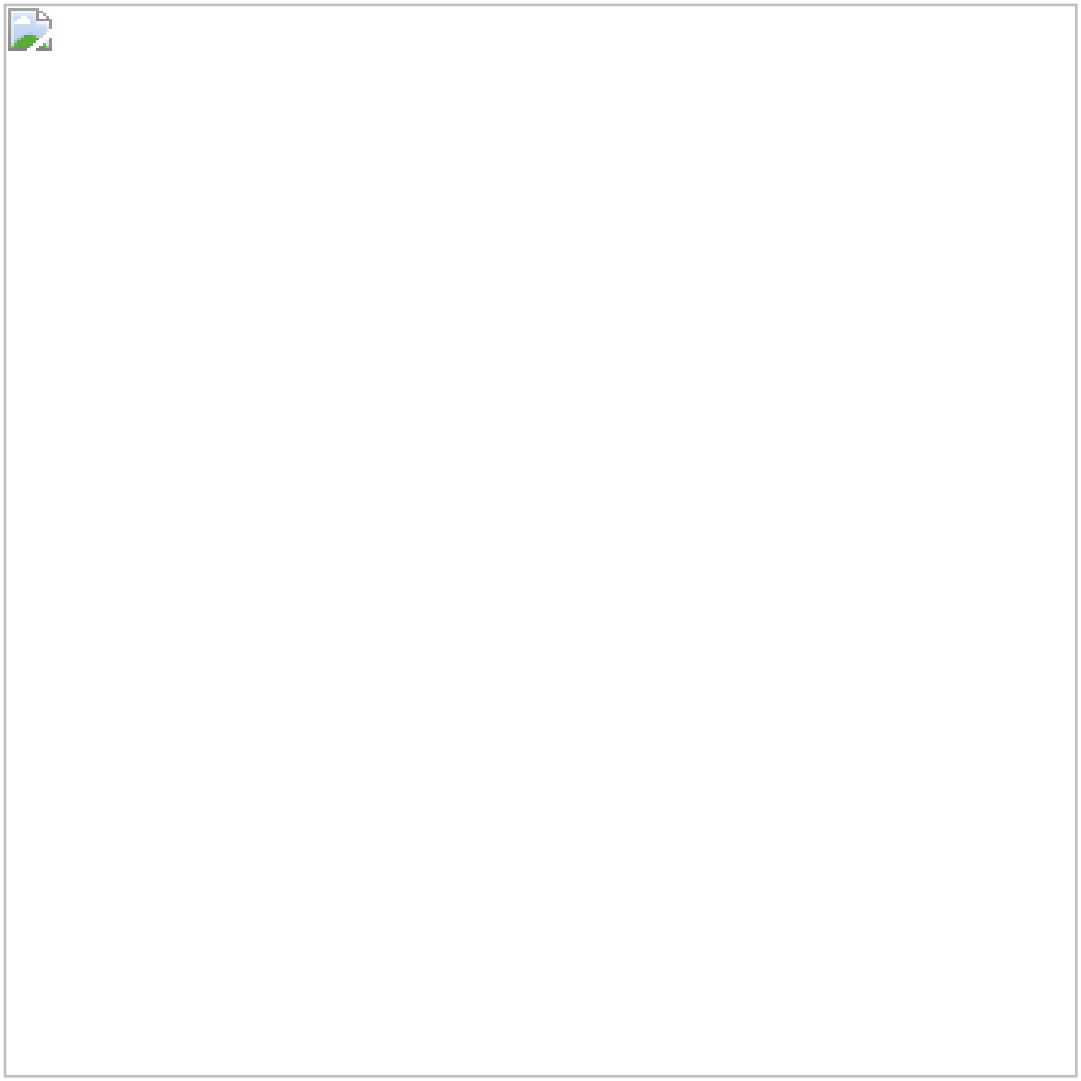


Mosquito Vitellogenin Genes: Comparative Sequence Analysis, Gene Duplication, and the Role of Rare Synonymous Codon Usage in Regulating Expression

Authors: Isoe, Jun, and Hagedorn, Henry H.

Source: Journal of Insect Science, 7(1) : 1-49

Published By: Entomological Society of America

URL: https://doi.org/10.1673/031.007.0101

Mosquito vitellogenin genes: Comparative sequence analysis, gene duplication, and the role of rare synonymous codon usage in regulating expression

Jun Isoe[1] and Henry H. Hagedorn

Department of Entomology and Center for Insect Science, University of Arizona, Tucson, AZ. 85721 USA
[1]Present address, Department of Biochemistry and Molecular Biophysics, University of Arizona. Tucson, AZ 85721 USA

Abstract

Comparative sequence analysis of mosquito vitellogenin (Vg) genes was carried out to gain a better understanding of their evolution. The genomic clones of vitellogenin genes were isolated and sequenced from all three subfamilies of the family Culicidae including Culicinae (*Aedes aegypti*, *Ochlerotatus atropalpus*, *Ae. polynesiensis*, *Ae. albopictus*, *Ochlerotatus triseriatus* and *Culex quinquefasciatus*), Toxorhynchitinae (*Toxorhynchites amboinensis*), and Anophelinae (*Anopheles albimanus*). Genomic clones of vitellogenin genes Vg-B and Vg-C were isolated from *Ae. aegypti* and sequenced. A comparison of Vg-B and Vg-C, with the previously characterized vitellogenin gene, Vg-A1, suggests that Vg-A1 and Vg-B probably arose by a recent gene duplication, and Vg-C apparently diverged from the two other members of the gene family in an earlier gene duplication event. Two vitellogenin genes orthologous to Vg-C were cloned from a *Cx. quinquefasciatus* DNA library, one of which is truncated at the N-terminal end. Single vitellogenin genes, orthologous to Vg-C, were cloned from the *An. albimanus* and *Tx. amboinensis* libraries. Incomplete sequences orthologous to Vg-B and Vg-C were isolated from the *Oc. atropalpus* library. Only partial sequences were isolated from *Ae. polynesiensis, Ae. albopictus* and *Oc. triseriatus*. Inferred phylogenetic relationships based on analysis of these sequences suggest that Vg-C was the ancestral gene and that a recent gene duplication gave rise to Vg-A1 and Vg-B after the separation of the genus *Aedes*.

The deduced amino acid composition of mosquito vitellogenin proteins exhibits higher tyrosine and phenylalanine composition than other mosquito proteins except for the hexamerin storage proteins. Analysis of vitellogenin coding sequences showed that a majority of amino acid substitutions were due to conserved and moderately conserved changes suggesting that the vitellogenins are under moderately selective constrains to maintain tertiary structure. The vitellogenin genes of the three anautogenous mosquitoes, that require a blood meal to develop eggs, had very high synonymous codon usage biases similar to highly expressed genes of other organisms. On the other hand, the vitellogenin genes of autogenous mosquitoes, that develop at least one batch of eggs without a blood meal, exhibited low synonymous codon usage bias. An unusual pattern of synonymous codon usage was observed in the first 15 amino acid residues encoding the signal peptide in the vitellogenin genes, where a high number of rarely used synonymous codons are present. It is hypothesized that rare synonymous codons have selectively accumulated in the signal peptide region to down-regulate the rate of translation initiation in the absence of a blood meal. Real-time PCR gene expression experiments showed that all three *Ae. aegypti* vitellogenin genes were highly expressed after a blood meal, and expressed in non-blood-fed females, males, larvae and pupae at trace levels. Sequences were deposited in GenBank (accession

numbers: *Ae. aegypti* Vg-B, AY380797, Vg-C, AY373377; *Oc. atropalpus* Vg-B, AY691321, Vg-C, AY691322; *Ae. polynesiensis* Vg-A1, AY691318, Vg-B, AY691319, Vg-C, AY691320; *Ae. albopictus Vg-A1,* AY691316, Vg-C, AY691317; *Oc. triseriatus* Vg-C, AY691323; *Cx. quinquefasciatus* Vg-C1, AY691324, Vg-C2, AY691325; *Tx. amboinensis* Vg-C, AY691326*; An. albimanus* Vg-C, AY691327).

Keywords: nucleotide sequence analysis, gene duplication, amino acid composition, nonsynonymous and synonymous amino acid substitution, autogeny, anautogeny, *Aedes aegypti*, *Ochlerotatus atropalpus*, *Aedes polynesiensis*, *Aedes albopictus*, *Ochlerotatus triseriatus*, *Culex quinquefasciatus*, *Toxorhynchites amboinensis*, *Anopheles albimanus*
Abbreviations: GC3: G+C content 3rd position of the synonymous codon, ENC: effective number of codons, Vg: vitellogenin, Fat body preparations: abdominal body wall with fat body
Correspondence: jisoe@ag.arizona.edu
Received: 25 October 2003 | Accepted: 30 June 2005 | Published: 12 January 2007

ISSN: 1536-2442 | Volume 7, Number 1

Cite this paper as:
Isoe J, Hagedorn HH. 2007. Mosquito vitellogenin genes: Comparative sequence analysis, gene duplication, and the role of rare synonymous codon usage in regulating expression. 49pp. *Journal of Insect Science* 7:01, available online: insectscience.org/7.01

Introduction

Vitellogenins are precursors of the yolk proteins, which are phosphoglycolipoproteins utilized in oviparous animals to provide nutrition for the developing embryo. Vitellogenins of many animals have multiple vitellogenin genes including the frog, *Xenopus laevis* (Germond et al. 1984), the chicken, *Gallus gallus*, the nematode, *Caenorhabditis elegans* (Spieth et al. 1991), and several insects including *Drosophila melanogaster* (Wahli 1988), and *Aedes aegypti* (Hamblin et al. 1987). The gene duplication events that gave rise to these gene families are not well understood. Duplicated genes could evolve into pseudogenes or into new genes with novel functions as a consequence of relaxation of functional constraints (Shimeld 1999; Krakauer and Nowak 1999). Pseudogenes encoding non-functional vitellogenin have been isolated from *G. gallus* (Silva et al. 1989), *C. elegans* (Spieth et al. 1985), and the shrimp, *Metapenaeus ensis* (Tsang et al. 2003). The entire open-reading frames of vitellogenin gene sequences in insects have been documented in Dictyoptera (*Blatella germanica*, Martin et al. 1998; *Leucophaea maderae,* Tufail and Takeda 2002; *Periplaneta americana*, Tufail et al. 2000), Hemiptera *(Plautia stali,*Lee et al. 2000; *Riptortus clavatus,* Hirai 1998), Lepidoptera (*Bombyx mori*, Yano et al. 1994; *Lymantria dispar*, Adamczyk et al. 1996), Coleoptera (*Anthonomus grandis,* Trewitt et al. 1992), Hymenoptera (*Apis mellifera,* Piulachs et al. 2003; *Athalia rosae*, Kageyama et al. 1994; *Pimpla nipponica* Nose et al. 1997), and Diptera, *Ae. aegypti* Chen et al. 1994; Romans et al. 1995; *Anopheles gambiae* Holt et al. 2002).

Comparative sequence analyses of vitellogenin genes from taxonomically diverse organisms have suggested that vertebrate and invertebrate vitellogenins share a common ancestry for several reasons. First, the positions of some of the cysteines in the primary structure of vitellogenins are highly conserved, especially at the C-terminal region (Trewitt et al. 1992; Hagedorn et al. 1998; Sappington and Raikhel 1998). The conserved positions of cysteines are likely involved in the formation of a complex tertiary structure of vitellogenins. The tertiary structure of Lamprey vitellogenin has been determined by X-ray crystallography (Raag et al. 1988; Thompson and Banaszak 2002). Second, the position of some introns in vertebrate and invertebrate vitellogenin nucleotide sequences are well conserved (Nardelli et al. 1987; Trewitt et al. 1992; Hagedorn et al. 1998). Third, although sequence analysis has revealed extensive sequence divergence between vitellogenins of all organisms, the existence of some limited conserved regions suggest that they are related (Wahli 1988; Byrne 1989, Hagedorn et al. 1998; Sappington and Raikhel 1998). However, many amino acid substitutions were found to be conserved changes in the physical and chemical properties of the amino acids, implying that the overall tertiary structure of vertebrate and invertebrate vitellogenin genes are highly constrained and conserved (Trewitt et al. 1992; Hagedorn et al. 1998). Fourth, polyserine regions are present in many vertebrate and invertebrate organisms (Romans et al. 1995; Hagedorn et al. 1998; Sappington and Raikhel 1998). Taken together, this evidence suggests that vertebrate and invertebrate vitellogenins share a common ancestry.

The regulation of expression of highly expressed proteins sometimes involves the use of specific synonymous codons. When numerous gene sequences in *E. coli* and *Saccharomyces cerevisiae* were analyzed, a pronounced bias in patterns of synonymous codon usage was observed, especially in highly expressed genes where many amino acid residues are selectively coded by a single preferential synonymous codon (Kurland 1991; Ikemura 1985). The corresponding isoaccepting tRNAs of preferentially used synonymous codons in highly expressed genes were also found to be very abundant compared to rarely used synonymous codons. The abundance of each isoaccepting tRNA differs significantly in *E. coli* (Gouy and Gautier 1982; Ikemura 1985), and *D. melanogaster* (Moriyama and Powell 1997; Powell and Moriyama 1997). The abundance of tRNA isoaccepting species can be different in different tissues in eukaryotes and can be regulated differentially in response to developmental and physiological processes in bacteria (Emilsson and Kurland 1990), in plants (Sreenagesh and Jayabaskaran 1996), in arthropods (Chevallier and Garel 1979; Candelas et al 1990), and in humans (Kanduc et al. 1997). Thus, as a general rule in both prokaryotes and eukaryotes, highly expressed proteins tend to be selected for preferential usage of optimal synonymous codons for translational efficiency and/or accuracy.

In this study, the entire coding region of the vitellogenin genes from three anautogenous

mosquitoes, *Ae. aegypti, Culex quinquefasciatus* and *Anopheles albimanus*, and two autogenous mosquitoes, *Ochlerotatus atropalpus* and *Toxyrhynchites amboinensis*, and several partial vitellogenin coding sequences from three other anautogenous mosquitoes (*Aedes polynesiensis*, *Aedes albopictus*, and *Ochlerotatus triseriatus*) were cloned and sequenced to gain a better understanding of 1) vitellogenin gene duplication events within the mosquito lineage, 2) amino acid composition, and 3) the degree of nonsynonymous and synonymous amino acid substitutions.

Table 1. Classification and reproductive strategies of mosquitoes

Species	Reproduction	Vg	L*	References
Family Culicidae				
Subfamily Culicinae				
Tribe Aedeomyiini				
Genus *Aedes*				
Subgenus *Stegomyia*				
Aedes aegypti	Anautogenous	Vg-A1	2148	Romans et al. (1995)
		Vg-B	2142	This study
		Vg-C	2087	This study
Aedes polynesiensis	Anautogenous	Vg-A1	79	This study
		Vg-B	785	This study
		Vg-C	735	This study
Aedes albopictus	Anautogenous	Vg-A1	130	This study
		Vg-C	157	This study
Subgenus *Ochlerotatus*				
Ochlerotatus atropalpus	Autogenous	Vg-B	2108	This study
		Vg-C	1825	This study
Subgenus *Protomacleaya*				
Ochlerotatus triseriatus	Anautogenous	Vg-C	60	This study
Tribe Culicini				
Genus *Culex*				
Subgenus *Culex*				
Cules quinquefasciatus	Anautogenous	Vg-C1	2111	This study
		Vg-C2	2043	This study
Subfamily Toxorhynchitinae				
Genus *Toxorhynchites*				
Subgenus *Toxorhynchites*				
Toxorhynchites amboinensis	Autogenous	Vg-C	2032	This study
Subfamily Anophelinae				
Genus *Anopheles*				
Subgenus *Nyssorhynchus*				
Anopheles albimanus	Anautogenous	Vg-C	2029	This study

*L indicates the number of amino acids sequenced

Materials and Methods

Mosquitoes used in this study were selected based on taxonomic relationships representing all three mosquito subfamilies with different reproductive strategies (i.e. autogenous and anautogenous reproduction). The taxonomic classification and reproductive strategies of the mosquitoes studied are shown in Table 1. Reinhart et al. (2004) have described a revision of the tribe Aedini and renamed some members of the genus *Aedes; Ae. aegypti* to *Stegomyia aegypti,* and *Ae. albopictus* to *St. albopictus*. This change has been controversial (Savage 2005) (see also http://wrbu.si.edu/forums/viewtopic.php?t=9). We have retained the earlier classification.

Genomic DNA library construction

Genomic DNA was isolated from one anautogenous mosquito species, *Cx. quinquefasciatus*, kindly provided by J. Chaney at the University of California, Riverside and two autogenous mosquito species, *Oc. atropalpus* and *Tx. amboinensis* that were kindly provided by D. Wheeler at the University of Arizona, Tucson, Arizona and F. Mahmood at Rutgers University, New Brunswick, New Jersey, respectively. Fresh or alcohol preserved mosquitoes were used to isolate genomic DNA (Sambrook et al. 1989). The high molecular weight genomic DNA was subjected to partial digestion with *Sau*3AI. The partially digested genomic DNA was separated either on a spharose gradient or low melting point agarose gels, from which bands between 6.0 and 10.0 kb were excised and purified using GELase (Epicentre Technologies, http://www.epicentre.com/main.asp). The purified genomic DNA was ligated using a *Bam*HI predigested λ ZapExpress vector and packaged using Gigapack III Gold Packaging Extract from Stratagene (http://www.stratagene.com/). All three unamplified libraries had ~5 X 106 plaque forming units. Genomic DNA libraries from *Ae. aegypti* constructed in Lambda Dash® II vector, from *An. albimanus* constructed in Lambda EMBL 3 vector, and from *Ae. polynesiensis*, *Ae. albopictus*, and *Oc. triseriatus* constructed in a λ ZapExpress vector were provided by A. A. James (University of California, Irvine), M. Kidwell (University of Arizona, Tucson, Arizona), and R. Nussenzveig (University of Arizona, Tucson, Arizona), respectively.

Screening genomic DNA libraries

The *Ae. aegypti* genomic DNA library was probed using DNA probes containing conserved vitellogenin coding sequences (Hagedorn et al. 1998): Vg-A1 nucleotides 2425-4338, GenBank accession # L41842 (Romans et al. 1995), labeled with digoxigenin-11-dUTP using the random priming method (Roche Molecular Biochemicals, www.roche.com). Approximately 50,000–300,000 plaques were screened. *Ae. aegypti* Vg-B and Vg-C DNA probes were chosen to contain the conserved region of vitellogenin genes described above that were then used to probe the *Oc. atropalpus*, *Cx. quinquefaciatus*, *Tx. amboinensis*, and *An. albimanus* DNA libraries. Approximately 50,000–300,000

plaques were screened. Approximately 50,000–200,000 plaques were screened for the *Ae. polynesiensis*, *Ae. albopictus*, and *Oc. triseriatus* libraries. Genomic DNA libraries were constructed in a Lambda Dash® II vector (Stratagene, www.stratagene.com) kindly provided by A. A. James (University of California, Irvine). The plaques were transferred to nylon membranes (Micron Separations Inc., www.stratagene.com). The membranes were denatured, neutralized and DNA cross-linked using a Strata-linker (Strategene). Hybridization was performed overnight at 55oC using a Gene Roller (Savant Instruments Inc, www.combichemlab.com). Positive plaques were detected colormetrically using anti-digoxygenin conjugated with alkaline phosphatase, with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate as substrates. All positive plaques were further screened to purify each positive plaque. Phage DNAs for each putative vitellogenin gene cloned were extracted and purified using methods described by Sambrook et al. (1989). Phages were precipitated with PEG 8000 and NaCl overnight instead of for one hour. Phage DNAs were subsequently digested with restriction enzymes and electrophoresed in 1.0% agarose gels in TAE buffer. DNA fragments encoding vitellogenin genes were subcloned into the pBluescript SK+ plasmid vector (Strategene), and plasmid DNAs were isolated by plasmid DNA Miniprep (Promega, www.promega.com/). Transformation was carried out using DH5α competent cells (Life Technologies Inc., www.lifetech.com). All sequencing was performed at the University of Arizona Sequencing Facility using an automatic sequencer (Model 373, Applied Biosystems, appliedbiosystems.com). Sequences were determined from both strands. Sequences were deposited in GenBank (accession numbers: *Ae. aegypti* Vg-B, AY380797, Vg-C, AY373377; *Oc. atropalpus* Vg-B, AY691321, Vg-C, AY691322; *Ae. polynesiensis* Vg-A1, AY691318, Vg-B, AY691319, Vg-C, AY691320; *Ae. albopictus Vg-A1,* AY691316, Vg-C, AY691317; *Oc. triseriatus* Vg-C, AY691323; *Cx. quinquefasciatus* Vg-C1, AY691324, Vg-C2, AY691325; *Tx. amboinensis* Vg-C, AY691326*; An. albimanus* Vg-C, AY691327).

Determination of intron and 3' untranslated region sequences

Positions of introns for each vitellogenin gene were determined by comparison of genomic and cDNA sequences. Fat body preparations (which include the abdominal epidermis, heart, nerve cord and associated tissues) were dissected from female mosquitoes 24 hours after a blood meal and were frozen in liquid nitrogen. Total RNA was isolated using the guanidine thiocyanate method and an RNaid Kit (BIO 101, www.qbiogene.com) and stored at –80oC until use. The potential genomic DNA contamination in total RNA preparations was degraded with DNaseI prior to reverse transcription, followed by enzyme denaturation at 65oC. 1 μg of the total RNA was subjected to reverse transcription using a universal primer for the three *Ae. aegypti* vitellogenin genes at 42oC (Table 2) using Superscript reverse transcriptase (Life Technologies, www.invitrogen.com). The RNA template was subsequently removed by RNase H (Life Technologies). Polymerase chain reaction (PCR) was performed using gene-specific primers that were designed to flank each potential intron (Table 2). The PCR products were separated on agarose gels and the expected bands were purified using a Sephaglas BandPrep Kit (Pharmacia, www.Pharmacia.com). The PCR products were inserted into TA cloning vectors (Invitrogen, www.invitrogen.com/) following the manufacture's instructions, and positive clones were sequenced.

Table 2. Oligonucleotide primer sequences used for analysis of *Ae. aegypti* vitellogenin genes.

Primer			Sequence
First intron			
RT universal primer 1			5' CCAAGCCTTCAGGAT 3'
	Vg-A1	Forward	5' CAACGAGAGGAGGAGAACA 3'
		Reverse	5' GTTCGGCATCCATGAGCTG 3'
	Vg-B	Forward	5' CAACGGTGAAGACATACTG 3'
		Reverse	5' GGTAACCGTACTCGAATC 3'
	Vg-C	Forward	5' CCAGGGAATCATGTTAGTG 3'
		Reverse	5' GCCACTTCAAACTGCGATT 3'
Second intron			
RT universal primer 2			5' GCCTTCTGTTGGTAG 3'
	Vg-A1	Forward	5' CGGTCTTCAATGACTACTC 3'
		Reverse	5' AATGGAACACCACCGTTGT 3'
	Vg-B	Forward	5' TCTTCCTTCGCTACCAGC 3'
		Reverse	5' GGAGTGTCAGCGTTATTG 3'
	Vg-C	Forward	5' CAACAGTTCGCCTATCCG 3'
		Reverse	5' TCCATCGAAGGTAGCAACA 3'
3' RACE			
RT universal primer 3			5' CGGAATTCTCTAGACTCGAG(dT)$_{183}$'
		Reverse	5' CGGAATTCTCTAGACTCGAG 3'
	Vg-A1	Forward	5' TTCGGAATTCAGCTCTCTG 3'
	Vg-B	Forward	5' CGGAATCGAAGTCTAAGAG 3'
	Vg-C	Forward	5' CATCATCCTCTGAATCTCG 3'
Real-time PCR			
	Vg-A1	Forward	5' CCACAGTATCCTTCATGCTCTTT 3'
		Reverse	5' GTTCGTACTCTTCAAATCCATTCT 3'
	Vg-B	Forward	5' TCGTAAACCACAGTATCCTTCCT 3'
		Reverse	5' CTCGAAGTCAGAGTTGTAGAAGT 3'
	Vg-C	Forward	5' CGCGGACAGTACCACCCATC 3'
		Reverse	5' GTCAGCGTTCTGGAAGTGTGA 3'

Rapid amplification of cDNA ends (3' RACE) was carried out to determine the 3' untranslated region for each vitellogenin gene. Total RNA was isolated from female fat body 24 hours after a blood meal as described above. The first strand cDNA was synthesized from the total RNAs using an Oligo-(dT) containing an adapter primer (Table 2). PCR was performed on cDNA using gene-specific primers and an adapter primer. The PCR products were analyzed and cloned as above.

