Advances in Biology & Earth Sciences
Vol.4, No.2, 2019, pp.107-118

THE MODIFICATION OF CODON USAGE DUE TO THE EVOLUTION OF EUKARYOTES

Konul Nuriyeva[1,2*]

[1]Department of Biophysics and Molecular Biology, Faculty of Biology
Baku State University, Baku, Azerbaijan
[2]Reference Laboratory for Toxicological and Chemical Analysis, Hygiene and
Epidemiology Center of Azerbaijan Republic, Baku, Azerbaijan

Abstract. In this study, we mainly focused the relationships among extracted gene sequences in order to analyze the factors affecting codon usage bias in eukaryotes in the course of evolutionary history. We aimed to investigate the variations in the choice of synonymous codons in family genes in order to identify codon usage bias among different biological species. The performed statistical analysis revealed that the codon usage patterns of family genes in different species are more conservative in comparison with gene families in same species. We concluded that, all eukaryotic genomes are largely conservative and the variation of occurrence of synonymous codons is barely noticeable on complete genome dataset whereas we observed a clear bias in choosing the synonymous codons by looking at the codon frequencies of various gene sequences and this bias becomes more trivial as a set of randomized gene sequences increases.

Keywords: *codon usage bias, synonymous codons, family genes,orthologous gene.*

*[*]**Corresponding Author:** Nuriyeva Konul, Department of Biophysics and Molecular Biology, Faculty of Biology, Baku State University, 23 Z. Khalilov Str., Baku, Azerbaijan, e-mail:* konuriyev@gmail.com

***Received**: 29 April 2019;* ***Accepted**: 20 May 2019;* ***Published**: 10 August 2019.*

1. Introduction

The main feature of genetic code is its redundancy. There are 64 codons and only 20 amino acids. The redundancy of genetic codes ensuresthe eighteen out twenty amino acids (except methionine and tryptophan) to be coded by multiple codons. The codons thatencode for the same amino acid are called synonymous. Although the choice between synonymous codons does not affect the sequence or function of the translated protein, it is well known that the use of alternative synonymous codons is a nonrandom process (Bennetzen & Hall, 1982; Bentele *et al*., 2013; Blakem *et al*., 2003). The term codon usage bias which represents the unequal usage of synonymous codons for encoding amino acids may vary significantly between genomes, between genes in the same genome, and within a single gene (Bentele *et al*., 2013; Blakem *et al*., 2003).

One strong factor which can be used to predict the codon bias between different species is the genomic GC content, the fraction of the two nucleotides guanine and cytosine in the genome (Plotkin *et al*., 2004). To date, the highest GC contents of land plants have been found in grasses (*Poaceae*) (Schnable *et al*., 2009; Šmarda *et al*., 2012; Flagel & Blackman, 2012). In contrast to grasses, the lowest GC contents so far reported in plants have been found in several species possessing holocentric chromosomes (i.e., in *Cyperaceae* and *Juncaceae*) (Lipnerová *et al*., 2012; Šmarda *et al*., 2012).

Variation in codon bias among genes from the same organism has been shown to depend on many parameters, including expression level (Grantham *et al*., 1981), amino acid composition (Collins & Jukes, 1993), gene length (Eyre-Walker, 1996), mRNA structure (Gambari *et al*., 1990; Huynen *et al*., 1992), and protein level noise considerations (Blake *et al*., 2003). It was observed that in *E. coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*, in a subset of the genes within each genome which are highly expressed, the strength and, for some amino acids, the choice of the abundant codon differs significantly from the rest of the genes (Bennetzen & Hall, 1982; Grantham *et al*., 1981).

Some deviations can also be observed regarding choosing synonymous codons in different positions of one gene sequences. Studies show a strong deviation from null hypothesis in synonymous codon substitutions in the beginning region of the genes in diverse organisms such as bacteria, yeast and fruit flies (Bentele *et al*., 2013; Qin *et al*., 2004; Tuller *et al*., 2010). When looking at this region we observe that there's a tendency for choosing the codons which are thought to be not recognized and translated at high speed.

Codon usage bias is explained from two points of view, mutational bias and natural selection (Harrison & Charlesworth, 2010; Ingvarsson, 2007). Even though there have been studies showing that mutational bias is a significant factor in shaping the codon bias (Comeron & Kreitman, 2002; Kanaya *et al*., 2001; Kudla *et al*., 2009) the fact that in almost all of the cases the preferred (most frequent) codon is the one with most abundant matching tRNA molecules, indicates that natural selection might play a role as well (Higgs & Ran, 2008; Kanaya *et al*., 2001). Contrary to some prokaryotes, particularly in organisms with extremely A + T-rich or G+C-rich genomes (e.g., *Mycoplasma capricolum*, *Micrococcus luteus* and *Streptomyces species*), and in mammals, mutation bias is the major determining factor in shaping the codon usage pattern (Francino & Ochman, 1999). Codon usage pattern in multicellular eukaryotes like *Drosophila* (Powell & Moriyama, 1997) and in some plants is influenced by translational selection (Chen *et al*., 2014).

