

RESEARCH ARTICLE

Open Access

Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution

Lei Wei¹, Jian He², Xian Jia¹, Qi Qi¹, Zhisheng Liang¹, Hao Zheng¹, Yao Ping¹, Shuyu Liu² and Jingchen Sun^{1*}

Abstract

Background: Synonymous codon usage bias (SCUB) is an inevitable phenomenon in organismic taxa, generally referring to differences in the occurrence frequency of codons across different species or within the genome of the same species. SCUB happens in various degrees under pressure from nature selection, mutation bias and other factors in different ways. It also attaches great significance to gene expression and species evolution, however, a systematic investigation towards the codon usage in *Bombyx mori* (*B. mori*) has not been reported yet. Moreover, it is still indistinct about the reasons contributing to the bias or the relationship between the bias and the evolution of *B. mori*.

Results: The comparison of the codon usage pattern between the genomic DNA (gDNA) and the mitochondrial DNA (mtDNA) from *B. mori* suggests that mtDNA has a higher level of codon bias. Furthermore, the correspondence analysis suggests that natural selection, such as gene length, gene function and translational selection, dominates the codon preference of mtDNA, while the composition constraints for mutation bias only plays a minor role. Additionally, the clustering results of the silkworm superfamily suggest a lack of explicitness in the relationship between the codon usage of mitogenome and species evolution.

Conclusions: Among the complicated influence factors leading to codon bias, natural selection is found to play a major role in shaping the high bias in the mtDNA of *B. mori* from our current data. Although the cluster analysis reveals that codon bias correlates little with the species evolution, furthermore, a detailed analysis of codon usage of mitogenome provides better insight into the evolutionary relationships in Lepidoptera. However, more new methods and data are needed to investigate the relationship between the mtDNA bias and evolution.

Keywords: *Bombyx mori*, Synonymous codon usage bias, Genomic DNA, Mitochondrial DNA, Evolution

Background

Synonymous codon usage bias (SCUB) refers to the different frequency of synonymous codons in coding DNA. The triplets coding for the same amino acid are not equally used for protein expression, either among different organisms or the genes from a single species. The bias ensures that the most frequently used codons, the optimal codons, can pair with the anticodons of the

most abundant tRNA genes [1]. And it also avoids the mis-incorporation of amino acids to reduce processing errors. It reveals a balance between natural selection (eg. translational selection, gene length, and gene function) and mutation bias (such as GC content and mutation position of base), as well as the influence of random genetic drift [2,3]. Therefore, understanding the codon usage bias can show the codon usage pattern of species, and provide evidence about the evolution of organisms. Many previous studies suggest that there are various factors related to SCUB, determined by mutational bias alone or by both mutation bias and natural selection [4]. For instance, the driving force of the bias in many mammals has been proved to be mutational bias [5], whereas

* Correspondence: cyfz@scau.edu.cn
Equal contributors

¹Subtropical Sericulture and Mulberry Resources Protection and Safety Engineering Research Center, Guangdong Provincial Key Laboratory of Agro-animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China
Full list of author information is available at the end of the article

some others believed that the natural selection determines the bias in eukaryotic organisms [6-8].

The domesticated silkworm, *Bombyx mori* (*B. mori*), which belongs to the Bombycidae family, is a well-studied model species of the Lepidopteran model system with a rich repertoire of genetic information about morphology, development, and behavior [9-11]. Previous studies investigating *B. mori* have primarily focused on the cloning, expression and characterization of some important genes [12-14], as well as applied research such as bioreactor [15] or antivirus [16-18]. In addition, there are also some studies on its evolution, which reveal that the domesticated silkworm derives from Chinese stock rather than Japanese or Korean stock [19-21]. With the completion of the whole genome sequence of *B. mori* [22-24], more molecular genetic resources can be used to illuminate the evolution history.

Originally, the classification of species was based on the observed morphology and behaviors. Later, a new taxonomic method called DNA barcoding was created that uses a short genetic marker in an organisms DNA to identify it as belonging to a particular species [25,26]. However, it has been found that genes coded by genomic DNA (gDNA) seldom fit the standards of DNA barcode when compared to the genes from mitochondrial DNA (mtDNA). As the only genetic material outside gDNA, mtDNA is relatively small in size, extremely easy to amplify [27] and highly conservative, lacks extensive recombination, has maternal inheritance and shows a relatively high evolutionary rate and non-tissue specificity [28,29]. Thus, mtDNA is well used as a molecular marker in species identification, population genetics, systematic phylogeny and evolutionary studies [30-32].

The mtDNA of insects is a self-replicating, circular DNA molecule of about 14-20 kb in size, encoding a conserved set of 37 genes (13 protein genes, 22 tRNA genes, and 2 rRNA genes) and an A + T-rich segment known as control region [33,34]. Currently there are more than 260 species of insects whose mtDNA sequences have been determined (GenBank: www.ncbi.nlm.nih.gov/genbank). Hence, a combination of mtDNA and codon bias in *B. mori* might help us to understand the evolutionary relationships among the Lepidopteran species from another angle.

To test the above possibility, the primary aim of this study is to determine the factors contributing to the codon preference of mtDNA in *B. mori*. The results show that natural selection plays a major role in shaping the codon usage bias, while mutation bias only plays a minor role. Additionally, the evolutionary relationship of the species shown in our study is different from that of the traditional classification. A detailed and modified analysis based on mtDNA bias might offer a better understanding into the evolutionary relationships among the Lepidopteran insects.

Results and discussion

The codon usage pattern of gDNA and mtDNA in *B. mori*

The RSCU values and the number of each codon in the gDNA and mtDNA of *B. mori* are shown in Table 1. The codon usage frequency and the correspondent RSCU values were acquired from the Codon Usage Database (www.kazusa.or.jp/codon). It is found that codon preference happens both in the gDNA and the mtDNA. The GC content of the whole 1180 CDSs (450,043 codons) in the gDNA is 48.12%, while the values for the GC1s, GC2s and GC3s are 52.29%, 41.19%, and 50.89%, respectively. However, the GC content of 48 CDSs (12,937 codons) is 19.01% in mtDNA, and only 7.24% for the GC3s. Clearly, there are multiple differences within the GC contents between the gDNA and the mtDNA. And the frequency of codons ending with A/T is higher than that of G/C in the mtDNA. This may be due to the structure of the AT rich segments, and in turn leads to the higher codon bias.

Normally, if the frequency value of a relative synonymous codon is over 60% or 1.5 times out to the average synonymous codon value of the group, it can be defined as the high frequency codon [35]. In the gDNA, only three high frequency codons (TTC, TAC and AGA) are found. At the same time, the other 19 preferentially used codons are also labeled (Table 1). However, 16 codons with high frequency are observed in the mtDNA. Among them, seven codons show a similar bias level, including AAA (Lys), CCT (Pro), ACA (Thr), GCT (Arg), AGA (Ala), GAA (Glu) and CAA (Gln). In the mtDNA, there are also some codons (eg. CTG, CCG, ACG, GTC, GTG, GCG, CGG, AGG, TCG, AGC and GAC) that are seldom used, and several codons which are never used, such as CTC, TAG and CGC (Table 1).