Gene expression pattern in *Ae. aegypti* by real-time PCR

Real-time RT-PCR was performed to quantify differences in vitellogenin gene expression in *Ae. aegypti* after a blood meal. Fat body preparations, midgut, and ovary tissues were dissected from unfed and fed mosquitoes at specific times after a blood meal. Isolated tissues were stored in RNAlater (Ambion, www.ambion.com) at –80°C until use. Total RNA was extracted TRIzol reagent (Invitrogen) according to the manufacture's instructions. Total RNA was also isolated from whole larvae, male and female pupae, and male adult. Prior to synthesizing cDNA from each sample, 1 µg total RNA was first treated with RNase-free DNaseI to ensure complete degradation of potential genomic DNA contamination, followed by the addition of 25 mM EDTA and heat denaturation of the enzyme. Reverse transcription was carried out using oligo-$(dT)_{20}$ primer and reverse transcriptase (NEB, www.neb.com) in a 20 µl reaction volume at 37°C for 1 hour followed by termination of the reaction at 70°C for 20 min. The final volume of cDNA was increased to 200 µl. To minimize potential variation in reverse transcriptase efficiency, all cDNA syntheses were carried out simultaneously. To design optimized gene-specific sense and antisense oligonucleotide primers without primer-dimer formation and self-priming formation, we used OLIGO software (version 6, Molecular Biology Insights, www.oligo.net/) for each vitellogenin gene (Table 2). Nucleotide pairwise comparison was performed using the GAP program in the GCG (Genetic Computer Group, www.accelrys.com/products/gcg/) to ensure that the primers anneal exclusively to the DNA of specific vitellogenin genes. Oligonucleotide primers were custom made by OPERON (http://www.operon.com/)

The real-time RT-PCR was carried out in the ABI PRISM 7700 Sequence Detection System (Applied Biosystems) according to the manufactures instructions in a 96-well microtiter plate with a 10 µl reaction volume containing 5 µl SYBR Green PCR Master Mix (Applied Biosystems), 3 µl of each primer set (0.5 µM final concentration), and 2 µl of cDNA templates. Each sample was run in triplicate. Negative controls without template were performed in each run. PCR conditions: preincubation was performed for 10 min at 95°C to denature the target DNA and activate AmpliTaq Gold DNA Polymerase; DNA was amplified for 40 cycles of 15 sec at 95°C and 1 min at 60°C. Data were analyzed by ABI software, Version 1 (www.appliedbiosystems.com/).

Sequence analysis

Multiple alignment of protein sequences from translation of nucleotide sequences was performed using PILEUP of GCG Wisconsin Software Package (version 10.0-UNIX) and subsequently improved by eye using MacClade (test version 4.0a11, Maddison and Maddison 1999). Pairwise comparisons was performed to estimate sequence divergence of *Ae. aegypti* vitellogenins using the PAUP* 4.0b2 package (Phylogenetic Analysis Using Parsimony, Swofford 2000). The extent of amino acid substitutions were categorized based on physicochemical properties of amino acids (Grantham 1974).

Additional gene sequences analyzed in this study were obtained from the GenBank database. Putative ribosomal protein genes were retrieved from the *Anopheles gambiae* genome by BLAST searches. The effective number of codons, ENC, was used to measure overall synonymous codon usage bias (Wright 1990). Values of ENC can range between 20, in an extremely biased gene, where only one codon is used for each amino acid, and 61, where all synonymous codons are used with the same probability. ENC, GC content, and GC3 (G+C content at the 3rd position of the synonymous codon) were calculated for each gene using CodonW (Peden 1997). A χ^2 test of homogeneity of base frequency across taxa was performed using the PAUP* 4.0b2 package (Swofford 2000) to determine the level of bias at each codon position.

Results

Cloning of vitellogenin genes

The screening of the *Ae. aegypti* genomic DNA library was conducted at low stringency to clone different members of the vitellogenin (Vg) gene family using an *Ae. aegypti* Vg-A1 DNA probe

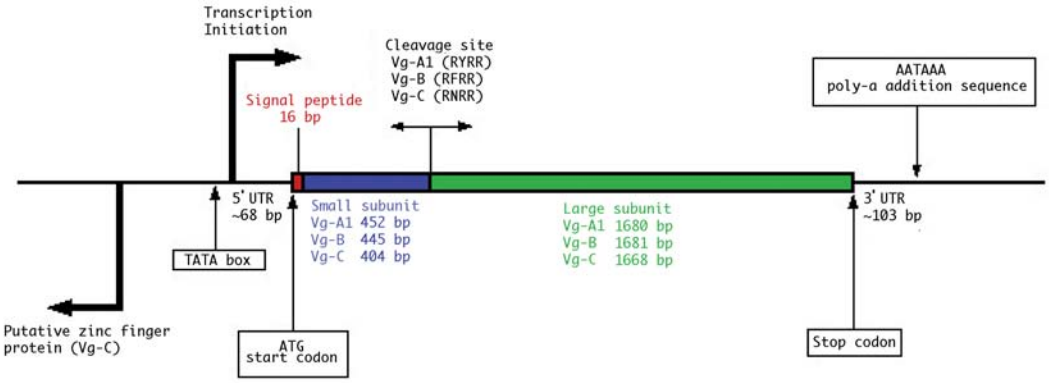


Figure 1. Generalized genomic organization of *Aedes aegypti* vitellogenin genes. The number in parenthesis indicates the number of amino acid residues.

corresponding to nucleotides 2425-4338

(accession # L41842, Romans et al. 1995) that contains the most conserved domains within the insect vitellogenin genes (Hagedorn et al. 1998). Four genes were previously isolated from an *Ae. aegypti* genomic DNA library and restriction mapped (Gemmill et al. 1986; Hamblin et al. 1987). Based on the comparisons of the restriction sites present in the four vitellogenin genes, Vg-A1, A2, B, and C, isolated in this study, and the previously mapped clones, the same four vitellogenin genes had been cloned.

Genomic libraries of *Ae. albopictus, Ae. polynesiensis, Oc. atropalpus, Oc. triseriatus, Cx. quinquefasciatus, Tx. amboinensis* and *An. albimanus* were subsequently screened at low stringency using *Ae. aegypti* Vg-B and Vg-C DNA probes, which were chosen to contain the conserved region of vitellogenin genes. Vitellogenin genes cloned from each species were also used to rescreen the library to clone possible additional members of the vitellogenin gene families. The vitellogenin genes sequenced from these mosquito species were classified as orthologous to one of the members of vitellogenin gene family in *Ae. aegypti* as judged by sequence identity, molecular signatures such as insertion and deletion events, and phylogenetic analysis of vitellogenin gene sequences.

Structure and sequence of the *Ae. aegypti* vitellogenin genes

DNA sequences in part of the coding and upstream regulatory regions of Vg-A1 and Vg-A2 of *Ae. aegypti* had high identity including the coding region (98.5% nucleotide identity with synonymous substitutions) and the upstream region (97.5%), suggesting that these two genes

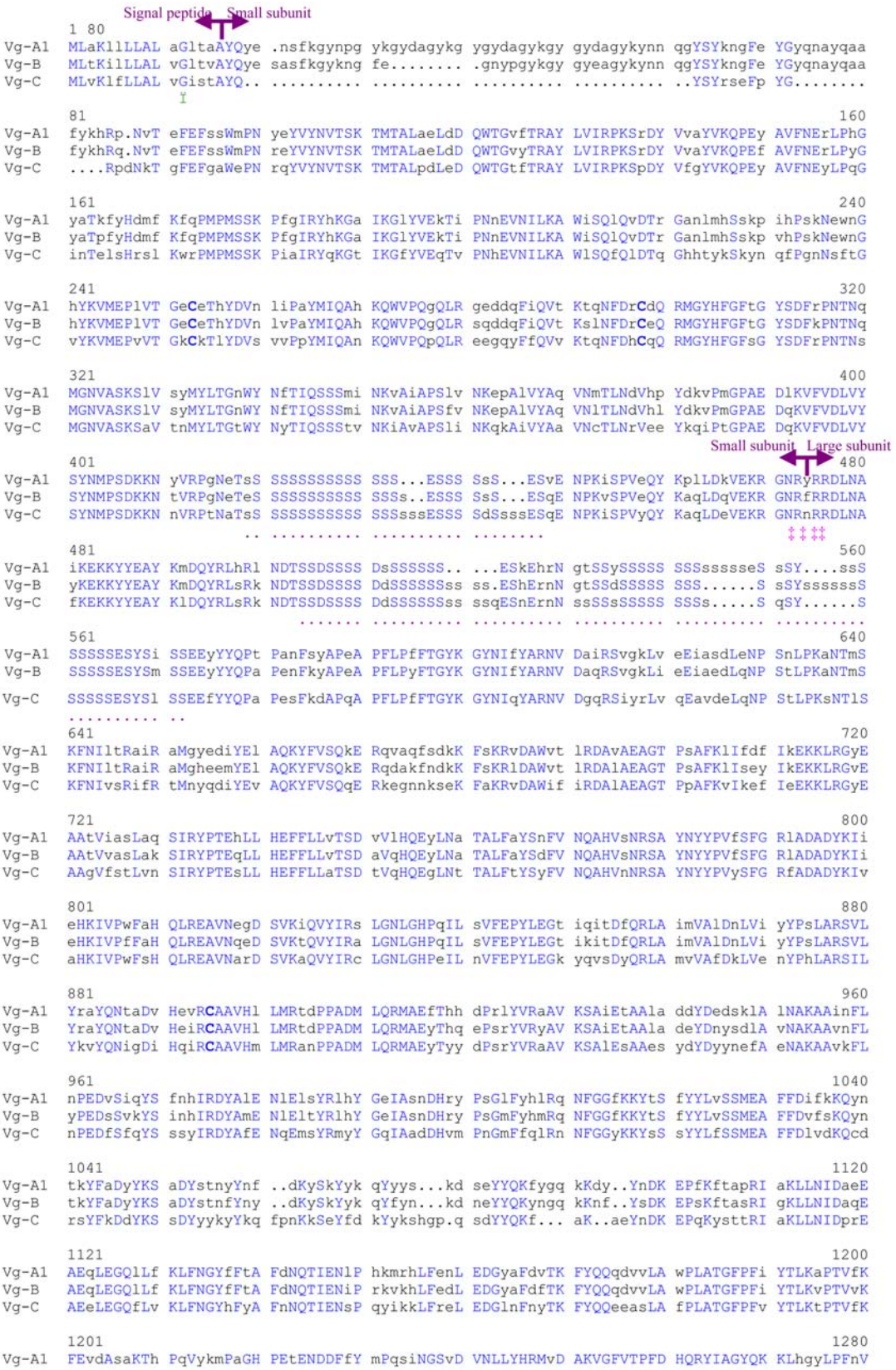
are allelic variants or derived from a recent gene duplication. Further analysis of Vg-A2 was not pursued.

Figure 1 shows the generalized structural organization of the mosquito vitellogenin genes. The intron and exon structures of Vg-B and Vg-C of *Ae. aegypti* were predicted based on their consensus 5' and 3' splice sites and by comparison to Vg-A1 cDNA (Chen et al. 1994). Total RNA isolated from fat body preparations of blood fed females was subjected to RT-PCR using gene-specific primers for Vg-B and Vg-C. The sequences derived from each PCR product established precise intron splicing sites for Vg-B and Vg-C, exactly as had been found previously for Vg-A1 (Romans et al. 1995). The 3' untranslated region of Vg-B and Vg-C was determined by 3' RACE. As described for Vg-A1, Vg-B and Vg-C had short 3' untraslated regions (110 and 84 bp, respectively), having no sequence similarity to one another or to Vg-A1. The transcription initiation site and 5' untranslated regions for Vg-B and Vg-C were predicted based on the sequence comparison with Vg-A1 (Figure 2).

Table 3. Uncorrected sequence identity (%) of three members of the vitellogenin gene family of *Aedes aegypti*

	1	2	3
1. *Vg-A1*	-	89.4	64.5
2. *Vg-B*	86.1	-	64.9
3. *Vg-C*	68.8	68.6	-

Amino acid and nucleotide sequence identity are shown above and below the diagonal, respectively.



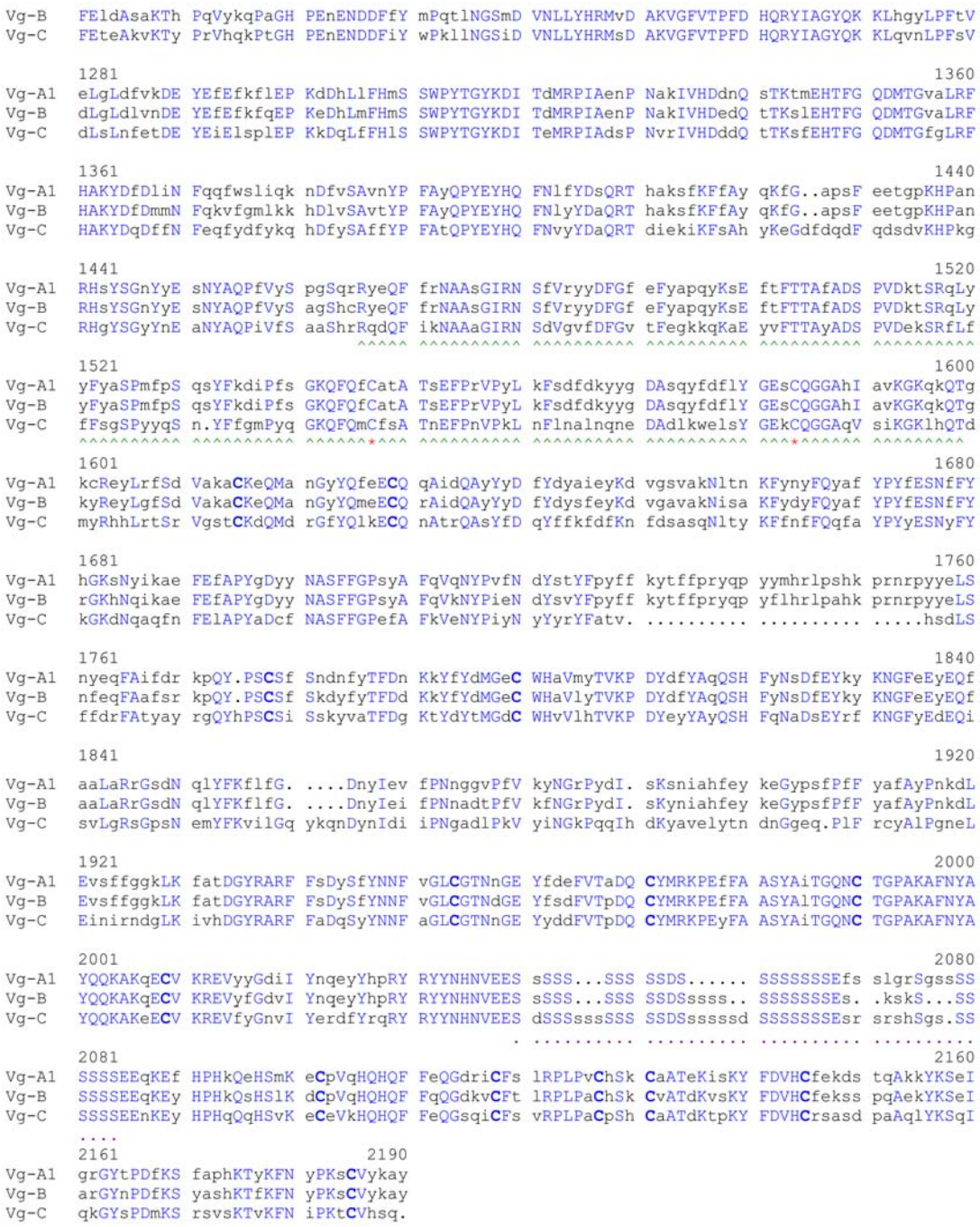


Figure 2. A multiple deduced amino acid sequence alignment of three members of vitellogenin family in *Aedes aegypti*. The conserved amino acid residues are capital letters. The conserved cysteine residues are marked with bold letters. An intron splice site is marked by Ĭ. Polyserine regions are underlined by dots (...). The symbol (^) indicates a region where no nonsynonymous substitutions occur between Vg-A1 and Vg-B. The first set of arrows indicate the end of the signal peptide and the beginning of the small subunit. The second set of arrows indicate the position of the cleavage site where the 'small' subunit ends and the 'large' subunit begins. ‡‡‡‡ indicate the cleavage sequence (RXRR) between the large and small subunits.

The deduced amino acid sequence alignment of Vg-A1, Vg-B and Vg-C of *Ae. aegypti*, is shown in Figure 2, and the uncorrected pairwise sequence divergence of amino acid and nucleotide sequences is presented in Table 3. Sequence comparison of the coding regions of Vg-A1 and Vg-B showed high identity (86.1% and 89.4%) for the nucleotides and deduced amino acids, respectively. Romans et al. (1995) showed that the genomic sequence of Vg-A1 and the cDNA

sequence of Chen et al. (1994) showed no significant differences except that the Vg-A1 sequence has an extra imperfect repeat of amino acid sequences near the N-terminal region of the large subunit. Vg-B has two imperfect repeats of this sequence. Sequence comparisons between Vg-C and other members of the vitellogenin gene family showed that Vg-C is more different from Vg-A1 and B, and that nucleotide identity was higher than amino acid identity, suggesting the presence of more amino acid substitutions (Table 3).

The vitellogenin genes of eight other mosquito species

The vitellogenin genes of eight other mosquitoes were also examined. Table 1 provides information about the relationships between these mosquitoes. Appendix 1 shows the deduced amino acid sequence alignment of these genes. Two incomplete sequences coding for vitellogenin genes orthologous to *Ae. aegypti* Vg-B and Vg-C were cloned from the autogenous *Oc. atropalpus* mosquito. Screening the *Cx. quinquefasciatus* library resulted in the cloning of two different members of the vitellogenin gene family that appear to be orthologous to Vg-C of *Ae. aegypti* and were designated as Vg-C1 and Vg-C2. The Vg-C1 clone contains a complete coding sequence of the gene, whereas the Vg-C2 clone has all of the molecular signatures of a vitellogenin gene but no frame shifts and stop codons are present. However, this gene is truncated in such a way that 39 amino acid residues at the N-terminal are missing from the coding region. Partial sequences coding for the N-terminal region of vitellogenin genes from three anautogenous mosquitoes, *Ae. polynesiensis* (orthologous to *Ae. aegypti* Vg-A1, Vg-B, and Vg-C), *Ae. albopictus* (orthologous to *Ae. aegypti* Vg-A1 and Vg-C) *and Oc. triseriatus* (orthologous to *Ae aegypti* Vg-C) were also sequenced (Appendix 1).

Intensive screening of *An. albimanus* and *Tx. amboinensis* libraries resulted in the isolation of a single vitellogenin gene for each of these species. Subsequent screening of both libraries with homologous probes did not yield additional members of vitellogenin genes of either species. The vitellogenin genes sequenced from *An. albimanus* and *Tx. amboinensis* appear to be orthologous to *Ae. aegypti* Vg-C (Appendix 1). For example, no duplicated imperfect repeats were found between Vg-C of *Ae. aegypti* and the vitellogenin genes from *An. albimanus* and *Tx. amboinensis*. Also the alignments show that there are several stretches of amino acid deletions and insertions in identical positions in Vg-C of *Ae. aegypti* and the vitellogenin genes in *An. albimanus* and *Tx. amboinensis* that are not present in Vg-A1 and Vg-B of *Ae. aegypti*, suggesting that the *Ae. aegypti* Vg-C and the vitellogenin gene cloned from these two species probably shared a recent common ancestor, and that Vg-A1 and Vg-B of *Ae. aegypti* are distantly related to them. Comparison of the incomplete pairwise sequence divergence of Vg-C genes orthologous to Vg-C of *Ae. aegypti* shows that they have relatively high homology with Vg-C of *Ae. aegypti* (Table 4).

Table 4. Uncorrected sequence identity (%) of six mosquito vitellogenin Vg-C genes.

	1	2	3	4	5	6
1 *Ae. aegypti C*	-	85.1	68.7	55.9	65.4	64.7
2 *Oc. atropalpus C*	79.8	-	69.8	54.9	66.5	65.6
3 *Cx. quinquefasciatus C1*	71.2	70.2	-	57.8	65.2	64.0
4 *Cx. quinquefasciatus C2*	61.7	60.0	63.6	-	54.9	57.3
5 *Tx. amboinensis C*	67.8	68.3	67.9	60.3	-	61.6
6 *An. albimanus C*	69.1	67.5	70.5	64.6	66.2	-

Amino acid and nucleotide sequence identity are shown above and below the diagonal, respectively.

Amino acid composition of vitellogenin proteins

There are no dramatic differences in the amino acid composition between the mosquito vitellogenins (Table 5). Three polyserine regions are present (Appendix 1). Most are rich in tyrosine (Y) and phenylalanine (F), however, Vg-C1 and Vg-C2 of *Cx. quinquefasciatus* have a lower number of these two amino acids compared to the other mosquitoes (Table 5). The reduction is more prominent in the truncated gene, Vg-C2, than in Vg-C1. A reduction in the number of serine residues in *An. albimanus* Vg-C by about 30% is also apparent. The deduced amino acid composition of vitellogenin proteins from other insects is also shown in Table 5. Interestingly, none of the vitellogenin proteins of other insects have a high content of tyrosine and phenylalanine.

Ae. aegypti vitellogenin proteins show a biased amino acid composition compared to 144 other *Ae. aegypti* proteins available from the GenBank database (Table 6). In general, the aromatic amino acids tyrosine and phenylalanine that are rich in most mosquito vitellogenin proteins contribute a relatively low proportion of amino acids in all other proteins examined, with the

Table 5. Amino acid composition of insect vitellogenin proteins, not including mitochondrial proteins. Blue = # of amino acids. Red = percent of total # of amino acids.