2. Methodology

Codon adaptation index

Here we used CAI in order to compare the relative codon usage of a gene and to measure the similarities between synonymous codon usage of the gene sequences. The sequences were obtained from widely used genome database browser called Ensembl (https://asia.ensembl.org/index.html). The first step of calculating CAI is measuring the value of relative synonymous codon usage (RSCU) index. RSCU value is simply the observed frequency of this codon divided to excepted frequency of the synonymous codons for an amino acid

$$RSCU_{ij} = \frac{n_{ij}}{\frac{1}{N_i}\sum_{j=1}^{N_i} n_{ij}} \qquad (1)$$

where, ij is the occurrence of j codon in the i amino acid, N_i is the number of all synonymous codon, n_{ij} the number of observed j codons which codes for i amino acid.

In the next step, $RSCU$ is used to generate relative adaptiveness of observed codon. The relative adaptiveness of a codon, ω, is the frequency of use of that codon compared to the frequency of the optimal codon for that amino acid:

$$\omega_{ij} = \frac{RSCU_{ij}}{RSCU_{i\,max}} = \frac{N_{ij}}{N_{i\,max}} \tag{2}$$

$i\,max$ indicates most frequently used codon for i amino acid compared with other synonymous codons. Finally, codon adaptation index of a specific gene can be calculated as

$$CAI = exp\left(\frac{1}{L}\sum_{k=1}^{L} ln\omega_k\right) \tag{3}$$

L is the number of codons, and ω_k is the relative adaptiveness of the k^{th} codon in the gene sequence.

The value of CAI ranges between 0 and 1. Lower CAI value which close to 0 shows the synonymous codons are more randomly used and the bias between codons is low, while higher CAI value indicates that the codon usage bias is stronger between genes and gene sequences have more optimal codons.

Correlation analyses

Correlation analysis was used to identify the relationship between the patterns of synonymous codon usage in obtained gene sequences. R values range between -1 and $+1$; positive r values correspond to correlated gene pairs, negative values correspond to anti-correlated gene pairs and values close to zero to non-correlated.

Genome database browser - Ensembl

In this study, we used Ensembl - genome database browser (https://asia.ensembl.org/index.html) in order to obtain coding DNA sequences (CDS) of biological species to further examine codon usage bias related to relationship between two gene sequences.

3. Results and discussion

The relationship between same gene sequences of different species

In this research paper, we mainly focused one of the housekeeping gene family named histone genes that are essential for the existence of a cell. For this matter, we collected gene sequences of nine genetically distinct species – homo sapiens (human), pan troglodytes (chimpanzee), mus musculus (house mouse), gallus gallus (chicken), anolis carolinensisanole (green anole), xenopus tropicalis (western clawed frog), danio rerio (zebrafish), caenorhabditis elegans (nematode), saccharomyces cerevisiae (baker's yeast) - and the statistical analyses were performed among whole obtained gene sequences which comprise the main six housekeeping genes in order to indicate the codon frequencies in different genes in mentioned species. Correlation analyses were used here to measure how strong a relationship is between two variables.

Histones are a family of essential proteins that associate with DNA in the nucleus and help condense it into chromatin. The correlation values of five histone genes (H2AFV, H2AFX, H2AFY, H3F3A and H1F0) were clearly estimated. Each of this selected housekeeping histone gene can be defined as orthologous gene that retains the same function with their ancestral gene that they evolved from and every arranged table shown below indicate the correlation values among various organisms for the specific orthologous gene. It appears from the tables that the codon usage of the single orthologous gene varies enormously across organisms.

The correlation value increases with the similar codon usage pattern of two organisms. Genetically close species indicate strong positive correlation between each other. As the species move away from each other the choice of synonymous codons differs significantly for the same genes. In this sense, mammals showed relatively higher correlation compared to other groups of organisms. Whole histone genes showed higher correlation value between human and chimpanzee, even this value is equal to 1 in H3F3A gene (see Table 1). It implies that the codon frequencies of synonymous codons in H3F3A gene are same for human and chimpanzee.

