International Journal of Pharmacy and Pharmaceutical Sciences

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

Full Proceeding Paper

WHICH EVOLUTIONARY FORCES DICTATE CODON USAGE IN HUMAN TESTIS SPECIFIC GENES?

MONISHA NATH CHOUDHURY[1], SUPRIYO CHAKRABORTY[2]

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

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

ABSTRACT

Objective: Unequal usage of synonymous codons encoding an amino acid is termed as codon usage bias. Synonymous codon usage bias is an inevitable phenomenon in organismic taxa across the three domains of life, *i.e.* plants, animals and microbes. Here we report the codon usage pattern in human testis-specific genes found in Y chromosome. Testis-specific genes are associated with several dysfunctions, such as gonadal sex reversion, infertility, gonadoblastoma and non-syndromic hearing impairment.

Methods: We used bioinformatics approaches to analyze codon usage bias in human testis-specific genes

Results: Highly significant negative correlation was found between ICDI and tAI ($r=-0.939^{**}$, $p<0.01$). Moreover, highly significant positive correlation between A% and A3% ($r=0.774^{*}$, $p<0.05$), T and T3% ($r=0.894^{**}$, $p<0.01$), GC% and GC3% ($r=0.897^{**}$, $p<0.01$) suggest that mutation pressure played an important role in codon usage pattern of these genes. However, significant positive correlation between G and G3 % ($r=0.936^{**}$, $p<0.01$), G and C3 ($r=0.557$, $p>0.05$) but negative correlation between GC and T3 % ($r=-0.960^{**}$, $p<0.01$) indicate the role of natural selection on codon bias. Variation of codon usage pattern was also evident in different genes from principal component analysis (PCA).

Conclusion: Codon usage bias in human testis-specific genes is low. These genes are rich in GC content. Both natural selection and mutation pressure affect the codon usage bias in these genes.

Keywords: Codon usage bias, Mutation pressure, Natural selection

INTRODUCTION

It is well known that genetic code comprises of 64 codons, out of which 61 code for 20 standard amino acids and 3 codons act as terminating codons. Alternative codons within the identical group coding for the same amino acid are often termed synonymous codons. The redundancy of the genetic code, in which most of the amino acids can be translated by more than one codon, embodies a crucial step in curbing the efficiency and accuracy of protein production while maintaining the same amino acid sequence of the protein. Then again, the unequal usage of synonymous codons during translation of a gene to protein is stated as codon usage bias [1, 2]. It has been studied in a wide range of organisms [3, 4]. Studies on codon usage have revealed several factors that could influence codon usage patterns, including natural or translational selection, mutational pressure, secondary protein structure, replication, hydrophobicity and hydrophilicity of the protein and the external environment. Among these, the foremost factors accountable for codon usage bias among diverse organisms are considered to be compositional constraints under mutational pressure and natural selection [3, 5].

Male sexual hormones and male gametes are produced by the testis which includes all the courses involved in the production of gametes and the enzymatic reactions leading to the production of male steroid hormones [6]. The complex gene expression in the testis can be explained by the complicated processes of spermatogenesis and steroidogenesis [7, 8]. Genes crucial for spermatogenesis and male fertility are often exclusively expressed in male germ cells and are called testis-specific genes [9].

Goal

Knowledge of the codon usage in testis-specific genes not only exposes information about molecular evolution but also improves our understanding of the regulation of gene expression and the design of synthetic gene. In the current study, we report the detailed codon usage data and analysis of various factors shaping the codon usage patterns in testis-specific genes.

MATERIALS AND METHODS

Coding sequence data

Using accession numbers, testis-specific genes were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/). Only those coding sequences (cds) were considered for analysis, which are exact multiples of three bases with a proper start and stops codon. The accession numbers of different genes are given in table 1.

