Published in final edited form as:

Insect Mol Biol. 2011 April ; 20(2): 177–187. doi:10.1111/j.1365-2583.2010.01055.x.

Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in *Aedes aegypti* and *Anopheles gambiae*

Susanta K. Behura and David W. Severson*

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

Abstract

The transfer RNAs (tRNA) are essential components of translational machinery. We determined that tRNA isoacceptors (tRNAs with different anticodons but incorporating the same amino acid in protein synthesis) show differential copy number abundance, genomic distribution patterns and sequence evolution between *Aedes aegypti* and *Anopheles gambiae* mosquitoes. The tRNA-Ala genes are present in unusually high copy number in the *Ae. aegypti* genome but not in *An. gambiae*. Many of the tRNA-Ala genes of *Ae. aegypti* are flanked by a highly conserved sequence that is not observed in *An. gambiae*. The relative abundance of tRNA isoacceptor genes is correlated with preferred (or optimal) and non-preferred (or rare) codons for ~2–4% of the predicted protein coding genes in both species. The majority (~74–85%) of these genes are related to pathways involved with translation, energy metabolism and carbohydrate metabolism. Our results suggest these genes and the related pathways may be under translational selection in these mosquitoes.

Keywords

gene cluster; sequence divergence; anticodon; selection; metabolic pathway

Introduction

The genome biology of *Anopheles* and *Aedes* mosquitoes is of great interest because of their roles as major vectors for malaria and dengue disease transmission, respectively. The complete genomes of *Anopheles gambiae* (Holt *et al.* 2000) and *Aedes aegypti* (Nene *et al.* 2007) have been reported. These projects have generated opportunities to better understand the comparative genomics and evolution of important mosquito genes (Waterhouse *et al.* 2007, 2008). In this regard, the transfer RNA (tRNA) genes are important for evolutionary investigations as they are key components of translational pathways. The tRNA genes transcribe 75–95 nt functional RNA molecules that are required at the ribosome for transfer of cognate amino acids to a growing polypeptide chain. Because of their ubiquitous nature and high sequence conservation, tRNAs have been attractive gene models for evolutionary studies (Cedergren *et al.* 1980; Kubli 1981; Percudani *et al.* 1997; Björk *et al.* 2001; Marck and Grosjean 2002; Goodenbour and Pan 2006; Sun and Caetano-Anollés 2008; Fujishima *et al.* 2008).

*Correspondence: David W. Severson, Phone: 574-631-3826, FAX: 574-631-7413, severson.1@nd.edu.

Supporting Information

The paper is accompanied with additional tables (Table S1-S8) and figures (Figure S1-S5) as supporting data.

Each tRNA is conservatively folded into an L-shaped tertiary structure. Such a structure facilitates base-pairing of its anticodon region to the corresponding codon on the mRNA, as well as attachment with the correct cognate amino acid at the CCA-3' end. The genes of different tRNA isotypes are differentially selected at the sequence level in different species. Isoacceptors are tRNAs that have different anticodons but still carry the same amino acid. Mutations in the anticodon site can generate new isoacceptor tRNA families. However, isoacceptor tRNA genes are not necessarily constrained within a monophyletic isotype group and may evolve in polyphyletic groups suggesting selection on other parts of the tRNAs than the anticodon region (Saks and Sampson 1995; Cognat *et al.* 2008).

One of the interesting and yet intriguing aspects of tRNA evolution is its role on codon usage bias. Codon usage bias is an evolutionary process in which particular codons are more frequently represented than other synonymous codons in certain types of genes. Correlation between codon usage bias and tRNA gene abundance has been demonstrated in many prokaryotes as well as some eukaryotes (Kanaya *et al.* 1999, 2001; Rocha 2004; Higgs and Ran 2008). Evidence for correlation between tRNA abundance and protein synthesis was initially observed from the cytoplasmic diffusion of the tRNA-EF-Tu-GTP complex to the ribosome as the rate-limiting step in the elongation of the polypeptide chain (Varenne *et al.* 1984). Subsequently, it was shown that the most favorable codons are the ones that correspond to the most abundant and efficient cognate aminoacylated-tRNA present in the cell (Andersson and Kurland 1990). In this context, it was realized that codon preference could be linked to translational efficiency of a select set of genes depending upon their functional roles. Since then, numerous reports have shown links between codon bias and selection for translational efficiency (Hershberg and Petrov 2008).

Generally, highly expressed genes (such as ribosomal protein genes) tend to reflect greater codon bias than lowly expressed genes (Hershberg and Petrov 2008). The up-regulated or down-regulated expression of tRNA genes, even when present in the same copy numbers in the genome, can account for differential representation of individual tRNA molecules. These can result in the unequal charging of tRNAs for synonymous codons. However, in some fast-growing organisms, the same result is achieved by duplicating the tRNA gene copies in the genome. The number of tRNA genes varies among species (Genomic tRNA Database: <http://lowelab.ucsc.edu/GtRNAdb/#eukarya>). The role of genomic tRNA gene abundance on codon usage bias is known for some organisms (Ikemura 1985; Kano *et al.* 1991; Yamao *et al.* 1991; Percudani *et al.* 1997; Kanaya *et al.* 1999, 2001; Duret 2000). The link between tRNA genes and codon bias is a dynamic evolutionary process that varies from species to species depending upon the species-specific translational efficiency of protein coding genes (Rocha 2004; Higgs and Ran 2008).

