Published in final edited form as:

Mol Genet Genomics. 2014 October ; 289(5): 885–894. doi:10.1007/s00438-014-0857-x.

Bicluster Pattern of Codon Context Usages between Flavivirus and Vector Mosquito *Aedes aegypti*: Relevance to Infection and Transcriptional Response of Mosquito Genes

Susanta K. Behura and David W. Severson*

Eck Institute for Global Health and Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA

Abstract

The mosquito *Aedes aegypti* is the primary vector of dengue virus (DENV) infection in most of the subtropical and tropical countries. Besides DENV, yellow fever virus (YFV) is also transmitted by *A. aegypti*. Susceptibility of *A. aegypti* to West Nile virus (WNV) has also been confirmed. Although studies have indicated correlation of codon bias between flaviviridae and their animal/insect hosts, it is not clear if codon sequences have any relation to susceptibility of *A. aegypti* to DENV, YFV and WNV. In the current study, usages of codon context sequences (codon pairs for neighboring amino acids) of the vector (*A. aegypti*) genome as well as the flaviviral genomes are investigated. We used bioinformatics methods to quantify codon context bias in a genome-wide manner of *A. aegypti* as well as DENV, WNV and YFV sequences. Mutual information statistics was applied to perform bicluster analysis of codon context bias between vector and flaviviral sequences. Functional relevance of the bicluster pattern was inferred from published microarray data. Our study shows that codon context bias of DENV, WNV and YFV sequences varies in a bicluster manner with that of specific sets of genes of *A. aegypti*. Many of these mosquito genes are known to be differentially expressed in response to flaviviral infection suggesting that codon context sequences of *A. aegypti* and the flaviviruses may play a role in the susceptible interaction between flaviviruses and this mosquito. The bias in usages of codon context sequences likely has a functional association with susceptibility of *A. aegypti* to flaviviral infection. The results from this study will allow us to conduct hypothesis driven tests to examine the role of codon contexts bias in evolution of vector-virus interactions at the molecular level.

Keywords

Flavivirus; codon context bias; coevolution; mosquito; translational selection

Introduction

Dengue virus (DENV), yellow fever virus (YFV) and West Nile virus (WNV) are flaviviruses that cause widespread vector-borne diseases in humans throughout the world (Solomon and Mallewa, 2001; Mackenzie et al. 2004; Hayes et al. 2005; Barrett and Higgs

*Correspondence David W. Severson, Eck Institute for Global Health and Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA Tel.: + 1 574-631-3826; Fax: + 1 574 631 7413; severson.1@nd.edu.

2007; Gould and Solomon 2008; Gardner and Ryman 2010; Rossi et al. 2010; Guzman et al. 2010; Suthar et al 2013, Bhatt et al. 2013). The mosquito *Aedes aegypti* is the principal vector that transmits DENV to humans. *Aedes aegypti* is also considered a principal vector for YFV and a potential secondary vector for WNV (CDC 2013, Turell et al. 2001; Vanlandingham et al. 2007). In the absence of effective vaccines against most flaviviral diseases (Pugachev et al. 2005; Heinz and Stiasny 2012), understanding the molecular mechanisms of vector-virus interactions is critical in order to design novel disease control strategies. *Aedes aegypti* susceptibility to dengue infection is modulated by differential expression of genes of the mosquito vector in response to infection (Xi et al. 2008; Souza-Neto et al. 2009; Ramirez and Dimopoulos 2010; Girard et al. 2010; Behura et al. 2011; Colpitts et al. 2011; Chauhan et al. 2012). The study by Colpitts *et al.* 2011 revealed that a common set of genes of *A. aegypti* are associated with differential expression in response to infection by each of these three flaviviruses. In addition, we have previously shown that codon bias of genes of *A. aegypti* has a significant association with expression changes of mosquito genes to DENV infection (Behura and Severson 2012a). Furthermore, the study by Lobo *et al.* 2009 showed that flaviviruses and the respective animal and insect hosts have common patterns of usage of sequence motifs in their respective genomes. Our primary aim of the present study is to investigate if codon usage bias has any relation to susceptibility of *A. aegypti* to DENV, YFV and WNV.

Codon usage bias results due to non-random usage of synonymous codons during gene translation and the extent of such bias varies within and between species (Behura and Severson 2013a). Our recent studies have revealed genome-wide differential usage of codon sequences in different insect species including *A. aegypti* (Behura and Severson 2012b). In addition to that, we have also shown that codon bias is a significant factor in genetic diversity of DENV populations (Behura and Severson 2013b). Specifically, codon usage bias is known to have an association with translational efficiency and/or accuracy (Rocha 2006; Behura and Severson 2011; Rodriguez et al. 2012; Novoa et al. 2012) as well as expression levels of protein coding genes (Akashi 2001; Najafabadi and Salavati 2008; Najafabadi et al. 2009; Camiolo et al. 2012). Codon context bias results due to non-random usage of codon sequences in the messenger RNA that correspond to neighboring amino acids in the protein. In the simplest term, codon context refers to sequential codon positions (or codon pairs) in the mRNA. The biased usage of codon context sequences is important, considering that they correspond to the A and P sites of the ribosome during translation. The A site binds to an aminoacyl-tRNA and the P site binds to a peptidyl-tRNA. Hence, biased usages of codon context sequences can affect translation selection of genes (Moura et al. 2005). Moreover, the role of codon context sequences in gene function has been suggested in several studies (Bossi and Ruth 1980; Buckingham 1990; Buckingham 1994; Irwin et al. 1995; Moura et al. 2007). The biased usage of codon context sequences of poliovirus plays an important role in interaction with human host cells (Coleman et al. 2008). However, no investigation has been performed to explore if codon context bias has any association with the transcriptional response of mosquito genes to viral infections.

