THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Codon usage behavior in *Helicobacter pylori* 35A, a gastrointestinal pathogen, from a bioinformatics perspective

Nishika Jaishee[1], Bhupender Singh[2] & Ayan Roy[2]*

[1]*Plant Biochemistry Laboratory, Department of Botany, University of North Bengal, Siliguri, India*
[2]*Department of Bioinformatics, Lovely Professional University, Punjab, India*
**E.mail: ayanroy.24373@lpu.co.in Tel:9163139658*

Abstract

Helicobacter pylori, residing in human gut, has been associated with several threatening diseases that include gastric and duodenal ulcers and gastric adenocarcinoma. Helicobacter pylori is an intimidating pathogen that resides in human gastrointestinal tract and causes gastritis, peptic ulcers and gastric adenocarcinoma on acquiring pathogenic phenotype. Present research endeavor has been targeted to deciphering the attributes of codon usage in Helicobacter pylori 35A. Factors like genomic compositional constraint, selection pressure for translational exactitude and gene length were imperative indices in influencing the mode of codon usage. AT compositional constraint was noted to be the most decisive factor. Present research approach confers information that might be beneficial for biological domains associated with pathogenomics of Helicobacter pylori and devise mutational experiments to combat associated infection.

Introduction

Helicobacter pylori, residing in human gut, has been associated with several threatening diseases that include gastric and duodenal ulcers [1] and gastric adenocarcinoma [2]. *Helicobacter pylori* is an intimidating pathogen that resides in human gastrointestinal tract and causes gastritis, peptic ulcers and gastric adenocarcinoma on acquiring pathogenic phenotype [3]. *H. pylori* is an alarming pathogen that resides in human gastrointestinal tract and has been reported to associated with severe intestinal disorders that include gastritis, peptic ulcers and gastric adenocarcinoma [2,3]. Identification of potential drug candidates has been achieved in various strains of *H. pylori* [4]. However, *H. pylori* 35A strain, a pathogenic resident of human gut, still demands to be explored and characterized from the drug discovery aspect. We aim an extensive

ISSN:0971-1260
Vol-22-Issue-37-December -2019

genomic/proteomic characterization of the *H. pylori* 35A strain for apt identification of factors influencing the codon usage mode.

The genus *Helicobacter* dominantly resides in the stomach and crucially influences the entire microbial composition of the gastric flora. It has been reported that when *Helicobacter pylori* resides as a commensal there is rich diversity of various microbial genera like *Streptococcus*, *Prevotella*, *Veillonella* and *Rothia* [5]. However, on acquiring pathogenic phenotype, *H. pylori* significantly administers the reduction of other microbial members. Large intestine harbors several bacterial members that are crucially associated with homeostasis.

Disturbances in gut microbial balance have also been found to be associated with obesity and type 2 diabetes [6]. *Helicobacter pylori* has been a notorious pathogen associated with gastritis, duodenal ulcers and gastric cancer and has generated a lot of concern [3,4]. Exact mechanisms underlying dysbiosis and gastrointestinal disorders still remain somewhat obscure. Various hypotheses have been put forward pertaining to the complex mechanisms. Some researchers have proposed the 'pathogen hypothesis' in which they have tried to correlate gastrointestinal diseases with infective potential of established and recognized pathogens.

Present research work aims to decipher the complexities of codon usage pattern in *H. pylori* 35A and underlying factors contributing to the variations. The work promises to be handy for research sector associated with pathogenomics of *H. pylori*.

Materials and Methods

Genomic data:

Genome sequence of *H. pylori* 35A was downloaded from NCBI database. Coding sequences with proper initiation and termination codons and were considered in our dataset to avoid stochastic variations [7].

Computation of codon based parameters:

Indices like GC3, GC and effective number of codons (Nc) were computed using CodonW software [8]. Relative synonymous codon usage was calculated using CodonW.

Principal component analysis of codon usage and associated statistical analysis:

Correspondence analysis was executed employing CodonW software package. Statistical analysis was performed SPSS software suite.

Computation of tRNA Adaptation Index (tAI):

tRNA Adaptation Index (tAI) is an efficient indicator of tRNA profile that represents magnitude

ISSN:0971-1260
Vol-22-Issue-37-December -2019

of co-adaptation between particular codon and corresponding tRNA pool [9]. Thus, tAI is considered to be an effective parameter for assessing codon bias as it directly correlates with protein abundance. tAI was calculated using R software.

