*Journal of Pharmaceutical Research International*

***33(60B): 400-410, 2021; Article no.JPRI.76241***
***ISSN: 2456-9119***
*(Past name: British Journal of Pharmaceutical Research, Past ISSN: 2231-2919, NLM ID: 101631759)*

# Comprehensive Analysis and Assessment on Codon Usage Pattern of Hemolytic Genes from Different Strains of *Leptospira interrogans*

**Chandan Dharmashekara [a], Sabhyata M. Gopal [a#], Gayatri Vaidya [b], Bhargav Shreevatsa [a], H. H. Bhavana [c], Chandrashekar Srinivasa [d], S. M. Gopinath [d], Shiva Prasad Kollur [e], Sharanagouda S. Patil [f], Kuralayanapalya Puttahonnappa Suresh [f], Raghu Ram Achar [g], Ashwini Prasad [c] and Chandan Shivamallu [a*]**

*[a] Department of Biotechnology and Bioinformatics, School of Life Sciences, JSS Academy of Higher Education & Research, Mysuru- 570015, India.*
*[b] Department of Studies in Food Technology, Davangere University, Davangere, 577007, India.*
*[c] Department of Microbiology and Tissue Culture, School of Life Sciences, JSS Academy of Higher Education & Research, Mysuru- 570015, India.*
*[d] Department of Studies in Biotechnology, Davangere University, Shivagangotri, Davangere-577007, India.*
*[e] Department of Sciences, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru Campus, Mysuru, Karnataka – 570 026, India.*
*[f] ICAR- National Institute of Veterinary Epidemiology and Disease Informaatics (NIVEDI), Yelahanka, Bengaluru-560064, India.*
*[g] Division of Biochemistry, School of Life Sciences, JSS Academy of Higher Education and Research, Mysuru, Karnataka 570015, India.*

***Authors' contributions***

*This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.*

***Article Information***

DOI: 10.9734/JPRI/2021/v33i60B34635

**Open Peer Review History:**
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/76241

***Original Research Article***

***Received 09 October 2021***
***Accepted 18 December 2021***
***Published 22 December 2021***

*[#]Equal contribution to first author.*
**Corresponding author: E-mail: chandans@jssuni.edu.in;*

## ABSTRACT

**Background**: Leptospirosis is a disease that is zoonotic in nature. It is easily able to survive in the environment and gets transmitted from the reservoir hosts such as rat to humans. Though the cause of *Leptospira interrogans* virulency is not determined, these studies expresses that the product of a *Leptospira interrogans* gene sphH is hemolytic in nature and is capable of lysing the erythrocyte cells and also few epithelial cell membranes.
**Aim:** In this study, we have analyzed the codon usage bias of the gene sphH, taken from 49 strains of *Leptospira interrogans*.
**Materials and Methods**: Data Collection: The full CDSs nucleotide of the gene sphH from 49 *Leptospira interrogans* was downloaded separately in FASTA format from the NCBI nucleotide database. Molecular Evolutionary Genetics Analysis (MEGA) program was used to edit and align the coding sequences. With the help of statistical techniques such as RSCU and ENC we have determined the codons mostly used to express the gene sphH and source of the bias with methods such as parity rule 2, neutrality plots.
**Result**: The nucleotide content of the , gene across 49 variants of leptospira divulged that the usage of A (36.03% ± 0.74) and T (31.81% ± 0.65) occurred more frequently than the usage of G (17.88% ± 0.57) and C (14.28% ± 0.66) making the gene composition AT more than GC. The composition of bases in the 3$^{rd}$ position saw base T most frequent than base A.
**Conclusion**: The codon and amino acid use patterns reflect this bias in genetic composition. Leptospiral strains have a similar general codon usage pattern and are slightly biassed. The majority of the frequently occurring codons are A- and U-ending, showing that mutational bias is the primary driver determining codon usage in this bacterium. There were significant changes in synonymous codon usage frequencies between *Leptospira interrogans* and with other virulent strains. Codon use preferences to account for these differences.


*Keywords: Leptospira interrogans; sphH genes; Codon.*

## ABBREVIATION

*RSCU : Relative Synonymous Codon Usage*
*ENC : Effective Number of Codons*
*PR2 : Parity rule 2*

