ECRONICON OPEN ACCESS

EC CLINICAL AND MEDICAL CASE REPORTS
Research Article

Stop Codon Usage Varies on CDS Length, Nucleotide Compositions, and Peptide Instability in Six *Escherichia coli* Strains

Shaun CH Chua[1] and Maurice HT Ling[1,2]*

[1]*School of Applied Sciences, Temasek Polytechnic, Singapore*

[2]*HOHY PTE LTD, Singapore*

***Corresponding Author:** Maurice HT Ling, School of Applied Sciences, Temasek Polytechnic and HOHY PTE LTD, Singapore.

Received: January 02, 2021; **Published:** January 30, 2021

Abstract

Prokaryotic stop codon usage has been shown to be influenced by GC content and release factor abundance. However, it is unlikely that GC content and release factors are the only factors influencing stop codon usage as nucleotide compositions and peptide properties have also been shown to influence codon usage biasness; of which, stop codon usage is a specific instance of codon usage bias. Here, the stop codon usage frequencies, nucleotide compositions, and peptide properties in six strains of *E. coli* (MG1655, W3110, BL21, O25b:H4, O157:H7 and 58-3) were examined. Our results suggest that the stop codon usage frequencies of pathogenic strains O157:H7 and O25b:H4 are significantly different to other strains (Chi-Square ≥ 7.241, p-value ≤ 2.7E-02), suggesting different evolutionary paths between pathogenic and non-pathogenic strains of *E. coli*. The average lengths, nucleotide compositions, and peptide instability between the stop codons are significantly different in all cases (F ≥ 3.07, p-value ≤ 4.7E-02) except for average thymine composition in *E. coli* 58-3. This suggests a relationship between stop codon usage and nucleotide compositions, other than GC content and/or peptide properties.

Keywords: *E. coli; CDS Length; Nucleotide; Compositions; Peptide Instability*

Introduction

Codon usage bias is an important feature of evolution [1-4] and has been shown to be explained by tRNA abundance in *Drosophila* [5]. Stop codon usage bias is a specific instance of codon usage bias focusing on stop codons, with Sharp and Bulmer [6] performing the first stop codon usage analysis using 323 genes across three species. In DNA, the three common stop codons are TAA, TAG, TGA; and their usage may be related to GC content in prokaryotes [7] but not in eukaryotes [8]. Release factors abundance have also been shown to influence stop codon usages [9] as prokaryotes encodes for three release factors; namely, RF1, RF2 and RF3. RF1 recognizes UAA and UAG stop codons whereas RF2 recognizes UAA and UGA [10]. RF3 increases the dissociation of RF1 and RF2 from the ribosome after the release of the peptide [11].

It is unlikely that GC content and release factors are the only factors influencing stop codon usage as several studies show that nucleotide compositions and peptide properties, such as isoelectric point and aromaticity, have an impact on codon usage biasness in eukaryotes [12], prokaryotes [13] and viruses [14]. Hence, it is conceivable that nucleotide compositions and peptide properties may also influence stop codon usage as stop codon usage bias is a specific instance of codon usage bias. There are several studies [9,15,16] on nucleotide compositions and peptide properties on stop codon usages but no studies specifically on *Escherichia coli*.

In this study, we examine the stop codon usage frequencies, nucleotide compositions, and peptide properties in six strains of *E. coli*; namely, MG1655, W3110, BL21, O25b:H4, O157:H7 and 58-3. Our results suggest that the stop codon usage frequencies of pathogenic strains O157:H7 and O25b:H4 are significantly different to other strains (Chi-Square ≥ 7.241, p-value ≤ 2.7E-02). The average lengths and nucleotide compositions between the stop codons are significantly different in all cases (F ≥ 3.07, p-value ≤ 4.7E-02) except for average thymine composition in *E. coli* 58-3. Of the peptide properties, only instability between the stop codons is significantly different in all strains (F ≥ 15.4, p-value ≤ 2.1E-07).

