Research Article
Open Access

Codon usage bias and peptide properties of *Pseudomonas balearica* DSM 6083^T

Abstract

Pseudomonas balearica DSM 6083^T has potential applications in bioremediation and its genome is recently sequenced. Codon usage bias is important in the study of evolutionary pressures on the organism and physical properties of peptides may elucidate functional peptides. However, both have not been studied for *P. balearica* DSM 6083^T. Here, we investigated the codon usage bias and peptide properties of the 4,050 coding sequences in *P. balearica*. Codon usage analysis suggests that all preferred codons were either G or C ending. There is a skew towards smaller peptides and all peptide properties (pI, aromaticity, hydropathy, and instability) are correlated ($|r| > 0.102$, p-value $< 7 \times 10^{-11}$). %GC is correlated ($|r| > 0.122$, p-value $< 6 \times 10^{-15}$) to peptide length, aromaticity, hydropathy, and instability. Peptide length is correlated ($|r| < 0.057$, p-value < 0.0003) to pI, aromaticity, and instability. Codon usage is correlated ($r < -0.042$, p-value < 0.0075) with all peptide properties while amino acid usage is correlated ($r < -0.084$, p-value $< 8 \times 10^{-8}$) to all peptide properties except instability. A substantial proportion (26.9%) of genes show significantly different codon and amino acid ratios compared to the genomic and proteomic averages respectively (p-value $< 1.2 \times 10^{-5}$), suggesting potential exogenous origins. These results suggest a complex interplay of metagenomic environment and various genomic/proteomic properties in shaping the evolution of *P. balearica* DSM 6083^T.

Volume 8 Issue 2 - 2019

Argho Maitra,[1,2] Maurice HT Ling[1,2,3]

[1]Department of Applied Sciences, Northumbria University, UK
[2]School of Life Sciences, Management Development Institute of Singapore, Singapore
[3]HOHY PTE LTD, Singapore

Correspondence: Maurice HT Ling, School of Life Sciences, Management Development Institute of Singapore, Singapore, Email mauricling@acm.org

Received: March 25, 2018 | **Published:** April 15, 2019

Introduction

Pseudomonas balearica is an environmental Gram-negative bacilliform bacterium with denitrifying capabilities and the ability to degrade several organic compounds; such as, naphthalene[1] and thiosulfate;[2] suggesting potential applications in bioremediation.[3] Biochemically, *P. balearica* and *Pseudomonas stutzeri* exhibit several common phenotypical traits;[4] such as, starch hydrolysis, maltose utilisation, arginine utilisation, and does not undergo gelatin hydrolysis. As a result, *P. balearica* was previously considered to be a genomovar of *P. stutzeri* but an analysis of 16S rRNA sequences[1] suggests that *P. balearica* should be considered distinct from *P. stutzeri*. Genotypical differentiation and phylogenetic proximity of *P. balearica* from *P. stutzeri* was established through comparative genomic sequence analysis using BLAST calculation of average nucleotide identity[5] and found 81.2% identity from the closest species, *P. stutzeri* ATCC 17588^T. The genome of *P. balearica* DSM 6083^T (CCUG 44595^T, SP1042^T) has been sequenced recently[6] – Accession number CP007511. This provides a resource to study the evolution, genomics and proteomics of *P. balearica.*

Codon usage bias (CUB) can be defined as the preferential bias creating a non-uniformity in the frequency of codon usage[7] and has been implicated in gene expression,[8–10] leading to the application of CUB in protein expression.[11] A study[12] demonstrates that CUBs of mammals, birds, insects, yeast, and bacteria correspond to evolutionary distance, suggesting that CUB is evolutionarily conserved. Besides selective pressure on codon usage;[13] other factors, such as guanine-cytosine (GC)-content;[14] and physical properties of the peptides (PPp), such as aromaticity and hydropathy; have been suggested to influence CUB.[15] PPp is an important aspect to study protein chemistry and elucidating potential functions.[16,17] However, CUB and PPp of *P. balearica* have not been studied.

In this study, we examine the CUB, %GC and PPp of *P. balearica* DSM 6083^T using its recently published sequence.[6] *P. balearica* is GC-rich with an average %GC of 64.6% with a preference for G and C ending codons, and multiple correlations between various genomic and proteomic properties in *P. balearica* DSM 6083^T. Significantly, a substantial proportion of genes appear exogenous, suggesting a significant role of horizontal gene transfer in its evolution.

Material and methods

Sequence data

A complete set of coding sequences (CDSes) was extracted from the genome sequence of *P. balearica* DSM 6083^T genome (Accession number CP007511.1), consisting of 4.38million base pairs. There are 4,126 genes; of which, 4,050 are CDS.

Biasness calculations

Codon usage bias can be calculated as Relative Synonymous Codon Usage (RSCU) or Codon Count Variation (CCV). RSCU is the ratio of observed to expected codon distribution, provided codon synonymy for identical amino acid holds true and is calculated for each CDS[11,18] as:

$$RSCU = \left(\frac{OF_{ij}}{\sum_j^{N_i} \left(OF_{ij} \right)} \right) N_i$$

where OF_{ij} is the observed codon distribution for the jth codon in the ith amino acid, N_i is the total number of codons encoding the ith amino acid and $\sum_j^{N_i} \left(OF_{ij} \right)$ is the overall expected codon distribution. RSCU value of 1 signifies no bias, while greater than or lesser than

1 signifies an increase or decrease in codon abundance respectively.[18] Unbiased codons (ATG for methionine and TGG for tryptophan) and stop codons were not considered in RSCU analysis. CCV is the codon usage deviation from expected codon usage for each amino acid[19] and can be calculated as the probability value using Chi-Square test with Bonferroni correction.[20] Similarly, Amino Acid Variation (AAV) is the deviation of amino acid count from expected distribution of amino acid count, which can also be calculated as the probability value using Chi-Square test with Bonferroni correction.[20]

Nucleotide composition

Two sets of nucleotide compositions were calculated. Firstly, each CDS was calculated for nucleotide composition regardless of position within a codon and is denoted as nucleotide percentage. For example, %GC refers to combined percentage of guanine and cytosine. Secondly, each CDS was calculated for nucleotide composition with regards to its position within a codon[21] and is denoted as positional nucleotide percentage. For example, %GC3 refers to combined percentage of guanine and cytosine at the third base of a codon. In extension, %GC12 refers to combined percentage of guanine and cytosine at the first and second base of a codon.

Physical Properties of Peptides (PPp)

Aromaticity refers to the relative abundance of aromatic amino acids in a peptide.[22] Hydropathy (GRAVY) refers to the overall hydrophobic/hydrophilic properties of a peptide.[23] Isoelectric point (pI) is the pH where a peptide is electrical neutrality.[24] Instability index refers to the stability of the peptide where high instability score suggests shorter half-life.[25] All four methods are available in Biopython library.[26]

Statistical analysis

Statistical analysis was performed using Pearson's Chi-Square test with Bonferroni correction[20] and regression analysis with Pearson's product moment correlation coefficient were carried out using Microsoft® Excel for Mac (version 16.22) in macOS Mojave (version 10.14.1). Significance of Pearson's correlation was carried out using t-test for correlation coefficient. Significance between two Pearson's correlations was carried out using Z-test for two correlation coefficients.

Results and discussion

Uneven distribution of genomic features

The *P. balearica* DSM 6083^T genome has a sequence length of 4,383,480bp. It lacks extra chromosomal elements and has 8,126 features; of which, there are 4,126 genes, 4,050 CDSes, 60 transfer RNAs, 3 non-coding RNAs, 12 ribosomal RNAs, 1 transfer-messenger RNA, 5 regulatory sequences, 1 miscellaneous feature and 2 repeats. Feature map across genome can provide an overview, showing distributions of various genomic features.[27] A visual inspection of the feature map (Figure 1) suggests that the genomic features are not likely to be evenly distributed across the genome, which is supported by studies showing non-randomness in genomic architecture.[28,29] For example, regulatory features appear to be clustered differently to transfer RNAs.

