Acta Tropica 176 (2017) 29–33

Contents lists available at ScienceDirect

Acta Tropica

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

New insights into the factors affecting synonymous codon usage in human infecting *Plasmodium* species

Shivani Gajbhiye, P.K. Patra, Manoj Kumar Yadav*

Department of Biochemistry, Pt. J.N.M. Medical College, Raipur, CG 492001, India

ARTICLE INFO

Keywords:
Codon usage
Malaria
Compositional bias
Translational selection
Phylogeny

ABSTRACT

Codon usage bias is due to the non-random usage of synonymous codons for coding amino acids. The synonymous sites are under weak selection, and codon usage bias is maintained by the equilibrium in mutational bias, genetic drift and selection pressure. The differential codon usage choices are also relevant to human infecting *Plasmodium* species. Recently, *P. knowlesi* switches its natural host, long-tailed macaques, and starts infecting humans. This review focuses on the comparative analysis of codon usage choices among human infecting *P. falciparum* and *P. vivax* along with *P. knowlesi* species taking their coding sequence data. The variation in GC content, amino acid frequencies, effective number of codons and other factors plays a crucial role in determining synonymous codon choices. Within species codon choices are more similar for *P. vivax* and *P. knowlesi* in comparison with *P. falciparum* species. This study suggests that synonymous codon choice modulates the gene expression level, mRNA stability, ribosome speed, protein folding, translation efficiency and its accuracy in *Plasmodium* species, and provides a valuable information regarding the codon usage pattern to facilitate gene cloning as well as expression and transfection studies for malaria causing species.

1. Introduction

Triplet codons are the basic units for the encoding of mRNA in all biological organisms. Since the genetic code is degenerate, more than one codons encode most amino acids. Such codons are called synonymous codons, and they usually differ by one nucleotide at the third codon position (or for some amino acids, in the second position). The frequency of occurrence of synonymous codons is different for most amino acids. This phenomenon is known as codon usage bias (CUB). There is a bias in the usage of synonymous codons and this trend is observed for most genes and genomes (Andersson and Kurland, 1990; Grantham et al., 1980). One of the most representative theories to explain codon usage bias is selection-mutation-drift, proposing that codon usage bias is dependent upon a number of factors, such as natural selection (e.g., gene expression level, tRNA abundance, protein length, gene translation initiation signals and protein structure) and mutational pressure (e.g., GC content, mutation frequency and pattern), as well as random genetic drift (Andersson and Kurland, 1990; Bulmer, 1991; Grantham et al., 1980). Codon usage bias plays an important role in predicting the optimum host of exogenous genes and can be used to improve the expression levels of exogenous genes via codon optimization. Understanding the extent and causes of differences in codon usage pattern provides clues about how the evolution and environmental adaptation for different organisms (Cannarozzi and Schneider, 2012; Duret, 2002).

Our understanding of the basic features of the molecular organization of genomes is greatly enhanced by the genome-wide study of codon usage patterns. In general, composition biased mutation pressure and/or selection pressure for accurate and efficient translation in various organisms are the main reasons for this bias (Knight et al., 2001). Any similarities in the codon usage pattern identify some degree of biological relationship, environmental adaptation, and evolution among *Plasmodium* species (Yadav and Swati, 2012). More recent studies have revealed that patterns of codon usage bias and nucleotide composition within many cellular genomes are far more complex than previously imagined, and the factors shaping their evolution are still not entirely understood (Mitreva et al., 2006; Peixoto et al., 2004; Rashmi and Swati, 2013; Salim and Cavalcanti, 2008). The recent advancement in sequencing technologies, allow studying the codon usage behaviour of disease-causing apicomplexan protozoans. Among them, the study of different human infecting Plasmodium species on the basis of their genome and proteome is an important one as they are responsible for high morbidity and mortality cases in the tropical and subtropical regions of the world. *P. falciparum* and *P. vivax* is responsible for the most severe cases of malaria (Carlton et al., 2008). In recent years, sporadic cases of traveler's malaria, due to *P. knowlesi*, has been reported which

* Corresponding author.
E-mail address: manojiids@gmail.com (M.K. Yadav).

http://dx.doi.org/10.1016/j.actatropica.2017.07.025
Received 6 June 2017; Received in revised form 18 July 2017; Accepted 21 July 2017
Available online 24 July 2017
0001-706X/

normally infects long tailed macaques. So, a detailed study of their genomic pattern is needed to understand the pattern of disease. This review focuses on the detailed study of the genomic pattern of human infecting *P. falciparum*, *P. vivax*, and *P. knowlesi* in the light of their codon usage pattern.

2. Codon bias is influenced by GC content

The GC composition of an organism may be divided into three levels: Overall GC content, local GC content, and the ratio of G/C or A/T content. Variation in the GC composition at all the three levels within an organism's genome can affect the synonymous codon choices, and this correlation is a widely accepted phenomenon (Mondal et al., 2016; Palidwor et al., 2010; Zhou et al., 2014). The GC content of the genome intergenic region shows close resemblance with the GC content of their coding part (Palidwor et al., 2010). Due to the degeneracy of the genetic code, the third base is less discriminatory for the amino acid than the other two bases. This third position in the codon is referred to as the wobble position. Most of the amino acids allow substitution of synonymous codons at the third codon position (Yu and Li, 2011). Many prokaryotes and fungi have shown that the favored codons are closely related to the species intergenic GC content (Hershberg and Petrov, 2009). In many bacteria, prokaryotes and eukaryotes the codon usages, as well as amino acid usage are highly influenced by GC compositional bias (Brbić et al., 2016; Esposito et al., 2016; Rashmi and Swati, 2013). *P. falciparum* shows a lower percentage of GC, about 28% in composition, resulting in higher usage of AT rich codons for coding amino acids (Rao et al., 2011). The coding sequences of *P. vivax* and *P. knowlesi* show comparatively higher GC composition, about 45%, and 40% respectively. The G/C ending codon usage increases with increase in genomic GC bias. This trend is also seen in *P. vivax* where more GC ended codons are present at wobble positions while *P. falciparum* and *P. knowlesi* show preferences towards A/T ending codons (Table 1).