Materials and Methods

Six strains of *Escherichia coli* were chosen for study: namely: (a) MG1655, Accession number NZ_CP032667.1 [17]; (b) W3110, Accession number AP009048.1 [18]; (c) BL21, Accession number CP060121.1; (d) O25b:H4, Accession number NZ_CP015085.1 [19]; (e) O157:H7, Accession number AE005174.2 [20] and (f) 58-3, Accession number CP050036.1. For each strain, complete set of nucleotide coding sequences and its corresponding peptide sequences were obtained. Stop codon usage frequencies between each strain were analyzed in pairwise manner (A vs B and B vs A) using Chi-Square test where the frequency of one strain was used to determine the expected frequency of another strain and the lower statistic with higher probability are taken. Nucleotide compositions (CDS length, GC content, adenine content, thymine content, guanine content, and cytosine content) and peptide properties [molecular weight, isoelectric point [21,22], aromaticity [23], instability [24], hydropathy [25]] for each coding sequence were determined using Biopython [26] via SeqProperties [27]. For translation of nucleotide sequence to peptide sequence for peptide properties determination, translation table 11 (bacterial, archaeal and plant plastid code) was used. The null hypothesis of no difference in averages between the three stop codons were tested using 1-way ANOVA for each of the nucleotide compositions and peptide properties. Comparisons between any two stop codons were performed using 2-samples t-test assuming unequal variance. For all statistical tests, a probability value of less than 5% was significant; hence, rejecting null hypothesis and accepting alternate hypothesis.

Results and Discussion

Codon frequencies, nucleotide compositions, and peptide properties across six *E. coli* strains were examined in this study.

Stop codon frequency: To ascertain the presence of stop codon usage biasness, the stop codon usage [28] frequencies of each strain was evaluated under the null hypothesis of uniform frequency. Our results suggest that stop codon usages are not uniform (Table 1; Chi-Square ≥ 1794.1, p-value < 1.0E-240), which is consistent with that of other prokaryotic species [15], with more than 59% of the peptides using TAA and less than 8.5% of the peptides using TAG as stop codons.

Strain	Accession Number	Stop Codon Count			Chi-Square Statistic	p-value
		TAA	TAG	TGA		
MG1655	NZ_CP032667.1	2750	315	1247	2100.4	<1E-240
W3110	AP009048.1	2756	330	1266	2063.8	<1E-240
BL21	CP060121.1	2711	303	1210	2100.9	<1E-240
O25b:H4	NZ_CP015085.1	3109	368	1560	2250.0	<1E-240
O157:H7	AE005174.2	2716	378	1481	1794.1	<1E-240
58-3	CP050036.1	2720	347	1242	1999.7	<1E-240

Table 1: *Stop codon usage frequencies against uniform distribution.*

However, stop codon frequencies are not consistent across the six *E. coli* strains (Figure 1). Our results suggest that *E. coli* O157:H7 and O25b:H4 strains have significantly different stop codon usage frequencies compared to the other four strains (Chi-Square ≥ 7.241, p-value < 2.7E-02) and between each other (Chi-Square = 12.744, p-value = 1.7E-03). This is supported by a study by Lukjancenko., *et al*. [29] showing that the pan-genome of BL21 sub-strains cluster with that of K-12 derived sub-strains, which includes MG1655 and W3110; but BL21/K-12 sub-strains cluster separately from that of O157:H7 sub-strains, which is also supported by Bachmann., *et al*. [30] showing that *E. coli* O157:H7 sub-strains are phylogenetically closer to *Shigella dysenteriae* than to *E. coli* MG1655 and W3110. In addition, another study also suggests that O25b:H4 belongs to a different phylogenetic group compared to MG1655 [31]. As *E. coli* O157:H7 [32,33] and O25b:H4 [28,34] are both pathogenic strains, this may suggests different evolutionary paths between pathogenic and non-pathogenic strains of *E. coli*.