In this study, we investigated the tRNA genes of two important mosquitoes, *Ae. aegypti* and *An. gambiae*, to better understand their genomic structure and molecular evolution. Although the tRNA genes for both species have been identified from the genome sequencing projects, their evolutionary features have not been studied. We also explored the role of tRNA gene copy number on codon usage patterns of protein coding genes of these mosquitoes. Our data identify co-evolutionary patterns between tRNA isoacceptor copy number and codon usage bias in genes associated with specific cellular functions such as translation, energy and carbohydrate metabolism.

Results

Variation in copy number of tRNA genes

A total of 906 and 441 tRNA genes were predicted from the genome annotation of *Ae. aegypti* and *An. gambiae*, respectively (Table 1). However, analysis with the tRNAScan-SE

program (Lowe and Eddy 1997) identified 111 of these annotated tRNA genes in *Ae. aegypti* and a single gene in *An. gambiae* as pseudogenes. Most of the tRNA genes showed higher copy number (except for Isoleucine and Valine) in *Ae. aegypti* than *An. gambiae*. The tRNA-Ala gene copy number was much higher in *Ae. aegypti* ($n = 206$) compared to that of *An. gambiae* ($n = 28$). Analyses of the 1 kb regions flanking each tRNA gene indicated that about half of the *Ae. aegypti* tRNA-Ala genes ($n=103$) contained flanking conserved sequences. Each of these genes was found to contain a moderately conserved ($\pi = 0.17$) flanking sequence at one end and a less conserved sequence ($\pi = 0.38$) at the other end. These conserved flanking regions contain remnants (18–59 bp) of tRNA-Ala-like sequences and are located eight bp from the tRNA gene in many cases. The *An. gambiae* tRNA-Ala genes are not associated with similar conserved flanking sequences. We did not identify examples of conserved sequences in the flanking regions for other tRNA genes in either species.

The Relative Abundance of Isoacceptor tRNA (RAIT) genes varies between the two mosquitoes (Table 2). The RAIT index represents the relative increase or decrease in the observed number of individual isoacceptor tRNA genes compared to the expected number if there were no bias in their copy number. Genes predicted as possible pseudogenes (Table S1) were not included in estimating RAIT values. The tRNAs with $RAIT > 1$ represent relatively higher abundance isoacceptors than expected, whereas tRNAs with $RAIT < 1$ represent lower abundance isoacceptors than expected in the genome. Although some tRNA genes had similar abundance of isoacceptor copies in both genomes, RAIT values varied from gene to gene (Table 2), and some isoacceptor tRNAs were abundant in one genome but not in the other.

Genomic distributions of tRNA genes

Examinations of tRNA gene locations throughout the genomes of both species indicated that some tRNA genes reside in small physical clusters. There are 61 clusters of tRNA genes (ranging from 2 to 13 tRNAs per cluster) in *Ae. aegypti* and 39 clusters (ranging from 2 to 37 tRNAs per cluster) in *An. gambiae* (Table S2). Thus, the *Ae. aegypti* tRNA clusters contain fewer numbers of tRNA genes within individual clusters than *An. gambiae*. In addition, the clustered genes are often localized in different orientations to each other in both genomes. There are 87 tRNA gene pairs that are in the same orientation and 66 pairs in opposite orientation in the *Ae. aegypti* genome. In the *An. gambiae* genome, we identified 107 tRNA gene pairs that are in the same orientation and 32 pairs in opposite orientation.

In *An. gambiae*, a large cluster of 37 tRNA-Val genes was identified in chromosome 2L (Figure 1A). Interestingly, 28 of these 37 genes are separated by an exact 162 bp intergenic region. The intergenic regions separating the remaining genes in this cluster were 163 bp, 166 bp, 158 bp or 99 bp. Another cluster consisting of nine tRNA-Pro genes was also observed in chromosome 3L (Figure 1B) where the intergenic region between most of these genes was 130 bp. The *Ae. aegypti* genome contains a tRNA-Met gene repeat structure where the intergenic distance between the clustered pairs was ~86 bp (Figure 1C). We also observed tRNA-Tyr gene clusters in both genomes, although with less defined structures (Figure 1D, E). We also observed several other examples of tRNA gene clusters in both genomes (Figure S1).

We observed that some tRNA genes were localized within expressed sequence tags (ESTs) in both genomes (Table S3). Twelve cDNAs were identified from the *Ae. aegypti* EST database that contained embedded tRNA gene sequences. Seven of these tRNA genes are in the opposite strand to their corresponding cDNA and the other five tRNAs are in the same strand as that of the cDNA. While tRNAs residing within cDNAs in an antisense orientation are, by definition, transcribed independently from the mRNA, the sense-oriented tRNAs

may be produced from the common transcript with their host mRNA. Interestingly, five of these 12 ESTs were identified as pseudogenes of tRNAs and all of these are in antisense orientation to the host mRNA. In *An. gambiae*, we identified three cDNAs with embedded complete tRNA genes, but none with a pseudogene. In addition, we also identified a tRNA-Met gene located in the sense orientation in the 3'-UTR of the protein coding gene AAEL003157 of *Ae. aegypti* (data not shown). Similarly a tRNA-Tyr gene was found in the 3'-UTR of the protein coding gene AGAP007440 of *An. gambiae*, but in antisense orientation.