## 1. INTRODUCTION

In eubacteria spirochetes form a major phylum because of their unique and flexeous shape. Liptospires are physiologically chemoheterotrophs and is the only genus other than *Borrelia, Treponema and Brachyspira* that causes infection in mammals [1]. *Leptospira interrogans* is known worldwide for the life-threatening infections they cause in mammals and their significant colonization in rodents. The organism has the ability to respond to a wide variety of stimulus in the environment due to the presence of genes coding for regulatory system, signal transduction and methyl-accepting chemotaxis proteins. The study of toxins, lipoproteins and several surface-exposed proteins related genes can serve as a target for vaccine [2]. The molecular processes underpinning leptospirosis pathogenicity poorly understood [3]. Whilst the cytotoxicity of the hemolytic protein in the Interrogans is known, it can be suspected that the pathogenicity lies in the ability of these species to hemolyze the erythrocytes. The evidence of genetic transfer of the hemolytic gene from *L.biflexa* to *L.interrogans* [4] might suggest that this gene perhaps is a very important factor of the organisms virulency. sphH is a gene present in *Leptospira interrogans*, that translated into a hemolytic protein. The sphH gene was shown to be substantially conserved in pathogenic leptospires, hence it appears to represent a key virulence component [5]. The time spent incubating together with the concentration of *Leptospira hemolysin* sphH were factors, that lead to mammalian cell injury.

The reason for choice of a particular codon over the other synonymous codons is an enigma. Any organism tends to skew towards a particular set of codons composition in the gene for the production of amino acid [6]. The codon usage can vary between organisms or throughout the genome of a single organism. Few studies suggest that there is a direct relation between that usage bias and the tRNA composition for that particular codon. Which means that the codon bias tends to influence the composition of tRNA in the cell [6,7]. This if thought of is a very

efficient way of functioning in a cell. Carrying the complementary tRNA for all the synonymous codons can be an energy and time-consuming task for the cells. Having codon bias will in a way indirectly increase the speed of translation and hence enhanced expression of protein. There are also theories which suggests that Codon usage bias is an effect of natural selection or of mutational pressure. The effect of the environmental conditions to the codon usage bias can be considered [8,6]. Regardless of their evolutionary variety, it is said that microorganisms residing in the same eco system have a similar affinity for codon use [9-11].

## 2. MATERIALS AND METHOD

### 2.1 Data Collection

Nucleotide sequences are the most significant aspect of the codon usage analysis. The full CDSs nucleotide of the gene sphH from 49 *Leptospira interrogans* was downloaded separately in FASTA format from the NCBI nucleotide database. All in all, 49 sequences were identified and examined. Molecular Evolutionary Genetics Analysis (MEGA) program was used to edit the coding sequences of each nucleotide individually. Further all the edited nucleotides were aligned using the same software.

### 2.2 Overall Nucleotide Content Analysis

The Nucleotides A, T, G, C are part of the genome. The nucleotide content was determined for the gene sphH. In addition to this a calculation was done to determine the overall prevelence of A, T, G, and C at the third position (A3, T3, G3, C3). This also included the complete GC content and the GC content located at the positions 1, 2, and 3 (GC1, GC2, GC3). This was done using the MEGA tool and also with the help of seqinr package In R languages[12].

### 2.3 Relative Dinucleotide Abundance Analysis

Some have hypothesized that the relative dinucleotide abundance can determine the preference of codon in the organism. It is possible to get 16 different combinations of dinucleotide, and the dinucleotide frequency outlines both mutational and selection pressure [13]. The calculations to measure the relative dinucleotide abundance of the gene sphH are made using the method specified by Karlin and Burge [14].

$$P_{xy} = F_{xy} / (F_x F_y)$$

Here the individual frequencies of nucleotides are denoted as "$F_X$" and "$F_Y$", dinucleotides denoted as "$F_{XY}$". With regard to standards, $P_{XY}$ is considered high at a $P_{XY}$ value of 1.23, and $P_{XY}$ is considered poor when $P_{XY}$ is less than 0.73 [14]. Programming tools called R Studio was used to determine dinucleotide frequencies, in order to use an external library known as “seqinR”

### 2.4 Relative Synonymous Codon Usage (RSCU) Analysis

The RSCU method is elucidated for a given amino acid as the ratio of the observed to contemplated value. The RSCU values are relatively unaffected by the amino acid frequency and also the length of the sequence the codon. Codons with a value greater than 1.6 are considered to be overrepresented, while codons with a value less than 0.6 are regarded to be underrepresented. Codon values between 1.6 and 0.6 are considered to be unbiased or arbitrarily used. To measure the RSCU values, the following formula has been used:

$$RSCU = g_{ij} / \sum_{j}^{i} g_{ij}$$

Where $g_{ij}$ is the number of the $i_{th}$ codon for the $j_{th}$ amino acid incorporating $n_i$ synonymous codons. The RSCU values were determined and visualized using the R Studio programming environment and the “seqinR” library.