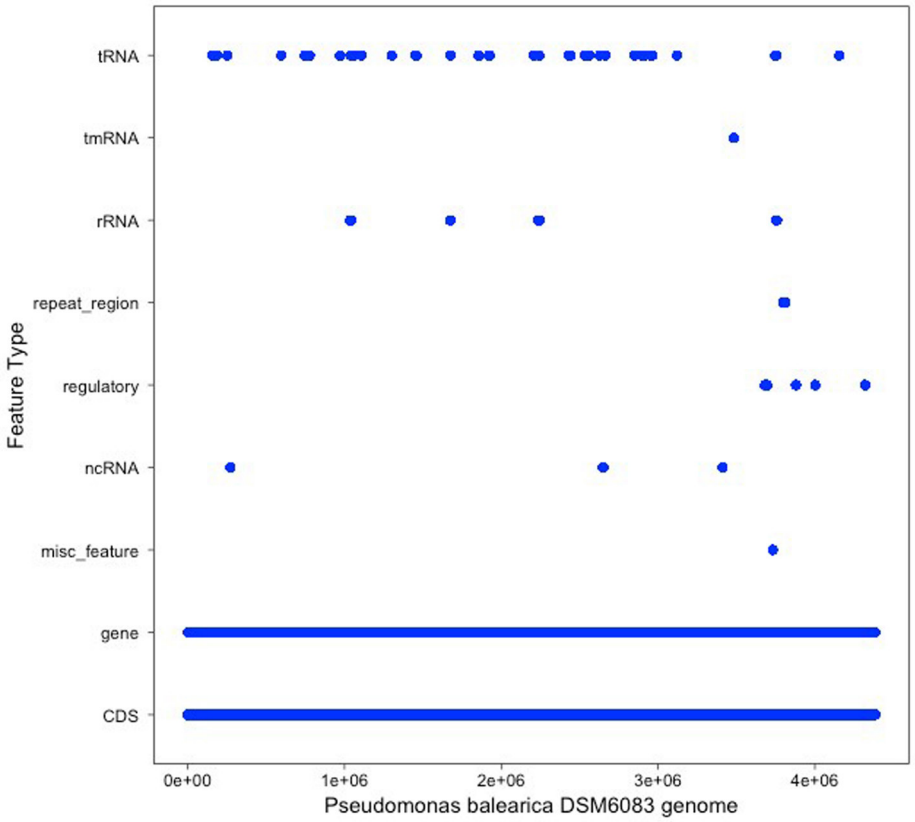


Figure 1 Feature map for the *P. balearica* DSM 6083^T genome. The features represent coding sequence (CDS), gene, miscellaneous features (misc_feature), non-coding RNA (ncRNA), regulatory sequences (regulatory), repeats (repeat_region), ribosomal RNA (rRNA), transfer-messenger RNA (tmRNA) and transfer RNA (tRNA).

Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas balearica DSM 6083^T. *MOJ Proteomics Bioinform.* 2019;8(2):27–39. DOI: 10.15406/mojpb.2019.08.00263

P. balearica is GC rich but CDSes show substantial GC variation

%GC ranges between 50.9% and 70.5% for all 4,126 *P. balearica* genes and the average %GC is 64.6% with a standard deviation of 3.34%. The %GC of the 4,050 CDSes ranges from 38.7% to 75.1% and the average %GC is 64.6% with a standard deviation of 4.44% (Figure 2). These results indicate a GC-rich bias in the genome (p-value $<1x10^{-200}$) and %GC variation across genome has also been shown.[30] Global %GC has been shown to be strongly associated with positional %GC, amino acid usage and CUB.[31,32] In addition, higher %GC in CDSes is indicative of greater bias in synonymous codon usage.[31,33]

P. balearica Prefers G and C Ending Codons

4,050 *P. balearica* CDSes provided a total of 1,297,664 codons, based on which RSCU values were tabulated (Table 1). Our results show substantial bias in codon usage in *P. balearica* where 44% (n=26) of the total codons show a high degree of abundance (RSCU >1), 12% (n=7) are less abundant (0.5<RSCU <1) and the remaining 44% (n=26) are under-represented (RSCU <0.5). 16 of the 26 highly abundant codons (RSCU >1.6); namely, AAG, AAC, ACC, CGC, AGC, TCG, ATC, CAG, CCG, CTG, GCC, GGC, GTC, GTG, TGC, TTC; are superior and over-represented.[34,335] However, all abundant codons (RSCU >1) are G and C ending codons where CTG has the highest RSCU value of 3.79, whereas the non-preferred codons end mostly with an A or T. These indicate that *P. balearica* prefers G and C ending codons. This is consistent with a study on extremophiles showing preference for G and C ending codons,[36] which is consistent to the environment where P. *balearica* may be found.[37] At the same time, compositional constraints of nucleotides (G or C) is one of the major factors that profoundly affect the overall CUB in a genome.[18]

Table 1 RSCU Table for *P. balearica* CDSes. AA stands for amino acids, N is the total number of codons encoding the respective amino acid and RSCU represents the frequency of codon occurrence. Cells with RSCU values greater than 1 (highlighted in red) show highly abundant codons and values greater than 1.6 (shown in bold) show superior and over-represented codons.

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Lys	AAA	6112	0.32	Leu	CTA	2864	0.11
	AAG	32495	1.68		CTC	35106	1.34
Asn	**AAC**	26421	1.56		**CTG**	99670	3.79
	AAT	7375	0.44		CTT	7072	0.27
Thr	ACA	2640	0.18		TTA	802	0.03
	ACC	36537	2.53		TTG	12196	0.46
	ACG	15327	1.06	Glu	GAA	29071	0.72
	ACT	3320	0.23		GAG	51494	1.28
Arg	AGA	972	0.06	Asp	GAC	47100	1.37
	AGG	2259	0.14		GAT	21668	0.63
	CGA	4443	0.28	Ala	GCA	12607	0.33
	CGC	57445	3.62		**GCC**	77448	2.05
	CGG	16391	1.03		GCG	50850	1.35
	CGT	13692	0.86		GCT	10257	0.27
Ser	**AGC**	29752	2.60	Gly	GGA	4211	0.16
	AGT	4492	0.39		**GGC**	70514	2.70
	TCA	2036	0.18		GGG	14566	0.56
	TCC	11497	1.01		GGT	15090	0.58
	TCG	18970	1.66	Val	GTA	5811	0.26
	TCT	1814	0.16		**GTC**	36870	1.62
Ile	ATA	1494	0.08		**GTG**	41388	1.82
	ATC	48025	2.49		GTT	6722	0.30
	ATT	8260	0.43	Tyr	TAC	22192	1.40
Met	ATG	28244	1.00		TAT	9488	0.60
Gln	CAA	7614	0.26	Trp	TGG	18715	1.00
	CAG	50082	1.74	Cys	TGT	2094	0.32
His	CAC	19065	1.30		**TGC**	11071	1.68
	CAT	10216	0.70	Phe	**TTC**	37612	1.65
Pro	CCA	4969	0.31		TTT	8015	0.35
	CCC	16234	1.01	Stop	TAA	591	0.44
	CCG	38530	2.40		TAG	525	0.39
	CCT	4357	0.27		**TGA**	2904	2.17

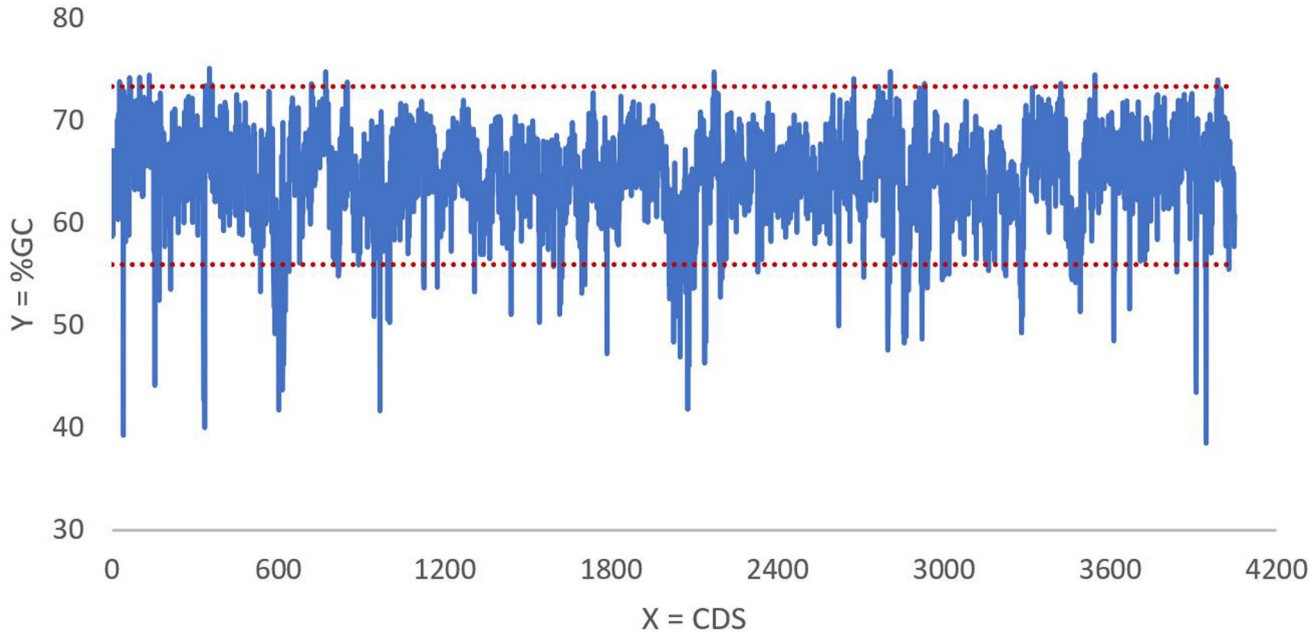


Figure 2 Varying %GC of 4,050 P. *balearica* CDSes. The dotted lines show 95% confidence interval of %GC.