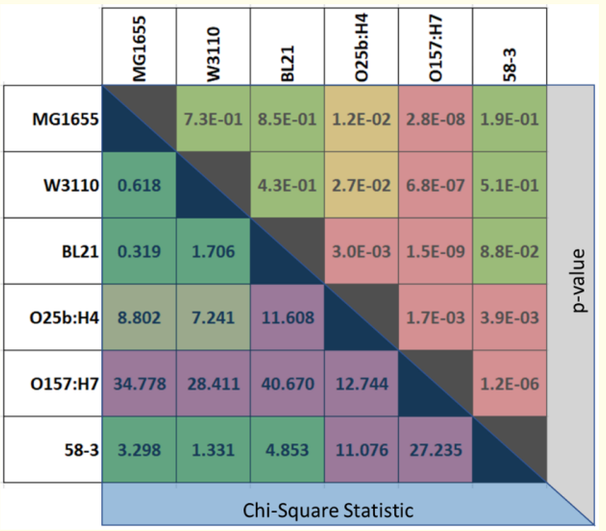

Figure 1: *Statistical comparison of stop codon usage frequencies between six strains of Escherichia coli. Values in the lower triangle are the Chi-Square statistics while values in the upper triangle are the corresponding p-values. Not significance is highlighted in green.*

CDS lengths and nucleotide compositions: Our results show unequal average CDS length between the stop codons (Table 2; F ≥ 3.07, p-value ≤ 4.7E-02) across the six strains with long CDSes preferring TAA, followed by TGA and short CDSes preferring TAG. This trend is consistent in all six strains. Although this may be surprising, this is consistent to that reported by Duret and Mouchiroud [35] showing gene length affects stop codon usage in eukaryotes.

Strain	Average Length	Average %GC	Average %A	Average %T	Average %G	Average %C
MG1655	F = 4.97 (7.0E-03)	F = 6.47 (1.6E-03)	F = 8.40 (2.3E-04)	F = 3.75 (2.4E-02)	F = 6.25 (1.9E-03)	F = 6.26 (1.9E-03)
W3110	F = 5.99 (2.5E-03)	F = 10.4 (3.1E-05)	F = 9.88 (5.3E-05)	F = 6.56 (1.4E-02)	F = 6.85 (1.1E-03)	F = 11.3 (1.2E-05)
BL21	F = 7.97 (3.5E-04)	F = 13.9 (9.6E-07)	F = 14.5 (5.1E-07)	F = 6.54 (1.5E-03)	F = 11.7 (8.7E-06)	F = 10.6 (2.5E-05)
O25b:H4	F = 5.74 (3.2E-3)	F = 39.3 (<1E-240)	F = 27.2 (1.8E-12)	F = 12.3 (4.6E-06)	F = 32.0 (1.5E-14)	F = 23.9 (4.6E-11)
O157:H7	F = 11.0 (1.7E-05)	F = 34.8 (1.0E-15)	F = 15.4 (2.1E-7)	F = 22.2 (2.5E-10)	F = 32.4 (1.1E-14)	F = 25.3 (1.2E-11)
58-3	F = 3.07 (4.7E-02)	F = 7.40 (6.2E-04)	F = 11.3 (1.2E-05)	F = 0.254 (7.8E-01)	F = 10.6 (2.6E-05)	F = 4.37 (1.3E-02)

Table 2: *Statistical comparison of average length and nucleotide compositions by stop codon usage. P-values are in brackets and non-significance is highlighted in green.*

In terms of nucleotide compositions, our results show unequal average %GC between the stop codons (Table 2; F ≥ 6.47, p-value ≤ 1.6E-03) with high GC content preferring TGA, followed by TAA and low GC content preferring TAA. This is consistent with studies suggesting that stop codon usage is correlated to GC content [7,16]. 23 of the 24 nucleotide compositions across all six strains of *E. coli* are significant (Table 2, F ≥ 3.75, p-value ≤ 2.4E-02) between stop codons except for average thymine composition in *E. coli* 58-3 (Table 2, F = 0.254, p-value = 0.78). Using pairwise t-test as post-hoc test, our results show consistent significant differences (Table 3, p-value ≤ 2.4E-02) between the average nucleotide compositions of the highest and lowest averages; except for average thymine composition in *E. coli* 58-3 (p-value ≥ 5.3E-01), which is consistent with our findings from 1-way ANOVA in table 2. Examining the trend for stop codon usage by nucleotide composition (Table 3), *E. coli* O157:H7 and O25b:H4 are consistent with each other. Interestingly, the trend of average nucleotide composition by stop codon is not consistent between *E. coli* MG1655 and W3110, which may be accounted for by the genomic differences [18] amounting to phenotypic differences [36] between these two K-12 derived strains. Taken together, these results suggest substantial stop codon usages between nucleotide compositions, which may suggest evolutionary pressures.