Our results reveal that the codon bias level of the gDNA is relatively lower than that of the mtDNA in *B. mori*. The high bias of the mtDNA might be derived from the evolution system of mitochondrion. As we know, the evolution patterns of mtDNA and gDNA are not identical. There are two main theories about the endosymbiotic origin of mitochondria at present. One of the theories suggests that the eukaryote engulfed the mitochondrion [36], while the other posits that the prokaryote host acquired the mitochondrion. As a relatively independent organelle inside the cell, early studies used to assume that mtDNA undergoes neutral evolution [37]. However, later findings have denied those findings, and it is now believed that mtDNA suffers from positive and negative selection [38-41]. An abundance of evidences supports the existence of co-evolution and co-adaption between the nuclear and mitochondrial genomes [42-49]. However, different substitution rates have been confirmed between them, where the mtDNA shows a higher rate of nucleotide substitution than that of the gDNA [50]. Thus, the higher rate of

Table 1 The RSCU values and used codon numbers in the gDNA and mtDNA of *B. mori*

AA	Codon	RSCU/number		AA	Codon	RSCU/number	
		gDNA	mtDNA			gDNA	mtDNA
Phe	<i>TTT*</i>	0.78/6894	<i>1.92/1439</i>	Tyr	<i>TAT*</i>	0.78/6250	<i>1.88/641</i>
	TTC	1.22/10823	0.08/63		TAC	1.22/9918	0.12/42
Asn	<i>AAT*</i>	0.90/9230	<i>1.76/790</i>	Cys	<i>TGT*</i>	0.88/3929	<i>1.76/790</i>
	AAC	1.10/11200	0.24/108		TGC	1.12/5094	0.24/108
Lys	AAA*	1.10/15283	<i>1.84/414</i>	Arg	CGT	1.04/4399	0.72/58
	AAG	0.90/12759	0.16/37		CGC	1.14/4711	0.00/0
Leu	<i>TTA*</i>	0.96/5885	<i>5.22/1640</i>	Ser	CGA	0.78/3153	1.26/97
	TTG	1.14/7124	0.36/114		CGG	0.60/2423	0.06/4
	CTT	0.84/5193	0.12/37		AGA	1.50/6325	<i>3.90/301</i>
	CTC	1.04/6570	0.00/0		AGG	0.96/4011	0.06/4
	CTA	0.66/3941	0.30/96		<i>TCT*</i>	1.08/5718	<i>2.70/378</i>
	CTG	1.32/8268	0.00/3		TCC	1.022/5399	0.30/39
	CCT*	1.08/6283	<i>1.80/165</i>		TCA	1.08/5846	2.04/292
Pro	CCC	0.88/5256	0.44/42	TCG	0.90/4969	0.00/2	
	CCA	1.04/6110	1.68/155	AGT	0.84/4632	0.90/128	
	CCG	1.00/5860	0.08/8	AGC	1.02/5618	0.06/9	
	Thr	ACT	1.08/7037	1.40/163	Asp	<i>GAT*</i>	0.94/11523
ACC		1.00/6390	0.28/34	GAC		1.06/12971	0.10/9
ACA*		1.12/7146	<i>2.24/265</i>	Glu		GAA*	1.16/16217
ACG	0.84/5350	0.08/8	GAG		0.84/11781	0.20/24	
Val	<i>GTT*</i>	1.04/7558	<i>2.20/254</i>	Gly	GGT	1.24/9761	<i>1.72/260</i>
	GTC	1.04/7502	0.04/4		GGC	1.08/8443	0.16/22
	GTA	0.76/5588	1.72/195		<i>GGA*</i>	1.24/9636	<i>1.80/275</i>
	GTG	1.20/8735	0.04/5		GGG	0.44/3465	0.32/51
Ala	GCT*	1.40/11368	<i>2.64/226</i>	Ile	<i>ATT</i>	1.02/8391	<i>1.77/1508</i>
	GCC	1.08/8860	0.20/16		ATC	1.11/9321	0.09/80
	GCA	0.80/6556	1.12/97		<i>ATA</i>	0.87/7173	1.14/956
	GCG	0.72/6066	0.04/4		Trp	TGG	1.00/5272
Met	ATG	1.00/10544	1.00/94	Gln		CAA	1.04/8409

Note: The preferentially used codons base on RSCU values for gDNA and mtDNA are in bold and italics, respectively. Asterisk (*) refers to the high frequency codons in mtDNA.

substitution might be related to the higher codon bias in the mtDNA.

Correspondence analysis of codon bias

There are various reasons contributing to codon usage bias, including gene function, compositional constraints of genes, translational selection, the secondary structure of proteins, natural selection, and mutation bias [2,6,8]. Different substitutions suggest that mtDNA has experienced the non-uniform selective pressures [45]. Thus, the correspondence analysis of codons usage may help us to investigate the possible factors that cause the bias

in the mtDNA of *B. mori*. Here, a set of indices were calculated, including G + C content, the content of bases in the third position (GC3s, G3s, T3s, C3s, and A3s etc.), and indices inflecting the codon bias level (eg. CBI, Fop, and ENc), as well as the gene function indices (eg. Gravy and Aromo).

Similar to the results of RSCU, the effective number of codons (ENc) in *B. mori* ranges from 23.14 to 44.00, with an average of 29.47, indicating a high-level codon bias in mtDNA. In addition, most of the tRNAs in mtDNA are known as inactive areas and have almost no ENc value, excluding the genes encoded tRNA-Thr, tRNA-Lys, tRNA-

lle. The absence of values might be a result of the extraordinarily short sequences which are unable to rightly encode enough amino acid, according to the definition of ENc and the calculation of CodonW [4].

Based on the 59 synonymous codons, a set of 37 genes were represented by the points in a super dimensional axes and carried on to the correspondence analysis. Contributions of the axes were shown (Figure 1) and we arrived at a four main contributors which were Axis1, Axis2, Axis3 and Axis4. The first two-main-dimensional coordinates, Axis1 and Axis2 (Figure 2) can explain 12.07% and 8.64% of the total variation, while Axis3 and Axis4 explain 8.31% and 7.58%, respectively, which led to the first axis as the major contributor to the codon bias. The location of the genes marked in the two-dimensional coordinate after the correspondence analysis of 37 genes. The position of the AT ended codons are more close to Axis1 than the GC ending codons with a concentrate distribution (Figure 2), indicating that the base composition for mutation bias might correlate to the codon bias. While the genes with different GC content (CDP average 19.02%, tRNA average 19.61%, rRNA average 16.03%) showed a relatively regular distribution, and lower GC content gene located more close to Axis1 (Figure 3), implying that GC content for mutation pressure probably influenced the bias. In addition, a considerable amount of the genes are in a discrete distribution, indicating that there are many other factors that exist, for example, natural selection.

In order to intuitively display the indices related to the four main contributors, correlations of the important indices were calculated to determine the important factors that result in codon bias (Table 2).

Significant correlations are found in the indices. GC3s shows the significant correlation with Axis1 and ENc ($r = -0.581$, $p < 0.01$; $r = -0.389$, $p < 0.05$) suggesting that the base composition for mutation bias might have an impact on the codon bias. The first axis also shows the significant linear correlation with L_aa ($r = 0.455$, $p < 0.01$). Meanwhile, there is also a significant correlation between L_aa and ENc ($r = 0.649$, $p < 0.01$), revealing that gene length might affect the codon bias too. In addition, L_aa contains a negative correlation with GC3s ($r = -0.545$, $p < 0.01$), but strong positive linear correlation with ENc ($r = 0.649$, $p < 0.01$), which might give an assumption to the usage of synonymous codons suffered from the natural selection in the exact gene.