Amino acid single letter code																					
GENES	**A**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**K**	**L**	**M**	**N**	**P**	**Q**	**R**	**S**	**T**	**V**	**W**	**Y**	**Total**
Mosquitoes																					
Aedes aegypti (Vg-A1)	155	20	116	123	174	102	61	69	148	114	39	116	104	102	79	213	80	104	12	217	2148
	7.2	0.9	5.4	5.7	8.1	4.7	2.8	3.2	6.9	5.3	1.8	5.4	4.8	4.7	3.7	9.9	3.7	4.8	0.6	10.1	
Aedes aegypti (Vg-B)	157	20	116	126	171	95	59	56	155	111	43	111	101	109	77	220	83	109	10	213	2142
	7.3	0.9	5.4	5.9	8.0	4.4	2.8	2.6	7.2	5.2	2.0	5.2	4.7	5.1	3.6	10.3	3.9	5.1	0.5	9.9	
Aedes aegypti (Vg-C)	131	24	118	120	151	93	52	68	140	103	32	122	97	127	86	230	86	111	12	184	2087
	6.3	1.1	5.7	5.7	7.2	4.5	2.5	3.3	6.7	4.9	1.5	5.8	4.6	6.1	4.1	11.0	4.1	5.3	0.6	8.8	
Ochlerotatus atropalpus (Vg-B)	148	18	127	126	177	92	61	58	138	103	33	109	96	114	73	218	82	107	12	216	2108
	7.0	0.9	6.0	6.0	8.4	4.4	2.9	2.8	6.5	4.9	1.6	5.2	4.6	5.4	3.5	10.3	3.9	5.1	0.6	10.2	
Ochlerotatus atropalpus (Vg-C)	118	15	105	120	152	78	45	58	116	92	33	114	88	100	67	172	76	94	14	168	1825
	6.5	0.8	5.8	6.6	8.3	4.3	2.5	3.2	6.4	5.0	1.8	6.2	4.8	5.5	3.7	9.4	4.2	5.2	0.8	9.2	
Culex quinquefasciatus (Vg-C1)	135	26	121	149	102	95	55	83	146	134	40	126	93	128	89	214	90	137	19	129	2111
	6.4	1.2	5.7	7.1	4.8	4.5	2.6	3.9	6.9	6.3	1.9	6.0	4.4	6.1	4.2	10.1	4.3	6.5	0.9	6.1	
Culex quinquefasciatus (Vg-C2)	161	24	108	145	59	77	58	95	140	147	38	118	94	115	88	235	95	139	28	78	2042
	7.9	1.2	5.3	7.1	2.9	3.8	2.8	4.7	6.9	7.2	1.9	5.8	4.6	5.6	4.3	11.5	4.7	6.8	1.4	3.8	
Toxyrhynchites amboinensis (Vg-C)	156	19	101	125	139	90	66	67	118	107	38	133	95	125	84	207	82	121	15	144	2032
	7.7	0.9	5.0	6.2	6.8	4.4	3.2	3.3	5.8	5.3	1.9	6.5	4.7	6.2	4.1	10.2	4.0	6.0	0.7	7.1	
Anopheles albimanus (Vg-C)	122	24	130	137	148	93	65	72	124	119	33	104	92	116	82	154	98	123	16	177	2029
	6.0	1.2	6.4	6.8	7.3	4.6	3.2	3.5	6.1	5.9	1.6	5.1	4.5	5.7	4.0	7.6	4.8	6.1	0.8	8.7	
Anopheles gambiae	126	23	128	134	154	92	59	63	129	116	43	114	92	124	80	174	94	121	17	168	2051
	6.1	1.1	6.2	6.5	7.5	4.5	2.9	3.1	6.3	5.7	2.1	5.6	4.5	6.0	3.9	8.5	4.6	5.9	0.8	8.2	
Other insects																					
Apis mellifera	76	21	109	126	66	79	39	103	138	158	66	108	83	60	82	150	110	129	18	49	1770
	4.3	1.2	6.2	7.1	3.7	4.5	2.2	5.8	7.8	8.9	3.7	6.1	4.7	3.4	4.6	8.5	6.2	7.3	1.0	2.8	
Athalia roase	178	23	89	117	61	108	36	81	105	122	34	119	89	140	78	215	83	102	19	73	1872
	9.5	1.2	4.8	6.3	3.3	5.8	1.9	4.3	5.6	6.5	1.8	6.4	4.8	7.5	4.2	11.5	4.4	5.4	1.0	3.9	
Antheraea pernyi	134	13	87	118	63	75	62	93	135	123	36	98	82	126	69	164	91	110	15	84	1778
	7.5	0.7	4.9	6.6	3.5	4.2	3.5	5.2	7.6	6.9	2.0	5.5	4.6	7.1	3.9	9.2	5.1	6.2	0.8	4.7	
Antheraea yamamai	129	15	87	118	58	74	59	93	130	127	38	102	84	125	71	164	94	112	15	83	1778
	7.3	0.8	4.9	6.6	3.3	4.2	3.3	5.2	7.3	7.1	2.1	5.7	4.7	7.0	4.0	9.2	5.3	6.3	0.8	4.7	
Apis mellifera	76	21	109	126	66	79	39	103	138	158	66	108	83	60	82	150	110	129	18	49	1770
	4.3	1.2	6.2	7.1	3.7	4.5	2.2	5.8	7.8	8.9	3.7	6.1	4.7	3.4	4.6	8.5	6.2	7.3	1.0	2.8	
Blattella germanica	92	19	94	104	85	55	69	93	126	155	53	124	85	86	78	187	112	141	19	85	1862
	4.9	1.0	5.0	5.6	4.6	3.0	3.7	5.0	6.8	8.3	2.8	6.7	4.6	4.6	4.2	10.0	6.0	7.6	1.0	4.6	
Bombyx mori	121	14	100	105	78	80	60	94	115	116	39	75	81	123	83	175	108	105	16	94	1782
	6.8	0.8	5.6	5.9	4.4	4.5	3.4	5.3	6.5	6.5	2.2	4.2	4.5	6.9	4.7	9.8	6.1	5.9	0.9	5.3	
Graptopsaltria nigrofuscata	140	21	95	100	83	91	43	97	125	144	42	152	99	95	89	224	105	142	15	85	1987
	7.0	1.1	4.8	5.0	4.2	4.6	2.2	4.9	6.3	7.2	2.1	7.6	5.0	4.8	4.5	11.3	5.3	7.1	0.8	4.3	
Leucophaea maderae	98	16	122	77	74	63	59	97	106	165	40	137	86	125	125	187	106	133	16	81	1913
	5.1	0.8	6.4	4.0	3.9	3.3	3.1	5.1	5.5	8.6	2.1	7.2	4.5	6.5	6.5	9.8	5.5	7.0	0.8	4.2	
Lymantria dispar	123	15	75	101	80	85	56	100	117	149	22	102	74	123	61	130	119	113	23	79	1747
	7.0	0.9	4.3	5.8	4.6	4.9	3.2	5.7	6.7	8.5	1.3	5.8	4.2	7.0	3.5	7.4	6.8	6.5	1.3	4.5	
Periplaneta americana (Vg-1)	150	15	113	116	54	68	45	122	86	176	54	137	105	74	128	187	101	101	19	45	1896
	7.9	0.8	6.0	6.1	2.8	3.6	2.4	6.4	4.5	9.3	2.8	7.2	5.5	3.9	6.8	9.9	5.3	5.3	1.0	2.4	
Periplaneta americana (Vg-2)	113	20	101	142	90	63	62	73	138	125	38	99	90	69	122	183	86	158	11	93	1876
	6.0	1.1	5.4	7.6	4.8	3.4	3.3	3.9	7.4	6.7	2.0	5.3	4.8	3.7	6.5	9.8	4.6	8.4	0.6	5.0	
Pimpla nipponica	136	20	128	126	76	87	46	83	147	133	48	91	89	67	89	152	104	111	16	58	1807
	7.5	1.1	7.1	7.0	4.2	4.8	2.5	4.6	8.1	7.4	2.7	5.0	4.9	3.7	4.9	8.4	5.8	6.1	0.9	3.2	
Plautia stali (Vg-1)	125	13	79	121	112	121	59	74	116	114	28	119	92	116	68	205	81	116	14	134	1907
	6.6	0.7	4.1	6.3	5.9	6.3	3.1	3.9	6.1	6.0	1.5	6.2	4.8	6.1	3.6	10.7	4.2	6.1	0.7	7.0	
Plautia stali (Vg-2)	140	15	69	116	88	105	57	74	105	110	29	114	84	135	68	192	92	119	14	130	1856
	7.5	0.8	3.7	6.3	4.7	5.7	3.1	4.0	5.7	5.9	1.6	6.1	4.5	7.3	3.7	10.3	5.0	6.4	0.8	7.0	
Plautia stali (Vg-3)	125	15	66	100	52	123	43	82	103	138	36	118	92	153	78	199	123	146	12	99	1903
	6.6	0.8	3.5	5.3	2.7	6.5	2.3	4.3	5.4	7.3	1.9	6.2	4.8	8.0	4.1	10.5	6.5	7.7	0.6	5.2	
Riptortus clavatus	143	14	61	118	57	97	46	83	99	126	34	128	78	191	84	218	80	113	15	91	1876
	7.6	0.7	3.3	6.3	3.0	5.2	2.5	4.4	5.3	6.7	1.8	6.8	4.2	10.2	4.5	11.6	4.3	6.0	0.8	4.9	
Samia cynthia ricini	132	13	83	124	57	79	53	89	143	116	40	93	83	126	68	162	97	116	15	90	1779
	7.4	0.7	4.7	7.0	3.2	4.4	3.0	5.0	8.0	6.5	2.2	5.2	4.7	7.1	3.8	9.1	5.5	6.5	0.8	5.1	

exception of the hexamerin proteins, which are also rich in methionine (Figure 3). The hexamerins of the mosquitoes, *Oc. atropalpus* (AAL29455), *An. gambiae* (U51225), and larval *Ae. aegypti* (Gordadze et al. 1999) are also high in tyrosine and phenylalanine.

Conservative and non-conservative amino acid substitutions in *Ae. aegypti* Vg

Grantham (1974) scored differences between amino acid substitutions based on the physical and chemical properties of amino acids. He categorized amino acid changes as conserved (score < 50), moderately conserved (50 < score < 100), moderately radical (100 < score < 150), and radical substitution (score > 150). Using this scoring scheme, the entire coding regions of vitellogenin genes were divided into 30 amino acid blocks each based on the alignment of the three *Ae. aegypti* vitellogenin proteins (Figure 2) and compared to assess the degree of amino acid substitutions.

Table 6. Amino acid composition of *Aedes aegypti* proteins, not including mitochondrial proteins. Blue = # of amino acids. Red = percent of total # of amino acids.

		Amino acid single letter code																				
GENES	**Accession #**	**A**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**K**	**L**	**M**	**N**	**P**	**Q**	**R**	**S**	**T**	**V**	**W**	**Y**	**Total**
Actin	U20287	28	6	23	25	10	28	10	28	20	27	15	11	18	10	20	29	23	25	4	16	376
		7.4	1.6	6.1	6.6	2.7	7.4	2.7	7.4	5.3	7.2	4.0	2.9	4.8	2.7	5.3	7.7	6.1	6.6	1.1	4.3	
Allatotropin	U65314	16	4	8	11	7	11	3	9	7	16	6	10	8	5	11	19	9	11	3	1	175
		9.1	2.3	4.6	6.3	4.0	6.3	1.7	5.1	4.0	9.1	3.4	5.7	4.6	2.9	6.3	10.9	5.1	6.3	1.7	0.6	
Amino Acid Transporter	AF543193	41	9	26	23	41	50	11	58	30	68	14	23	22	17	17	54	32	53	23	28	640
		6.4	1.4	4.1	3.6	6.4	7.8	1.7	9.1	4.7	10.6	2.2	3.6	3.4	2.7	2.7	8.4	5.0	8.3	3.6	4.4	
Aminopeptidase N	AY064079	83	5	57	48	36	52	25	43	36	99	19	62	34	40	45	69	70	65	20	47	955
		8.7	0.5	6.0	5.0	3.8	5.4	2.6	4.5	3.8	10.4	2.0	6.5	3.6	4.2	4.7	7.2	7.3	6.8	2.1	4.9	
Amylase I	AF000569	54	13	34	27	43	53	13	45	26	46	14	68	47	38	34	65	35	56	7	19	737
		7.3	1.8	4.6	3.7	5.8	7.2	1.8	6.1	3.5	6.2	1.9	9.2	6.4	5.2	4.6	8.8	4.7	7.6	0.9	2.6	
Amylase II	AF000568	33	11	26	20	25	53	17	29	10	30	13	38	23	18	28	30	16	34	15	17	486
		6.8	2.3	5.3	4.1	5.1	10.9	3.5	6.0	2.1	6.2	2.7	7.8	4.7	3.7	5.8	6.2	3.3	7.0	3.1	3.5	
Angiopoietin-Like Protein	AF466609	13	3	19	18	17	23	6	11	17	26	8	12	8	12	15	29	17	17	9	16	296
		4.4	1.0	6.4	6.1	5.7	7.8	2.0	3.7	5.7	8.8	2.7	4.1	2.7	4.1	5.1	9.8	5.7	5.7	3.0	5.4	
Apyrase	L41391	44	8	46	29	18	36	12	46	56	42	7	25	27	16	24	29	28	44	5	21	563
		7.8	1.4	8.2	5.2	3.2	6.4	2.1	8.2	9.9	7.5	1.2	4.4	4.8	2.8	4.3	5.2	5.0	7.8	0.9	3.7	
Aquaporin	AF218314	51	3	24	25	17	43	6	20	35	39	10	21	16	8	14	24	15	36	2	6	415
		12.3	0.7	5.8	6.0	4.1	10.4	1.4	4.8	8.4	9.4	2.4	5.1	3.9	1.9	3.4	5.8	3.6	8.7	0.5	1.4	
Asparagine Synthetase	U84118	40	12	35	43	25	41	14	33	31	49	14	18	26	17	33	32	26	36	7	30	562
		7.1	2.1	6.2	7.7	4.4	7.3	2.5	5.9	5.5	8.7	2.5	3.2	4.6	3.0	5.9	5.7	4.6	6.4	1.2	5.3	
Carbonic Anhydrase	AF395662	25	2	15	13	11	21	14	16	10	29	6	14	19	11	12	23	22	21	4	10	298
		8.4	0.7	5.0	4.4	3.7	7.0	4.7	5.4	3.4	9.7	2.0	4.7	6.4	3.7	4.0	7.7	7.4	7.0	1.3	3.4	
Carboxypeptidase A	AF165923	28	6	22	31	16	35	11	21	23	37	4	27	17	13	18	33	23	26	8	28	427
		6.6	1.4	5.2	7.3	3.7	8.2	2.6	4.9	5.4	8.7	0.9	6.3	4.0	3.0	4.2	7.7	5.4	6.1	1.9	6.6	
Cathepsin B-Like Thiol Protease	L41940	30	14	17	13	12	38	9	18	16	19	7	17	15	17	16	23	11	20	12	18	342
		8.8	4.1	5.0	3.8	3.5	11.1	2.6	5.3	4.7	5.6	2.0	5.0	4.4	5.0	4.7	6.7	3.2	5.8	3.5	5.3	
Cellular Retinaldehyde-Binding Protein	AY064091	14	4	19	21	19	12	9	15	25	38	7	12	17	9	16	13	10	17	3	10	290
		4.8	1.4	6.6	7.2	6.6	4.1	3.1	5.2	8.6	13.1	2.4	4.1	5.9	3.1	5.5	4.5	3.4	5.9	1.0	3.4	
Cellular Retinaldehyde-Binding Protein	AF329893	13	4	19	21	19	12	9	15	25	38	7	13	17	9	16	12	11	17	3	10	290
		4.5	1.4	6.6	7.2	6.6	4.1	3.1	5.2	8.6	13.1	2.4	4.5	5.9	3.1	5.5	4.1	3.8	5.9	1.0	3.4	
Chitin Synthase	AF223577	104	13	67	86	49	90	27	58	79	108	36	45	62	46	65	65	54	101	4	36	1195
		8.7	1.1	5.6	7.2	4.1	7.5	2.3	4.9	6.6	9.0	3.0	3.8	5.2	3.8	5.4	5.4	4.5	8.5	0.3	3.0	
Chitinase 1	AF026491	45	14	46	37	26	45	9	22	31	25	13	20	40	21	26	29	44	39	13	29	574
		7.8	2.4	8.0	6.4	4.5	7.8	1.6	3.8	5.4	4.4	2.3	3.5	7.0	3.7	4.5	5.1	7.7	6.8	2.3	5.1	
Chitinase 2	AF026492	109	40	121	90	58	119	36	62	112	99	28	69	87	44	88	94	137	100	53	89	1635
		6.7	2.4	7.4	5.5	3.5	7.3	2.2	3.8	6.9	6.1	1.7	4.2	5.3	2.7	5.4	5.7	8.4	6.1	3.2	5.4	
Chymotrypsin	AF487334	18	6	16	11	8	28	10	17	14	22	1	14	9	3	12	20	15	19	5	9	257
		7.0	2.3	6.2	4.3	3.1	10.9	3.9	6.6	5.4	8.6	0.4	5.4	3.5	1.2	4.7	7.8	5.8	7.4	1.9	3.5	
Chymotrypsin-Like Protease Precursor	U56423	27	6	18	17	12	21	7	9	13	20	2	8	12	12	8	21	12	32	5	6	268
		10.1	2.2	6.7	6.3	4.5	7.8	2.6	3.4	4.9	7.5	0.7	3.0	4.5	4.5	3.0	7.8	4.5	11.9	1.9	2.2	
Cyclic-AMP Response Element Binding Protein	AY083158	21	4	14	16	1	27	9	15	15	23	6	18	14	28	15	35	12	17	0	5	295
		7.1	1.4	4.7	5.4	0.3	9.2	3.1	5.1	5.1	7.8	2.0	6.1	4.7	9.5	5.1	11.9	4.1	5.8	0.0	1.7	
2-Cys Thioredoxin Peroxidase	AF323181	12	3	12	13	14	15	3	9	14	16	3	7	13	8	8	11	10	18	2	5	196
		6.1	1.5	6.1	6.6	7.1	7.7	1.5	4.6	7.1	8.2	1.5	3.6	6.6	4.1	4.1	5.6	5.1	9.2	1.0	2.6	
Cytochrome P450	AY064093	39	5	32	35	42	23	10	27	41	53	16	19	24	14	28	33	31	37	4	23	536
		7.3	0.9	6.0	6.5	7.8	4.3	1.9	5.0	7.6	9.9	3.0	3.5	4.5	2.6	5.2	6.2	5.8	6.9	0.7	4.3	
D7 Protein	M33157	23	10	22	24	16	11	5	9	41	25	6	15	16	16	10	18	14	22	3	15	321
		7.2	3.1	6.9	7.5	5.0	3.4	1.6	2.8	12.8	7.8	1.9	4.7	5.0	5.0	3.1	5.6	4.4	6.9	0.9	4.7	
Defensin Isoform A1	AF156088	12	8	5	5	5	7	1	4	4	8	1	5	4	2	6	7	5	6	0	3	98
		12.2	8.2	5.1	5.1	5.1	7.1	1.0	4.1	4.1	8.2	1.0	5.1	4.1	2.0	6.1	7.1	5.1	6.1	0.0	3.1	
Defensin Isoform B1	AF156090	13	8	4	6	5	6	1	4	3	8	1	5	4	3	6	6	5	7	0	3	98
		13.3	8.2	4.1	6.1	5.1	6.1	1.0	4.1	3.1	8.2	1.0	5.1	4.1	3.1	6.1	6.1	5.1	7.1	0.0	3.1	
Defensin Isoform C1	AF156092	13	8	5	6	5	6	1	3	3	10	1	5	2	1	8	7	4	8	0	3	99
		13.1	8.1	5.1	6.1	5.1	6.1	1.0	3.0	3.0	10.1	1.0	5.1	2.0	1.0	8.1	7.1	4.0	8.1	0.0	3.0	
Dopa Carboxylase	U27581	35	6	14	28	14	23	8	19	16	30	10	14	22	9	12	19	16	21	9	12	337
		10.4	1.8	4.2	8.3	4.2	6.8	2.4	5.6	4.7	8.9	3.0	4.2	6.5	2.7	3.6	5.6	4.7	6.2	2.7	3.6	
Dopachrome Conversion Enzyme	AY064099	20	5	33	21	27	30	9	26	19	34	8	25	34	19	28	29	24	42	10	20	463
		4.3	1.1	7.1	4.5	5.8	6.5	1.9	5.6	4.1	7.3	1.7	5.4	7.3	4.1	6.0	6.3	5.2	9.1	2.2	4.3	
E74A	AF435023	28	3	10	6	13	33	4	23	7	18	9	6	8	6	8	14	17	25	5	6	249
		11.2	1.2	4.0	2.4	5.2	13.3	1.6	9.2	2.8	7.2	3.6	2.4	3.2	2.4	3.2	5.6	6.8	10.0	2.0	2.4	
E74B	AF435022	45	7	44	37	11	69	55	20	28	61	16	20	80	49	61	113	40	33	4	34	827
		5.4	0.8	5.3	4.5	1.3	8.3	6.7	2.4	3.4	7.4	1.9	2.4	9.7	5.9	7.4	13.7	4.8	4.0	0.5	4.1	
Early Trypsin	X64362	31	32	22	20	8	17	17	10	13	20	1	19	52	29	1	59	77	25	9	24	486
		6.4	6.6	4.5	4.1	1.6	3.5	3.5	2.1	2.7	4.1	0.2	3.9	10.7	6.0	0.2	12.1	15.8	5.1	1.9	4.9	
Ecdysteroid Receptor	U02021	37	18	28	38	14	45	20	35	30	62	27	54	33	42	32	76	31	30	4	19	675
		5.5	2.7	4.1	5.6	2.1	6.7	3.0	5.2	4.4	9.2	4.0	8.0	4.9	6.2	4.7	11.3	4.6	4.4	0.6	2.8	
Elongation Factor 2	AY064104	61	17	66	53	33	58	16	49	65	62	32	31	42	26	43	38	38	80	7	27	844
		7.2	2.0	7.8	6.3	3.9	6.9	1.9	5.8	7.7	7.3	3.8	3.7	5.0	3.1	5.1	4.5	4.5	9.5	0.8	3.2	
Enhancer of Rudimentary	U66869	4	3	6	8	2	3	2	6	6	9	2	6	5	5	5	6	10	6	1	8	103
		3.9	2.9	5.8	7.8	1.9	2.9	1.9	5.8	5.8	8.7	1.9	5.8	4.9	4.9	4.9	5.8	9.7	5.8	1.0	7.8	
Eye Pigment Transporter	U88851	51	17	27	34	42	42	8	39	36	75	20	28	33	24	33	59	41	49	9	25	692
		7.4	2.5	3.9	4.9	6.1	6.1	1.2	5.6	5.2	10.8	2.9	4.0	4.8	3.5	4.8	8.5	5.9	7.1	1.3	3.6	
Ferritin Subunit	L37082	23	2	12	20	12	11	8	12	19	17	5	6	4	11	6	8	11	12	1	9	209
		11.0	1.0	5.7	9.6	5.7	5.3	3.8	5.7	9.1	8.1	2.4	2.9	1.9	5.3	2.9	3.8	5.3	5.7	0.5	4.3	
Fxa-Directed Anticoagulant Precursor	AF050133	24	6	22	16	31	17	8	24	28	39	17	19	20	21	22	36	21	24	5	15	415
		5.8	1.4	5.3	3.9	7.5	4.1	1.9	5.8	6.7	9.4	4.1	4.6	4.8	5.1	5.3	8.7	5.1	5.8	1.2	3.6	
GABA Receptor Subunit	U28803	24	7	22	22	25	54	20	38	20	44	16	16	35	20	27	43	31	42	6	21	533