Strain	Composition	Average Composition			Pairwise t-test p-values		
		TAA	TAG	TGA	TAA vs TAG	TAA vs TGA	TAG vs TGA
MG1655	%A	24.875%	24.701%	24.299%	5.1E-01	4.2E-05	1.5E-01
	%T	24.430%	25.030%	24.469%	1.4E-02	7.5E-01	2.9E-02
	%G	26.754%	26.861%	27.144%	6.3E-01	3.9E-04	2.3E-01
	%C	23.942%	23.409%	24.088%	5.8E-03	1.7E-01	9.3E-04
W3110	%A	24.820%	24.867%	24.228%	8.6E-01	1.2E-05	2.4E-02
	%T	24.247%	24.973%	24.318%	2.0E-02	5.5E-02	7.8E-03
	%G	26.882%	26.812%	27.256%	7.6E-02	3.8E-02	6.6E-02
	%C	24.052%	23.348%	24.199%	1.6E-02	1.4E-01	1.7E-05
BL21	%A	24.886%	25.007%	24.154%	6.5E-02	2.3E-07	2.5E-03
	%T	24.490%	25.302%	24.539%	1.4E-03	7.1E-01	4.4E-03
	%G	26.705%	26.501%	27.208%	3.9E-01	7.8E-06	4.5E-03
	%C	23.918%	23.190%	24.099%	5.3E-04	9.8E-02	4.7E-05
O25b:H4	%A	25.018%	25.219%	24.160%	4.3E-01	8.5E-13	7.6E-05
	%T	24.315%	25.097%	24.062%	4.5E-04	2.3E-02	9.2E-06
	%G	26.802%	26.606%	27.534%	3.5E-01	4.5E-14	2.7E-05
	%C	23.866%	23.077%	24.245%	4.2E-05	4.1E-05	6.8E-09
O157:H7	%A	25.508%	26.370%	24.992%	2.7E-03	3.5E-04	4.1E-06
	%T	24.538%	25.741%	24.255%	1.4E-06	2.3E-02	1.2E-08
	%G	26.530%	25.770%	27.271%	1.4E-03	5.7E-10	2.1E-09
	%C	23.416%	22.116%	23.468%	1.7E-08	6.4E-01	1.6E-08
58-3	%A	24.907%	24.755%	24.238%	5.2E-01	2.1E-06	4.2E-02
	%T	24.429%	24.576%	24.432%	5.3E-01	9.9E-01	5.6E-01
	%G	26.757%	27.112%	27.242%	8.6E-02	7.0E-06	5.5E-01
	%C	23.907%	23.556%	24.089%	5.2E-02	8.6E-02	5.8E-03

Table 3: *Post-Hoc test of average nucleotide composition for each stop codon. Listing of mean nucleotide composition for each stop codon; for example, the mean adenosine percentage of the CDS that uses TAA as stop codon in E. coli MG1655 is 24.875%. The highest and lowest percentage of each nucleotide composition is shaded in yellow and green, respectively. Significant p-values of less than 0.05 are shaded in red.*

Peptide properties: Molecular weight of peptides is highly correlated to its CDS length (R-square ≥ 0.9497, F ≥ 82787, p-value ≤ 1E-240). As our earlier findings show potential preferential stop codon usage by CDS length, we expect preferential stop codon usage by molecular weights of peptide (Table 4), which is supported by Duret and Mouchiroud [35] by proxy due to high correlation between CDS lengths and molecular weight of peptides. Our results indicate that four of the six strains (W3110, BL21, O25b:H4, and 58-3) are significant in average isoelectric point [21,22] by stop codons (Table 4; F ≥ 3.16, p-value ≤ 4.3E-02). An early work by Boucherie., *et al*. [37] indicated that basic proteins in yeast do not correspond to genes with high codon usage index; thus, suggest a possible link between codon usage and stop codon usage by extension to isoelectric point. Interesting, all strains are significant in average instability [24] by stop codons (Table 4; F ≥ 15.4, p-value ≤ 2.1E-07), which may suggest selective pressures between gene level and peptide level. However, both average aromaticity [23] and average hydropathy [25] are not significant in any strains, suggesting the possibility of a non-straightforward selective pressure between gene level and peptide level.