Natural selection influences the codon bias of mtDNA

A neutrality plot was drawn to estimate the extent of directional mutation pressure against selection in the codon usage bias of mtDNA. The neutrality plot (Figure 4) reveals the results of equilibrium coefficient of mutation and selection. GC12 is the average value of GC1 and GC2. The regression coefficient of GC12 to GC3 of mtDNA is 0.244 0.074, indicating the relative neutrality is 24.40 % while the relative constraint is 75.60 % for GC3. In Figure 2, the points are not located to the diagonal distribution (red line) and the values of GC3 are in a narrow distribution, indicating GC12 and GC3 are definitely not the mutation bias model [25]. On the other hand, the regression curve (green line) tended to be sloped or parallel to the horizontal axis. The subsequent correlation analysis reveals little correlation between GC12 and GC3. Accordingly, mutation bias might just play a minor role in

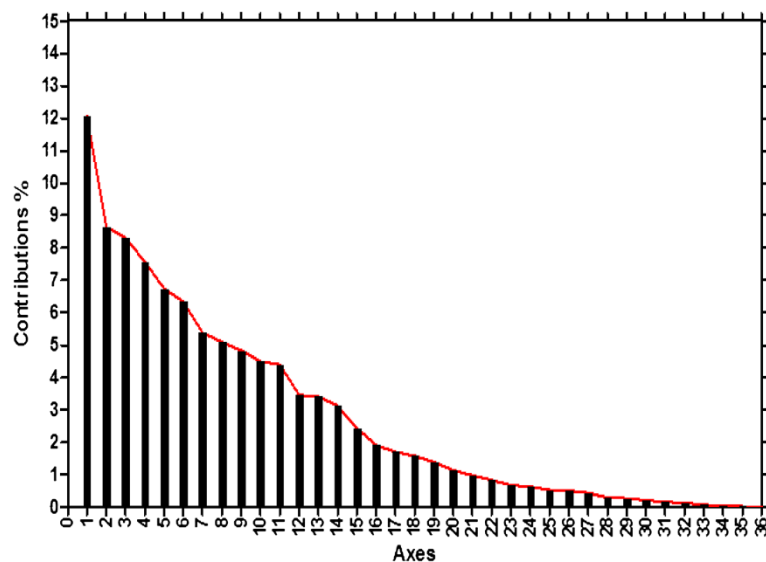
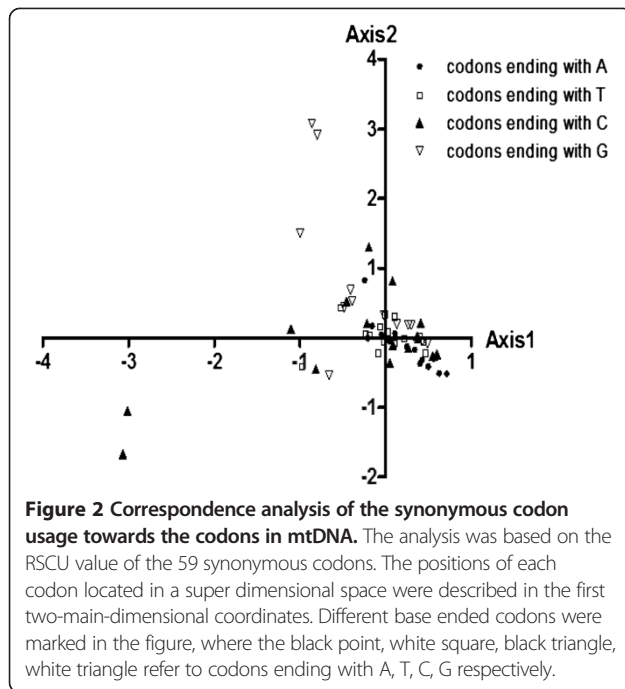


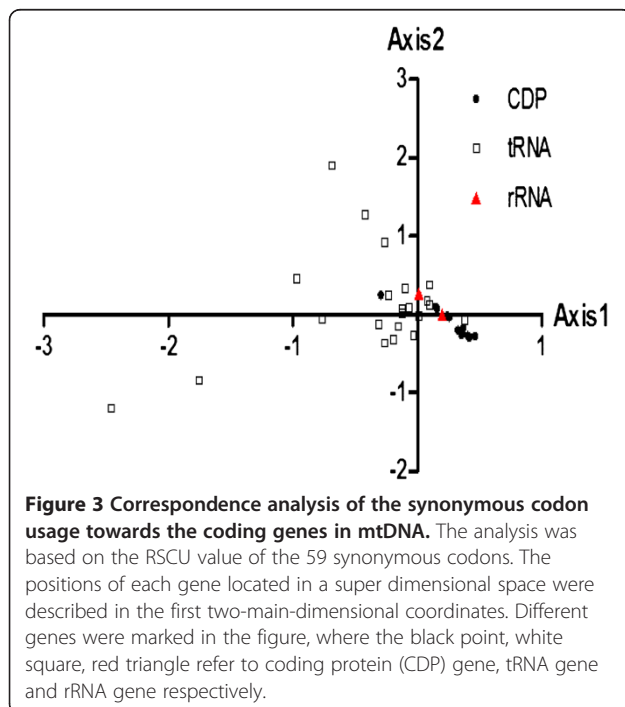
Figure 1 Contributions of the axes. Contributions of 36 axes were shown based on the COA.



shaping the codon bias, whereas natural selection seems to probably dominate the codon bias.

Mutation bias plays a minor role in the codon usage bias of mtDNA

To further confirm whether mutation bias plays a relatively minor role in effecting the codon bias as deduced



above, the ENc values were plotted against the GC3s values in ENc-plot [51] (Figure 5). The standard curve shows the functional relationship between ENc and GC3s is under mutation pressure rather than selection. In this case, if codon usage bias is completely based on GC3s, all the points will lie exactly on the standard curve (corresponding to the ENc values). For instance, the point of the gene 1-rRNA is in the plot just located on the curve, showing its codon bias is only or mainly affected by mutation pressure. However, most of the points do not lie close to the standard curve, which indicates that the GC3s for mutation bias is not the main factor in shaping the codon bias. For the points corresponding to different genes which show a discrete distribution, it implies other factors such as natural selection for gene length and gene expression level may also influence the codon bias to a certain extent.

It is generally acknowledged that codon preference reflects a balance between mutational bias and natural selection for translational optimization. In other words, natural selection and mutational bias are the two main factors leading to codon usage bias [52,53]. So far, combined with the neutrality plot, it shows that the SCUB of mtDNA is mainly dominated by natural selection, and mutational pressure only lightly affects the usage bias. Similarly, in many other species, such as *Arabidopsis thaliana*, *Drosophila melanogaster* or *Caenorhabditis elegans*, natural selection also plays a more important role in forming the codon bias of the whole genome [54,55].

Gene length correlated to expression level causing the codon bias of mtDNA

Ribosome genes of *B. mori* were used as a reference set to calculate the CAI value of the genes. Correlation analysis indicates that CAI is positive correlation against Axis1 ($r = 0.504$, $p < 0.01$) and strong negative correlation with ENc values ($r = -0.851$, $p < 0.01$). Thus, gene expression level is thought to be a factor for causing the bias. Furthermore, the correlation between CAI and gene length is significant ($r = 0.683$, $p < 0.01$) (Figure 6), indicating that gene length is related to expression level, although a few people believed that CAI and ENc were not correlated to the length of genes [56]. Additionally, the average CAI values of 13 protein-coded genes range from 0.5 to 0.7, a level close to 1.0, which might give an assumption that a relatively high expression level of the genes according to the index definition. The results are similar to previous studies [57]. However, it was necessary to investigate more data in order to identify that the 13 protein-coding genes in mitogenome of *B. mori* are highly expressed.