### 2.5 Effective Number of Codons (ENC) Analysis

An ENC analysis represents the codon's divergence from random selection. The effective number of codons is usually between 20 and 61[15]. The meaning 20 denotes an extremely skewed situation in which each amino acid is coded by a single codon. Value 61 indicates that there is no bias and that all codons have been used equally. If the ENC values are less than 45, the codon consumption is considered moderately biased [13]. The ENC value was computed using the equation below:

$$ENC = 2 + 9/F_2 + 1/F_3 + 5/F_4 + 5/F_6$$

Where $F_i$, i being 2,3, 4, 6 denotes the average $F_i$ in the *i*- fold degenerate amino acid family.

Here the $F_i$ value is calculated using:

$$F_i = \left. n \sum_{j=1}^{i} \left(n_j / n\right)^2 - 1 \right/ n - 1$$

Here,

n = sum of codons for particular aminoacid.
$n_j$= sum of the observed $^{jth}$ codon for a particular amino acid. The ENC were calculated in R language using the library, "vhica" [15].

The ENC plot was created to show the relationship between an efficient number of codons and GC3 that is the number of G&C nucleotides in the 3rd position. This approach identifies and quantifies gene or genome codon utilization bias. ENC values are calculated using the formula

$$ENC^{expected} = 2 + S + \left(29 / S^2 + (1 - S)^2\right)$$

Where S = the sum of G&C nucleotide at the third position. ENC values are located on the normal curve, they indicate that mutation strain has an effect on codon use. Values that are under the normal curve mean that the values are limited by some other aspect, such as natural selection.

## 2.6 Neutrality Plot Analysis

The neutrality plot analysis is conducted to ascertain the effects of mutational pressure and natural selection on the pattern of codon usage. In order to exemplify the neutrality plot, the GC3 values were plotted against the mean of GC12. GC3 values seem to be important when GC3 is closer to 1. Since this is the case, it seems that the degree of mutational constraint is also significant in building the codon use pattern. If the regression slope equals zero, natural selection is likely to be a significant factor [15].

## 2.7 Parity Rule 2 (PR2) Plot Analysis

The PR2 or Parity rule 2 analysis was carried out by representing the GC bias on the x-axis [G3/(G3+C3)] and the AT bias on the y-axis [A3/(A3+T3)]. Typically, depending on the genome structure, the study shows the relative magnitudes of natural selection and mutation pressure. All axes would have an origin of 0.5 (X= 0.5, Y= 0.5). This implies that A=T, G=C. points located on the origin imply that natural selection and mutational pressure are identical.

# 3. RESULTS

## 3.1 Nucleotide Content Analysis of sphH gene in *Leptospira interrogans*

To assess the extent of codon, use bias in the sphH gene, the nucleotide compositions A, T, G, C, and the G+C contents that is GC content, first codon position with GC content, second codon position GC content, and third codon position GC content were determined. To test the consequence of nucleotide composition on codon use trends, the nucleotide composition was determined. The nucleotide frequency values were as follows. The nucleotide content of the sphH gene across 49 variants of leptospira divulged that the usage of A (36.03% ± 0.74) and T (31.81% ± 0.65) occurred more frequently than the usage of G (17.88% ± 0.57) and C (14.28% ± 0.66) making the gene composition AT more than GC. The composition of bases in the 3$^{rd}$ position saw base T most frequent than base A, while base G and C were the least frequent. This implied that though the overall content of nucleotide A was the most, nucleotide T(U) took most of the 3$^{rd}$ position. The overall GC content was lesser at (32.15% ± 0.16). This study corelated with the reported overall GC content of Spirochaetes at 40.6% making it GC-poor (Table 1).

## 3.2 Relative Dinucleotide Abundance Frequency Analysis

The use of the dinucleotide bias will affect the bias in codon use. R Studio software was used to measure the relative abundance of total 16 dinucleotides of gene sphH. As compared to a theoretical value, the intensity of frequency of each section was found to be less consistent that is equal to 1.0 (Fig. 2). Overall abundance level is calculated depending on how many times it is overrepresented (i.e.>1.23) and how few times it is underrepresented (i.e. <0.78) (Table 2) (Fig. 2).

**Table 1. Nucleotide composition of sphH Gene in *Leptospira interorgan* nucleotide abundance frequency**

| | | | | Dinucleotide | | | | |
|---|---|---|---|---|---|---|---|---|
| Frequency | 1.18616 | 0.929776 | 0.732551 | 0.970432 | 1.000392 | 0.956642 | 0.906515 | 1.104526 |
| Dinucleotide | GA | GC | GG | GT | TA | TC | TG | TT |
| Frequency | 1.080043 | 0.858804 | 1.427725 | 0.741657 | 0.758316 | 1.144935 | 1.107111 | 1.128854 |
| Dinucleotide | AA | AC | AG | AT | CA | CC | CG | CT |