Table 1: Accession number of testis-specific genes

CDS	Accession No	Genes
CDS 1	NM_003140	Homo sapiens sex-determining region Y (SRY), mRNA, complete cds
CDS 2	>M98525	Homo sapiens testicular protein (TSPY) mRNA, complete cds
CDS 3	>U58096	Human testis-specific protein (TSPY) mRNA, complete cds
CDS 4	>U21663	Homo sapiens DAZ protein (DAZ) mRNA, complete cds
CDS 5	>AF000988	Homo sapiens testis-specific PTP-BL Related Y protein (PRY) mRNA, complete cds
CDS 6	>AF000997	Homo sapiens testis-specific XK Related Y (XKRY) mRNA, complete cds
CDS 7	>AF000979	Homo sapiens testis-specific Basic Protein Y 1 (BPY1) mRNA, complete cds

Indices of codon usage bias

Relative synonymous codon usage (RSCU) was calculated for the 59 synonymous codons for understanding the pattern of codon usage in testis-specific genes. RSCU>1.6 indicated that codons were over-represented while the RSCU values>1.0 indicated that the codon is more frequently used [10]. The formula used to estimate RSCU is as follows

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

Where, X $_{ij}$ is the frequency of occurrence of the jth codon for ith 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 ith amino acid (ith codon family).

The codon adaptation index (CAI) was used to estimate the extent of gene expression of a single gene. The CAI value ranged between 0 and 1.0, and high value of CAI indicates high gene expression [11]. The CAI is calculated as

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

Where ωk is the relative addictiveness of the kth codon, and L is the number of synonymous codons in the gene.

The codon bias index (CBI) measures the extent to which preferred codons are used in a gene. The formula used to calculate CBI is as follows

$$\mathrm{CBI} = \frac{\mathrm{N_{opt}} - \mathrm{N_{ran}}}{\mathrm{N_{tot}} - \mathrm{N_{ran}}}$$

Where N$_{opt}$ is the number of preferred optimal codons, N$_{tot}$ is the total number of codons, and N$_{ran}$ is the expected number of optimal codons if random codon assignments were made for each amino acid [12].

GRAVY (Grand Average of Hydropathicity) values are the sum of the hydropathy values of all the amino acids in the encoded protein of the gene divided by the number of residues in the sequence [13].

Aromo stands for aromaticity and refers to the frequency of aromatic amino acids (Phe, Tyr, Trp) in the translated gene product [14].

The frequency of overall A, T, G, C and their frequency at third codon position, overall GC content, and GC contents at first, second and third (GC1, GC2, GC3) positions were calculated using a perl script as developed by SC (corresponding author). GC3s was used as an excellent indicator for compositional constraint bias.

Analysis tools

Codon usage parameters and compositional dynamics were calculated (excluding the codons for Met, Trp, and the termination codons) using the Perl script developed by corresponding author SC. We used correspondence analysis (COA) which is a multivariate statistical analysis used to analyse the variation in codon usage pattern using XLSTAT. Correspondence analysis uses RSCU values, and its axes 1 and 2 contribute to total variation. Correlation and regression analysis were carried out by using the multi-analysis software SPSS 21.0.

RESULTS

Nucleotide composition

Codon usage bias can be mainly influenced by the general nucleotide composition of the genomes [15]. Therefore, we have first analyzed the nucleotide composition of testis specific genes (fig. 1), the mean G% was the highest, followed by the similar composition of A% and C%, with the T% being the lowest. This appears to advocate that there might be unequal distribution of A, T, G, and C nucleotides among codons of testis specific genes, with potentially more preference towards G-ended codons followed by A/C-ended codons. The overall GC% was 52.13 *i.e.* gene is GC rich.