GENES	Accession #	Amino acid single letter code																				
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
		4.5	1.3	4.1	4.1	4.7	10.1	3.8	7.1	3.8	8.3	3.0	3.0	6.6	3.8	5.1	8.1	5.8	7.9	1.1	3.9	
GATA Transcription Factor	AJ400338	73	13	36	41	11	51	67	27	24	46	14	67	47	49	32	146	44	45	6	29	868
		8.4	1.5	4.1	4.7	1.3	5.9	7.7	3.1	2.8	5.3	1.6	7.7	5.4	5.6	3.7	16.8	5.1	5.2	0.7	3.3	
Glucosamine-Fructose-6-Phosphate Aminotransferase	AF399922	42	16	36	48	22	46	22	50	36	62	20	23	23	26	38	41	44	55	1	24	675
		6.2	2.4	5.3	7.1	3.3	6.8	3.3	7.4	5.3	9.2	3.0	3.4	3.4	3.9	5.6	6.1	6.5	8.1	0.1	3.6	
Glutamine Synthetase	AF004351	35	8	29	21	10	33	7	18	25	30	8	18	25	14	24	23	19	24	9	20	400
		8.8	2.0	7.3	5.3	2.5	8.3	1.8	4.5	6.3	7.5	2.0	4.5	6.3	3.5	6.0	5.8	4.8	6.0	2.3	5.0	
Glutathione S-Transferase	AF384858	20	2	14	16	7	13	4	9	10	28	6	9	14	9	7	13	11	16	3	7	218
		9.2	0.9	6.4	7.3	3.2	6.0	1.8	4.1	4.6	12.8	2.8	4.1	6.4	4.1	3.2	6.0	5.0	7.3	1.4	3.2	
Head Peptide I	AF155738	15	2	16	7	8	26	3	5	11	19	3	6	8	4	18	22	3	7	1	13	197
		7.6	1.0	8.1	3.6	4.1	13.2	1.5	2.5	5.6	9.6	1.5	3.0	4.1	2.0	9.1	11.2	1.5	3.6	0.5	6.6	
Heavy Clathrin Chain	S:1076	149	32	89	130	69	70	40	106	99	183	45	91	60	84	82	88	69	108	17	71	1682
		8.9	1.9	5.3	7.7	4.1	4.2	2.4	6.3	5.9	10.9	2.7	5.4	3.6	5.0	4.9	5.2	4.1	6.4	1.0	4.2	
Hexamerin 1 Gamma	U86079	13	4	55	30	63	33	9	39	56	56	58	32	23	11	23	31	39	36	14	67	692
		1.9	0.6	7.9	4.3	9.1	4.8	1.3	5.6	8.1	8.1	8.4	4.6	3.3	1.6	3.3	4.5	5.6	5.2	2.0	9.7	
Hexamerin 2 Alpha	U86080	36	2	47	45	47	40	30	26	47	48	20	30	35	22	24	25	34	71	11	72	712
		5.1	0.3	6.6	6.3	6.6	5.6	4.2	3.7	6.6	6.7	2.8	4.2	4.9	3.1	3.4	3.5	4.8	10.0	1.5	10.1	
HNF4 Isoform A	AF059026	36	17	46	30	18	23	16	34	26	55	16	39	25	28	35	57	25	23	2	14	565
		6.4	3.0	8.1	5.3	3.2	4.1	2.8	6.0	4.6	9.7	2.8	6.9	4.4	5.0	6.2	10.1	4.4	4.1	0.4	2.5	
Insulin Receptor	U72939	68	43	89	87	46	88	28	85	83	116	36	77	81	47	77	86	86	96	15	56	1390
		4.9	3.1	6.4	6.3	3.3	6.3	2.0	6.1	6.0	8.3	2.6	5.5	5.8	3.4	5.5	6.2	6.2	6.9	1.1	4.0	
Juvenile Hormone Epoxide Hydrolase	AF517544	26	2	28	28	27	33	8	26	27	51	10	18	23	19	26	20	22	33	7	27	461
		5.6	0.4	6.1	6.1	5.9	7.2	1.7	5.6	5.9	11.1	2.2	3.9	5.0	4.1	5.6	4.3	4.8	7.2	1.5	5.9	
Kynurenine Aminotransferase	AF395204	38	8	20	30	21	31	11	25	31	40	16	24	25	19	18	31	28	33	9	19	477
		8.0	1.7	4.2	6.3	4.4	6.5	2.3	5.2	6.5	8.4	3.4	5.0	5.2	4.0	3.8	6.5	5.9	6.9	1.9	4.0	
Kynurenine Monooxygenase	AF325508	35	9	30	21	22	33	16	25	32	50	16	23	17	13	25	25	25	26	6	27	476
		7.4	1.9	6.3	4.4	4.6	6.9	3.4	5.3	6.7	10.5	3.4	4.8	3.6	2.7	5.3	5.3	5.3	5.5	1.3	5.7	
Late Trypsin Precursor	M77814	20	6	13	12	13	26	4	3	14	25	2	12	12	12	7	17	21	30	3	5	257
		7.8	2.3	5.1	4.7	5.1	10.1	1.6	1.2	5.4	9.7	0.8	4.7	4.7	4.7	2.7	6.6	8.2	11.7	1.2	1.9	
Lipase-Like Protein	AF303984	23	14	13	8	7	30	11	19	8	21	8	17	22	5	17	12	12	14	4	13	278
		8.3	5.0	4.7	2.9	2.5	10.8	4.0	6.8	2.9	7.6	2.9	6.1	7.9	1.8	6.1	4.3	4.3	5.0	1.4	4.7	
Lipophorin Receptor	AF355595	53	72	103	74	31	75	37	54	71	64	24	57	64	30	54	89	96	57	16	35	1156
		4.6	6.2	8.9	6.4	2.7	6.5	3.2	4.7	6.1	5.5	2.1	4.9	5.5	2.6	4.7	7.7	8.3	4.9	1.4	3.0	
Long Form D7Bclu1 Salivary Protein	AF420272	22	10	21	22	20	16	7	16	45	22	10	10	14	12	17	21	12	16	5	14	332
		6.6	3.0	6.3	6.6	6.0	4.8	2.1	4.8	13.6	6.6	3.0	3.0	4.2	3.6	5.1	6.3	3.6	4.8	1.5	4.2	
Lysosomal Aspartic Protease	M95187	27	7	22	15	24	40	4	26	22	29	9	19	18	11	8	32	22	32	3	17	387
		7.0	1.8	5.7	3.9	6.2	10.3	1.0	6.7	5.7	7.5	2.3	4.9	4.7	2.8	2.1	8.3	5.7	8.3	0.8	4.4	
Lysozyme	AJ290428	10	8	10	5	4	8	1	5	21	19	2	14	2	4	7	9	3	7	5	4	148
		6.8	5.4	6.8	3.4	2.7	5.4	0.7	3.4	14.2	12.8	1.4	9.5	1.4	2.7	4.7	6.1	2.0	4.7	3.4	2.7	
Maltase-Like I	M30442	29	2	48	31	25	37	10	30	37	47	11	35	28	21	31	38	29	43	17	30	579
		5.0	0.3	8.3	5.4	4.3	6.4	1.7	5.2	6.4	8.1	1.9	6.0	4.8	3.6	5.4	6.6	5.0	7.4	2.9	5.2	
Microvilli Membrane Protein	AY050565	19	0	17	14	25	12	4	3	16	28	1	8	5	7	4	7	9	20	2	6	207
		9.2	0.0	8.2	6.8	12.1	5.8	1.9	1.4	7.7	13.5	0.5	3.9	2.4	3.4	1.9	3.4	4.3	9.7	1.0	2.9	
MRGG	U84248	52	25	36	29	17	57	21	33	28	43	11	19	34	22	34	91	49	42	8	12	663
		7.8	3.8	5.4	4.4	2.6	8.6	3.2	5.0	4.2	6.5	1.7	2.9	5.1	3.3	5.1	13.7	7.4	6.3	1.2	1.8	
Mucin-Like Protein	AF125984	12	11	7	8	10	13	8	3	6	9	2	8	21	6	1	19	19	14	4	2	183
		6.6	6.0	3.8	4.4	5.5	7.1	4.4	1.6	3.3	4.9	1.1	4.4	11.5	3.3	0.5	10.4	10.4	7.7	2.2	1.1	
Multivitamin Transporter	AY063742	55	7	23	11	39	48	3	41	24	74	23	15	24	22	21	47	34	38	12	31	592
		9.3	1.2	3.9	1.9	6.6	8.1	0.5	6.9	4.1	12.5	3.9	2.5	4.1	3.7	3.5	7.9	5.7	6.4	2.0	5.2	
Na+/H+ Antiporter	AF187723	81	8	61	67	59	86	32	75	56	100	32	58	55	33	64	110	66	96	12	28	1179
		6.9	0.7	5.2	5.7	5.0	7.3	2.7	6.4	4.7	8.5	2.7	4.9	4.7	2.8	5.4	9.3	5.6	8.1	1.0	2.4	
Neuropeptide F Preproprotein	AF474405	8	1	8	5	5	3	2	2	2	10	1	1	4	6	8	10	8	3	1	2	90
		8.9	1.1	8.9	5.6	5.6	3.3	2.2	2.2	2.2	11.1	1.1	1.1	4.4	6.7	8.9	11.1	8.9	3.3	1.1	2.2	
Nuclear Hormone Receptor	AF165528	50	16	21	25	22	26	18	18	26	46	7	20	46	30	30	57	31	33	1	22	545
		9.2	2.9	3.9	4.6	4.0	4.8	3.3	3.3	4.8	8.4	1.3	3.7	8.4	5.5	5.5	10.5	5.7	6.1	0.2	4.0	
Nuclear Hormone Receptor	AF303224	32	17	18	20	21	38	13	20	17	48	13	20	33	31	28	55	23	28	4	14	493
		6.5	3.4	3.7	4.1	4.3	7.7	2.6	4.1	3.4	9.7	2.6	4.1	6.7	6.3	5.7	11.2	4.7	5.7	0.8	2.8	
Nuclear Receptor 3	AF230281	26	11	29	32	19	22	11	28	25	41	21	17	15	28	26	41	26	29	2	19	468
		5.6	2.4	6.2	6.8	4.1	4.7	2.4	6.0	5.3	8.8	4.5	3.6	3.2	6.0	5.6	8.8	5.6	6.2	0.4	4.1	
Nuclear Receptor FTZ-F1 Protein	AF274870	61	18	40	33	23	74	33	45	33	73	17	36	39	56	25	100	69	43	4	18	840
		7.3	2.1	4.8	3.9	2.7	8.8	3.9	5.4	3.9	8.7	2.0	4.3	4.6	6.7	3.0	11.9	8.2	5.1	0.5	2.1	
Odorant-Binding Protein-Related Protein	AY062131	10	7	10	8	6	6	2	6	16	18	5	5	3	4	7	10	5	7	0	5	140
		7.1	5.0	7.1	5.7	4.3	4.3	1.4	4.3	11.4	12.9	3.6	3.6	2.1	2.9	5.0	7.1	3.6	5.0	0.0	3.6	
Ovarian Ecdysteroidogenic Hormone I	U96112	8	12	3	5	3	20	4	5	7	17	6	5	10	7	3	14	4	9	1	6	149
		5.4	8.1	2.0	3.4	2.0	13.4	2.7	3.4	4.7	11.4	4.0	3.4	6.7	4.7	2.0	9.4	2.7	6.0	0.7	4.0	
Peritrophin	AY050566	65	11	32	29	16	64	35	16	24	85	11	40	53	76	29	91	42	37	5	17	778
		8.4	1.4	4.1	3.7	2.1	8.2	4.5	2.1	3.1	10.9	1.4	5.1	6.8	9.8	3.7	11.7	5.4	4.8	0.6	2.2	
Peroxidase	AF098717	53	15	45	34	33	39	25	30	24	54	15	41	40	29	40	45	49	37	10	25	683
		7.8	2.2	6.6	5.0	4.8	5.7	3.7	4.4	3.5	7.9	2.2	6.0	5.9	4.2	5.9	6.6	7.2	5.4	1.5	3.7	
Phenylalanine Hydroxylase	AY099427	32	8	24	37	29	26	14	22	20	40	6	21	27	21	25	30	18	21	3	23	447
		7.2	1.8	5.4	8.3	6.5	5.8	3.1	4.9	4.5	8.9	1.3	4.7	6.0	4.7	5.6	6.7	4.0	4.7	0.7	5.1	
Phosphoglycerate Kinase	AY064545	25	19	27	23	14	36	12	26	25	43	15	21	25	31	33	38	13	20	2	11	459
		5.4	4.1	5.9	5.0	3.1	7.8	2.6	5.7	5.4	9.4	3.3	4.6	5.4	6.8	7.2	8.3	2.8	4.4	0.4	2.4	
Preproallatostatin	U66841	30	0	29	47	11	33	8	11	52	10	5	43	17	21	32	19	31	10	3	7	419
		7.2	0.0	6.9	11.2	2.6	7.9	1.9	2.6	12.4	2.4	1.2	10.3	4.1	5.0	7.6	4.5	7.4	2.4	0.7	1.7	
Preproleukokinin	U66832	17	5	12	13	11	11	4	8	16	18	4	24	16	9	11	17	5	14	7	6	228
		7.5	2.2	5.3	5.7	4.8	4.8	1.8	3.5	7.0	7.9	1.8	10.5	7.0	3.9	4.8	7.5	2.2	6.1	3.1	2.6	
Probable Transport Protein Sec61 Alpha Subunit	AY064125	36	8	13	19	29	47	6	53	18	56	16	13	19	18	18	30	20	34	5	18	476
		7.6	1.7	2.7	4.0	6.1	9.9	1.3	11.1	3.8	11.8	3.4	2.7	4.0	3.8	3.8	6.3	4.2	7.1	1.1	3.8	

GENES	Accession #	Amino acid single letter code																				
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
Prophenol Oxidase Propo3	AF310673	46	5	54	33	43	37	15	40	25	57	16	34	42	30	56	41	38	43	6	26	687
		6.7	0.7	7.9	4.8	6.3	5.4	2.2	5.8	3.6	8.3	2.3	4.9	6.1	4.4	8.2	6.0	5.5	6.3	0.9	3.8	
Prophenol Oxidase Propo4	AF327409	31	7	42	34	42	43	19	35	17	49	19	51	44	40	60	44	34	43	7	25	686
		4.5	1.0	6.1	5.0	6.1	6.3	2.8	5.1	2.5	7.1	2.8	7.4	6.4	5.8	8.7	6.4	5.0	6.3	1.0	3.6	
Prophenoloxidase Propo1	AF292114	39	6	45	33	39	41	15	36	14	56	17	47	46	35	61	36	43	44	6	25	684
		5.7	0.9	6.6	4.8	5.7	6.0	2.2	5.3	2.0	8.2	2.5	6.9	6.7	5.1	8.9	5.3	6.3	6.4	0.9	3.7	
Putative 16.9 Kda Secreted Protein	AF466612	9	8	11	10	11	6	1	6	18	12	8	9	3	7	1	8	7	8	1	4	148
		6.1	5.4	7.4	6.8	7.4	4.1	0.7	4.1	12.2	8.1	5.4	6.1	2.0	4.7	0.7	5.4	4.7	5.4	0.7	2.7	
Putative 18.2 Kda Secreted Protein	AF466613	12	6	10	11	9	7	2	16	14	11	5	10	7	4	6	7	7	5	2	7	158
		7.6	3.8	6.3	7.0	5.7	4.4	1.3	10.1	8.9	7.0	3.2	6.3	4.4	2.5	3.8	4.4	4.4	3.2	1.3	4.4	
Putative 18.6 Kda Secreted Protein	AF466604	21	0	8	17	4	11	2	11	10	16	3	8	4	9	9	12	4	11	2	6	168
		12.5	0.0	4.8	10.1	2.4	6.5	1.2	6.5	6.0	9.5	1.8	4.8	2.4	5.4	5.4	7.1	2.4	6.5	1.2	3.6	
Putative 30 Kda Allergen-Like Protein	AF466608	16	4	26	23	4	14	1	9	25	10	5	9	8	8	7	20	11	11	0	4	215
		7.4	1.9	12.1	10.7	1.9	6.5	0.5	4.2	11.6	4.7	2.3	4.2	3.7	3.7	3.3	9.3	5.1	5.1	0.0	1.9	
Putative 30.5 Kda Secreted Protein	AF466607	21	7	12	3	6	10	0	15	11	34	11	21	9	29	3	23	23	19	3	15	275
		7.6	2.5	4.4	1.1	2.2	3.6	0.0	5.5	4.0	12.4	4.0	7.6	3.3	10.5	1.1	8.4	8.4	6.9	1.1	5.5	
Putative 34 Kda Secreted Protein	AF466595	21	4	22	26	11	11	3	18	26	30	14	17	7	20	14	18	22	21	3	8	316
		6.6	1.3	7.0	8.2	3.5	3.5	0.9	5.7	8.2	9.5	4.4	5.4	2.2	6.3	4.4	5.7	7.0	6.6	0.9	2.5	
Putative 56.5 Kda Secreted Protein	AF466593	50	6	26	20	19	27	2	31	12	56	11	43	14	28	33	44	36	49	0	23	530
		9.4	1.1	4.9	3.8	3.6	5.1	0.4	5.8	2.3	10.6	2.1	8.1	2.6	5.3	6.2	8.3	6.8	9.2	0.0	4.3	
Putative 7.8 Kda Secreted Protein	AF466588	8	6	3	6	6	9	0	5	10	4	2	4	1	3	3	5	5	6	2	7	95
		8.4	6.3	3.2	6.3	6.3	9.5	0.0	5.3	10.5	4.2	2.1	4.2	1.1	3.2	3.2	5.3	5.3	6.3	2.1	7.4	
Putative 8.3 Kda Secreted Protein	AF466592	9	1	3	2	1	5	7	5	1	5	2	7	5	6	3	19	6	7	2	0	96
		9.4	1.0	3.1	2.1	1.0	5.2	7.3	5.2	1.0	5.2	2.1	7.3	5.2	6.3	3.1	19.8	6.3	7.3	2.1	0.0	
Putative Adenosine Deaminase	AF466610	40	1	31	33	34	27	17	33	40	54	9	23	24	24	20	39	17	34	9	21	530
		7.5	0.2	5.8	6.2	6.4	5.1	3.2	6.2	7.5	10.2	1.7	4.3	4.5	4.5	3.8	7.4	3.2	6.4	1.7	4.0	
Putative C-Type Lectin	AF466606	15	7	9	10	7	8	4	5	6	12	4	7	4	7	8	11	4	13	8	5	154
		9.7	4.5	5.8	6.5	4.5	5.2	2.6	3.2	3.9	7.8	2.6	4.5	2.6	4.5	5.2	7.1	2.6	8.4	5.2	3.2	
Putative Calreticulin	AF466603	25	5	52	51	18	21	10	16	45	20	5	16	22	10	6	14	18	30	10	13	407
		6.1	1.2	12.8	12.5	4.4	5.2	2.5	3.9	11.1	4.9	1.2	3.9	5.4	2.5	1.5	3.4	4.4	7.4	2.5	3.2	
Putative Carboxylesterase	AF466590	44	8	34	28	30	61	7	43	40	55	10	33	36	24	28	49	35	51	6	14	636
		6.9	1.3	5.3	4.4	4.7	9.6	1.1	6.8	6.3	8.6	1.6	5.2	5.7	3.8	4.4	7.7	5.5	8.0	0.9	2.2	
Putative Chymotrypsin-Like Serine Protease	AF466611	21	8	15	12	13	26	7	16	5	29	5	11	12	11	28	15	23	19	8	4	288
		7.3	2.8	5.2	4.2	4.5	9.0	2.4	5.6	1.7	10.1	1.7	3.8	4.2	3.8	9.7	5.2	8.0	6.6	2.8	1.4	
Putative Lysozyme	AF466591	13	9	6	4	8	10	3	7	15	9	3	13	0	3	2	13	10	7	3	6	144
		9.0	6.3	4.2	2.8	5.6	6.9	2.1	4.9	10.4	6.3	2.1	9.0	0.0	2.1	1.4	9.0	6.9	4.9	2.1	4.2	
Putative mRNA Binding Protein	AY064109	73	27	77	51	26	99	40	45	35	69	14	51	54	46	61	118	58	69	20	27	1060
		6.9	2.5	7.3	4.8	2.5	9.3	3.8	4.2	3.3	6.5	1.3	4.8	5.1	4.3	5.8	11.1	5.5	6.5	1.9	2.5	
Putative Prophenoloxidase Activating Protein	AF466614	20	17	27	20	16	40	11	29	15	33	3	15	29	13	22	35	21	28	5	14	413
		4.8	4.1	6.5	4.8	3.9	9.7	2.7	7.0	3.6	8.0	0.7	3.6	7.0	3.1	5.3	8.5	5.1	6.8	1.2	3.4	
Putative Protein G12	AY038042	19	1	17	13	25	12	4	3	17	28	1	9	5	7	4	7	8	20	2	5	207
		9.2	0.5	8.2	6.3	12.1	5.8	1.9	1.4	8.2	13.5	0.5	4.3	2.4	3.4	1.9	3.4	3.9	9.7	1.0	2.4	
Putative Secreted Protein	AF466586	13	12	14	12	10	13	9	11	26	16	9	13	6	8	11	27	14	10	6	15	255
		5.1	4.7	5.5	4.7	3.9	5.1	3.5	4.3	10.2	6.3	3.5	5.1	2.4	3.1	4.3	10.6	5.5	3.9	2.4	5.9	
Putative Secreted Protein	AF466601	24	14	7	6	11	20	5	18	10	15	6	22	9	15	14	17	20	14	2	10	259
		9.3	5.4	2.7	2.3	4.2	7.7	1.9	6.9	3.9	5.8	2.3	8.5	3.5	5.8	5.4	6.6	7.7	5.4	0.8	3.9	
Putative Secreted Protein	AF466589	26	13	9	12	13	17	10	12	17	11	10	15	11	12	16	16	10	10	3	12	255
		10.2	5.1	3.5	4.7	5.1	6.7	3.9	4.7	6.7	4.3	3.9	5.9	4.3	4.7	6.3	6.3	3.9	3.9	1.2	4.7	
Putative Secreted Protein	AF466605	21	13	8	9	13	15	13	16	25	17	13	18	7	7	11	19	15	8	4	9	261
		8.0	5.0	3.1	3.4	5.0	5.7	5.0	6.1	9.6	6.5	5.0	6.9	2.7	2.7	4.2	7.3	5.7	3.1	1.5	3.4	
Putative Secreted Protein	AF466596	18	5	14	24	15	25	5	15	13	18	11	15	6	15	15	16	20	16	10	16	292
		6.2	1.7	4.8	8.2	5.1	8.6	1.7	5.1	4.5	6.2	3.8	5.1	2.1	5.1	5.1	5.5	6.8	5.5	3.4	5.5	
Putative Secreted Protein	AF466594	20	7	16	13	20	39	5	17	20	24	6	30	21	16	23	27	15	22	16	14	371
		5.4	1.9	4.3	3.5	5.4	10.5	1.3	4.6	5.4	6.5	1.6	8.1	5.7	4.3	6.2	7.3	4.0	5.9	4.3	3.8	
Putative Secreted Protein	AF466597	38	8	42	41	17	22	11	31	53	60	22	39	15	25	21	37	32	38	5	23	580
		6.6	1.4	7.2	7.1	2.9	3.8	1.9	5.3	9.1	10.3	3.8	6.7	2.6	4.3	3.6	6.4	5.5	6.6	0.9	4.0	
Putative Secreted Protein	AF466598	32	10	37	46	20	23	13	32	57	59	16	38	14	18	18	38	32	40	3	20	566
		5.7	1.8	6.5	8.1	3.5	4.1	2.3	5.7	10.1	10.4	2.8	6.7	2.5	3.2	3.2	6.7	5.7	7.1	0.5	3.5	
Putative Secreted Protein	AF466602	9	9	3	2	2	8	0	4	5	6	1	1	0	3	5	9	5	6	0	7	85
		10.6	10.6	3.5	2.4	2.4	9.4	0.0	4.7	5.9	7.1	1.2	1.2	0.0	3.5	5.9	10.6	5.9	7.1	0.0	8.2	
Putative Serine Protease	AF466600	21	8	15	8	6	25	8	23	8	21	6	13	14	8	19	27	18	32	6	10	296
		7.1	2.7	5.1	2.7	2.0	8.4	2.7	7.8	2.7	7.1	2.0	4.4	4.7	2.7	6.4	9.1	6.1	10.8	2.0	3.4	
Pyruvate Carboxylase	L36530	24	8	13	10	8	25	4	14	7	17	4	11	11	9	12	24	15	27	4	7	254
		9.4	3.1	5.1	3.9	3.1	9.8	1.6	5.5	2.8	6.7	1.6	4.3	4.3	3.5	4.7	9.4	5.9	10.6	1.6	2.8	
Relish	AF498105	61	19	51	86	36	37	39	47	53	105	30	66	42	47	43	91	45	62	6	29	995
		6.1	1.9	5.1	8.6	3.6	3.7	3.9	4.7	5.3	10.6	3.0	6.6	4.2	4.7	4.3	9.1	4.5	6.2	0.6	2.9	
Ribonucleotide Reductase 1	AF411296	65	18	48	55	28	48	22	58	61	66	35	53	29	27	37	55	47	41	11	40	844
		7.7	2.1	5.7	6.5	3.3	5.7	2.6	6.9	7.2	7.8	4.1	6.3	3.4	3.2	4.4	6.5	5.6	4.9	1.3	4.7	
Ribonucleotide Reductase 2	AF411297	28	4	23	40	29	16	5	27	33	36	12	20	11	15	16	27	13	24	7	11	397
		7.1	1.0	5.8	10.1	7.3	4.0	1.3	6.8	8.3	9.1	3.0	5.0	2.8	3.8	4.0	6.8	3.3	6.0	1.8	2.8	
Ribosomal Protein L17A	AY064121	16	2	5	4	5	19	1	9	12	8	4	8	7	1	13	6	3	14	1	2	140
		11.4	1.4	3.6	2.9	3.6	13.6	0.7	6.4	8.6	5.7	2.9	5.7	5.0	0.7	9.3	4.3	2.1	10.0	0.7	1.4	
Ribosomal Protein L31	AF324863	7	1	6	10	3	3	2	6	14	8	2	9	5	1	16	5	11	11	1	3	124
		5.6	0.8	4.8	8.1	2.4	2.4	1.6	4.8	11.3	6.5	1.6	7.3	4.0	0.8	12.9	4.0	8.9	8.9	0.8	2.4	
Ribosomal Protein S11	AY133344	8	3	6	5	9	11	6	10	20	11	3	8	4	5	16	5	7	12	0	3	152
		5.3	2.0	3.9	3.3	5.9	7.2	3.9	6.6	13.2	7.2	2.0	5.3	2.6	3.3	10.5	3.3	4.6	7.9	0.0	2.0	
Ribosomal Protein S6	AF154067	38	2	17	28	6	19	6	11	68	21	6	7	14	9	36	14	16	22	1	5	346
		11.0	0.6	4.9	8.1	1.7	5.5	1.7	3.2	19.7	6.1	1.7	2.0	4.0	2.6	10.4	4.0	4.6	6.4	0.3	1.4	
Ribosomal Protein S24	AY064700	8	2	5	4	9	10	3	4	22	9	6	3	4	3	14	3	11	8	0	2	130

