GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Synonymous codon usage patterns in different parasitic platyhelminth mitochondrial genomes

L. Chen[1], D.Y. Yang[1], T.F. Liu[2], X. Nong[1], X. Huang[1], Y. Xie[1], Y. Fu[1], W.P. Zheng[1], R.H. Zhang[1], X.H. Wu[1], X.B. Gu[1], S.X. Wang[1], X.R. Peng[2] and G.Y. Yang[1]

[1]Department of Parasitology, College of Veterinary Medicine, Sichuan Agricultural University, Ya'an, China
[2]Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou, China
[3]College of Life and Basic Science, Sichuan Agricultural University, Ya'an, China

Corresponding author: G.Y. Yang
E-mail: guangyou1963@yahoo.com.cn

Genet. Mol. Res. 12 (1): 587-596 (2013)
Received September 19, 2012
Accepted December 21, 2012
Published February 27, 2013
DOI http://dx.doi.org/10.4238/2013.February.27.8

ABSTRACT. We analyzed synonymous codon usage patterns of the mitochondrial genomes of 43 parasitic platyhelminth species. The relative synonymous codon usage, the effective number of codons (NC) and the frequency of G+C at the third synonymously variable coding position were calculated. Correspondence analysis was used to determine the major variation trends shaping the codon usage patterns. Among the mitochondrial genomes of 19 trematode species, the GC content of third codon positions varied from 0.151 to 0.592, with a mean of 0.295 ± 0.116. In cestodes, the mean GC content of third codon positions was 0.254 ± 0.044. A comparison of the nucleotide composition at 4-fold synonymous sites revealed that, on average, there was a greater abundance of codons ending on U (51.9%) or A (22.7%) than on C (6.3%) or G (19.14%). Twenty-two codons, including UUU, UUA and UUG, were frequently used. In the NC-plot, most of points were distributed well below or around the expected NC curve. In

addition to compositional constraints, the degree of hydrophobicity and the aromatic amino acids also influenced codon usage in the mitochondrial genomes of these 43 parasitic platyhelminth species.

Key words: Platyhelminths; Mitochondrial genomes; Optimal codons; Codon usage bias; Correspondence analysis

INTRODUCTION

Most amino acids are encoded by multiple codons. In most cases, the synonymous codons are not used randomly in different genomes (Grantham et al., 1980; Lloyd and Sharp, 1992). Furthermore, the pattern of codon usage can vary considerably among organisms, and also among genes from the same genome. Codon usage bias among synonymous codons of many genes has been documented in many species. It has been reported that synonymous codon usage bias may be associated with various factors, such as mutational pressure and compositional constraints (Osawa et al., 1988), translational selection (Sharp and Li, 1986), gene function (Chiapello et al., 1998; Ma et al., 2002), mRNA secondary structure (Oresic and Shalloway, 1998; Xie and Ding, 1998), tRNA abundance (Ikemura, 1981), gene length (Moriyama and Powell, 1998; Duret and Mouchiroud, 1999), hydropathy of each encoded protein and degree of amino acid conservation (Romero et al., 2000). It also has been reported that codon usage bias could be influenced by temperature (Sau and Deb, 2009). Among these factors, compositional constraints and translational selection are thought to be the main factors accounting for codon usage variation between genes in different organisms (Karlin and Mrazek, 1996). Analysis of the synonymous codon usage pattern can improve our understanding of the mechanisms of biased usage of synonymous codons (Powell and Moriyama, 1997). The profiles of synonymous codon usage can reveal information about the molecular evolution of individual genes. They can also provide data to train genome-specific gene recognition algorithms, which detect protein-coding regions in uncharacterized genomic DNA (Fickett, 1982).

Mitochondria are the cellular energy providers of all eukaryotic organisms. The mitochondrial genome often accumulates nucleotide substitutions at faster rates than does the nuclear genome, especially for mammalian mitochondrial DNA (Brown et al., 1979; Martin et al., 1992; Ramirez et al., 1993). Various studies have analyzed different organisms, such as viruses (Zhang et al., 2011), bacteria (Luo et al., 2011), yeast (Whittle et al., 2011), plants (Liu et al., 2010), and animals (Musto et al., 2001; Fadiel et al., 2002). However, codon usage in platyhelminth mitochondrial genomes has only rarely been studied. In this study, we analyzed the codon usage bias in 43 published mitochondrial genome sequences of parasitic platyhelminths using multivariate statistical analysis and correlation analysis. Knowledge of the codon usage patterns may therefore provide a basis for understanding the mechanisms of biased synonymous codon usage.

