Mitochondrion xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Mitochondrion

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

Codon usage and expression level of human mitochondrial 13 protein coding genes across six continents

Supriyo Chakraborty[a,*], Arif Uddin[b,*], Tarikul Huda Mazumder[a], Monisha Nath Choudhury[a], Arup Kumar Malakar[a], Prosenjit Paul[a], Binata Halder[a], Himangshu Deka[a], Gulshana Akthar Mazumder[a], Riazul Ahmed Barbhuiya[a], Masuk Ahmed Barbhuiya[a], Warepam Jesmi Devi[a]

[a] Department of Biotechnology, Assam University, Silchar, Assam 788001, India
[b] Department of Zoology, Moinul Hoque Choudhury memorial Science College, Algapur, Hailakandi, Assam 788150, India

ARTICLE INFO

Keywords:
Codon usage bias
Evolutionary relationships
Mitochondrial genes
Mutation pressure
Natural selection

ABSTRACT

The study of codon usage coupled with phylogenetic analysis is an important tool to understand the genetic and evolutionary relationship of a gene. The 13 protein coding genes of human mitochondria are involved in electron transport chain for the generation of energy currency (ATP). However, no work has yet been reported on the codon usage of the mitochondrial protein coding genes across six continents. To understand the patterns of codon usage in mitochondrial genes across six different continents, we used bioinformatic analyses to analyze the protein coding genes. The codon usage bias was low as revealed from high ENC value. Correlation between codon usage and GC3 suggested that all the codons ending with G/C were positively correlated with GC3 but vice versa for A/T ending codons with the exception of *ND4L* and *ND5* genes. Neutrality plot revealed that for the genes *ATP6, COI, COIII, CYB, ND4* and *ND4L*, natural selection might have played a major role while mutation pressure might have played a dominant role in the codon usage bias of *ATP8, COII, ND1, ND2, ND3, ND5* and *ND6* genes. Phylogenetic analysis indicated that evolutionary relationships in each of 13 protein coding genes of human mitochondria were different across six continents and further suggested that geographical distance was an important factor for the origin and evolution of 13 protein coding genes of human mitochondria.

1. Introduction

A codon is a set of three nucleotides that encode a specific amino acid residue in a polypeptide chain or for the termination of translation process. Due to degeneracy of the genetic code, a single amino acid is encoded by more than one codon except for two amino acids viz. methionine and tryptophan in standard genetic code. These codons are known as synonymous codons for an amino acid. The synonymous codons are not used with equal frequencies in the mature mRNA and this unequal relative codon usage frequency leads to codon bias. Thus, codon usage bias is the phenomenon involving non-uniform usage of synonymous codons encoding the same amino acid during the translation of mRNA to protein (Behura and Severson, 2012). Codon bias is a unique property of the genome of an organism and species specific but may vary significantly among the genes within the same organism (Grantham et al., 1980; Marin et al., 1989; Prat et al., 2009). It is well evident that the variation in codon usage within the same synonymous codon family dictates the translational efficiency of a gene, thus having a great impact in shaping genome evolution (Bentele et al., 2013). The increased translational efficiency of genes helps organisms adapt to the changing conditions and is found related to the lifestyle of the organisms (Botzman and Margalit, 2011). Previously, many studies have reported synonymous codon usage bias in different organisms from simple prokaryotes to higher eukaryotes. It was further reported that many genomic factors such as gene length, GC content, gene expression level, structure of mRNA and its stability mediate the codon usage bias in different organisms (Chen et al., 2004; Mazumder and Chakraborty, 2015). However, in general two major factors namely mutational pressure and weak natural selection are involved in guiding the codon usage bias in different organisms (Butt et al., 2014; Hershberg and Petrov, 2008).

The codon bias study has acquired renewed attention of the scientific community with the inception of whole genome sequencing in different organism. It acquires significance in molecular biology for understanding the patterns of synonymous codon usage, analysing the level of gene expression, genome characterization and also for

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

https://doi.org/10.1016/j.mito.2017.11.006
Received 14 July 2016; Received in revised form 9 October 2017; Accepted 27 November 2017

Table 1
Accession number of human mitochondrial genomes used for retrieval of complete coding sequences of 13 genes and their lengths (nucleotides).

Continents		Africa	Asia	Europe	N. America	L. America	Australia
Accession no.		KC533522.1	AP008824.1	DQ112955.3	KF874379.1	KM656464.1	DQ112752.3
Gene Length	ATPase6	786	786	783	786	786	783
	ATPase8	681	681	681	681	681	681
	COI	684	684	684	684	684	684
	COII	207	207	207	207	207	207
	COII	348	348	348	348	348	348
	NADH1	1044	957	1044	1044	1044	1044
	NADH2	1542	1044	1542	1542	1542	1542
	NADH3	297	348	297	297	297	297
	NADH4	1812	1380	1812	1812	1812	1812
	NADH4L	1380	297	1380	1380	1380	1380
	NADH5	525	1812	525	525	525	525
	NADH6	1143	525	1137	1143	1143	1137
	CYB	1143	1143	1137	1143	1143	1137

exploring the selective forces that shape their evolution in an organism (Mazumder and Chakraborty, 2015).

