Research Paper

Volume : 3 | Issue : 9 | September 2014 • ISSN No 2277 - 8179

Biotechnology

KEYWORDS : Codon usage, Synonymous codons, Indian one-horned rhino

Molecular Biology of Codon Usage in One-Horned Rhino (*Rhinoceros unicornis* L)

*** SUPRIYO CHAKRABORTY** — Associate Professor, Department of Biotechnology, Assam University, Silchar 788011, Assam, India * Corresponding Author

ARIF UDDIN — Research Scholar, Department of Biotechnology, Assam University, Silchar 788011, Assam, India

ABSTRACT

Background: This is the first report on the codon usage pattern in Indian one-horned rhino (Rhinoceros unicornis L). No literature is available on codon usage bias in Indian one-horned rhino till to date. Indian one-horned rhino is the pride of Assam, being the state animal of Assam. The phenomenon of unequal use of synonymous codons i.e. some codons are more preferred than others in genes is known as codon usage bias. Synonymous codons which vary from two to six codons encode a particular amino acid and these generally differ in the 3rd codon position. However, the usage pattern of synonymous codons differs not only among the organisms but also among the genes within an organism. The codon bias is commonly found in highly expressed genes.

*Result and Discussion: In this study, we analyzed the codon usage pattern in some selected nuclear genes of one-horned rhino using bioinformatics approach. We found that the overall GC% ranged from 41.3 to 57.2 % with an average of 51.46 %. The GC3 % of rhino genes ranged from 53.3 to 73.6 % with an average of 65.2 %. ENC value varied from 50 to 60 with a mean value of 55. We found a significant negative correlation of ENC with GC (r=-0.934**, p< 0.01) and ENC with GC3 (r=-0.979**, p<0.01). In addition, a significant positive correlation was found between GC12 and GC3 (r=0.896*, p<0.05) of the genes. Further we observed significant positive correlation between A and A3 % (r=0.964** p<0.01), C and C3 % (r=0.951**, p<0.001), GC and GC3% (r=0.959**, p<0.001).*

Conclusion: We found that the codon usage bias in the analyzed coding sequences (cds) of rhino is not remarkable. Codon usage bias prefers GC bases at the 3rd codon position and most frequent codons end with G. Nucleotide constraint as well as compositional constraint under mutation pressure might have made major contribution in the codon usage bias in the coding sequences of the nuclear genes in one-horned rhino.

Introduction

Indian one-horned rhino (Rhinoceros unicornis L) is a highly endangered species belonging to the Schedule-I of animals of the Indian wildlife (Wildlife Protection Act, 1972) and it is widely distributed in the tall, wet grassland and riverside forests, foot hills of the Himalayas, particularly confined to the protected areas in Assam, West Bengal and Uttar Pradesh within India [1]. Rhinoceros unicornis is the pride of Assam and it is also the state animal of Assam. Chromosome number of rhino is 2n=82 or 84 in which the male is heterogametic with XY allosomes but the female is homogametic with XX chromosomes [2]. For establishing the relationship between Rhinoceros unicornis and other mammals, the mitochondrial DNA sequences were used [3].

The sequence of three nitrogen bases which codes for a particular amino acid is the basis of genetic code. Using the genetic code biological information is transferred from DNA to protein. Sixty one codons of the genetic code encode 20 standard amino acids but only three codons act as stop codons. Due to degeneracy of the genetic code, some amino acids are encoded by more than one codon. The level of degeneracy varies from two to six and the synonymous codons generally differ in the 3rd position (wobble position). Some amino acids are encoded by only one codon (met, trp). Synonymous codons are those which code for the same amino acid. The phenomenon of non uniform usage of codons in which some codons are more preferred than others is the codon usage bias. It is mainly caused by mutation and translational selection [4][5][6][7]. Optimal codons are the more frequently used codons whereas non-optimal codons are the less frequently used codons. Non optimal codons usually correspond to less abundant tRNA molecules in the cell than optimal codons do [8][9][10][11]12] and the pause of translational machinery there [13]. For efficiency in translation process, the optimal codons may be selected for and are used frequently in highly expressed genes [14][15]. Several factors which affect the codon usage bias in an organism are gene expression level, composition bias (%GC content and GC-skew), gene length, RNA stability and recombination rate [16][17][18]. The objective of the present study is to analyze the codon usage bias in the available nuclear genes of one-horned rhino as no literature is available on this aspect in rhino.