Both EST-based and new quantitative measures of gene expression (MPSS) suggest that codon preference is strongly associated with gene expression derived from the information on tRNA abundance [58]. Analysis

Table 2 Correlation coefficients of the factors that influence codon bias in mtDNA

	ENc	GC3s	GC	L_aa	Gravy	Axis1	Axis2	Axis3
GC3s	-0.389*							
GC	0.053	0.351*						
L_aa	0.649**	-0.515**	0.063					
Gravy	0.327*	-0.505**	-0.396*	0.423**		0.177		
Axis1	0.374*	-0.581**	0.005	0.455**				
Axis2	-0.032	0.143	0.121	-0.178	-0.015	0.120		
Axis3	-0.365*	0.393*	0.081	-0.285	-0.335*	0.063	-0.045	
Axis4	-0.021	0.132	0.022	-0.086	0.052	-0.053	0.038	-0.070

Note: One asterisk (*) and two asterisks (**) indicate correlations at a level of 0.05 and 0.01, respectively.

about the bias might be of great importance to study gene expression, gene function, and even the evolution of the species since SCUB is considered to be an outcome of natural selection, mutation bias or genetic drift [59]. Thus, our analysis reveals that the high expression levels of the coded proteins making up the respiratory chain in mitochondrion meet well with the high energy demand to fit the life system. The process is recognized by the species containing abundant tRNA [60]. Large amount of tRNAs existing in the mitochondrion can also explain the high codon bias of mtDNA of *B. mori*. In addition, translational selection and translational efficiency also play a role in the natural selection towards codon bias. It is observed that long genes are highly expressed in *Yersinia pestis* [61], while the shorter genes have a better expression in *Drosophila melanogaster* [62,63]. This indicates the existence of translational selection and probably that the translational efficiency acts negatively in *Yersinia pestis*, but positively in *Drosophila*

melanogaster in shaping the SCUB for the reason that high codon bias always appears in the highly expressed genes [6]. Our analysis suggests that gene length does play a role in shaping the bias and shows strong correlation with the gene expression level. Therefore we conclude that the translational efficiency is possibly a positive factor to cause the bias and the translational selection might also influence SCUB in mtDNA of *B. mori*.

Factors such as gene length, translational selection and translational efficiency can be due to the natural selection pressure that causes the SCUB in mtDNA of *B. mori*. The index Gravy correlated to the Axis3 might also be suggested as an aspect of the selection pressure in shaping the bias. However, indices which reflect the mutation pressure show little correlation with the primary axes, except GC3s. Combined with the neutrality plot and ENc- plot, the results suggest that it is natural selection rather than mutation pressure that dominates the SCUB in mtDNA of *B. mori*.

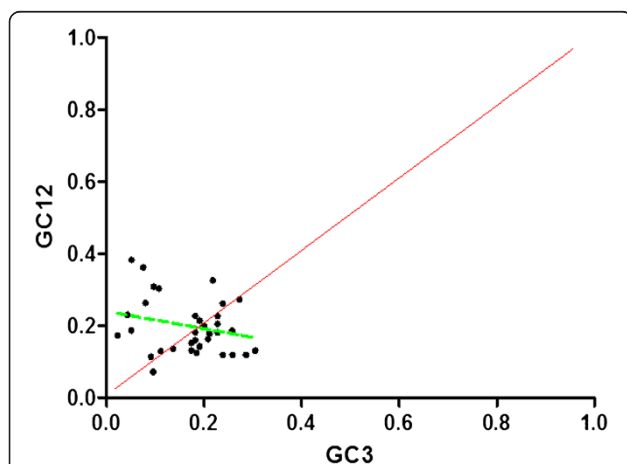
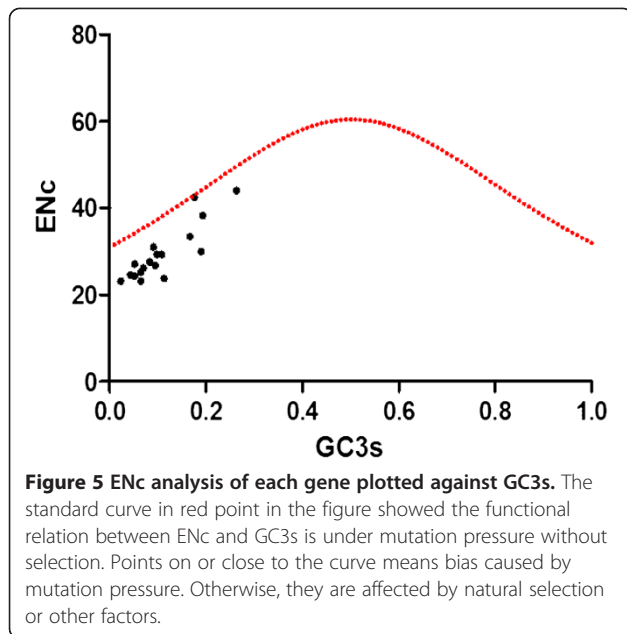


Figure 4 GC12 against GC3 value of each gene. GC12 stands for the average value of GC content in the first and second position of the codons (GC1 and GC2), while GC3 refers to the GC content in the third position. The diagonal line and the regression curve were colored in red and green, respectively. The regression curve can describe as $y = -0.242x + 0.24$, $R^2 = 0.0597$, and GC12 showed no correlation with GC3.

The species classification based on the mtDNA SCUB

30 Neolepidoptera species were used for the cluster analysis, included in five superfamilies (Bombycoidea, Geometroidea, Hepialoidea, Noctuoidea, and Papilionoidea) and 13 families, using the RSCU values from correspondence analysis of the 13 CDSs in each mitogenome. The cluster analysis to *Bombyx* and other species with a relatively near genetic distance in Lepidoptera reveals that all the species are divided into two big clusters at the evolution distance of about 23 (Figure 7). *B. mori* and *Bombyx mandarina* stay the closest and are isolated almost from the root. While the other species in the Bombycoidea family known as have close relationship with the *Bombyx* are divided into another branch. Thus another branch is divided into several clades in turn. However, species in the same color are in a discrete distribution, although several small numbers of species gather together. The results appear largely different from the previous studies [64].

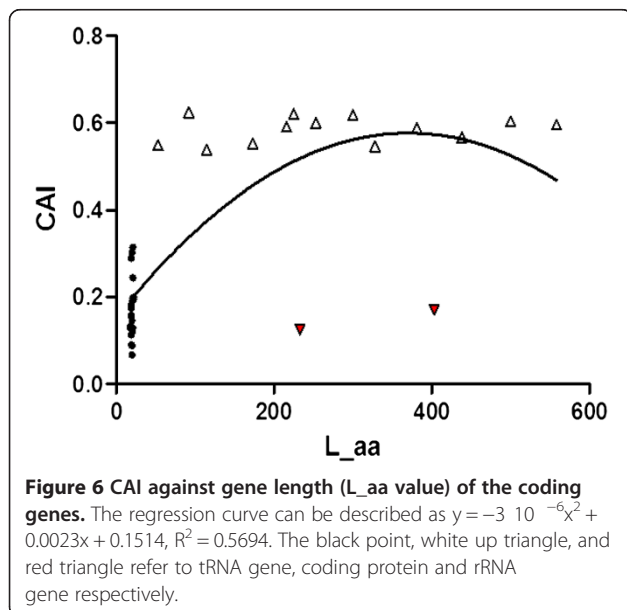
However, almost all of the species in different families are in a similar mean value of ENc, GC content and GC3s (Table 3). Hence, we can hardly classify the species



according to the codon usage indices, except the clade of *B. mori* and *Bombyx mandarina* which has extremely lower ENc value from the other species. Thus, the codon usage of mitogenome might trace a different evolutionary path from the way of species evolution. It could be also assumed that codon bias might determine the mitochondrial genome in a specific way and transcend the boundary of phylogenetic relationships of species [65].