Sequence variation of tRNA genes

The tRNA genes show differential patterns of sequence evolution between *Ae. aegypti* and *An. gambiae*. The Trp, His, Cys, Phe and Asp tRNA genes are highly conserved whereas the Ser, Ala, Met and Thr tRNA genes are least conserved between the two species (Figure 2). The remaining tRNA genes are moderately conserved. A sliding window sequence analysis was performed to measure nucleotide differences in different parts of genes in both species and at least 65% of sites were found conserved even in the least conserved tRNA genes. A significant correlation ($P < 0.05$) was observed between sequence diversity (π - values) of the anticodon arm and the A-box (internal promoter) region in all tRNA genes within each genome. In addition, sequence variation in the B-box promoter showed significant correlation ($P < 0.05$) with a triplet (63–65) just downstream of this promoter and a part of the T ψ C arm. Potentially, the promoters and the associated regions may have some co-evolutionary relationship that is instrumental for transcriptional control of tRNAs.

Each of the Tyr-tRNA genes contains an intron in both species, while the tRNA-Ile genes and tRNA-Leu genes are present in both intron-containing and intron-less forms (Table S4). The introns are inserted at position 38 in all Tyr-tRNA genes and at position 39 in the tRNA-Ile and tRNA-Leu genes, positions that correspond to one nucleotide following the anticodon triplet. Though the intron sequences are variable among gene copies, phylogenetic analysis of the full-length tRNA-Tyr gene sequences suggests possible common origins of these genes between *Ae. aegypti* and *An. gambiae* (Figure 3). Moreover, in spite of differences in the intron sequences, the intron-containing tRNA genes show remarkably similar structure at the intron junctions in all copies of these genes in both species (Figure S2).

tRNA genes and codon usage bias

We investigated whether tRNA gene copy number was correlated with codon usage. We determined total counts of codons, amino acids and the relative synonymous codon usage (RSCU) for the annotated gene sets of both species (see experimental procedures and Table S5). Our genome-wide analysis of RSCU values for the predicted protein coding genes indicated that only a select set of codons were optimized in each genome. We identified subsets of optimal codons and rare codons in both mosquitoes based on the genome-specific frequencies of these codons (Table S6). We further determined that the cognate tRNAs to the optimal codons were relatively more abundant in each genome compared to cognate tRNAs to the rare codons. We found a significant association ($P < 0.05$) between codon abundance and cognate tRNA gene copy numbers (Figure S3) for both *Ae. aegypti* ($R^2 = 0.21$) and *An. gambiae* ($R^2 = 0.36$).

To test the effect of codon bias on gene expression, we first calculated the codon bias index using codonW (CBI; Bennetzen and Hall 1982) for each predicted protein coding gene in both genomes (the data can be provided upon request). We then compared the CBI values with the observed gene expression levels as determined from whole transcriptome microarray data for *Ae. aegypti* (GEO Accession # GSE16563) and *An. gambiae* (Cassone *et*

al., 2008). We observed a significant positive Spearman correlation between the CBI and gene expression levels in *Ae. aegypti* ($r = 0.64$, $P < 0.0001$) and *An. gambiae* ($r = 0.66$, $P < 0.0001$). For both mosquitoes, the highly expressed genes showed relatively higher CBI values than lowly expressed genes (Figure S4), thus supporting an expectation of higher codon bias in highly expressed genes.

Pathway analysis

To determine if genes in particular pathways showed an association between codon usage bias and genome abundance of tRNA genes, we compared the RSCU values of individual genes in the Koyota Encyclopedia of Genes and Genomes (KEGG) pathways with the corresponding RAIT values in both the *Ae. aegypti* and *An. gambiae* genomes. To declare a gene as showing a positive association between RSCU and RAIT, we arbitrarily selected the observed genome-wide Spearman correlation coefficient as the threshold value ($r = 0.331$ for *Ae. aegypti* and $r = 0.538$ for *An. gambiae*). Our analyses indicated that ~2–4% of the genes among individual pathways (113 of 5,227 genes in *Ae. aegypti* and 149 of 3665 genes in *An. gambiae*) showed a positive association between codon usage and tRNA gene abundance (Table 3). Of these, 97 genes in *Ae. aegypti* and 111 genes in *An. gambiae* (~74–85 % of the positive genes) were related to pathways involved with translation, energy metabolism and carbohydrate metabolism for both species. Further, genes in pathways showing an association tended to have greater within gene variation in codon bias than observed with genes in pathways where no association was evident. For example, many genes in the translational pathway showed an association between codon usage and tRNA gene numbers whereas no association was observed for genes in the mTOR pathway; codons within genes in the translational pathway showed greater variation in usage bias than those within the mTOR pathway (Figure S5). Within a pathway, some genes may reflect greater bias in codon usage than others; our analyses showed that genes within a pathway with more biased codon usage consistently showed an association between tRNA gene number and codon counts, unlike genes with generally unbiased codon usage (Table S7). More highly conserved genes within a family were generally more prone to show an association between codon usage and tRNA gene abundance; this phenomena was evident from our analysis of the highly conserved actin gene family that is essential for regulation of the actin cytoskeleton. A phylogenetic analysis of *Ae. aegypti* and *An. gambiae* actin genes showed that genes showing codon usage/tRNA gene number associations were better conserved than genes that lack an association (Figure 4). This suggests that specific genes within a gene family may be subject to selection for biased codon usage for translational optimization.