Results and Discussion

Fashion of codon usage in Helicobacter pylori:

AT rich genomic composition of *Helicobacter pylori* was prominent from the investigation of overall codon frequency. The average genome size of the *H. pylori* strains was found to be around 1.59±0.04 millions of base pairs (Mbp) and the average AT composition was observed to be around 61.06±0.25%. It was distinct from overall codon usage pattern of *Helicobacter pylori* 35A (Table 1) that 14 amino acids preferred to be encoded by AT rich codons. Moreover, out of 28 codons with high relative synonymous codon usage (> 1.00), 16 were e AT rich. Thus, there was a marked impact of AT compositional constraint on the overall codon usage pattern of *H. pylori*. Pronounced effect of AT compositional bias has already been inferred in *H. pylori* [10] and our results were in complete agreement with the proposed fact.

Codon usage based roseplot of *Helicobacter pylori* 35A (Figure 1) suggested that the AT rich codons like AAA, GAA, TTA, ATT, AAT and GAT were used in higher frequencies (reflected by sharp peaks). Roseplot analysis revealed that *H. pylori* 35A were inclined for AT richness.

Table 1. Codon frequency of *H. pylori* 35A

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	**UUU**	20777	1.59	**Ser**	**UCU**	7617	1.40
	UUC	5393	0.41		UCC	2611	0.48
Leu	**UUA**	21149	2.35		UCA	3202	0.59
	UUG	14265	1.58		UCG	1812	0.33
	CUU	8109	0.90		AGU	4669	0.86
	CUC	4625	0.51		AGC	12773	2.34
	CUA	3969	0.44	Cys	UGU	1854	0.68
	CUG	1961	0.22		UGC	3619	1.32
Tyr	**UAU**	12328	1.41	Pro	CCU	7667	1.94
	UAC	5120	0.59		CCC	4182	1.06
His	**CAU**	6992	1.36		CCA	2427	0.61
	CAC	3278	0.64		CCG	1544	0.39
Gln	**CAA**	14701	1.70	**Arg**	CGU	2315	0.82

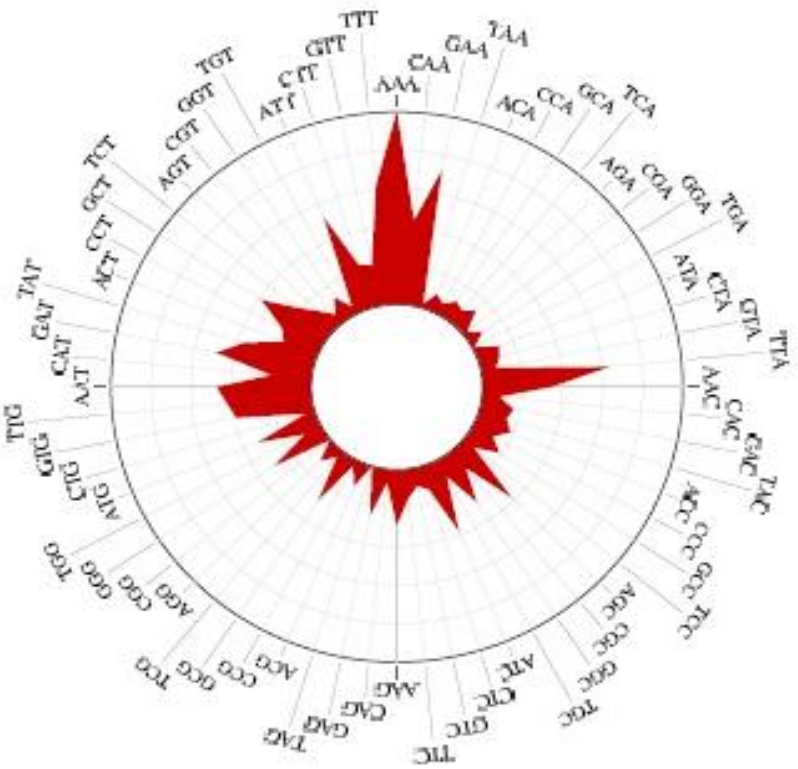
ISSN:0971-1260
Vol-22-Issue-37-December -2019

