Հայաստանի Գիտությունների Ազգային Ակադեմիա, Հայաստանի Կենսաբանական Հանդես
Национальная Академия Наук Армении, Биологический Журнал Армении
National Academy of Sciences of Armenia, Biological Journal of Armenia

Биолог. журн. Армении, 3-4 (57), 2005

УДК 575:599.9

EXTREME CODON BIAS DIFFERENCES IN MITOCHONDRIAL GENOMES OF *TETRAHYMENA* AND *PARAMECIUM*

M.M. MORADIAN

Primex Cliniacal Laboratory, CA, USA, mmoradia@ucla.edu

The codon usage differences in mitochondrial genomes of two ciliated protozoa, *Tetrahymena* and *Paramecium* have been investigated. Although they contain the same tRNA genes with the same anticodon and a very similar amino acid composition in their shared mitochondrial proteins, their codon preferences are extremely different. Codons, which end in A or T, are used 91% of the time in *Tetrahymena* where their incidence in *Paramecium* mitochondrial proteins is only 43%. Basically in *Tetrahymena*, with over 80% AT content, all amino acids with no exception prefer synonymous codon(s) that end in either A or T. Also, the GC contents at all three codon positions were significantly higher in *Paramecium* mitochondrial proteins. Differences in nucleotide composition at the second and third codon positions clustered mitochondrial proteins into two distinct groups. There was a clear relationship between extreme codon usage bias quantified by the Effective Number of Codons (ENC) and the GC content at the third codon positions in *Tetrahymena*. However, no relationships were observed between the ENC and gene lengths in either mitochondrial genome. tRNA availability, which is selectively imported from cytosol into mitochondria is a possible driving force for such biased codon usage in *Tetrahymena* has been postulated.

Extreme codon - mitochodriae genome - ciliated protozoa

The number of completely sequenced genomes is rapidly on the rise and they have recently provided an extensive sample of evolutionary adjustment of codon usage and meaning spanning their history [15], which is true for both nuclear and organelle genomes. The mitochondrial and nuclear DNA cannot be distinguished by physical and chemical (e.g., buoyant density and binding affinity) characteristics and in this sense the mitochondrial DNA is not unique. However, the metabolism of mitochondrial DNA in ciliate *Tetrahymena* is reported to be different from that of nuclear DNA suggesting that it is produced via a different metabolic pathway than nuclear DNA [3]. Therefore, despite their physical and chemical similarities nuclear and mitochondrial DNA are produced by different DNA synthetic systems. Transcription machinery may also be different in mt DNA in ciliates since there are no distinct promoter regions for individual genes. A comparative sequence analysis of an intergenic region among five *Tetrahymena* species revealed a conserved GC promoter region (data not published). This region is particularly interesting since its ability to initiate a bidirectional transcription suggests a polycistronic transcription in these mt genomes. Such differences inspired us to study the codon usage choices in ciliate

mitochondrial genes and present the mechanisms responsible for them.

Codon usage appears to be non-random in synonymous codon families in all life forms. There are several processes considered to cause biased codon usage such as natural selection [1, 25] and directional mutation pressure [11, 22, 13]. Also, a positive correlation between tRNA frequencies and frequently used codons [10, 19] and between gene expression level and codon bias [9, 18] has been reported. Furthermore, possible roles for gene length [7, 19], gene function [8], secondary protein structure [21, 29], the origin of genes [14, 17], and the chromosomal regional location of genes [6, 26] in determining the codon choices have been proposed. The codon usage bias is a very well researched field, yet surprisingly an overwhelming number of these studies are on nuclear genomes. The focus of mitochondrial projects has been on genetic code definition and variation, which explain how codons are lost and reassigned in mitochondria. Although there are several variables that may affect the codon usage bias, there is no universal mechanism that can be delineated in every genome. Thus, analyzing more genomes with different natures may reveal novel information and unveil new mechanisms for their codon usage choices.

Mitochondrial genomes may possess unique mechanisms for codon usage since the factors responsible for codon bias may act differently in mitochondrial and nuclear DNA. Purifying selection for translational efficiency and accuracy is strong on highly expressed genes in nucleus resulting in a distinct codon usage bias. On the other hand, in lowly expresses genes selection is relatively weak, so the usage pattern is mainly affected by mutation pressure and random genetic drift and may be less skewed [25]. Genes in mitochondrial DNA, especially in ciliates, have equal expression levels because they transcribe the entire genome at once. Thus a variable expression level and selective pressure may not play a major role in mitochondrial codon usage choices. Codons, as DNA sequences, are subject to mutational pressure acting on all DNA sequences in any organism. In prokaryotes, in general, genome wide codon bias is determined primarily by mutational pressure, where GC content variation is the most important factor in determining the codon usage choices between different organisms [5]. Also, species-specific codon bias is strongly correlated with overall genome percentage GC content [12]. The G+C nucleotides content in the third codon position has been used as an indirect measure of the extent of bias in different synonymous codons and is significantly correlated with codon bias in many organisms, however it is not universal [25]. Since AT mutational pressure acts on majority of the mitochondrial genomes it is plausible to suggest that the AT-rich genomes tend to use AT-rich codons and vice versa. However AT-richness may be due only to an abundance of T, which may range between 13% at the third codon position, as in the bird *Aythya americana* to 68%, as in nematode *Onchocerca volvulus* [15]. Nevertheless, mutational pressure may be considered a potential cause for codon usage bias in nuclear as well as in mitochondrial genome.