Mitochondria are cell organelles found in all nucleated cells. It is called the power house of the cell since it generates ATP, the 'energy currency' of the cell (Schon et al., 2012). A mitochondrion contains two membranes i.e. outer membrane and inner membrane. The human mitochondrial genome is 16.6 kb in size, and contains only 37 genes, of which 2 are ribosomal RNAs (rRNAs), 22 tRNAs and 13 protein-coding genes. The protein-coding genes are involved in different complexes of oxidative phosphorylation (Schon et al., 2012). In oxidative phosphorylation, there are five complex systems in which complex I contains *ND* genes (*ND1, ND2, ND3, ND4, ND4l, ND5* and *ND6*); complex III contains *CO* genes (*COI, COII* and *COIII*), complex IV contains *CYB* gene and complex V contains *ATP* genes (*ATP6* and *ATP8*) (Weber, 2001).

Mitochondria contain their own genetic material, the mitochondrial DNA (mtDNA), located in the mitochondrial matrix (Schon et al., 2012). The mtDNA is maternally inherited and usually occurs in multiple copies within a cell (Harrison, 1989). Two theories were proposed to explain the endosymbiotic origin of mitochondria in eukaryotes. One theory states that the eukaryote engulfed the mitochondrion (Embley and Martin, 2006), whereas the other suggests that the prokaryote host obtained the mitochondrion. Previous studies suggested that mtDNA undergoes neutral evolution (Wise et al., 1998). However, later findings reported that both positive and negative selection acted on mtDNA (Rand et al., 1994). The mtDNA illustrates a higher rate of nucleotide substitution than that of the gDNA (Pesole et al., 1999).

The frequency of mutations in mtDNA is approximately tenfold greater than that in nuclear DNA (Grossman and Shoubridge, 1996). Mutations in mtDNA are associated with several diseases, including Leber's hereditary optic neuropathy (LHON) (Wallace et al., 1988), Leigh syndrome (Santorelli et al., 1993), mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS) (Koo et al., 1993), myoclonic epilepsy and ragged red fibres (MERRF) (Berkovic et al., 1989), and cancer (Higuchi, 2007). The common mtDNA-related disorder is Leber's hereditary optic neuropathy (LHON) in which mutation occurs in one of the three genes ND1, ND4 and ND6. The disease causes sub acute loss of central vision in young adults (Wallace et al., 1988). The Leigh syndrome is caused by the mutation in subunits of complex I or also due to assembly factors of complex IV. It is characterized by delay in developmental respiratory problems, ataxia, dystonia etc. (Santorelli et al., 1993).

This study focuses on the comparative analysis of codon usage bias and codon context patterns of thirteen protein coding genes of human mitochondria among six continents. Since codon bias study in human mitochondrial genes is very rare, little information is available on the codon usage bias of human mitochondrial protein-coding genes and cancer as reported by Uddin (2016b). Very interestingly, no detailed study across continents on the comparative analysis of codon usage bias of human mitochondrial genes has been reported till now, so this study will certainly provide novel insights into the codon usage pattern, translational efficiency of gene expression as well as the evolutionary significance of thirteen protein coding genes of human mitochondria representing six continents.

2. Methods

2.1. Retrieval of data

Complete nucleotide coding sequences of thirteen protein coding genes of human mitochondria from six continents of the world namely Africa, Asia, Europe, North-America, Latin-America and Australia were retrieved from National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) GenBank database. The Antarctica was excluded from the present analysis since no data is available. In our analysis of codon usage bias we selected only those coding sequences (cds) which are exact multiple of three bases having perfect start and stop codon without any unknown base (N) in the entire sequence. The genes with proper accession number that are available for each region of the world in NCBI database were considered for the present study and are listed in a tabular form (Table 1).

2.2. Nucleotide constraint analysis

The frequency of G and C at the third position of codon is a good indicator of the degree of base composition bias (Zhou et al., 2005). Hence the frequencies of each of the nucleotides (A, T, G, C) and that of its 3rd codon position in the coding sequences of each human mitochondrial gene in six continents of the world were analyzed and the occurrence of overall GC% at different codon positions namely the first (P1), second (P2) and third position (P3) were calculated to study the relationship between codon usage variation and base compositional constraints. Furthermore, to explore the nucleotide bias, GC skews and AT were calculated as $GC_{skew} = \frac{G-C}{G+C}$ and $AT_{skew} = \frac{A-T}{A+T}$ (Tillier and Collins, 2000).

2.3. Effective number of codons (ENC)

The ENC is usually used to measure the codon usage bias of a gene and is independent of the gene length and number of amino acids (Wright, 1990). The value of ENC ranges from 20 to 61. For a gene in which only one codon is used for each amino acid out of the whole synonymous codon family, ENC value would be 20, meaning that the particular codon is highly biased in usage. When all codons of synonymous codon families are uniformly used the ENC value would be 61, meaning no biasness exists for codon usage. The ENC value closer to 20 indicates a strong codon usage bias in the gene and these biased genes are generally highly expressed (Wright, 1990).