The Journal of
Eukaryotic Microbiology

Published by the International Society of Protistologists

J. Eukaryot. Microbiol., 59(3), 2012 pp. 287–289

DOI: 10.1111/j.1550-7408.2011.00613.x

Factors Affecting Codon Bias in the Mitochondrial Genomes of the Streptophyte *Mesostigma viride* and the Chlorophyte *Chlamydomonas reinhardtii*

JIMENG HUA and ROBERT W. LEE

Department of Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4R2, Canada

ABSTRACT. Mitochondrial genomes typically show genome-wide patterns of synonymous codon usage bias. In animals and land plants, mutation appears more dominant than selection in shaping this bias, while in green algae the relative importance of these factors is not well studied. Based on our analysis of mitochondrial DNA sequence from the green algae *Mesostigma viride* (NIES-296) and *Chlamydomonas reinhardtii* (CC-277) and a closely related relative of each, we conclude that both mutation and selection are important in shaping synonymous codon usage bias in their mitochondrial genomes, with selection being more dominant. The possible confounding influence of mutational context dependence on our analyses is discussed.

Key Words. Green algae, mutation, selection, synonymous codon usage.

LIKE other genomes, those of mitochondria commonly show genome-wide synonymous codon usage bias, which can result from natural selection and mutation (Bulmer 1991; Lynch 2007). Complex likelihood-based tests on different models and estimates of the average intensity of selection on synonymous sites (ΔF) suggest that in animal and land plant mitochondria, mutation dominates over selection in shaping this bias (Jia and Higgs 2008; Sloan and Taylor 2010). Although mitochondria from the two main lineages of green algae (Streptophyta and Chlorophyta) show synonymous codon usage bias (Wang et al. 2011), the relative importance of adaptive vs. neutral forces in shaping this bias is not well explained. The first and best studied green algal mitochondrial genome, that of the chlorophyte *Chlamydomonas reinhardtii*, shows an excess of T and A nucleotides at 4-fold degenerate (FFD) sites (Boer and Gray 1988), which has been explained in part by background mutation pressure evident in both intergenic and intronic noncoding regions (Popescu and Lee 2007) and by selection for preferred codons that match the anticodons of the mitochondrial tRNAs, most of which are imported from the cytosol (Vinogradova et al. 2009). The GC bias evident in mitochondrial DNA (mtDNA) at FFD sites and intergenic regions in the chlorophyte green alga *Polytomella capuana* has been explained by biased gene conversion (Smith and Lee 2008). Finally, based on a broad survey of chlorophyte and streptophyte green algal mitochondrial genomes, Wang et al. (2011) discuss evidence for selection contributing to the unequal frequencies of synonymous codons in the former group and mutation being the dominant force in the latter group. These conclusions are suspect, however, because they are based mainly on plots of effective number of codons vs. the GC content at the third synonymous codon sites (i.e. *Nc-GC3s* plots) (Wright 1990) of protein-coding genes (PCGs), which cannot always distinguish the relative effects of mutation and selection on the base frequencies at synonymous sites (Wright 1990).

In this study, we chose to estimate the overall relative impact of selection and mutation on synonymous codon usage bias in the mtDNA of the chlorophyte *Chlamydomonas* and the streptophyte *Mesostigma*, which represent the only green algal genera where appropriate mtDNA divergence data are available (Hua et al. 2012; Popescu and Lee 2007).

Corresponding Author: Robert W. Lee, Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2—Telephone number: +902 494 2554; FAX number: +902 494 3736; e-mail: robert.lee@dal.ca

MATERIALS AND METHODS

DNA sequences. The DNA sequences were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank/)—NC008240 and EU306622 for the complete mitochondrial genomes of *Mesostigma viride* NIES-296 and *C. reinhardtii* CC-277; HQ667981 and DQ373068 for the partial mitochondrial genomes of *M. viride* SAG 50-1 and *Chlamydomonas incerta* SAG 7.73.