Materials and Methods.

Plots.

Tetrahymena and *Paramecium* mitochondrial genes were converted to codons through using

DNA_Strider [16]. DNA Strider is a new integrated DNA and Protein sequence analysis program written with the C language for the Macintosh Plus, SE and II computers. It has been designed as an easy to learn and use program as well as a fast and efficient tool for the day-to-day sequence analysis work. The PR2 bias plots and neutrality plots [28] were made using the nucleotide compositions in different positions of the codons. The strength of the selection on a given gene relative to the mutation pressure can be estimated by the method of the relative neutrality plot (RNP), which gives indications on how 'neutral' a coding sequence can be considered. The method consists of plotting the G+C content at the constrained (or nonsynonymous) positions (that is, first and second positions) of the codons against the G+C content at the relaxed (or synonymous) position (that is, third position). The slope of the resulting linear correlation gives evidence on how the protein sequence is affected by the mutational bias acting on the nucleotide sequence, and thus on how strongly the selection pressure acting on the protein can counteract this bias. Note that the effect measured is relative to the translational selection acting on the third position of codons and that the strength of this pressure is supposed to be weak compared to the selection on the protein sequence. The slope is expected to be equal to one if the protein sequences are under no selective constraints, and to decrease with the strength of the selection acting at the protein level. Translational selection is also expected to reduce the correlation, though to a lower extent. For this study, I analyzed the correlations according to different mitochondrial genomes to determine whether there were differences in the relative selection pressures in each of the genomes. In a PR2-bias plot, the value of the AT-bias A/(A+T)] is plotted as the ordinate and the value of the GC-bias [G/(G+C)] is plotted as the abscissa [28]. In this plot, the center of the plot, where both coordinates are 0.5, is the place where A= T and G=C, holding PR2. A vector from the center represents the extent and direction of biases from PR2. PR2 bias plots are particularly informative when PR2 biases at the third codon position of the four-codon amino acids of individual genes are plotted. In this case, A3/(A3+T3) and 'G3/(G3+C3) are plotted as the ordinate and abscissa respectively. A3, T3, G3 and C3 are fractions of the corresponding nucleotides at the third codon position, where A3+T3+G3+C3=1. PR2 biases at the third codon position are presented for all codons in both *Tetrahymena* and *Paramecium* mitochondrial genes.

ENC, CBI, Chi2 Calculations.

Effective number of codons, codon bias index, and Chi square values were calculated using DNASP. DnaSP, DNA Sequence Polymorphism, is a software package for the analysis of nucleotide polymorphism from aligned DNA sequence data. DnaSP can estimate several measures of DNA sequence variation within and between populations (in noncoding, synonymous or nonsynonymous sites, or in various sorts of codon positions), as well as linkage disequilibrium, recombination, gene flow and gene conversion parameters [23]. One method to quantify variations in base composition patterns is to determine how similar the observed frequencies are to the expected frequencies. The predicted usage of codons can be estimated for any given GC-content at third codon position if the patterns are determined solely by mutations. One such plot is the ENC-GC3 plot. ENC refers to the effective number of codons used at a given GC-content at third codon position. The effective number of codons can be shown as if all 61 codons are used at equal frequencies the ENC value is 61 (for 61 codons), but if only one codon on the average is used for each amino acid the ENC value approaches 20 (for 20 amino acids and one codon per amino acid). As the GC-contents at third codon positions change from 50% to either a higher or a lower GC-content, the number of codons used per amino acid is reduced. In AT-rich genomes, codons ending in A or T are primarily used. In GC-rich genomes, codons ending in G or C are primarily used. By determining the extent to which the observed number of effective codons deviates from the predicted number of effective codons, it is possible to estimate the extent to which the genes within a dataset are subjected to constraints other than mutational effects [30]. Codon bias index is another measure of directional codon bias, it measures the extent to which a gene uses a set of optimal codons. In a gene with absolute codon bias, CBI will equal 1.0, in a gene with random codon usage CBI will equal 0.0. It is possible that the number of optimal codons be less than expected number of codons by random change. This results in a negative value for CBI [28]. The scaled Chi 2 is a measure of departure from equal use of synonymous codons estimated by a Chi 2 statistic scaled by dividing it by the number of codons analyzed; the higher the values, the higher the degree of bias, and 0 indicates a perfectly uniform usage. In the present study scaled Chi 2 was estimated with the correction for continuity, consisting in subtracting 0.5 from the absolute value of the deviation between observed and expected frequencies, when the observed number of synonymous codons is less than 5. Equivalent analyses of the scaled Chi 2 were also performed without correcting for continuity.