**Table 2. Frequency of the 16 Dinucleotide in the Gene sphH**

| Nucleotides | T(U) | C | A | G | T-3 | C-3 | A-3 | G-3 | GC | GC1 | GC2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average with | 31.81% ± | 14.28% ± | 36.03% | 17.88% ± | 44.73% ± | 10.57% ± | 34.25% ± | 10.44% ± | 32.15% ± | 42.74% ± | 32.69% ± |
| SD | 0.65 | 0.66 | ± 0.74 | 0.57 | 1.76 | 2.70 | 1.81 | 1.09 | 0.16 | 3.16 | 0.65 |

**Table 3. RSCU Analysis of the gene sphH. Red labeled indicates the over represented codons, and blue labeled are underrepresented codons.**

| AA | CODON | RSCU | AA | CODON | RSCU | AA | CODON | RSCU |
|---|---|---|---|---|---|---|---|---|
| Lys | Aaa | 1.863636 | Pro | Cca | 0.705882 | Gly | Ggc | 0.102564 |
| Asn | Aac | 0.391304 | Pro | Ccc | 0.235294 | Gly | Ggg | 0.102564 |
| Lys | Aag | 0.136364 | Pro | Ccg | 0.705882 | Gly | Ggt | 1.435897 |
| Asn | aat | 1.608696 | Pro | Cct | 2.352941 | Val | Gta | 1.866667 |
| Thr | aca | 1.142857 | Arg | cga | 1.2 | Val | gtc | 0.266667 |
| Thr | acc | 0.571429 | Arg | cgc | 0.6 | Val | gtg | 0.4 |
| Thr | acg | 0.142857 | Arg | cgg | 0.6 | Val | gtt | 1.466667 |
| Thr | act | 2.142857 | Arg | cgt | 0.3 | Tyr | tac | 0.52381 |
| Arg | aga | 3 | Leu | Cta | 0.837209 | Tyr | tat | 1.47619 |
| Ser | agc | 0.176471 | Leu | Ctc | 0.55814 | Ser | tca | 0.882353 |
| Arg | agg | 0.3 | Leu | Ctg | 0.837209 | Ser | tcc | 0.705882 |
| Ser | Agt | 1.411765 | Leu | Ctt | 1.395349 | Ser | tcg | 0.705882 |
| Ile | Ata | 1 | Glu | gaa | 1.75 | Ser | tct | 2.117647 |
| Ile | Atc | 0.333333 | Asp | gac | 0.315789 | Trp | tgg | 1 |
| Met | Atg | 1 | Glu | gag | 0.25 | Cys | tgt | 2 |
| Ile | Att | 1.666667 | Asp | gat | 1.684211 | Leu | tta | 1.534884 |
| Gln | caa | 1.84 | Ala | gca | 2.086957 | Phe | ttc | 0.344828 |
| His | cac | 0.4 | Ala | gcc | 0.173913 | Leu | ttg | 0.837209 |
| Gln | cag | 0.16 | Ala | gct | 1.73913 | Phe | ttt | 1.655172 |
| His | cat | 1.6 | Gly | gga | 2.358974 | | | |

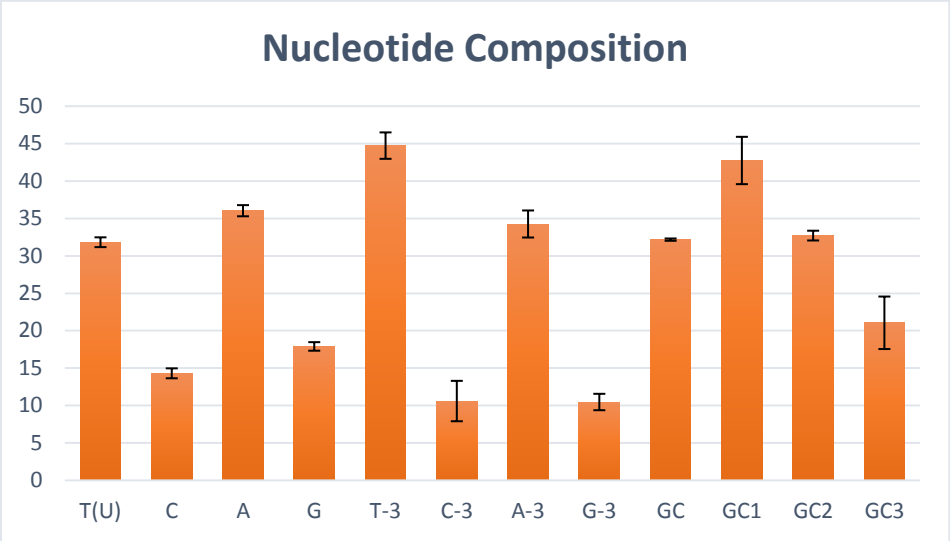


**Fig. 1. Graph showing the nucleotide composition in the gene sphH and along with overall GC contents and GC content at first, second and third position**

