

Codon bias and gene expression of mitochondrial ND2 gene in chordates

Arif Uddin, Tarikul Huda Mazumder, Monisha Nath Choudhury & Supriyo Chakraborty*

Department of Biotechnology, Assam University, Silchar-788011, Assam, India; Supriyo Chakraborty - Email: supriyoch_2008@rediffmail.com; *Corresponding author

Received August 05, 2015; Accepted August 24, 2015; Published August 31, 2015

Abstract:

Background: Mitochondrial *ND* gene, which encodes NADH dehydrogenase, is the first enzyme of the mitochondrial electron transport chain. Leigh syndrome, a neurodegenerative disease caused by mutation in the *ND2* gene (T4681C), is associated with bilateral symmetric lesions in basal ganglia and subcortical brain regions. Therefore, it is of interest to analyze mitochondrial DNA to glean information for evolutionary relationship. This study highlights on the analysis of compositional dynamics and selection pressure in shaping the codon usage patterns in the coding sequence of *MT-ND2* gene across pisces, aves and mammals by using bioinformatics tools like effective number of codons (ENC), codon adaptation index (CAI), relative synonymous codon usage (RSCU) etc. **Results:** We observed a low codon usage bias as reflected by high ENC values in *MT-ND2* gene among pisces, aves and mammals. The most frequently used codons were ending with A/C at the 3rd position of codon and the gene was AT rich in all the three classes. The codons TCA, CTA, CGA and TGA were over represented in all three classes. The F1 correspondence showed significant positive correlation with G, T3 and CAI while the F2 axis showed significant negative correlation with A and T but significant positive correlation with G, C, G3, C3, ENC, GC, GC1, GC2 and GC3. **Conclusions:** The codon usage bias in *MT-ND2* gene is not associated with expression level. Mutation pressure and natural selection affect the codon usage pattern in *MT-ND2* gene.

Key words: Codon usage, *MT-ND2* gene, natural selection, mutation pressure

Background

The mitochondrial genome is ideal as the molecular marker for species identification as well as systematic phylogenetic studies due to its small size. It is easily amplified and mostly conserved in gene content and characterized by lack of recombination, maternal inheritance and high evolutionary rate [1]. The respiratory chains of mitochondrial genome comprise of four complexes (complex, I-IV) and are encoded by 37 genes consisting of two ribosomal RNA (rRNA), twenty-two transfer RNA (tRNA) and thirteen protein coding genes. The complex-I of mitochondrial respiratory chain includes the first enzyme NADH dehydrogenase and its seven subunits (ND1-6 & ND4L) play a pivotal role in diverse pathological processes [2]. The subunit 2 of NADH dehydrogenase is encoded by *ND2* gene and its function is not yet fully understood. However, literature suggests that a mutation in the *ND2* (T4681C) gene was found in patients with Leigh syndrome, a neurodegenerative disease

characterized by bilateral symmetric lesions in basal ganglia and subcortical brain regions [3].

Urrutia and Hurst (2003) reported that the codon usage in human is positively related to gene expression but is inversely related to the rate of synonymous substitution [4]. Several genomic factors such as gene expression level, protein secondary structure, and translational preferences balancing between the mutational pressure and natural selection contribute to the synonymous codon usage variation in different organisms [5,6]. Therefore, gaining the information on the synonymous codon usage pattern provides significant insights pertaining to the prediction, classification, and evolution of a gene at molecular level and also helps in designing highly expressed genes. In the present study we have carried out a comparative analysis of the *ND2* gene codon usage and codon context patterns among the mitochondrial

genomes of three chordate classes (pisces, aves and mammals) in order to understand the molecular mechanism along with functional conservation of gene expression during the period of evolution using several bioinformatics tools.