To test the above possibility, our primary aim of this study is to determine if codon context sequences of *A. aegypti* and the three flaviviruses (DENV, WNV and YFV) have any

similarity in usage patterns. We are particularly interested in the genes that have been identified as common response genes in *A. aegypti* to infection by these three viruses (Colpitts et al. 2011). The results of our study show that *A. aegypti* genes that commonly respond to these flaviviral infections show a bicluster pattern in the usages of codon context sequences with that of the flaviviruses indicating that codon contexts bias may have a role in the vector-virus interactions.

Materials and Methods

Data

The coding sequences of *A. aegypti* genes (AaegL1.4) were obtained from VectorBase (<https://www.vectorbase.org/>) using the Biomart tool. The coding sequences of the complete genome sequences of WNV, DENV and YFV were downloaded (May 2013) from the nucleotide database of NCBI (<http://www.ncbi.nlm.nih.gov/>). The list of strains and GenBank accession numbers of the three flaviviruses that are investigated in this study can be provided upon request. The set of genes in *A. aegypti* showing common patterns of transcriptional response to infection of DENV, WNV and YFV were obtained from the Colpitts *et al.* (2011) study. The orthologs of these *A. aegypti* genes in two other important vector mosquito species, *Culex quinquefasciatus* and *Anopheles gambiae*, were used for comparisons of codon context sequences. The coding sequences of these orthologs were obtained from VectorBase using the Biomart tool.

Analysis of codon context sequences

The coding sequences of *A. aegypti* genes and that of the individual virus genomes were quantified for codon context usages by the ‘Anaconda’ software (Moura et al. 2005). In order to determine the biased usage of the synonymous codons of the context sequences, the relative synonymous codon context usage (RSSCU) of each codon pair was calculated using the Anaconda tool.

Bicluster analysis of codon context bias by mutual information

The bicluster analysis of codon context sequences of *A. aegypti* and the three flaviviruses was performed using mutual information (MI) of the RSSCU values generated from Anaconda (Moura et al. 2005). MI is a probabilistic measure of mutual dependence between two variables that is estimated from the joint probability divided by the product of marginal probabilities of the variables (in our case the variables being codon context bias between virus and mosquito). As codon context bias is not an independent variable, we resorted to a shuffle-and-sample approach in order to estimate the MI values from variation of codon context usages between the virus and vector. The R code ‘shuffleAndExtract’ (<http://tata-box-blog.blogspot.com/search/label/R>) was used to shuffle the coding sequences (*A. aegypti* as well as the three flaviviruses) and then to extract equal numbers of random samples (n= 25 sequences) each from *A. aegypti* and the three flaviviruses. The shuffle-and-extract procedure was repeated 1,000 times.

Each of the 1,000 randomized datasets was subjected to calculation of RSSCU of each codon context by Anaconda. The MIs of codon context variation among randomized

sequences were calculated using the *R* package 'BioDist'. The MI values corresponding to each of the 1,000 sequence datasets were then subjected to bicluster analysis in *R* by using a three-step procedure as described by Gupta and Aggarwal (2008). In the first step, an initial subset of coding sequences was identified that showed similar variation in codon context usages among *A. aegypti* and flaviviral genomes. In the second step, the subset of codon contexts was identified that showed similar variation among the sequences identified in the first step. In the third step, a final subset of sequences was identified that showed similar variation in codon contexts usage identified from the second step. In each step, a minimum threshold value (user dependent) was used to compare the MI values to select the reduced set of sequences or codon contexts. We set 0.281 as the threshold, which is the average MI of codon contexts and sequences among the 1,000 randomized datasets. The *A. aegypti* gene(s) showing higher MI than the set threshold were considered as bicluster genes (henceforth 'BCGs') of the mosquito with the codon context bias of the three flaviviruses.

Analysis of functional relevance of BCGs

The BCGs were further investigated to know if they are also differentially expressed in *A. aegypti* in response to flaviviral infections. This was accomplished by comparisons with the results of a published microarray study (Colpitts et al. 2011). Yates Chi square (2x2 contingency) tests were performed to determine if the number of response/ non-response genes identified from the microarray experiments (Colpitts et al. 2011) were significantly associated with the number of BCGs/ non-BCG genes identified from the present study. Also, we analyzed the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways associated with the BCGs by using the KEGG mapper tool available at <http://www.genome.jp/kegg/mapper.html>. The mapped pathways were grouped into different modules based on relatedness among pathways (for example, genes associated with glycolysis, TCA cycle, galactose-, pentose- and fructose- metabolisms were collectively referred to as carbohydrate metabolism genes). The correlation of codon context variation of DENV, WNV and YFV with that of genes belonging to individual pathways was estimated by Pearson correlation and the pathway modules of BCGs were predicted from the rank order of the correlations.

Multiple matrix regression with Mantel permutations

The Mantel test (Mantel 1967) was used to determine significance of linear regression of codon context variation of *A. aegypti* BCGs and non-BCGs with that of the three flaviviruses. The RSSCU values were converted to distance matrix forms which were then subjected to linear regression using the Multi-Mantel (<http://anolis.oeb.harvard.edu/~liam/programs/>). The program implements the Mantel procedure (Mantel 1967) for multiple matrix regression and evaluates significance by data permutation. In our case, the data was permuted 10,000 times to test significance of analysis of variance in codon context bias between *A. aegypti* and the flaviviral sequences. The similar matrix regression procedure was also adopted to test significance of association of codon context bias of flaviviruses with codon context bias of orthologous BCGs and non-BCGs of *C. quinquefasciatus* and *A. gambiae*.