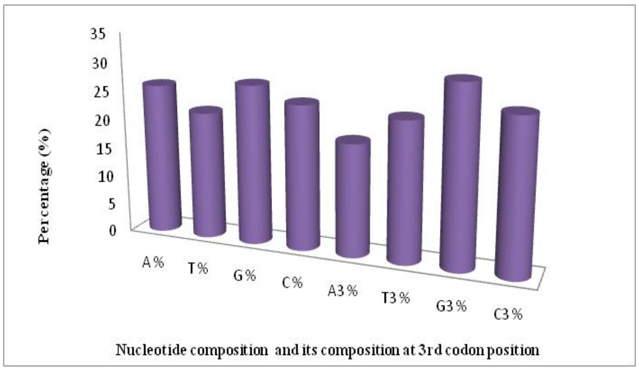

Fig. 1: Distribution of nucleotides

However, overall nucleotide composition that could influence the codon usage preference in testis specific genes emerged from the analysis of the nucleotide composition of the third position of codons namely A3, T3, G3, C3 (fig. 1) and of GC1, GC2, GC3. The mean G3 and C3 were the highest, followed by T3 and A3. From the table 2, it was found that GC3 % was the highest, followed by GC1% and, with the GC2 % being the lowest supporting the result of Butt *et. al.* [16]. Therefore, from the initial nucleotide composition analysis, it is expected that G/C-ended codons might be preferred over A/T-ended codons in testis specific genes.

Table 2: Overall GC content with GC content at 1st, 2nd and 3rd codon position

GC%	GC1 %	GC2 %	GC3 %
50.24	48.8	42.4	59.5
58.25	62.5	40.1	72.2
57.0	61.8	39.4	69.7
44.1	52.3	44.1	35.7
49.6	50	45.9	52.7
40.0	39.4	36.3	44.4
65.9	65.9	60.3	71.4
52.131429	54.385714	44.071429	57.94285714

Codon usage pattern in testis-specific genes

The CBI values of these testis-specific genes were lower, varying from 0.16 to 0.57 with a mean value of 0.2742. This result indicates that codon usage bias is weak in testis-specific genes and is maintained at a constant level.

To understand the pattern of synonymous codon usage in testis-specific genes, relative synonymous codon usage (RSCU) of individual codons was compared among the seven coding sequences. RSCU value zero means that the particular codon is absent, RSCU<0.06 represents the codons which are under-represented, and RSCU>1 represents the codons that are used more frequently than expected as shown in fig 2.

Correspondence analysis (COA)

To investigate RSCU variation, COA was performed using the seven testis-specific genes as a single dataset. The distribution of genes on the COA axis was used to identify the source of the variation among a set of multivariate data points. A major trend in the first axis (F1') accounted for 39.78 % of total synonymous codon usage variation, and the second major trend in the second axis (F2') accounted for 20.10 % of the total variation as shown in fig. 3.

