International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 8, Suppl 1, 2016

Full Proceeding Paper

CODON USAGE IN HUMAN MITOCHONDRIAL GENES IN THE CONTEXT OF CANCER

ARIF UDDIN[1], SUPRIYO CHAKRABORTY[2]

[1,2]Department of Biotechnology, Assam University, Silchar 788011, Assam, India
Email: supriyoch_2008@rediffmail.com

Received: 01 Dec 2015 Revised and Accepted: 12 Feb 2016

ABSTRACT

Objective: Mitochondria are the powerhouse of the cell**.** Mitochondrial DNA is more susceptible to oxidative damage due to the lack of histone protein and chromatin structure. The alteration in the level of gene expression in cytochrome c oxidase gene is associated with cancer. The expression of *coxiii* gene was found to be lower in human colonic carcinoma. However, a systematic analysis of codon usage in human mitochondrial protein-coding genes has not been reported yet. This study gives an insight into the understanding of the pattern of codon usage and expression in human mitochondrial genes.

Methods: We used a bioinformatics approach to analyse the codon usage parameters by using bioinformatics tools like an effective number of codons (ENC), codon adaptation index (CAI), relative synonymous codon usage (RSCU) etc.

Results: The comparison of codon usage pattern among different mitochondrial genes suggests that mitochondrial genes have a lower level of codon usage bias and high expression level. Highly significant positive correlation between ENC and GC3 ($r=0.782^{**}$, $p<0.01$), nucleobases C and C3 ($r=0.655^{*}$, $p<0.05$), GC and GC3 ($r=0.690^{**}$, $p<0.01$) suggest that mutation pressure played an important role in codon usage bias. Highly significant positive correlation was found between ENC and CAI ($r=0.762^{**}$, $p<0.01$). The over-represented codons are TCA, TCC, CTA, CTC, CAA, CGC, TGA, ATA, AAA, GTA, GCC, GAA and GGC while the under-represented codons are TCG, AGT, CTG, CCG, CAG, CGT, ACG, AAT, GTG, GAT, GGG and ATG.

Conclusion: Mutation pressure is found to play major roles in shaping the low bias in the protein-coding genes of human mitochondrial DNA, although codon usage bias is weak. The over-represented and under-represented codons are used to increase or decrease the expression level. In addition, codon usage bias has influenced the gene expression in human mitochondrial genes.

Keywords: Mitochondrial DNA, Synonymous codon usage bias, Gene expression

INTRODUCTION

In the genetic code, there are 64 codons, which encode 20 standard amino acids and termination signals, with all its amino acids being encoded by two to six synonymous codons. Synonymous codons that encode the same amino acid do not show the same frequency in coding sequences and during the translation of gene to protein, the unequal usage of synonymous codons in coding sequences is the codon usage bias (CUB) [1]. CUB exists in an extensive variety of organisms, from prokaryotes to even multicellular eukaryotes [2, 3]. The pattern of the usage of the codon usually differs among the genes of an organism and also among the different organisms [4]. In few prokaryotes and unicellular eukaryotes, CUB is influenced by the equilibrium between mutation bias and natural selection [5, 6]. It has been found that in some prokaryotes and mammals having high AT or GC contents, codon usage variation is mainly influenced by the mutational pressure[7]. However, in the case of *Drosophila*, translational selection plays the important role in shaping codon usage pattern [8].

Since the beginning of genome sequencing of different organisms, analysis of codon usage bias has gained its renewed attention [9]. However, the analysis of CUB has many other important applied aspects, such as heterologous gene expression, prediction of gene expression level, determination of the origins of species, the design of degenerate primers, as well as the prediction of gene functions [10].