Cluster analysis of codon context usages

The Anaconda software was used to generate the cluster tree of codon context usages among the BCGs and the three flaviviruses, along with the non-BCGs and their orthologs in *C. quinquefasciatus* and *A. gambiae*. The variation of RSCCU values among the individual coding sequences of the flaviviruses and the mosquito genes were assessed by the complete linkage clustering method based on Euclidean distances, using the Cluster 3.0 program (de Hoon et al. 2004). The purpose of the cluster analysis was to determine if species codon context sequences are used (either preferentially or avoided) both in flavivirus as well as *A. aegypti* BCGs.

Canonical correlation of codon context bias between *A. aegypti* and flavivirus

Canonical correlation was performed using the principal coordinates method (Anderson and Willis 2003). This was accomplished by partitioning the variation of codon context usages data into the principal coordinates and then testing by permutation if the variation was significant between vector and virus. First, distance matrices were generated separately from the codon context data of *A. aegypti* BCGs and then for DENV, YFV and WNV sequences. They were then decomposed into component eigenvalues by principal coordinate analysis. The canonical axes scores (position of multivariate points on the canonical axes) were used to determine correlations of each of the original variables with the canonical axes. A permutation test was performed to determine if the correlation of codon context bias between *A. aegypti* and the flavivirus was significant. These procedures were implemented in a FORTRAN computer program named 'canonical analysis of principal coordinates (CAP)' freely available at <https://www.stat.auckland.ac.nz/~mja/Programs.htm>. To further validate the relationship of codon context variation between *A. aegypti* and flaviviruses, we compared the codon context bias of *A. aegypti* BCGs with that of tick-borne encephalitis virus as a negative control (they don't infect mosquitoes) and cell fusing agent virus as positive control that is highly adapted to *A. aegypti* (Cook et al. 2006). The codon sequences of these viruses were also obtained from NCBI.

Pair-wise *a posteriori* comparisons

To identify significant association of codon context sequences between *A. aegypti* and the flaviviruses, a permutation based Analysis of Variance (ANOVA) method (Anderson 2001) was used. The method is based on partitioning of multivariate variation (defined by a distance measure) according to individual factors. The Euclidean distance was used as the measure of variation of codon context bias in our case. The distances between each pair of observation units (i.e. codon contexts) were calculated to obtain a distance matrix which was then used to perform the test-statistics in pair-wise manner. The pair-wise *a posteriori* comparisons were performed between *A. aegypti* BCGs and flaviviruses for individual amino acids classified as 2-fold, 3-fold, 4-fold and 6-fold degenerate (Tryptophan and Methionine were excluded as they have no synonymous codons). No data transformation or standardization was done prior to analysis. Unrestricted permutation of the data was allowed to perform the permutation tests (4,999 times) to determine the statistical significance of the analyses. The *p*-value < 0.05 was considered as the threshold of significance for association

of codon context sequences between flaviviruses and vector sequences, unless stated otherwise.

Results

Quantification of codon context bias

The analysis of coding sequences of the 16,018 predicted genes of *A. aegypti* (www.vectorbase.org) revealed the usage patterns of individual codon context sequences as shown in Figure 1. The residual values of each codon context, based on its usage frequency, were calculated using the Anaconda software (Moura et al. 2005) wherein the positive and negative values showed the contrasting patterns of codon context usages in *A. aegypti* genes (Figure 1). The red color in Figure 1 represents the negative values (bad context of codons) and green color represents the positive values (good context of codons) in each analysis. The values that are not statistically significant, calculated by χ^2 test of independence as described in (Moura et al. 2005), are indicated in black (-3 to $+3$). The usage patterns of codon context sequences of flaviviruses were determined from complete coding sequences of downloaded from NCBI (<http://www.ncbi.nlm.nih.gov>). All the complete genomes (3,014 for DENV, 399 for WNV and 48 for YFV) of the three viruses available at NCBI at the time of this investigation were included. The patterns of codon context usages in these viruses are shown in Figure 2.

Bicluster patterns of codon context usages between flavivirus and *A. aegypti*

To compare patterns of codon context usages between *A. aegypti* and the three flaviviruses, the relative synonymous codon context usages (RSCCU) were calculated. We used a mutual information (MI) based bi-clustering method (Gupta and Aggarwal 2008) to determine if a specific set of genes of *A. aegypti* tend to have similar variation in the codon context usages of the flaviviruses. We adopted a sequence shuffling approach wherein equal numbers of coding sequences ($n = 25$) were chosen randomly from the mosquito and the three flaviviruses. The shuffling was repeated 1,000 times to generate randomized coding sequences consisting of 25 sequences each from the four original FASTA sequence datasets. These randomized sequence data sets of flaviviruses and *A. aegypti* genes were then analyzed separately to quantify codon context patterns using the software 'Anaconda' (Moura et al. 2005). The MI values that measure the mutual dependence of codon content variation among the coding genes of *A. aegypti* and the DENV, YFV and WNV sequences were estimated for each of the 1,000 randomized data sets. The bicluster predictions were performed using a predefined threshold value (0.281) which was the mean value of the MI scores of among all the randomized sequences. Using this threshold, we identified specific coding sequences of *A. aegypti* and the flaviviruses (Table S1) that qualified as showing the 'bicluster' pattern in codon context usages.

