Molecular & Biochemical Parasitology 219 (2018) 24–32

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

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

Analysis of codon usage pattern of mitochondrial protein-coding genes in different hookworms

Bornali Deb[a], Arif Uddin[b,*], Gulshana Akthar Mazumder[a], Supriyo Chakraborty[a,*]

[a] Department of Biotechnology, Assam University, Silchar 788011, Assam, India
[b] Department of Zoology, Moinul Hoque Choudhury Memorial Science College, Algapur, Hailakandi 788150, Assam, India

ARTICLE INFO

Keywords:
Codon usage bias
GC content
AT content
Mutation pressure
Natural selection

ABSTRACT

The phenomenon of unequal usage of synonymous codons encoding an amino acid in which some codons are more preferred to others is the codon usage bias (CUB) and it is species specific. Analysis of CUB helps in understanding evolution at molecular level and acquires significance in mRNA translation, design of transgenes and new gene discovery. In our current study, we analyzed synonymous codon usage pattern and the factors influencing it on mitochondrial protein coding genes of 6 different hookworms *i.e. Ancylostoma ceylanicum, Ancylostoma duodenale, Necator americanus, Ancylostoma tubaeforme, Ancylostoma caninum* and *Uncinaria sanguinis* as no work was reported yet. The effective number of codons for mitochondrial genes suggested that codon usage bias was high in most species. The GC content was lower than AT content *i.e.* genes were AT rich as indicated by nucleotide composition analysis. The overall nucleotide composition along with its composition at 3rd codon position and correspondence analysis suggested that both natural selection and mutation pressure might have affected the codon usage bias in mitochondrial genes. However, neutrality plot revealed that mutation pressure might have played a major role in *A. ceylanicum* while natural selection might have played the dominant role in *Ancylostoma duodenale, Necator americanus, Ancylostoma tubaeforme, Ancylostoma caninum* and *Uncinaria sanguinis.*

1. Introduction

It is well known that 61 codons out of 64 codons encode 20 standard amino acids in standard genetic code. But mitochondrial genetic code of nemathelminths follows translation table 14 of National Center for Biotechnology Information (NCBI). It consists of 63 sense codons and only one stop codon (TAG). The degeneracy level of serine is 8 with the codons TCT, TCC, TCA, TCG, AGT, AGC, AGA and AGG. The amino acid leucine has six codons namely CTA, CTC, CTG, CTT, TTA and TTG while arginine has four codons *viz.* CGA, CGC, CGG and CGT. In this genetic code, tryptophan has two codons *i.e.* TGG and TGA. But two amino acids namely methionine and lysine are encoded by single codon each.The different codons that encode the same amino acid are called synonymous codons. Previous studies reported that the non uniform usage of synonymous codons in which some codons are used more preferentially than others is due to a feature known as codon usage bias (CUB) [1,2]. Earlier investigations suggested that CUB varies within the genome as well as between the genomes and it may help in understanding the genome evolution among related species [3].

Several hypotheses have been proposed to elucidate the origin of CUB. Amongst these hypotheses, the neutral theory [4] and the selection-mutation-drift balance model [5] are very important. According to neutral theory, mutations at degenerate codon positions should be selectively neutral which would result in non-uniform preference of synonymous codons of an amino acid and hence failure of natural selection pressure. According to selection-mutation-drift model, CUB is assumed to be balanced by mutation pressure, genetic drift and weak selection. On the other hand, if the gene expression level is high indicating high selective pressure, then it may result from stronger CUB [6]. However, with the completion of whole genome projects of many organisms, these two hypotheses are no longer enough to explain the observed CUB. Several important factors have been reported to affect the CUB such as compositional bias (GC% and GC skew), mutation pressure, natural selection, expression level, gene length, RNA stability, replication, hydrophobicity and hydrophilicity of the protein [7–10]. Of these, the compositional constraints in the presence of natural selection and mutation pressure are considered as the major factors which may vary across species [5,6,11]. The study of codon usage is essential to predict and optimize protein expression levels, to recognize protein coding genes, and to detect lateral gene transfer [12]. In the case of

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

https://doi.org/10.1016/j.molbiopara.2017.11.005
Received 14 May 2017; Received in revised form 14 November 2017; Accepted 16 November 2017
Available online 20 November 2017
0166-6851/

lateral gene transfer, CUB analysis also helps in prediction of gene expression level, elucidation of over-represented and under-represented codons in genes [13].

Many previous studies on codon usage in several viruses have suggested mutational pressure as a major force in shaping codon usage patterns [14]. Information about molecular evolution can be revealed with the study of codon usage bias in viruses; it can also improve our understanding of the regulation of gene expression and would aid vaccine design [15]. From the study it was also found that the codon usage bias in CHIKV is slightly biased, and the mutational pressure was the major factor that contributed to shape the codon usage pattern [16]. Jia et al. [17] reported the CUB of *B. mori* and found that the codon bias was relatively weak, and the influencing factors were nucleotide composition, mutation pressure, natural selection and expression level [17]. Of these, natural selection was considered to be the major factor [17]. Also the nucleobases G and C were found to encode the optimal codons of *B. mori*, which provides useful information for enhancing the gene expression in *B. mori* through codon optimization [17]. The study of codon usage in parasites suggested that the pattern of codon usage differed significantly between *W. bancrofti* and *S. haematobium,* two major human parasites [18]. Various factors such as nucleotide composition, natural selection and mutation pressure were found to affect the codon usage pattern in *W. bancrofti* and *S. haematobium* [18].