GENES	Accession #	Amino acid single letter code																			
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Total
		6.2	1.5	3.8	3.1	6.9	7.7	2.3	3.1	16.9	6.9	4.6	2.3	3.1	2.3	10.8	2.3	8.5	6.2	0.0	1.5	
30 Kda Salivary Gland Allergen	AF001927	19	5	29	42	5	29	3	8	25	12	4	6	7	5	6	22	8	14	0	4	253
		7.5	2.0	11.5	16.6	2.0	11.5	1.2	3.2	9.9	4.7	1.6	2.4	2.8	2.0	2.4	8.7	3.2	5.5	0.0	1.6	
Salivary Purine Nucleosidase	AY038045	26	5	23	22	12	23	5	23	15	33	14	27	16	7	18	9	13	32	5	10	338
		7.7	1.5	6.8	6.5	3.6	6.8	1.5	6.8	4.4	9.8	4.1	8.0	4.7	2.1	5.3	2.7	3.8	9.5	1.5	3.0	
Salivary Vasodilatory Protein Precursor	AF108099	6	0	10	6	4	5	0	3	7	6	4	2	4	3	2	7	5	7	0	4	85
		7.1	0.0	11.8	7.1	4.7	5.9	0.0	3.5	8.2	7.1	4.7	2.4	4.7	3.5	2.4	8.2	5.9	8.2	0.0	4.7	
Serine Protease Inhibitor	AF466599	29	4	27	29	24	17	3	22	33	42	14	21	10	18	14	37	20	38	3	13	418
		6.9	1.0	6.5	6.9	5.7	4.1	0.7	5.3	7.9	10.0	3.3	5.0	2.4	4.3	3.3	8.9	4.8	9.1	0.7	3.1	
Short Form D7Cclu23 Salivary Protein	AF420273	5	8	8	11	15	7	2	5	19	12	7	5	3	2	8	10	6	8	2	6	149
		3.4	5.4	5.4	7.4	10.1	4.7	1.3	3.4	12.8	8.1	4.7	3.4	2.0	1.3	5.4	6.7	4.0	5.4	1.3	4.0	
Short-Chain Dehydrogenease/Reductase	AY064123	26	4	11	11	9	22	6	18	17	23	5	12	9	8	10	16	20	24	0	3	254
		10.2	1.6	4.3	4.3	3.5	8.7	2.4	7.1	6.7	9.1	2.0	4.7	3.5	3.1	3.9	6.3	7.9	9.4	0.0	1.2	
Sialokinin I Preproprotein	AF108101	6	0	9	7	2	5	0	3	5	8	4	4	3	3	1	7	7	7	0	4	85
		7.1	0.0	10.6	8.2	2.4	5.9	0.0	3.5	5.9	9.4	4.7	4.7	3.5	3.5	1.2	8.2	8.2	8.2	0.0	4.7	
Sterol Carrier Protein 2 Variant 1	AF539987	11	0	8	9	5	5	0	8	12	9	7	4	3	6	4	4	4	10	1	1	111
		9.9	0.0	7.2	8.1	4.5	4.5	0.0	7.2	10.8	8.1	6.3	3.6	2.7	5.4	3.6	3.6	3.6	9.0	0.9	0.9	
Sterol Carrier Protein 2 Variant 2	AF539988	12	0	7	11	5	5	1	7	11	9	7	4	3	4	3	5	4	10	1	1	110
		10.9	0.0	6.4	10.0	4.5	4.5	0.9	6.4	10.0	8.2	6.4	3.6	2.7	3.6	2.7	4.5	3.6	9.1	0.9	0.9	
Takeout	AF458100	18	2	14	8	14	20	4	17	22	25	10	18	10	11	4	9	14	20	2	6	248
		7.3	0.8	5.6	3.2	5.6	8.1	1.6	6.9	8.9	10.1	4.0	7.3	4.0	4.4	1.6	3.6	5.6	8.1	0.8	2.4	
Thioredoxin	AF394233	10	3	9	8	7	4	1	7	14	8	3	2	3	3	0	5	2	11	1	5	106
		9.4	2.8	8.5	7.5	6.6	3.8	0.9	6.6	13.2	7.5	2.8	1.9	2.8	2.8	0.0	4.7	1.9	10.4	0.9	4.7	
Toll-Related Protein	AY124195	36	20	59	71	50	29	25	61	51	156	14	81	42	47	40	84	52	55	8	26	1007
		3.6	2.0	5.9	7.1	5.0	2.9	2.5	6.1	5.1	15.5	1.4	8.0	4.2	4.7	4.0	8.3	5.2	5.5	0.8	2.6	
Toll1	AY124196	40	21	57	70	50	59	38	68	52	155	18	86	51	49	46	77	71	68	14	34	1124
		3.6	1.9	5.1	6.2	4.4	5.2	3.4	6.0	4.6	13.8	1.6	7.7	4.5	4.4	4.1	6.9	6.3	6.0	1.2	3.0	
Toll1B	AY124197	52	23	62	60	53	43	33	83	40	147	20	73	45	52	61	91	40	58	16	24	1076
		4.8	2.1	5.8	5.6	4.9	4.0	3.1	7.7	3.7	13.7	1.9	6.8	4.2	4.8	5.7	8.5	3.7	5.4	1.5	2.2	
Transfferin Precursor	AF019117	59	20	47	37	26	41	8	25	49	53	14	31	28	29	37	31	20	52	3	23	633
		9.3	3.2	7.4	5.8	4.1	6.5	1.3	3.9	7.7	8.4	2.2	4.9	4.4	4.6	5.8	4.9	3.2	8.2	0.5	3.6	
Tryptophan Oxygenase	AF325458	20	3	20	31	20	21	12	22	19	47	16	14	15	21	26	31	16	21	5	13	393
		5.1	0.8	5.1	7.9	5.1	5.3	3.1	5.6	4.8	12.0	4.1	3.6	3.8	5.3	6.6	7.9	4.1	5.3	1.3	3.3	
Ubiquitin-52-Amino-Acid Fusion Protein	AF418984	5	5	6	7	2	7	3	11	17	14	2	6	6	7	10	4	9	4	0	3	128
		3.9	3.9	4.7	5.5	1.6	5.5	2.3	8.6	13.3	10.9	1.6	4.7	4.7	5.5	7.8	3.1	7.0	3.1	0.0	2.3	
Ultraspiracle	S:13554	12	2	5	10	5	6	1	3	9	5	4	5	9	3	17	14	4	9	2	3	128
		9.4	1.6	3.9	7.8	3.9	4.7	0.8	2.3	7.0	3.9	3.1	3.9	7.0	2.3	13.3	10.9	3.1	7.0	1.6	2.3	
V-Atpase 110 Kda Integral Membrane Subunit	AF173554	43	10	29	57	55	48	18	65	44	86	33	38	31	30	31	53	40	49	15	29	804
		5.3	1.2	3.6	7.1	6.8	6.0	2.2	8.1	5.5	10.7	4.1	4.7	3.9	3.7	3.9	6.6	5.0	6.1	1.9	3.6	
Vacuolar ATPase A Subunit	AF008922	39	6	35	45	22	50	7	37	33	55	20	19	31	20	36	47	32	47	6	28	615
		6.3	1.0	5.7	7.3	3.6	8.1	1.1	6.0	5.4	8.9	3.3	3.1	5.0	3.3	5.9	7.6	5.2	7.6	1.0	4.6	
Vacuolar ATPase B Subunit	AF092934	38	5	27	36	22	34	10	40	21	41	14	22	27	21	29	30	26	36	2	15	496
		7.7	1.0	5.4	7.3	4.4	6.9	2.0	8.1	4.2	8.3	2.8	4.4	5.4	4.2	5.8	6.0	5.2	7.3	0.4	3.0	
Vacuolar ATPase C Subunit	AF008924	24	0	2	4	7	23	1	19	6	17	7	0	7	2	3	7	6	14	0	8	157
		15.3	0.0	1.3	2.5	4.5	14.6	0.6	12.1	3.8	10.8	4.5	0.0	4.5	1.3	1.9	4.5	3.8	8.9	0.0	5.1	
Vitelline Membrane 15a1	U91680	20	3	3	3	4	6	13	3	3	6	2	2	21	1	1	3	1	8	0	5	108
		18.5	2.8	2.8	2.8	3.7	5.6	12.0	2.8	2.8	5.6	1.9	1.9	19.4	0.9	0.9	2.8	0.9	7.4	0.0	4.6	
Vitelline Membrane 15a2	U91681	17	3	1	0	1	3	10	3	4	6	1	2	28	0	0	1	2	10	0	5	97
		17.5	3.1	1.0	0.0	1.0	3.1	10.3	3.1	4.1	6.2	1.0	2.1	28.9	0.0	0.0	1.0	2.1	10.3	0.0	5.2	
Vitelline Membrane 15a3	U91682	21	3	2	3	3	6	16	2	3	5	2	3	32	1	1	4	1	8	0	6	122
		17.2	2.5	1.6	2.5	2.5	4.9	13.1	1.6	2.5	4.1	1.6	2.5	26.2	0.8	0.8	3.3	0.8	6.6	0.0	4.9	
Vitellogenic Carboxypeptidase	L46594	26	4	27	30	30	38	17	22	29	42	15	28	24	17	19	33	16	24	6	24	471
		5.5	0.8	5.7	6.4	6.4	8.1	3.6	4.7	6.2	8.9	3.2	5.9	5.1	3.6	4.0	7.0	3.4	5.1	1.3	5.1	
Vitellogenic Cathepsin-B Like Protease	AF127592	31	16	16	21	16	43	11	12	17	26	6	20	21	17	25	29	11	17	12	19	386
		8.0	4.1	4.1	5.4	4.1	11.1	2.8	3.1	4.4	6.7	1.6	5.2	5.4	4.4	6.5	7.5	2.8	4.4	3.1	4.9	
Vitellogenin Convertase	L46373	101	25	88	92	87	69	23	95	93	153	64	70	70	65	75	108	82	113	33	58	1564
		6.5	1.6	5.6	5.9	5.6	4.4	1.5	6.1	5.9	9.8	4.1	4.5	4.5	4.2	4.8	6.9	5.2	7.2	2.1	3.7	
Vitellogenin Receptor	L77800	86	131	163	106	62	139	64	107	107	134	40	91	77	57	80	128	99	106	19	51	1847
		4.7	7.1	8.8	5.7	3.4	7.5	3.5	5.8	5.8	7.3	2.2	4.9	4.2	3.1	4.3	6.9	5.4	5.7	1.0	2.8	
Vitellogenin Vg-A1	L41842	155	20	116	123	174	102	61	69	148	114	39	116	104	102	79	213	80	104	12	217	2148
		7.2	0.9	5.4	5.7	8.1	4.7	2.8	3.2	6.9	5.3	1.8	5.4	4.8	4.7	3.7	9.9	3.7	4.8	0.6	10.1	
Vitellogenin Vg-B	AY380797	157	20	116	126	171	95	59	56	155	111	43	111	101	109	77	220	83	109	10	213	2142
		7.3	0.9	5.4	5.9	8.0	4.4	2.8	2.6	7.2	5.2	2.0	5.2	4.7	5.1	3.6	10.3	3.9	5.1	0.5	9.9	
Vitellogenin Vg-C	AY373377	131	24	118	120	151	93	52	68	140	103	32	122	97	127	86	230	86	111	12	184	2087
		6.3	1.1	5.7	5.7	7.2	4.5	2.5	3.3	6.7	4.9	1.5	5.8	4.6	6.1	4.1	11.0	4.1	5.3	0.6	8.8	

A substitutional comparison of Vg-A1 and Vg-B of *Ae. aegypti* showed that among 227 amino acid changes, the majority (84.1%) resulted in conserved/moderately conserved substitutions (Table 7). Moderately radical and radical amino acid substitutions (15.9%) were distributed throughout the sequences. Nonsynonymous amino acid substitutions were absent from a 136 amino acid sequence of Vg-A1 and Vg-B (residues 1466~1599 in Figure 2) suggesting the action of selective constraints in this region. A substitutional comparison of Vg-A1 and Vg-B to Vg-C also showed that most amino acid changes were due to conserved and moderately conserved changes in amino acid properties (Table 7). However, the number of amino acid substitutions

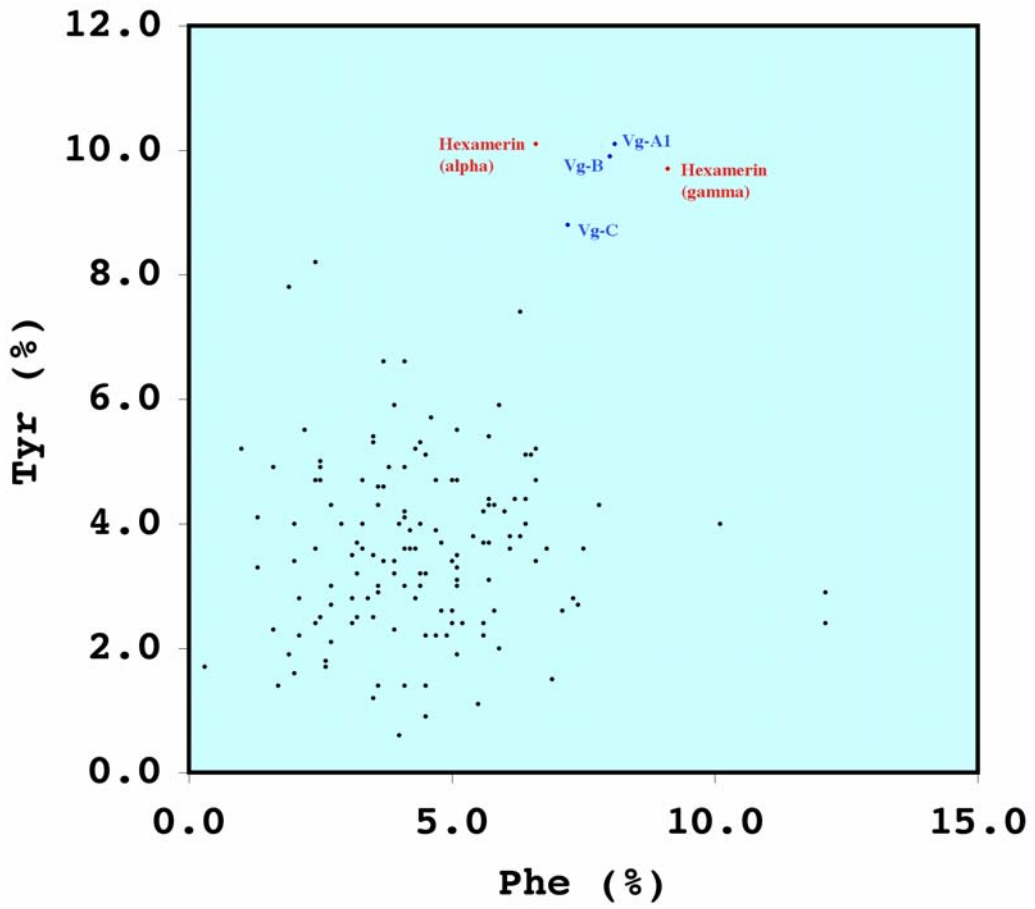


Figure 3. Distribution of the compositional percentage of phenylalanine and tyrosine aromatic amino acid residues of *Aedes aegypti* proteins. The location of the vitellogenins and hexamerin proteins of *Ae. aegypti* are shown (Gordadze et al. 1999).

increased dramatically from 226 amino acid changes in Vg-A1 and Vg-B (Figure 4a) to an average of 725 changes in Vg-C, indicating that Vg-C is significantly different from Vg-A1 and Vg-B (Figure 4b, c). Vg-C showed radical and moderately radical amino acid substitutions distributed throughout the protein, except in the least conserved region (1041~1101) where a large number of these amino acid changes were present, suggesting that this region is less constrained (Figure 4b, c).

The extent of nonsynonymous substitutions was also examined between the Vg-C orthologs of different mosquito species. As shown in Table 7, most nonsynonymous substitutions resulted from conserved and moderately conserved substitutions (average of 43.6 and 41.3%, respectively). Only an average of 2.7% of changes were due to radical amino acid substitutions, suggesting that dramatic changes in amino acid properties in most positions could be detrimental to the function of the vitellogenin proteins. An analysis of intragenic positions of nonsynonymous substitutions shows that most radical and moderately radical changes appear to be randomly distributed.

Synonymous codon usage

Synonymous codon usage bias was examined in vitellogenin genes from insects and other organisms (Table 8) by determining the effective number of codons (ENC). Vitellogenin genes of mosquitoes have a codon bias that ranged from ENC 32.3 (*An. albimanus)* to 50.7 (*Tx. amboinensis)*. Other insect vitellogenin and yolk protein genes also have a wide range of codon usage bias, varying from *D. melanogaster* yolk protein 1 (ENC: 32.6) to *S. cynthia ricini* (ENC: 56.4), with varying GC content at the third codon position. Vitellogenin genes from animals other

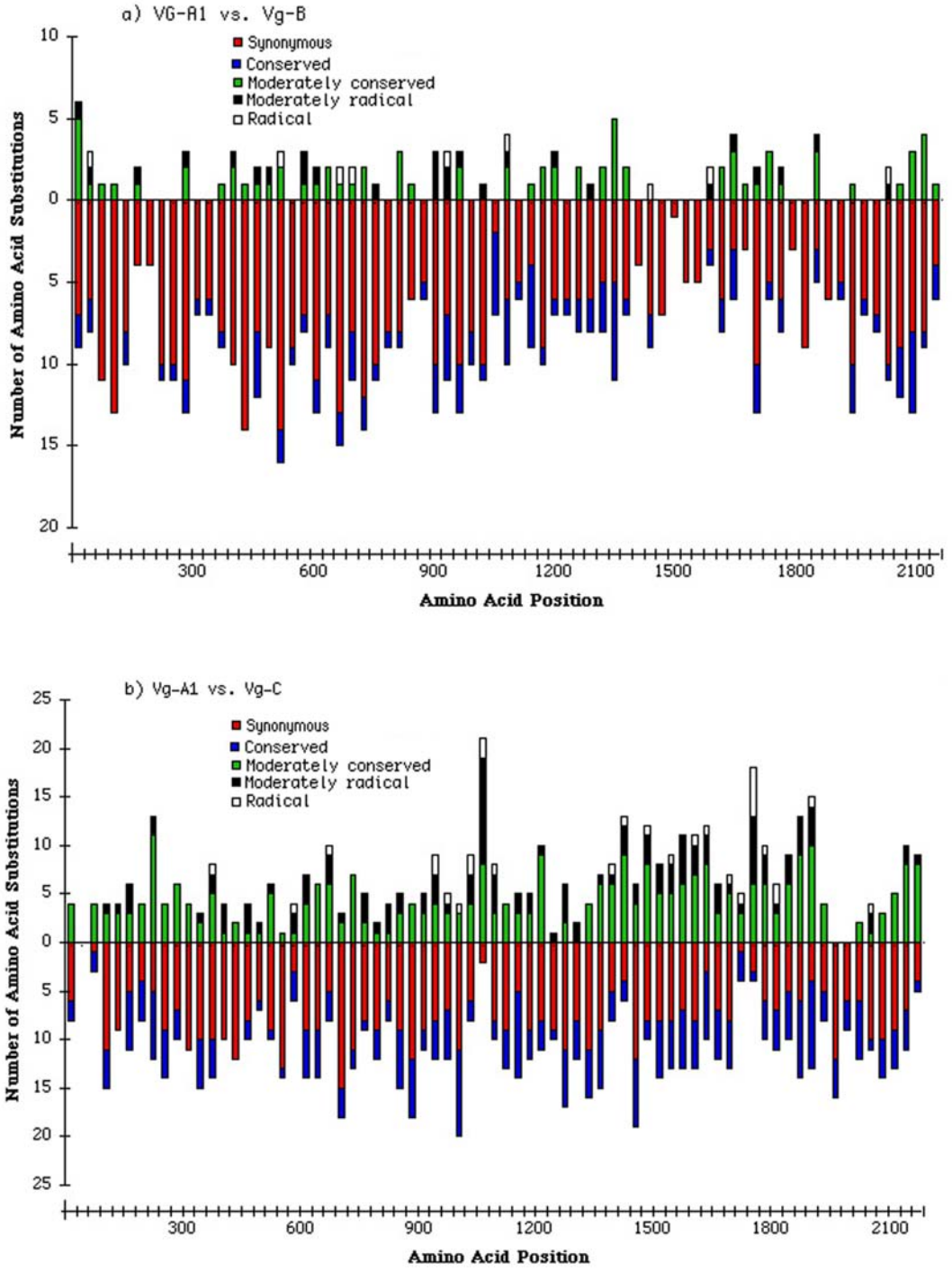


Figure 4 (Part 1). Degree of nonsynonymous substitutions in pairwise comparison of the *Aedes aegypti* vitellogenin genes; (a) Vg-A1 and Vg-B, (b) Vg-A1 and Vg-B, and (c) Vg-B and Vg-C. The meaning of the bars is indicated above the figure.

than insects also show varying degrees of codon usage bias. The intragenic position of synonymous codon usage throughout the entire coding region of the vitellogenin genes was examined. A bias for codon usage was observed in the region encoding the signal peptide of *Ae. aegypti* (Figure 5), and all of the other mosquito vitellogenin genes. In this region, which includes 11–15 codons, excluding the first methionine, a high number of rarely used synonymous codons were present. None of the other, non-mosquito, vitellogenin genes shown in Table 8 had a cluster of rare

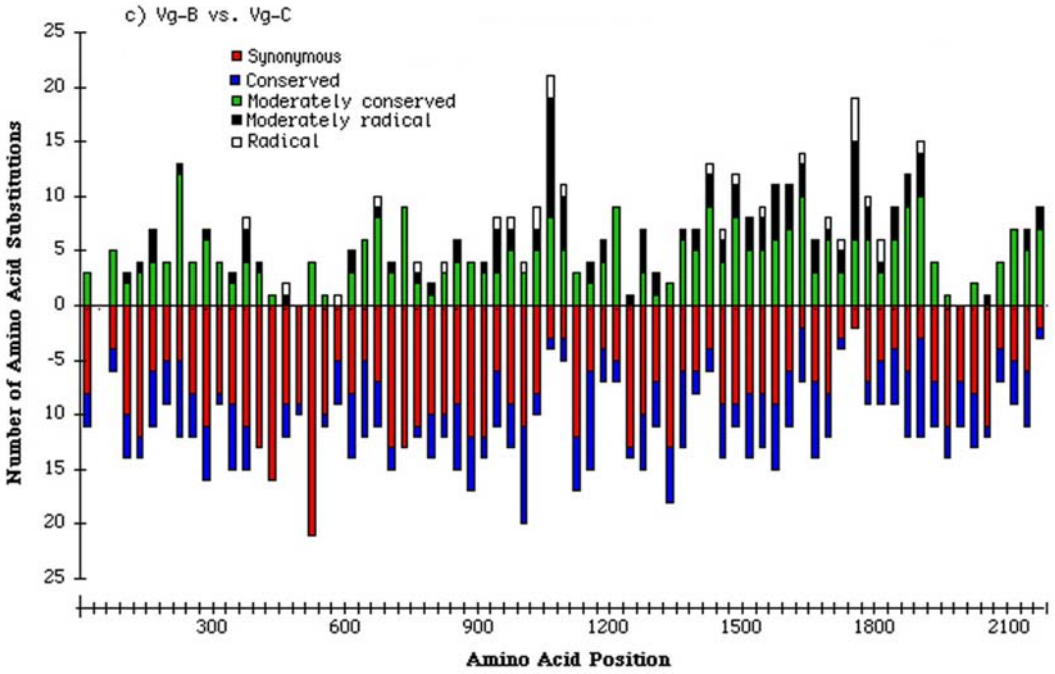


Figure 4 (Part 2). Degree of nonsynonymous substitutions in pairwise comparison of the *Aedes aegypti* vitellogenin genes; (a) Vg-A1 and Vg-B, (b) Vg-A1 and Vg-B, and (c) Vg-B and Vg-C. The meaning of the bars is indicated above the figure.

synonymous codons in the signal peptide region, except for the N-terminal region of *C. elegans* vit-5, which has a few rare synonymous codons (Speith et al. 1985).