Marginal but significant correlations between peptide physical properties

Our results show that majority of the peptides in the *P. balearica* genome are less than 500 amino acids in length (Figure 3). This suggests a preference for shorter peptides, which is supported by other studies[7,38] suggesting a drawback in longer peptides over the shorter ones due to energy expenditure in event where both are homogenous in function. 54% of peptides have electrical neutrality between the pH 5 to 7 (Figure 4A), which is consistent to a study 5,029 proteomes.[39] 65% of the peptides have an aromaticity index between 0.05 and 0.1 (Figure 4B) and 89% show hydrophobic character varying from -1 to 0.6 (Figure 4C). 70% of the peptides have an instability index between 30 and 50 (Figure 4D). An instability index above 40 suggests unstable peptide and is indicative of short half-life based on the documentation regarding instability index in Biopython library.[26] This suggests that about half of the proteins in *P. balearica* may have a short half-life.

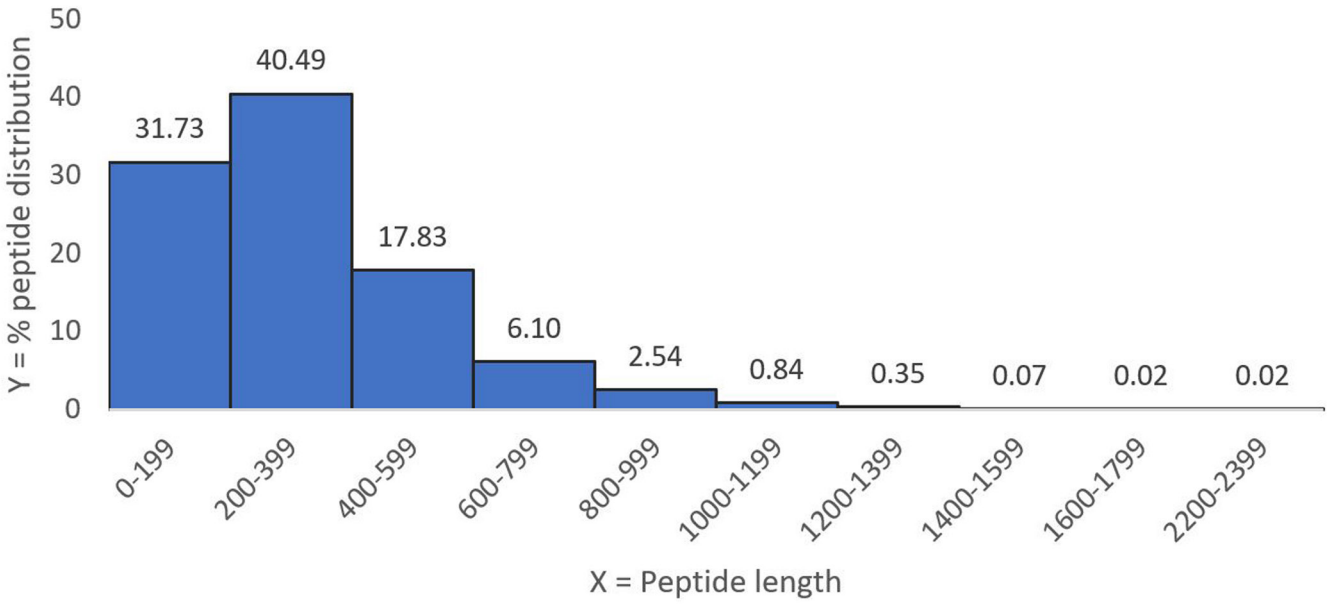


Figure 3 Distributions of Peptides by Length.

Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas balearica DSM 6083[T]. *MOJ Proteomics Bioinform.* 2019;8(2):27–39. DOI: 10.15406/mojpb.2019.08.00263

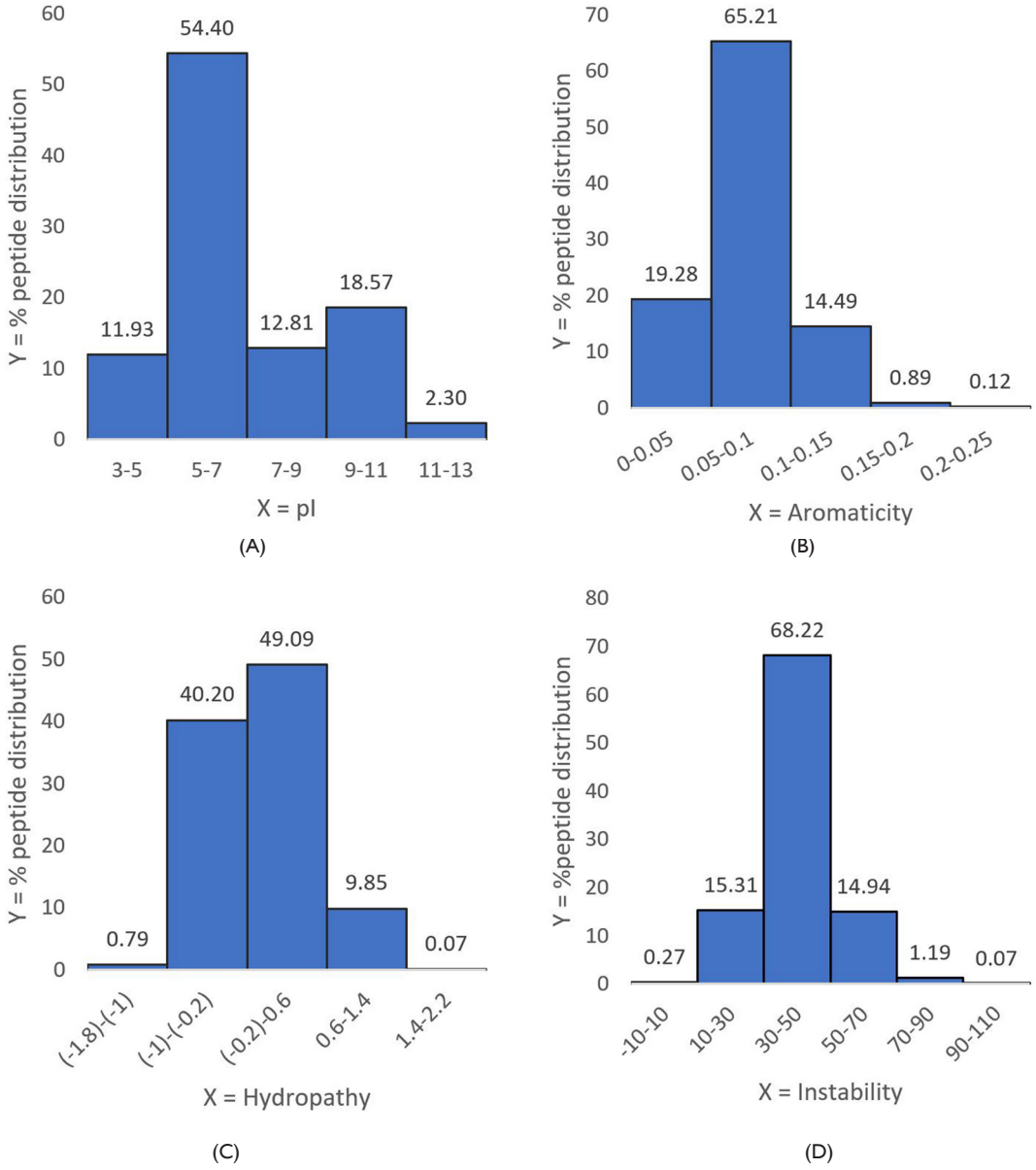


Figure 4 Distribution of Peptides by Physical Properties. (A) Shows the distribution of peptides by pI. (B) Shows the distribution of peptides by aromaticity. (C) Shows the distribution of peptides by hydropathy. (D) Shows the distribution of peptides by instability.