Discussion

This study investigated the comparative genome structure and molecular evolution of nuclear tRNA genes of *Ae. aegypti* and *An. gambiae* mosquitoes and explored the potential association of tRNA evolution with codon usage bias. Although the tRNA gene complements for both species were reported with the respective genome sequencing efforts (Holt *et al.* 2002; Nene *et al.* 2007), to date no comparative analyses of tRNA gene architecture or relationships with codon usage have been reported. The tRNAs function as adapter molecules where the anticodon loop of the tRNA binds to the complementary codon triplet in the mRNA to transfer specific amino acids to a growing polypeptide chain at the ribosome. This process depends on the availability of the individual amino acids and their cognate tRNAs, and also their ready availability at the appropriate time and tissue required for specific protein production in a cell. Of note, the *Ae. aegypti* genome contains more than 2X the number of tRNA genes than *An. gambiae* (906 vs. 441), although the roles that individual gene copy number, sequence variability and genome location play in their respective expression levels remain to be investigated. About 12% of the *Ae. aegypti* tRNA

gene sequences were determined to be pseudogenes, while only a single putative pseudogene was identified with *An. gambiae*. For only two amino acids (Ile and Val) were the tRNA gene copy numbers greater in *An. gambiae* compared to *Ae. aegypti*. Of note, however, despite differences in total tRNA gene numbers, comparisons of RAIT values indicated that the high abundance and low abundance isoacceptor tRNA genes for most amino acids were the same for both species, while only a small number showed high abundance in only one of the species. Further, although considerable sequence diversity was evident among the tRNA-Met gene copies, distinct initiator and elongator copies of these genes were not evident. That is, no tRNA-Met gene copies indicative of an initiator gene where the anticodon sequence CAU was preceded by C instead of U (Marck and Grosjean 2002) were identified in either species.

Genome locations for some tRNA genes were often non-random in both *Ae. aegypti* and *An. gambiae*. First, several interesting tRNA gene cluster orientations were evident, wherein genes resided in physical clusters comprised of 2 to 37 individual genes. These clusters often contained copies of the same tRNA gene, although in some cases clusters involved repeated groups of tRNA genes coding for 2–3 different amino acids in tandem or near-tandem linear arrays. Further, for several clusters the sizes of the intergenic regions were frequently the same or nearly the same, wherein the most extreme example was a 37 gene tRNA-Val cluster in *An. gambiae* with an ~162 bp intergenic region. Second, pairs of tRNA gene copies were most often located in the same orientation, although many examples wherein they were located on opposite strands were also evident. Finally, several examples of tRNA genes being embedded within protein coding gene sequences were identified in both species, with roughly equal representation of those that are in sense vs. anti-sense orientations. We do not know if the primary RNAs of these mRNA/tRNA gene pairs are transcribed by polymerase II (that produces mRNAs) or polymerase III (that produces tRNAs) or by cooperative involvement of both (Wang and Stumph, 1995; Raha *et al.*, 2010). However, in insects it is known that tRNA expression is dependent on both internal and external regulatory elements (Trivedi *et al.* 1999; Ouyang *et al.* 2000), so the impact of tRNA gene cluster orientation on individual gene expression remains unknown.

The observed extreme difference in tRNA-Ala gene copy number between *Ae. aegypti* and *An. gambiae* (n = 206 vs. 28, respectively) invokes speculation that this may reflect the over-representation of alanine in *Ae. aegypti* proteins and a corresponding need for increased expression of these tRNAs for efficient translation. However, comparison of total codons in the respective annotated gene sets indicated that the relative frequency of codons for alanine was quite similar in the two species and actually slightly higher in *An. gambiae* (0.07 in *Ae. aegypti* vs. 0.08 in *An. gambiae*, Tables S5, S8). Indeed, the observed relative frequencies of codons for all amino acids were remarkably similar between them.

Mature tRNAs universally reflect multiple modifications at several residues, and these modifications are known to influence nuclear transport, translation and tRNA stability (Hopper and Phizicky 2003). Thus, knowing the genomic sequences of individual tRNA genes is only an approximate guide to determine which tRNAs are ultimately optimized for specific codon(s). For example, purine nucleoside phosphorylase converts the wobble A in tRNA to inosine that can pair with A, C, U, thus disrupting the expected codon-anticodon relationship (Gustilo *et al.* 2008). Also, tRNAs may show variability in translation efficiency depending on the wobble cost of the codon (Xia 2008). Although comparative investigations on the specific base modifications in *Ae. aegypti* and *An. gambiae* tRNAs are outside the scope of the present study, some indirect evidence for base modifications within the anticodon is reflected in the identification of tRNA genes containing introns. It is well-established that introns serve as target sites to guide nuclear enzymes responsible for nucleoside modification. Here we observed the presence of introns in some Ile- and Leu-

tRNAs and all Tyr-tRNAs in both species; base modifications within the anticodon in these three tRNA classes have previously been shown to be dependent on presence of an intron (Grosjean *et al.* 1997). This suggests the potential for differential anticodon modification of Ile- and Leu-tRNAs (intron-containing vs. intron-less copies) in both species and that all Tyr-tRNAs in both species are likely subject to anticodon modification.

We investigated whether tRNA gene abundance in the *Ae. aegypti* and *An. gambiae* genomes was correlated with codon usage patterns, and subsequent gene expression levels. Our genome-wide analyses confirmed previous studies for both species involving relatively small numbers of genes (Argentine and James, 1993; Besansky, 1993; Morlais and Severson, 2003) that showed a broad continuum in RSCU values and the highest codon bias among putative highly expressed genes. This was expected as, for example, codon usage bias has been confirmed among drosophilids (Heger and Pointing, 2007) and a diversity of eukaryotes (dos Reis and Wernisch, 2009). As observed with some organisms including bacteria, yeast, *Caenorhabditis*, and *Drosophila* (Moriyama and Powell, 1997; Higgs and Ran, 2008) our results also demonstrated a significant association between codon usage and cognate tRNA gene copy numbers.