Besides, one noticeable relationship can be seen in H3F3A gene (Table 1) is that the relationship of chicken with the rest of the species is relatively weak, where chicken only holds relatively close nucleotide coding sequences with fish compared with other selected species, which $r = 0{,}792$. This reveals that the choice of the synonymous codon in chicken differs significantly from the rest of the species and chicken did not conserve codon frequencies belonging to other species which diverged before and after chicken in the course of evolution. Similarly, yeast also obviously highly correlated with frog ($R = 0{,}628$), while evidently lower values were observed in the correlation of yeast with other species. The reason of such relationship is unknown, but according to all these results it can be assumed that both chicken or frog tried to keep their nucleotide composition of H3F3A gene conserved which passed down from their common ancestor, as a result both of them showed higher correlations with their ancestor species (chicken with fish or frog with yeast).

Table 1. A variation of codon usage patterns for H3F3A genes

H3F3A	Human	Chimpanzee	Mouse	Chicken	Anole	Frog	Fish	Yeast
Human	1							
Chimpanzee	1	1						
Mouse	0,950978	0,950978	1					
Chicken	0,429346	0,4293456	0,45529	1				
Anole	0,746767	0,7467674	0,73009	0,418551	1			
Frog	0,632419	0,6324193	0,62479	0,296443	0,715428	1		
Fish	0,675034	0,6750338	0,65518	0,792883	0,642792	0,45159	1	
Yeast	0,183436	0,1834357	0,21773	0,106355	0,342074	0,62826	0,173725	1

Moreover, the noticeable no (negligible) and negative relationships are represented in Table 2 (H2AFX gene) in correlation of yeast with five species (human, chimpanzee, mouse, chicken and frog), $r = -0{,}052; -0{,}051; 0{,}081; -0{,}05$; and 0,02 respectively, as well as between nematode and fish in H2AFY (Table 2), $r = -0{,}053$. No(negligible) or negative correlation value arises with the presence of totally different nucleotide sequences between two gene sets as a result preference of totally different synonymous codons of two genetically distant species. Such weakest or negative relationships of selected gene sequences are good indicator to further explaining the implication of orthologous genes. We see that, the same gene that operates in two different organisms possess the striking different nucleotide sequences, but it always retains its common function associated with its ancestral gene. Similar relationships can also be observed in the rest of the tables which mainly highlighted with red.

Table 2. Variations of codon usage patterns for H2AFX and H2AFY genes

H2AFX	Human	Chimpanzee	Mouse	Chicken	Frog	Fish	Yeast
Human	1						
Chimpanzee	0,99880281	1					
Mouse	0,966700842	0,97002196	1				
Chicken	0,938527665	0,93761801	0,92966116	1			
Frog	0,7320164	0,73272799	0,72486995	0,77670781	1		
Fish	0,691242988	0,68263263	0,69222888	0,62177911	0,59732691	1	
Yeast	0,097556517	0,09658819	-0,1244014	0,10275853	0,00374589	0,2032101	1

H2AFY	Human	Chimpanzee	Mouse	Chicken	Anole	Frog	Fish	Nematode	Yeast
Human	1								
Chimpanzee	0,999524	1							
Mouse	0,930258	0,93199	1						
Chicken	0,83964	0,83789	0,82762	1					
Anole	0,669403	0,66658	0,58749	0,823941	1				
Frog	0,665623	0,66224	0,59706	0,821548	0,90648	1			
Fish	0,769945	0,77194	0,76148	0,619567	0,35844	0,32304	1		
Nematode	0,245633	0,24249	0,19047	0,435988	0,65268	0,68305	-0,052596	1	
Yeast	0,17005	0,16513	0,14264	0,296408	0,39183	0,41533	0,0591683	0,373146	1

Table 3. Variations of codon usage patterns for H1F0 and H2AFV genes

H1F0	Human	Chimpanzee	Mouse	Chicken	Frog	Fish	Nematode	Yeast
Human	1							
Chimpanzee	0,99807	1						
Mouse	0,98869	0,989532	1					
Chicken	0,93003	0,922071	0,922806	1				
Frog	0,7775	0,795159	0,792043	0,656744	1			
Fish	0,81542	0,828897	0,848447	0,750475	0,857931	1		
Nematode	0,82461	0,826829	0,832597	0,742045	0,709528	0,649506	1	
Yeast	0,84111	0,851582	0,841851	0,760521	0,755723	0,65074	0,9484652	1

H2AFV	Human	Chimpanzee	Mouse	Chicken	Anole	Frog	Zebrafish	Nematode	Yeast
Human	1								
Chimpanzee	0,9968	1							
Mouse	0,89227	0,88854	1						
Chicken	0,60066	0,60258	0,70465	1					
Anole	0,81961	0,81588	0,73248	0,33552	1				
Frog	0,84066	0,84335	0,74411	0,49142	0,79253	1			
Zebrafish	0,71569	0,71797	0,68569	0,43823	0,72728	0,73459	1		
Nematode	0,46806	0,47175	0,45426	0,41984	0,49714	0,59278	0,46673	1	
Yeast	0,54149	0,54611	0,47021	0,13669	0,67481	0,60608	0,56107	0,25156	1

H1F0 gene indicated possessing relatively highly conserved nucleotide sequences regarding the strong correlation values among species compared other genes. It is appeared from the table 3 that the weakest value in H1F0 gene is equal to nearly 0,65 was measured in a correlation of fish with nematode and yeast as well as between chicken and frog. The high correlation between nematode and yeast in H1F0 gene

($R = 0{,}94$) can be used to infer that they have relatively similar codon usage composition.

