Gene 586 (2016) 105–114

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Codon usage trend in mitochondrial CYB gene

Arif Uddin, Supriyo Chakraborty *

Departments of Biotechnology, Assam University, Silchar 788011, Assam, India

ARTICLE INFO

Article history:
Received 4 August 2015
Received in revised form 11 March 2016
Accepted 2 April 2016
Available online 6 April 2016

Keywords:
Codon usage
MT-CYB gene
Mutation pressure
Natural selection
Pisces
Aves
Mammals

ABSTRACT

Here we reported the pattern of codon usage and the factors which influenced the codon usage pattern in mitochondrial cytochrome B (MT-CYB) gene among pisces, aves and mammals. The F1 axis of correspondence analysis showed highly significant positive correlation with nucleobases A3, C and C3 and significant negative correlation with T and T3 while F2 of correspondence analysis showed significant positive correlation with C and C3 and significant negative correlation with A and A3. From the neutrality plot, it was evident that the GC12 was influenced by mutation pressure and natural selection with a ratio of 0.10/0.90 = 0.11 in pisces, 0.024/0.976 = 0.0245 in aves and in mammals 0.215/0.785 = 0.273, which indicated that the role of natural selection was more than mutation pressure on structuring the bases at the first and second codon positions. Natural selection played the major role; but compositional constraint and mutation pressure also played a significant role in codon usage pattern. Analysis of codon usage pattern has contributed to the better understanding of the mechanism of distribution of codons and the evolution of MT-CYB gene.

1. Introduction

Synonymous codons encoding a particular amino acid are not used with equal frequency regardless of the degeneracy of the genetic code due to a phenomenon known as codon bias (Ikemura, 1981). It is an evolutionary relic. The frequencies of codon usage are found to be species specific and also specific across genomes and within the same genome. Its role is significant to understand the evolution of genome (Jenkins and Holmes, 2003). Codon usage pattern is affected by various factors such as compositional bias (GC% and GC skew), mutation pressure, natural selection, gene length, expression level, replication, RNA stability, hydrophobicity and hydrophilicity of the protein (Akashi, 1997; Moriyama and Powell, 1998; Powell and Moriyama, 1997; Powell et al., 2003). Among these, the compositional constraints in the presence of mutation pressure and natural selection are the major factors which vary across species (Sharp et al., 1986, 1993). The modifications of biochemical mechanism *i.e.* more frequent changes of certain bases than others cause mutational biases (Francino and Ochman, 2001; Green et al., 2003). Mutation pressure is mainly responsible for codon usage pattern in some prokaryotes and in many mammals with high AT or GC contents (Sharp et al., 1993; Zhao et al., 2007). However, in *Drosophila* and in some plants, the codon usage pattern is mainly governed by translational selection (Liu et al., 2004). The nonsynonymous substitution is driven by selection because it alters the amino acids and thus affects protein's biochemical nature (Plotkin and Kudla, 2011).

In fast-growing organisms with huge population size, the codon usage pattern is mainly driven by selection (Green et al., 2003; Ikemura, 1982, 1985; Sharp and Li, 1987). However, the effect of natural selection in codon usage in the mammalian genome is considered to be low (Duret, 2002; Sharp et al., 1995). This is due to small population size in many mammalian species, and the codon usage pattern is due to the effect of genetic drift (Keightley et al., 2005; Sharp et al., 1995). But with the exception, in non mammalian species highly expressed genes with high codon usage bias are under selection pressure to diminish the error in expression level (Hershberg and Petrov, 2008). Essentially, the efficiency of gene expression is due to the redundancy of genetic code tuned by selective forces (Gingold and Pilpel, 2011). Moreover, codon usage declines the proofreading expenses by reducing the time and energy required to discard the non-cognate tRNAs (Bulmer, 1991). Use of unpreferred codons would increase proofreading expenses and would result in a net decline in the protein levels.

The association between the codon bias and the level of gene expression has been experimentally established in *Esherichia coli* (Andersson and Kurland, 1990). Moreover, the *in-vitro* expression proficiency has been shown to be significantly increased by using the preferred codons of the host cell in heterologous genes in cultured eukaryotic cells (Kim et al., 1997).

The mitochondrial genes are a subset of the frequently expressed genes in eukaryotes, and its genome is usually ideal as the molecular marker for species identification, systematic phylogeny, and evolutionary studies. The choice of mitochondrial DNA in all these studies is primarily

Abbreviations: MT-CYB, mitochondrial cytochrome b gene; RSCU, relative synonymous codon usage; CAI, codon adaptation index; ENC, effective number of codon.

* Corresponding author.
E-mail addresses: arif.uddin29@gmail.com (A. Uddin), supriyoch_2008@rediffmail.com (S. Chakraborty).

http://dx.doi.org/10.1016/j.gene.2016.04.005
0378-1119/

due to its small size, easy amplification and conserved gene content, lack of recombination, maternal inheritance pattern and high evolutionary rate (Harrison, 1989). The metazoan mitochondrial DNA is circular, 16 kb in size, covalently closed and consists of 37 genes. The majority of the mitochondrial proteins are encoded by nuclear genes while only 2 rRNA, 22 tRNAs and 13 proteins involved in the respiratory chain are encoded by mitochondrial genomes (Wolstenholme, 1992). Further the genetic code of mitochondria differs from that of standard genetic code. The standard genetic code consists of 64 codons, wherein 61 sense codons encode 20 standard amino acids and three codons namely TAA, TAG and TGA act as termination signal. In mitochondrial genetic code there are four termination codons such as TAA, TAG, AGA and AGG. Patterns of codon usage in nuclear genomes are extensively studied whereas studies on mitochondrial genomes or genes are very scanty.