Results and discussion.

Amino acid composition and tRNA content.

Table 1. **tRNA content in mitochondrial genomes of *Tetrahymena* and *Paramecium***

Paramecium tRNAs			*Tetrahymena* tRNAs		
AA	anticodon	codon	AA	anticodon	codon
TAC tyr Y	GUA	UAC	TAC tyr Y	GUA	UAC
TTC phe F	GAA	UUC	TTC phe F	GAA	UUC
TGA trp W	UCA	UGA	TGA trp W	UCA	UGA
ATG met M	CAU	ATG	ATG met M	CAU	AUG
			CAC *his H*	GUG	CAC
			GAA glu E	UUC	GAA
			TTA leu L	UAA	UUA

Table 2. **Amino acid composition of genes in *Tetrahymena* and *Paramecium* mt genomes. The Analysis of Variance (ANOVA) between these two genomes indicates the similarity of their amino acid compositions (pvalue=0.96).**

Percentage Amino Acid	*Tetrahymena* Genes	*Paramecium* Genes
F	9,8	13,4
L	12,4	12,5
I	11,1	4,8
M	2,2	3,1
V	4,4	6,2
S	7,3	7,5
P	2,1	2,6
T	4,6	5,0
A	3,0	5,5
Y	6,9	5,0
H	1,6	1,8
Q	1,9	2,0
N	9,3	5,1
K	8,0	6,8
D	2,7	3,0
E	3,0	3,4
C	0,9	1,6
W	1,9	2,0
R	3,2	3,9
G	3,7	4,7

Table 1 shows the list of tRNAs shared by *Tetrahymena* and *Paramecium* mt genomes. These four tRNA genes (i.e., Tyrosin (Tyr), Phenylalanine (Phe), Tryptophan (Trp), and Methionine (Met)) have the same anticodon sequence in both species, suggesting that these mt genomes may behave similarly in using synonymous codons. Besides, *Tetrahymena* mitochondria contain three additional tRNAs (i.e., Leucine(Leu), Glutamic acid (Glu), and Histidine (His)). The overall amino acid composition of the mt genes in these two genomes was significantly similar, suggesting that the transmembrane and ribosomal proteins coded by these genomes evolved in parallel with regards to their amino acid usage (Table 2). This striking similarity may imply that either the amino acid composition of *Tetrahymena* and *Paramecium* mt genomes has not been significantly diverged from their latest common ancestor or they have been subject to a similar evolution.

Synonymous codon usage and nucleotide composition.

Despite similarities mentioned earlier, the synonymous codon usage in mt genomes of *Tetrahymena* and *Paramecium* are significantly different (Table 3). There is a strong bias for Adenine and Thymine (A and T) rich codons in *Tetrahymena* mt genome where all of the amino acids with no exceptions overwhelmingly use the synonymous codons, which contain more A and T. More specifically, *Tetrahymena* mt genes prefer codons, which almost exclusively end

Table 3. **Synonymous codon usage in *Tetrahymena* and *Paramecium* mt genome for individual amino acid**

codons	*Tetrahymena*	*Paramecium*	*Tetrahymena* and *Paramecium* Codon Comparison Table (Usage Percentage)			codons	*Tetrahymena*	*Paramecium*
TTT phe F	89	63	codons	*Tetrahymena*	*Paramecium*	GAA glu E	91	32
TTC phe F	11	37	TAT tyr Y	86	36	GAG glu E	9	68
TTA leu L	83	16	TAC tyr Y	14	64	CGT arg R	8	6
TTG leu L	3	13				CGC arg R	0	15
CTT leu L	5	19	TCT ser S	32	26	CGA arg R	2	11
CTC leu L	0	30	TCC ser S	2	21	CGG arg R	0	4
CTA leu L	8	14	TCA ser S	37	12	AGA arg R	86	16
CTG leu L	0	8	TCG ser S	2	9	AGG arg R	3	48
			AGT ser S	23	8			
CCT pro P	51	28	AGC ser S	4	25	GGT gly G	80	17
CCC pro P	2	51				GGC gly G	2	36
CCA pro P	45	19	ACT thr T	50	24	GGA gly G	15	25
CCG pro P	3	3	ACC thr T	5	36	GGG gly G	4	23
			ACA thr T	45	19			
GTT val V	49	34	ACG thr T	0	21	GCT ala A	56	31
GTC val V	3	28				GCC ala A	4	38
GTA val V	44	21	ATT ile I	92	43	GCA ala A	35	24
GTG val V	3	17	ATC ile I	8	57	GCG ala A	4	7
CAT his H	79	33	ATA met M	75	38	TGT cys C	88	30
CAC his H	21	67	ATG met M	25	62	TGC cys C	12	70
CAA gln Q	97	41	AAT asn N	87	36	TGA trp W	97	50
CAG gln Q	3	59	AAC asn N	13	64	TGG trp W	3	50
GAT asp D	88	40	AAA lys K	97	39	TAA Stop	100	57
GAC asp D	12	60	AAG lys K	3	61	TAG Stop	0	53