Data analysis. MEGA version 4.0 (Tamura et al. 2007) was used to analyze the nucleotide composition. CodonW (Peden 1999) was used to calculate the relative synonymous codon usage. Likelihood ratio tests (LRT) were conducted to test the fitness of the FMutSel/FMutSel0 models in PAML version 4.3 (Yang and Nielsen 2008) on the divergence data within *Mesostigma* and *Chlamydomonas*. The ΔF values were estimated by averaging the absolute values of the differences in codon fitness, provided by the FMutSel model, between all possible pairs of synonymous codons differing by a single nucleotide (Sloan and Taylor 2010). The relative abundance of dinucleotides was estimated by calculating the ratio of observed to expected frequencies (Fedorov et al. 2002; see Supporting Information Table S1).

RESULTS AND DISCUSSION

Nucleotide composition at FFD codon and noncoding sites. To test if mutation and/or selection are involved in shaping the biased synonymous codon usage in the mtDNA of *M. viride* and *C. reinhardtii* (see Supporting Information Table S2), we compared the nucleotide composition at FFD-codon and noncoding sites in these DNAs. In both instances regardless of the DNA strand, the FFD sites of standard PCGs and the noncoding sites are biased toward A and T (Table 1). Theoretically, if the nucleotide composition was determined only by mutation and the rates of all 12 possible base-substitution mutations were identical, the percentage of each of the four nucleotides would be 25% (Lynch 2007), and consequently, the AT content would be 50%. If we assume that the nucleotide composition of noncoding regions has evolved neutrally, the AT-richness of these regions should reflect a genomic background mutation bias toward AT. The FFD sites are even AT-richer than the noncoding regions, which may indicate that selection has driven the synonymous sites to be more biased. Although, the noncoding mtDNA of *M. viride* is AT-richer than that of *C. reinhardtii*, the increase of AT content from noncoding sites to FFD sites in *M. viride* (from ~ 73% to ~ 87%) is less than that in *C. reinhardtii* (from ~ 60% to ~ 79%), which may reveal that selection is stronger in the latter taxon.

Table **1.** Nucleotide composition at fourfold degenerate (FFD) sites in protein-coding sequences and in noncoding regions of the *Mesostigma viride* (NIES-296) and *Chlamydomonas reinhardtii* (CC-277) mitochondrial genomes.

DNA sites[a]	DNA strand[b]	No. of sites	A (%)	T (%)	G (%)	C (%)	AT (%)
Mesostigma viride							
FFD							
Standard genes	+	1,239	32.3	53.8	6.7	7.2	86.1
	−	2,173	36.1	52.0	5.9	6.0	88.1
Noncoding							
Intronic	+	941	40.4	31.8	15.1	12.8	72.2
	−	1,488	39.9	32.9	14.7	12.5	72.8
Intergenic	+	1,502	37.4	36.4	13.0	13.2	73.8
	−	2,347	38.2	34.5	12.8	14.5	72.7
Chlamydomonas reinhardtii							
FFD							
Standard genes	+	633	26.2	52.1	2.8	18.8	78.3
	−	658	26.4	52.6	1.5	19.5	79.0
rtl	+	163	24.5	39.9	14.7	20.9	64.4
Noncoding							
Intronic[c]	−	361	30.5	33.0	20.2	16.3	63.5
Intergenic	+	1,096	27.8	31.3	18.7	22.2	59.1
	−	432	27.3	31.9	19.0	21.8	59.2

[a]All sites were obtained from concatenated sequences in each category except in the case of the reverse transcriptase-like coding sequence (*rtl*).

[b]The complete mitochondrial genome sequences directly downloaded from GenBank were designated as + strands and the complementary strands were designated as − strands.

[c]Data from another geographical isolate of C. *reinhardtii* (CC-1373); GenBank accession number EU306617.