Various parameters have been used to study the CUB. RSCU (relative synonymous codon usage), a parameter of CUB, is the number of times a codon appears in a gene to its expected occurrence. The RSCU values of codons will be equal to 1, if the synonymous codons of an amino acid are used with equal frequencies [19]. The effective number of codons (ENC) is used to quantify how much the codon usage of a gene departs from equal usage of synonymous codons. Low ENC value means high codon usage bias and vice-versa [20].

Hookworms are blood-sucking parasitic nematodes infecting over 740 million people in developing countries [21] with almost 135,000 death rates annually [22]. It sustains itself in the host's digestive system and burrows into its intestinal wall and thus into the victim's bloodstream, causing mental retardation and growth deficiencies [23,24].

Ancylostoma ceylanicum is a hookworm which produces potent infections to humans only leading to heavy infection with concurrent anemia [25]. *Ancylostoma duodenale* and *Necator americanus* are primarily two hookworms infecting humans [26] by the oral as well as the percutaneous route [27] causing great blood loss and iron deficiency [28]. *Ancylostoma tubaeforme* is endemic to domestic cats. It penetrates skin at third-stage of larvae [29] and causes mild enteritis, but in case of heavy infections it can lead to anemia which might be fatal [30]. *Ancylostoma caninum*, is a pathogenic gastrointestinal parasite of dogs [31], possessing significant risk in public health too, causing dermatitis and eosinophilic enteritis [32]. *Uncinaria sanguinis* is an haematophagous parasitic nematode of the small intestine associated with reduced growth rate, anaemia, and mortality of pups in several otariid species [33,34].

Analysis of CUB of mitochondrial protein coding genes of different hookworms would help us in understanding the process of molecular evolution and the factors influencing it. In mitochondria, the oxidative phosphorylation is the process in which oxygen is used by the eukaryotic cells to synthesize ATP. The 13 mitochondrial protein-coding genes encode for protein subunits of the different complexes of oxidative phosphorylation. The complex I is found to be formed by seven mitochondrially encoded proteins *viz*., ND1, ND2, ND3, ND4, ND4L, ND5, ND6 [35]. In mitochondria, the nuclear encoded proteins form the complex II while CYB protein forms the complex III. CO proteins form the complex IV whereas ATPase proteins form the complex V of respiratory chain.

Analysis of codon usage bias promotes the understanding of the genetic and evolutionary relationship of a species, and in this case mitochondrial genes are considered to be the most ideal tool for evolutionary studies. Here, we have investigated the compositional properties and the pattern of codon usage bias in mitochondrial genes of 6 different hookworms to understand the molecular biology and mechanism of codon distribution as no work was reported yet. This study also sheds insight into the evolutionary forces which influence the codon usage bias of the mitochondrial protein coding genes.

Table 1
Accession number of mitochondrial genomes of different hookworm species.

Sl. No.	Species	Accession No.	References
1	*Ancylostoma ceylanicum*	AP017674.1	Bradbury et al. [36]
2	*Ancylostoma duodenale*	NC_003415.1	Inpankaew et al. [37]
3	*Necator americanus*	NC_003416.2	Xie et al. [38]
4	*Ancylostoma tubaeforme*	NC_034289.1	Burrows [39]
5	*Ancylostom acaninum*	FJ483518.1	Jex et al. [40]
6	*Uncinaria sanguinis*	NC_025267.1	Xie et al. [38]

2. Materials and methodology

2.1. Sequence data access

The mitochondrial protein coding genes of six hookworms were retrieved from the National Centre for Biotechnology Information, USA (http://www.ncbi.nlm.nih.gov/). All coding sequences (cds) used in this analysis had correct start and stop codons with exact multiple of three bases (Table 1). The analysis was based on translation table 14 of NCBI.

2.2. Nucleotide composition

Overall nucleotide composition (AC, T and G%) and its composition at the third codon position of each coding sequence (A3, C3, T3 and G3%) were analyzed. The GC, GC1, GC2 and GC3% were estimated for each coding sequence using a perl script.

2.3. Relative synonymous codon usage

Relative synonymous codon usage (RSCU) is the ratio of observed codon usage frequency of a gene divided by the expected codon usage frequency in the synonymous family given that all the codons for the particular amino acid are used equally. If the synonymous codons of an amino acid are used equally, their RSCU values will equal 1. If the RSCU value of a codon is < 1.0, it indicates that the corresponding codon is used less frequently than the expected frequency and vice versa. Moreover, if the RSCU value of a codon is > 1.6, it reveals that the particular codon is over represented while the RSCU value < 0.6 indicates that the codon is underrepresented in the coding sequence [41].

$$\mathrm{RSCU_{ij}} = \frac{X_{ij}}{\frac{1}{ni}\sum_{j=1}^{ni} Xij}$$

Here, X_{ij} is the number of 'ith' codon for the 'jth' amino acid, and 'ni' is the total number of synonymous codons that encode the 'jth' amino acid.

2.4. Effective number of codons (ENC)

The effective number of codons (ENC) is a frequently used quantitative parameter for the measurement of the codon usage bias. It measures the extent how much the codon usage of a gene varies from the equal usage of synonymous codons. The ENC value ranges from 20 to 61.While a single codon for each particular amino acid is used in the cds, ENC value is 20. The ENC value increases while multiple codons for an amino acid are used in the cds. Higher ENC value signifies low codon usage bias and vice versa. ENC value less than 35 for a cds is generally considered as significant codon usage bias [20]. ENC value of a cds for