Regression analyses were carried out between pI, aromaticity index, hydropathy, and instability index on the CDSes. Our results show significant correlation between pI and aromaticity (r=0.161, p-value=6 x 10^{-25}, Figure 5A), pI and hydropathy (r=0.212, p-value=2 x 10^{-42}, Figure 5B), pI and instability (r=0.102, p-value=7 x 10^{-11}, Figure 5C), and aromaticity and hydropathy (r=0.259, p-value=4 x 10^{-63}, Figure 5D), aromaticity and instability (r=-0.104, p-value=3x10^{-11}, Figure 5E), and hydropathy and instability (r=-0.296, p-value=1x10^{-82}, Figure 5F). Scatterplots of pI (Figures 5A to 5E) is consistent with bimodal distribution of peptides with low fractions at pI close to 7.4.[40]

% GC is correlated to hydropathy, aromaticity and instability but not to pI

Global GC content and PPp of a genome are strongly correlated with CUB and amino acid usage;[41] however, their interrelation has not been studied in *P. balearica*. Regression analyses were conducted to investigate the relationship between *P. balearica* genome and proteome. Our findings revealed that there are significant positive correlations between %GC and hydropathy (r=0.146, p-value=9 x 10^{-21}, Figure 6C), %GC and instability index (r=0.122, p-value=6 x 10^{-15}, Figure 6D). Studies on prokaryotes have suggested positive correlation between %GC3 and hydropathy,[42] and between %GC and %GC3;[43] hence, inferring a potential positive correlation between %GC and hydropathy. However, despite our results showing positive correlation between hydropathy and aromaticity (r=0.259), %GC and aromaticity is significantly negative correlated (r=-0.170, p-value=1 x 10^{-27}, Figure 6B). In addition, pI is not correlated to %GC (r=0.026, p-value=0.098, Figure 6A).

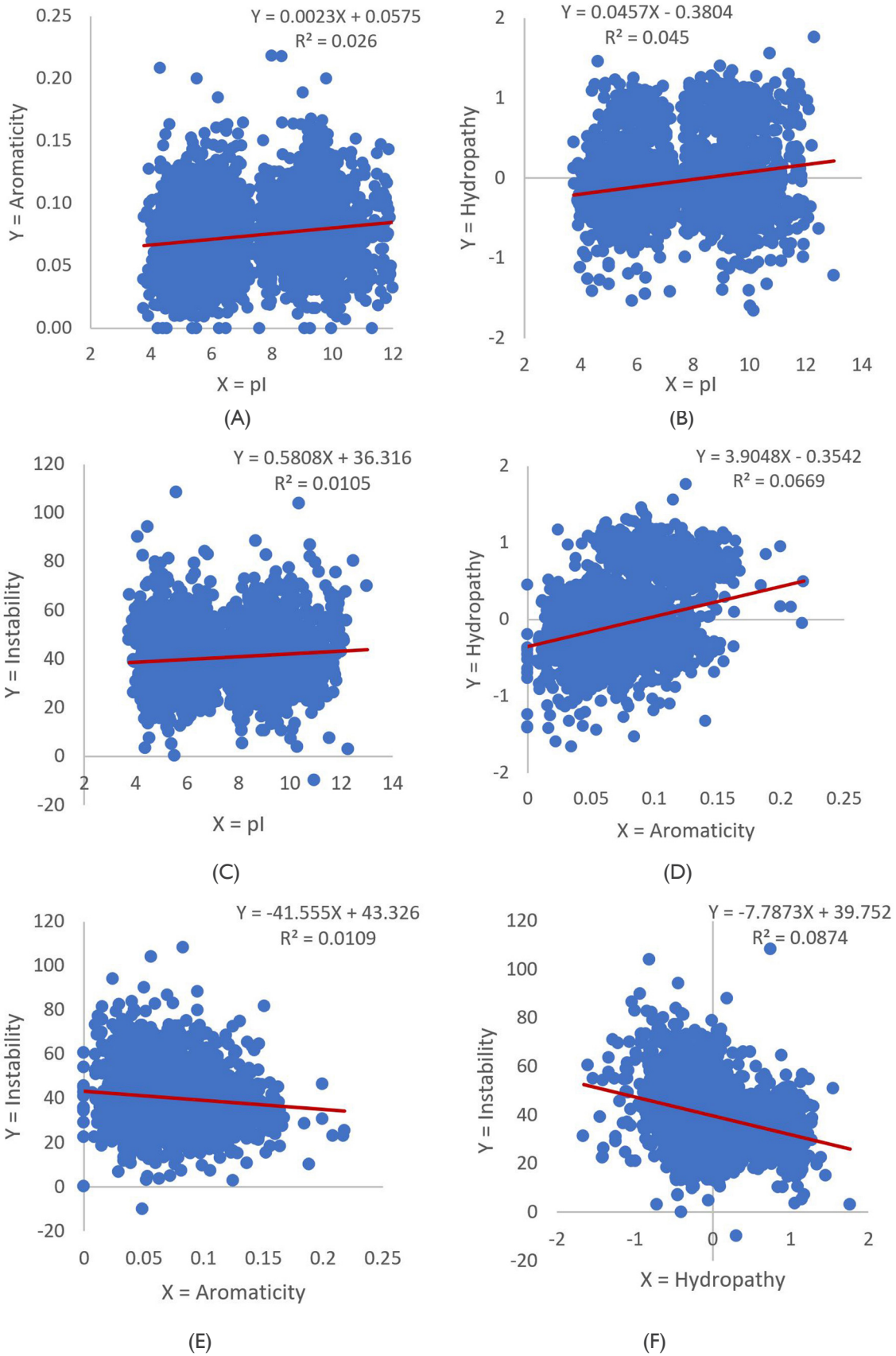


Figure 5 Relationships of Physical Properties of Peptides (n =4,050 CDSes). (A) Is the relationship between pI and aromaticity. (B) Is the relationship between pI and hydropathy. (C) Is the relationship between pI and instability. (D) Is the relationship between aromaticity and hydropathy. (E) Is the relationship between aromaticity and instability. (F) Is the relationship between hydropathy and instability.

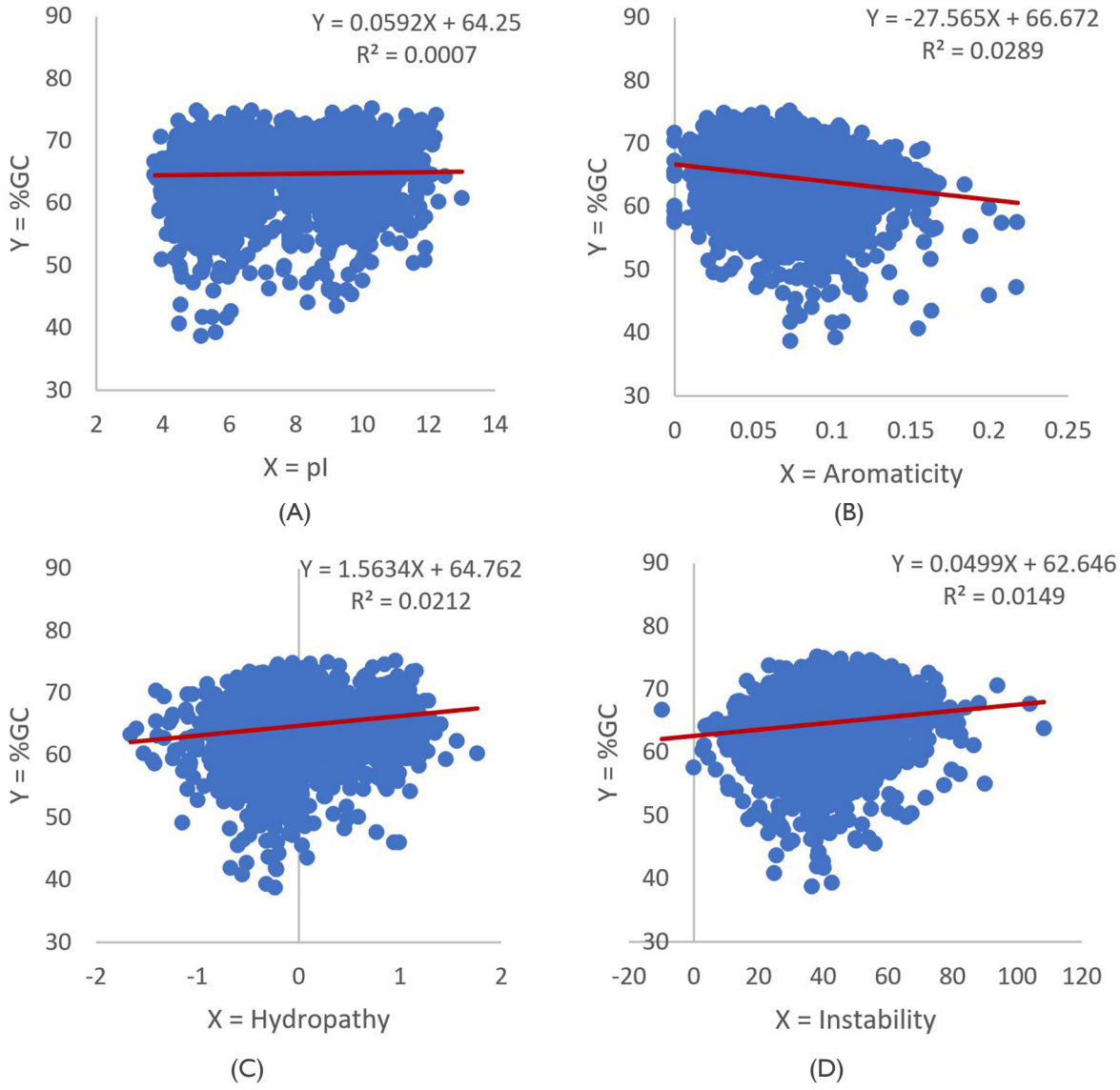


Figure 6 Relationship between Nucleotide Composition and Peptide Physical Properties (n = 4,050 CDSes). (A) Is the relationship between pI and %GC. (B) Is the relationship between aromaticity and %GC. (C) Is the relationship between hydropathy and %GC. (D) Is the relationship between instability and %GC.