Conclusions

Many factors can result in the synonymous codon usage bias of organisms. For the mtDNA of *B. mori*, natural selection is found to dominate the high SCUB. And we



believe that mutation bias only plays a relatively minor role. Moreover, our study provides new insight into the exploration of setting up new methods for species taxonomy, though a trial still needs to be conducted in the future.

Methods

Codon usage indices

For codon bias, both mutational pressure and natural selection are two core factors. The first step is to assess the codon bias level in analysis, and then determine which one is the driving force. At present, many statistical methods have been proposed and all the common indices are utilized in our study.

Measurement indices of codon bias level

Effective number of codons (ENc) It is often used to measure the codon bias in one specific gene [51], which is an assessment of non-uniformity of usage within synonymous groups of codons. ENc values can vary from 20 (extreme bias where only one codon is used per amino acid) to 61 (without bias where codons are randomly used).

Relative synonymous codon usage (RSCU) This index for a particular codon is given by the observed frequency divided by the expected frequency. If all synonymous codons coding the same amino acid are used equally, RSCU values are close to 1.0, indicating a lack of bias.

Codon bias index (CBI) and frequency of optimal codons (Fop) They are another two common indicators for assessing codon bias [66].

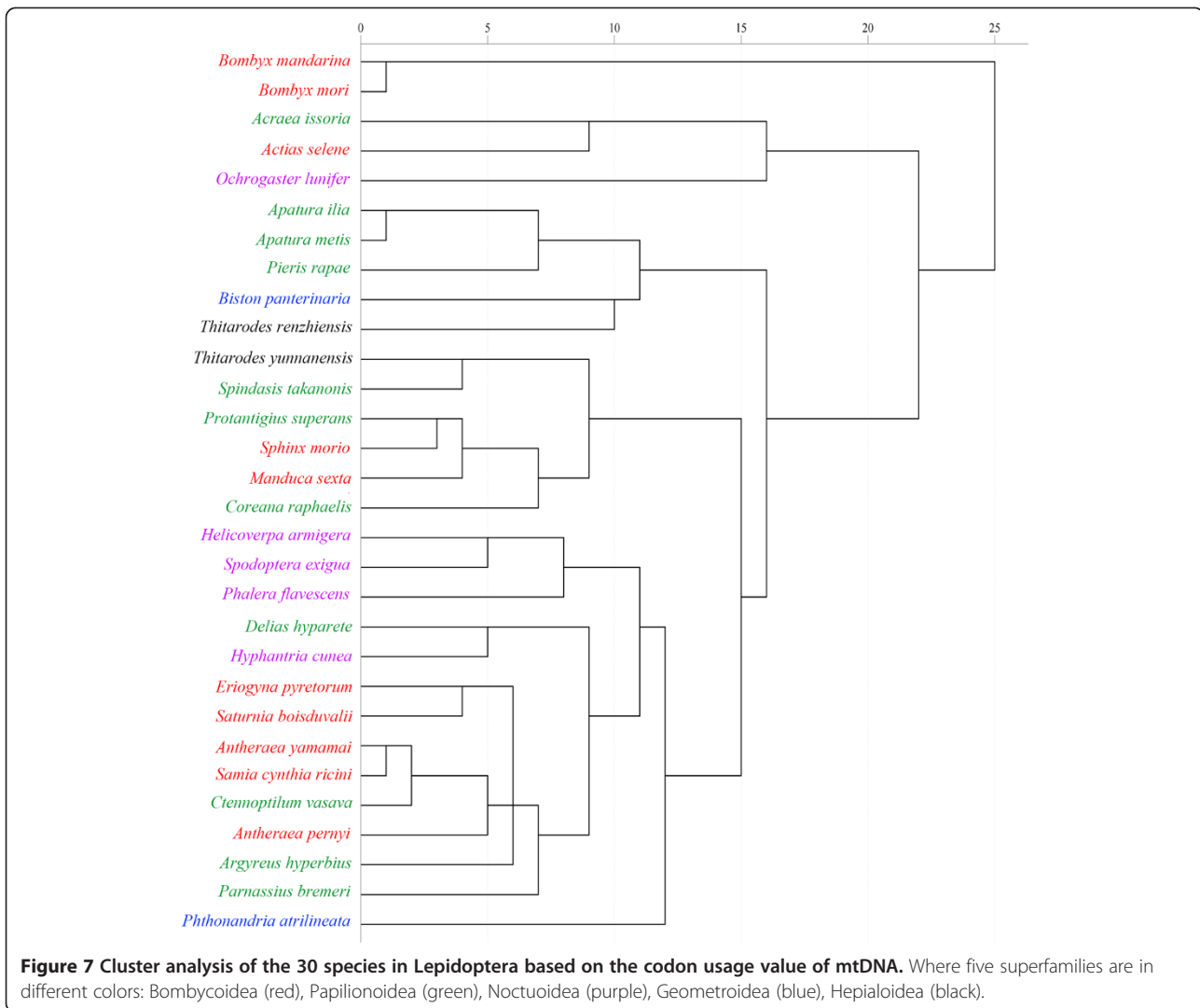
Indices related with mutation bias

GC3s It presents the frequency of use of G + C in the synonymously variable third positions of the sense codon (i.e., excluding Met, Trp and termination codons).

Indices related with natural selection

Codon adaptation index (CAI) The CAI index uses a reference set of highly expressed genes from a species to assess the relative merits of each codon, and a score for a gene is calculated from using the frequency of all codons in that gene. The index assesses the level to which selection has been effectively molding the pattern of codon usage. The CAI values can range from 0 to 1, and the larger ones reveal a higher gene expression level as well as codon bias. In that respect, CAI is useful for predicting the expression level of a particular gene [67].

Length of synonymous codons (L_sym) and length of amino acids (L_aa) The two indices represent the number of synonymous codons and the number of translatable codons, respectively.



General average hydropathicity (GRAVY) and Amoro

The GRAVY index is used for the hypothetical translated gene product, calculated as the arithmetic mean of the sum of the hydropathic indices of each amino acid [68]. Its score range from -2 to 2: positive value shows hydrophobic protein, while negative one indicates hydrophilic protein. Amoro refers to aromaticity, which represents the frequency of aromatic amino acids (Phe, Tyr, Trp) in the hypothetical translated gene product. Both the hydropathicity and aromaticity protein scores are indices of amino acid usage. In *E. coli* genes, the strongest trend in the variation in the amino acid composition is correlated with protein hydropathicity, the second trend is correlated with gene expression, and the third is correlated with aromaticity [69]. The variation in amino acid composition can influence the analysis results of codon usage.

Multivariate statistical analysis

Several methods for the multivariate statistical analysis were used here:

Principal component analysis (PCA)

This method is usually used to analyze genes by site in a 59-dimensional space based on the RSCU values [70]. PCA can transform some of large correlated variables (eg. RSCU values) into major smaller ones. PCA was used here to find the factors determining the major trends in codon usage bias, such as mutation bias or natural selection.