	CAG	2639	0.30		CGC	4129	1.46
Ile	**AUU**	17471	1.51		CGA	1262	0.45
	AUC	12817	1.10		CGG	542	0.19
	AUA	4530	0.39		**AGA**	4651	1.65
Asn	**AAU**	15803	1.16		**AGG**	4063	1.44
	AAC	11511	0.84	**Thr**	**ACU**	6384	1.22
Lys	**AAA**	33001	1.54		ACC	6493	1.24
	AAG	9858	0.46		ACA	3469	0.66
Val	**GUU**	7410	1.10		ACG	4567	0.87
	GUC	3556	0.53	Ala	GCU	12949	1.58
	GUA	2846	0.42		GCC	6476	0.79
	GUG	13146	1.95		GCA	3929	0.48
Asp	**GAU**	16419	1.48		GCG	9450	1.15
	GAC	5824	0.52	Gly	GGU	4700	0.67
Glu	**GAA**	23752	1.47		GGC	9808	1.40
	GAG	8511	0.53		GGA	3165	0.45
					GGG	10432	1.48

AT rich codons are represented in bold

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Figure 1: Roseplot for *H. pylori* 35A displaying relative frequency of the codons

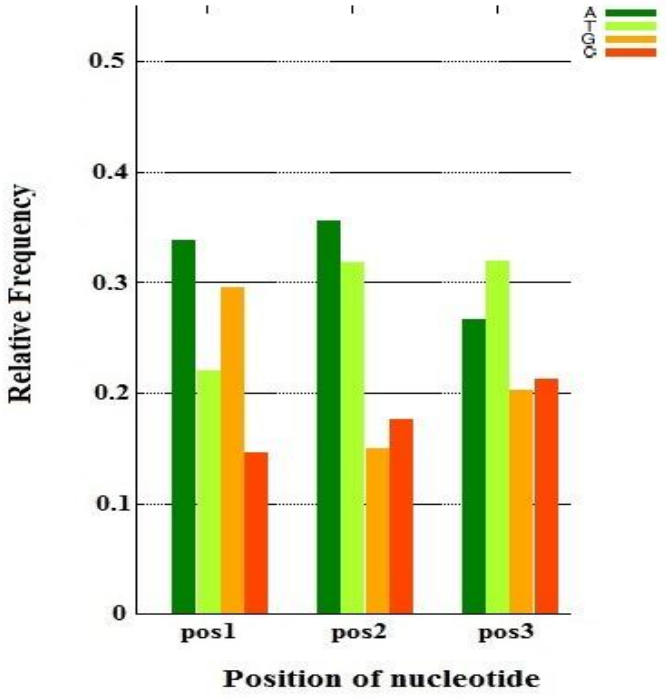


ISSN:0971-1260
Vol-22-Issue-37-December -2019

Barplot analysis showing AT richness:

It was evident from barplot analysis (Figure 2) that the wobble nucleotide was AT rich with a distinct preference for Thymine, in *H. pylori* 35A. Thus, it could be inferred the *H. pylori* 35A was under a strong influence of AT compositional constraint.

Figure 2. Barplot for *Helicobacter pylori* 35A displaying frequency of nucleotides at different codon positions



ISSN:0971-1260
Vol-22-Issue-37-December -2019

Factors underlying codon usage:

Principal component analysis revealed that Axes 1 and 2 produced major codon usage variations observed in the 59-dimensional hyperspace (indicated by high values of inertia). Substantial correlations of Axis 1 of RSCU data with genomic (AT/GC) composition, in all concerned strains of *H. pylori*, further signified the marked influence of compositional bias (Table 2).

Table 2. Correlation matrix of indices for *H. pylori* 35A

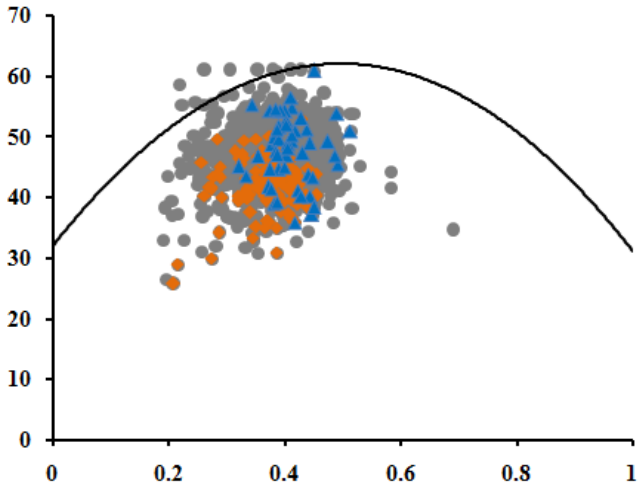
Organism		Axis1	Axis2	GC3	Nc	Length	GC	tAI
Helicobacter pylori 35A	**Axis1**	1.00	0.15**	0.48**	-0.15**	0.23**	0.22**	0.29*
	Axis2		1.00	-0.05	0.16**	-0.07**	-0.09**	0.01
	GC3			1.00	0.05*	0.25**	0.73**	0.40**
	Nc				1.00	0.06*	0.06*	-0.08
	Length					1.00	0.23**	0.27**
	GC						1.00	0.27**
	tAI							1.00