with either A or T. On the other hand, *Paramecium* has less biased and more balanced codon usage and-especially in two fold degenerate codons-acts opposite to *Tetrahymena*. *Paramecium* mt genes do not discriminate against codons ending in guanine (G) or cytosine (C) and use these codons at approximately 50% of the time, especially for amino acids that are encoded by four and/or six codons. Unlike in *Tetrahymena, Paramecium* mt genes prefer codons that end in G or C in two fold degenerate codons with the exception of Phe (Table 3). These two mt genomes share four tRNAs with the same anticodon sequence (i.e., Tyr, Phe, Trp, and Met), where the first three of these common tRNAs are capable of wobbling. In *Tetrahymena* tRNAs Tyr, Phe, and Trp strongly and effectively use wobbling where they choose to pair a synonymous codon with higher AT content rather than a codon with complementary sequence to the mitochondrial encoded tRNA anticodons. In contrast *Paramecium* mt genes do not discriminate against codons with low AT content, in fact, many amino acids use synonymous codons, which possess higher GC content (Table 3). In addition tRNA His, which is only coded in *Tetrahymena* mitochondria acts in the same manner as the common ones.

A few codons have disappeared from *Tetrahymena* mt genome where as all codons are used in *Paramecium* mt genes. They include two of the Leu (CTC and CTG) and two of the Arginine (Arg) codons (CGC and CGG) along with a Threonine (Thr) codon (ACG). Codon (TAG), which is a stop signal in *Tetrahymena* nucleus and *Paramecium* mitochondria, is also missing in *Tetrahymena*. Such significant differences between *Tetrahymena* and *Paramecium* mitochondrial codon usage may suggest different grounds and mechanisms for codon usage bias

between mitochondria of these two ciliates.

The nucleotide composition of *Tetrahymena* and *Paramecium* mitochondrial genomes are conspicuously different. In *Tetrahymena* mitochondria, mutational bias is clearly towards an AT rich genome where over 80% of the nucleotides are either A or T. Yet just 59% of *Paramecium* mitochondrial genome is A or T. Also, codons ending in A or T comprise 91% of the codons used in *Tetrahymena* mt genes compared to 43% in *Paramecium*, yet interestingly in both species, T is used slightly more than A at the third codon position (Table 4). To investigate whether these differences in nucleotide composition affect codon choices and whether they are restricted in any of the codon positions, the relation between G and C and between A and T content in all codon families was analyzed by PR2 bias plots [25]. Plots in figure 1 indicate that the nucleotide content of mitochondrial genomes in *Tetrahymena* and *Paramecium* is quite diverse and may affect their codon choices. This is more obvious in second and third codon positions where the genes of the two species are assembled into two distinct clusters (Figure 1b and 1c). *Paramecium* codons are particularly interesting since their second codon position is almost exclusively comprised of G and the major base of the third position is C. Hence, in all codon groups except Arg the codon that ends in C is most frequently used (Table 3).

Table 4. **AT composition (percentages) at the third codon position of *Tetrahymena* and *Paramecium***

Codon Percent	*Tetrahymena*	*Paramecium*
Third pos AT	91	43
Third pos A	47	44
Third pos T	53	56

GC contents in *Tetrahymena* and *Paramecium* mitochondrial genes analyzed in this study indicate that the average GC3s are significantly higher in *Paramecium* (ANOVA pvalue= 2.1*E-20). GC1 and GC2 are also significantly higher in *Paramecium* (Table 5). These results are consistent with the higher GC

Table 5. **Summary of codon usage in *Tetrahymena* and *Paramecium* mitochondria**