Correspondence analysis (COA)

It is utilized to compare and analyze two or more categories of variable data in one contingency table. Variable categories (rows and the columns) are presented in a low dimension space diagram, in which relatively simple data

Table 3 List of taxa in the phylogenetic and correspondence analysis with their GenBank accession numbers

Species	GenBank No.	ENC SD	GC3s% SD	GC% SD
<i>Bombyx mandarina</i>	NC_003395	27.70 2.04	7.08 0.02	19.13 0.05
<i>Bombyx mori</i>	NC_002355	27.73 2.72	7.23 0.03	19.35 0.05
<i>Actias selene</i>	NC_002355	32.87 2.51	10.32 0.02	21.57 0.05
<i>Antheraea pernyi</i>	NC_018133	31.26 4.05	7.20 0.03	20.22 0.06
<i>Antheraea yamamai</i>	NC_012739	30.46 2.83	6.17 0.02	19.51 0.06
<i>Eriogyna pyretorum</i>	NC_012727	30.08 2.50	5.62 0.02	18.98 0.06
<i>Samia cynthia ricini</i>	NC_017869	31.30 1.87	7.92 0.02	20.23 0.06
<i>Saturnia boisduvalii</i>	NC_010613	31.20 2.58	6.64 0.02	19.25 0.06
<i>Manduca sexta</i>	NC_010266	28.92 2.68	5.12 0.02	18.43 0.05
<i>Sphinx morio</i>	NC_020780	30.18 2.48	6.75 0.03	19.25 0.04
<i>Biston panterinaria</i>	NC_020004	30.79 3.33	7.78 0.02	19.60 0.05
<i>Phthonandria atrilineata</i>	NC_010522	30.19 2.87	7.72 0.03	19.43 0.06
<i>Thitarodes yunnanensis</i>	NC_018095	28.89 1.69	4.46 0.01	17.82 0.05
<i>Thitarodes renzhiensis</i>	NC_018094	31.81 2.11	8.57 0.03	19.58 0.05
<i>Hyphantria cunea</i>	NC_014058	30.24 2.32	8.18 0.01	20.10 0.05
<i>Helicoverpa armigera</i>	NC_014668	30.10 1.98	6.65 0.02	19.36 0.05
<i>Spodoptera exigua</i>	NC_019622	29.66 2.49	5.83 0.02	19.30 0.05
<i>Ochrogaster lunifer</i>	NC_011128	33.80 3.22	14.63 0.03	23.02 0.06
<i>Phalera flavescens</i>	NC_016067	31.47 4.08	7.24 0.04	19.93 0.06
<i>Ctenoptilum vasava</i>	NC_016704	31.43 3.51	7.33 0.02	19.49 0.06
<i>Coreana raphaelis</i>	NC_007976	29.06 2.17	3.44 0.02	17.27 0.04
<i>Protantigius superans</i>	NC_016016	29.41 2.77	5.11 0.02	18.28 0.05
<i>Spindasis takanonis</i>	NC_016018	29.19 2.03	5.44 0.02	17.61 0.05
<i>Apatura ilia</i>	NC_016062	30.99 2.90	8.60 0.03	19.67 0.05
<i>Apatura metis</i>	NC_015537	31.45 2.69	8.48 0.02	19.64 0.05
<i>Argyreus hyperbius</i>	NC_015988	30.79 3.33	7.78 0.02	19.60 0.05
<i>Parnassius bremeri</i>	NC_014053	30.65 3.97	5.66 0.02	18.61 0.05
<i>Delias hyparete</i>	NC_020428	31.68 3.03	7.72 0.03	20.02 0.06
<i>Pieris rapae</i>	NC_015895	31.90 2.14	8.94 0.02	20.20 0.06
<i>Acraea issoria</i>	NC_013604	32.65 1.20	9.78 0.03	20.52 0.05

Note: All of the ENC, GC3s%, GC% are mean values.

is displayed to replace the original data, eliminating excess noise and complex data structure, and providing visual results [71-73].

Linear regression analysis (LRA) and factor analysis (FA)

The two methods were also used to show the correlations of the indices and the codon usage pattern of the mtDNA in *B. mori*s.

Cluster analysis (CA)

In the analysis, data is classified into different categories according to their similarities. The clustering analysis to *B. mori* was made among 8 representative organisms using the method of squared Euclidean distance. And the final results came from the clustering analysis of the most

debated classification between Bombycidae and Saturniidae based on the whole mtDNA sequence and the codon frequency in mitochondrial, respectively.

ENC-plot mapping analysis

The analysis was employed in order to find out the determining factor in influencing the codon usage bias, in which the ENC values were plotted against the GC3s values in Nc-plot [51]. The standard curve shows the functional relation between ENC and GC3s was under mutation pressure rather than selection. In this case, if the codon usage is optimal, the analyzed points will lie on or just below the expected curve (composed of ENC values). Otherwise, genes with serious codon bias and significant correlation

with gene expression will be low in ENc values and show points far lower than the standard curve.

Neutrality plot mapping analysis

The frequent mutations of the synonymous codons usually happen in the third position, whereas some mutations occurring in the first or the second positions are non-synonymous. The frequency of these non-synonymous mutations are relatively small, caused by the gene function or gene activity. This means that if there is no outside pressure, mutations should occur in random rather than in a certain direction under the condition of pressure. Thus, there is no difference between three codon positions, and base content composition is also similar. However, bases preferences of the three positions are different in the presence of certain selection pressures [74]. When drawing, GC12 works as the ordinate while GC3 works as the abscissa, and each dot represents an independent gene. If all the points lie along the diagonal distribution, it indicates that no significant difference exists on the three codon positions, and there is no or only a weak external selection pressure. On the other hand, if the regression curve tends to be sloped or parallel to the horizontal axis, it means that the variation correlation between GC12 and GC3 is very low. Therefore, the regression curve works well to measure the neutral degree when selecting effect dominates the evolution [74,75].

Software used

All indices of codon usage bias above were calculated in the data set using the program CodonW 1.4.4 (<http://codonw.sourceforge.net/>) [4]. Clustering analysis and correlation analysis were carried out using the statistical software SPSS 19.0. Graphs were generated in GraphPad Prism 5.0 (<http://www.graphpad.com/scientific-software/prism/>).

Availability of supporting data

The codon usage of the 48 CDSs (12937 codons) in mtDNA and the 1180 CDSs (450043 codons) in the gDNA in Table 1 were downloaded from the Codon Usage Database (<http://www.kazusa.or.jp/codon/>). The codon usage of *B. mori* ribosome used in calculating the CAI value was also downloaded from the Codon Usage Database. The sequence set of the 37 genes in the mtDNA of *B. mori* (NC_002355) used in correspondence analysis were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank>). And all of the mtDNA CDSs of the 30 species used in clustering analysis were achieved from GenBank.

Abbreviations

SCUB: Synonymous codon usage bias; gDNA: Genomic DNA; mtDNA: Mitochondrial DNA; CDS: Coding sequence; CDP: Coding protein; ENc: Effective number of codons; RSCU: Relative synonymous codon usage; CAI: Codon adaptation index; CBI: Codon bias index; Fop: Frequency of optimal codons; GRAVY: General average hydropathicity; Amoro: Amoromaticity; PCA: Principal component analysis;

COA: Correspondence analysis; LRA: Linear regression analysis; FA: Factor analysis; Lei Wei: Jian He and Xian Jia contributed equally to this work.

Competing interests

The authors declare that they have no competing interests.

Authors contributions

LW, JH, XJ and JS conceived the study. LW, JH, XJ and JS carried out the manuscript and writing work. LW, JH, ZL and QQ performed the data analysis. QQ, HZ and YP coordinated the sampling of material. QQ, HZ and SL interpreted the context of results. All authors wrote, read and approved the final manuscript.