MATERIAL AND METHODS

Data

The published mitochondrial genome sequences of 43 parasitic platyhelminths, including 19 Trematoda and 24 Cestoidea, were retrieved from GenBank (Table 1). The annotated coding sequences (CDSs) were extracted with a PERL script. The dataset obtained was

then manually checked to correct existing errors, and the 5'- or 3'-partial CDSs were removed. A total of 479 genes from all 43 mitochondrial genome sequences were selected for this study.

Data analysis

The frequency of 59 codons encoding 18 amino acids (excluding Met, Trp and stop codons) was examined using three different codon indices: relative synonymous codon usage (RSCU), effective number of codons (NC) and frequency of G+C at the third synonymously variable coding position (GC3s). The G+C contents of the first, second and third codon positions (GC1, GC2 and GC3, respectively) were also calculated. RSCU values are calculated by dividing the observed codon usage by the expected rate when all codons for the same amino acid are used equally. If all synonyms encoding this amino acid are used equally, RSCU values are close to 1.0, indicating a lack of bias for that codon. NC is often used to measure the magnitude of codon bias for an individual gene, which is essentially independent of gene length. Values of NC range from 20 (for a gene with extreme bias, where only one codon per amino acid is used) to 61 (for a gene with no bias, where all synonymous codons are used equally). The GC3s value is the frequency of G+C at the third synonymously variable coding position. It is a good indicator of the extent of base composition bias, and the expected NC values from GC3s were computed according to Wright (1990). The CodonW 1.4.2 program was used for calculating the indices of codon usage, and SPSS12.0 and Excel 2007 software packages were implemented for statistical analysis.

Correspondence analysis (COA) of codon usage

COA is the most popular multivariate analysis method for codon usage analysis. It can be used to determine the major variation trends using these RSCU values and genes ordered according to their positions along the major axis.

RESULTS

GC content variation in 43 parasitic platyhelminth mitochondrial genomes

Taenia pisiformis had the smallest of the 43 mitochondrial genomes, with only 13,387 bp. The largest mitochondrial genomes were of *Schistosoma haematobium* and *S. spindale*, with up to 16,901 bp. Most of the genomes were between 13.0 and 15.0 kb in length. The complete mtDNA sequences of cestodes were 13.3-14.2 kb in length and highly compact. Fluke mt genome sizes ranged from 13.9 to 16.9 kb with an AT richness of 60-70%, except for *Paragonimus westermani* with only 51.5% AT content (Table 1).

Base composition has been found to influence both codon usage and gene function (Fadiel et al., 2001). General information about codon usage and gene function, including the nucleotide composition of the 43 genomes, is summarized in Table 1. The mean GC content of the third codon position (GC3) was 0.272, indicating that the A+T content was higher than G+C content. Among the 19 trematode species, GC3 varied from 0.151 in *Benedenia seriolae* minor to 0.592 in *P. westermani*, with a mean of 0.295 ± 0.116 (Table 2). In cestodes, the mean GC3 was 0.254 ± 0.044, with a smaller range than in Trematoda. The mean NC of Trematoda was 40.6 ± 7.0, higher than 37.0 ± 7.0 in Cestoidea.

Table 1. General feature of 43 parasitic platyhelminth mitochondrial genomes.