	# Codons	GCall	GC1	GC2	GC3	ENC	CBI	SChi2
Tetrahymena	7439	0,2208	0,1279	0,2951	0,0805	30,16	0,764	1,297
Paramecium	7190	0,4445	0,6524	0,3416	0,5598	50,35	0,398	0,417
ANOVA pvalue	-	3.0*E-15	2.8*E-18	0,05	2.1*E-20	5.4*E-15	1.9*E-15	4.1*E-16

contents in *Paramecium* mt genome [4]. Differences in GC contents are greatest at the third codon position followed by the first and the second positions, suggesting that a different GC mutation bias leads to different codon choice despite little changes in proteins' amino acid composition. To determine the nature of the relations among three codon positions, neutrality plots [27] are analyzed for *Tetrahymena* and *Paramecium* mitochondrial genes (Figure 2). These plots (GC12 vs GC3) indicate that *Paramecium* has a wide range of GC3, whereas *Tetrahymena* has narrow GC3 distribution. If the slope of the regression line is near zero then one may suggest that there are low mutation bias and high conservation of GC contents throughout the genome. In *Paramecium* there is a correlation be-

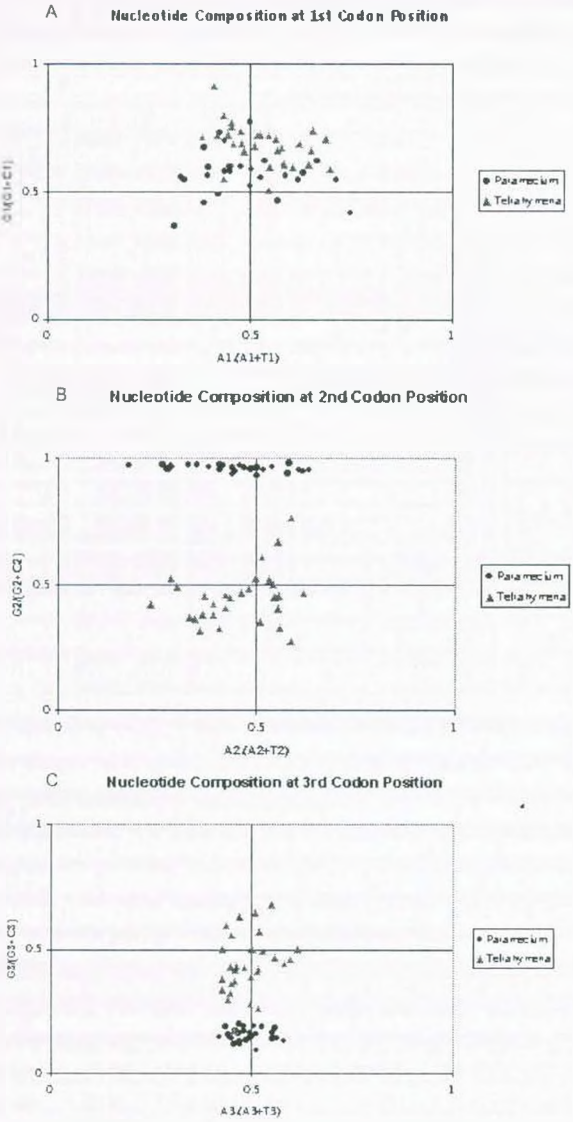

Fig. 1. PR2-bias plots: A-first position [A1/(A1+T1) against G1/(G1+C1)] with average position of x=0.4930 ± 0.1110, y=0.5715 ± 0.0859 for *Paramecium* and x=0.5394 ± 0.0813, y=0.6593 ± 0.0896 for *Tetrahymena*. B-second codon position [A2/(A2+T2) against G2/(G2+C2)] with average position of x=0.4317 ± 0.1175, y=0.9597 ± 0.0122 for *Paramecium* and x=0.4637 ± 0.0940, y=0.4336 ± 0.1137 for *Tetrahymena*. C-third codon position with average position of x=0.4921 ± 0.0406, y=0.1554 ± 0.0250 for *Paramecium* and x=0.4892 ± 0.0481, y=0.4085 ± 0.1114 for *Tetrahymena* [A3/(A3+T3) against G3/(G3+C3)].

tween GC contents and the slope of the regression line is approximately 0.25 suggesting that the intra-genomic GC mutation bias affect the GC contents somewhat similarly among all positions of codons in mitochondrial genes. In contrast in *Tetrahymena* the slope of the regression line is quite high (i.e., 1.91), which may indicate high mutation bias and low conservation of GC contents throughout the mitochondrial genome (Figure 2).

Interestingly, the frequencies of bases A and T at the second codon position for each of the transmembrane (Tm) and ribosomal (Rp) proteins in *Tetrahymena* mt genome generate two distinct clusters (Fig. 3). It can be seen that the two groups are separated in a region around the center of the graph (30, 30) by a line T2 = A2. Proteins that lie above the line T2 = A2, can be categorized as Tm and the ones below the line T2 = A2 will be Rp proteins. For bases Guanine (G) and Cytosine (C), the variation in their frequencies is not so marked and the distribution is not shown here. Tm proteins in *Tetrahymena* mitochondria have more codons with T at their second codon position than those with A at this position. Since Tm and Rp proteins in *Paramecium* do not generate any marked clustering I suggest that a strong mutational pressure and a biased codon usage is necessary, and may be used, to cluster proteins based on their function. Thus, I can classify the proteins in *Tetrahymena* mt genome into one of the two clusters according to the frequencies

of adenine and thymine at site 2 of their codons.