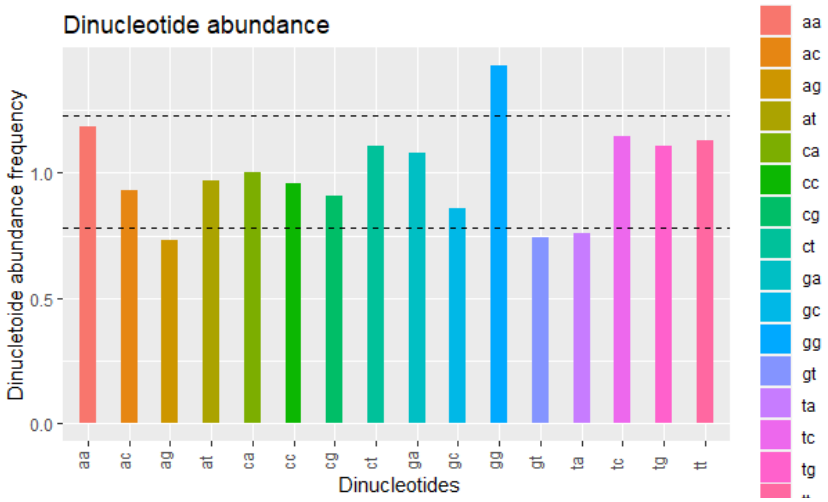


**Fig. 2. Graph showing Dinucleotide abundance frequency of the gene sphH. With the dotted line representing a limit for underrepresented dinucleotides at 0.78 and overrepresented dinucleotides at 1.23**

### 3.3 Relative Synonymous Codon Usage (RSCU) Analysis

The RSCU measures fall between 0.6 and 1.6. Values that are less than 0.6 are considered to be underrepresented, while values that are greater than 1.6 are considered to be overrepresented. The more 1.0 the codon gets, the more positive the codon bias; conversely, the less 1.0 the codon gets, the more negative the codon bias. The following conclusion was reached in relation to the Relative Synonymous Codon Use of each segment (Table 3) (Fig. 3).

### 3.4 Effective Number of Codons (ENC) Analysis

To measure the frequency of the pattern of codon use, the effective number of codon values was calculated. The ENC values ranged between 42.4 and 49.2. The mean of a normal distribution with a standard deviation of 42.92 ± 1.18 As one plots an ENC-GC3 graph, all points are seen to be under the standard curve, showing the possibilities of strong mutational strain. The GC3S values were plotted against ENC to derive an ENC-plot (Fig. 4). This plot revealed that all spots congregated somewhat below on the left

side of the predicted curve, demonstrating that mutational pressure might not be the sole factor impacting codon usage bias, but other factors, like as natural selection, may well be at play.

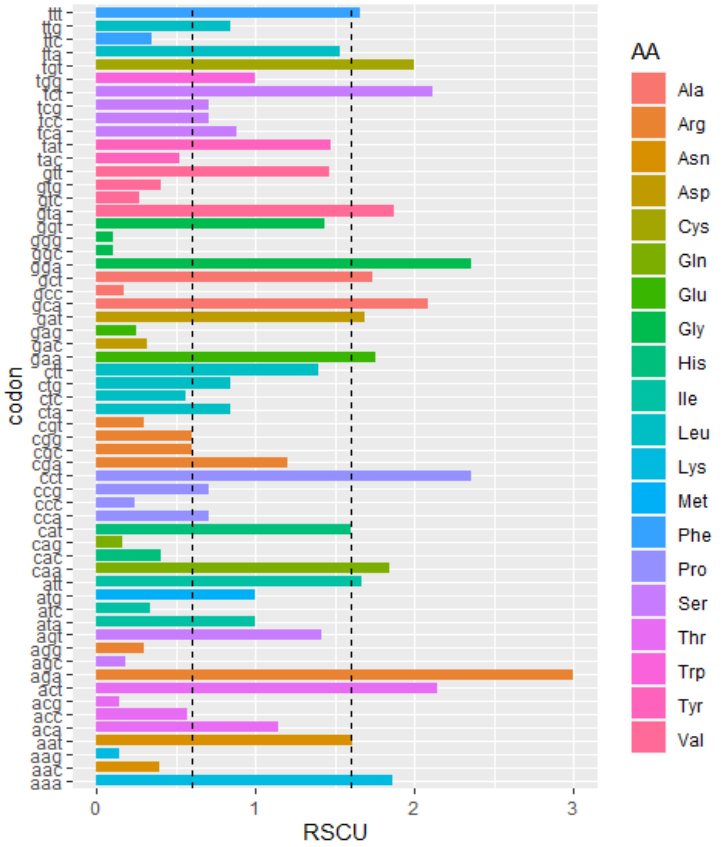


**Fig. 3. Graph showing RSCU analysis of gene SphH. With overrepresented codons in the area above the dotted line (>1.6) and underrepresented in the area below the dotted lines (<0.6)**