Codon preferences in a set of genes is thought to arise as a result of balance between forces of selection and mutation within finite populations following the selection-mutation-drift theory and predicts stronger selection for translation efficiency (Bulmer, 1991), that includes regulating translational accuracy of genes (Akashi, 1994; Rocha, 2004). In addition, selection among high vs. low expression genes has been shown to drive coevolution between codon usage and tRNA gene copy number (Higgs and Ran, 2008). The ribosomal protein genes (RPGs) are generally highly expressed and biased in codon usage and have proven useful in other organisms as reference genes for identifying codon bias (Bennetzen and Hall 1982, Sharp *et al.* 1986, Duret 2000, Rocha 2004, Behura *et al.* 2010). Our comparisons of CBI values for individual genes and whole transcriptome assays for *Ae. aegypti* (Behura *et al.* unpublished; GEO Accession #GSE16563) and *An. gambiae* (Cassone *et al.*, 2008) revealed moderate evidence for positive correlation between CBI values and expression levels (Figure S4). We observed a general trend where higher codon bias was associated with higher expression of some genes, although many genes were found to deviate from this trend, suggesting that selection for translational efficiency and expression may have driven codon bias and tRNA gene copy number in a subset of genes. However, we note too that our comparisons of CBI values with single transcriptome assays are unlikely to have captured the true magnitude of this relationship (Coghlan and Wolfe, 2000).

We mapped genes showing correlation coefficients between RSCU and RAIT that were above the genome-wide mean to individual KEGG pathways and determined that the majority of these genes were linked with pathways involved in translation, energy metabolism and carbohydrate metabolism for *Ae. aegypti* and *An. gambiae*. These pathways regulate the major cellular housekeeping functions and because of this role many of the associated genes are under strong selection for optimization of the translation process to maximize growth (Rocha, 2004). This phenomena is most evident among fast-growing organisms with corresponding high metabolic demands like bacteria (Andersson and Kurland 1990; Rocha, 2004). However, most mosquitoes live only 2 to 4 weeks in natural field conditions depending on the species and other factors such as temperature, humidity, food and predators, etc. Our results indicate that the translational process seems to be under selective pressure in order to respond to metabolic demands associated with rapid changes in biotic and abiotic conditions typically encountered within their relatively short life spans. Thus, translational efficiency of protein coding genes and the employment of tRNA abundance in this process may have evolved as a survival strategy in these mosquitoes.

Experimental Procedures

The tRNA genes used in this study were originally annotated in the genome sequencing projects of *An. gambiae* and *Ae. aegypti*, and were downloaded from Biomart via VectorBase (www.vectorbase.org). These were re-validated using the tRNAScan-SE program (Lowe and Eddy 1997) and also identified those that were pseudogenes and genes with intron sequences. The Relative Abundance of Isoacceptor tRNA (RAIT) was calculated as the ratio of observed number of tRNA genes to the expected number of tRNA genes of an isoacceptor family assuming no bias (putative pseudogenes were excluded). The expected number of tRNAs of an isoacceptor family was determined as the total number of tRNA genes that charge the same amino acid divided by the number of different anticodons these genes represent.

The genome distribution of tRNA genes was determined based on their start and end coordinates in the assembled genome sequences. For identification of tRNA gene clusters, we included only those tRNA genes that were located within 1 kb of each other. The relative orientations of tRNA genes in each cluster were compared based on the strand assignment of each gene. The flanking sequences (1 kb) up-stream and downstream of each tRNA gene were retrieved through Biomart and the homologies of flanking sequences with the respective tRNA gene were identified by BLASTN searches at GenBank as well as locally. Sequence analyses of tRNA genes between *Ae. aegypti* and *An. gambiae* were performed using the DnaSP program (Rozas *et al.* 2003). The average nucleotide differences were used to compare each tRNA isotype between the two species. Phylogenetic trees were constructed by the neighbor joining method using MEGA 3.0 (Kumar *et al.* 2004) and bootstrap tests of the inferred phylogenies were performed with 1000 replications. ClustalW incorporated in MEGA 3.0 was used, with the default parameters, to perform multiple alignments of DNA sequences. The EST sequences of *Ae. aegypti* and *An. gambiae* were downloaded from VectorBase and subjected to a local BLASTN search using the mosquito tRNA gene sequences as the query. The complete tRNA gene sequences within significant EST hits (with e-value < 0.0001 as a cutoff) were identified using the tRNAScan-SE program.