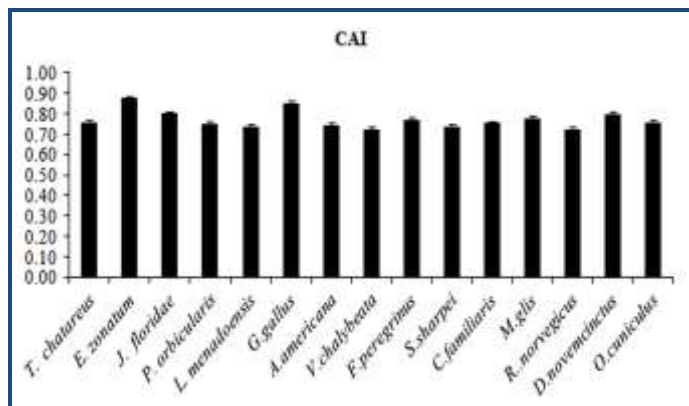


Figure 1: Distribution of CAI in MT-ND2 gene among different species

Methodology:

Retrieval of Sequence data

The coding sequences (cds) of MT-ND2 gene from five species of pisces, aves and mammals each were retrieved from National Center for Biotechnology Information, USA (<http://www.ncbi.nlm.nih.gov/>) using the following accession numbers. The accession numbers of different species are AP006806, AP006813, AP006778, AP006825, AP006858, X52392, AF090337, AF090341, AF090338, AF090340, U96639, AJ001562, X14848, Y11832 and AJ001588. A perl programme was used to analyze the compositional features and codon usage bias parameters.

Compositional properties

The overall composition of A, T, G, C bases and its composition at 3rd position along with GC, GC1, GC2 and GC3 contents were calculated using the perl script.

Codon adaptation Index (CAI)

Codon adaptation index (CAI) is used to estimate gene expression level. The CAI is calculated as

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

Where, ω_k is the relative adaptiveness of the k_{th} codon and L is the number of synonymous codons in the gene [7].

Effective Number of Codons (ENC)

The effective number of codons (ENC) is the most extensively used parameter to measure the usage bias of the synonymous codons [8]. The ENC value ranges from 20 (when only one codon is used for each amino acid) to 61 (when all codons are used randomly). It is calculated as:

$$ENC = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

Where $F_k(k= 2,3,4,6)$ is the mean of F_k values for the k-fold degenerate amino acids.