By way of comparison, the complete sequences of 150 *Ae. aegypti* genes were analyzed for synonymous codon usage bias (Table 9). The vitellogenin genes showed a pattern of codon usage that is among the most highly biased of all known complete *Ae. aegypti* genes based on ENC. A positive correlation (r^2 = 0.5235) was found between GC content and the degree of synonymous codon usage bias measured by ENC

Table 7. Degree of nonsynonymous substitutions between mosquito vitellogenin genes.

Mosquito Vg genes compared			Nonsynonymous substitutions					Synonymous substitutions	L
			Conserved	Moderately conserved	Moderately radical	Radical	Total		
Comparisons between *Ae. aegypti* genes									
Ae. aegypti Vg-A1	v.s.	*Ae. aegypti* Vg-B	115 (50.9%)	75 (33.2%)	27 (11.9%)	9 (4.0%)	226	527	2128
Ae. aegypti Vg-A1	v.s.	*Ae. aegypti* Vg-C	268 (36.8%)	300 (41.2%)	131 (18.0%)	30 (4.1%)	729	551	2058
Ae. aegypti Vg-B	v.s.	*Ae. aegypti* Vg-C	265 (36.7%)	308 (42.7%)	120 (16.6%)	29 (4.0%)	722	573	2062
Comparisons between mosquito Vg-C genes									
Ae. aegypti	v.s.	*Ae. atropalpus*	137 (50.7%)	102 (37.8%)	19 (7.0%)	12 (4.4%)	270	639	1816
Ae. aegypti	v.s.	*Cx. quinquefasciatus 1*	279 (42.9%)	282 (43.4%)	71 (10.9%)	18 (2.8%)	650	591	2076
Ae. aegypti	v.s.	*Cx. quinquefasciatus 2*	346 (39.1%)	377 (42.6%)	138 (15.6%)	25 (2.8%)	886	514	2009
Ae. aegypti	v.s.	*Tx. amboinensis*	312 (44.7%)	284 (40.7%)	87 (12.5%)	15 (2.1%)	698	631	2018
Ae. aegypti	v.s.	*An. albimanus*	309 (43.3%)	293 (41.9%)	89 (12.5%)	23 (3.2%)	714	546	2021
Ae. atropalpus	v.s.	*Cx. quinquefasciatus 1*	262 (47.6%)	222 (40.4%)	50 (9.1%)	16 (2.9%)	550	597	1818
Ae. atropalpus	v.s.	*Cx. quinquefasciatus 2*	307 (38.9%)	342 (43.3%)	118 (15.0%)	22 (2.8%)	789	510	1749
Ae. atropalpus	v.s.	*Tx. amboinensis*	269 (45.4%)	243 (41.0%)	66 (11.1%)	15 (2.5%)	593	581	1771
Ae. atropalpus	v.s.	*An. albimanus*	275 (44.8%)	251 (40.9%)	72 (11.7%)	16 (2.6%)	614	576	1783
Cx. quinquefasciatus 1	v.s.	*Cx. quinquefasciatus 2*	319 (37.3%)	385 (45.0%)	132 (15.4%)	19 (2.2%)	855	469	2024
Cx. quinquefasciatus 1	v.s.	*Tx. amboinensis*	329 (46.7%)	278 (39.4%)	84 (11.9%)	14 (2.0%)	705	627	2029
Cx. quinquefasciatus 1	v.s.	*An. albimanus*	326 (44.7%)	295 (40.4%)	90 (12.3%)	19 (2.6%)	730	457	2027
Cx. quinquefasciatus 2	v.s.	*Tx. amboinensis*	359 (40.7%)	371 (42.1%)	133 (15.1%)	19 (2.2%)	882	572	1958
Cx. quinquefasciatus 2	v.s.	*An. albimanus*	348 (41.7%)	340 (40.8%)	125 (15.0%)	21 (2.5%)	834	459	1952
Tx. amboinensis	v.s.	*An. albimanus*	350 (45.8%)	306 (40.0%)	89 (11.6%)	20 (2.6%)	765	603	1995

Nonsynonymous amino acid substitutions were divided into 4 categories based on Grantham (1974); conserved (score≤50), moderately conserved (50<scores≤100), moderately radical (100<score≤150), and radical (score>150).
L indicates the number of amino acid compared.

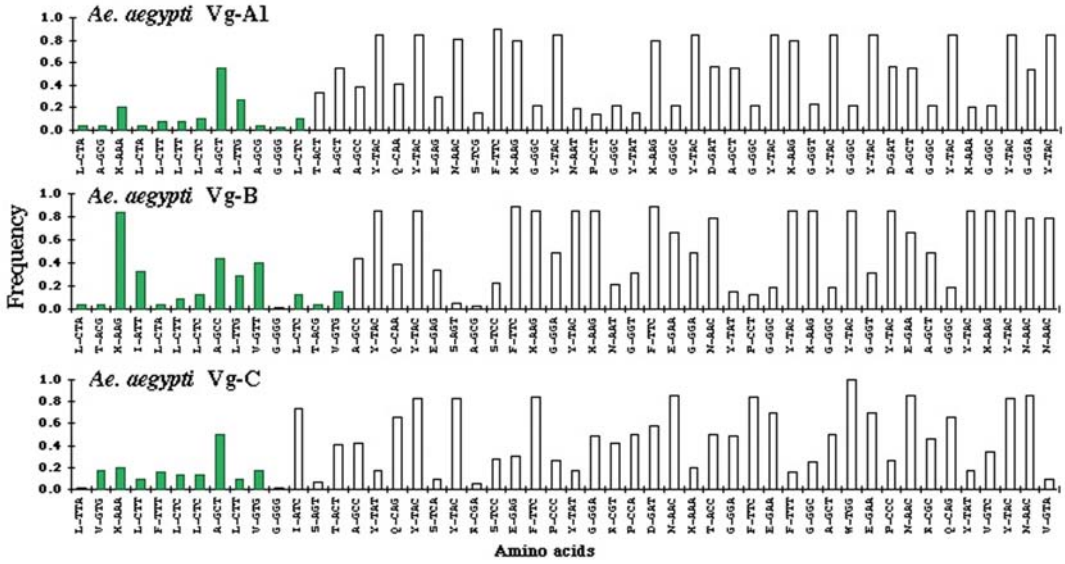


Figure 5. Observed frequency of the first 50 codons in the *Aedes aegypti* vitellogenin genes, excluding the first initiator methionine. Green bars show the codons of the signal peptide, the first 11–14 amino acids, excluding the first methionine. Short bars indicate rare codons.

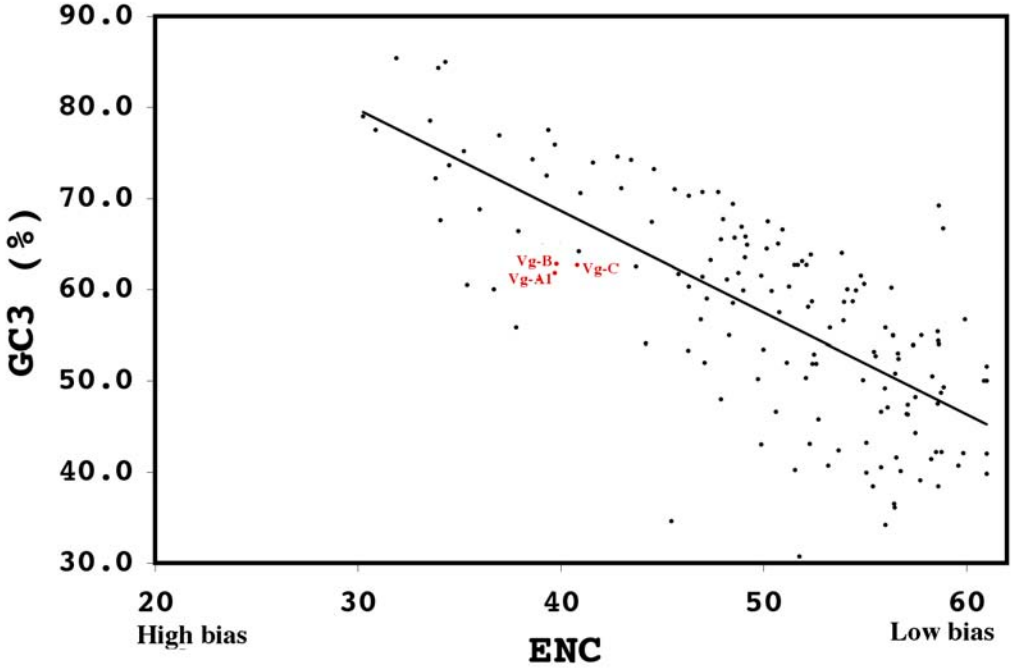


Figure 6. Correlation of synonymous codon usage in 150 *Aedes aegypti* genes. ENC = effective number of codons, GC3= GC content in the third position. Shown in red are the vitellogenin genes. $r^2 = 0.5235$.

Table 8. Effective number of synonymous codons of vitellogenin genes of insects and other organisms. Ordered by ENC value.

Organisms	Order	Family	# Accession	ENC	GC3s	GC%	# AA
Anopheles albimanus Vg-C	Diptera	Culicidae	AY691327	32.3	83.5	56.8	2029
Caenorhabditis elegans vit-5	Nematoda	Rhabditidae	X03044	32.5	57.2	49.5	1651
Drosophila melanogaster YP1	Diptera	Drosophilidae	V00248	32.6	79.7	59.4	439
Drosophila melanogaster YP3	Diptera	Drosophilidae	M15898	33.2	80.4	60.3	420
Drosophila melanogaster YP2	Diptera	Drosophilidae	AY061042	33.4	78.8	59.6	442
Athalia roase	Hymenoptera	Tenthredinidae	AB007850	36.2	75.4	57.7	1872
Anopheles gambiae	Diptera	Culicidae	AF281078	36.2	79.2	55.2	2051
Culex quinquefasciatus Vg-C1	Diptera	Culicidae	AY691324	36.5	71.0	53.6	2111
Aedes aegypti Vg-A1	Diptera	Culicidae	L4182	39.7	61.8	48.7	2148
Aedes aegypti Vg-B	Diptera	Culicidae	AY380797	39.8	62.8	49.1	2142
Pimpla nipponica	Hymenoptera	Ichneumonidae	AF026789	40.0	80.0	57.6	1807
Culex quinquefasciatus Vg-C2	Diptera	Culicidae	AY691325	40.7	66.5	53.6	2099
Ichthyomyzon unicuspus	Vertebrata	Petromyzontidae	M88749	40.7	79.5	59.5	1823
Aedes aegypti Vg-C	Diptera	Culicidae	AY373377	40.8	62.7	49.3	2087
Ochlerotatus atropalpus Vg-B	Diptera	Culicidae	AY691321	45.4	53.9	45.8	2108
Ochlerotatus atropalpus Vg-C	Diptera	Culicidae	AY691322	46.1	52.6	45.4	1825
Anthonomus grandis	Coleoptera	Curculionidae	M72980	46.7	30.1	38.5	1790
Fundulus heteroclitus VGT1	Vertebrata	Fundulidae	U07055	48.4	53.0	48.6	1704
Apis mellifera	Hymenoptera	Apidae	AJ517411	49.0	69.3	51.0	1770
Graptopsaltria nigrofuscata	Hemiptera	Cicadidae	AB026848	49.1	30.3	39.4	1987
Riprotus clavatus	Hemiptera	Alydidae	U97277	49.5	39.4	43.7	1876
Toxorhynchites amboinensis Vg-C	Diptera	Culicidae	AY691326	50.7	57.6	48.7	2032
Gallus gallus	Vertebrata	Phasianidae	D89547	50.9	40.3	44.8	1912
Plautia stali Vg-2	Hemiptera	Pentatomidae	AB033499	50.9	38.7	42.7	1856
Leucophaea maderae	Dictyoptera	Blaberidae	AB052640	51.3	40.4	42.6	1913
Fundulus heteroclitus VTG2	Vertebrata	Fundulidae	U70826	51.5	56.2	50.4	1687
Blattella germanica	Dictyoptera	Blattellidae	AJ005115	51.5	38.0	40.5	1862
Plautia stali Vg-1	Hemiptera	Pentatomidae	AB033498	52.1	43.5	43.9	1907
Plautia stali Vg-3	Hemiptera	Pentatomidae	AB033500	53.9	44.5	45.7	1903
Lymantria dispar	Lepidoptera	Lymantriidae	U60186	54.2	34.7	40.9	1747
Periplaneta americana Vg-1	Dictyoptera	Blattidae	AB034804	54.7	41.6	45.0	1896
Antheraea yamamai	Lepidoptera	Saturniidae	AB055843	55.0	40.2	43.1	1778
Periplaneta americana Vg-2	Dictyoptera	Blattidae	AB047401	55.2	45.9	44.7	1876
Antheraea pernyi	Lepidoptera	Saturniidae	AB049631	55.8	40.6	43.4	1778
Bombyx mori	Lepidoptera	Bombycidae	D13160	56.0	43.5	44.4	1782
Samia cynthia ricini	Lepidoptera	Saturniidae	AB055844	56.4	40.9	43.6	1779

ENC = effective number of codons
GC3s = GC content at third synonymous position
GC% = % of overall GC content
#AA = total number of amino acids

(Figure 6). Thus, the synonymous codon usage bias of vitellogenins probably reflects the biased GC3. Highly expressed proteins in general have higher selective constraints on synonymous codon choices for translational efficiency and/or accuracy. For example, most of the highly expressed ribosomal genes of *An. gambiae* also show a high synonymous codon usage bias with a biased GC3 (r^2 = 0.4606) (Table 10 and Figure 7).

The pattern of synonymous codon usage is also of interest. Since the overall synonymous codon usage provides less informative data, Table 9 includes gene products with various expression levels. Preferred and rare codons were determined from the top 15% of all genes based on ENC and vitellogenin genes, assuming that these gene products are highly expressed. Amino acids such as asparagine, lysine, glutamine, phenylalanine, and tyrosine are preferentially encoded by one codon and not by other synonymous codons. Genes that use synonynous codons more equally (have high ENC) are probably under weak translational selective constraints, as those gene products may not be highly expressed. Synonymous codon usage in *An. gambiae* ribosomal genes (ENC ≤ 40.8) was also examined (Table 11). An analysis of interspecific synonymous codon choice in vitellogenins and ribosomal proteins revealed that when bias is present most of the preferential codons were conserved, except for glutamic acid; *An. gambiae* ribosomal genes preferentially use GAG over GAA, and *Ae. aegypti* vitellogenin genes preferentially use GAA over GAG. Extremely rarely used synonymous codons such as TTA (leucine), CTA (leucine), ATA (isoleucine), GGG (glycine) are also evolutionary conserved, suggesting that the usage of these codons may be disadvantageous to the translational rate of highly expressed proteins.

The patterns of synonymous codon usage in the autogenous and anautogenous mosquito vitellogenin genes were also examined (Table 12**)**. All anautogenous mosquito vitellogenin genes (*An. albimanus, Ae. aegypti* and *Cx. quinquefaciatus*, shown in red in Table 12)

Table 9. Synonymous codon usage bias of 150 *Aedes aegypti* genes, not including mitochondrial genes. Orange = top 15%. Blue = lower 15%.

Genes	Accession #	ENC	GC3 (%)	GC (%)	# AA
Ribosomal protein S24	AY064700	30.3	79.0	57.4	130
Sterol carrier protein 2 variant 2	AF539988	30.9	77.5	55.8	110
Sterol carrier protein 2 variant 1	AF539987	31.9	85.4	57.7	111
Ribosomal protein S11	AY133344	33.6	78.5	56.4	152
Putative calreticulin	AF466603	33.8	72.2	54.4	407
Thioredoxin	AF394233	34.0	84.3	54.7	106
Elongation factor 2	AY064104	34.1	67.6	54.5	844
Enhancer of rudimentary	U66869	34.3	85.0	57.3	103
Peroxidase	AF098717	34.5	73.6	58.3	683
Ribosomal protein S6	AF154067	35.2	75.2	58.5	346
Actin	U20287	35.4	60.5	53.3	376
Vacuolar ATPase B subunit	AF092934	36.0	68.8	54.8	496
Vacuolar ATPase C subunit	AF008924	36.7	60.0	55.4	157
Ribosomal protein L31	AF324863	37.0	76.9	57.5	124
Vitelline membrane 15a3	U91682	37.8	55.8	62.6	122
Vacuolar ATPase A subunit	AF008922	37.9	66.4	54.9	615
2-Cys thioredoxin peroxidase	AF323181	38.6	74.3	57.3	196
Hexamerin 2 alpha	U86080	39.3	72.5	52.5	712
Takeout	AF458100	39.4	77.5	54.6	248
Ferritin subunit	L37082	39.7	75.9	55.2	209
Vitellogenin Vg-A1	L41842	39.7	61.8	48.7	2148
Vitellogenin Vg-B	AY380797	39.8	62.8	49.1	2142
Vitellogenin Vg-C	AY373377	40.8	62.7	49.3	2087
Vitelline membrane 15a1	U91680	40.9	64.2	63.3	108
Ubiquitin-52-amino-acid fusion protein	AF418984	41.0	70.6	52.6	128
Ecdysteroid receptor	U02021	41.6	73.9	56.3	675
Insulin receptor	U72939	42.8	74.6	56.2	1390
Hexamerin 1 gamma	U86079	43.0	71.1	46.5	692
Dopachrome conversion enzyme	AY064099	43.5	74.2	56.9	463
Cathepsin B-like thiol protease	L41940	43.7	62.5	55.5	342
Defensin isoform C1	AF156092	44.2	54.1	54.5	99
Transfferin precursor	AF019117	44.5	67.4	55.2	633
Ovarian ecdysteroidogenic hormone I	U96112	44.6	73.2	59.1	149
Putative secreted protein	AF466586	45.5	34.6	38.7	255
Toll1	AY124196	45.6	71.0	53.1	1124
Chymotrypsin-like protease precursor	U56423	45.8	61.7	55.5	268
Putative 8.3 Kda secreted protein	AF466592	46.3	53.3	53.1	96
Microvilli membrane protein	AY050565	46.3	60.3	49.4	207
Nuclear receptor 3	AF230281	46.3	70.3	54.1	468
Probable transport protein Sec61 alpha subunit	AY064125	46.9	56.7	49.9	476
Cellular retinaldehyde-binding protein	AY064091	47.0	70.7	53.2	290
Vitellogenic cathepsin-B like protease	AF127592	47.0	61.4	56.0	386
Glutamine synthetase	AF004351	47.1	52.0	51.8	400
Preproleukokinin	U66832	47.2	59.0	50.3	228
Glutathione S-transferase	AF384858	47.4	63.2	54.6	218
Cellular retinaldehyde-binding protein	AF329893	47.8	70.7	53.0	290
Phenylalanine hydroxylase	AY099427	47.9	65.5	53.6	447
Lysosomal aspartic protease	M95187	47.9	48.0	46.4	387
Vitellogenin receptor	L77800	48.0	67.7	54.3	1847
Vitellogenic carboxypeptidase	L46594	48.2	61.1	50.2	471
Dopa carboxylase	U27581	48.3	55.0	52.6	337
Ribosomal protein L17A	AY064121	48.5	58.5	56.2	140
Lipophorin receptor	AF355595	48.5	69.4	55.5	1156
Nuclear hormone receptor	AF165528	48.6	65.7	57.2	545
Putative protein G12	AY038042	48.8	61.8	49.8	207
Putative C-type lectin	AF466606	48.9	66.9	55.2	154
Late trypsin precursor	M77814	49.0	59.9	54.5	257
Carbonic anhydrase	AF395662	49.1	63.5	55.7	298
MRGG	U84248	49.1	65.8	57.9	663
Defensin isoform A1	AF156088	49.2	64.9	57.5	98
Glucosamine-fructose-6-phosphate aminotransferase	AF399922	49.7	50.2	48.0	675
Amylase II	AF000568	49.9	43.0	47.8	486
Vitelline membrane 15a2	U91681	49.9	61.5	64.6	97
Putative secreted protein	AF466601	50.0	53.4	48.5	259
Nuclear hormone receptor	AF303224	50.2	64.5	56.3	493
Nuclear receptor FTZ-F1 protein	AF274870	50.2	67.5	56.6	840
Defensin isoform B1	AF156090	50.4	59.8	56.5	98
Neuropeptide F preproprotein	AF474405	50.6	46.6	51.5	90
Relish	AF498105	50.7	65.0	51.8	995
V-Atpase 110 Kda integral membrane subunit	AF173554	50.8	57.5	47.5	804
Putative 34 Kda secreted protein	AF466595	50.9	66.6	50.5	316
Salivary purine nucleosidase	AY038045	51.2	52.0	47.5	338
Ribonucleotide reductase 1	AF411296	51.3	60.3	49.1	844
Ultraspiracle	AF305214	51.5	62.7	54.3	459
Head peptide I	AF155738	51.6	40.2	50.0	128
Preproallatostatin	U66841	51.7	62.7	56.7	197
Putative mRNA binding protein	AY064109	51.8	30.7	41.6	419
Vitellogenin convertase	L46373	51.9	63.1	56.7	1060
Chitin synthase	AF223577	52.1	62.7	50.5	1564
Pyruvate carboxylase	L36530	52.1	50.3	50.2	1195