The oxidative phosphorylation is one of the most important biochemical processes operating in the mitochondria in which the aerobic eukaryotic cell uses oxygen to synthesize ATP. The 13 protein-coding genes of mitochondria are universal and encode for the protein subunits of the different complexes of oxidative phosphorylation. The oxidative phosphorylation is the multienzymatic system which creates the proton gradient required for ATP synthesis (or heat generation). The seven mitochondrially encoded proteins (Nd1, Nd2, Nd3, Nd4, Nd4l, Nd5, Nd6) form the complex I in which Nd1 and Nd2 play an essential structural role between the membrane-embedded and peripheral arms of the complex whereas the role of Nd2, Nd4, and Nd5 is to transfer electron (da Fonseca et al., 2008). The nuclear encoded proteins form the complex II while complex III is formed by CYB gene, the only mitochondrial protein-coding gene.

Mitochondrial cytochrome B (MT-CYB) gene contains more conservative as well as rapidly evolving codon position and variable region and so this gene is also used in systematics (Meyer and Wilson, 1990; Moritz et al., 1992). The catalytic core of the complex III of electron transport chain is formed by CYB protein along with cytochrome c1 and it helps in the assembly and function of the complex. Moreover, the pattern of codon usage in MT-CYB gene is yet to be reported despite being a vital protein of complex III, coded by mitochondrial DNA. In addition, many reports have indicated the phylogenetic usefulness of MT-CYB gene in different vertebrates (Degli Esposti et al., 1993; Irwin et al., 1991).

The mitochondrial respiratory chain plays a central role in satisfying the energy demand of an organism. It will be interesting to analyze the pattern of codon usage in MT-CYB gene among pisces, aves and mammals residing at different habitats with diverse energy requirements. The pisces, aves and mammals are the three classes of chordates, which live in three entirely different environments namely aquatic, aerial and terrestrial. The mode of respiration and the demand of energy in these chordates are also different (Ellington, 2001). Therefore, understanding the patterns of synonymous codon usage in three chordates would improve our understanding of the mechanisms underlying the distribution of codons and their differential usage in MT-CYB gene and would elucidate the factors affecting the codon usage pattern.

Analysis of codon usage is a useful technique to understand the genetic and evolutionary relationship of different species belonging to diverse habitats. Moreover, mitochondrial genes are very significant and suitable tools for such studies. In the current study, we investigated the codon usage pattern in MT-CYB gene among 15 species each of pisces, aves and mammals thriving in different habitats to understand the pattern of codon usage. Moreover, this study would give insight in to the factors influencing the codon usage pattern among the species under study.

2. Materials and methods

2.1. Sequence data

The coding sequences of MT-CYB gene for 15 different species each of pisces, aves and mammals (in FASTA format) were retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/). The study was carried out on 45 species. The sequences having correct start and stop codons with an exact multiple of three bases were used in this analysis. The details of the accession numbers of MT-CYB gene from 45 species of pisces, aves and mammals were shown in S1.

2.2. Compositional properties

The compositional properties of MT-CYB gene such as overall nucleotide composition (A%, C%, T% and G%), nucleotide composition at the third position of each codon (A3%, C3%, T3% and G3%), overall GC content and GC content at the 1st, 2nd and 3rd codon positions, AT, GC, purine, pyrimidine, amino and keto skew were analyzed for pisces, aves and mammals using a perl script developed by SC (corresponding author).

2.3. Measures of synonymous codon usage bias

Some of the most important and widely used indices of the codon usage bias that were analyzed in this study are discussed below.

2.3.1. Relative synonymous codon usage (RSCU)

Relative synonymous codon usage (RSCU) is the observed frequency of a codon to the expected frequency if all synonymous codons of a particular amino acid are used evenly. RSCU value >1.0 indicates that the corresponding codons are used more frequently than the expected frequency whereas the RSCU values <1.0 indicate that the particular codons were used less frequently. Besides, the RSCU value >1.6 was treated as over represented codon while RSCU value <0.6 was treated as under-represented codon (Behura and Severson, 2012; Sharp and Li, 1986a).

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

where, X_{ij} is the frequency of occurrence of the jth codon for the ith amino acid and n_i is the number of codons for the ith amino acid (ith codon family).

2.3.2. Effective number of codons (ENC)

The effective number of codons (ENC) is the commonly used parameter to measure the usage bias of synonymous codons (Wright, 1990). The ENC value ranges from 20 (when only one codon is used for each amino acid) to 61 (when all codons are used randomly). A higher ENC value means low codon usage bias and vice-versa. ENC values <35 are generally considered as the significant codon usage bias. The ENC is measured as

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

where s represents the given (G + C) 3% value (Wright, 1990)

2.3.3. Codon adaptation index (CAI)

The codon adaptation index (CAI) is a very extensively used parameter to measure the codon usage bias and the gene expression level. Its value ranges from 0 to 1; with high value indicating a higher proportion of the most abundant codons coupled with high expression level and vice-versa. CAI is a measure of the relative adaptedness of the codon usage of a gene to the codon usage of the highly expressed genes