Mitochondrial genome (about 16.6kb) is a covalently closed-circular and a double-helical molecule which encodes two rRNAs, 22 tRNAs and 13 polypeptides which are involved in respiration [11]. As the mitochondrial DNA (mtDNA) possesses no histones as well as no introns, hence it is found to be more prone to oxidative damage due to reactive oxygen species (ROS) which are produced as a by-product of electron transport system (ETS) [12]. ETS is located in the inner membrane of mitochondria consisting of four respiratory enzyme complexes. These are complex I which contains seven subunits of respiratory enzyme, complex III contains one subunits, complex IV contains three subunits and complex V two subunits, whereas, the nuclear genes encode all other mitochondrial proteins involved in replication, transcription and translation of mtDNA [12].

Cancer, a group of diseases, is associated with abnormal continuous cell growth and spreads to other body parts. Various previous studies reported that cancer is caused by the mitochondrial dysfunction, where due to an impaired respiratory capacity, various tumors are formed [13]. It was also reported that the expression of these mitochondrial protein-coding genes increases in breast cancer, ovarian cancer, colon cancer and prostate cancer [14].

In the present study, we have carried out analysis of the codon usage to elucidate the over and under-represented codons in mitochondrial protein-coding genes in order to understand the molecular mechanism along with functional conservation of gene expression during the period of evolution using several bioinformatics tools.

METHODS

Availability of sequences data

The coding sequence (cds) of mitochondrial DNA in human was retrieved from National Center for Biotechnology Information, USA (http://www.ncbi.nlm.nih.gov/).

Compositional properties

The overall nucleotide composition (A, C, T and G %) and nucleotide composition in codon 3rd position, the overall GC % and the GC % at the 1st, 2nd and 3rd position of codon were calculated. All the calculations were done using a Perl script developed by SC (Corresponding author).

Synonymous codon usage bias indices

Some of the most relevant and widely used measures of codon usage bias analyzed in this study are discussed below.

Relative Synonymous Codon Usage (RSCU)

Relative synonymous codon usage (RSCU) was calculated as the observed frequency of a codon divided by the expected frequency if all the synonymous codons of a particular amino acid are used equally. The RSCU is calculated as,

$$RSCUij = \frac{Xij}{\frac{1}{ni}\sum_{j=1}^{ni} Xij}$$

where X_{ij} is the frequency of occurrence of the j^{th} codon for i^{th} amino acid (any X_{ij} with a value of zero is arbitrarily assigned a value of 0.5), and n_i is the number of codons for the i^{th} amino acid (i^{th} codon family) [15].

Effective number of Codons (ENC)

The effective number of codons (ENC) is used to quantify the codon usage bias of a gene. ENC value ranges from 20 (when only one codon is used for each amino acid) to 61 (when all codons are used randomly). It is calculated as,

$$ENC = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

Where Fk (k= 2, 3, 4, 6) is the mean of Fk values for the k-fold degenerate amino acids [16].

Codon adaptation index (CAI)

The codon adaptation index (CAI) is a very extensively used measure of gene expression. CAI values range from 0 to 1; with higher values indicating a higher proportion of the most abundant codons. The CAI is calculated as,

$$CAI = \exp\left(\frac{1}{L}\sum_{k=1}^{L} \ln \omega k\right)$$

where ωk is the relative adaptiveness of the kth codon, and L is the number of synonymous codons in the gene [15].

Statistical analysis

Correlation analysis was done to identify the relationship between overall nucleotide composition and each base at 3rd position of the codon. All the statistical analyses were done using the SPSS software 21(SPSS Inc., Chicago, IL).

Software used

A novel software developed by Prof (Dr) Supriyo Chakraborty (corresponding author), using Perl script was used to calculate all the codon usage bias parameters used in the present study.

RESULTS

Nucleotide composition

The bases C and A, occurred more frequently than T and G, in the13 protein-coding genes of mitochondrial genomes. At the third codon position C, occurred most frequently, while G was the least (fig 1). The overall nucleotide composition and the nucleotide composition at the third codon position in mitochondrial genomes suggest that the codon usage pattern of this genome might be influenced by the compositional constraint supporting the result of Butt *et. al.* [17].