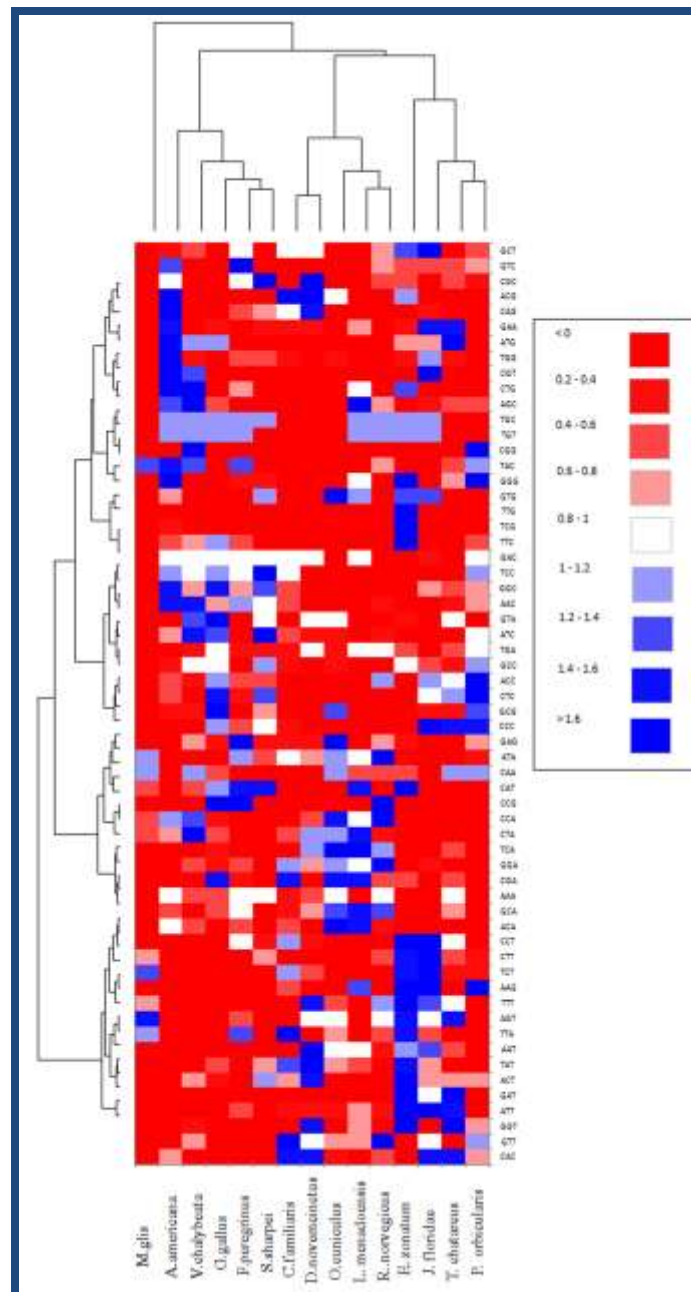


Figure 2: Hierarchical clustering of RSCU value in different species of MT-ND2 gene

Relative Synonymous Codon Usage (RSCU)

Relative synonymous codon usage was calculated as the ratio of the observed frequency of a codon to its expected frequency if all the synonymous codons of a particular amino acid are used equally [9]. The RSCU value is calculated using the formula

$$RSCU_{ij} = \frac{X_{ij}}{\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}}$$

where, X_{ij} is the frequency of occurrence of the j_{th} codon for i_{th} amino acid (any X_{ij} with a value of zero is arbitrarily assigned a

We calculated the codon adaptation index (CAI) values for *MT-ND2* gene in order to find out the expression level among pisces, aves and mammals (Figure 1). In our analysis, the CAI values were (Mean±SD) 0.7851±0.05, 0.7667±0.05, 0.7635±0.02 in pisces, aves and mammals, respectively. We used unpaired t test between pisces and aves as well as between pisces and mammals but the difference was not statistically significant. Wei *et.al* 2014, also reported the average value of CAI in mitochondrial protein coding genes ranged from 0.5-0.7 in *B.mori* [16]. In addition, we performed a correlation analysis between ENC and gene expression level as measured by CAI and found no significant relationship suggesting that the codon usage bias in *MT-ND2* gene is not associated with expression level among the three classes.

Moreover, we calculated the relative synonymous codon usage (RSCU) values in the coding sequences of *MT-ND2* gene among pisces, aves and mammals Table 3 (see supplementary material). In our analysis, the RSCU value > 1 means the codon is more frequently used, RSCU value <1 as less frequently used codon. But RSCU value >1.6 means over represented codon whereas the RSCU value <0.06 as under represented codon. The RSCU values of 60 codons also indicated that the codon usage bias of *MT-ND2* gene is low. Approximately half of the codons were frequently used i.e. 30 codons in pisces, 28 in aves and 21 in mammals. In all three classes the most frequent codons ended with A or C at the 3rd codon position (A/T ended: G/C ended = 16:14 in pisces, 13: 15 in aves and 14: 7 in mammals). The RSCU values were analyzed using heat map (Figure 2) which represented the more frequently, less frequently, over-represented and under-represented codons. The results showed that the most preferred codons were TTC, CTA and TAC in all the three classes. In addition, the four codons namely TCA, CTA, CGA and TGA encoding the amino acid serine, leucine, arginine and tryptophan respectively were over represented in *MT-ND2* gene among pisces, aves and mammals. This result suggests that compositional features played an important role in codon usage in *MT-ND2* gene [13].

The overall percentage of GC contents at different codon positions were calculated (see supplementary material Table S2). In order to find out the role of mutation pressure and natural selection, we constructed a neutrality plot of GC12 against GC3 (Figure 3, a-c) [17]. The linear regression coefficient of GC12 on GC3 indicated that natural selection plays a major role while mutation pressure plays a minor role in shaping the codon usage patterns in *MT-ND2* gene. Our result was similar to the findings of Wei *et.al* (2014) in the mitochondrial DNA codon usage analysis of *B.mori* [16].