Peptide length is correlated to %GC, %GC1, and %GC3, but not to %GC2, and %GC12

Regression analyses were performed to examine the relationship between nucleotide composition and peptide length. Our results showed that there is a positive significant correlation between peptide length and %GC (r=0.145, p-value=1 x 10^{-20}, Figure 7A), %GC1 (r=0.054, p-value=0.0006, Figure 7B) and %GC3 (r=0.210, p-value=1 x 10^{-41}, Figure 7D). However, %GC2 (r=-0.03, p-value=0.056, Figure 7C) and %GC12 (r=0.017, p-value=0.279, Figure 7E) are not significantly correlated to peptide length. D'Onofrio et al.[42] have suggested higher correlation (r=0.95) between %GC3 and %GC1 than %GC3 and %GC2 (r=0.89) in prokaryotes.

Peptide length is correlated to pI, aromaticity, and instability, but not to hydropathy

Regression analyses were conducted to examine the relationship between peptide length and PPp in *P. balearica* proteome. Our results showed that there is a significantly negative correlation between pI and peptide length (r=-0.160, p-value=1 x 10^{-24}, Figure 8A) and instability and peptide length (r=-0.072, p-value=4 x 10^{-6}, Figure 8D). Relationship between peptide size and pI[40] and peptide size and its stability[44] have been suggested. Our results suggest a significantly positive correlation between aromaticity and peptide length (r=0.057, p-value=0.0003, Figure 8B). Hydropathy showed no correlation to peptide length (p-value=0.913, Figure 8C).

Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas balearica DSM 6083[T]. *MOJ Proteomics Bioinform.* 2019;8(2):27–39. DOI: 10.15406/mojpb.2019.08.00263

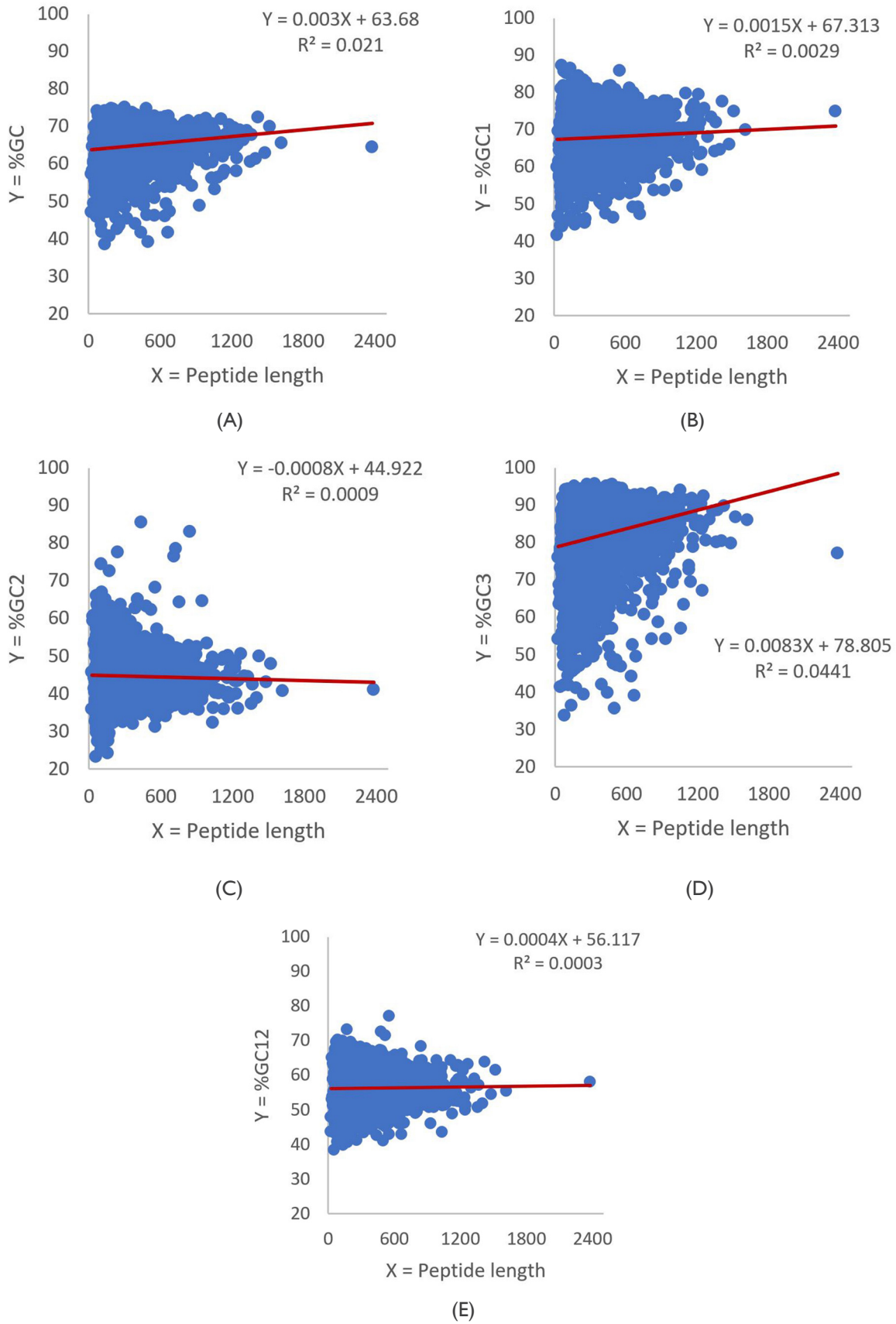


Figure 7 Relationships between Nucleotide Composition and Peptide Length (n=4,050 CDSes). (A) Is the relationship between peptide length and %GC. (B) Is the relationship between peptide length and %GC1. (C) Is the relationship between peptide length and %GC2. (D) Is the relationship between peptide length and %GC3. (E) Is the relationship between peptide length and %GC12.

Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas balearica DSM 6083^T. *MOJ Proteomics Bioinform.* 2019;8(2):27–39. DOI: 10.15406/mojpb.2019.08.00263