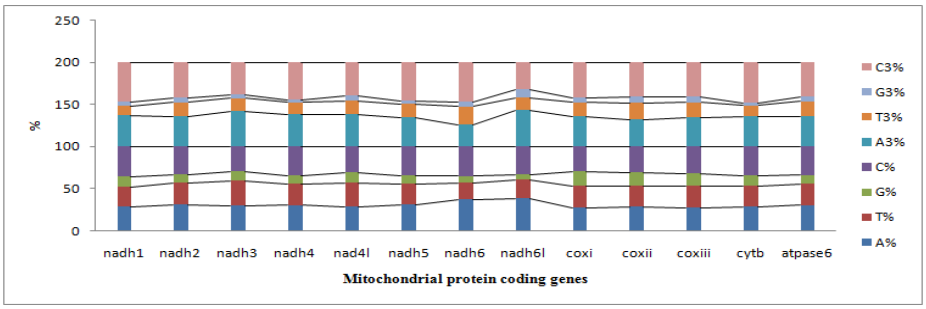

Fig. 1: Distribution of nucleotide composition

It was found that from the fig. (fig 2), the overall GC content is lower than the GC content at the 1st and 3rdcodon position and the GC content at the 2nd codon position is the lowest among all.

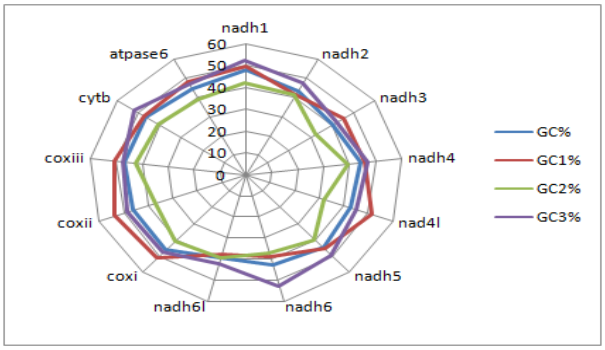

Fig. 2: GC content and its 1st, 2nd and 3rd codon position

Codon usage pattern

Relative synonymous codon usage (RSCU) values of 60 codons also supported the idea that mitochondrial protein-coding genes had a weak codon bias; this is because approximately half of the codons (26/60) were used more frequently. Further, RSCU values in mitochondrial protein-coding genes indicated that at the third codon position, A and C occurred most frequently. The over-represented and the under-represented codons (fig 3 and fig 4) which play a role in increasing or decreasing the gene expression level supporting the result of Carlini *et. al.* [18]. These lead us to hypothesize that in the codon usage pattern of mitochondrial protein-coding genes, the compositional constraint is an essential contributing factor supporting the result of Wei et. al. [19]. After combining the nucleotide composition and the RSCU analysis, we found that the compositional constraints mostly influenced the preferred codon's selection, which strongly suggests the presence of mutational pressure.

Relation between codon usage bias and gene expression level

The ENC values of mitochondrial genes varied from 59 to 60, with a mean value of 56.21. This high ENC value indicates that in mitochondrial protein-coding genes, codon usage bias is weak and is maintained at a stable level. The mean CAI value for all mitochondrial genes was found to be 0.6729, which indicates that mitochondrial genes, in general, have a high expression level. To understand the nucleotide composition variation and codon selection for mitochondrial protein-coding genes, a correlation analysis was done between ENC and CAI. Significant positive correlation was found between ENC and CAI (r=0. 762**, p<0.01) (fig 5) which suggests a very

distinct relationship between codon usage bias and nucleotide composition for mitochondrial protein-coding genes [10].