Variation of codon context usages between BCGs and flaviviruses

Our results further showed that the patterns of codon context usages among the three flaviviruses were more closely related to that of the BCGs than the non-BCGs of *A. aegypti* (Figure 3). The BCGs of *A. aegypti* also revealed the most similarity in codon context usages with the flaviviruses. Based on canonical correlation test (see Methods), the codon

context usages of the BCGs of *A. aegypti* showed significant correlation with codon context usages of the flaviviruses (Figure 4). Using matrix regression test (see Methods), it was further observed that the codon context bias of *A. aegypti* BCGs, but not any other gene groups (Figure 3) has statistically significant correlation ($p < 0.01$) with that of the three flaviviruses. Moreover, the correlation coefficient of BCGs with the flaviviruses was almost 3-fold higher than that of non-BCGs in *A. aegypti*, a distinction that was absent in correlations observed with *C. quinquefasciatus* and *A. gambiae* genes (Figure 3).

From the canonical correlation test, it was observed that the codon context biases of *A. aegypti* BCGs varied canonically with that of the flaviviruses, where more than 80% of the correlated changes in codon context usages could be explained by two principal axes (Figure 4). The canonical correlation was statistically significant ($p < 0.01$) in each case. The correlated variation of codon context sequences of *A. aegypti* with that of the flaviviruses is most likely associated with natural adaptation of these viruses to the mosquito. This was also evident from significant correlation of codon context bias ($p < 0.01$) between *A. aegypti* BCGs and cell fusing agent viruses but not with tick-borne encephalitis viruses. To further confirm this relationship, we also performed hierarchical cluster analysis of codon context usages in flaviviruses and *A. aegypti* BCGs. We observed that specific codon context sequences are used in biased manner in both mosquito and virus sequences in the form of distinct clusters. This is evident in Figure 5 where the *A. aegypti* sequences (shown by solid vertical line on the right) and the flaviviral sequences (shown by dotted vertical line on the right) coexist in four separate clusters.

Relevance of the BCGs to *A. aegypti* susceptibility to flaviviruses

When compared with the microarray expression data obtained for the Rockefeller strain of *A. aegypti* (Colpitts *et al.* 2011), a total of 85 of the significantly differentially expressed genes ($n = 405$) identified from the microarray study were also represented in the BCGs (Table S1). Of these 85 genes, 51 genes were upregulated at least 2 fold in response to each of the three flavivirus at all the three post-infection time points examined in the Colpitts *et al.* 2011 study. Taking into account the total number of coding genes in *A. aegypti* genome assembly as 16,018 (AaegL1.4), Yates Chi-square tests shows statistically significant association ($\chi^2 = 1190.9$, $p < 0.0001$) between the differentially expressed genes from the microarray study [21] and the BCGs identified from the current study. To determine if the observed correlation is specific to the Rockefeller strain or may also be associated other strains of *A. aegypti*, the dengue response genes identified in the DENV susceptible Moyo-S strain (Chauhan *et al.* 2012) were investigated. The comparison of codon context variation of BCGs with that of dengue response genes in Moyo-S also revealed significant association ($p < 0.0001$) suggesting that a bicluster pattern between *A. aegypti* and flaviviruses is not specific to the Rockefeller genotype but may be a general feature of the *A. aegypti* genome. Furthermore, we also identified expressed sequence tags (ESTs) that unambiguously mapped to a subset of BCGs (~48%). These ESTs were sequenced from different strains of *A. aegypti* by various labs and submitted to the EST database of NCBI. After generating the codon alignment of ESTs with that of BCGs (using blastx method described at http://cgpdb.ucdavis.edu/COS_Arabidopsis/Codon_Usage_Pipeline.html), the codon context variation of ESTs was compared with that of the flaviviruses. It was observed that the EST

sequences also showed similar levels of correlation with the flaviviruses as that of BCGs (data not shown) indicating that association of codon context bias between *A. aegypti* and flaviviruses is statistically significant across different strains of *A. aegypti*.

To further understand the functional significance of the BCGs, we mapped the BCGs to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways of *A. aegypti*. It was expected that genes with similar bias in codon usages would tend to map to the same pathways (Najafabadi et al. 2009). We observed three specific pathway modules 1) transport, 2) signaling and 3) glycan biosynthesis that were identified as top ranking functions of the BCGs (Table 1). This assessment was based on the rank order correlation of codon context bias of KEGG pathway genes with the coding sequences of the three flaviviruses. From this analysis it was also observed that each pathway retained the same rank order of correlation with each of the three flaviviruses, although the rank orders generally varied among the pathways.

Role of Arg and Ser residues in variation of codon context bias between *A. aegypti* and the flaviviruses

It was further observed that different codon context sequences had differential association between *A. aegypti* and flavivirus based on the PERMANOVA (Anderson 2001) analysis (Table 2). The results of this analysis showed that biased usages of specific codon contexts of each of the three flaviviruses vary in a significantly correlated manner ($p < 0.05$) with the codon context biases of either the commonly up-regulated genes or the commonly down-regulated genes of the mosquito (see Table S1). It was found that codon context bias of AGA and CGC (coding for Arg) and AGC, UCC and UCU (coding for Ser) was statistically significant between the differentially expressed genes of *A. aegypti* and the flavivirus (Table 3) suggesting their possible role in translational selection of genes in the mosquito vector and the flaviviruses.