We observed that there is no negative or no(negligible) relationship exist in H1F0 and H2AFV genes in relationship of yeast with other species compared to other genes. The weakest relationship was calculated between chicken and baker's yeast in H2AFV gene (Table 3), with the value of $0{,}136$.

The relationship between different gene sequences of same species

We also calculated the frequency of the occurrence of synonymous codons among various gene sequences in each separate organism. The calculated dataset was listed in the tables shown below and the correlation values for each obtained housekeeping gene were discussed.

The same gene sequences were obtained as done previously. In contrast, here we aimed to investigate codon usage pattern in terms of gene families within one species. Each of the gene family contains several separate gene sequences and these are homologous genes that have diverged within one species. Unlike orthologous genes, gene sequences that included selected gene family hold different functions and catalyze the different reactions.

As represented in the tables, different relationships were observed between same histone genes in various organisms. Each table listed in Table 4, exemplify the differences of the codon usage patterns of histone gene family sequences within one species.

Firstly, obvious lower rates observed in the relationships among gene sequences in same species compared to orthologous genes. It appeared from the tables that, there is no strong correlation among these parameters, because most of the correlation results are almost lower than $0{,}75$. The indicated family genes were originated from the different ancestors, consequently, they acquired new functions which catalyzed the new reaction. But, nevertheless, it is important to note that the functions of translated new proteins are similar, but not identical and they are generally all acting for the creation of histone enzymes. The collaboration of such biased gene sequences for the common purpose is interestingly undeniable.

Interestingly, the choice of synonymous codon in chicken's histone gene families varies obviously among genes. For example, the strongest relationship exists between H2AFX and H3F3A genes in chicken ($R = 0{,}839$) where these genes are assumed as weakly correlated genes in mammals with regard to non-conservative codon usage patterns, where R value is just under $0{,}35$. Similar relationships can also be observed between H2AFV and H2AFX genes. In contrast, relatively lower relationship exists between chicken's H2AFY and H1F0 genes ($R = 0{,}51$) compared to mammals.

Examining the frequency of codons in western clawed frog revealed considerable lower correlation rates in terms of H2AFX gene. In this findings, R value was measured about $0{,}199$ between H2AFX and H2AFY genes, the complement relationship is obviously strong in the rest of the species where R value is equal to $0{,}955$ in yeast. Similar declines have been also observed in the correlation of H2AFX gene with H1F0 gene in frog.

Table 4. Variations of fivehistone gene sequences (H2AFV, H2AFX, H2AFY, H3F3A, H1F0) in selected species (human, chimpanzee, house mouse, chicken, green anole, western clawed frog, zebrafish, nematode, baker's yeast).

Human	H2AFV	H2AFX	H2AFY	H1F0	H3F3A
H2AFV	1				
H2AFX	0,377	1			
H2AFY	0,563	0,651	1		
H1F0	0,477	0,545	0,779	1	
H3F3A	0,606	0,282	0,436	0,326	1

Chimpanzee	H2AFV	H2AFX	H2AFY	H1F0	H3F3A
H2AFV	1				
H2AFX	0,367	1			
H2AFY	0,551	0,644	1		
H1F0	0,492	0,537	0,788	1	
H3F3A	0,611	0,275	0,434	0,343	1

Mouse	H2AFV	H2AFX	H2AFY	H1F0	H3F3A
H2AFV	1				
H2AFX	0,511	1			
H2AFY	0,688	0,629	1		
H1F0	0,551	0,508	0,789	1	
H3F3A	0,511	0,317	0,389	0,352	1

Chicken	H2AFV	H2AFX	H2AFY	H1F0	H3F3A
H2AFV	1				
H2AFX	0,758	1			
H2AFY	0,492	0,425	1		
H1F0	0,701	0,639	0,510	1	
H3F3A	0,780	0,839	0,352	0,576	1