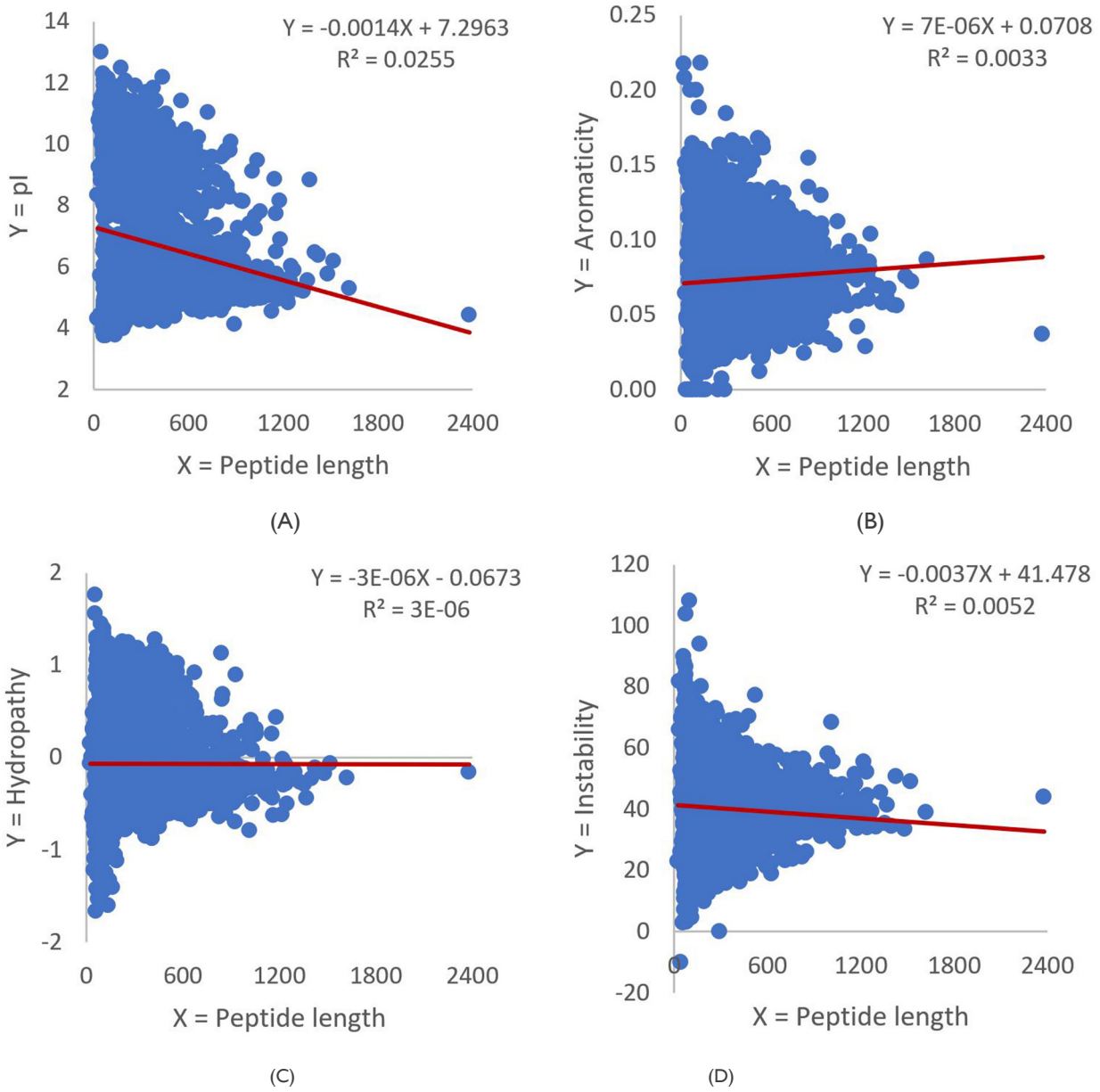


Figure 8 Relationships between Peptide Length and Physical Properties of Peptides (n=4,050 CDSes). (A) Is the relationship between peptide length and pI. (B) is the relationship between peptide length and aromaticity. (C) Is the relationship between peptide length and hydropathy. (D) Is the relationship between peptide length and instability.

Usage biases are negatively correlated to peptide length

Regression analyses show a non-linear relationship between peptide lengths and usage biases. There are significantly negative correlations between peptide length with amino acid variation ($|r|=6 \times 10^{-64}$, Figure 9A) and codon count variation ($|r|=8 \times 10^{-286}$, Figure 9B). These findings are consistent with *Taenia solium*[7] suggesting inverse relationship between the length of peptide to both CUB and amino acid usage.

Using the average codon count and amino count across all 4,050 CDSes as null hypotheses in CCV and AAV respectively, a substantial proportion of the CDSes (26.9%) show both significant CCV and AAV (p-value threshold=1.2×10^{-5} after Bonferroni correction). Statistically significant deviation of feature properties from the rest of the genome, also known as composition based[45] or parametric[46] approach, can be a method to identify horizontal transferred genes (HTG) as deviation without statistical support has been shown to be insufficient.[47] This led to the development of composition based statistical approaches to identify HGT,[45,48] which is the principle behind our CCV and AAV methods. Hence, our result may suggest a substantial proportion of the exogenous genes in *P. balearica* as HGT has been considered prevalent in bacteria.[49,50]

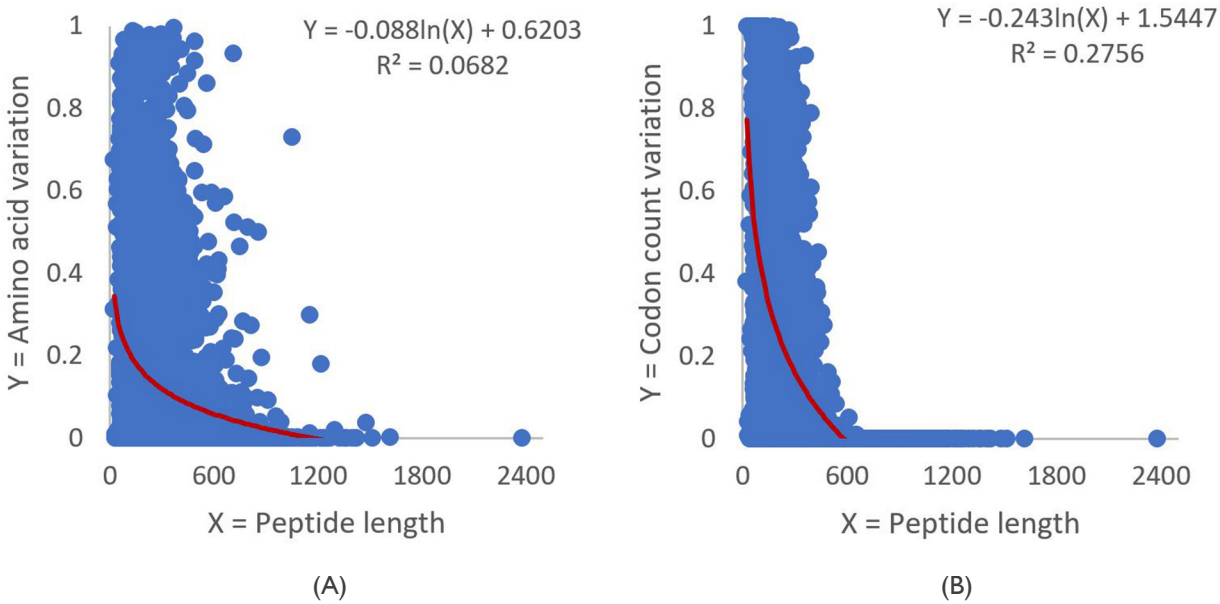


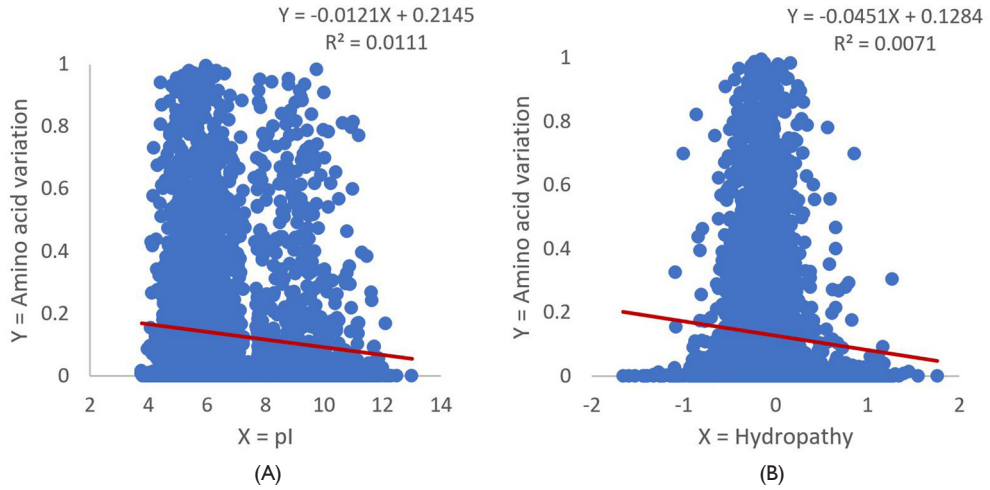
Figure 9 Relationship between Peptide Length and Usage Bias. (A) Shows relationship between peptide length and amino acid variation. (B) Shows relationship between peptide length and codon count variation.

Usage biases are correlated to peptide physical properties except AAV-instability pair

Relationship between PPp to codon count variation (CCV) and amino acid variation (AAV) were analyzed. Our results show significant correlations between CCV to all PPp. There are significantly negative correlations between CCV and pI (r=-0.042, p-value=0.0075, Figure 10E), aromaticity (r=-0.108, p-value=5 x 10^{-12}, Figure 10G), and hydropathy (r=-0.090, p-value=9 x 10^{-9}, Figure 10F) but significantly positive correlation with instability (r=0.093, p-value=3 x 10^{-9}, Figure 10H). These are consistent with platyhelminth mitochondrial genome analyses,[51] suggesting that hydrophobicity and aromaticity are significantly associated with CUB patterns.

In terms of amino acid usage, our results show significant negative correlations to all PPp except instability (r=0.0007, p-value=0.964, Figure 10D). There are significantly negative correlations between AAV and pI (r=-0.105, p-value=2 x 10^{-11}, Figure 10A), aromaticity (r=-0.086, p-value=4 x 10^{-8}, Figure 10C) and hydropathy (r=-0.084, p-value=8 x 10^{-8}, Figure 10B). These are consistent with studies on *Ginkgo biloba*[52] and chicken proteome analysis,[41] suggesting that variability in the amino acid usage can be attributed to the global hydrophobicity and aromatic amino-acid content of proteins. In addition, Lobry et al.[22] identified global hydrophobicity and aromaticity of proteins as the two essential factors that drives the bias in the amino acid usage. Besides these factors, instability and pI also show marginal but significant association with CUB.