Discussion

The present study was inspired by the results obtained from the study by Colpitts *et al.* 2011 which suggested that specific genes of *A. aegypti* commonly respond to infection with DENV, YFV and WNV. Hence, we wanted to know if these common response genes might have similar codon context biases with that of the flavivirus's. Because the *A. aegypti* translational machinery is exploited by the virus to synthesize viral proteins to complete its life cycle within the mosquito, translational selection of *A. aegypti* and flaviviral coding sequences may affect molecular interactions between the mosquito and virus. It has been experimentally verified in poliovirus where replacement of optimized codons of the virus with rare synonymous codons attenuated the virus replication in the infected cells (Mueller et al. 2006; Coleman et al. 2008). Moreover, biased usage of synonymous codons is predicted to have a significant role in transcriptional responses of *A. aegypti* genes to DENV infection (Behura and Severson, 2012a). Hence, it was reasonable to investigate if the common response genes of *A. aegypti* to the flaviviral infections might have similar patterns of codon context usages with that of the virus.

Considering our results that the BCGs show significant association with the response genes of *A. aegypti* to infection of the flaviviruses, it is likely that the observed bicluster pattern of codon context usages has a functional role in the interaction between mosquito and flavivirus. Furthermore, the predicted functions of BCGs are also consistent with the predicted functions of the differentially-expressed *A. aegypti* genes identified from microarray studies (Souza-Neto et al. 2009; Behura et al. 2011; Colpitts et al. 2011; Chauhan et al. 2012). The 'transport' function is predicted as one of the major functions of the BCGs from this study. The earlier studies (Behura et al. 2011; Colpitts et al. 2011; Sim and Dimopoulos 2010) also identified such function of the genes from microarray studies. It is likely that a susceptible vector-virus interaction is dependent upon genes related to transport (such as vacuolar trafficking processes) in the infected cell (Hsieh et al. 2008). Similarly, signal transduction is an important feature of molecular interactions between flavivirus and *A. aegypti*, which is also predicted by the BCGs as well as from previous investigations (Souza-Neto et al. 2009; Behura et al. 2011).

The codon contexts that are GC-rich are generally used in low frequency in the BCGs of *A. aegypti* as well as the flaviviral sequences (data not shown). The GC-rich genes are generally more biased in codon usages than GC-poor genes of *A. aegypti* (Behura and Severson 2012b). Hence, the result of the bicluster analysis is consistent to our earlier report that high codon bias has a negative marginal effect on vector gene responsiveness to dengue infection (Behura and Severson 2012a). The avoidance of GC-rich codons in flaviviral sequences was also evident from an earlier study (Behura and Severson 2013a). It was observed from that study that AT-rich contexts [particularly, (A)(A/T)(A)-(A)(A/T)(A) pairs] are highly abundant codon contexts, whereas the GC-rich contexts [particularly, (C/G)(C/A)(C/G)-(C/G)(C/A)(C/G) pairs] are generally avoided in the DENV genome.

The results of the present study further indicate that Serine and Arginine have significant association in the relative usage of the codon context sequences between flavivirus and the mosquito. This is an interesting observation because investigation of the DENV genome (Behura and Severson 2013a) also revealed that intracodon variation of Leu, Arg and Ser coding sequences was higher compared to the codons of other amino acids. It was further discovered from that study that the 1st position of Ser and Arg codons, but not Leu codons, was under selection constraint. Most of the variation in Arg and Ser codons in the DENV genome were associated with synonymous changes contributing to biased usage of alternate codons of Ser and Arg. In agreement with the results of current investigation, it is likely that Arg and Ser codons may play roles in translational selection of coding sequences in *A. aegypti* and the flaviviruses.

This study showed that codon context bias varies in a bicluster manner between flaviviruses and genes of the *A. aegypti* mosquito. The bias in usages of codon context sequences has a functional association of *A. aegypti* interaction with DENV, YFV and WNV. Many of these *A. aegypti* bicluster genes are known to be differentially expressed in response to infection by these flaviviruses suggesting that codon context sequences may have a role in the susceptible interaction between flaviviruses and this mosquito. Although varying degrees of susceptibility of different mosquito populations to dengue virus is known, the role of codon context bias in variation of mosquito-virus interaction has not been tested. The results from

this study may promote such research to better understand the molecular evolution of vector-virus interactions. In particular, the results from this study will allow us to conduct experiments to examine the role of codon contexts bias in evolution of vector-virus interactions at the molecular level.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are thankful to Casey Hill for critically reading the manuscript. This study was supported in part from grant RO3-TW008138, Fogarty International Research Collaboration Award (FIRCA), National Institutes of Health (NIH), USA.

References

- Akashi H. Gene expression and molecular evolution. *Curr Opin Genet Dev.* 2001; 11:660–666. [PubMed: 11682310]
- Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Aust Ecol.* 2001; 26:32–46.
- Anderson MJ, Willis TJ. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology.* 2003; 84:511–525.
- Barrett AD, Higgs S. Yellow fever: a disease that has yet to be conquered. *Annu Rev Entomol.* 2007; 52:209–229. [PubMed: 16913829]
- Behura SK, Gomez-Machorro C, Harker BW, deBruyn B, Lovin DD, Hemme RR, Mori A, Romero-Severson J, Severson DW. Global cross-talk of genes of the mosquito *Aedes aegypti* in response to dengue virus infection. *PLoS Negl Trop Dis.* 2011; 5:e1385. [PubMed: 22102922]
- Behura SK, Severson DW. Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in *Aedes aegypti* and *Anopheles gambiae*. *Insect Mol Biol.* 2011; 20:177–187. [PubMed: 21040044]
- Behura SK, Severson DW. Intrinsic features of *Aedes aegypti* genes affect transcriptional responsiveness of mosquito genes to dengue virus infection. *Infect Genet Evol.* 2012a; 12:1413–1418. [PubMed: 22579482]
- Behura SK, Severson DW. Comparative analysis of codon usage bias and codon context patterns between Dipteran and Hymenopteran sequenced genomes. *PLoS One.* 2012b; 7:e43111. [PubMed: 22912801]
- Behura SK, Severson DW. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol Rev.* 2013a; 88:49–61. [PubMed: 22889422]
- Behura SK, Severson DW. Nucleotide substitutions in dengue virus serotypes from Asian and American countries: insights into intracodon recombination and purifying selection. *BMC Microbiol.* 2013b; 13:37. [PubMed: 23410119]
- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI. The global distribution and burden of dengue. *Nature.* 2013; 496:504–507. [PubMed: 23563266]
- Bossi L, Ruth JR. The influence of codon context on genetic code translation. *Nature.* 1980; 286:123–127. [PubMed: 7402305]
- Buckingham RH. Codon context. *Experientia.* 1990; 46:1126–1133. [PubMed: 2253710]
- Buckingham RH. Codon context and protein synthesis: enhancements of the genetic code. *Biochimie.* 1994; 76:351–354. [PubMed: 7849098]
- Camiolo S, Farina L, Porceddu A. The relation of codon bias to tissue-specific gene expression in *Arabidopsis thaliana*. *Genetics.* 2012; 192:641–649. [PubMed: 22865738]