Name of species (strains)	Size (bp)	Accession No.	CDS	Codons	GC1	GC2	GC3
Trematoda							
Clonorchis sinensis	13,875	FJ381664	12	3403	0.433	0.367	0.376
Clonorchis sinensis (China strain)	13,879	JF729303	10	2948	0.441	0.367	0.377
Fasciola hepatica	14,462	NC_002546	11	3277	0.395	0.366	0.334
Opisthorchis felineus	14,277	NC_011127	11	2980	0.428	0.367	0.391
Paragonimus westermani	14,965	NC_002354	11	3224	0.476	0.388	0.592
Schistosoma haematobium	16,901	NC_008074	11	3002	0.299	0.315	0.202
S. japonicum	14,085	NC_002544	11	2860	0.314	0.311	0.217
S. mansoni	14,415	NC_002545	11	2834	0.344	0.319	0.263
S. mekongi	14,072	NC_002529	11	2882	0.306	0.308	0.187
S. spindale	16,901	NC_008067	12	3364	0.300	0.314	0.185
Trichobilharzia regenti	14,838	NC_009680	12	3347	0.344	0.336	0.269
Benedenia hoshinai	13,554	NC_014591	12	3320	0.295	0.304	0.192
B. seriolae	13,498	NC_014291	12	3315	0.283	0.307	0.151
Gyrodactylus derjavinoides	14,741	NC_010976	10	2902	0.347	0.330	0.298
G. salaries	14,790	NC_008815	11	3029	0.395	0.345	0.448
G. thymalli	14,788	NC_009682	11	3029	0.393	0.343	0.443
Microcotyle sebastis	14,407	NC_009055	8	1690	0.307	0.316	0.243
Pseudochauhanea macrorchis	15,031	NC_016950	12	3410	0.332	0.336	0.227
Polylabris halichoeres	15,527	NC_016057	12	3445	0.324	0.326	0.212
Cestoidea							
Echinococcus granulosus (G1)	13,588	NC_008075	12	3373	0.353	0.331	0.315
E. granulosus (G4/horse strain)	13,598	AF346403	9	1884	0.354	0.313	0.307
E. granulosus (G5/cattle strain)	13,717	NC_011122	12	3362	0.348	0.332	0.298
E. granulosus (G6/camel strain)	13,721	NC_011121	12	3362	0.347	0.332	0.302
E. granulosus (G7/pig strain)	13,719	AB235847	12	3362	0.345	0.331	0.302
E. granulosus (G8/cervid strain)	13,717	AB235848	12	3362	0.346	0.330	0.309
E. multilocularis	13,733	NC_000928	12	3351	0.337	0.331	0.265
E. oligarthrus	13,791	NC_009461	12	3371	0.338	0.329	0.263
E. shiquicus	13,807	NC_009460	12	3368	0.340	0.330	0.285
E. vogeli	13,750	NC_009462	11	2828	0.349	0.329	0.299
Hydatigera taeniaeformis	13,647	NC_014768	9	2553	0.300	0.319	0.189
Taenia (*Multiceps*) *multiceps*	13,693	NC_012894	12	3365	0.310	0.330	0.218
T. asiatica	13,703	NC_004826	11	3064	0.312	0.325	0.215
T. crassiceps	13,503	NC_002547	12	3363	0.285	0.309	0.187
T. hydatigena	13,492	NC_012896	12	3363	0.317	0.323	0.230
T. pisiformis	13,387	NC_013844	9	2860	0.292	0.331	0.178
T. saginata	13,670	NC_009938	12	3368	0.307	0.330	0.211
T. solium	13,709	NC_004022	11	3052	0.310	0.320	0.205
Hymenolepis diminuta	13,900	NC_002767	11	2830	0.299	0.327	0.233
Diphyllobothrium latum	13,608	NC_008945	12	3357	0.349	0.354	0.257
D. nihonkaiense	13,747	NC_009463	9	2724	0.355	0.352	0.278
Spirometra erinaceieuropaei	13,643	NC_011037	10	3026	0.360	0.354	0.283
Diplogonoporus balaenopterae	13,724	NC_017613	11	3140	0.347	0.352	0.239
D. grandis	13,725	NC_017615	11	3140	0.348	0.352	0.236

Genome size, total number of coding sequences (CDS) and codons used, GC content of first, second and third codon positions (GC1, GC2 and GC3) are shown.

Table 2. GC contents of 43 parasitic platyhelminth mitochondrial genomes among different classes.

	GC1	GC2	GC3	NC
Trematoda (19 species or strains)	0.356 ± 0.058	0.335 ± 0.025	0.295 ± 0.116	40.6 ± 7.0
Cestoidea (24 species or strains)	0.331 ± 0.023	0.332 ± 0.012	0.254 ± 0.044	37.0 ± 7.0

NC = effective number of codons.