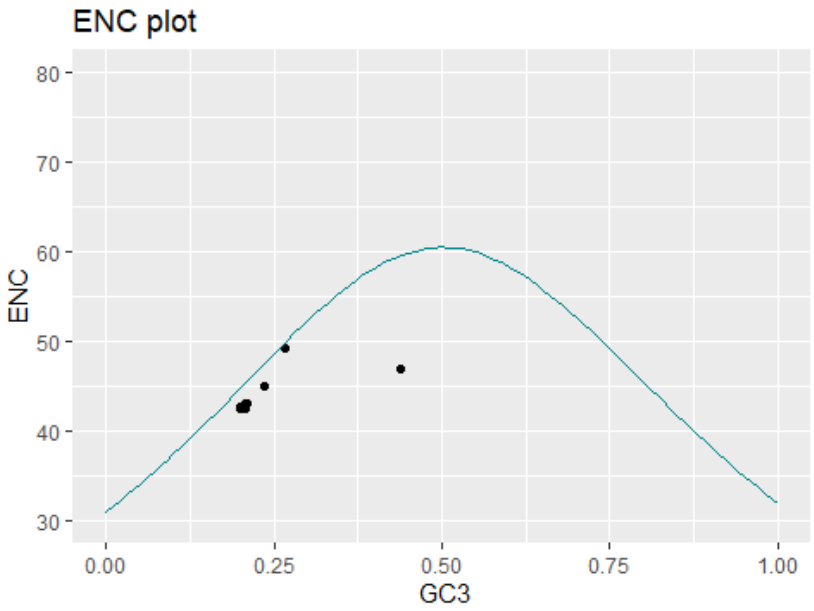


**Fig. 4. An ENC plot showing all the points plotted below the curve**


**Fig. 5. Parity rule 2Plot of gene sphH**

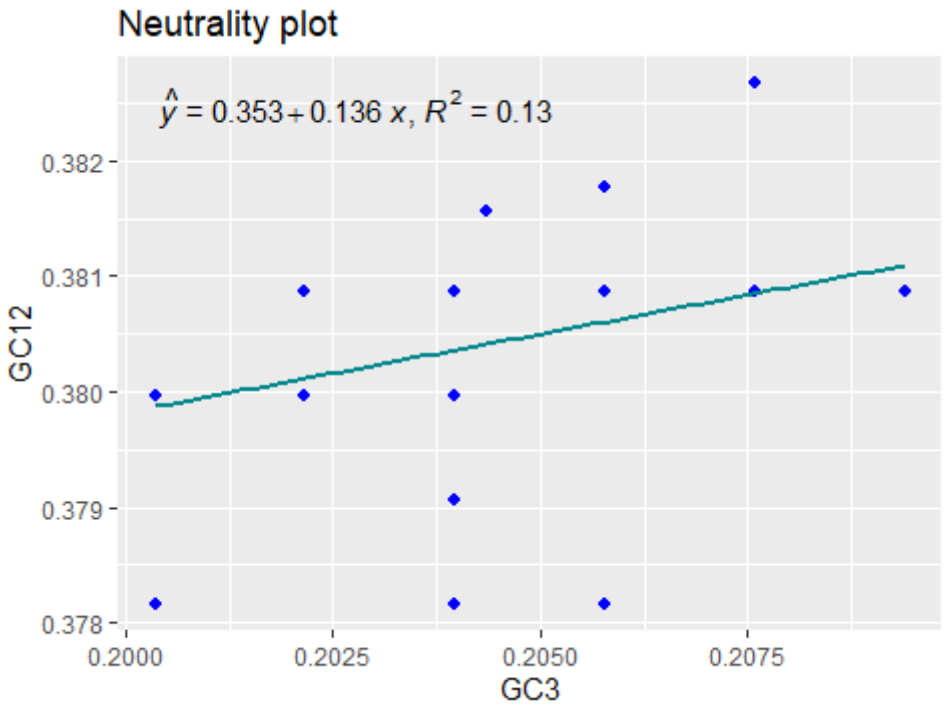


**Fig. 6. Neutrality plot of gene sphH having a slope of 0.136 and an $R^2$ value of 0.113**

### 3.5 Parity Rule 2 (PR2) Plot Analysis

The results of the Parity Rule 2 research were employed to study the influence of selection and mutation strain. To demonstrate the PR2 curve, the values of the AT bias of 3rd position and the GC bias of 3rd positions were used as weights to a PR2 scale. For X- coordinates, [G3/(G3+C3)]

represent; for Y- coordinates, [A3/(A3+T3)] represent. The GC and AT bias of the gene SphH is approximately equal to the mean value of 0.49 and 0.43 respectively, recommending pyrimidine choice instead of purine preference. Parity rule 2 analysis is done to understand the force that is driving this codon usage bias. It was observed that A≠T(U) and G≠C which means that there is some kind of bias in the usage of codons (Fig. 5).