Relative strength of selection and mutation on synonymous codon sites. Independent of the comparison of nucleotide composition, the method of Sloan and Taylor (2010) employing the FMutSel/FMutSel0 models in PAML (Yang and Nielsen 2008), which needs pairwise sequence data, also enables us to investigate the relative roles of mutation and selection on patterns of synonymous codon usage. *Mesostigma viride* NIES-296 and SAG 50-1, which appear to represent deeply isolated lineages, and *C. reinhardtii* and *C. incerta*, are the only two identified green algal pairs with sufficient but nonsaturated mitochondrial nucleotide sequence divergence data (Hua et al. 2012; Popescu and Lee 2007). LRT on the FMutSel/FMutSel0 models for both sets of divergence data show that the more complex FMutSel model fits our datasets ($P < 0.001$, in each case); this provides statistically significant evidence that selection acts on the synonymous codon sites (Yang and Nielsen 2008). Then, we calculated the intensity of selection on the synonymous codon sites (ΔF) using the parameters inferred from the FMutSel model. The ΔF is 1.99 for the *Mesostigma* pair and 4.77 for the *Chlamydomonas* pair. In the case of the *Chlamydomonas* analysis, if the reverse transcriptase-like (*rtl*) coding sequence is included in the concatenation, the FMutSel model still fits ($P < 0.001$) and the ΔF is 3.78. A ΔF of 1 is generally viewed as an approximate threshold above which selection begins to dominate over neutral forces (Sloan and Taylor 2010). Thus, in the mitochondrial genomes of both pairs of *Mesostigma* and *Chlamydomonas*, the higher-than-1 ΔF values suggest that selection is the dominant force affecting the frequency of alternative synonymous codons; the greater ΔF for the *Chlamydomonas* pair supports selection being stronger than for the *Mesostigma* pair.

Context dependence. The nucleotide composition of a site can be influenced by the nucleotides at surrounding sites; this phenomenon is known as context dependence and can result from mutation and/or selection (Berg and Silva 1997). Our χ^2 tests show that for the FFD codons in both *M. viride* and *C. reinhardtii* mtDNA, the nucleotide frequency at the third synonymous codon position (Z) is not independent of the second codon position (Y) and the first codon position of the following codon (N_1) (Supporting Information Table S3). The existence of mutational context dependence may affect our above interpretations that both selective and mutational forces are driving synonymous FFD-codon-usage bias in *Mesostigma* and *Chlamydomonas* mtDNA, because our explanations are based on the assumption that mutations are not context dependent. The difference in nucleotide composition between the synonymous sites in FFD codons and the noncoding regions could be due to different contextual influences rather than selection as discussed by Morton (2003), and the FMutSel/FMutSel0 models assume that single mutations occur independently at nucleotide sites (Yang and Nielsen 2008). Therefore, to explore further the causes of context dependence in our case, we calculated the relative abundance of dinucleotides in intergenic regions (r), dinucleotides YZ in FFD codons ($R(YZ)$), and dinucleotides ZN_1 where Z belongs to FFD codons ($R(ZN_1)$) (see Supporting Information Table S1); R or r values ≤ 0.78 or ≥ 1.23 were considered significantly biased (Karlin and Mrázek 1996). We found that in *M. viride* 39.6% of the $R(YZ)$ values, 46.9% of the $R(ZN_1)$ values, and 40.1% of the r values are significantly biased, and in *C. reinhardtii* 83.2% of the $R(YZ)$ values, 41.8% of the $R(ZN_1)$ values, and 21.9% of the r values are significantly biased. The biased r, $R(YZ)$, and $R(ZN_1)$ values show that the observed frequencies of these dinucleotides deviate from their expected frequencies based on the nucleotide composition of the corresponding region, which means there is context dependence between the neighboring sites investigated. Furthermore, we found that in *M. viride*, only 9 of the 38 significantly biased $R(YZ)$ and 33 of the 120 significantly biased $R(ZN_1)$ values have corresponding same-direction-biased r values, and these proportions in *C. reinhardtii* are 3 of 79 for $R(YZ)$ and 11 of 107 for $R(ZN_1)$. Under the assumption that the context dependence in the

intergenic regions is caused purely by mutational effects, these data suggest that the context dependence of the FFD sites and their neighboring nucleotides in both mtDNAs are caused mainly by selection. However, some of the $R(YZ)$ and $R(ZN_1)$ values are based on small occurrence numbers of corresponding contexts, which seems inevitable because of the small size of these mtDNAs, and there is a possibility that some noncoding sites are constrained by selection.