Anole	H2AFV	H2AFY	H3F3A	H2AFY2
H2AFV	1			
H2AFY	0,591	1		
H3F3A	0,533	0,489	1	
H2AFY2	0,663	0,840	0,461	1

Frog	H2AFV	H2AFX	H2AFY	H1F0	H3F3A
H2AFV	1				
H2AFX	0,326	1			
H2AFY	0,595	0,199	1		
H1F0	0,597	0,316	0,766	1	
H3F3A	0,605	0,296	0,542	0,527	1

Zebrafish	H2AFV	H2AFX	H2AFY	H1F0	H3F3A
H2AFV	1				
H2AFX	0,558	1			
H2AFY	0,388	0,710	1		
H1F0	0,506	0,536	0,559	1	
H3F3A	0,396	0,707	0,655	0,498	1

Yeast	H2AFV	H2AFX	H2AFY	H1F0	H3F3A
H2AFV	1				
H2AFX	0,714	1			
H2AFY	0,738	0,955	1		
H1F0	0,319	0,389	0,291	1	
H3F3A	0,474	0,736	0,697	0,503	1

Analyzing the variation of codon usage patterns among gene datasets

We previously examined the divergence in the use of codon patterns across obtained family genes. Here we aimed to investigate how the strength of codon usage bias changes across individual gene sequences compared to whole genome dataset. In this manner, the vast quantities of gene datasets were extracted from Ensembl. For this matter, a dataset of vast amount of gene sequences were generated for glycine max (soybean) species. Afterwards, the statistical analyses were employed in order to figure out the mean correlation value for wide range of gene sequences included in the gene datasets. In order to view codon preference across massive gene sequences in the dataset a graphical method was employed. This method provides an easy way for clearly spotting codon usage bias across the huge gene datasets.

We analyzed that patterns of codon usage can vary widely across gene sequences depending on the amount of genes in each dataset. We generated 20 datasets which each individual dataset comprises 1 to 900 gene sequences.

The graphs here were devised to look at codon usage heterogeneity and here one can clearly see differences in codon usage patterns within different number of gene sequences in 20 datasets. Each of the gene dataset was denoted with different colored lines. Looking at the codon frequencies in different gene dataset for minor number of gene sequences (1 gene and 10 genes), we observed a clear bias in choosing the synonymous codons indicated with entangled lines and this assumption can be proven by the lower mean correlation value, represented below the graphs. Mean correlation value here indicate how the use of codons altering across species depending on the number of gene sequences. Therefore, it was seen that the mean correlation values for selected genes do indeed enhance in conjunction with the increased number of gene sequences.

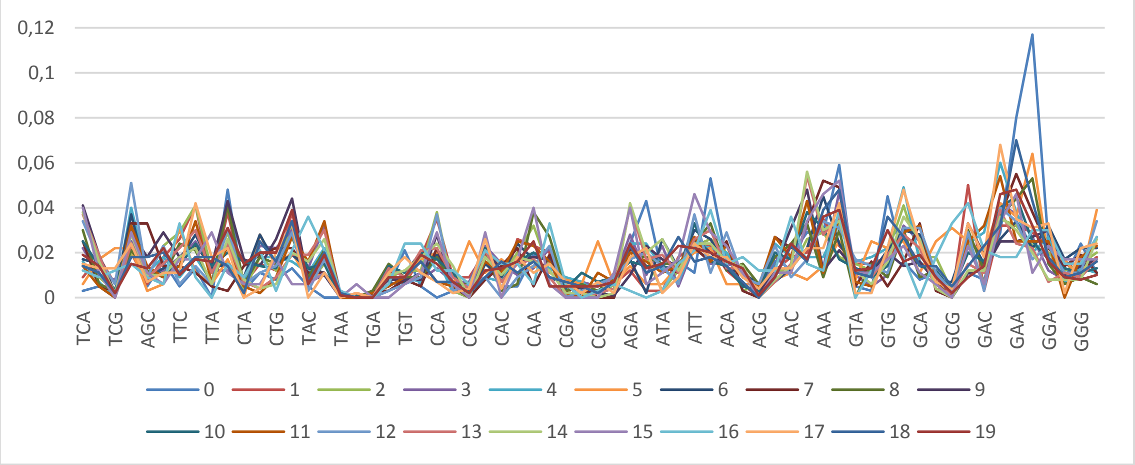

Figure 1. The graph shows the variation of codon usage for 1 gene in a set in glycine max (soybean) Mean correlation value for 1 gene in a set: $R = 0{,}565$