Acknowledgements

We would like to thank Dr. Maria Anisimova, Mr. Joselito Acosta and three anonymous reviewers for providing critical and helpful comments on the manuscript. We would also like to give special thanks to Anthony Dinardo, Han Jiang and Dr. Aguang Dai (Florida State University) for revising the grammar of the manuscript. This research was supported by the National Natural Science Foundation of China (31372373, 31172263 to JS), the Natural Science Foundation of Guangdong Province, China (S2013010016750 to JS) and Specialized Research Fund for the Doctoral Program of Higher Education, China (20134404110023 to JS).

Author details

¹Subtropical Sericulture and Mulberry Resources Protection and Safety Engineering Research Center, Guangdong Provincial Key Laboratory of Agro-animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China. ²Guangzhou East Campus Lab Center, Sun Yat-sen University, Guangzhou 510006, China.

Received: 20 August 2014 Accepted: 9 December 2014

Published online: 17 December 2014

References

1. Sun Z, Wan D, Murphy RW, Ma L, Zhang X, Huang D: Comparison of base composition and codon usage in insect mitochondrial genomes. *Genes Genom* 2009, **31**(1):65-71.
2. Sharp MP, Li W: Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rare codons. *Nucleic Acids Research* 1986, **14**(19):7737-7749.
3. Bulmer M: The selection-mutation-drift theory of synonymous codon usage. *Genetics* 1991, **129**(3):897-907.
4. Peden JF: *Analysis of Codon Usage*. PhD thesis. Nottingham University, Department of Genetics; 1999.
5. Francino HP, Ochman H: Isochores result from mutation not selection. *Nature* 1999, **400**:30-31.
6. Powell JR, Moriyama EN: Evolution of codon usage bias in *Drosophila*. *Proc Natl Acad Sci U S A* 1997, **94**:7784-7790.
7. Wang HC, Hickey DA: Rapid divergence of codon usage patterns within the rice genome. *BMC Evol Biol* 2007, **7**(1):S6.
8. Ingvarsson PK: Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol Biol Evol* 2007, **24**:836-844.
9. Zhu B, Liu Q, Dai L, Wang L, Sun Y, Lin K, Wei G, Liu C: Characterization of the complete mitochondrial genome of *Diaphania pyloalis* (Lepidoptera: Pyralidae). *Gene* 2013, **527**(1):283-291.
10. Sima Y, Chen M, Yao R, Li Y, Liu T, Jin X, Wang L, Su J, Li X, Liu Y: The complete mitochondrial genome of the Ailanthus silkworm, *Samia cynthia cynthia* (Lepidoptera: Saturniidae). *Gene* 2013, **526**(2):309-317.
11. Dai L, Zhu B, Liu Q, Wei G, Liu C: Characterization of the complete mitochondrial genome of *Bombyx mori* strain H9 (Lepidoptera: Bombycidae). *Gene* 2013, **519**(2):326-334.
12. Kobayashi N, Takahashi M, Kihara S, Niimi T, Yamashita O, Yaginuma T: Cloning of cDNA encoding a *Bombyx mori* homolog of human oxidation resistance 1 (OXR1) protein from diapause eggs, and analyses of its expression and function. *J Insect Physiol* 2014, **68**:58-68.
13. Ihara H, Okada T, Ikeda Y: Cloning, expression and characterization of *Bombyx mori* α 1,6-fucosyltransferase. *Biochem Biophys Res Commun* 2014, **450**(2):953-960.