- Center for Disease Control and Prevention. [21 December 2013] West Nile virus in the United States: Guidelines for surveillance, prevention, and control. 2013. <http://www.cdc.gov/ncidod/dvbid/westnile/mosquitospecies.htm>.
- Chauhan C, Behura SK, Debruyne B, Lovin DD, Harker BW, Gomez-Machorro C, Mori A, Romero-Severson J, Severson DW. Comparative expression profiles of midgut genes in dengue virus refractory and susceptible *Aedes aegypti* across critical period for virus infection. *PLoS One*. 2012; 7:e47350. [PubMed: 23077596]
- Coleman JR, Papamichail D, Skiena S, Fitcher B, Wimmer E, Mueller S. Virus attenuation by genome-scale changes in codon pair bias. *Science*. 2008; 320:1784–1787. [PubMed: 18583614]
- Colpitts TM, Cox J, Vanlandingham DL, Feitosa FM, Cheng G, Kurscheid S, Wang P, Krishnan MN, Higgs S, Fikrig E. Alterations in the *Aedes aegypti* transcriptome during infection with West Nile, dengue and yellow fever viruses. *PLoS Pathog*. 2011; 7:e1002189. [PubMed: 21909258]
- Cook S, Bennett SN, Holmes EC, De Chesse R, Moureau G, de Lamballerie X. Isolation of a new strain of the flavivirus cell fusing agent virus in a natural mosquito population from Puerto Rico. *J Gen Virol*. 2006; 87:735–748. [PubMed: 16528021]
- de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004; 20:1453–1454. [PubMed: 14871861]
- Gardner CL, Ryman KD. Yellow fever: a reemerging threat. *Clin Lab Med*. 2010; 30:237–260. [PubMed: 20513550]
- Girard YA, Mayhew GF, Fuchs JF, Li H, Schneider BS, McGee CE, Rocheleau TA, Helmy H, Christensen BM, Higgs S, Bartholomay LC. Transcriptome changes in *Culex quinquefasciatus* (Diptera: Culicidae) salivary glands during West Nile virus. *J Med Entomol*. 2010; 47:421–435. [PubMed: 20496590]
- Gould EA, Solomon T. Pathogenic flaviviruses. *Lancet*. 2008; 371:500–509. [PubMed: 18262042]
- Gupta N, Aggarwal S. MIBiClus: Mutual Information based biclustering algorithm. *Int J Comp Sc*. 2008; 3:2.
- Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, Hunsperger E, Kroeger A, Margolis HS, Martínez E, Nathan MB, Pelegrino JL, Simmons C, Yoksan S, Peeling RW. Dengue: a continuing global threat. *Nat Rev Microbiol*. 2010; 8:S7–16. [PubMed: 21079655]
- Hayes EB, Sejvar JJ, Zaki SR, Lanciotti RS, Bode AV, Campbell GL. Virology, pathology, and clinical manifestations of West Nile virus disease. *Emerg Infect Dis*. 2005; 11:1174–1179. [PubMed: 16102303]
- Heinz FX, Stiasny K. Flaviviruses and flavivirus vaccines. *Vaccine*. 2012; 30:4301–4306. [PubMed: 22682286]
- Hsieh SC, Liu JJ, King CC, Chang GJ, Wang WK. A strong endoplasmic reticulum retention signal in the stem-anchor region of envelope glycoprotein of dengue virus type 2 affects the production of virus-like particles. *Virology*. 2008; 374:338–350. [PubMed: 18252258]
- Irwin B, Heck JD, Hatfield GW. Codon pair utilization biases influence translational elongation step times. *J Biol Chem*. 1995; 270:22801–22806. [PubMed: 7559409]
- Lobo FP, Mota BE, Pena SD, Azevedo V, Macedo AM, Tauch A, Machado CR, Franco GR. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One*. 2009; 4:e6282. [PubMed: 19617912]
- Mackenzie JS, Gubler DJ, Petersen LR. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nat Med*. 2004; 10:S98–109. [PubMed: 15577938]
- Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. 1967; 27:209–220. [PubMed: 6018555]
- Moura G, Pinheiro M, Arrais J, Gomes AC, Carreto L, Freitas A, Oliveira JL, Santos MA. Large scale comparative codon-pair context analysis unveils general rules that fine-tune evolution of mRNA primary structure. *PLoS One*. 2007; 2:e847. [PubMed: 17786218]
- Moura G, Pinheiro M, Silva R, Miranda I, Afreixo V, Dias G, Freitas A, Oliveira JL, Santos MA. Comparative context analysis of codon pairs on an ORFeome scale. *Genome Biol*. 2005; 6:R28. [PubMed: 15774029]

- Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol*. 2006; 80:9687–9696. [PubMed: 16973573]
- Najafabadi HS, Goodarzi H, Salavati R. Universal function-specificity of codon usage. *Nucleic Acids Res*. 2009; 37:7014–7023. [PubMed: 19773421]
- Najafabadi HS, Salavati R. Sequence-based prediction of protein-protein interactions by means of codon usage. *Genome Biol*. 2008; 9:R87. [PubMed: 18501006]
- Novoa EM, Ribas de Pouplana L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet*. 2012; 28:574–581. [PubMed: 22921354]
- Pugachev KV, Guirakhoo F, Monath TP. New developments in flavivirus vaccines with special attention to yellow fever. *Curr Opin Infect Dis*. 2005; 18:387–394. [PubMed: 16148524]
- Ramirez JL, Dimopoulos G. The Toll immune signaling pathway control conserved anti-dengue defenses across diverse *Ae. aegypti* strains and against multiple dengue virus serotypes. *Dev Comp Immunol*. 2010; 34:625–629. [PubMed: 20079370]
- Rocha EP. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res*. 2004; 14:2279–2286. [PubMed: 15479947]
- Rodriguez O, Singh BK, Severson DW, Behura SK. Translational selection of genes coding for perfectly conserved proteins among three mosquito vectors. *Infect Genet Evol*. 2012; 12:1535–1542. [PubMed: 22705463]
- Rossi SL, Ross TM, Evans JD. West Nile virus. *Clin Lab Med*. 2010; 30:47–65. [PubMed: 20513541]
- Sim S, Dimopoulos G. Dengue virus inhibits immune responses in *Aedes aegypti* cells. *PLoS One*. 2010; 5:e10678. [PubMed: 20502529]
- Solomon T, Mallewa M. Dengue and other emerging flaviviruses. *J Infect*. 2001; 42:104–115. [PubMed: 11531316]
- Souza-Neto JA, Sim S, Dimopoulos G. An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. *Proc Natl Acad Sci U S A*. 2009; 106:17841–17846. [PubMed: 19805194]
- Suthar MS, Diamond MS, Gale M Jr. West Nile virus infection and immunity. *Nat Rev Microbiol*. 2013; 11:115–128. [PubMed: 23321534]
- Turell MJ, O'Guinn ML, Dohm DJ, Jones JW. Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus. *J Med Entomol*. 2001; 8:130–134. [PubMed: 11296813]
- Vanlandingham DL, McGee CE, Klinger KA, Vessey N, Fredregillo C, Higgs S. Relative susceptibilities of South Texas mosquitoes to infection with West Nile virus. *Am J Trop Med Hyg*. 2007; 77:925–928. [PubMed: 17984355]
- Xi Z, Ramirez JL, Dimopoulos G. The *Aedes aegypti* toll pathway controls dengue virus infection. *PLoS Pathog*. 2008; 4:e1000098. [PubMed: 18604274]

Figure 1.

Variation of codon context bias in *A. aegypti* genome. The codon pairs are shown in rows (3' codon) and columns (5' codon). The corresponding amino acids of these codons are also shown. The color codes represent the residual values of codon context usages where the red represents the negative values (bad context) and green the positive values (good context). The grey color represents values of zero. The values that are not statistically significant are indicated in black (-3 to $+3$). The scale of the color code is shown below the figure.

Figure 2.

Variation of codon context bias in yellow fever virus, West Nile virus and dengue virus. The codon pairs are shown in rows (3' codon) and columns (5' codon). The corresponding amino acids of these codons are also shown. The color codes represent the residual values of codon context usages where the red represents the negative values (bad context) and green the positive values (good context). The grey color represents values of zero. The values that are not statistically significant are indicated in black (-3 to $+3$). The scale of the color code is shown below each figure.

Figure 3.

Cluster tree of flaviviruses and mosquito genes based on variation of codon context usages. The mosquito names are abbreviated as Aaeg (*A. aegypti*), Agam (*A. gambiae*) and Cqui (*C. quinquefasciatus*). The BCGs of *A. aegypti* and the three flaviviral sequences are those listed in Table S1. The bicluster and non-bicluster genes of *A. gambiae* and *C. quinquefasciatus* are the orthologs of bicluster genes and non-bicluster genes of *A. aegypti*, respectively. The bar graph on the right of the cluster tree shows differential correlation in eigenvalues of codon context usages of the mosquito gene groups with that of the three flaviviruses (collectively represented by a vertical line). The horizontal scale below the bar graph shows the axis of correlation values. The corresponding correlation values for each bar are also shown.

Figure 4.

Canonical correlation of codon context bias of *A. aegypti* genes with codon context bias of flaviviruses. The eigenvalues (shown in y-axis) of correlation of codon context bias of *A. aegypti* response genes (see Table S1) with codon context bias of the flaviviruses (shown in x-axis) are shown. The dark and the light bars in the graph represent the first and the second principal coordinates of the canonical correlation test respectively.

Figure 5.

Patterns of biased usage of synonymous codon context bias of flaviviruses and the *A. aegypti* BCGs identified from mutual information based bicluster method (see Table S1). The variation among codon context biases is indicated by differential color; red to black: red being high bias and black being low bias. The patterns were identified by hierarchical cluster analysis of the relative synonymous codon context usages (RSCCU) of each codon pair. The 5' codon of each context sequence is shown on the top of the cluster. The 3'-codon of the contexts correspond to the tips of the cluster tree shown on the right (the codon sequences for each tip are not shown). Four major clusters were identified as indicated by A through D. The *A. aegypti* and flaviviral codon contexts that co-cluster within each grouping are shown by either a solid or a dashed vertical line, respectively. The start and end of these lines match to the node in the cluster tree that separate the mosquito and virus codon sequences.