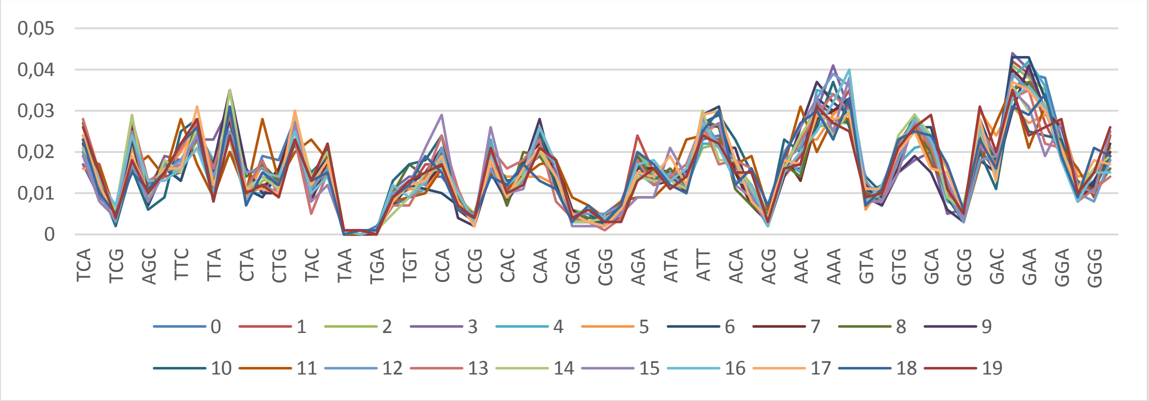

Figure 2. The graph shows the variation of codon usage for 10 genes in a set in glycine max (soybean) Mean correlation value for 10 genes in a set: $R = 0{,}885$

In contrast to minor number of genes, we can barely observe fluctuations in the different colored gene sequenceslines and noticeable tidiness was observed in the demonstration of codons variations as illustrated by steady curve line with respect to the mass amount of gene sequences (100 and 900 genes). The mean correlation value has been found to rise considerable with the higher number of gene datasets. While mean correlation value is equal to 0,565 for 1 genes in a set in glycine max (Figure 1), it is estimated about 0,997 for 400 genes in a set (Figure 4) which have contributed evidently similar codon usage patterns across gene datasets. According to these findings it can be assumed that the significant switches that responsible highly biased codon usage patterns was mainly appeared within separate gene sequences compared the whole genome dataset of an organisms.

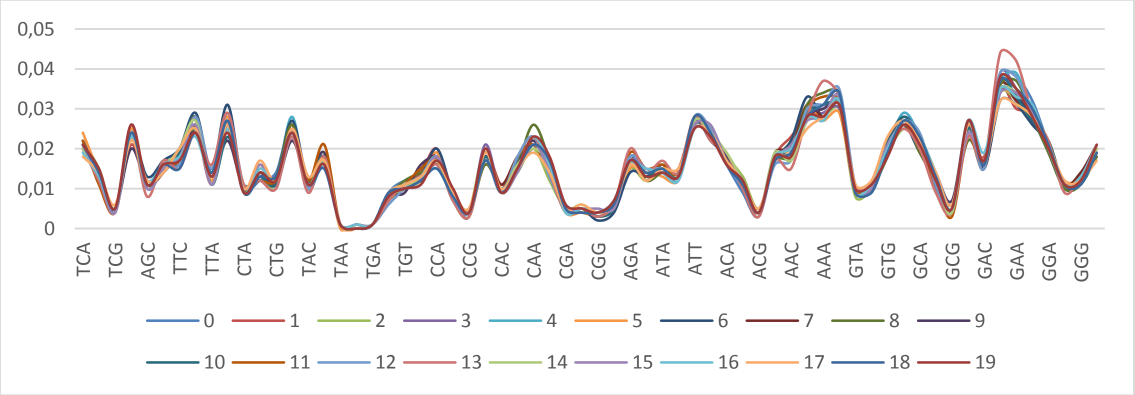

Figure 3. The graph shows the variation of codon usage for 100 genes in a set in glycine max (soybean) Mean correlation value for 100 genes in a set: R=0,979