To compare the GC content and synonymous codon usage bias content between different protein-coding areas in the 43 mitochondrial genomes, the average GC1, GC2, GC3

contents, and NC of protein-coding genes were determined (Table 3). For almost all the genes, the codon GC contents, of the first, second and third codon positions ranged from 0.259 (GC3 of *ATP6*) to 0.425 (GC1 of *Cox3*). Except for *NAD3* and *NAD4L*, all GC1 and GC2 contents of coding genes were higher than GC3. The NC values ranged from 35.86 (*NAD3*) to 40.44 (*NAD5*), which means that *NAD3* had a stronger codon usage bias than other genes.

Table 3. Average GC1, GC2, GC3 contents, and effective number of codons (NC) of 43 parasitic platyhelminth mitochondrial genes.

Genes	CDS	Codons	GC1	GC2	GC3	NC
ATP6	85	7380	0.354	0.315	0.259	38.92
Cytb	83	15088	0.372	0.328	0.284	38.35
Cox1	71	18729	0.371	0.381	0.263	36.31
Cox2	89	8841	0.425	0.328	0.268	39.90
Cox3	77	8441	0.365	0.327	0.269	37.27
NAD1	77	11467	0.338	0.349	0.280	38.46
NAD2	85	12396	0.286	0.314	0.270	38.94
NAD3	75	4344.5	0.298	0.257	0.275	35.86
NAD4	85	17680	0.327	0.339	0.262	39.86
NAD4L	85	3687	0.307	0.249	0.265	36.54
NAD5	83	21805	0.327	0.341	0.275	40.44
NAD6	83	6356	0.289	0.309	0.268	37.87

CDS = coding sequences.

As in almost all genomes, mitochondrial genes do not use synonymous codons with similar frequency. A comparison of nucleotide composition at 4-fold synonymous sites revealed that there was on average a much greater abundance of codons ending in U (51.9%) or A (22.7%) than in C (6.3%) or G (19.14%), reflecting the bias toward U or A in the third position of the codons on the coding strand. At 2-fold synonymous sites, the mean frequency of U was 86.77% in U- or C-ending codons, and the mean frequency of A was 47.91% in A- or G-ending codons (Figure 1; Figure S1; Figure S2).

Preferential codons in the protein-coding genes of 43 mitochondrial genomes

To examine the overall preferential codons in the protein-coding genes of 43 platyhelminth mitochondrial genome, we concatenated all 479 coding sequences and calculated the overall codon bias level. The overall codon usage of 479 coding sequences is presented in Table 4. Twenty-two codons, including UUU, UUA and UUG, were frequently used codons. The preferred terminating codon was UGA. The RSCU values showed that U-ending codons were predominantly used. Most (15/24) of the optimal codons ended with U. This result was consistent with the AU richness of the third codon position in these mitochondrial genomes.

Codon usage analysis

In trematodes and cestodes, almost all 43 genes distributed well below or around the expected NC curve. NC-GC3s plots are shown got four representative taxa, namely *Clonorchis sinensis*, *Benedenia hoshinai*, *Echinococcus granulosus* (*G1*), and *Diphyllobothrium latum* (Figure 2). The NC-GC3s plots of the remaining species are available in Figure S3.

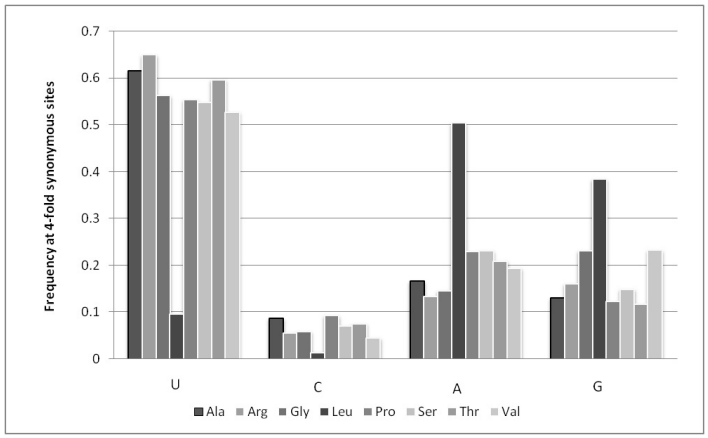

Figure 1. Nucleotide frequencies at 4-fold synonymous site of protein-coding areas in 43 mitochondrial genomes. Each bar represents the average third position nucleotide frequencies for a set of codons encoding the same amino acid (from left to right: Ala, Arg, Gly, Leu, Pro, Ser, Thr, Val). The data presented are based on the protein-coding genes found in all 43 mitochondrial genomes. Nucleotide frequencies at 2-fold synonymous sites are shown as Supplementary material (Figure S1 and Figure S2).