### 3.6 Neutrality Plot Analysis

The plot investigated the relationship and primary factors (mutational pressure and natural selection) between GC12 and GC3, which were shown using the mean values of the first and second positions of GC.

For the most part, the most important factors deciding the codon use in the sphH gene were established by neutrality plot analysis. It was discovered that GC3s (the previous version of GC12s) had a similarity of 0.13 with GC12s (the new version of GC12s). As can be seen in the graph, the regression line slope was just 0.136, which shows that natural selection was the primary driver in the codon consumption trend of the sphH gene, with mutation pressure playing a minor role. The neutrality plot seen between GC3s and GC12s values was designed to assess the magnitude of the two evolutionary influences on the codon usage pattern of sphH gene. The neutrality plot saw a positive correlation, of y= 0.353+0.136x the pearsons coefficient($R^2$) was a positive 0.13 indicating that natural selection may have been the dominant factor, although mutation pressure had a small effect in determining the CUB of the sphH gene (Fig. 6).

## 4. DISCUSSION

In this study, codon usage bias for the sphH gene in leptopirainterrogans is analyzed. The nucleotide content of the sphH gene across 49 variants of leptospira divulged that the usage of A (36.03% ± 0.74) and T (31.81% ± 0.65) occurred more frequently than the usage of G (17.88% ± 0.57) and C (14.28% ± 0.66) making the gene composition AT more than GC. The composition of bases in the 3rd position saw base T most frequent than base A, while base G and C were the least frequently. This implied that though the overall content of nucleotide A was the most, nucleotide T(U) took most of the 3rd position. The overall GC content was lesser at (32.15% ± 0.16). This study corelated with the reported overall GC content of spirochaetes at 40.6% making it GC-poor.

The RSCU analysis was done to testify if the extent of A/T in the 3rd position is preferred. Out of 62 codons, there were 16 codons that were overrepresented, and 21 codons underrepresented. The overrepresented codons were aaa, aat, act, aga, att, caa, cct, gaa, gat, gca, gct, gga, gta, tct, tgt, ttt. All the over expressed codons had T or an A in the 3rd position with none having a G or a C. Precisely out of the 16 over represented codons 9 of them were Thymine-ending the rest 7 were Adenine-ending.

We examined over-representative and under-representative dinucleotides with favored and under-representative codons to see how dinucleotide use affects codon use bias. In the dinucleotide frequency there was more abundance of GpG with frequency of 1.43 and the least abundant were ApG, GpT and TpA with frequencies 0.73, 0.74, 0.75 respectively. Out of seven ApG-containing codons, five codons: AAG, AGC, AGG, CAG and GAG had RSCU values less than 0.6, suggesting that dinucleotide ApG were inhibited. Furthermore, amongst all eight GpT dinucleotides 3 codons GTA, GTG, CGT were under expressed and 3 other codons had RSCU values lesser than 1.6, also there were six expressions of the dinucleotide TpA out of which the only codon coding for Valenine-GTA was over expressed, and the five codons had the values lesser than 1.6, suggesting that all the three codons were probably under expressed. There was an interesting phenomenon observed for the dinucleotide GpG, this dinuclotide as mentioned was expressed in large numbers, but out of seven codons only one codon GGA coding for Glycine was over expressed, while the rest six of them were under expressed.

To evaluate the forces shaping the codon usage patterns in the gene SphH, PR2 bias, ENC plots, and neutrality analyses were carried out. Parity rule 2 analysis is done to understand the force that is driving this codon usage bias. It was observed that A≠T(U) and G≠C which means that there is some kind of bias in the usage of codons. The GC3S values were plotted against ENC to derive an ENC-plot. This plot revealed that all spots congregated somewhat below on the left side of the predicted curve, demonstrating that mutational pressure might not be the sole factor impacting codon usage bias,

but other factors, like as natural selection, may well be at play. The neutrality plot seen between GC3s and GC12s values was designed to assess the magnitude of the two evolutionary influences on the codon usage pattern of sphH gene. The neutrality plot saw a positive correlation, of y= 0.353+0.136x the pearsons coefficient ($R^2$) was a positive 0.13. indicating that natural selection may have been the dominant factor, although mutation pressure had a small effect in determining the CUB of the sphH gene.