Codon usage of Tetrahymena and Paramecium mt genes.

We analyzed the codon usage of genes common to *Tetrahymena* and *Paramecium* mt genomes and demonstrated the results in table 5. The average values are used for each genus and they indicate significant differences in codon usage preferences. ENC measure [30] quantifies the "effective" number of codons that are used in a gene. For a genetic code, the value of ENC ranges from 20 (only one codon is used for each amino acid; i.e., the codon bias is maximum) to 61 (all synonymous codons for each amino acid are equally used; i.e., no codon bias). CBI [20] is a measure of the deviation from the equal use of synonymous codons. CBI values range from 0 (uniform use of synonymous codons) to 1 (maximum codon bias). The calculated ENC and CBI parameters reveal significantly (see ANOVA p-values) greater biased codon usage in *Tetrahymena* than in *Paramecium* mt genomes (Table 5). Also, the Chi2 values, which are the "scaled" 2 measures based on the chi square statistics; i.e., based on the difference between the observed number of codons and those expected from equal usage of codons (Shields et al. 1988), indicate radical deviation of codon usage in *Tetrahymena* (Table 4). The relationship between GC3 and ENC was analyzed to determine whether the codon usage differences between *Tetrahymena* and *Paramecium* mt genomes were related to the differences in GC contents at the third codon position (Nc plot, [25]). For comparison simplicity the values for both genome are incorporated in one Nc plot. The patterns of Nc plot are different for proteins in these two ciliate mt genomes (Figure 4A). Proteins in *Tetrahymena* mitochondria appear to cluster around 5%

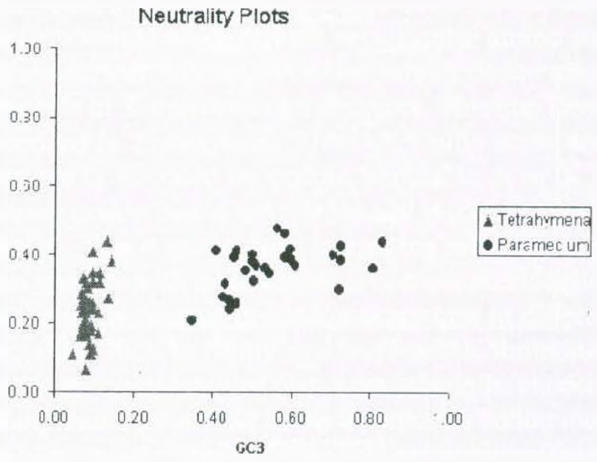

Fig. 2. Neutrality plot (GC12 vs GC3) for *Tetrahymena* and *Paramecium* mt genes. The regression line for *Tetrahymena* is y=1.914x + 0.064, R2=0.22, and for *Paramecium* is y=0.251x + 0.221, R2=0.21.

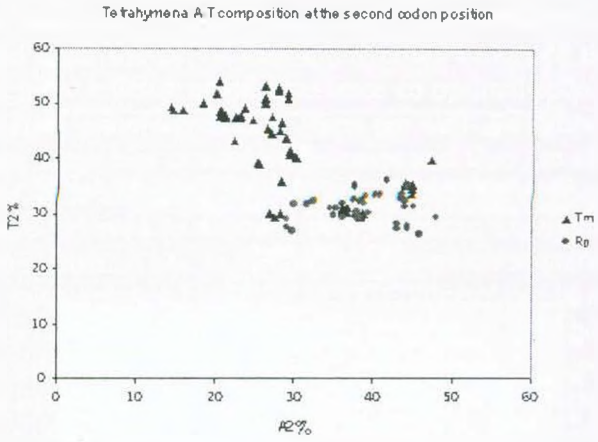

Fig. 3. The occurrence frequency of bases A and T at the second codon position for proteins in Tetrahymena mitochondrial genome.