The coding sequences for all predicted protein coding genes in the *Ae. aegypti* (Aaeg1.1) and *An. gambiae* (AgamP3.4) genomes were downloaded from Biomart via VectorBase (<http://www.vectorbase.org>). The sequences were used to calculate the total numbers of amino acids and codons, as well as relative synonymous codon usage (RSCU) using the Anaconda software (Pinheiro *et al.* 2006). Anaconda program parameters were set to exclude genes where the coding sequences had no universal start codon or no universal stop codon(s) or had stop codons within the predicted coding sequences. Also, genes were excluded where predicted coding sequences were not multiples of three or had undetermined nucleotides (N). The Anaconda program was also used to identify subsets of rare or optimal codons among the predicted protein coding genes for both mosquitoes. The codon bias index (CBI) for each gene was calculated using the codonW program (<http://codonw.sourceforge.net/>). CBI measures the extent to which a gene uses a subset of optimal codons. Ribosomal protein genes (RPGs) were used as reference genes for relative evaluations of codon bias in predicted protein coding gene sets for both species, as RPGs have often used as reference genes for studying codon usage bias in other organisms (Bennetzen and Hall 1982, Sharp *et al.* 1986, Duret 2000, Rocha 2004, Behura *et al.* 2010). The RPGs of *An. gambiae* were downloaded from the Ribosomal Protein Gene Database (<http://ribosome.med.miyazaki-u.ac.jp/>) and the corresponding RPGs of *Ae. aegypti* were manually identified and downloaded at VectorBase. We determined Spearman correlations between CBI values and the median normalized expression values of individual genes with *R* using the 'stats' package. The expression data were derived from microarray experiments

with *Ae. aegypti* (Behura *et al.* unpublished; GEO Accession #GSE16563) and *An. gambiae* (Cassone *et al.*, 2008).

To identify pathways enriched with genes showing associations between codon usage and tRNA gene abundance, we analyzed a total 5527 genes in *Ae. aegypti* and 3665 genes in *An. gambiae* that have been assigned to different pathways by Koyota Encyclopedia of Genes and Genomes (<http://www.genome.ad.jp/kegg/-October-2008> KEGG, <http://www.genome.ad.jp/kegg/>, October 3, 2008 release). The RSCU values (which are variable for each gene) were compared with the RAIT values (which are constant within a genome) in both species. First, we determined genome-wide Spearman correlation between RSCU values of codons (stop codons were excluded) of all genes in each genome with RAIT values. Then, we determined Spearman correlation between RSCU and RAIT values for each gene. If the correlation coefficient obtained for an individual gene was greater than the genome-wide correlation coefficient, we considered that gene as showing a positive association between RSCU and RAIT. Finally, we mapped these genes to the respective KEGG pathway.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grant RO1-AI059342 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA.

Literature cited

- Akashi H. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* 1995;139:1067–1076. [PubMed: 7713409]
- Andersson SG, Kurland CG. Codon preferences in free-living microorganisms. *Microbiol Rev* 1990;54:198–210. [PubMed: 2194095]
- Argentine JA, James AA. Codon preference of *Aedes aegypti* and *Aedes albopictus*. *Insect Mol Biol* 1993;1:189–194. [PubMed: 8269097]
- Behura SK, Stanke M, Desjardins CA, Werren JH, Severson DW. Comparative analysis of nuclear tRNA genes of *Nasonia vitripennis* and other arthropods, and relationships to codon usage bias. *Insect Mol Biol* 2010;19(Suppl 1):49–58. [PubMed: 20167017]
- Bennetzen JL, Hall BD. Codon selection in yeast. *J Biol Chem* 1982;257:3026–3031. [PubMed: 7037777]
- Besansky NJ. Codon usage patterns in chromosomal and retrotransposon genes of the mosquito *Anopheles gambiae*. *Insect Mol Biol* 1993;1:171–178. [PubMed: 8269095]
- Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* 1991;129:897–907. [PubMed: 1752426]
- Cassone BJ, Mouline K, Hahn MW, White BJ, Pombi M, Simard F, Costantini C, Besansky NJ. Differential gene expression in incipient species of *Anopheles gambiae*. *Mol Ecol* 2008;17:2491–2504. [PubMed: 18430144]
- Cedergren RJ, LaRue B, Sankoff D, Lapalme G, Grosjean H. Convergence and minimal mutation criteria for evaluating early events in tRNA evolution. *Proc Natl Acad Sci USA* 1980;77:2791–2795. [PubMed: 6930666]
- Coghlan A, Wolfe KH. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast* 2000;16:1131–1145. [PubMed: 10953085]
- Cognat V, Deragon JM, Vinogradova E, Salinas T, Remacle C, Maréchal-Drouard L. On the evolution and expression of *Chlamydomonas reinhardtii* nucleus-encoded transfer RNA genes. *Genetics* 2008;179:113–123. [PubMed: 18493044]