By comparing the proportion of amino acids to various kingdoms,[39] the amino acid ratios in *P. balearica* proteome is leucine and alanine dominant (Figure 11) and most correlated to eubacteria (r=0.906) compared to archaebacteria (r=0.850, p-value=0.468), eukaryotes (r=0.878, p-value=0.688) or viruses (r=0.819, p-value=0.306). This suggests a "universal prevalence" of amino acid usage ratios across biotic life. In the *P. balearica* proteome, aromatic amino acids like tryptophan, tyrosine and histidine occur less abundantly as compared to hydrophobic amino acids like alanine, leucine and valine. This suggests that the variability in the amino acid usage patterns may be dominated by the hydrophobicity of proteins. Taken together, this suggests the diversity of evolutionary pressures that may act on *P. balearica*.



Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas balearica DSM 6083[T]. *MOJ Proteomics Bioinform.* 2019;8(2):27–39. DOI: 10.15406/mojpb.2019.08.00263

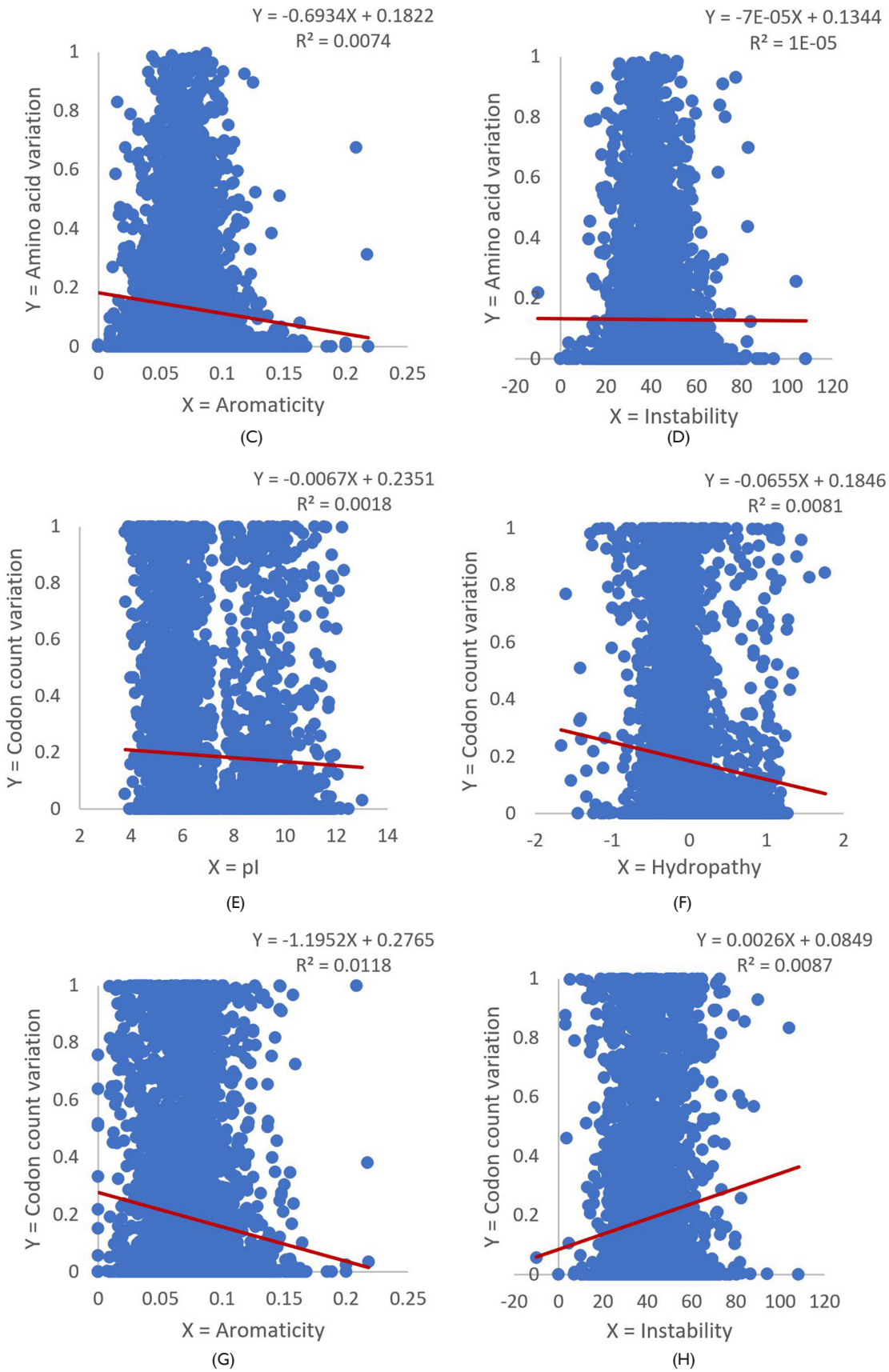


Figure 10 Relationships between Usage Bias and Physical Properties of Peptides (n = 4,050 CDSes). (A) Shows the relationship between pI and amino acid variation. (B) Shows the relationship between hydropathy and amino acid variation. (C) Shows the relationship between aromaticity and amino acid variation. (D) Shows the relationship between instability and amino acid variation. (E) Shows the relationship between pI and codon count variation. (F) Shows the relationship between hydropathy and codon count variation. (G) Shows the relationship between aromaticity and codon count variation. (H) Shows the relationship between instability and codon count variation.

Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas balearica DSM 6083[T]. *MOJ Proteomics Bioinform.* 2019;8(2):27–39. DOI: 10.15406/mojpb.2019.08.00263

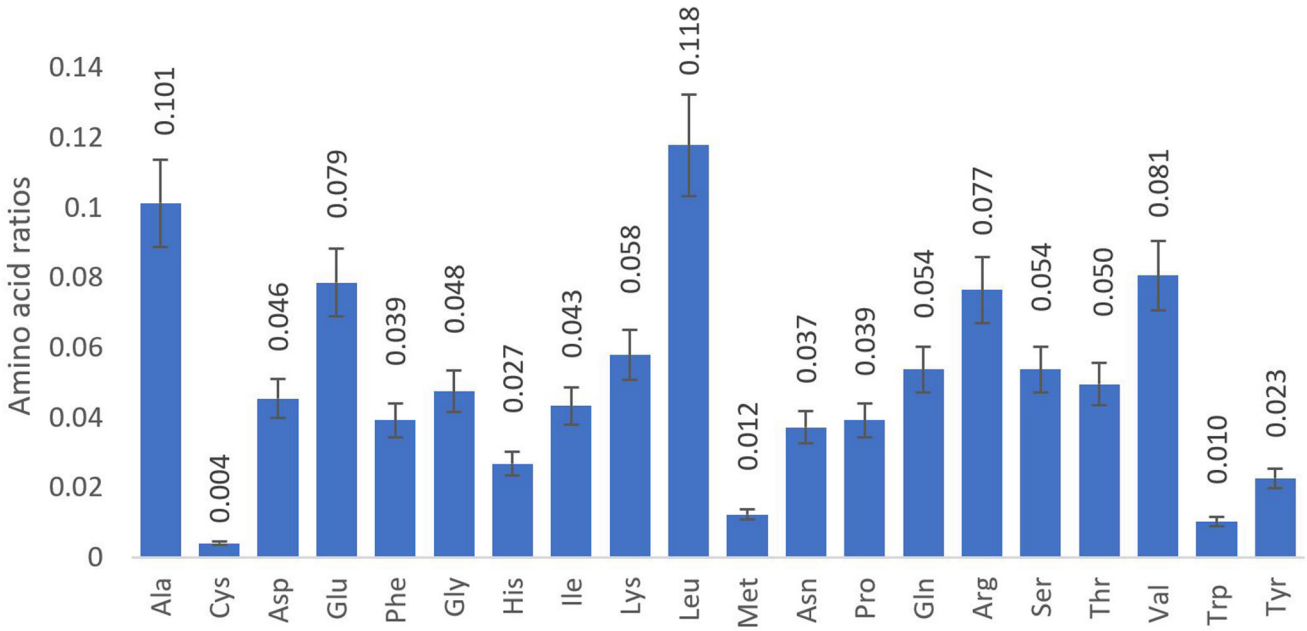


Figure 11 Proportion of amino acid usage.

Acknowledgments

None.

Conflicts of interest

The authors declare no conflict of interest.