**: significant at 1%; *: significant at 5%; Length: Gene Length; Axis1: Axis1 of RSCU; Axis2– Axis2 of RSCU.

GC3 versus Nc (effective number of codons) plots have been suggested to be useful in proper investigation of codon usage data [7]. [7] proposed that when codon usage of an organism is entirely influenced by composition bias, all coding sequences tend to lie over the Nc plot curve. In case of *H. pylori* 35A (Figure 3) significant share of genes were found to lie over the Nc plot curve. Our finding accentuated the undeniable effect of AT composition on *H. pylori* 35A. Interestingly, the ribosomal proteins and the genes with high expression (PHX) were found to lie below the continuous Nc plot curve which suggested the influence of other factors, apart from compositional bias.

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Figure 3. GC3 vs Nc plots for *Helicobacter pylori* 35A. x-axis-GC3; y-axis-Nc; gray colored circles– overall genes; orange colored diamonds– PHX genes; blue colored triangles– ribosomal proteins



In order to validate such an observation, we correlated tAI with Nc (Table 2). Insignificant correlation emphasized the lack of translational exactitude on the genes with high expression. Similar results have been reported in *H. pylori* where low tRNA count (only a small set of 36 tRNA genes) has been attributed to the absence of efficacy for translation [11].

Gene length was noted to correlate significantly with Axis 1 and thus, could be attributed as a contributory factor for the variations observed in codon usage.

Thus, AT compositional bias was found to dominate codon usage heterogeneity of *H. pylori* 35A. Level of gene expression and gene length played significant role apart from mutational bias.

ISSN:0971-1260
Vol-22-Issue-37-December -2019

References

1. K.M. Fock, D.Y. Graham and P. Malfertheiner, “Helicobacter pylori research: historical insights and future directions,” Nat Rev Gastroenterol Hepatol, vol. ED- 10, pp. 495-500, Aug. 2013.
2. D.Y. Graham, “Helicobacter pylori infection is the primary cause of gastric cancer,” Journal of gastroenterology, vol. ED- 35, pp. 90-97, Dec. 2000.
3. N.R. Neelapu and T. Pavani, “Identification of novel drug targets in HpB38, HpP12, HpG27, Hpshi470, HpSJM180 strains of Helicobacter pylori : an in silico approach for therapeutic intervention,” Curr Drug Targets, vol. ED- 14, pp. 601-611, May 2013.
4. D. Nammi, R.C. Srimath-Tirumala-Peddinti, and N.R. Neelapu, “Identification of Drug Targets in Helicobacter pylori by in silico Analysis: Possible Therapeutic Implications for Gastric cancer,” Curr Cancer Drug Targets vol. ED- 16, pp. 79-98, Jan. 2016.
5. G.E. Andersson and P.M. Sharp (1996), “Codon usage in the Mycobacterium tuberculosis complex,” Microbiology, vol. ED- 142, pp. 915-925, Apr. 1996.
6. R.E. Ley, “Obesity and the human microbiome,” Curr Opin Gastroenterol, vol. ED- 26, pp. 5-11, Jan. 2010.
7. F. Wright, “The 'effective number of codons' used in a gene,” Gene, vol. ED- 87, pp. 23-29, Mar. 1990.
8. J.F. Peden, “Analysis of codon usage,” University of Nottingham, UK. PhD, Aug. 1999.
9. T. Tuller, I. Veksler-Lublinsky, N. Gazit, M. Kupiec, E. Ruppin and M. Ziv-Ukelson, “Composite effects of gene determinants on the translation speed and density of ribosomes,” Genome Biol, vol. ED- 12, pp. 110, Nov. 2011.
10. B. Lafay, J.C. Atherton and P.M. Sharp, “Absence of translationally selected synonymous codon usage bias in Helicobacter pylori,” Microbiology, vol. ED- 146, pp. 851-860, Apr. 2000.
11. M. dos Reis, R. Savva and L. Wernisch, “Solving the riddle of codon usage preferences: a test for translational selection,” Nucleic Acids Res, vol. ED- 32, pp. 5036-5044, Jan. 2004.