Table 1. Correlation coefficient between overall nucleotide and the nucleotide at 3rd codon position in rhino

Nucleotide (%)	A3 %	T3 %	G3 %	C3 %	GC3 %
A %	0.964**	0.417	-0.832*	-0.688	-0.846*
T %	0.308	0.975**	0.651	-0.723	-0.753
G %	-0.825*	-0.603	-0.970**	0.546	0.865*
C %	-0.759	-0.706	0.689	0.951**	0.884*
GC %	-0.868*	-0.719	0.908*	0.824*	0.959*

*** p=0.05, ** p=0.01, respectively**

Materials and Methods

Sequence data

The complete coding sequences (CDS) of six genes (only these are available in database) encoding the proteins (viz. estrogen receptor alpha, estrogen receptor beta, CD28, HFE protein, gamma-interferon and interleukin-4) in Indian one-horned rhino (Rhinoceros unicornis) were retrieved from the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/Genbank/). The cds sequences having exact multiple of three bases are considered for this analysis using the standard genetic code.

Compositional properties

Overall nucleotide composition (A, C, T and G%) and nucleotide composition at the wobble position of each codon (A3, C3, T3 and G3%) were analyzed for one-horned rhino coding sequences using an in-house perl script developed by S. Chakraborty. The GC and GC3 indices referred to the overall GC content and that at the wobble position of synonymous codons in the cds sequences (excluding the codons encoding single amino acids met and trp and the three codons responsible for polypeptide termination).

Measures of synonymous codon usage bias

A couple of indices for measuring the codon usage bias in genes have been proposed by several workers. Some of the most widely

used measures for measuring the synonymous codon usage bias are discussed here.

Relative Synonymous Codon Usage (RSCU)

Relative synonymous codon usage (RSCU) is analysed as the ratio of the observed frequency of a codon to the expected frequency if all the synonymous codons of a particular amino acid are used equally. If the RSCU value of a codon is greater than 1.0, it indicates that the particular codon is used more frequently than the expected frequency whereas the reverse is true for RSCU values less than 1.0 [14].

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

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).

Effective Number of Codons (ENC)

The effective number of codons used by a gene (ENC) is generally used to measure the codon usage bias of synonymous codons [19]. The value of ENC ranges from 20 (for each amino acid, when only one codon is used) to 61 (when all synonymous codons for an amino are randomly used). If the calculated ENC value is greater than 61 (because codon usage is more evenly distributed than expected), it is adjusted to 61.

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

Codon Adaptation Index (CAI)

The codon adaptation index (CAI) [28] is the most extensively used measure of codon bias in prokaryotes [20][21][22] and in eukaryotes [23][24][25]. CAI is widely used as a measure of gene expression. CAI is based on relative adaptiveness of codons. The relative adaptiveness (ω) of a codon is the ratio of the usage of each codon divided by the most abundant codon within the same synonymous family. CAI value ranges from 0 to 1, with higher values indicating a higher proportion of the most abundant codons [26]. The analysis for codon usage bias was done by using an in-house perl script developed by S. Chakraborty.

Statistical analysis

Correlation analysis was performed to identify the relationship between overall nucleotide composition and each base at 3rd codon position and correlation of ENC with GC, GC3 and CAI. All the statistical analyses were done using the SPSS software.

Results

Compositional properties

In the coding sequences of genes in one-horned rhino, it was found that the nucleotides, A and C occurred more frequently than G and T in the overall composition. The base C occurred most frequently (average C_3 %=33.77) and A occurred least frequently at the 3rd codon position (average A_3 %=16.20) as shown in Fig 1. The nucleotide composition in general and that at 3rd codon position for nuclear genes in one-horned rhino suggests that compositional constraints might be influencing the pattern of codon usage bias in the coding sequences of rhino. The overall GC % of rhino nuclear genes ranged from 41.3 to 57.2 % with an average of 51.46 % but the $GC_{3\,\%}$ of nuclear genes ranged from 53.3 to 73.6 % with an average value of 65.2%. These differences could be attributed to the increased number of C at 3rd codon position in all the cds.