In the future, the expression levels of PCGs encoded in the mtDNA of *C. reinhardtii* and *M. viride* might be worth studying because the effects of selectional and mutational context dependence might be untangled if these genes vary in expression at the protein level (Berg and Silva 1997; Ran and Higgs 2010) and in codon usage patterns.

ACKNOWLEDGMENTS

We thank Huaichun Wang for helpful discussion. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to R.W.L.

LITERATURE CITED

Berg, O. G. & Silva, P. J. 1997. Codon bias in *Escherichia coli*: the influence of codon context on mutation and selection. *Nucleic Acids Res.*, **25**:1397–1404.

Boer, P. H. & Gray, M. W. 1988. Transfer RNA genes and the genetic code in *Chlamydomonas reinhardtii* mitochondria. *Curr. Genet.*, **14**:583–590.

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

Fedorov, A., Saxonov, S. & Gilbert, W. 2002. Regularities of context-dependent codon bias in eukaryotic genes. *Nucleic Acids Res.*, **30**:1192–1197.

Hua, J., Smith, D. R., Borza, T. & Lee, R. W. 2012. Similar relative mutation rates in the three genetic compartments of *Mesostigma* and *Chlamydomonas*. *Protist*, **163**:105–115.

Jia, W. & Higgs, P. G. 2008. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Mol. Biol. Evol.*, **25**:339–351.

Karlin, S. & Mrázek, J. 1996. What drives codon choices in human genes? *J. Mol. Evol.*, **262**:459–472.

Lynch, M. 2007. The Origins of Genome Architecture. Sinauer Assocs. Inc., Sunderland, MA.

Morton, B. R. 2003. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. *J. Mol. Evol.*, **56**:616–629.

Peden, J. F. 1999. Analysis of Codon Usage. Doctoral Thesis. Department of Genetics, University of Nottingham, UK.

Popescu, C. E. & Lee, R. W. 2007. Mitochondrial genome sequence evolution in *Chlamydomonas*. *Genetics*, **175**:819–826.

Ran, W. & Higgs, P. G. 2010. The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria. *Mol. Biol. Evol.*, **27**:2129–2140.

Sloan, D. B. & Taylor, D. R. 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.

Smith, D. R. & Lee, R. W. 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. *Mol. Biol. Evol.*, **25**:487–496.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, **24**:1596–1599.

Vinogradova, E., Salinas, T., Cognat, V., Remacle, C. & Marechal-Drouard, L. 2009. Steady-state levels of imported tRNAs in *Chlamydomonas* mitochondria are correlated with both cytosolic and mitochondrial codon usages. *Nucleic Acids Res.*, **37**:1521–1528.

Wang, B., Yuan, J., Liu, J., Jin, L. & Chen, J. Q. 2011. Codon usage bias and determining forces in green plant mitochondrial genomes. *J. Integr. Plant Biol.*, **53**:324–334.

Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene*, **87**:23–29.

Yang, Z. & Nielsen, R. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol. Biol. Evol.*, **25**:568–579.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. The relative abundance of dinucleotides in intergenic regions (r), dinucleotides YZ in 4-fold degenerate (FFD) codons ($R(YZ)$), and dinucleotides ZN_1 where Z belongs to FFD codons ($R(ZN_1)$) in the mitochondrial genomes of *Mesostigma viride* (NIES-296) and *Chlamydomonas reinhardtii* (CC-277).

Table S2. Codon usage in the mitochondrial genomes of *Mesostigma viride* (NIES-296) and *Chlamydomonas reinhardtii* (CC-277).

Table S3. Independence tests (χ^2 tests) between the second (Y) and the third codon positions (Z) of the 4-fold degenerate (FFD) codons, and between the third position of FFD codons and the first codon position (N_1) of the following codon in the mitochondrial genomes of *Mesostigma viride* (NIES-296) and *Chlamydomonas reinhardtii* (CC-277).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received: 11/21/11; accepted: 01/11/12