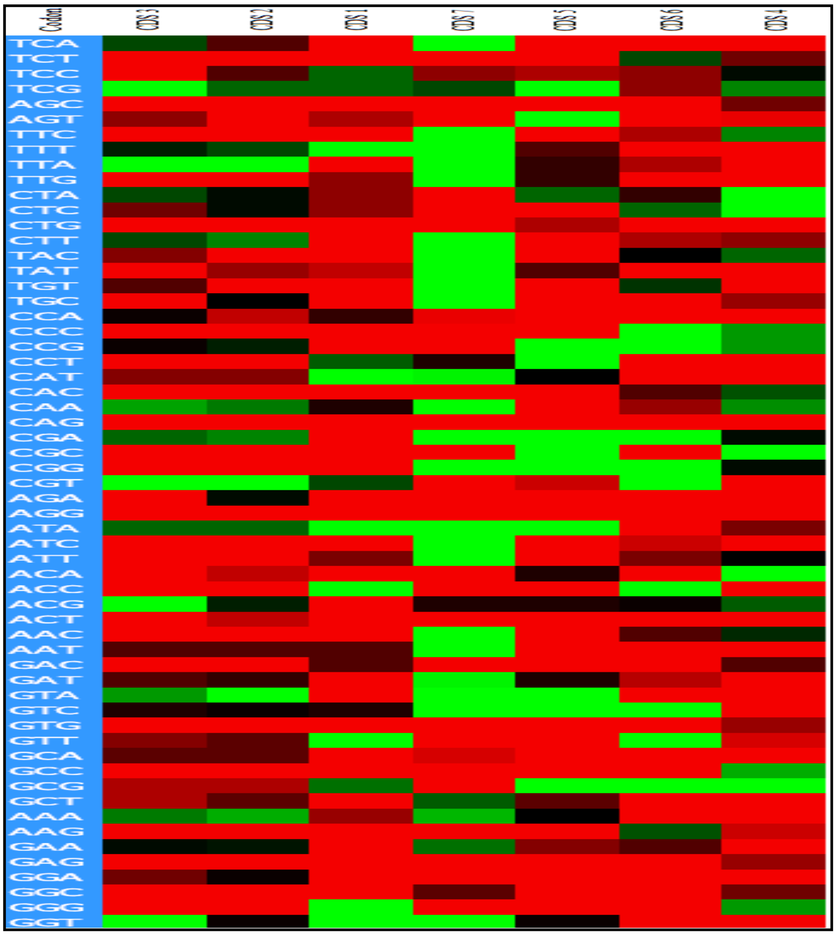

Fig. 2: UPGMA hierarchal clustering of RSCU values of each codon among the 7 testis-specific genes using Eulidean distance metric. Each square on the self-organizing map represents the RSCU value of a codon (shown in columns) corresponding to the CDS (shown in rows). The color coding varies from green to red with low to high values of the RSCU respectively. Green means RSCU value zero; dark green means RSCU value<0.06, dark red means RSCU value<1 and red indicate RSCU value>1

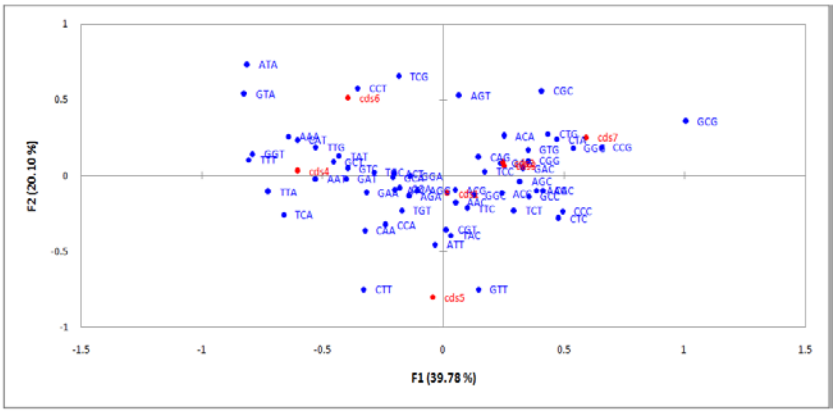

Fig. 3: Correspondence analysis of testis-specific genes

Effect of mutational bias on codon usage variation

To investigate whether the evolution of codon usage bias in testis-specific genes had been determined by mutation pressure alone or whether natural selection also has contributed to it, we first compared the correlation between nucleotide composition (A%, T%, G%, C%, GC%) and nucleotide composition at the third codon position (A3%, T3%, G3%, C3 %, GC3%) using the Pearson's correlation analysis method (table 3). A significant positive correlation was observed between A% and A3%, GC% and GC3% and significant negative correlation was observed for most of the heterogeneous nucleotide comparisons.

Table 3: Summary of correlation analysis between nucleotide constraints in testis specific genes

	A3 %	T3 %	G3 %	C3 %	GC3 %
A %	0.774*	0.314	-0.508	-0.323	-0.541
T %	0.688	0.894**	-0.905**	-0.406	-0.844*
G %	-0.916**	-0.877**	0.936**	0.557	0.943**
C %	-0.531	-0.742	0.812*	0.168*	0.673
GC%	-0.829*	-0.874*	0.941**	0.449	0.897**

* indicates p<0.05, ** indicates p<0.01

Further, correlation analysis was performed among the first two principal axes (F1’ and F2’) of COA and A3%, T3%, G3%, C3%, GC, GC1, GC2 and GC3% (table 3). The first principal axis (F1’) exhibited a significant positive correlation with G3%, C3%, GC%, GC1, GC3 and a negative correlation with A3%, T3%. It was interesting to note that, the second principle axis (F2’) had no correlation with any nucleotide content. Taken together, it is evident that mutation pressure might influence the codon usage pattern in testis-specific genes.