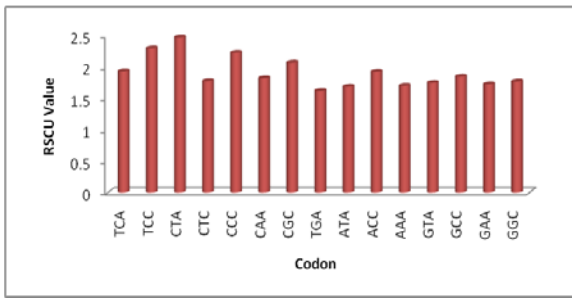

Fig. 3: Frequency of over-represented codons

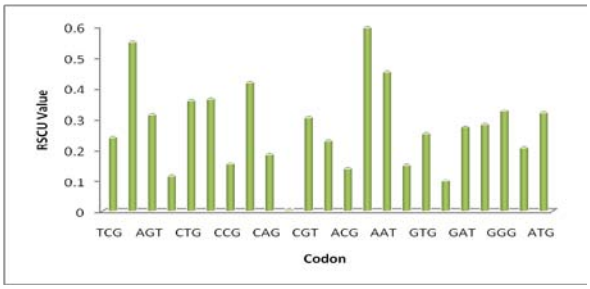

Fig. 4: Frequency of under-represented codons

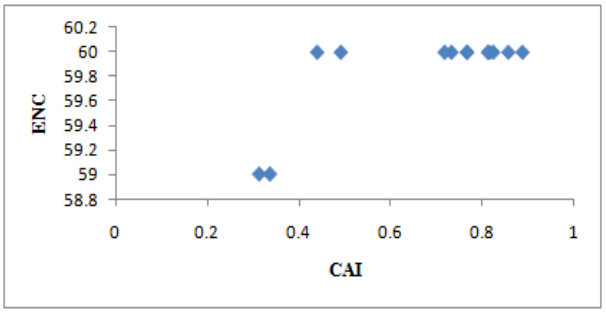

Fig. 5: Relationship between ENC and CAI

Mutation pressure affects the codon usage pattern

To identify whether the evolution of pattern of codon usage in mitochondrial protein-coding genes had been driven alone by mutation pressure or also contributed by the natural selection, we compared the correlation between overall nucleotide composition (A%, T%, G%, C%, GC%) and nucleotide composition at the third position of codon (A3%, T3%, G3%, C3%, GC3%) using the Spearman's rank correlation analysis (table 4). Significant positive correlation was found between ENC and GC3 ($r=0.782^{**}$, $p<0.01$), nucleobases C and C3 ($r=0.655^{*}$, $p<0.05$), GC and GC3 ($r=0.690^{**}$, $p<0.01$) which suggest that mutational pressure played a role in codon usage bias. These results also suggest that compositional constraints under mutational pressure determine the codon usage pattern for mitochondrial protein-coding genes supporting the result of Butt *et. al.*[17].

Table 1: Correlation between overall nucleotide composition (A%, T%, G%, C%, GC %) and nucleotide composition at the third position of codon (A3%, T3%, G3%, C3%, GC3%)

	A3%	T3%	G3%	C3%	GC3%
A%	-.077	.245	.492	-.286	-.104
T%	.502	-.207	-.386	-.200	-.477
G%	-.158	.062	-.204	.203	.152
C%	-.318	-.251	-.084	.508	.623*
GC%	-.448	-.133	-.312	.657*	.690**

DISCUSSION

Synonymous codons are not used uniformly during protein biosynthesis, and different genes from the same species or from different species have an obvious pattern of codon usage bias.

Several factors are involved in codon usage bias, such as GC composition, expression level, gene length, mutational bias, and natural selection [20, 21]. Codon usage pattern is a genetic feature of a variety of organisms. The overall GC% content was lower, and the genes are found to be AT rich. Based on RSCU, the most frequent codons were found to end with A or C supporting the result of Zhang *et. al.* [22].

From RSCU analysis, the over-represented and the under-represented codons in the genes were elucidated. The role of natural selection in codon usage bias is evident from the use of preferred codons that match the most abundant tRNA. It results in an increase in translational efficiency and accuracy. Carlini and Stephan (2001) introduced unpreferred codons into the coding sequences of alcohol dehydrogenase gene (*Adh*) in *Drosophila* and observed a significant decrease in ADH protein production with increasing number of unpreferred codons [18].