Table 4. Overall codon usage of 43 parasitic platyhelminth mitochondrial genome protein-coding genes.

AA	Codons	N	RSCU	AA	Codons	N	RSCU
Phe	**UUU**	13,707	1.82	Tyr	**UAU**	6,724	1.69
	UUC	1,381	0.18		UAC	1,228	0.31
Leu	**UUA**	9,062	2.64	TER	UAA	225	0.94
	UUG	7,235	2.11		**UAG**	254	1.06
	CUU	1,892	0.55	His	**CAU**	1,746	1.67
	CUC	297	0.09		CAC	346	0.33
	CUA	1,277	0.37	Gln	CAA	466	0.91
	CUG	843	0.25		**CAG**	554	1.09
Ile	**AUU**	5,586	1.79	Asn	**AAU**	2,657	1.71
	AUC	651	0.21		AAC	445	0.29
Met	**AUA**	4,503	1.09	Lys	AAA	1,745	0.95
	AUG	3,784	0.91		**AAG**	1,945	1.05
Val	**GUU**	7,490	2.08	Asp	**GAU**	2,561	1.75
	GUC	660	0.18		GAC	373	0.25
	GUA	2,837	0.79	Glu	GAA	1,050	0.76
	GUG	3,443	0.95		**GAG**	1,705	1.24
Ser	**UCU**	3,903	2.22	Cys	**UGU**	4,112	1.75
	UCC	428	0.24		UGC	585	0.25
	UCA	1,461	0.83	Trp	UGA	1,797	0.96
	UCG	722	0.41		**UGG**	1,929	1.04
Pro	**CCU**	1,678	2.18	Arg	**CGU**	1,399	2.60
	CCC	287	0.37		CGC	121	0.22
	CCA	716	0.93		CGA	278	0.52
	CCG	397	0.52		CGG	357	0.66
Thr	**ACU**	2,315	2.38	Ser	**AGU**	3,802	2.16
	ACC	289	0.30		AGC	556	0.32
	ACA	807	0.83		AGA	1,766	1.00
	ACG	479	0.49		AGG	1,452	0.82
Ala	**GCU**	2,300	2.45	Gly	**GGU**	5,093	2.26
	GCC	331	0.35		GGC	534	0.24
	GCA	618	0.66		GGA	1,267	0.56
	GCG	513	0.55		GGG	2,125	0.94

RSCU = relative synonymous codon usage; AA = amino acids; N = number of codons. The preferentially used codons for each amino acid are displayed in bold.