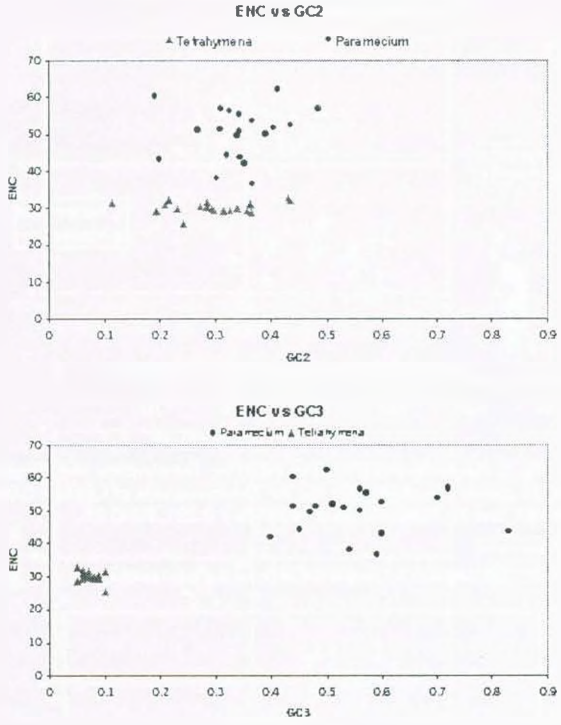

Fig. 4. Relations between GC content and ENCs (Nc plot). A) ENCs are plotted against GC3. B) ENCs are plotted against GC2,

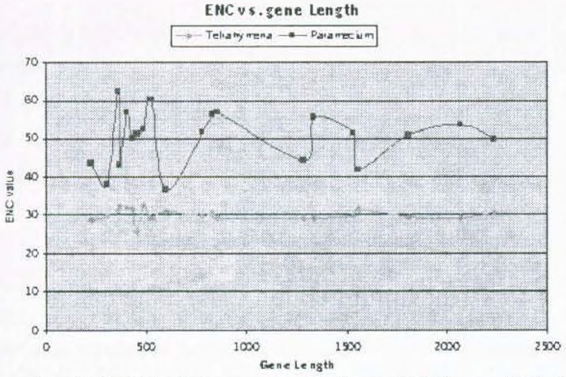

Fig. 5. Relations between gene lengths and ENCs are shown for *Tetrahymena* and *Paramecium* proteins.

to 10% GC3, whereas their *Paramecium* counterparts show a wide distribution of 40% to 80%. Unlike GC3, when the ENC is compared with GC2, I observe that GC contents at the second codon position have a fairly similar distribution (Figure 4B). This indicates that the GC content at the third codon position is more likely to be responsible for a biased codon usage in *Tetrahymena*. Also, the relationship between ENCs and the gene lengths were analyzed to determine whether the length of the proteins affect the codon choices. Figure 5 demonstrates these relationships for *Tetrahymena* and *Paramecium* mt genomes. It is quite clear that in *Tetrahymena* mt genome the ENCs, which have a narrow distribution, have no correlation with gene lengths suggesting no apparent role for protein lengths in determining the codon choices. Although the ENCs for *Paramecium* proteins vary relatively more and swing between 40 and 60, the gene length is probably not affecting the codon usage since there are short genes (length < 500 bp) with high and long genes (length > 1500 bp) with low ENC values (Figure 5).

It is widely accepted that codon usage in most organisms is controlled by a complex set of rules. Majority of the studies on codon preferences are on nuclear genomes, and consequently the rules that direct the codon usage are shown to be applicable to nuclear genomes. In this project I have identified codon usage bias in mitochondrial genomes of six different species of ciliates *Tetrahymena* and *Paramecium* (see materials and methods) to reveal insights in how mitochondria define their codon usage. The mitochondrial genomes of these two genera share

four tRNAs with identical anticodon sequences. They also code for the same group of transmembrane and ribosomal proteins with very similar amino acid compositions. Although such similarities may suggest a comparable codon usage, actual results indicate a strong bias in *Tetrahymena* mt genomes. This bias results in exclusive use of AT rich codons in *Tetrahymena* where all amino acids, with the exception of Phe, are primarily encoded by codons ending in A or T. In contrast, *Paramecium* mt genome codon bias is less intense and is geared towards balanced use of GC rich codons. This is particularly interesting since mt genes from these two ciliates express proteins with quite similar sequences and identical functions. Although there are various reasons for codon usage bias in different organisms and gene families, a few such as highly expressed genes, mutational pressure, and tRNA availability are considered to play a major role in choosing the optimal codons. Rapidly expressed genes tend to use a particular codon more often than the other synonymous one(s), which allows them to supply sufficient copies of metabolically important proteins in shorter periods of time. In these genes codons may be under or over represented resulting in a different codon usage. Also, in some organism codon preferences may follow mutational pressure unique to that organism and consequently prefer codons with higher contents of a particular base(s). In several cases the tRNA availability could influence the codon choices and provide a more efficient translation since an optimal codon-anticodon pairing may accelerate protein synthesis. These are well- established reasons responsible for codon usage bias in nuclear genomes and they are also very likely to cause preferential codon usage in mt genomes. Ciliates *Tetrahymena* and *Paramecium* mt genes are believed to follow a poly-cystronic expression suggesting a single transcription initiation site. In fact, using comparative sequence analysis in five *Tetrahymena* mitochondria I have found a GC control box in an intergenic region where bidirectional transcription initiates (data not published). Therefore, expression levels may not be the major factor to generate a biased codon usage in *Tetrahymena* since I suppose that the entire mt genes may transcribe in one step. Also, since *Paramecium* mt genes are believed to go through the same poly-cystronic expression and have a single transcription initiation site, they too, should have an extreme preferential usage of codons. However this is not the case and *Paramecium* mitochondrion, which has a fairly balanced codon usage. An AT mutational pressure is quite clear in *Tetrahymena* mt genome, which reflects in its codon preferences where majority of the used codons are AT rich or end in one of these bases. There could be several reasons for this pressure such as different DNA polymerase activity, selection, and so on. Yet since it is shown that *Tetrahymena* mitochondria should import tRNA from cytosol [24], tRNA availability becomes a significant candidate responsible for codon usage bias in this organism and should not be ignored.