14. Cai X, Yu J, Yu H, Liu Y, Fang Y, Ren Z, Jia J, Zhang G, Guo X, Jin B, Gui Z: **Core promoter regulates the expression of cathepsin B gene in the fat body of *Bombyx mori*.** *Gene* 2014, **542**(2):232-239.
15. Zhou Y, Chen H, Li X, Wang Y, Chen K, Zhang S, Meng X, Lee EYC, Lee MYWT: **Production of recombinant human DNA polymerase delta in a *Bombyx Mori* bioreactor.** *PLoS One* 2011, **6**(7):e22224.
16. Jiang L, Zhao P, Cheng T, Sun Q, Peng Z, Dang Y, Wu X, Wang G, Jin S, Lin P, Xia Q: **A transgenic animal with antiviral properties that might inhibit multiple stages of infection.** *Antivir Res* 2013, **98**:171-173.
17. Jiang L, Zhao P, Wang G, Cheng T, Yang Q, Jin S, Lin P, Xiao Y, Sun Q, Xia Q: **Comparison of factors that may affect the inhibitory efficacy of transgenic RNAi targeting of baculoviral genes in silkworm.** *Bombyx mori Antivir Res* 2013, **97**(3):255-263.
18. Jiang L, Xia Q: **The progress and future of enhancing antiviral capacity by transgenic technology in the silkworm *Bombyx mori*.** *Insect Biochem Molec* 2014, **48**:1-7.
19. Arunkumar KP, Metta M, Nagaraju J: **Molecular phylogeny of silkmoths reveals the origin of domesticated silkmoth, *Bombyx mori* from Chinese *Bombyx mandarina* and paternal inheritance of *Antheraea proylei* mitochondrial DNA.** *Mol Phylogenet Evol* 2006, **40**(2):419-427.
20. Maekawa H, Takada N, Mikitani K, Ogura T, Miyajima N, Fujiwara H, Kobayashi M, Ninaki O: **Nucleolus organizers in the wild silkworm *Bombyx mandarina* and the domesticated silkworm *B. mori*.** *Chromosoma* 1988, **96**(4):263-269.
21. Li D, Guo Y, Shao H, Tellier LC, Wang J, Xiang Z, Xia Q: **Genetic diversity, molecular phylogeny and selection evidence of the silkworm mitochondria implicated by complete resequencing of 41 genomes.** *BMC Evol Biol* 2010, **10**:81.
22. Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C, Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W, et al: **A draft sequence for the genome of the domesticated silkworm (*Bombyx Mori*).** *Science* 2004, **306**(5703):1937-1940.
23. Mita K, Yasukochi Y, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-I T, Hiroaki A, Shimada T, Morishita S, Sasaki T: **The genome sequence of silkworm, *Bombyx mori*.** *DNA research: an international journal for rapid publication of reports on genes and genomes* 2004, **11**(1):27-35.
24. Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, Cheng T, Jiang T, Becquet C, Xu X, Liu C, Zha X, Fan W, Lin Y, Shen Y, Jiang L, Jensen J, Hellmann I, Tang S, Zhao P, Xu H, Yu C, Zhang G, Li J, Cao J, Liu SHe N, et al: **Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*).** *Science* 2009, **326**(5951):433-436.
25. Hebert PD, Cywinska A, Ball SL, de Waard JR: **Biological identifications through DNA barcodes.** *Proc Biol Sci* 2003, **270**:313-321.
26. Zhu LX, Wu XB: **Phylogenetic evaluation of the taxonomic status of *Papilio maackii* and *P. sfyanius* (Lepidoptera: Papilionidae).** *Zool Res* 2011, **32**:248-254.
27. Gissi C, Iannelli F, Pesole G: **Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species.** *Heredity* 2008, **101**:301-320.
28. Curole JP, Kocher TD: **Mitogenomics: digging deeper with complete mitochondrial genomes.** *Trends Ecol Evol* 1999, **14**(10):394-398.
29. Barton N, Jones JS: **Evolutionary biology: Mitochondrial DNA: new clues about evolution.** *Nature* 1983, **306**:317-318.
30. Galtier N, Nabholz B, Glemin S, Hurst GDD: **Mitochondrial DNA as a marker of molecular diversity: a reappraisal.** *Mol Ecol* 2009, **18**(22):4541-4550.
31. Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y: **Significant levels of sequence divergence and gene Rearrangements have occurred between the mitochondrial Genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*.** *Mol Biol Evol* 2002, **19**(8):1385-1389.
32. Wei S, Li Q, van Achtenberg K, Chen X: **Two mitochondrial genomes from the families *Bethylidae* and *Mutillidae*: Independent rearrangement of protein-coding genes and higher-level phylogeny of the Hymenoptera.** *Mol Phylogenet Evol* 2014, **77**:1-10.
33. Harrison RG: **Animal mitochondrial DNA as a genetic marker in population and evolutionary biology.** *Trends Ecol Evol* 1989, **4**(1):6-11.
34. Boore JL: **Animal mitochondrial genomes.** *Nucleic Acids Res* 1999, **27**(8):1767-1780.
35. Kawabe A, Miyashita NT: **Patterns of codon usage bias in three dicot and four monocot plant species.** *Genes Genet Syst* 2003, **78**(5):343-352.
36. Embley TM, Martin W: **Eukaryotic evolution, changes and challenges.** *Nature* 2006, **440**(7084):623-630.
37. Wise CA, Sraml M, Easteal S: **Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees.** *Genetics* 1998, **148**(1):409-421.
38. Rand DM, Kann LM: **Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA.** *Genetica (Dordrecht)* 1998, **102-103**:393-407.
39. David MR, Michele D, Lisa MK: **Neutral and nonneutral evolution of *Drosophila* mitochondrial DNA.** *Genetics* 1994, **138**:741-756.
40. Nachman MW, Brown WM, Stoneking M, Aquadro CF: **Nonneutral mitochondrial DNA variation in humans and chimpanzees.** *Genetics* 1996, **142**(3):953-963.
41. Ballard J, Kreitman JW: **Unraveling selection in the mitochondrial genome of *Drosophila*.** *Genetics* 1994, **138**(3):757-772.
42. Rand DM, Haney RA, Fry AJ: **Cytosuclear coevolution: the genomics of cooperation.** *Trends Ecol Evol* 2004, **19**(12):645-653.
43. Dowling D, Friberg U, Lindell J: **Evolutionary implications of non-neutral mitochondrial genetic variation.** *Trends Ecol Evol* 2008, **23**(10):546-554.
44. Gemmell NJ, Metcalf VJ, Allendorf FW: **Mothers curse the effect of mtDNA on individual fitness and population viability.** *Trends in Ecology and Evolution* 2004, **19**(5):238-244.
45. Blier PU, Dufresne F, Burton RS: **Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation.** *Trends Genet* 2001, **17**(7):400-406.
46. Meiklejohn CD, Montooth KL, Rand DM: **Positive and negative selection on the mitochondrial genome.** *Trends Genet* 2007, **23**(6):259-263.
47. Ballard J, Rand DM: **The population biology of mitochondrial DNA and its phylogenetic implications.** *Annu Rev Ecol Syst* 2005, **36**:621-642.
48. Burton RS, Ellison CK, Harrison JS: **The sorry state of F2 hybrids: consequences of rapid mitochondrial DNA evolution in allopatric populations.** *Am Nat* 2006, **168**(Suppl 6):S14-S24.
49. Rand DM: **The Units of Selection of Mitochondrial DNA.** *Annu Rev Ecol Syst* 2001, **32**:415-448.
50. Pesole G, Gissi C, De Chirico A, Saccone C: **Nucleotide substitution rate of mammalian mitochondrial genomes.** *J Mol Evol* 1999, **48**(4):427-434.
51. Wright F: **The effective number of codons used in a gene.** *Gene* 1990, **87**(1):23-29.
52. Tao P, Dai L, Luo M, Tang F, Tien P, Pan Z: **Analysis of synonymous codon usage in classical swine fever virus.** *Virus Genes* 2009, **38**(1):104-112.
53. Liu YS, Zhou JH, Chen HT, Ma LN, Pejsak Z, Ding YZ, Zhang J: **The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern.** *Infect Genet Evol* 2011, **11**(5):1168-1173.
54. Duret L, Mouchiroud D: **Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*.** *Proc Natl Acad Sci U S A* 1999, **96**(8):4482-4487.
55. Qiu S, Bergero R, Zeng K, Charlesworth D: **Patterns of codon usage bias in *Silene latifolia*.** *Mol Biol Evol* 2010, **28**(11):771-780.
56. Comeran JM, Aguade M: **An evaluation of measures of synonymous codon usage bias.** *J Mol Evol* 1998, **47**(3):268-274.
57. Jia W, Higgs PG: **Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection.** *Mol Biol Evol* 2008, **25**(2):339-351.
58. Wright SI: **Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*.** *Mol Biol Evol* 2004, **21**(9):1719-1726.
59. Zhang Z, Li J, Cui P, Ding F, Li A, Townsend JP, Yu J: **Codon deviation coefficient: a novel measure for estimating codon usage bias and its statistical significance.** *BMC Bioinformatics* 2012, **13**:43.
60. Ikemura T: **Codon usage and tRNA content in unicellular and multicellular organisms.** *Mol Biol Evol* 1985, **2**(1):13-34.
61. Hou ZC, Yang N: **Factors affecting codon usage in *Yersinia pestis*.** *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* 2003, **35**(6):580-586.
62. Moriyama EN, Powell JR: **Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*.** *Nucleic Acids Res* 1998, **26**(13):3188-3193.
63. Miyasaka H: **Translation initiation AUG context varies with codon usage bias and gene length in *Drosophila melanogaster*.** *J Mol Evol* 2002, **55**(1):52-64.

64. Liu S, Xue D, Cheng R, Han H: The complete mitogenome of *Apocheima cinerarius* (Lepidoptera: Geometridae: Ennominae) and comparison with that of other Lepidopteran insects. *Gene* 2014, **547**:136-144.
65. Roychoudhury S, Mukherjee D: A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res* 2010, **148**(1-2):31-43.
66. Bennetzen JL, Hall BD: Codon selection in yeast. *The Journal of biological chemistry* 1982, **257**(6):3026-3031.
67. Sharp PM, Tuohy TM, Mosurski KR: Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 1986, **14**(13):5125-5143.
68. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982, **157**(1):105-132.
69. Lobry JR, Gautier C: Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res* 1994, **22**(15):3174-3180.
70. Gupta SK, Ghosh TC: Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*. *Gene* 2001, **273**(1):63-70.
71. Grantham R, Gautier C, Gouy M, Mercier R, Pavé A: Codon catalog usage and the genome hypothesis. *Nucleic Acids Res* 1980, **8**(1):r49-r62.
72. Sharp PM, Devine KM: Codon usage and gene expression level in *Dictyostelium discoideum*: highly expressed genes do prefer optimal codons. *Nucleic Acids Res* 1989, **17**(13):5029-5039.
73. Shields DC, Sharp PM: Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Res* 1987, **15**(19):8023-8040.
74. Sueoka N: Directional mutation pressure and neutral molecular evolution. *P Natl Acad Sci USA* 1988, **85**(8):2653-2657.
75. Sueoka N: Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *J Mol Evol* 1999, **49**(1):49-62.

doi:10.1186/s12862-014-0262-4

Cite this article as: Wei et al.: Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evolutionary Biology* 2014 **14**:262.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