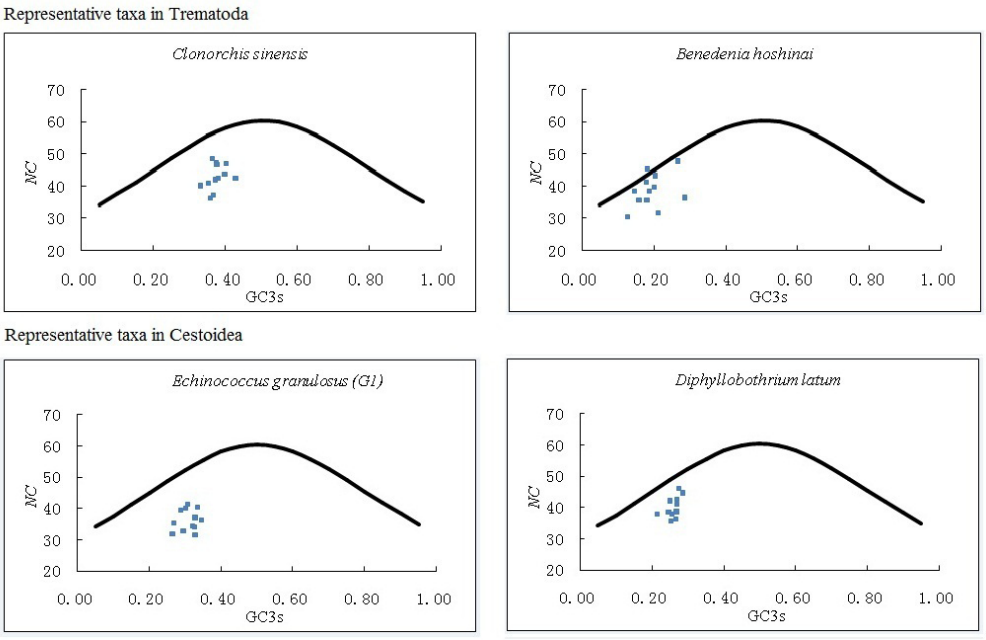

Figure 2. Representative NC-GC3s plots for 43 mitochondrial genes. The NC-GC3s plots of rest mitochondrial genes were available in the Supplementary Material (Figure S3).

COA of codon usage

COA of codon usage was used to determine the major source of variation among the protein-coding genes. Each gene was encoded by a vector of 59 variables, which represents the number of codons with synonyms. COA displayed these genes in a multidimensional space of 59 axes. Among these vectors, the axes representing the most prominent factors contributing to the variation between genes were plotted. The variation in the first and the second dimensions explained 16.6 and 9.9% of the total codon variation, respectively. Most genes near the origins of the axes clustered together to form an ellipsoid cloud in a range of -0.526 to 1.14 for the first axis and -0.550 to 0.862 for the second axis (Figure 3A). In the figure showing the corresponding positions of synonymous codons on these axes (Figure 3B), most of the C/G-ending codons were above the A/U-ending codons. This further supports the view that the nucleotide composition of genes was the basis for the difference in synonymous codon usage between genes.

To explore the correlation between the codon usage bias and the nucleotide composition, we calculated the correlation coefficient between the positions of genes along the first major axis, GC, GC3s, and NC values. The position of each gene along the first axis was strongly correlated with its GC, GC3s, and NC value (r = 0.275, 0.367 and 0.769, respectively). Furthermore, the NC was significantly correlated with GC3s and GC contents (r = 0.469 and 0.408, respectively). The results suggested that compositional constraint was one of the major factors in shaping codon usage in the genomes.

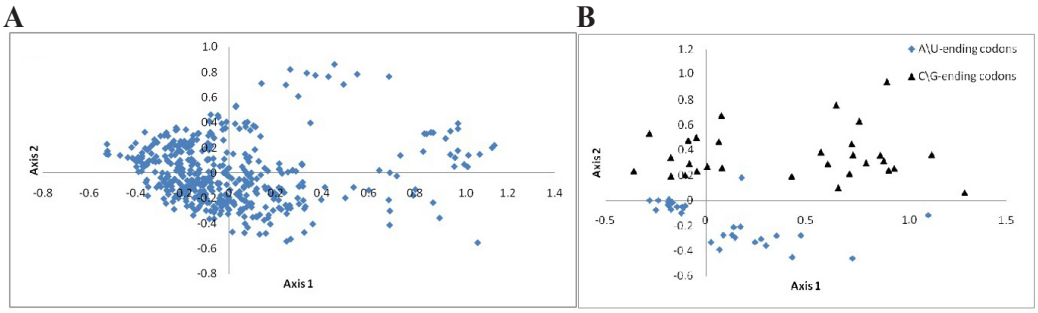

Figure 3. A. Distribution of the protein-coding genes of 43 platyhelminth mitochondrial genomes on the plane corresponds to the coordinates on the first and second axes produced by the correspondence analysis on codon usage. **B.** Distribution of the codons on the plane corresponds to the coordinates along the first and second axes produced by the correspondence analysis on codon usage. Codons ending with A and U are shown as blue squares, and codons ending with G and C are shown as black triangles.

Effect of the hydrophobicity and aromaticity of encoded protein on synonymous codon usage bias

We performed a correlation analysis to investigate whether other factors could explain codon usage. This was done to evaluate whether Gravy and Aromo values were related to NC values. The correlation analyses between the hydrophobicity of each protein and NC value showed that the correlation coefficients ($r = -0.281$, $P < 0.01$) were significantly correlated. The aromaticity of each protein was also significantly correlated with NC ($r = -0.153$, $P < 0.01$). These results indicated that the degree of hydrophobicity and the aromatic amino acids were associated with codon usage variation.