REFERENCES

1. *Akashi H.* Genetics 136: 927-935, 1994.
2. *Bennetzen J.L., Hall B.D.* J. Biol. Chem. 257: 3026-3031, 1982.

3. *Brunk C.F., Hanawalt P.C.* Experimental Cell Research 54:143-149, 1969.
4. *Burger G., Zhu Y., Littlejohn Tim G., Greenwood S.J., Schnare M.N., Lang B.F., Gray M.W.* J. Mol. Biol. 297(2). March 24, 365-380, 2000.
5. *Chen S.L., Lee W., Hottes A.K., Shapiro L., McAdams H.H.* Proc. Natl. Acad. Sci. U S A. Mar 9; 101 (10) 3480-5, 2004.
6. *Chiapello H., Ollivier E., Lande's-Devauchelle C., Nitschker P., Risler J.-L.* Nucleic Acids Res. 27, 2848–2851, 1999.
7. *Duret L., Mouchiroud D.* Proc. Natl. Acad. Sci. 96:4482–7, 1999.
8. *Fuglsang A.* Strong associations between gene function and codon usage. APMIS 111: 843-847, 2003.
9. *Grantham R., Gautier C., Gouy M., Jacobzone M., Mercier R.* Nucleic Acids Res. Jan 10;9(1):r43-74, 1981.
10. *Ikemura T.* J. Mol. Biol. Feb 15;146(1):1-21, 1981.
11. *Jukes T.H., Bhushar V.* J. Mol. Evol. 24: 39-44, 1986.
12. *Kanaya S., Kinouchi M., Abe T., Kudo Y., Yamada Y., Nishi T., Mori H., Ikemura T.* Gene. Oct 3; 276 (1-2) 89-99, 2001.
13. *Kano A., Andachi Y., Ohama T., Osawa S.* J. Mol. Biol. Sep 20;221(2):387-401, 1991.
14. *Karlin S., Mrazek J., Campbell A.M.* Mol. Microbiol. 29, 1341–1355, 1998.
15. *Knight R.D., Landweber L.F., Yarus M.* J. Mol. Evol. Oct-Nov; 53(4-5):299-313, 2001.
16. *Marck C.* Nucleic Acids Res. 1988 Mar 11;16(5):1829-36, 1988.
17. *Medigue C., Rouxel T., Vigier P., Henaut A., Danchin A.* J. Mol. Biol. 222, 851–856, 1991.
18. *Moriyama E.N., Hartl D.L.* Genetics. Jul; 134 (3):847-58, 1993.
19. *Moriyama E.N., Powell J.R.* Nucleic Acids Res;26:3188–93, 1998.
20. *Morton B.R.* J. Mol. Evol. 37:273–280, 1993.
21. *Oresic M., Shalloway D.* J. Mol. Biol. 281: 31–48, 1998.
22. *Osawa S., Ohama T., Yamao F., Muto A., Jukes T.H., Ozeki H., Umesono K.* Proc Natl Acad Sci U S A. Feb;85(4):1124-8, 1988.
23. *Rozas J., Rozas R.* Comput. Applic. Biosci. 13: 307-311, 1997.
24. *Rusconi C.P., Cech T.R.* J. 1996 Jul 1;15(13):3286-95, 1996A.
25. *Sharp P.M., Li W.H.* J. Mol. Evol. 24: 24-38, 1986.
26. *Sharp P.M., Lloyd A.T.* Nucleic Acids Res. 21, 179–183, 1993.
27. *Sueoka N.* Proc. Natl. Acad. Sci. U S A. Apr;85(8):2653-7, 1988.
28. *Sueoka N.* J. Mol. Evol. Jul;49(1):49-62, 1999.
29. *Tao X., Dafu D.* FEBS Lett. 434:93–6, 1998.
30. *Wright F.* Gene. Mar 1; 87 (1): 23-9, 1990.

Поступила 25.VII.2005