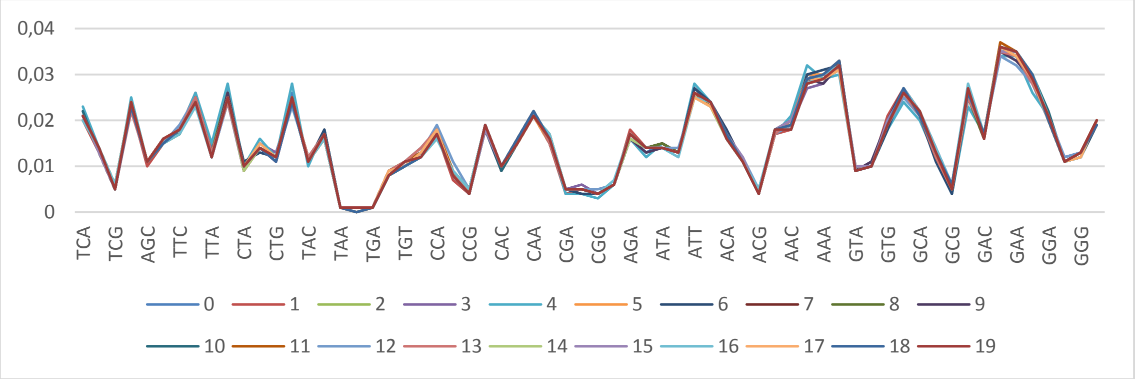

Figure 4. The graph shows the variation of codon usage for 900 genes in a set in glycine max (soybean) Mean correlation value for 900 genes in a set: $R = 0{,}997$

The graph shown below was allowed a more in depth look at the mean values for entire gene datasets (Figure 5). X axe from the graphs displays the number of the genes whilst Y axe is an indicator of mean correlation value. It was found from the scatter plot that the mean correlation value is highly proportional with the number of the genes.

Generally, identifying the trends in the curve line elucidated that the mean values that reflecting codon usage bias for small number of genes in a set are obviously low with R values observed between 0,55 and 0,85 for genes ranges from 1 to 5. As an amount of the gene sequences increased the mean correlation value went up gradually, peaked up with the highest number of genes and remained steady where R value reached to its upward trend value equal roughly to 1. This trends can therefore be inferred that the fully amount of gene sequences included in the gene datasets of same organisms comprise the evidently similar (almost same) codon usage patterns regardless of the individual biased different functional gene sequences.

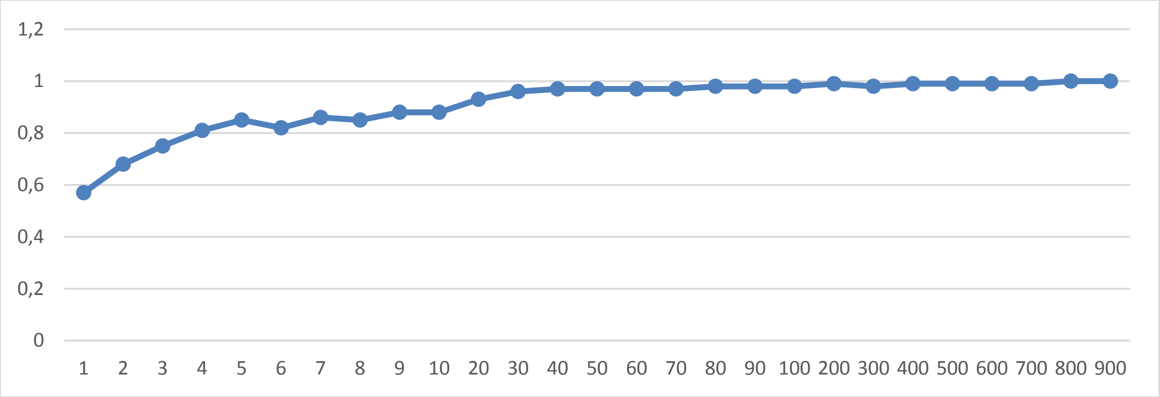

Figure 5. The graph indicates the mean correlation values for a different number of genes in datasets in glycine max (soybean)

3. Conclusion and discussion

Here we demonstrated that the codon usage variations are important factor in the evolution of eukaryote genomes. In particular, it has been shown that the choice of synonymous codon differs significantly among family genes in same and different species and family genes in same species indicated a relatively higher biased codon usage patterns among gene sequences compared to family genes in different species - orthologous genes. Therefore, genetically close species indicated a higher positive correlation values in terms of orthologous genes, however several negligible or negative relationships were also detected between some genetically distant biological species.

Consequently, it turns out that the biased relationship between biological species for individual gene sequences is obviously discernible as measured for orthologous genes in contrast to complete genome dataset. We assumed that every biological species possesses common codon usage patterns and the switches that appeared in family genes which were passed down from their proper ancestral genes were distributed among whole complete genome instead of in individual gene sequence. Hereby, such separation of major changes of common codon usage patterns caused minor switches in separate gene sequences and biological species are largely correlated in terms of genome database compared to individual family genes. It can also be assumed that the divergence in the frequency of choosing synonymous codon in family genes between species was appeared in response to different environmental conditions that living organisms evolved and the variation in the occurrence of codon bias among family genes in same species can be appeared a result of non-random evolutionary causes.

In general, we concluded that all biological species share the similar properties which were passed down from their common ancestor, regardless the obvious variation among individual gene sequences with respect to evident switches in codon preference. Because living organisms try to keep their common codon usage patterns conservative, hence codon usage patterns of species were distributed among family genes instead of single gene sequences. Thereby, relatively higher correlation values were observed in orthologous genes in comparison with family genes in same species reflects the consistent codon usage patterns that species tries to conserve.