- dos Reis M, Wernish L. Estimating translational selection in eukaryotic genomes. *Mol Biol Evol* 2009;26:451–461. [PubMed: 19033257]
- Duret L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet* 2000;16:287–289. [PubMed: 10858656]
- Fujishima K, Sugahara J, Tomita M, Kanai A. evidence in the archaeal genomes that tRNAs emerged through the combination of ancestral genes as 5' and 3' tRNA halves. *PLoS ONE* 2008;3:e1622. [PubMed: 18286179]
- Goodenbour JM, Pan T. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res* 2006;34:6137–6146. [PubMed: 17088292]
- Grosjean H, Szweykowska-Kulinska Z, Motorin Y, Fasiolo F, Simos G. Intron-dependent enzymatic formation of modified nucleosides in eukaryotic tRNAs: a review. *Biochimie* 1997;79:293–302. [PubMed: 9258438]
- Gustilo EM, Vendeix FA, Agris PF. tRNA's modifications bring order to gene expression. *Curr Opin Microbiol* 2008;11:134–140. [PubMed: 18378185]
- Heger A, Ponting CP. Variable strength of translational selection among 12 *Drosophila* species. *Genetics* 2007;177:1337–348. [PubMed: 18039870]
- Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet* 2008;42:287–299. [PubMed: 18983258]
- Higgs PG, Ran W. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Mol Biol Evol* 2008;25:2279–2291. [PubMed: 18687657]
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 2002;298:129–149. [PubMed: 12364791]
- Hopper AK, Phizicky EM. tRNA transfers to the limelight. *Genes Dev* 2003;17:162–180. [PubMed: 12533506]
- Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985;2:13–34. [PubMed: 3916708]
- Kanaya S, Yamada Y, Kudo Y, Ikemura T. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 1999;238:143–155. [PubMed: 10570992]
- Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol* 2001;53:290–298. [PubMed: 11675589]
- Kano A, Andachi Y, Ohama T, Osawa S. Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic G + C content. Correlation with codon usage. *J Mol Biol* 1991;221:387–401. [PubMed: 1717697]
- Kubli E. The structure and function of tRNA genes of higher eukaryotes. *Experientia* 1981;37:1–9. [PubMed: 6781925]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 2004;5:150–163. [PubMed: 15260895]
- Lehmann J, Libchaber A. Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. *RNA* 2008;14:1264–1269.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997;25:955–964. [PubMed: 9023104]
- Marck C, Grosjean H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* 2002;8:1189–1232. [PubMed: 12403461]
- Moriyama EN, Powell JR. Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol* 1997;45:514–523. [PubMed: 9342399]
- Morlais I, Severson DW. Intraspecific DNA variation in nuclear genes of the mosquito *Aedes aegypti*. *Insect Mol Biol* 2003;12:631–639. [PubMed: 14986924]

- Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 2007;316:1718–1723. [PubMed: 17510324]
- Ouyang C, Martinez MJ, Young LS, Sprague KU. TATA-binding protein – TATA interaction is a key determinant of differential transcription of silkworm constitutive and silk gland-specific tRNA^{Ala} genes. *Mol Cell Biol* 2000;20:1329–1343. [PubMed: 10648619]
- Percudani R, Pavesi A, Ottonello S. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J Mol Biol* 1997;268:322–330. [PubMed: 9159473]
- Pinheiro M, Afreixo V, Moura G, Freitas A, Santos MA, Oliveira JL. Statistical, computational and visualization methodologies to unveil gene primary structure features. *Methods Inf Med* 2006;45:163–168. [PubMed: 16538282]
- Raha D, Wang Z, Moqtaderi Z, Wu L, Zhong G, Gerstein M, Struhl K, Snyder M. Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc Natl Acad Sci USA* 2010;107:3639–3644. [PubMed: 20139302]
- 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]
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003;19:2496–2497. [PubMed: 14668244]
- Saks ME, Sampson JR. Evolution of tRNA recognition systems and tRNA gene sequences. *J Mol Evol* 1995;40:509–518. [PubMed: 7540216]
- Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucl Acids Res* 1986;14:5125–5143. [PubMed: 3526280]
- Sun FJ, Caetano-Anollés G. Transfer RNA and the origins of diversified. *Life Sci Prog* 2008;91:265–284.
- Trivedi A, Young LS, Ouyang C, Johnson DL, Spragu KU. A TATA element is required for tRNA promoter activity and confers TATA-binding protein responsiveness in *Drosophila* Schneider-2 cells. *J Biol Chem* 1999;274:11369–11375. [PubMed: 10196229]
- Varenne S, Buc J, Lloubes R, Lazdunski C. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* 1984;180:549–576. [PubMed: 6084718]
- Wang Y, Stumph WE. RNA polymerase II/III transcription specificity determined by TATA box orientation. *Proc Natl Acad Sci U S A* 1995;92:8606–8610. [PubMed: 7567983]
- Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 2007;316:1738–1743. [PubMed: 17588928]
- Waterhouse RM, Wyder S, Zdobnov EM. The *Aedes aegypti* genome: a comparative perspective. *Insect Mol Biol* 2008;17:1–8. [PubMed: 18237279]
- Xia X. The cost of wobble translation in fungal mitochondrial genomes: integration of two traditional hypotheses. *BMC Evol Biol* 2008;8:211. [PubMed: 18638409]
- Yamao F, Andachi Y, Muto A, Ikemura T, Osawa S. Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins. *Nucl Acids Res* 1991;19:6119–6122. [PubMed: 1956771]

Figure 1. Cluster patterns of tRNA genes in both genomes. **A)** Val-tRNAs in *An. gambiae*, **B)** Pro-tRNA genes in *An. gambiae*. **C)** Met-tRNA genes in *Ae. aegypti*. **D)** Tyr-tRNA genes in *An. gambiae*. **E)** Tyr-tRNA genes in *Ae. aegypti*. The chromosome/supercontig location, along with start and end coordinates, of each cluster is indicated. The numbers on the right side of each cluster indicates the distance between neighboring tRNA genes.

Figure 2.

Sequence variation within tRNA genes. The tRNA genes of both species show varying levels of average nucleotide differences. Note that tRNA genes analyzed here are all intron-less copies. The tRNA-Tyr genes are not included in this analysis as each of these genes contains an intron in both species.

Figure 3. Phylogenetic (neighbor-joining tree) relationships of Tyr-tRNA genes of *Ae. aegypti* (shown as gene IDs starting AAELs) and *An. gambiae* (shown as gene IDs starting AGAPs). The scale is shown by the vertical line.