We performed correspondence analysis (CoA) based on RSCU values to analyze the codon usage variation in *MT-ND2* gene among pisces, aves and mammals. In our analysis, the 1st axis (F1) accounted for 34.50% of the total variation and the 2nd axis

accounted for 12.51% of the total variation (Figure 4). Further, correlation analysis was done to determine the interrelationships between the first two principle axes (F1 and F2), nucleotide constraints and indices of natural selection (CAI, Gravy, Aromo) on *MT-ND2* gene. The F1 axis showed significant positive correlation with G, T3 and CAI whereas the F2 showed significant positive correlation with G, C, G3, C3, ENC, GC and GC1-3 but significant negative correlation with A and T Table 4 (see supplementary material). These results suggest both compositional constraint under mutation pressure and natural selection affect the codon usage pattern in *MT-ND2*. The results were similar to the findings of Butt *et.al*. [18].

Conclusion:

The codon usage bias in *MT-ND2* gene is weak with high expression level. It is found that natural selection and mutation pressure affect the codon usage pattern in *MT-ND 2* gene.

Acknowledgment:

We are thankful to Assam University, Silchar, Assam, India, for providing the necessary infrastructural facilities for this work.

References:

- [1] Modica-Napolitano JS & Singh KK, *Mitochondrion*. 2004 **4**: 755 [PMID: 16120430]
- [2] Bellance N *et al.* *Front Biosci*. 2009 **14**: 4015 [PMID: 19273331]
- [3] Ugalde C *et al.* *Mol Genet Metab*. 2007 **90**: 10 [PMID: 16996290]
- [4] Urrutia AO & Hurst LD, *Genome Res*. 2003 **13**: 2260 [PMID: 12975314]
- [5] Sharp PM *et al.* *Biochem Soc Trans*. 1993 **21**: 835 [PMID: 8132077]
- [6] Stenico M *et al.* *Nucleic Acids Res*. 1994 **22**: 2437 [PMID: 8041603]
- [7] Sharp PM *et al.* *Nucleic Acids Res*. 1986 **14**: 5125 [PMID: 3526280]
- [8] Wright F, *Gene*. 1990 **87**: 23 [PMID: 2110097]
- [9] Gupta SK & Ghosh TC, *Gene* 2001 **273**: 63 [PMID: 11483361]
- [10] Kyte J & Doolittle RF, *J Mol Biol* . 1982 **157**: 105 [PMID: 7108955]
- [11] Lobry JR & Gautier C, *Nucleic Acids Res*. 1994 **22**: 3174 [PMID: 8065933]
- [12] Shields DC & Sharp PM, *Nucleic Acids Res*. 1987 **15**: 8023 [PMID: 3118331]
- [13] Zhang Z *et al.* *PloS one*. 2013 **8**: e81469 [PMID: 24303050]
- [14] Jia X *et al.* *BMC Genomics*. 2015 **16**: 356 [PMID: 25943559]
- [15] Peixoto L *et al.* *Parasitology*. 2004 **128**: 245 [PMID: 15074874]
- [16] Wei L *et al.* *BMC Evol Biol*. 2014 **14**: 262 [PMID: 25515024]
- [17] Sueoka N, *Proc Natl Acad Sci U S A*. 1988 **85**: 2653 [PMID: 3357886]
- [18] Butt AM *et al.* *PloS one*. 2014 **9**: e90905 [PMID: 24595095]

Edited by P Kanguane

Citation: Uddin *et al.* *Bioinformation* 11(8): 407-412 (2015)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.

