthomonas* genomes shows a high degree of conservation; *In Silico Biol.*7:547-558.

Sharp PM, Li WH. 1987. The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications. *Nucl Acid Res.* 15: 1281-1295.

Sharp PM, Li WH. 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Mol Biol Evol.* 4: 223–230.

Sharp PM , Wen-Hsiung. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucl Acid Res.* 15(3): 1281-1295.

Simon M, Zieg J, Silverman M, Mandel G, Doolittle R, 1980. Phase variation: evolution of a controlling element. *Science* .209: 1370-1374.

Thibault FM, Hernandez E, Vidal DR, Girardet M, Cavallo JD. 2004. Antibiotic susceptibility of 65 isolates of *Burkholderia pseudomallei* and *Burkholderia mallei* to 35 antimicrobial agents. *The Journal of Antimicrobial Chemotherapy.* 54(6): 1134 - 1138.

Wright F. 1990. The effective number of codons used in a gene. *Gene.* 87: 23-29.

Wu G, Culley DE, W.Zhang. 2005. Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. *Microbiology Engl Tr.* 151:2175-2187.