The ENC values calculated for the mitochondrial protein-coding genes indicated that the codon usage bias of these genes was weak. Behura and Severson also reported that codon usage bias in Dipteran and Hymenopteran sequenced genomes was also weak [20].

We found mutational pressure affects the codon usage pattern in mitochondrial protein-coding genes. Shacklton *et. al.* also revealed that ENC was highly correlated with the overall GC content in DNA virus, which suggests that mutation pressure mainly influences the codon usage bias [23].

CONCLUSION

The overall codon usage bias of mitochondrial genes was found to be low, and the expression level of the gene was high. Our study also elucidated the under-represented codons which could reduce the gene expression level and thereby hold potential applications in cancer biology.

ACKNOWLEDGEMENT

We are thankful to Assam University, Silchar for providing necessary facilities for this research work. No fund was received from DBT or DST, Govt. of India

CONFLICT OF INTERESTS

We do not have any conflict of interest

REFERENCES

1. Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J Mol Biol 1981;146:1-21.
2. Akashi H, Eyre-Walker A. Translational selection and molecular evolution. Curr Opin Genet Dev 1998;8:688-93.
3. Akashi H. Gene expression and molecular evolution. Curr Opin Genet Dev 2001;11:660-6.
4. Roth A, Anisimova M, Cannarozzi GM. Measuring codon usage bias. Codon evolution: mechanisms and models New York: Oxford University Press Inc; 2012. p. 189-17.

5. Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res 2005;33:1141-53.
6. Bulmer M. Are codon usage patterns in unicellular organisms determined by selection-mutation balance? J Evol Biol 1988;1:15-26.
7. Francino MP, Ochman H. Isochores result from mutation not selection. Nature 1999;400:30-1.
8. Powell JR, Moriyama EN. Evolution of codon usage bias in Drosophila. Proc Natl Acad Sci USA 1997;94:7784-90.
9. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet 2011;12:32-42.
10. Sharp PM, Emery LR, Zeng K. Forces that influence the evolution of codon bias. Philosophical Transactions of the Royal Society B: Bioll Sci 2010;365:1203-12.
11. Singh KK. Mitochondrial DNA mutations in aging, disease and cancer: Springer New York; 1998. p. 1-412.
12. Schatz G. The protein import system of mitochondria. J Biol Chem 1996;271:31763-6.
13. Warburn O, Dickens F. The metabolism of tumors. Am J Med Sci 1931;182:23.
14. Modica-Napolitano JS, Kulawiec M, Singh KK. Mitochondria and human cancer. Curr Mol Med 2007;7:121-31.
15. Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1987;15:1281-95.
16. Wright F. The 'effective number of codons' used in a gene. Gene 1990;87:23-9.
17. Butt AM, Nasrullah I, Tong Y. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PloS One 2014;9:e90905. Doi: 10.1371/journal.pone.0090905. [Article in Press]
18. Carlini DB, Chen Y, Stephan W. The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes *Adh and Adhr*. Genetics 2001;159:623-33.
19. Wei L, He J, Jia X, Qi Q, Liang Z, Zheng Hao, *et al.* Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. BMC Evol Biol 2014;14:262.
20. Behura SK, Severson DW. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. PloS One 2012;7:e43111. Doi:10.1371/journal.pone.0043111. [Article in Press]
21. Sueoka N, Kawanishi Y. DNA G+C content of the third codon position and codon usage biases of human genes. Gene 2000;261:53-62.
22. Zhang Z, Dai W, Dai D. Synonymous codon usage in TTSuV2: analysis and comparison with TTSuV1. PloS One 2013;8:e81469. Doi: 10.1371/journal.pone.0081469. [Article in Press]
23. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J Mol Evol 2006;62:551-63.