References

Bennetzen, J.L. & Hall, B.D. (1982). Codon selection in yeast. *Journal of Biological Chemistry*, *257*(6), 3026-3031.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. *Molecular systems biology*, *9*(1).

Blake, W.J., Kærn, M., Cantor, C.R., & Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature*, *422*(6932), 633.

Chen, H., Sun, S., Norenburg, J. L., & Sundberg, P. (2014). Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). *PloS one*, *9*(1), e85631.

Collins, D.W., & Jukes, T.H. (1993). Relationship between G+ C in silent sites of codons and amino acid composition of human proteins. *Journal of Molecular Evolution*, *36*(3), 201-213.

Comeron, J.M., & Kreitman, M. (2002). Population, evolutionary and genomic consequences of interference selection. *Genetics*, *161*(1), 389-410.

Eyre-Walker A. (1996). Isochore evolution in mammals: a human-like ancestral structure. *Mol. Biol. Evol.*, 13, 864-872.

Flagel, L.E., & Blackman, B.K. (2012). The first ten years of plant genome sequencing and prospects for the next decade. In *Plant Genome Diversity*, Volume 1 (pp. 1-15). Springer, Vienna.

Francino, M.P., & Ochman, H. (1999). Isochores result from mutation not selection. *Nature*, *400*(6739), 30.

Gambari, R., Nastruzzi, C., & Barbieri, R. (1990). Codon usage and secondary structure of the rabbit alpha-globin mRNA: a hypothesis. *Biomedica Biochimica Acta*, *49*(2-3), S88.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., & Mercier, R. (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic acids research*, *9*(1), 213-213.

Harrison, R.J., & Charlesworth, B. (2010). Biased gene conversion affects patterns of codon usage and amino acid usage in the *Saccharomyces sensu stricto* group of yeasts. *Molecular biology and evolution*, *28*(1), 117-129.

Hershberg, R. & Petrov, D.A. (2008). Selection on codon bias. *Annual Review of Genetics*, 42, 287-299.

Higgs, P.G., & Ran, W. (2008). Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Molecular biology and evolution*, *25*(11), 2279-2291.

Hooper, S.D., & Berg, O.G. (2000). Gradients in nucleotide and codon usage along *Escherichia coli* genes. *Nucleic Acids Research*, *28*(18), 3517-3523.

Huynen, M.A., Konings, D.A. & Hogeweg, P. (1992). Equal G and C contents in histone genes indicate selection pressures on mRNA secondary structure. *Journal of Molecular Evolution*, *34*(4), 280-291.

Ingvarsson, P.K. (2007). Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Molecular biology and evolution*, *24*(3), 836-844.

Kanaya, S., Yamada, Y., Kinouchi, M., Kudo, Y. & Ikemura, T. (2001). Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *Journal of Molecular Evolution*, *53*(4-5), 290-298.

Kudla, G., Murray, A.W., Tollervey, D. & Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *Science*, *324*(5924), 255-258.

Lavner, Y. & Kotlar, D. (2005). Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene*, *345*(1), 127-138.

Lipnerová, I., Bureš, P., Horová, L., & Šmarda, P. (2012). Evolution of genome size in Carex (Cyperaceae) in relation to chromosome number and genomic base composition. *Annals of Botany*, *111*(1), 79-94.

Plotkin, J.B. & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, *12*(1), 32.

Plotkin, J.B., Robins, H., & Levine, A.J. (2004). Tissue-specific codon usage and the expression of human genes. *Proceedings of the National Academy of Sciences, 101*(34), 12588-12591.

Powell, J.R. & Moriyama, E.N. (1997). Evolution of codon usage bias in *Drosophila*. *Proceedings of the National Academy of Sciences, 94*(15), 7784-7790.

Qin, H., Wu, W. B., Comeron, J. M., Kreitman, M., & Li, W. H. (2004). Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. *Genetics*, *168*(4), 2245-2260.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., ... & Minx, P. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science, 326*(5956), 1112-1115.

Sharp, P.M., Emery, L.R., & Zeng, K. (2010). Forces that influence the evolution of codon bias. Philosophical *Transactions of the Royal Society B: Biological Sciences*, *365*(1544), 1203-1212.

Šmarda, P., Bureš, P., Šmerda, J., & Horová, L. (2012). Measurements of genomic GC content in plant genomes with flow cytometry: a test for reliability. *New Phytologist*, *193*(2), 513-521.

Tuller, T., Waldman, Y.Y., Kupiec, M. & Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences, 107*(8), 3645-3650.