3. Codon usage bias and differential gene expression

Gene expression is a process by which the genetic information embedded in the DNA sequence is utilized to produce functional proteins. The genes may be divided into highly expressed genes and lowly expressed genes according to their expression level. Highly expressed genes are expressed more often, producing greater than average levels of protein compared to other genes (Peixoto et al., 2004). The selection in accuracy is required to prevent misfolding errors leading to loss of functional proteins. Highly expressed genes use more preferred codons than rest of the genes of the genome (Quax et al., 2015). Translation selection is one of the important factors responsible for shaping codon usage choices of highly expressed genes (Brule and Grayhack, 2017; Gout et al., 2010; Mondal et al., 2016; Presnyak et al., 2015).

'Effective number of codons' (ENc) is a measure of species-independent synonymous codon bias in genes (Fuglsang, 2004; Wright, 1990). The differential expression of protein coding genes can be directly or indirectly classified by their ENc values. There is an inverse relation between the gene expression and ENc values. The ENc value for highly expressed genes are less, and more for lowly expressed genes. The average ENc value for all protein coding genes is low for *P. falciparum* compared to that of *P. vivax* and *P. knowlesi* (Hardison, 2003; Šali et al., 1995; Yadav and Swati, 2012). This shows that on an average, the codon usage bias is more in highly expressed genes of *P. falciparum*.

Table 1
Comparison of genomic and coding sequence characteristics in *Plasmodium* species.

Species	Chromosomes	Genome size	CDS count	Genomic GC%	CDS GC%	CDS GC3%
P. falciparum	14	23.3	5818	19.41	28	17.3
P. vivax	14	27. 01	5610	42.3	45	48.15
P. knowlesi	14	23.46	5484	37.5	40	41.17

Table 2
Average value of ENc and CAI for *P. falciparum, P. vivax* and *P. knowlesi.*

Species	ENc_{ave}	CAI_{ave}
P. falciparum	39.06	0.15
P. vivax	53.72	0.187
P. knowlesi	54.25	0.186

Highly expressed genes show a tendency of high bias towards some codons and tend to use those codons frequently. Codon Adaptation Index (CAI) is another measure to identify the differential gene expression of genes using codon choices (Carbone et al., 2003; Supek, 2016). CAI assumes that the difference in gene expression is the result of translational selection pressure, due to which there is a difference in the usage of synonymous codons between highly expressed genes and rest of the genes in a genome. The low CAI_{ave} for *P. falciparum* shows that highly expressed genes are under more translation selection pressure for shaping their codon usage choices compared to other studied *Plasmodium* species (Table 2).

4. Synonymous codon choice affects the rate of translation

The differential rate of translation depends upon the concentration of tRNA, mainly in the interaction of the anticodon and wobble position of codon (Endres et al., 2015). The slower rates of translation are caused by codons, decoded by rare tRNA. This decreases the translation efficiency and accuracy, and results in the reduction of mRNA stability. In general, the optimized codons refer to the codons with highest tRNA copy numbers. The frequency of optimal codons in a gene can be used as an indicative measure to check if the codons are optimized for efficient translation (Lavner and Kotlar, 2005; Salim and Cavalcanti, 2008). The optimal codons are used more often than other codons causing ribosomes to move faster along mRNA and results in their elevated expression. Optimal codons not only used in higher proportion for coding essential proteins, also increase translation efficiency and accuracy (Brule and Grayhack, 2017). Highly expressed genes of *P. falciparum*, *P. vivax*, and *P. knowlesi* show higher usage of 26, 30 and 29 optimal codons respectively. These optimal codons are rich in A/T bases at the wobble position in all the three species irrespective of their genomic composition. The higher usage of A/T ending optimal codons in *P. falciparum* is the result of higher compositional bias while G/C ending optimal codon usage indicates translational selection pressure acting on them (Chanda et al., 2005). Comparatively GC-rich *P.vivax* genome shows higher usage of A/T at the third position of optimal codons in highly expressed genes indicates the role of translational selection pressure in shaping their codon usage. *P. knowlesi* and *P. falciparum* show a higher percentage of AT as compared to GC, this indicates codon usage bias towards AT rich codons, albeit there is a large difference in AT content of *P. knowlesi* and *P. falciparum*.

5. Synonymous codon choices regulate the protein folding

The genetic code has the ability to regulate the protein synthesis and its folding (Spencer, 2012). Protein folding is correlated with the translation rate (Tuller et al., 2010). Variations in synonymous codon choices confer the variable rate of polypeptide discloses from ribosome's, which may influence the folding capacity of proteins (Ran and Higgs, 2010). The translational rate is inversely proportional to folding efficiency of proteins. An increase in the rate of translation results in a decrease of folding efficiency of proteins (Yu et al., 2015). Rare codons play an important role in translation pause, and allow the folding of proteins efficiently (Komar et al., 1999; Shabalina et al., 2013). *Plasmodium* species uses a different strategy where codon preferences were