Codon usage bias and peptide properties of Pseudomonas Balearica DSM 6083T

Abstract

Pseudomonas balearica DSM 6083T 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 6083T. 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 (pl, aromaticity, hydrophathy, and instability) are correlated (|r|>0.102, p-value<7 x 10^-15). %GC is correlated (|r|>0.122, p-value<6 x 10^-13) to peptide length, aromatic capacity, hydrophathy, and instability. Peptide length is correlated (|r|<0.057, p-value<0.0003) to pl, aromatic capacity, 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 x 10^-15) 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 x 10^-14), 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 6083T.

Introduction

Pseudomonas balearica is an environmental Gram-negative bacilliform bacterium with denitrifying capabilities and the ability to degrade several organic compounds; such as, naphthalene and thiosulfate; suggesting potential applications in bioremediation. Biochemically, P. balearica and Pseudomonas stutzeri exhibit several common phenotypical traits; 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 suggests that P. balearica should be considered distinct from P. stutzeri. Genotypitical differentiation and phylogenetic proximity of P. balearica from P. stutzeri was established through comparative genomic sequence analysis using BLAST calculation of average nucleotide identity and found 81.2% identity from the closest species, P. stutzeri ATCC 17588T. The genome of P. balearica DSM 6083T (CCUG 44595T, SP1042T) has been sequenced recently – 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 and has been implicated in gene expression, leading to the application of CUB in protein expression. A study 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; other factors, such as guanine-cytosine (GC)-content; and physical properties of the peptides (PPp), such as aromaticity and hydrophathy; have been suggested to influence CUB. PPp is an important aspect to study protein chemistry and elucidating potential functions. 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 6083T using its recently published sequence. 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 6083T. 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 6083T 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 as:

$$RSCU = \frac{OF_i}{\sum_{j=1}^{N} (OF_j)}$$

where $OF_i$ 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 $\sum_{j=1}^{N} (OF_j)$ is the overall expected codon distribution. RSCU value of 1 signifies no bias, while greater than or lesser than...
I signifies an increase or decrease in codon abundance respectively.\textsuperscript{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\textsuperscript{19} and can be calculated as the probability value using Chi-Square test with Bonferroni correction.\textsuperscript{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.\textsuperscript{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\textsuperscript{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.\textsuperscript{22} Hydropathy (GRAVY) refers to the overall hydrophobic/hydrophilic properties of a peptide.\textsuperscript{23} Isoelectric point (pI) is the pH where a peptide is electrical neutrality.\textsuperscript{24} Instability index refers to the stability of the peptide where high instability score suggests shorter half-life.\textsuperscript{25} All four methods are available in Biopython library.\textsuperscript{26}

**Statistical analysis**

Statistical analysis was performed using Pearson’s Chi-Square test with Bonferroni correction\textsuperscript{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\textsuperscript{7} 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.\textsuperscript{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.\textsuperscript{28,29} For example, regulatory features appear to be clustered differently to transfer RNAs.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{feature_map.png}
\caption{Feature map for the *P. balearica* DSM 6083\textsuperscript{3} 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).}
\end{figure}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Feature & Description & Count \\
\hline
CDS & Coding sequence & 4,050 \\
\hline
Gene & Gene & 4,126 \\
\hline
Miscellaneous feature & Misc. feature & 1 \\
\hline
Transfer RNA & T. RNA & 1 \\
\hline
Ribosomal RNA & Rib. RNA & 12 \\
\hline
Regulatory sequence & Regulatory & 5 \\
\hline
Repeat region & Repeat & 2 \\
\hline
ncRNA & Non-coding RNA & 3 \\
\hline
\end{tabular}
\caption{Summary of genomic features in *P. balearica* DSM 6083\textsuperscript{3}.}
\end{table}

However, all abundant codons (RSCU >1) are G and C ending codons (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.\textsuperscript{34,35} 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,\textsuperscript{36} which is consistent to the environment where *P. balearica* may be found.\textsuperscript{19} 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.\textsuperscript{18}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Codon & Amino Acid & Frequency \\
\hline
ATG & Methionine & 570 \\
\hline
TGG & Tryptophan & 329 \\
\hline

Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas Balearica DSM 6083\textsuperscript{3}. MOJ Proteomics Bioinform. 2019;8(2):27–36.

DOI: 10.15406/mojpb.2019.08.00263
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.

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

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 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 proteomes. 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. This suggests that about half of the proteins in *P. balearica* may have a short half-life.

Figure 3 Distributions of Peptides by Length.

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 pl, aromaticity index, hydropathy, and instability index on the CDSes. Our results show significant correlation between pl and aromaticity (r=0.161, p-value=6 x 10^-29, Figure 5A), pl and hydropathy (r=0.212, p-value=2 x 10^-42, Figure 5B), pl 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^-61, Figure 5D), aromaticity and instability (r=-0.104, p-value=3 x 10^-11, Figure 5E), and hydropathy and instability (r=-0.296, p-value=1 x 10^-45, Figure 5F). Scatterplots of pl (Figures 5A to 5E) is consistent with bimodal distribution of peptides with low fractions at pl close to 7.4.40

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

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^-23, 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^-23, Figure 6B). In addition, pl is not correlated to %GC (r=0.026, p-value=0.098, Figure 6A).

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^-22, Figure 7A), %GC1 (r=0.154, 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.41 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 pl, 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 pl 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^-8, Figure 8D). Relationship between peptide size and pl40 and peptide size and its stability44 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).

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=-0.261, p-value=6 x 10^-41, Figure 9A) and codon count variation (r=-0.525, p-value=8 x 10^-36, Figure 9B). These findings are consistent with Taenia solium 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 CCCV and AAV respectively, a substantial proportion of the CDSes (26.9%) show both significant CCCV and AAV (p-value threshold=1.2 x 10^-3 after Bonferroni correction). Statistically significant deviation of feature properties from the rest of the genome, also known as composition based46 or parametric46 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 HTG,48 which is the principle behind our CCCV and AAV methods. Hence, our result may suggest a substantial proportion of the exogenous genes in P. balearica as HTG has been considered prevalent in bacteria.48,49

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

Relationship between PPp to codon count variation (CCCV) and amino acid variation (AAV) were analyzed. Our results show significant correlations between CCCV to all PPp. There are significantly negative correlations between CCCV and pl (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 pl (r=-0.105, p-value=2 x 10^-41, 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 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, Loby et al.42 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 pl also show marginal but significant association with CUB.

By comparing the proportion of amino acids to various kingdoms,50 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 archaeabacteria (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.
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.

Citation: Maitra A, Ling MHT. Codon usage bias and peptide properties of Pseudomonas Balearica DSM 6083'. MOJ Proteomics Bioinform. 2019;8(2):27-36. 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.

Figure 8 Relationships between Peptide Length and Physical Properties of Peptides (n=4,050 CDSes). (A) Is the relationship between peptide length and pl. (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.

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.

Figure 10 Relationships between Usage Bias and Physical Properties of Peptides (n = 4,050 CDSes). (A) Shows the relationship between pl and amino acid variation. (B) Shows the relationship between hydrophathy 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 pl and codon count variation. (F) Shows the relationship between hydrophathy 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'. MOJ Proteomics Bioinform. 2019;8(2):27–36. 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