Table 2. Correlation coefficients between parameters of codon usage bias in rhino

No.	Correlation between parameters	Correlation coefficient value
1	ENC and GC	-0.934**
2	ENC and GC3	-0.979**
3	GC1 and GC3	0.817*
4	GC2 and GC3	0.799
5	GC12 and GC3	0.896*

*** p=0.05, ** p=0.01, respectively**

ENC values of rhino cds varied from 50 to 60 with a mean of 55 and were usually of higher magnitude. This indicates that the codon usage bias is possibly not remarkable in rhino and is maintained apparently at a stable level.

Codon Usage in Rhino

The overall RSCU values for 59 codons in rhino cds indicate that G and C occurred most frequently at the 3rd codon position as shown in Fig. 3. Fourteen codons end with C and these codons are TCC, AGC, TTC, CTC, TAC, TGC, CCC, CAC, ATC, ACC, AAC, GAC, GCC, GGC encoding the amino acid serine, serine, phenylalanine, leucine, tyrosine, cysteine, proline, histidine, isoleucine, threonine, asparagine, aspartate, alanine and glycine, respectively. Seven codons end with G and these are CTG, CAG, CGG, AGG, GTG, AAG, GAG encoding leucine, glutamine, arginine, arginine, valine, lysine and glutamate, respectively. Four codons ending with A nucleotide are CCA, AGA, GCA, GGA which encode the amino acid proline, arginine, Alanine and glycine, respectively. Codon ending in T is ATT which encodes the amino acid isoleucine.

For the rhino cds under study, we plotted the ENC value against overall GC, GC_3 content and found significant negative correlation of ENC with GC (r= -0.934**, p< 0.01), ENC with GC_3 (r= -0.979**, p<0.01). In addition, we observed a significant positive correlation between GC_1 and GC_3 (r=0.817* p<0.05) and between GC_{12} and GC_3 (r=0.896* p<0.05). These results suggest that nucleotide constraint under mutation pressure is the major factor for codon usage bias in these coding sequences of one-horned rhino. No significant correlation between ENC and CAI was observed in the present study.

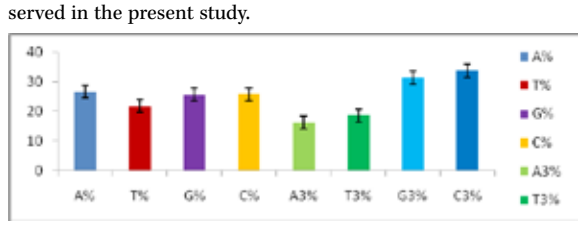

Fig. 1 Distribution of overall nucleotides (A, T, G, C%) and nucleotides at 3rd codon position (A3, T3, G3, C3%) in coding sequences of rhino

Effect of Mutation pressure on codon usage bias

To test the hypothesis that mutation pressure has no role in the evolution of codon usage bias in rhino, we compared the correlation coefficient between overall nucleotide composition A, T, G, and C% and its composition at 3rd codon position A_3, T_3, G_3, and C_3% using Karl Pearson's correlation coefficient. Significant positive correlation was found between A and A_3 (r=0.964**, p<0.01), C and C_3 (r=0.951**, p<0.001), GC and GC_3 (r=0.959**, p<0.001) whereas significant negative correlation was observed for other nucleotides. This suggests that compositional constraint under mutation pressure might affect the codon usage pattern in rhino. However, significant positive correlation was found between T and T_3 (r=0.975**, p<0.001) but significant negative correlation between G and G_3 (-0.970**, p<0.001). Moreover, no significant correlation was found between G and C_3 (r=0.546,

Volume : 3 | Issue : 9 | September 2014 • ISSN No 2277 - 8179

p>0.05) and between GC and T3(r=-0.719, p>0.05) suggesting that natural selection probably did not play a major role in codon usage pattern in rhino.