DISCUSSION

Our analysis of codon usage patterns revealed a large number of genes with a high AT content among the protein-coding genes of 43 platyhelminth mitochondrial genomes. This means that the content of A and T nucleotides was higher than the G and C content. Most of the preferred codons also ended in A or T, which is similar to the mitochondrial genomes of many higher plants (Sloan and Taylor, 2010; Wang et al., 2011). However, there are many differences between plant and animal mitochondria, such as their size or the usage of the genetic codes. In this study, the complete mtDNA sequences of cestodes were only 13.3-14.2 kb in size and the fluke mt genome sizes ranged from 13.9 to 16.9 kb, whereas the sizes of higher plant mitochondrial genomes range from 200 to 2400 kb. Mitochondrial protein-coding regions of flatworms are based on translation tables of the genetic codons, while higher plants are based on the universal genetic code.

Originally, the mitochondrion was thought to be a simple bacterium that entered the eukaryotic cell in its early developmental stage. Thus, like other genomes, mitochondria commonly show genome-wide synonymous codon usage bias (Bulmer, 1991). The codon usage bias among species is an important evolutionary phenomenon, which is affected by many factors. Many hypotheses have been put forward to explain the origin of codon usage bias, such as the neutral theory and natural selection theory. If mutation occurred at the third codon

position at a neutral rate, it would result in random synonymous codon choice where GC and AT would be used proportionally among the degenerate codon groups in a gene. In contrast, if translational selection pressure influenced the shaping of codon usage, the bias would be significantly positively correlated to expression levels, and some translation-preferred codons would appear to be used more frequently than others. Natural selection played a crucial role but nucleotide mutational bias and amino acid composition only in a minor way in shaping codon usage in plant mitochondrial genomes (Liu et al., 2004; Zhou and Li, 2009). Complex mutation and weak selection together are thought to determine codon usage in bryophyte mitochondrial genomes (Wang et al., 2010). When 36 different green plant mitochondrial genomes were compared, it was found that plant codon usage in organelle genomes was less likely to be shaped by selection after they had colonized the land (Wang et al., 2011). Complex likelihood-based tests on different models and estimates of the average intensity of selection on synonymous sites suggest that in animal and land plant mitochondria, mutation dominates over selection in shaping codon usage bias (Jia and Higgs, 2008; Sloan and Taylor, 2010). Analyses of mitochondrial DNA sequences from the green algae *Mesostigma viride* (NIES-296) and *Chlamydomonas reinhardtii* (CC-277) suggest that both mutation and selection are important in shaping synonymous codon usage bias in their mitochondrial genomes, with selection being more dominant (Hua and Lee, 2012). In our study, we found that codon usage in parasitic platyhelminth mitochondrial genomes was affected not only by compositional constraints, but also by other factors, such as the degree of hydrophobicity and aromatic amino acids. Compositional constraints and translational selection are thought to be the main factors accounting for codon usage variation between genes in different organisms (Karlin and Mrazek, 1996). Further studies should be performed to investigate the role of translational selection in codon usage bias in the mitochondrial genome of parasitic platyhelminths.

ACKNOWLEDGMENTS

Research supported by the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT) (#IRT0848). We thank Hanmei Liu, Kunfeng Sun, Ning Yan, Jiahai Wang, and Yilan Liang for their help with statistical analysis and also Yun Sun, Xuan Zhou and Ning Wang for comments on the findings.

Supplementary material

REFERENCES

Brown WM, George M Jr and Wilson AC (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 76: 1967-1971.

Bulmer M (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129: 897-907.

Chiapello H, Lisacek F, Caboche M and Henaut A (1998). Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene* 209: GC1-GC38.

Duret L and Mouchiroud D (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 96: 4482-4487.

Fadiel A, Lithwick S, Wanas MQ and Cuticchia AJ (2001). Influence of intercodon and base frequencies on codon usage in filarial parasites. *Genomics* 74: 197-210.

Fadiel AA, Lithwick S and Gamra MM (2002). Codon usage analysis of *Ascaris* species influence of base and intercodon frequencies on the synonymous codon usage. *J. Egypt. Soc. Parasitol.* 32: 625-638.