Strain	Average Molecular Weight	Average Isoelectric Point	Average Aromaticity	Average Instability	Average Hydropathy
MG1655	F = 5.77 (3.2E-03)	F = 2.76 (6.3E-02)	F = 0.451 (6.4E-01)	F = 15.8 (1.5E-07)	F = 0.139 (8.7E-01)
W3110	F = 7.25 (7.2E-04)	F = 3.16 (4.3E-02)	F = 1.43 (2.4E-01)	F = 18.3 (1.2E-08)	F = 0.494 (6.1E-01)
BL21	F = 9.56 (7.2E-05)	F = 6.74 (1.2E-03)	F = 1.31 (2.7E-01)	F = 15.4 (2.1E-07)	F = 1.32 (2.7E-01)
O25b:H4	F = 6.83 (1.1E-03)	F = 7.12 (8.2E-04)	F = 1.72 (1.8E-01)	F = 17.6 (2.5E-08)	F = 0.223 (8.0E-01)
O157:H7	F = 10.0 (4.5E-05)	F = 1.97 (1.4E-01)	F = 2.32 (9.9E-02)	F = 28.6 (4.5E-13)	F = 2.51 (8.1E-02)
58-3	F = 4.03 (1.8E-02)	F = 9.50 (7.6E-05)	F = 0.178 (8.4E-01)	F = 22.1 (2.7E-10)	F = 2.28 (1.0E-01)

Table 4: *Statistical comparison of peptide properties by stop codon usage. P-values are in brackets and non-significance is highlighted in green.*

Conclusion

In this study, the stop codon usage frequencies, nucleotide compositions, and peptide properties in six strains of *E. coli*. Our results suggest that the stop codon usage frequencies of pathogenic strains O157:H7 and O25b:H4 are significantly different to other strains (Chi-Square ≥ 7.241, p-value ≤ 2.7E-02). The average nucleotide compositions and peptide instability are significantly different (F ≥ 15.4, p-value ≤ 2.1E-07) between each stop codons in most cases; suggesting a relationship between stop codon usage and nucleotide compositions, other than GC content, and/or peptide properties.

Data Availability

Data files for this study is available as a zipped file at https://bit.ly/ECO_Stop.

Conflict of Interest

The authors declare no conflict of interest.

Appendix

		58-3	BL21	MG1655	O25b:H4	O157:H7	W3110
Number of CDS	Count	4385	4312	4353	5061	5349	4337
CDS Length	Mean	943.1	930.0	939.7	924.9	905.2	950.7
	Sigma	629.13	631.62	634.90	663.48	692.99	640.65
Number of Adenine	Mean	227.8	224.2	226.6	224.6	222.1	229.2
	Sigma	152.58	153.05	154.32	162.43	169.86	156.52
Number of Thymine	Mean	226.7	224.3	226.2	221.7	216.8	228.5
	Sigma	146.15	147.63	147.65	153.24	156.68	148.69
Number of Guanine	Mean	257.6	253.8	256.6	252.8	247.1	259.8
	Sigma	179.07	179.04	180.19	188.22	199.07	181.50
Number of Cytosine	Mean	231.0	227.7	230.3	225.8	218.9	233.2
	Sigma	164.15	146.10	165.21	172.22	181.66	156.52
Peptide Molecular Weight (Daltons)	Mean	34130.2	33694.9	34228.3	33689.9	31098.9	34718.4
	Sigma	23128.45	23338.07	23278.89	24305.39	22655.19	23084.62
Isoelectric Point	Mean	6.93	6.96	6.92	6.97	6.98	6.93
	Sigma	1.874	1.890	1.860	1.885	1.900	1.859
Aromaticity Index	Mean	0.0820	0.0827	0.0825	0.0819	0.0825	0.0821
	Sigma	0.03296	0.03345	0.03216	0.03244	0.03226	0.03152
Instability Index	Mean	37.7	37.7	37.5	37.8	38.4	37.7
	Sigma	11.43	11.55	11.40	11.61	11.95	10.84
GRAVY (Hydropathy)	Mean	-0.0643	-0.0538	-0.0569	-0.0789	-0.1070	-0.0711
	Sigma	0.46087	0.46493	0.46832	0.45481	0.44272	0.44105