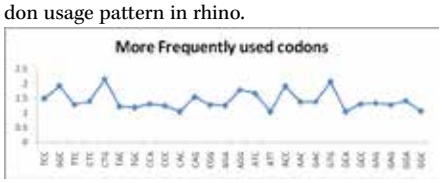

Fig. 2 Codons used more frequently than expected (RSCU > 1) in rhino

Discussion

The whole genome sequencing of many organisms and the easily accessible DNA data in publicly available nucleotide data bases like NCBI, EMBL and DDBJ attracted the attention of researchers to study the codon usage bias in several organisms. Synonymous codons are used non-uniformly during translation. So it is important to understand the role of codon usage bias in translation. Analysis of codon usage bias has amply contributed to the better understanding of molecular evolution of genes.

Genetic analysis of endangered species such as Indian one-horned rhino is useful not only for acquiring knowledge on the biology of animals but also for enhancing the conservation strategies [27][28]. Different types of DNA repeats are found in higher eukaryotic genomes [29]. For the development of desired genetic marker, information about the organization of satellite DNA and VNTR loci of species and subspecies of rhino would be relevant [30].

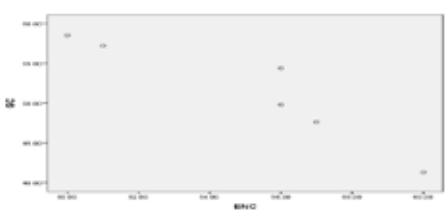

Fig. 3 Scatter diagram of ENC and GC values in gene coding sequences of rhino

We analyzed the pattern of codon usage bias in the nuclear genes of Rhinoceros unicornis L. Most frequent codons had C or G at its 3rd codon position. Codon usage bias in the analyzed cds is not remarkable as the ENC values are high. From this study, the overall nucleotide composition and the composition at 3rd codon position are shown in Table 1. Our analysis suggests that nucleotide composition in rhino could be under mutation pressure. Highly significant correlation between GC_1 and GC_3 and between GC12 and GC_3 suggest that mutation pressure is an important factor in shaping the codon usage in this species. Further, highly significant correlation between ENC with GC and ENC with GC_3 support the hypothesis that compositional constraint under mutation pressure has influenced the codon usage bias. From our earlier work on mitochondrial DNA of mammals, we have found the same result but the only difference is that mitochondrial DNA in mammals prefers AT richness [31].

Conclusion

The present study is the first endeavor on the codon usage analysis in Indian one-horned rhino (Rhinoceros unicornis L). We have found that codon usage bias in rhino is not very remarkable. Codon usage bias prefers G or C nucleotide at the 3rd position of codons. Most frequent codons in the cds end with C at 3rd codon position. Our results support the fact that both nucleotide constraint and compositional constraint might have played a major role in the codon usage bias in the nuclear genes of Indian one-horned rhino.

Acknowledgement:

We are thankful to Assam University, Silchar for providing the necessary facilities for this research work. This work has not been funded by any external funding agency.

Abbreviations:

Coding sequence (CDS), Effective number of codon (ENC), Mitochondrial DNA (mt DNA)

Conflict of Interest:

We do not have any conflict of interest.