Effect of natural selection in shaping the codon usage patterns in testis-specific genes

It has been suggested that when codon bias is affected by mutational pressure alone, then the frequency of nucleotides, A and T should be equal to that of C and G at the third codon position [19]. However, in the case of testis-specific genes, firstly, variations in nucleotide base compositions were noted (fig. 1) indicating that other factors, such as natural selection, could also influence overall synonymous codon bias. Secondly, a significant positive correlation between GC% and T 3% and no correlation between C% and C3% and thirdly, first principal axis showed significant positive correlation with aromaticity as shown in table 4. It suggested that natural selection might have played a significant role in shaping the codon usage pattern in testis-specific genes supporting the result of Butt *et al* [16].

Table 4: Summary of correlation between the first two principle axes and nucleotide constraints

	A3	T3	G3	C3	GC	GC1	GC2	GC3	CAI	Aromaticity	Gravy Score
F1 r	-0.805*	-0.960**	0.903**	0.648	0.950**	0.779*	0.543	0.956**	-0.607	0.930**	-0.252
P	0.029	0.001	0.005	0.116	0.001	0.039	0.207	0.001	0.148	0.002	0.586
F2 r	0.000	-0.057	0.247	-0.409	0.007	0.023	-0.094	0.052	-0.056	-0.152	0.079
P	1.000	0.904	0.594	0.362	0.989	0.961	0.840	0.912	0.905	0.744	0.867

* indicates p<0.05, ** indicates p<0.001

DISCUSSION

Codon usage bias analysis is an established technique for genetic information and evolutionary relationships. The whole genome sequencing of many organisms is gaining the attention of researchers to study the pattern of codon usage [20]. The studies of codon usage pattern are helpful for better understanding of evolution, mRNA translation, the design of transgene, new gene discovery and many more [21-22]. Analyzing the codon usage patterns in testis-specific genes is likely to give a better understanding of the characteristics and molecular evolution of these genes.

Here we have analyzed the synonymous codon usage pattern in different testis-specific genes in human. The average G % was the highest, and the T% was the lowest. The nucleobase G at the 3rd position was the highest and A the lowest which support the result of Butt *et. al*. The overall GC content was higher than AT content, *i.e.* testis-specific genes are GC rich.

CBI values of these testis-specific genes are lower, which indicate that codon usage bias is weak in testis-specific genes and is apparently maintained at a stable level. Hoda *et. al,* also reported the same result, *i.e.* codon usage bias in human albumin superfamily is low [23].

The heat map shows the over-represented and the under-represented codons and the different CDS having different over-represented and the under-represented codons as revealed from RSCU analysis. From nucleotide composition and RSCU analysis, it is clear that selection of preferred codons has been mainly influenced by compositional constraints, under mutational pressure. Although, compositional constraints may not be the only cause connected to the pattern of codon usage in testis-specific genes, the overall RSCU values could reveal the codon usage pattern for the genes, and these may hide the codon usage difference among various genes supporting the result of Behura *et. al.* [17].

Correspondence analysis shows the variation of codon usage bias among testis-specific genes. Most of the codons are more close to the axes, indicating that the base composition for mutation bias might correlate to the codon bias. Some codons are in detached distribution which suggests other factors such as natural selection might affect the codon usage pattern supporting the result of Wei *et. al.* [18].