Supplementary material:

Table 1: Nucleotide composition in MT-ND2 gene

Species	A %	T %	G %	C %	A3 %	T3 %	G3 %	C3 %
<i>T. chatareus</i>	29.2	24.9	11.4	34.5	41	19.2	3.7	36.1
<i>E. zonatum</i>	28.7	28.7	13.8	28.8	37.8	27.8	8.9	25.5
<i>J. floridae</i>	25.2	28.9	11.6	34.3	29.8	28.1	4.9	37.2
<i>P. orbicularis</i>	26.3	22.9	11.5	39.3	33.8	15.2	4.3	46.7
<i>L. menadoensis</i>	38.2	22.6	10.8	28.4	55.9	13.5	5.4	25.2
Mean±SD	29.52±5.12	25.60±±3.05	11.82±1.14	33.06±4.53	39.66±10	20.76±6.88	5.44±2.03	34.14±9.02
<i>G.gallus</i>	32.6	23	8.6	35.8	42.4	8.9	4.3	44.4
<i>A.americana</i>	29.1	21.8	13.1	36	38	9.2	10.7	42.1
<i>V.chalybeata</i>	31	23	11.9	34.1	45.5	11	6.1	37.4
<i>F.peregrinus</i>	32.2	25.1	9.4	33.3	43.5	14.1	4.6	37.8
<i>S.sharpei</i>	29.7	24.5	8.7	37.1	37.8	15.6	3.2	43.4
Mean±SD	30.92±1.52	23.48±1.31	10.34±2.04	35.26±1.53	41.44±3.41	11.76±2.98	5.78±2.93	41.02±3.22
<i>C.familiaris</i>	35.4	28	9.2	27.4	46	18.1	5.2	30.7
<i>M.glis</i>	33.2	30.1	9.3	27.4	42.7	23.1	3.5	30.7
<i>R.norvegicus</i>	36	25	7.9	31.1	47.1	15.9	2	35
<i>D.novemcinctus</i>	39.1	26.1	8.3	26.5	45.7	22.7	4	27.6
<i>O.cuniculus</i>	35	28.3	9.1	27.6	54.6	14.4	4.6	26.4
Mean±SD	35.74±2.14	27.50±±1.99	8.76±0.62	28±1.78	47.22±4.43	18.84±3.93	3.86±1.21	30.08±3.34

Table 2: GC contents, ENC and CAI values

Species	ENC	GC %	GC1 %	GC2 %	GC3 %	CAI
<i>T. chatareus</i>	58	45.8	52.1	45.6	39.8	0.7587
<i>E. zonatum</i>	55	42.7	48.1	45.6	34.4	0.8768
<i>J. floridae</i>	59	45.8	50.1	45.3	42.1	0.8011
<i>P. orbicularis</i>	60	50.8	55.9	45.6	51	0.7493
<i>L. menadoensis</i>	53	39.2	43.6	43.3	30.7	0.7396
Mean±SD	57±2.91	44.86±4.29	49.86±4.57	45.08±1.00	39.60±7.78	0.7851±0.05
<i>G.gallus</i>	60	44.5	41.8	42.9	48.7	0.8546
<i>A.americana</i>	60	49.1	51.9	42.7	52.7	0.7433
<i>V.chalybeata</i>	60	46	51.9	42.7	43.5	0.7251
<i>F.peregrinus</i>	59	42.7	41.5	44.4	42.4	0.7733
<i>S.sharpei</i>	60	45.8	46.7	44.1	46.7	0.7372
Mean±SD	59.80±0.44	45.62±2.34	46.76±5.12	43.36±0.82	46.80±4.14	0.7667±0.05
<i>C.familiaris</i>	56	36.6	36.2	37.6	35.9	0.7544
<i>M.glis</i>	55	36.7	36.3	39.5	34.3	0.7799
<i>R.norvegicus</i>	57	39	40.8	39.3	37	0.7247
<i>D.novemcinctus</i>	54	36.8	40.8	37.9	31.6	0.7988
<i>O.cuniculus</i>	53	34.8	35.6	37.6	31	0.7597
Mean±SD	55±1.58	36.78±1.49	37.94±2.62	38.38±0.94	33.96±2.61	0.7635±0.02

Table 3: Preferred codons of each amino acid in pisces, aves and mammals in MT-ND2