Figure 4. Neighbor-joining phylogenetic tree of actin genes show distinct phylogenetic grouping of genes that predominantly show RAIT-dependent codon usage (in bold branching). The genes that show RAIT-dependent codon usage are labeled RAIT after the gene ID.

Table 1

Number of tRNA gene copies in each genome

tRNA Gene	<i>Ae. aegypti</i>	<i>An. gambiae</i>
Ala	206	28
Arg	40	23
Asn	21	12
Asp	48	26
Cys	12	5
Gln	24	15
Glu	44	27
Gly	44	25
His	34	23
Ile	4	15
Leu	49	26
Lys	52	25
Met	34	19
Phe	17	9
Pro	36	32
Ser	32	22
Thr	18	15
Trp	13	7
Tyr	26	22
Val	41	64
Pseudogenes	111	1
Total	906	441

Table 2

Relative abundance of isoacceptor tRNAs (RAIT) classified in four groups based on relative abundance in each genome

Anticodon	<i>Ae. aegypti</i>	<i>An. gambiae</i>
<u>Abundant isoacceptors in both <i>Ae. aegypti</i> and <i>An. gambiae</i></u>		
Ala-AGC	3.7	1.8
Arg-TCG	1.1	1.4
Arg-ACG	2.2	1.9
Gln-CTG	1.4	1.5
Gly-GCC	2.0	1.3
Ile-AAT	1.5	1.7
Leu-CAG	2.0	2.3
Lys-CTT	1.5	1.4
Ser-AGA	1.8	1.1
Thr-AGT	1.9	1.5
Val-AAC	3.5	1.5
<u>Abundant isoacceptors in <i>Ae. aegypti</i> but not in <i>An. gambiae</i></u>		
Glu-TTC	1.0	0.8
Pro-TGG	1.5	0.8
<u>Abundant isoacceptors in <i>An. gambiae</i> but not in <i>Ae. aegypti</i></u>		
Glu-CTC	0.98	1.19
Pro-CGG	0.57	1.41
Ser-CGA	0.75	1.45
Thr-CGT	0.55	1.13
Val-CAC	0.18	1.36
<u>Low-abundance isoacceptors in both <i>Ae. aegypti</i> and <i>An. gambiae</i></u>		
Ala-CGC	0.08	0.86
Ala-TGC	0.16	0.32
Arg-TCT	0.30	0.35
Arg-CCT	0.40	0.35
Gln-TTG	0.58	0.53
Gly-TCC	0.91	0.72
Ile-TAT	0.50	0.27
Leu-TAA	0.57	0.38
Leu-TAG	0.57	0.77
Leu-CAA	0.68	0.58
Leu-AAG	0.79	0.96
Lys-TTT	0.45	0.64
Pro-AGG	0.97	0.75
Ser-TGA	0.50	0.36
Thr-TGT	0.55	0.38
Val-TAC	0.27	0.14

Table 3

Number of genes of *Aedes aegypti* (Aaeg) and *Anopheles gambiae* (Agam) of specific KEGG pathways that show correlation between RSCU and RAIT

Pathway	n_Aaeg	N_Aaeg	n_Agam	N_Agam	Functionality
Cysteine metabolism	1	11	0	13	Amino Acid Metabolism
Lysine degradation	0	44	4	40	Amino Acid Metabolism
Valine, leucine and isoleucine biosynthesis	1	14	0	11	Amino Acid Metabolism
Valine, leucine and isoleucine degradation	3	38	7	31	Amino Acid Metabolism
Regulation of actin cytoskeleton	7	40	9	42	Cell Motility
Oxidative phosphorylation	22	112	32	100	Energy Metabolism
Methane metabolism	3	21	3	18	Energy Metabolism
Ubiquitin mediated proteolysis	0	86	4	78	Folding/Sorting/Degradation
Peptidoglycan biosynthesis	3	9	5	9	Glycan Biosynthesis
Glycerolipid metabolism	1	58	0	44	Lipid Metabolism
Glutathione metabolism	0	58	4	46	Other Amino Acids syn
Wnt signaling pathway	0	59	4	55	Signal Transduction
Ribosome/Translation	43	94	47	86	Translation
Citrate cycle (TCA cycle)	7	29	12	25	Carbohydrate Metabolism
Glycolysis/Gluconeogenesis	8	40	8	35	Carbohydrate Metabolism
Glyoxylate and dicarboxylate metabolism	5	21	4	15	Carbohydrate Metabolism
Pentose and glucuronate interconversions	1	26	0	23	Carbohydrate Metabolism
Pyruvate metabolism	8	40	5	31	Carbohydrate Metabolism
n	113	-	148	-	
n*	97		111		
N	-	5227	-	3665	
N*	-	800	-	702	

n_Aaeg: Number of positive genes of *Aedes aegypti* where a correspondence was observed between RSCU and RAIT

N_Aaeg: Total number genes of *Aedes aegypti* mapped to the specific pathway

n_Agam: Number of positive genes of *Anopheles gambiae* where a correspondence was observed between RSCU and RAIT

N_Agam: Total number genes of *Anopheles gambiae* mapped to the specific pathway

n = Total number of positive genes showing correspondence between RSCU and RAIT

n* = Number of positive genes related to translation, energy and carbohydrate metabolism

N = Total number of genes analyzed. Note that the pathways are not shown where all of its genes showed absence of correspondence between RSCU and RAIT

N* = Total number of genes in individual pathways;

- = not applicable