Correlation of overall nucleotide composition and its composition at 3rd position suggest that compositional constraints under mutation pressure and natural selection determine the codon usage pattern for testis-specific genes. Highly significant correlation between axes and nucleobase, further suggest that mutation pressure and natural selection both influenced the codon usage pattern of testis-specific genes supporting the result of Butt *et. al.*[16].

CONCLUSION

Codon usage bias in testis-specific genes is low, and the genes are GC rich. The over-represented and under-represented codons are different in different genes, so the pattern of codon usage is different among human testis-specific genes. Both mutation pressure and natural selection affect the codon usage pattern of these genes.

ACKNOWLEDGEMENT

We are thankful to the Department of Biotechnology, Assam University, Silchar for providing the lab facility to carry out this research work. This work was not funded by any funding agency like DBT or DST, Government of India, Delhi.

CONFLICT OF INTERESTS

The authors declare no conflict of interests in this work

REFERENCES

1. Grantham R, Gautier C, Gouy M, Mercier R, Pave A. Codon catalog usage and the genome hypothesis. Nucleic Acids Res 1980;8:49–62.
2. Marin A, Bertranpetit J, Oliver JL, Medina JR. Variation in G+C-content and codon choice: differences among synonymous codon groups in vertebrate genes. Nucleic Acids Res 1989;17:6181–9.
3. Gu W, Zhou T, Ma J, Sun X, Lu Z. Analysis of synonymous codon usage in SARS coronavirus and other viruses in the Nidovirales. Virus Res 2004;101:155–61.
4. Liu YS, Zhou JH, Chen HT, Ma LN, Pejsak Z. The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. Infect Genet Evol 2011;11:1168–73.

5. Sharp PM, Li WH. Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons. Nucleic Acids Res 1986;14:7737–49.
6. Swerdloff RS, Wang C, Bhasin S. Developments in the control of the testicular function. Baillieres Clin Endocrinol Metab 1992;6:451-83.
7. Dufau ML, Tsai-Morris C, Tang Khanum A. Regulation of steroidogenic enzymes and a novel testicular RNA helicase. J Steroid Biochem Mol Biol 2001;76:187-97.
8. Walker WH. Molecular mechanisms of testosterone action in spermatogenesis. Steroids 2009;74:602-7.
9. Liu FJ, Jin SH, Li N, Liu X, Wang HY, Li JY. Comparative and functional analysis of testis-specific genes**.** Biol Pharm Bull 2011;34:28-35.
10. Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1986;14:7749.
11. Wright F. The effective number of codons' used in a gene. Gene 1990;87:23.
12. Carbone A, Zinovyev A, Képès F. Codon adaptation index, as a measure of dominating codon bias. Bioinformatics 2003;19:2005-15.
13. Sur S, Sen A, Bothra AK. Mutational drift prevails over translational efficiency in Frankia nif operons. Indian J Biotechnol 2007;6:321-8.
14. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of protein. J Mol Biol 1982;157:105-32.
15. Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res 2003;92:1–7.
16. 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]
17. 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]
18. Wei L, He J, Jia X, Qi Q, Liang Z. Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution. BMC Evol Biol 2014;14:262.
19. Aramouni M, Segalés J, Sibila M, Martin-Valls GE, Nieto D. Torque teno sus virus 1 and 2 viral loads in postweaning multisystemic wasting syndrome (PMWS) and porcine dermatitis and nephropathy syndrome (PDNS) affected pigs. Vet Microbiol 2011;153:377-81.
20. Powell JR, Sezzi E, Moriyama EN, Gleason JM, Caccone A. Analysis of a shift in codon usage in *Drosophila*. J Mol Evol 2003;57:214–25.
21. Wright F. The effective number of codons used in a gene. Gene 1990;87:23–9.
22. 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.
23. Hoda Mirsafian, Adiratna MP, Singh A, Hwan PT, Merican AF, Mohamad SB. A comparative analysis of synonymous codon usage bias pattern in human albumin superfamily. Scientific World Journal 2014. Doi.org/10.1155/2014/639682. [Article in Press].