References

1. Bennasar A, Rosselló–Mora R, Lalucat J, et al. 16S rRNA gene sequence analysis relative to genomovars of Pseudomonas stutzeri and proposal of Pseudomonas balearica sp. nov. *Int J Syst Bacteriol*. 1996;46(1):200–205.
2. Sorokin D, Teske A, Robertson L, et al. Anaerobic oxidation of thiosulfate to tetrathionate by obligately heterotrophic bacteria, belonging to the Pseudomonas stutzeri group. *FEMS Microbiol Ecol*. 1999;30(2):113–123.
3. ElBestawy E, Sabir J, Mansy A, et al. Comparison among the Efficiency of Different Bioremediation Technologies of Atrazine–Contaminated Soils. *Journal of Bioremediation & Biodegradation*. 2014;05:237.
4. Lalucat J, Bennasar A, Bosch R, et al. Biology of Pseudomonas stutzeri. *Microbiol Mol Biol Rev*. 2006;70(2):510–547.
5. Richter M, Rosselló–Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA*. 2009;106(45):19126–19131.
6. Bennasar–Figueras A, Salvà–Serra F, Jaén–Luchoro D, et al. Complete Genome Sequence of Pseudomonas balearica DSM 6083T. *Genome Announc*. 2016;4(2): e00217.
7. Yang X, Ma X, Luo X, et al. Codon Usage Bias and Determining Forces in Taenia solium Genome. *Korean J Parasitol*. 2015;53(6):689–697.
8. Lin K, Kuang Y, Joseph JS, et al. Conserved codon composition of ribosomal protein coding genes in Escherichia coli, Mycobacterium tuberculosis and Saccharomyces cerevisiae: lessons from supervised machine learning in functional genomics. *Nucleic Acids Res*. 2002;30(11):2599–607.
9. 9. Sabi R, Tuller T. Modelling the efficiency of codon–tRNA interactions based on codon usage bias. *DNA Res*. 2014;21(5):511–26.
10. Roymondal U, Das S, Sahoo S. Predicting gene expression level from relative codon usage bias: an application to Escherichia coli genome. DNA research : an international journal for rapid publication of reports on genes and genomes. *DNA Res*. 2009;16(1):13–30.
11. Villada JC, Brustolini OJB, Batista da Silveira W. Integrated analysis of individual codon contribution to protein biosynthesis reveals a new approach to improving the basis of rational gene design. *DNA Res*. 2017;24(4):419–434.
12. Keng BM, Chan OY, Ling MH. Codon usage bias is evolutionarily conserved. *Asia Pacific Journal of Life Sciences*. 2014;7(3):233–242.
13. Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet*. 2008;42:287–299.
14. Sueoka N, Kawanishi Y. DNA G+C content of the third codon position and codon usage biases of human genes. *Gene*. 2000;261(1):53–62.
15. Huang X, Xu J, Chen L, et al. Analysis of transcriptome data reveals multifactor constraint on codon usage in Taenia multiceps. *BMC Genomics*. 2017;18(1):308.
16. Liang L, Tan X, Juarez S, et al. Systems biology approach predicts antibody signature associated with Brucella melitensis infection in humans. *J Proteome Res*. 2011;10(10):4813–4824.
17. Mayers C, Duffield M, Rowe S, et al. Analysis of known bacterial protein vaccine antigens reveals biased physical properties and amino acid composition. *Comp Funct Genomics*. 2003;4(5):468–478.
18. Nasrullah I, Butt AM, Tahir S, et al. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evol Biol*. 2015;15:174.
19. Shields DC, Sharp PM, Higgins DG, et al. "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol*. 1988;5(6):704–716.
20. VanderWeele TJ, Mathur MB. Some Desirable Properties of the Bonferroni Correction: Is the Bonferroni Correction Really So Bad? *Am J Epidemiol*. 2019;188(3):617–618.

21. Halder B, Malakar AK, Chakraborty S. Nucleotide composition determines the role of translational efficiency in human genes. *Bioinformation*. 2017;13(2):46–53.

22. Lobry JR, Gautier C. Hydrophobicity, expressivity and aromaticity are the major trends of amino–acid usage in 999 Escherichia coli chromosome–encoded genes. *Nucleic Acids Research*. 1994;22(15):3174–3180.

23. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157(1):105–132.

24. Bjellqvist B, Hughes GJ, Pasquali C, et al. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*. 1993;14(10):1023–1031.

25. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*. 1990;4(2):155–161.

26. Cock PJA, Antao T, Chang JT, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422–1423.

27. Ling MHT. Back–of–the–Envelope Guide (A Tutorial) to 10 Intracellular Landscapes. *MOJ Proteomics & Bioinformatics*. 2018;7(1):31–36.

28. Rocha EPC. The organization of the bacterial genome. *Annu Rev Genet*. 2008;42:211–233.

29. Repar J, Warnecke T. Non–Random Inversion Landscapes in Prokaryotic Genomes Are Shaped by Heterogeneous Selection Pressures. *Mol Biol Evol*. 2017;34(8):1902–1911.

30. Bohlin J, Eldholm V, Pettersson JHO, et al. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC Genomics*. 2017;18(1):151.

31. Li J, Zhou J, Wu Y, et al. GC–Content of Synonymous Codons Profoundly Influences Amino Acid Usage. *G3 (Bethesda)*. 2015;5(10):2027–2036.

32. Wan XF, Xu D, Kleinhofs A, et al. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol*. 2004;4:19.

33. Palidwor GA, Perkins TJ, Xia X. A general model of codon bias due to GC mutational bias. *PLoS One*. 2010;5(10):e13431.

34. Gun L, Yumiao R, Haixian P, et al. Comprehensive Analysis and Comparison on the Codon Usage Pattern of Whole Mycobacterium tuberculosis Coding Genome from Different Area. *Biomed Res Int*. 2018;2018:1–7.

35. Nath Choudhury M, Uddin A, Chakraborty S. Codon usage bias and its influencing factors for Y–linked genes in human. *Comput Biol Chem*. 2017;69:77–86.

36. Khan MF, Patra S. Deciphering the rationale behind specific codon usage pattern in extremophiles. *Sci Rep*. 2018;8(1):15548.

37. Salvà–Serra F, Jakobsson HE, Busquets A, et al. Genome Sequences of Two Naphthalene–Degrading Strains of Pseudomonas balearica, Isolated from Polluted Marine Sediment and from an Oil Refinery Site. *Genome Announc*. 2017;5(14):e00116–e00117.

38. Moriyama EN, Powell JR. Codon usage bias and tRNA abundance in Drosophila. *J Mol Evol*. 1997;45(5):514–523.

39. Kozlowski LP. Proteome–pI: proteome isoelectric point database. *Nucleic Acids Res*. 2017;45(D1):D1112–126.

40. Kiraga J, Mackiewicz P, Mackiewicz D, et al. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC Genomics*. 2007;8:163–163.

41. Rao Y, Wang Z, Chai X, et al. Hydrophobicity and aromaticity are primary factors shaping variation in amino acid usage of chicken proteome. *PloS One*. 2014;9(10):e110381–e110381.

42. D'Onofrio G, Jabbari K, Musto H, et al. The correlation of protein hydropathy with the base composition of coding sequences. *Gene*. 1999;238(1):3–14.

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

44. Rahman M, Sadygov RG. Predicting the protein half–life in tissue from its cellular properties. *PloS One*. 2017;12(7):e0180428.

45. Adato O, Ninyo N, Gophna U, Snir S. Detecting Horizontal Gene Transfer between Closely Related Taxa. *PLoS Comput Biol*. 2015;11(10):e1004408.

46. Ravenhall M, Škunca N, Lassalle F, et al. Inferring horizontal gene transfer. *PLoS Comput Biol*. 2015;11(5):e1004095.

47. Wang B. Limitations of Compositional Approach to Identifying Horizontally Transferred Genes. *Journal of Molecular Evolution*. 2001;53(3):244–250.

48. Nguyen M, Ekstrom A, Li X, et al. HGT–Finder: A New Tool for Horizontal Gene Transfer Finding and Application to Aspergillus genomes. *Toxins*. 2015;7(10):4035–4053.

49. Fuchsman CA, Collins RE, Rocap G, et al. Effect of the environment on horizontal gene transfer between bacteria and archaea. *Peer J*. 2017;5:e3865.

50. Koonin EV. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000 Research*. 2016;5:F1000 Faculty Rev–1805.

51. Chen L, Yang D, Liu T, et al. Synonymous codon usage patterns in different parasitic platyhelminth mitochondrial genomes. *Genet Mol Res*. 2013;12(1):587–96.

52. He B, Dong H, Jiang C, et al. Analysis of codon usage patterns in Ginkgo biloba reveals codon usage tendency from A/U–ending to G/C–ending. *Scientific Reports*. 2016;6:35927.