Bibliography

1. Song H., *et al.* "Comprehensive Analysis of Codon Usage Bias in Seven Epichloë Species and Their Peramine-Coding Genes". *Frontiers in Microbiology* 8 (2017): 1419.

2. Sharp PM., *et al.* "Forces that Influence the Evolution of Codon Bias". *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1544 (2010): 1203-1212.

3. Keng BM., *et al.* "Codon Usage Bias is Evolutionarily Conserved". *Asia-Pacific Journal of Life Sciences* 7.3 (2014): 233-242.

4. Trotta E. "Selective forces and mutational biases drive stop codon usage in the human genome: a comparison with sense codon usage". *BMC Genomics* 17 (2016): 366.

5. Moriyama EN and Powell JR. "Codon Usage Bias and tRNA Abundance in Drosophila". *Journal of Molecular Evolution* 45.5 (1997): 514-523.

6. Sharp PM and Bulmer M. "Selective Differences Among Translation Termination Codons". *Gene* 63.1 (1988): 141-145.

7. Povolotskaya IS., *et al.* "Stop Codons in Bacteria are Not Selectively Equivalent". *Biology Direct* 7 (2012): 30.

8. Sun J., *et al.* "Relationships Among Stop Codon Usage Bias, Its Context, Isochores, and Gene Expression Level in Various Eukaryotes". *Journal of Molecular Evolution* 61.4 (2005): 437-444.

9. Korkmaz G., *et al.* "Comprehensive Analysis of Stop Codon Usage in Bacteria and Its Correlation with Release Factor Abundance". *Journal of Biological Chemistry* 289.44 (2014): 30334-30342.

10. Scolnick E., *et al.* "Release Factors Differing in Specificity for Terminator Codons". *Proceedings of the National Academy of Sciences of the United States of America* 61.2 (1968): 768-774.

11. Freistroffer DV., *et al.* "Release Factor RF3 in E. coli Accelerates the Dissociation of Release Factors RF1 and RF2 from the Ribosome in a GTP-Dependent Manner". *The EMBO Journal* 16.13 (1997): 4126-4133.

12. Yang X., *et al.* "Codon Usage Bias and Determining Forces in Taenia solium Genome". *The Korean Journal of Parasitology* 53.6 (2015): 689-697.

13. Maitra A and Ling MH. "Codon Usage Bias and Peptide Properties of Pseudomonas balearica DSM 6083T". *MOJ Proteomics and Bioinformatics* 8.2 (2019): 27-39.

14. Sheikh A., *et al.* "Analysis of Preferred Codon Usage in the Coronavirus N Genes and Their Implications for Genome Evolution and Vaccine Design". *Journal of Virological Methods* 277 (2020): 113806.

15. Wei Y., *et al.* "Coevolution between Stop Codon Usage and Release Factors in Bacterial Species". *Molecular Biology and Evolution* 33.9 (2016): 2357-2367.

16. Belinky F., *et al.* "Purifying and Positive Selection in the Evolution of Stop Codons". *Scientific Reports* 8.1 (2018): 9260.

17. Earl JP., *et al.* "Species-Level Bacterial Community Profiling of the Healthy Sinonasal Microbiome Using Pacific Biosciences Sequencing of Full-Length 16S rRNA Genes". *Microbiome* 6.1 (2018): 190.