Genes	Accession #	ENC	GC3 (%)	GC (%)	# AA
Early trypsin	X64362	52.2	58.1	55.0	254
Peritrophin	AY050566	52.3	43.1	49.5	486
E74A	AF435023	52.3	63.8	58.0	778
Aquaporin	AF218314	52.4	58.7	54.8	249
Phosphoglycerate kinase	AY064545	52.4	51.9	50.9	415
Chitinase 1	AF026491	52.5	52.9	51.9	574
Heavy clathrin chain	S:1076	52.6	51.9	47.2	1682
Prophenoloxidase propo1	AF292114	52.7	45.8	48.2	684
Apyrase	L41391	53.2	40.7	42.8	563
Ribonucleotide reductase 2	AF411297	53.3	55.8	46.8	397
Prophenol oxidase propo3	AF310673	53.7	42.4	45.9	687
Cyclic-AMP response element binding protein	AY083158	53.9	64.0	56.0	295
Angiopoietin-like protein	AF466609	54.0	56.6	49.4	296
Na+/H+ antiporter	AF187723	54.0	58.6	51.3	1179
E74B	AF435022	54.1	60.0	58.2	827
Cytochrome P450	AY064093	54.4	58.7	48.8	536
Putative secreted protein	AF466589	54.6	59.9	51.8	255
Kynurenine aminotransferase	AF395204	54.8	61.5	51.7	477
Maltase-like I	M30442	54.9	50.1	46.7	579
Kynurenine monooxygenase	AF325508	55.0	60.6	50.6	476
Putative 16.9 Kda secreted protein	AF466612	55.1	43.2	39.6	148
Putative lysozyme	AF466591	55.1	39.9	40.0	144
Chitinase 2	AF026492	55.4	38.4	45.5	1635
Tryptophan oxygenase	AF325458	55.5	52.7	48.3	393
GATA transcription factor	AJ400338	55.5	53.2	51.5	868
Putative 18.6 Kda secreted protein	AF466604	55.8	46.6	48.8	168
Putative serine protease	AF466600	55.8	40.5	46.7	296
Lipase-like protein	AF303984	56.0	49.2	51.6	278
Asparagine synthetase	U84118	56.0	55.8	50.3	562
HNF4 isoform A	AF059026	56.0	34.2	41.3	565
Prophenol oxidase propo4	AF327409	56.1	47.1	47.9	686
Toll-related protein	AY124195	56.3	60.2	47.1	1007
Putative secreted protein	AF466605	56.4	36.5	40.1	261
Eye pigment transporter	U88851	56.4	54.9	49.7	692
Putative secreted protein	AF466596	56.4	55.0	49.2	292
Putative carboxylesterase	AF466590	56.5	36.1	43.8	636
Long form D7Bclu1 salivary protein	AF420272	56.5	41.6	42.0	332
Mucin-like protein	AF125984	56.5	50.8	53.0	183
Putative 56.5 Kda secreted protein	AF466593	56.6	53.0	48.4	530
Toll1B	AY124197	56.7	52.4	47.0	1076
Putative secreted protein	AF466594	56.8	40.1	46.2	371
Serine protease inhibitor	AF466599	57.1	46.4	43.2	418
Carboxypeptidase A	AF165923	57.1	46.3	45.9	427
GABA receptor subunit	U28803	57.1	47.4	48.4	533
Aminopeptidase N	AY064079	57.4	53.9	49.7	955
Short-chain dehydrogenease/reductase	AY064123	57.5	48.2	48.4	254
Fxa-directed anticoagulant precursor	AF050133	57.5	44.3	42.7	415
Putative secreted protein	AF466597	57.7	39.1	40.3	580
Juvenile hormone epoxide hydrolase	AF517544	57.8	55.0	48.7	461
Putative secreted protein	AF466598	58.3	50.5	43.3	566
Chymotrypsin	AF487334	58.3	41.4	46.8	257
Amylase I	AF000569	58.5	42.2	45.3	737
Putative 30.5 Kda secreted protein	AF466607	58.6	54.4	46.2	275
Putative 18.2 Kda secreted protein	AF466613	58.6	38.4	38.6	158
Multivitamin transporter	AY063742	58.6	54.0	49.6	592
Allatotropin	U65314	58.6	55.4	51.8	175
Putative adenosine deaminase	AF466610	58.6	47.5	44.6	530
Putative 7.8 Kda secreted protein	AF466588	58.7	69.2	51.2	95
Putative chymotrypsin-like serine protease	AF466611	58.7	48.7	51.9	288
Putative prophenoloxidase activating protein	AF466614	58.8	42.2	47.9	413
Odorant-binding protein-related protein	AY062131	58.9	66.7	49.3	140
Short form D7Cclu23 salivary protein	AF420273	58.9	49.3	41.2	149
Salivary vasodilatory protein precursor	AF108099	59.6	40.7	43.5	85
Amino acid transporter	AF543193	59.9	42.1	43.6	640
lysozyme	AJ290428	59.9	56.7	46.2	148
Putative secreted protein	AF466602	60.9	50.0	49.4	85
Putative 30 Kda allergen-like protein	AF466608	61.0	50.0	47.4	215
30 Kda salivary gland allergen	AF001927	61.0	39.8	46.9	253
D7 Protein	M33157	61.0	51.6	45.2	321
Sialokinin I preproprotein	AF108101	61.0	42.0	44.3	85

revealed a high synonymous codon usage bias, preferentially using one or two synonymous codons over others. For example, amino acids such as asparagine, lysine, glutamine, phenylalanine, and tyrosine are almost exclusively encoded by one synonymous codon and not by other synonymous codons. An analysis of interspecific codon choice in anautogenous mosquito vitellogenin genes revealed that when bias is present most of the preferential codons were conserved in all three species. The exception is glutamic acid where *An. albimanus* Vg-C preferentially uses GAG over GAA, and all three

Table 10. Synonymous codon usage bias for *Anopheles gambiae* ribosomal protein genes

Ribosomal Proteins	Accession #	ENC	GC3 (%)	GC (%)	AA length
L7A	agCP5980	29.7	88.5	61.1	271
S4	agCP3558	31.1	85.2	60.7	262
Sa	agCP9554	31.3	82.7	63.4	285
L13	agCP12116	31.9	83.9	61.4	217
S2	agCP13883	32.1	74.2	61.7	274
S19	agCP10609	32.2	81.9	63.3	158
S6	agCP1429	32.2	85.9	65.0	372
L7	agCP4788	32.5	85.8	60.0	276
L18a	agCP9264	32.7	86.0	58.9	177
L1	agCP1226	32.8	85.1	59.8	276
P2	agCP1641	32.8	70.6	62.2	112
L3	agCP7843	33.1	75.9	58.3	417
L36	agCP1535	33.4	81.7	59.3	113
S23	agCP12166	33.4	81.4	61.5	143
L29	agCP5781	33.5	79.7	55.7	76
S24	agCP14988	33.5	87.1	60.0	130
S29	agCP12540	33.5	71.7	54.2	56
L8	agCP5982	33.7	79.2	63.3	261
S15a	agCP13543	33.8	87.2	59.5	130
S25	agCP14068	33.9	78.6	57.1	119
L10	agCP4253	33.9	83.5	61.8	219
S13	agCP6643	34.2	80.0	58.1	151
L34	agCP10269	34.3	78.0	61.2	134
L9	agCP14553	34.3	85.9	58.1	190
P0	agCP7923	34.9	81.0	59.8	314
L37a	agCP8681	35.0	65.9	57.7	89
S3a	agCP7766	35.1	78.8	57.2	269
L11	agCP6972	35.1	81.6	59.3	190
L32	agCP12827	35.2	81.1	59.5	134
S3	agCP12518	35.2	78.9	61.7	260
L30	agCP13921	35.3	77.0	57.2	116
S14a	agCP1637	35.3	77.0	63.8	152
S26	agCP8880	36.3	77.0	58.8	115
S12	agCP7049	36.6	79.9	60.1	137
L46	agCP9893	37.5	72.3	54.2	51
L44	ebiP415	37.6	72.3	53.5	104
S20	agCP11873	37.9	67.2	55.9	121
S5	agCP11155	38.4	71.2	58.4	233
L19	agCP14063	38.7	69.7	55.2	204
S18	agCP10687	38.8	67.3	55.0	152
S7	agCP8386	40.2	78.0	58.9	227
L28	agCP5722	40.4	79.1	62.6	157
P1	agCP11536	40.5	55.5	56.6	113
L12	agCP8207	40.8	65.2	56.4	165
L31	agCP9995	41.0	80.2	58.3	124
S9	agCP14911	41.0	69.3	56.9	195
S14b	agCP7464	41.6	66.2	60.3	152
S21	agCP13055	42.2	75.0	55.4	83
S22	agCP3614	43.7	77.3	57.2	375
L40	agCP11398	44.7	73.8	54.2	128
S15	agCP5590	45.6	74.5	54.4	278
S10	agCP5806	46.8	74.0	58.7	159
S17	agCP4384	49.1	60.6	49.4	131
L23	agCP8340	51.1	54.1	54.8	140

Ordered by increasing ENC value
Green shading = the highly biased *Anopheles gambiae* ribosomal proteins shown in Table 8.

vitellogenin genes in *Ae. aegypti* use GAA over GAG.

In contrast, vitellogenin genes from autogenous mosquitoes(*Oc. atropalpus* and *Tx. amboinesis*, shown in blue in Table 12) showed a relatively low synonymous codon usage bias. *Tx. amboinensis* Vg-C, and Vg-B and Vg-C of *Oc. atropalpus,* showed the lowest bias among all mosquito vitellogenin genes examined. The average ENC for the two autogenous species was 47.4, and for the three anautogenous species was 38.6. A positive correlation was found between GC content and the degree of synonymous codon usage bias measured by ENC (Figure 9). Thus, the highest synonymous codon usage bias of *An. albimanus* Vg-C probably reflects the biased GC3s (83.5%). Complete sequences of 54 *Ae. aegypti* genes coding for other proteins were analyzed for synonymous codon usage bias. The vitellogenin genes show a pattern of codon usage that is among the most highly biased of all known complete *Ae. aegypti* genes based on ENC (Figure 9).

The nucleotide composition of the mosquito vitellogenin genes were also examined. A χ^2 test of homogeneity of base frequency across taxa was

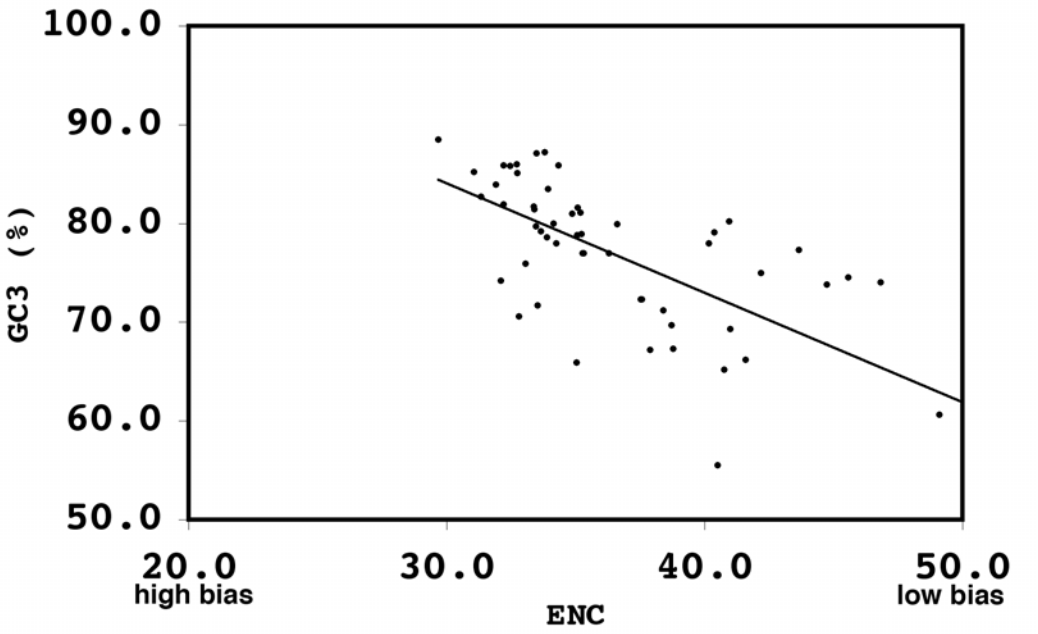


Figure 7. The correlation of synonymous codon usage bias and GC3 of *Anopheles gambiae* ribosomal genes. r^2 = 0.4606.

performed using the PAUP* program to determine the level of bias at each codon position. This analysis showed that the base frequency for the first and third codon positions varied significantly across taxa ($p < 0.01$), probably due to different codon usage preferences among mosquito species.

Vitellogenin gene expression in *Ae. aegypti*

Previous studies by Northern analysis and RNA dot blot analyses using a probe that most likely

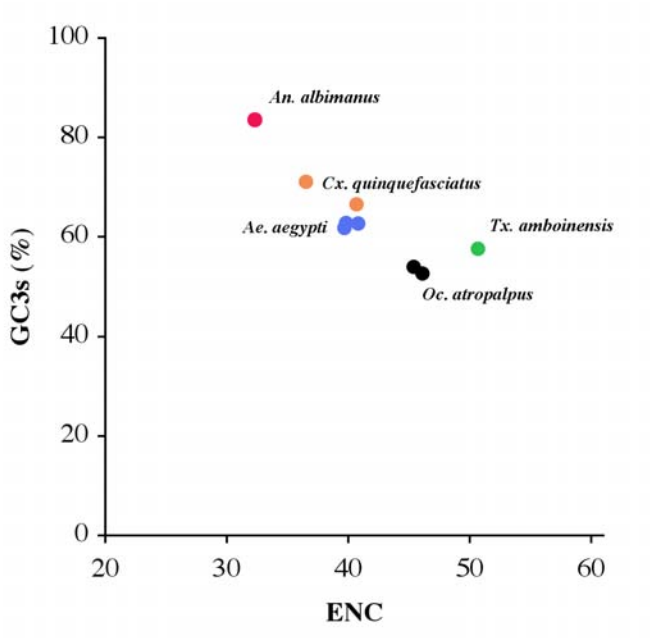


Figure 8. The correlation of synonymous codon usage bias and GC3 of anautogenous and autogenous mosquito vitellogenin genes.

Table 11. Synonymous codon usage of *Aedes aegypti* genes and *Anopheles gambiae* ribosomal protein genes, not including mitochondrial genes.

				Aedes aegypti									*Anopheles gambiae*		
		Overall		Bottom 15%			TOP 15%			Vitellogenins			Ribosomal protein genes		
Codons	**N**	**(%)**	**RSCU***	**N**	**(%)**	**RSCU**	**N**	**(%)**	**RSCU**	**N**	**(%)**	**RSCU**	**N**	**(%)**	**RSCU**
TTT phe F	870	25.4	0.51	152	44.2	0.88	82	10.6	0.21	60	12.1	0.24	12	5.0	0.10
TTC phe F	2551	74.6	1.49	192	55.8	1.12	695	89.4	1.79	436	87.9	1.76	228	95.0	1.90
TTA leu L	370	6.0	0.36	80	11.8	0.71	2	0.3	0.02	1	0.3	0.02	1	0.2	0.01
TTG leu L	1360	22.1	1.33	173	25.6	1.53	159	19.9	1.19	89	27.1	1.63	41	7.0	0.42
CTT leu L	678	11.0	0.66	103	15.2	0.91	48	6.0	0.36	28	8.5	0.51	19	3.3	0.20
CTC leu L	849	13.8	0.83	78	11.5	0.69	92	11.5	0.69	39	11.9	0.71	104	17.9	1.07
CTA leu L	522	8.5	0.51	78	11.5	0.69	25	3.1	0.19	9	2.7	0.16	6	1.0	0.06
CTG leu L	2376	38.6	2.32	165	24.4	1.46	473	59.2	3.55	162	49.4	2.96	411	70.6	4.24
ATT ile I	1279	33.6	1.01	166	39.1	1.17	138	25.7	0.77	63	32.6	0.98	99	20.8	0.62
ATC ile I	2041	53.7	1.61	151	35.5	1.07	397	73.9	2.22	130	67.4	2.02	375	78.8	2.36
ATA ile I	481	12.7	0.38	108	25.4	0.76	2	0.4	0.01	0	0.0	0.00	2	0.4	0.01
ATG met M	1827	100.0	1.00	179	100.0	1.00	298	100.0	1.00	114	100.0	1.00	188	100.0	1.00
GTT val V	1233	26.0	1.04	126	25.8	1.03	216	26.6	1.06	132	40.7	1.63	71	11.9	0.48
GTC val V	1390	29.3	1.17	76	15.6	0.62	358	44.1	1.76	119	36.7	1.47	196	32.9	1.32
GTA val V	681	14.4	0.57	99	20.3	0.81	51	6.3	0.25	19	5.9	0.23	24	4.0	0.16
GTG val V	1435	30.3	1.21	187	38.3	1.53	187	23.0	0.92	54	16.7	0.67	304	51.1	2.04
TCT ser S	640	11.4	0.69	64	11.9	0.72	153	15.7	0.94	124	18.7	1.12	17	4.4	0.26
TCC ser S	1179	21.1	1.26	81	15.1	0.91	306	31.3	1.88	172	25.9	1.56	100	25.8	1.55
TCA ser S	647	11.6	0.69	105	19.6	1.18	62	6.3	0.38	56	8.4	0.51	5	1.3	0.08
TCG ser S	1290	23.0	1.38	105	19.6	1.18	217	22.2	1.33	122	18.4	1.10	187	48.2	2.89
CCT pro P	547	15.3	0.61	57	19.7	0.79	69	11.1	0.44	35	11.6	0.46	20	5.4	0.21
CCC pro P	714	20.0	0.80	52	17.9	0.72	160	25.6	1.03	56	18.5	0.74	55	14.7	0.59
CCA pro P	1173	32.9	1.32	105	36.2	1.45	292	46.8	1.87	175	57.9	2.32	35	9.4	0.38
CCG pro P	1131	31.7	1.27	76	26.2	1.05	103	16.5	0.66	36	11.9	0.48	263	70.5	2.82
ACT thr T	801	20.7	0.83	88	23.7	0.95	146	25.3	1.01	98	39.4	1.57	23	6.0	0.24
ACC thr T	1489	38.5	1.54	99	26.7	1.07	359	62.3	2.49	133	53.4	2.14	215	56.4	2.26
ACA thr T	635	16.4	0.66	92	24.8	0.99	28	4.9	0.19	10	4.0	0.16	10	2.6	0.10
ACG thr T	947	24.5	0.98	92	24.8	0.99	43	7.5	0.30	8	3.2	0.13	133	34.9	1.40
GCT ala A	1487	29.2	1.17	148	29.2	1.17	347	37.8	1.51	229	51.7	2.07	125	17.3	0.69
GCC ala A	1889	37.1	1.49	123	24.3	0.97	475	51.7	2.07	184	41.5	1.66	394	54.5	2.18
GCA ala A	881	17.3	0.69	133	26.3	1.05	42	4.6	0.18	20	4.5	0.18	45	6.2	0.25
GCG ala A	828	16.3	0.65	102	20.2	0.81	55	6.0	0.24	10	2.3	0.09	159	22.0	0.88
TAT tyr Y	904	29.2	0.58	116	41.6	0.83	130	14.9	0.30	96	15.6	0.31	23	10.8	0.22
TAC tyr Y	2197	70.8	1.42	163	58.4	1.17	742	85.1	1.70	518	84.4	1.69	189	89.2	1.78
TAA OCH Z	70	46.7	1.40	10	43.5	1.30	17	73.9	2.22	1	33.3	1.00	41	93.2	2.80
TAG AMB Z	38	25.3	0.76	3	13.0	0.39	5	21.7	0.65	2	66.7	2.00	2	4.5	0.14
CAT his H	734	38.7	0.77	62	50.4	1.01	81	24.3	0.49	52	30.2	0.60	45	20.9	0.42
CAC his H	1163	61.3	1.23	61	49.6	0.99	252	75.7	1.51	120	69.8	1.40	170	79.1	1.58
CAA gln Q	1206	39.5	0.79	140	50.2	1.00	156	28.6	0.57	128	37.9	0.76	22	8.0	0.16
CAG gln Q	1845	60.5	1.21	139	49.8	1.00	389	71.4	1.43	210	62.1	1.24	253	92.0	1.84
AAT asn N	1411	37.6	0.75	193	52.2	1.04	91	15.3	0.31	63	18.1	0.36	35	11.9	0.24
AAC asn N	2342	62.4	1.25	177	47.8	0.96	505	84.7	1.69	286	81.9	1.64	258	88.1	1.76
AAA lys K	1604	37.1	0.74	267	53.7	1.07	135	14.5	0.29	83	18.7	0.37	81	8.6	0.17
AAG lys K	2725	62.9	1.26	230	46.3	0.93	794	85.5	1.71	360	81.3	1.63	862	91.4	1.83
GAT asp D	2355	54.8	1.10	283	64.5	1.29	379	50.5	1.01	200	57.1	1.14	145	45.2	0.90
GAC asp D	1946	45.2	0.90	156	35.5	0.71	372	49.5	0.99	150	42.9	0.86	176	54.8	1.10
GAA glu E	2425	57.1	1.14	250	60.8	1.22	473	59.9	1.20	252	68.3	1.37	104	24.4	0.49
GAG glu E	1823	42.9	0.86	161	39.2	0.78	316	40.1	0.80	117	31.7	0.63	323	75.6	1.51
TGT cys C	572	37.5	0.75	75	48.4	0.97	34	23.6	0.47	14	21.9	0.44	21	16.5	0.33
TGC cys C	952	62.5	1.25	80	51.6	1.03	110	76.4	1.53	50	78.1	1.56	106	83.5	1.67
TGA OPA Z	42	28.0	0.84	10	43.5	1.30	1	4.3	0.13	0	0.0	0.00	1	2.3	0.07
TGG trp W	905	100.0	1.00	104	100.0	1.00	95	100.0	1.00	34	100.0	1.00	67	100.0	1.00
CGT arg R	882	25.0	1.50	41	13.9	0.83	275	49.5	2.97	116	47.9	2.88	247	34.2	2.05
CGC arg R	769	21.8	1.31	44	14.9	0.89	209	37.6	2.26	105	43.4	2.60	348	48.1	2.89
CGA arg R	637	18.1	1.08	60	20.3	1.22	22	4.0	0.24	8	3.3	0.20	48	6.6	0.40
CGG arg R	573	16.3	0.98	49	16.6	1.00	19	3.4	0.21	1	0.4	0.02	52	7.2	0.43
AGT ser S	802	14.3	0.86	100	18.7	1.12	45	4.6	0.28	35	5.3	0.32	11	2.8	0.17
AGC ser S	1039	18.6	1.11	81	15.1	0.91	194	19.9	1.19	154	23.2	1.39	68	17.5	1.05
AGA arg R	393	11.1	0.67	54	18.3	1.10	22	4.0	0.24	11	4.5	0.27	11	1.5	0.09
AGG arg R	272	7.7	0.46	47	15.9	0.96	9	1.6	0.10	1	0.4	0.02	17	2.4	0.14
GGT gly G	1318	27.8	1.11	119	24.8	0.99	254	36.5	1.46	77	26.6	1.06	232	40.5	1.62
GGC gly G	1229	25.9	1.04	121	25.2	1.01	161	23.2	0.93	63	21.7	0.87	233	40.7	1.63
GGA gly G	1765	37.3	1.49	188	39.2	1.57	263	37.8	1.51	146	50.3	2.01	95	16.6	0.66
GGG gly G	426	9.0	0.36	52	10.8	0.43	17	2.4	0.10	4	1.4	0.06	13	2.3	0.09

*RSCU= relative synonymous codon usage

cross-hybridized with all members of the vitellogenin genes demonstrated that vitellogenin gene(s) were expressed only in fat body preparations of blood-fed female mosquitoes (Racciopi et al. 1986; Gemmill et al. 1986). To determine if all three genes were expressed, and to confirm the sex-and-stage specificity of their expression, real-time PCR was used. These experiments demonstrated that mRNAs from Vg-A1, Vg-B and Vg-C showed peak expression by 36 hours, and negligible by 72 hours after a blood meal (Fig. 11). Expression in whole body extracts of non-blood fed females, males, larvae and pupae of all three genes occurred but at 4 to 6 orders of

Table 12. Synonymous codon usage in autogenous and anautogenous mosquito vitellogenin genes. Red = anautogenous. Blue = autogenous.