## 5. CONCLUSION

The findings of this research revealed that the genomic composition of the *Leptospiral* strains studied differs from that of other sphH gene strains. The usage patterns of codons and amino acids reflect this unique genetic composition. The majority of the frequently occurring codons are A- and U-ending, implying that mutational bias is the primary driver determining codon usage in these bacteria. There were significant differences in RSCU frequencies between *Leptospira* and with other virulent strains. Codon use preferences to account for these differences.

## CONSENT

It is not applicable.

## ETHICAL APPROVAL

It is not applicable.

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES


1. Ren, Shuang-Xi, Gang Fu, Xiu-Gao Jiang, Rong Zeng, You-Gang Miao, Hai Xu, Yi-Xuan Zhang, et al. Unique physiological and pathogenic features of leptospira interrogans revealed by whole-genome sequencing. *Nature. 2003;422:6934*-422(6934):888–93. DOI:https://doi.org/10.1038/nature01597.
2. Nascimento ALTO, Verjovski-Almeida S, Van Sluys MA, Monteiro-Vitorello CB, Camargo LEA, Digiampietri LA, Harstkeerl RA, et al. Genome features of leptospira interrogans serovar copenhageni. Brazilian Journal of Medical and Biological Research. 2004;37(4):459–78. DOI:https://doi.org/10.1590/s0100-879x2004000400003.
3. Lehmann JS, Matthias MA, Vinetz JM, Fouts DE. Leptospiral pathogenomics. Pathogens (Basel, Switzerland). 2014;3(2):280–308. DOI:https://doi.org/10.3390/pathogens3020280
4. Chadha T, Alexandre Trindade A. Phylogenetic analysis of genetic diversity of hemolysins in Leptospira. Journal of Proteomics and Bioinformatics. 2012;5(7):152-154. DOI:https://doi.org/10.4172/jpb.1000228
5. Lee SH, Kim KA, Park YG, Seong IW, Kim MJ, Lee YJ. Identification and partial characterization of a novel hemolysin from Leptospira interrogans serovar lai. Gene. 2000; 254(1-2):19–28. DOI:https://doi.org/10.1016/s0378-1119(00)00293-6
6. Quax TE, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Molecular cell. 2015;59(2):149–161. DOI:https://doi.org/10.1016/j.molcel.2015.05.035
7. Rocha EP. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Research. 2004;14(11):2279–2286. DOI:https://doi.org/10.1101/gr.2896904
8. Franzo G, Tucciarone CM, Legnardi M, et al. Effect of genome composition and codon bias on infectious bronchitis virus evolution and adaptation to target tissues. BMC Genomics. 2021;22: 244. DOI:https://doi.org/10.1186/s12864-021-07559-5
9. Hart, Andrew, Cortés, María, Latorre, Mauricio, Martinez, Servet. Codon usage bias reveals genomic adaptations to environmental conditions in an acidophilic consortium. PLOS ONE. 2018;13:e0195869. DOI:10.1371/journal.pone.0195869.
10. Fuglsang A. The 'effective number of codons' revisited. Biochemical and Biophysical Research Communications. 2004;317(3):957–964. DOI:https://doi.org/10.1016/j.bbrc.2004.03.138.
11. Yengkhom S, Uddin A, Chakraborty S. Deciphering codon usage patterns and

evolutionary forces in chloroplast genes of *Camellia sinensis* var. assamica and *Camellia sinensis* var. sinensis in comparison to *Camellia pubicosta*. Journal of Integrative Agriculture. 2019;18:2771-2785.

12. Deb B, Uddin A, Chakraborty S. Codon usage pattern and its influencing factors in different genomes of hepadnaviruses. Archives of Virology. 2020;165(3):557–570. DOI:https://doi.org/10.1007/s00705-020-04533-6
13. Khandia R, Singhal S, Kumar U, Ansari A, Tiwari R, Dhama K, Das J, Munjal A, Singh RK. Analysis of Nipah virus codon usage and adaptation to hosts. Frontiers in Microbiology. 2019;10:886. https://doi.org/10.3389/fmicb.2019.00886
14. Tao P, Dai L, Luo M, Tang F, Tien P, Pan Z. Analysis of synonymous codon usage in classical swine fever virus. Virus Genes. 2009;38(1):104–112. DOI:https://doi.org/10.1007/s11262-008-0296-z
15. Pan S, Mou C, Wu H, Chen Z. Phylogenetic and codon usage analysis of atypical porcine pestivirus (APPV). Virulence. 2020;11(1):916–926. DOI:https://doi.org/10.1080/21505594.2020.1790282



*Peer-review history:*
*The peer review history for this paper can be accessed here:*
*https://www.sdiarticle5.com/review-history/76241*