18. Hayashi K., *et al.* "Highly Accurate Genome Sequences of Escherichia coli K-12 Strains MG1655 and W3110". *Molecular Systems Biology* 2 (2006): 2006.0007.

19. Alyamani E., *et al.* "Genome Sequence and Comparative Pathogenic Determinants of Multidrug Resistant Uropathogenic Escherichia coli O25b:H4, A Clinical Isolate from Saudi Arabia". *Journal of Pure and Applied Microbiology* 10.4 (2016): 2475-2484.

20. Perna NT., *et al.* "Genome Sequence of Enterohaemorrhagic Escherichia coli O157:H7". *Nature* 409.6819 (2001): 529-533.

21. Bjellqvist B., *et al.* "The Focusing Positions of Polypeptides in Immobilized pH Gradients can be Predicted from Their Amino Acid Sequences". *Electrophoresis* 14.10 (1993): 1023-1031.

22. Bjellqvist B., *et al.* "Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions". *Electrophoresis* 15.1 (1994): 529-539.

23. Lobry JR and Gautier C. "Hydrophobicity, Expressivity and Aromaticity are the Major Trends of Amino-Acid Usage in 999 Escherichia coli Chromosome-Encoded Genes". *Nucleic Acids Research* 22.15 (1994): 3174-3180.

24. Guruprasad K., *et al.* "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 Engineering Design and Selection* 4.2 (1990): 155-161.

25. Kyte J and Doolittle RF. "A Simple Method for Displaying the Hydropathic Character of a Protein". *Journal of Molecular Biology* 157.1 (1982): 105-132.

26. Cock PJA., *et al.* "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics". *Bioinformatics* 25.11 (2009): 1422-1423.

27. Ling MHT. "Seq Properties: A Python Command-Line Tool for Basic Sequence Analysis". *Acta Scientific Microbiology* 3.6 (2020): 103-106.

28. Flament-Simon S-C., *et al.* "Whole Genome Sequencing and Characteristics of mcr-1-Harboring Plasmids of Porcine Escherichia coli Isolates Belonging to the High-Risk Clone O25b:H4-ST131 Clade B". *Frontiers in Microbiology* 11 (2020): 387.

29. Lukjancenko O., *et al.* "Comparison of 61 Sequenced Escherichia coli Genomes". *Microbial Ecology* 60.4 (2010): 708-720.

30. Bachmann NL., *et al.* "Genomic Comparison of Translocating and Non-Translocating Escherichia coli". *PLOS ONE* 10.8 (2015): e0137131.

31. Vimont S., *et al.* "The CTX-M-15-Producing Escherichia coli Clone O25b:H4-ST131 has High Intestine Colonization and Urinary Tract Infection Abilities". *PLOS ONE* 7.9 (2012): e46547.

32. Piérard D., *et al.* "O157:H7 and O104:H4 Vero/Shiga Toxin-Producing Escherichia coli Outbreaks: Respective Role of Cattle and Humans". *Veterinary Research* 43 (2012): 13.

33. Byrne L., *et al.* "Association between Shiga Toxin-Producing Escherichia coli O157:H7 stx Gene Subtype and Disease Severity, England, 2009-2019". *Emerging Infectious Diseases* 26.10 (2020): 2394-2400.

34. Forde BM., *et al.* "The Complete Genome Sequence of Escherichia coli EC958: A High Quality Reference Sequence for the Globally Disseminated Multidrug Resistant E. coli O25b:H4-ST131 Clone". *PLOS ONE* 9.8 (2014): e104400.

35. Duret L and Mouchiroud D. "Expression Pattern and, Surprisingly, Gene Length Shape Codon Usage in Caenorhabditis, Drosophila, and Arabidopsis". *Proceedings of the National Academy of Sciences of the United States of America* 96.8 (1999): 4482-4487.

36. Monk JM., *et al.* "Multi-Omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes". *Cell Systems* 3.3 (2016): 238-251.

37. Boucherie H., *et al.* "Two-Dimensional Protein Map of Saccharomyces cerevisiae: Construction of a Gene-Protein Index". *Yeast* 11.4 (1995): 601-613.

Volume 4 Issue 2 February 2021