REFERENCE

1. Ali S, Azfer MA, Bashamboo A, Mathur PK, Malik PK, Mathur VB, Raha AK , Ansari S. 1999. Characterization of a species-specific repetitive DNA from a highly endangered wild animal, Rhinoceros unicornis, and assessment of genetic polymorphism by microsatellite associated sequence amplification (MASA). Gene 228: 33–42. | 2. Wurster DH, Benirschke K. 1968. The chromosomes of the Great Indian Rhinoceros (Rhinoceros unicornis L). Experimentia 24: 5–11. | 3. Xu X, Janke A, Arnason U.1996. The complete mtDNA sequence of the greater Indian rhinoceros, Rhinoceros unicornis, the phylogenetic relationship among Carnivora, Perissodac- tyla, and Artiodactyla (+Cetacea). Mol Biol Evol 13: 1167– 1173. | 4. Bulmer M. 1988. Are codon usage patterns in unicellular organisms determined by selection- mutation balance? J Evol Biol 1: 15–26. | 5. Sharp PM, Stenico M, Peden JF, Lloyd AT.1993. Codon usage: mutational bias, translational selection, or both? Biochem Soc Trans 21: 835–841. | 6. Kliman RM, Hey J. 2003. Hill-Robertson interference in Drosophila melanogaster: reply to Marais, Mouchiroud and Duret. Genet Res 81: 89–90. | 7. Akashi H. 1997. Codon bias evolution in Drosophila. Population genetics of mutation- selection drift. Gene 205: 269–278. | 8. Ikemura T. 1981. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J Mol Biol 151: 389–409. | 9. Ikemura T. 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. Mol Biol 158: 573–597. | 10. Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2: 13–34. | 11. Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. Nature 325: 728– 730. | 12. Akashi H. 2001. Gene expression and molecular evolution. Curr Opin Genet Dev 11: 660– 666. | 13. Kurland CG. 1992. Translational accuracy and the fitness of bacteria. Ann Rev Genet 26: 29–50. | 14. Sharp PM, Li WH. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 24: 28–38. | 15. Shields DC, Sharp PM. 1987. Synonymous codon usage in Bacillus subtilis reflects both translational selection and mutational biases. Nucleic Acids Res.15: 8023–8040. | 16. Powell JR, Moriyama EN. 1997. Evolution of codon usage bias in Drosophila. Proc Natl Acad Sci USA. 94: 7784–7790. | 17. Moriyama EN, Powell JR. 1998. Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. Nucleic Acids Res 26: 3188– 3193. | 18. Powell JR, Sezzi E, Moriyama EN, Gleason JM, Caccone A. 2003. Analysis of a shift in codon usage in Drosophila. J Mol Evol 57 Suppl. 1: S214–225. | 19. Wright F. 1990. The effective number of codons used in a gene. Gene 87: 23–29. | 20. Eyre-Walker A, Bulmer M. 1993. Reduced synonymous substitution rate at the start of Enterobacterial genes. Nucleic Acids Res 21: 4599-4603. | 21. Gutierrez GJ, Casadesus J, Oliver L, Marin A. 1994. Compositional heterogeneity of the Escherichia coli genome - a role for VSP repair. J Mol Evol 3: 340-346. | 22. Perriere G, Gouy M, Gojobori T. 1994. NRSUB a nonredundant database for the Bacillus subtilis genome. Nucleic Acids Res 22: 5525-5529. | 23. Akashi H. 1994. Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. Genetics 1363: 927–935. | 24. Frohlich DR, Wells MA. 1994. Codon usage patterns among genes for Lepidopteran hemolymph proteins. J Mol Evol 38: 476-481. | 25. Morton BR. 1994. Codon use and the rate of divergence of land plant chloroplast genes. Mol Biol Evol 11: 231-238. | 26. Sharp PM, Li WH. 1987. The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15: 1281-1295. | 27. Ali S, Hasnain SE. 1999. Conservation, management and documentation of livestock genetics resources and biodiversity: a biotechnology perspective. Curr Sci 77: 1154-1158. | 28. Ali S, Ansary S, Ehtesham NZ, Azfer MA, Homkar U, Gopal R, Hasnain S.E.1998. Analysis of the evolutionary conserved repeat motifs in the genome of the highly endangered central Indian swamp deer. Gene 223: 361-367. | 29. Cox R, Mirkin S.M. 1997. Characteristics enrichment of DNA repeats in different genomes. Proc Natl Acad Sci USA 94: 5237-5242. | 30. Ali S, Muller CR, Epplen JT. 1986. DNA fingerprinting by oligonucleotide probes specific for simple repeats. Hum Genet 74: 239-242. | 31. Uddin A, Chakraborty S. 2014. Analysis of codon usage bias in mitochondrial atpase6 in some mammals. Int J Recent Sci Res 5(4): 883-888. |