Table 1

Rank order correlation of codon context bias of *Aedes aegypti* pathway genes with codon context bias of flavivirus sequences.

<i>A. aegypti</i> pathway genes	Flavivirus
Glycan Biosynthesis and Metabolism	0.161
Membrane Transport	0.161
Signaling Molecules and Interaction	0.160
Endocrine System	0.147
Metabolism of Terpenoids and Polyketides	0.139
Metabolism of Cofactors and Vitamins	0.133
Amino Acid Metabolism	0.125
Carbohydrate Metabolism	0.123
Transport and Catabolism	0.122
Translation	0.121
Signal Transduction	0.121
Folding, Sorting and Degradation	0.120
Lipid Metabolism	0.120
Nucleotide Metabolism	0.115
Energy Metabolism	0.105
Transcription	0.105
Sensory System	0.103
Replication and Repair	0.099
Development	0.096
Environmental Adaptation	0.095
Metabolism of Other Amino Acids	0.094
Xenobiotics Biodegradation and Metabolism	0.092

Table 2

Pair-wise *a posteriori* comparison of codon context bias of different amino acids between *A. aegypti* BCGs and the flavivirus sequences (see Table S1). The amino acids are shown by three letters abbreviations. The t-statistic values and the *p*-values of significance (within parenthesis) are shown for each test. The entries in bold (Arg and Ser) show significant variation of their codon contexts between mosquito and the flavivirus.

Amino acid	<i>A. aegypti</i> up-regulated genes			<i>A. aegypti</i> down-regulated genes		
	DENV	WNV	YFV	DENV	WNV	YFV
Asn	1.3414 (0.3347)	1.164 (0.3167)	1.0197 (0.6873)	1.4308 (0.3437)	1.322 (0.3347)	1.1471 (0.3237)
Asp	1.3769 (0.3267)	1.2644 (0.3377)	1.2128 (0.3247)	1.262 (0.3556)	1.1536 (0.3337)	1.1534 (0.3147)
Cys	1.005 (0.6474)	1.0669 (0.3317)	1.1494 (0.3247)	1.1648 (0.3237)	1.275 (0.3477)	1.3136 (0.3397)
Gln	1.4482 (0.3487)	1.2657 (0.3457)	1.4128 (0.3047)	1.3939 (0.3407)	1.2017 (0.3247)	1.4091 (0.3247)
Glu	1.1091 (0.6753)	1.4281 (0.3037)	1.1505 (0.3357)	1.13 (0.3307)	1.5278 (0.3267)	1.1212 (0.3457)
His	1.1024 (0.3417)	0.9518 (0.6563)	1.2003 (0.3447)	1.1494 (0.3287)	0.9882 (1)	1.227 (0.3397)
Lys	0.3367 (0.2767)	1.3212 (0.3487)	1.4027 (0.3606)	1.3297 (0.3277)	1.3258 (0.3646)	1.39 (0.3576)
Phe	0.3157 (0.3467)	1.1609 (0.3337)	1.1478 (0.6853)	1.1894 (0.3437)	1.1088 (0.3506)	1.0583 (0.6533)
Tyr	0.3237 (0.3097)	1.1311 (0.3417)	0.9433 (0.6573)	1.2105 (0.3506)	1.2067 (0.3556)	1.0033 (0.6484)
Ile	0.2897 (0.2997)	1.0819 (0.3017)	0.9189 (0.7113)	1.2239 (0.1209)	1.1572 (0.2907)	0.8934 (0.6833)
Ala	0.023 (0.0809)	1.3392 (0.0549)	1.3089 (0.059)	1.4519 (0.057)	1.4289 (0.052)	1.2965 (0.0529)
Gly	0.2128 (0.2388)	1.2229 (0.1139)	1.1195 (0.2557)	1.1613 (0.3117)	1.0997 (0.3177)	0.9997 (0.3467)
Pro	0.1099 (0.1429)	1.2712 (0.1009)	1.4645 (0.0609)	1.4371 (0.1369)	1.2146 (0.1978)	1.4035 (0.1289)
Thr	0.0769 (0.1618)	1.1645 (0.1798)	1.1154 (0.2557)	1.6111 (0.0839)	1.4801 (0.052)	1.4165 (0.059)
Val	0.0599 (0.0539)	1.2599 (0.1479)	1.1948 (0.2208)	1.6348 (0.051)	1.3754 (0.1079)	1.2943 (0.1978)
Arg	0.021 (0.025)	1.7179 (0.02)	1.5935 (0.021)	2.0242 (0.014)	1.8696 (0.015)	1.7818 (0.003)
Leu	0.1369 (0.1908)	1.1233 (0.2218)	1.0797 (0.2937)	1.395 (0.0659)	1.3261 (0.1019)	1.3206 (0.0889)
Ser	0.015 (0.013)	1.7179 (0.002)	1.6569 (0.002)	1.9814 (0.001)	1.9263 (0.002)	1.8109 (0.003)

Table 3

PERMANOVA results of codon context bias of Arginine and Serine amino acids of *A. aegypti* BCGs and the three flaviviruses (see Table S1).

Comparison	t-stat	p-value
AGA vs. CGC	1.2429	0.0459
AGC vs. UCC	1.2811	0.025
AGC vs. UCU	1.2572	0.015