Fickett JW (1982). Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res.* 10: 5303-5318.
Grantham R, Gautier C, Gouy M, Mercier R, et al. (1980). Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 8: r49-r62.
Hua J and Lee RW (2012). Factors affecting codon bias in the mitochondrial genomes of the streptophyte *Mesostigma viride* and the chlorophyte *Chlamydomonas reinhardtii*. *J. Eukaryot. Microbiol.* 59: 287-289.
Ikemura T (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* 151: 389-409.
Jia W and Higgs PG (2008). Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Mol. Biol. Evol.* 25: 339-351.
Karlin S and Mrazek J (1996). What drives codon choices in human genes? *J. Mol. Biol.* 262: 459-472.
Liu H, He R, Zhang H, Huang Y, et al. (2010). Analysis of synonymous codon usage in *Zea mays*. *Mol. Biol. Rep.* 37: 677-684.
Liu Q, Feng Y and Xue Q (2004). Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*. *Mitochondrion* 4: 313-320.
Lloyd AT and Sharp PM (1992). Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 20: 5289-5295.
Luo XL, Xu JG and Ye CY (2011). Analysis of synonymous codon usage in *Shigella flexneri* 2a strain 301 and other *Shigella* and *Escherichia coli* strains. *Can. J. Microbiol.* 57: 1016-1023.
Ma J, Zhou T, Gu W, Sun X, et al. (2002). Cluster analysis of the codon use frequency of MHC genes from different species. *Biosystems* 65: 199-207.
Martin AP, Naylor GJ and Palumbi SR (1992). Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* 357: 153-155.
Moriyama EN and Powell JR (1998). Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res.* 26: 3188-3193.
Musto H, Cruveiller S, D'Onofrio G, Romero H, et al. (2001). Translational selection on codon usage in *Xenopus laevis*. *Mol. Biol. Evol.* 18: 1703-1707.
Oresic M and Shalloway D (1998). Specific correlations between relative synonymous codon usage and protein secondary structure. *J. Mol. Biol.* 281: 31-48.
Osawa S, Ohama T, Yamao F, Muto A, et al. (1988). Directional mutation pressure and transfer RNA in choice of the third nucleotide of synonymous two-codon sets. *Proc. Natl. Acad. Sci. U. S. A.* 85: 1124-1128.
Powell JR and Moriyama EN (1997). Evolution of codon usage bias in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 94: 7784-7790.
Ramirez V, Savoie P and Morais R (1993). Molecular characterization and evolution of a duck mitochondrial genome. *J. Mol. Evol.* 37: 296-310.
Romero H, Zavala A and Musto H (2000). Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res.* 28: 2084-2090.
Sau K and Deb A (2009). Temperature influences synonymous codon and amino acid usage biases in the phages infecting extremely thermophilic prokaryotes. *In Silico Biol.* 9: 1-9.
Sharp PM and Li WH (1986). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res.* 14: 7737-7749.
Sloan DB and Taylor DR (2010). Testing for selection on synonymous sites in plant mitochondrial DNA: the role of codon bias and RNA editing. *J. Mol. Evol.* 70: 479-491.
Wang B, Liu J, Jin L, Feng XY, et al. (2010). Complex mutation and weak selection together determined the codon usage bias in bryophyte mitochondrial genomes. *J. Integr. Plant Biol.* 52: 1100-1108.
Wang B, Yuan J, Liu J, Jin L, et al. (2011). Codon usage bias and determining forces in green plant mitochondrial genomes. *J. Integr. Plant Biol.* 53: 324-334.
Whittle CA, Sun Y and Johannesson H (2011). Evolution of synonymous codon usage in *Neurospora tetrasperma* and *Neurospora discreta*. *Genome Biol. Evol.* 3: 332-343.
Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29.
Xie T and Ding D (1998). The relationship between synonymous codon usage and protein structure. *FEBS Lett.* 434: 93-96.
Zhang Y, Liu Y, Liu W, Zhou J, et al. (2011). Analysis of synonymous codon usage in hepatitis A virus. *Virol. J.* 8: 174.
Zhou M and Li X (2009). Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. *Mol. Biol. Rep.* 36: 2039-2046.