Amino acid	Pisces			Aves		Mammals	
	Codon	Mean	Preferred codons	Mean	Preferred codons	Mean	Preferred codons
Ser	TCA	2.28	TCA	1.956	TCC	2.832	TCA
	TCT	1.032		0.456		0.924	
	TCC	1.476		2.52		1.536	
	TCG	0.68		0.144		0.048	
	AGC	0.888		0.888		0.456	
Phe	AGT	0.168		0.036		0.204	
	TTC	1.648	TTC	1.668	TTC	1.032	TTC
	TTT	0.912		0.332		0.968	
Leu	TTA	0.868	CTA	0.576	CTA	1.104	CTA
	TTG	1.1		0.012		0.048	
	CTA	2.028		2.784		2.88	
	CTC	1.44		1.608		0.972	
	CTG	0.36		0.468		0.168	
Tyr	CTT	1.152		0.54		0.864	
	TAC	1.272	TAC	1.508	TAC	1.172	TAC
	TAT	0.732		0.496		0.828	
Cys	TGT	0.012	TGC	0.016	TGC	0.004	TGC
	TGC	1.2		2		0.4	
Pro	CCA	1.272	CCC	1.88	CCA	2.12	CCA
	CCC	1.8		1.56		1.288	

	CCG	0		0.08		0.04	
	CCT	0.92		0.472		0.568	
His	CAT	0.376	CAC	0.596	CAC	0.248	CAC
	CAC	1.624		1.404		1.752	
Gln	CAA	1.876	CAA	1.588	CAA	1.756	CAA
	CAG	0.124		0.412		0.244	
Arg	CGA	2.88	CGA	2.264	CGA	2.626	CGA
	CGC	0.56		1.064		0.612	
	CGG	0.16		0.2		0	
	CGT	0.4		0.464		0.012	
Met	ATG	0.452	ATA	0.496	ATA	0.156	ATA
	ATA	1.548		1.504		1.844	
Ile	ATC	0.916	ATT	1.388	ATC	0.986	ATC
	ATT	1.084		0.612		0.714	
Thr	ACA	1.312	ACC	1.528	ACC	1.506	ACA
	ACC	1.68		1.648		0.984	
	ACG	0.08		0.104		0.188	
	ACT	0.92		0.72		0.58	
Asn	AAC	1.432	AAC	1.872	AAC	1.16	AAC
	AAT	0.568		0.128		0.54	
Asp	GAC	1.204	GAC	2	GAC	0.8	GAC
	GAT	0.804		0.02		0.008	
Val	GTA	1.408	GTC	1.912	GTA	1.314	GTA
	GTC	1.576		1.664		0.848	
	GTG	0.456		0.288		0.27	
	GTT	0.552		0.136		0.818	
Ala	GCA	1.232	GCC	1.312	GCC	1.288	GCC
	GCC	1.856		1.912		1.332	
	GCG	0.096		0.16		0.048	
	GCT	0.808		0.608		0.574	
Lys	AAA	1.592	AAA	1.936	AAA	1.58	AAA
	AAG	0.408		0.064		0.12	
Glu	GAA	1.764	GAA	1.58	GAA	1.224	GAA
	GAG	0.236		0.42		0.476	
Gly	GGA	1.48	GGA	1.712	GGC	2.088	GGA
	GGC	1.336		1.952		0.882	
	GGG	0.552		0.272		0.056	
	GGT	0.64		0.064		0.232	
Trp	TGG	0.14	TGA	0.288	TGA	0.09	TGA
	TGA	1.86		1.712		1.61	

Table 4: Correlation among the first two principle axes, nucleotide constraints and indices of natural selection in MT-ND2 gene

	A	T	G	C	A3	T3	G3	C3	ENC	GC	GC1	GC2	GC3	CAI	AR	GR
F1	-0.253	0.407	0.544*	-0.249	-0.227	0.587*	0.454	-0.414	-0.271	-0.005	0.135	0.327	-0.268	0.663**	0.175	0.035
F2	-0.559*	-.620*	.522*	0.724**	-0.507	-0.346	0.550*	0.618*	0.751**	0.820**	0.720**	0.625*	0.765**	-0.026	-0.182	-0.066

AR: aromaticity, GR: gravity score