		Ae. aegypti			*Oc. atropalpus*		*Cx. quinquefasciatus*		*Tx. amboinensis*	*An. albimanus*
Amino Acid	**Codon**	**Vg-A1**	**Vg-B**	**Vg-C**	**Vg-B*a***	**Vg-C*a***	**Vg-C1**	**Vg-C2*b***	**Vg-C**	**Vg-C**
Phe	TTT	18	19	23	43	39	19	17	23	11
	TTC	156	152	128	134	113	83	42	116	137
Leu	TTA	0	0	1	3	2	0	0	3	0
	TTG	30	32	27	39	28	12	27	25	6
Leu	CTT	9	10	9	17	14	13	13	12	9
	CTC	12	14	13	10	10	15	17	17	19
	CTA	4	4	1	2	9	2	1	7	3
	CTG	59	51	52	32	29	92	89	43	82
Ile	ATT	27	18	18	30	26	21	40	28	10
	ATC	42	38	50	25	31	62	55	28	62
	ATA	0	0	0	3	1	0	0	11	0
Met	ATG	39	43	32	33	33	40	38	38	33
Val	GTT	45	43	44	45	39	36	43	50	20
	GTC	36	45	38	31	26	55	60	25	61
	GTA	5	4	10	10	14	1	3	12	1
	GTG	18	17	19	21	15	45	33	34	41
Ser	TCT	39	39	46	41	36	44	38	19	19
	TCC	59	49	64	30	27	42	61	45	26
	TCA	18	18	20	26	21	8	15	27	1
	TCG	34	47	41	44	36	55	51	36	57
	AGT	10	11	14	19	17	4	5	31	0
	AGC	53	56	45	58	35	61	65	49	51
Pro	CCT	15	13	7	20	22	16	12	10	7
	CCC	16	15	25	15	16	21	21	22	34
	CCA	63	63	49	44	33	11	24	34	1
	CCG	10	10	16	17	17	45	37	29	50
Thr	ACT	27	36	35	34	31	26	29	19	13
	ACC	48	42	43	36	30	62	55	44	74
	ACA	3	2	5	9	11	0	1	9	1
	ACG	2	3	3	3	4	2	10	10	10
Ala	GCT	86	77	66	74	64	54	67	62	32
	GCC	60	68	56	55	42	77	79	56	81
	GCA	4	8	8	14	10	0	7	19	1
	GCG	5	4	1	5	2	4	8	19	8
Tyr	TAT	32	32	32	66	39	16	7	46	11
	TAC	185	181	152	150	129	113	71	98	166
His	CAT	20	17	15	31	24	14	17	28	9
	CAC	41	42	37	30	21	41	41	38	56
Gln	CAA	42	42	44	55	45	40	36	47	4
	CAG	60	67	83	59	55	88	79	78	112
Asn	AAT	22	23	18	32	48	8	15	59	1
	AAC	94	88	104	77	66	118	103	74	103
Lys	AAA	31	24	28	42	42	12	13	41	2
	AAG	117	131	112	96	74	134	127	77	122
Asp	GAT	65	67	68	72	59	57	59	55	48
	GAC	51	49	50	55	46	64	49	46	82
Glu	GAA	86	83	83	92	88	70	74	71	25
	GAG	37	43	37	34	32	79	71	54	112
Cys	TGT	5	5	4	7	5	4	5	8	3
	TGC	15	15	20	11	10	22	19	11	21
Trp	TGG	12	10	12	12	14	19	28	15	16
Arg	CGT	41	39	36	41	39	34	43	19	28
	CGC	31	34	40	23	20	48	30	34	52
	CGA	2	2	4	4	3	2	6	8	0
	CGG	0	0	1	2	2	0	3	7	1
Arg	AGA	4	2	5	3	3	4	6	11	0
	AGG	1	0	0	0	0	1	0	5	1
Gly	GGT	23	30	24	25	15	24	31	35	24
	GGC	22	18	23	16	17	14	10	16	25
	GGA	55	46	45	48	44	56	36	35	43
	GGG	2	1	1	3	2	1	0	4	1
Total		2148	2142	2087	2108	1825	2111	2042	2032	2029

Ae. atropalpus of Vg-B and Vg-C were not determined
The n-terminal end of the *Cx. quinquefasciatus*VgC2 gene was truncated (39 amino acid residues)

magnitude lower levels; expression of Vg-A1 at 36 hours was 4 orders, Vg-B was 5 orders, and Vg-C was 6 orders of magnitude higher than in fat body of non-blood fed females.

Discussion

Sequence analysis of vitellogenin genes in *Ae. aegypti*

To gain a better understanding of the evolution of vitellogenin genes in mosquitoes, genomic DNA libraries were constructed from several distantly related mosquitoes with different reproductive strategies, and vitellogenin genes were cloned and

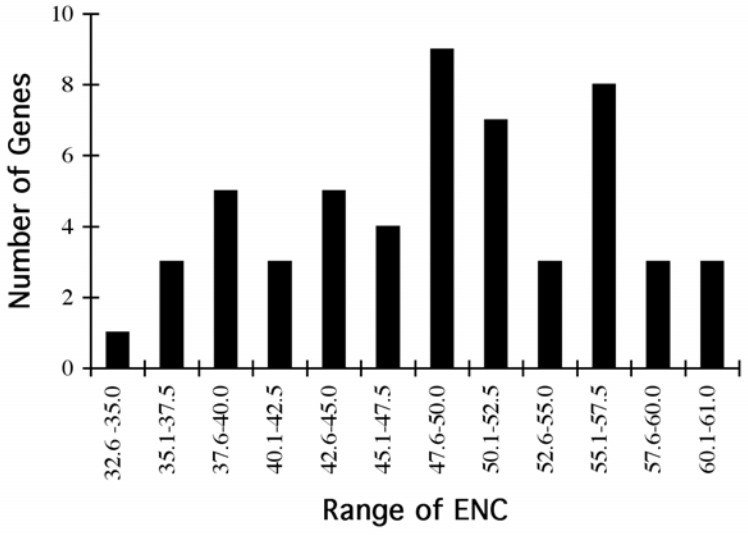


Figure 9. Distribution of *Aedes aegypti* genes based on synonymous codon usage bias. ENC = effective number of codons. Vg-A1, 39.7, Vg-B 39.8. Vg-C 40.8.

sequenced from them to conduct a comparative sequence analysis of mosquito vitellogenin genes.

Four vitellogenin genes of *Ae. aegypti* had been previously cloned (Hamblin et al. 1987), and one of them, Vg-A1, was sequenced and extensively studied (Gemmill et al. 1986; Racioppi et al. 1986, Chen et al. 1994; Romans et al. 1995). However, the clones of the other genes were lost. To gain a better understanding of the evolution of these genes, genomic DNA libraries were screened. Four vitellogenin genes, Vg-A1, Vg-A2, Vg-B and Vg-C, were again cloned. Vg-A2, Vg-B and Vg-C were sequenced and a comparative sequence analysis was conducted. Sequence analyses of Vg-A1 and Vg-A2 showed that they are possibly allelic to each other, Vg-A1 and Vg-B are closely related, and possibly arose by a recent gene

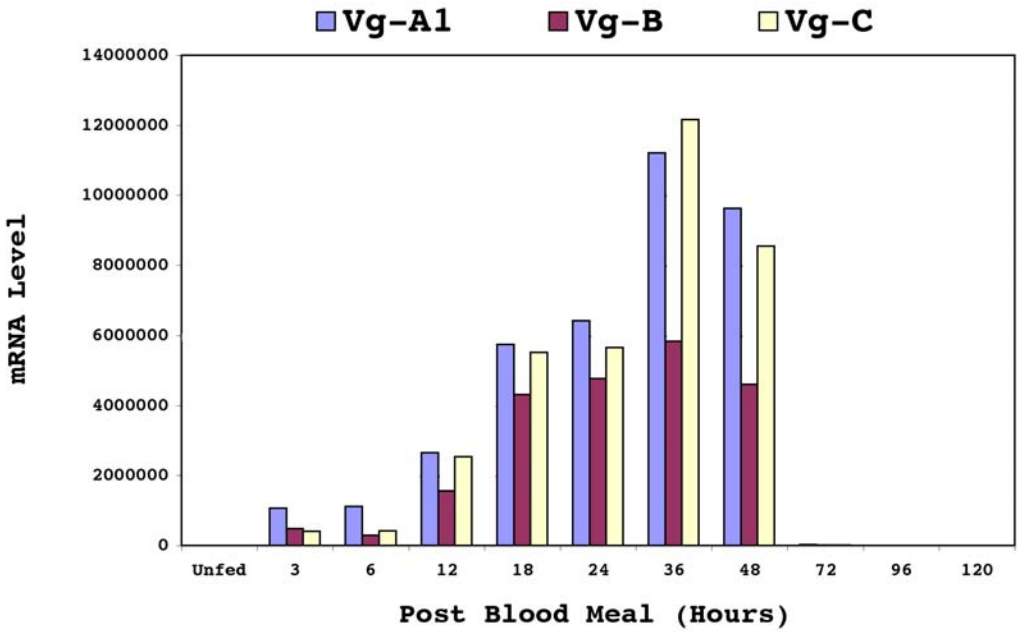


Figure 10. Expression of *Aedes aegypti* vitellogenin genes. mRNA levels were measured by real-time PCR.

duplication event, and Vg-C is distantly related to Vg-A1 and Vg-B lineage, and possibly arose by an earlier gene duplication event.

Vg-B and Vg-C were relatively conserved in length with minor insertions and deletions of amino acids, especially within the polyserine regions, relative to Vg-A1. Molecular signatures previously characterized in Vg-A1 were all present and conserved. These include two short introns, two polyserine regions, highly conserved cysteine amino acid residues, and a cleavage site between the two vitellogenin subunits. The vitellogenins are postranslationally cleaved in the fat body into two small and large subunits (Romans et al. 1995) (Figure 1) by a vitellogenin convertase (Chen and Raikhel 1996). The predicted cleavage sequence (RXRR, Hagedorn et al 1998; Sappington and Raikhel 1998) was present in all three vitellogenins. The position of cysteine residues are also well conserved. There are 19 evolutionarily conserved cysteines out of the 20 residues in Vg-A1 and Vg-B, and out of 24 in Vg-C. In general, disulfide bonds between cysteine amino acid residues stabilize protein structure. When vitellogenin protein sequences from other insect taxa were compared to the *Ae. aegypti* sequences for the position of cysteine residues, 6 positions near the C-terminal were found to be highly conserved, suggesting that the cysteines at these positions may play a role in maintaining the structure of the vitellogenin proteins of insects.

Sequence analysis of vitellogenin genes in other mosquitoes

Two distinct members of the vitellogenin gene family, Vg-C1 and Vg-C2, were cloned from *Cx. quinquefasciatus*. A complete coding sequence was obtained for Vg-C1, but the Vg-C2 sequence is truncated at the N-terminal end such that 39 amino acid residues are deleted. The nucleotide position at the truncation differs from the conserved first intron site of other mosquito genes, and the nucleotide sequence at the truncation also differs from the TAG or CAG that are always present in the 3' intron sequence of all mosquito vitellogenins, suggesting that it is unlikely that the truncation is due to the acquisition of a new intron site. Other organisms such as *G. gallus* (Silva et al. 1989) and *C. elegans* (Spieth et al. 1985) have pseudogenes encoding non-functional vitellogenins due to stop codons and frame shifts in the coding sequences. However, in *Cx. quinquefasciatus* Vg-C2, the rest of the coding sequence has no stop codon or frame shift, suggesting that the gene could be still active. The appearance of the truncation could be due to the insertion of transposable elements that are relatively abundant in the genome of mosquitoes (Tu 1999). Sequencing the upstream region of the truncation of Vg-C2, and expression studies either by a reverse transcription coupled with polymerase chain reaction, or 5' rapid amplification of cDNA ends using gene specific primers, would reveal whether *Cx. quinquefasciatus* Vg-C2 is expressed. If it is indeed expressed, it would be interesting to compare the pattern of expression with Vg-C1 that does not have a truncation.

Screening of *An. albimanus* and *Tx. amboinensis* genomic libraries resulted in positive clones that were all Vg-C orthologs based on the presence of molecular signatures. Rescreening with Vg-C probes derived from both species at a low stringency did not identify other members of the vitellogenin gene family. It is not certain whether the Vg-A1/B members became pseudogene(s), or the rate of mutation in this lineage is significantly different from others such that screening of the distantly related *Cx. quinquefasciatus*, *Tx, amboinensis,* and *An. albimanus* libraries failed. This is unlike other organisms such as *X. laevis* (Germond et al. 1984), *C. elegans (*Spieth et al. 1991), *D. melanogaster* (Wahli 1988), and *Ae. aegypti* (Hamblin et al. 1987), in which several members of the vitellogenin gene family exist.

However the Hymenoptera have a single vitellogenin gene (Kageyama 1994; Nose 1997; Piulachs et al. 2003; Donnell 2004).

All vitellogenin genes with complete coding sequences cloned in this study were relatively conserved in length with minor insertions and deletions of amino acids, especially within the polyserine regions, relative to their orthologs in *Ae. aegypti*. Molecular signatures previously characterized in *Ae. aegypti* were all present and conserved in the vitellogenin genes of these mosquitoes.

Amino acid composition

The vitellogenin proteins provide amino acid building blocks for embryonic development and therefore the precise amino acid sequence could be of minor importance. However, the comparative sequence analysis of vitellogenin genes suggests that a majority of amino acid substitutions are due to conserved and moderately conserved changes in amino acid physical-chemical properties. As described above, the position of cysteine residues in all vitellogenin

genes is highly conserved, consistent with other comparative studies including distantly diverged organisms (Wahli 1988; Byrne 1989, Trewitt et al. 1992; Hagedorn et al. 1998), suggesting that the overall tertiary structure of the vitellogenin proteins is important. If the amino acid sequence were of minor importance, there would likely be low discrimination between the extent of amino acid substitutions and their physical and chemical properties. Taken together, the analysis of *Ae. aegypti* vitellogenin proteins suggest that they are under moderate selective constraints to maintain tertiary structure in parts of the molecule, particularly by conserved cysteine residues. This conclusion is strongly supported by the anaysis of the tertiary structure of Lamprey vitellogenin by X-ray crystallography (Raag et al. 1988; Sharrock et al. 1992; Thompson and Banaszak 2002), which revealed a dimer that formed a pocket that could contain lipid. This region of the protein is highly conserved in the vitellogenins (Hagedorn et al. 1998).

Amino acid compositional analysis of the *Ae. aegypti* vitellogenin proteins revealed specific amino acid usage. All mosquito vitellogenin proteins, except for two vitellogenins cloned from *Cx. quinquefasciatus*, have a higher content of the aromatic amino acid residues tyrosine

and phenylalanine than vitellogenins of other organisms. It is also interesting that the amino acid composition of all complete sequences available for non-vitellogenin proteins of *Ae. aegypti* shows that none of them exhibits the high content of aromatic residues seen in vitellogenins, except for the hexamerin storage proteins. The biological roles of storage proteins have been investigated, especially in several holometabolous insects. Two important physiological processes, metamorphosis and female reproduction, have been shown to be linked with the utilization of hexamerin storage proteins (Haunerland 1996). In holometabolous insects, storage proteins accumulate before the metamorphic molt of the last larval instar and are known to play an important role in adult cuticle formation by providing tyrosine residues as precursors for tanning.

A female-specific hexamerin storage protein has been identified from autogenous *Oc. atropalpus* mosquitoes, and this protein is believed to provide amino acid residues for vitellogenin protein synthesis by the fat body in the absence of protein from a blood meal (Wheeler and Buck 1996). Assuming that autogenous mosquitoes and other autogenous nematoceran insects, such as chaoborid midges, share common ancestors as is suggested by molecular data (Miller et al. 1997), these ancestors must have relied on storage proteins accumulated from the larval stages for vitellogenin synthesis. Once mosquitoes secured the anautogenous mode of reproduction, the proteinaceous blood meal was utilized for vitellogenin synthesis, and hexamerin storage proteins were no longer utilized. The evolutionarily and structurally unrelated storage proteins and vitellogenins may have a common biological role in providing precursors for embryonic cuticle formation in autogenous and anautogenous mosquitoes as shown in their high content of tyrosine and phenylalanine, although the vitellogenin genes of other insects are not high in these aromatic amino acids. However, why *Cx. quinquefasciatus* vitellogenins break the pattern of high aromatic amino acids in mosquitoes remains an unresolved question.

Synonymous codon usage

The mosquito vitellogenin genes were shown to have high synonymous codon usage bias, preferentially using one or two "optimal" synonymous codons over others. This codon bias

phenomenon is common to highly expressed genes in other organisms. As has been suggested in unicellular organisms such as *E. coli* and *S. cerevisiae*, species-specific optimal codons have been used selectively for efficient translation elongation in highly expressed genes since the abundance of each isoaccepting tRNA reflects the corresponding optimal codon in a species-specific manner (Ikemura 1981; Gouy and Gauthier 1982; Benzene and Hall 1982). Thus, the results obtained in this study are in a good agreement with results from other highly expressed genes in suggesting that the highly expressed vitellogenin genes also exhibit high synonymous codon usage bias (Kurland 1991; Ikemura 1985). The GC content at the third codon position has been regarded as one of the major factors influencing synonymous codon usage bias in *D. melanogaster* (Shields at al. 1988; Powell and Moriyama 1997).

The single vitellogenin gene (Vg-C) cloned from *An. albimanus* shows the highest codon bias among all mosquito vitellogenin genes with an ENC of 32.2. Among four other complete gene sequences available from *An. albimanus*, one of those, heat shock protein 82 gene (hsp82,

accession number: L47285), also exhibits a similar degree of synonymous codon usage bias (ENC: 32.8) with very similar patterns of optimal codon usage. The GC content at the third codon position has been regarded as one of the major factors influencing synonymous codon usage bias in *D. melanogaster* (Shields at al. 1988; Powell and Moriyama 1997). This also appears to be the case with the mosquito vitellogenin genes. An extreme codon usage bias was found in the codons for asparagine, lysine, glutamine, phenylalanine, and tyrosine amino acid residues in Vg-C where over 90% of these amino acids were coded by a single optimal synonymous codon with G or C in the third position. Thus, the bias in highly expressed *An. albimanus* Vg-C and heat shock protein 82 probably resulted from the biased GC content at the third codon position, which is 83.5% and 81.3%, respectively. Although more complete sequences from highly expressed genes in *An. albimanus* are required to deduce optimal codons in this species because the abundance of each tRNA has not been investigated from mosquitoes, other Anopheline species such as *An. gambiae* also show very high GC content at the third codon position (Besansky 1993; Caccone et al. 1999). Vitellogenin genes of other anautogenous mosquitoes, *Ae. aegypti* and *Cx. quinquefasciatus*, also show high codon usage bias and high GC content at the third codon position.

However, vitellogenin genes of two autogenous mosquito species, *Oc. atropalpus* and *Tx. amboinensis*, show lower codon usage bias and GC content at the third codon position compared to those of anautogenous mosquitoes. The reason for the low synonymous codon usage bias of vitellogenin genes in autogenous mosquitoes could be that the amount of vitellogenin synthesized in these autogenous mosquitoes might be significantly smaller than anautogenous egg production, as shown in lower fecundity of facultatively autogenous mosquitoes (Magnarelli 1978; Jones 1993). Therefore, there may be no selective advantage to accumulate optimal codons for faster translation of vitellogenin mRNA. In the case of *Tx. amboinensis,* which live up to 80 days with continuous oviposition (Linley 1987; Tikasingh and Martinez 1991), the translational rate of vitellogenin could be much slower. Thus, the vitellogenin genes in these autogenous mosquitoes may be translated at a lower rate and therefore show lower synonymous codon usage bias.

When the intragenic position of rarely used synonymous codons in mosquito vitellogenin genes was examined, it was found that these codons are not randomly distributed. The conserved position of an extremely rarely used GGG synonymous codon coding for a glycine residue was found in the 12th position within the signal peptide region in all mosquito species examined, except in Vg-C of the autogenous mosquito *Tx. amboinensis* vitellogenin gene, where it was replaced by a preferential codon, GGA, coding for the same glycine residue. While there are about 100 glycine residues coded by the three other synonymous codons present throughout the protein, only one or two GGG codons were used, showing that the isoaccepting tRNA for GGG is very rare in mosquito vitellogenins. Subsequent analysis of intragenic positions of synonymous codons of each mosquito vitellogenin gene revealed that other rarely used synonymous codons such as AAA, ATA, TTA, and CTA have accumulated in the signal peptides of vitellogenin genes, suggesting that selective constraints act on this region to accumulate a cluster of rare synonymous codons in mosquito vitellogenin genes. Although it is not certain whether the presence of a cluster of rarely used synonymous codons in the signal peptide of mosquito vitellogenin genes is the result of selective functional constraints, or of site-specific differences in mutational rates, a potential selective mechanism that could explain the accumulation of rarely used synonymous codons at the N-terminal region of mosquito vitellogenin genes is presented below.

Vitellogenin gene expression in *Ae. aegypti*

The fact that all three members of the vitellogenin gene family in *Ae. aegypti* are expressed in non-blood-fed females, males, larvae and pupae at trace levels, 4 to 6 orders of magnitude lower than in blood fed females, suggests that vitellogenin genes may be constitutively expressed in all stages prior to blood feeding. If there are not enough amino acids available to support the completion of oocyte maturation, it is energetically disadvantageous for female mosquitoes to utilize limited resources to synthesize an amount of vitellogenin proteins that would be insufficient for egg development. Thus, it is possible that a cluster of rare synonymous codons have selectively accumulated at the 5' end of vitellogenin genes in anautogenous mosquitoes to regulate the rate of translation initiation that would down-regulate the level of vitellogenin

protein synthesis in the absence of a blood meal. Once a blood meal is acquired and digested, there are enough amino acid residues available to be charged by tRNAs in lower abundance. In vertebrates, an estrogen-induced increase in specific tRNA for protein synthesis has been observed. For example, the level of specific serine isoaccepting tRNAs are differentially regulated in the liver of the chicken by estrogen during the synthesis of phosvitin, which contains a high number of polyserine residues (Maenpaa and Bernfield 1975). Thus, by extension, it is possible that ecdysone stimulates an increase in tRNA levels, particularly those required to increase translation of rare codons, in addition to its role in inducing transcription. Given the presence of polyserine regions in mosquitoes a similar up regulation of serine tRNA may also occur in response to ecdysone.

Editor's Note

Dr. H. Fred Nijhout, Duke University, acted as editor for this paper.

Acknowledgments

The authors thank Dr. H. Fred Nijhout, Department of Biology, Duke University, for acting as editor for this paper. We would also like to thank two anonymous reviewers for their helpful suggestions. This work was supported by NIH grant HD24869 to H. H. Hagedorn and A. M. Fallon.

References

Adamczyk JJ, Fescemyer HW, Heckel DG, Gahan LJ, Davis RE, Kelly TJ. 1996. Sex-specific and hormone-controlled expression of a vitellogenin-encoding gene in the gypsy moth. *Archives of Insect Biochemistry and Physiology* 31: 237-256.

Besansky NJ. 1993. Codon usage patterns in chromosomal and retrotransposon genes of the mosquito *Anopheles gambiae*. *Insect Molecular Biology* 1:171-178.

Burns DM, Beacham IR. 1985. Rare codons in *Escherichia coli* and s-typhimurium signal sequences. *FEBS Letters* 189: 318-323.

Byrne BM, Gruber M, AB G. 1989. The evolution of yolk proteins. *Progress in Biophysics and Molecular Biology* 53: 33-69.

Caccone A, Garcia BA, Mathiopoulos KD, Min GS, Moriyama EN, Powell JR. 1999. Characterization of the soluble guanylyl cyclase beta-subunit gene in the mosquito *Anopheles gambiae*. *Insect Molecular Biology* 8: 23-30.

Candelas GC, Arroyo G, Carrasco C, Dompenciel R. 1990. Spider silk glands contain a tissue-specific alanine tRNA that accumulates in vitro in response to the stimulus for silk protein synthesis. *Developmental Biology* 140: 215-220.

Cho W-L, Kapitskaya MZ, Raikhel AS. 1995. Mosquito ecdysteroid receptor: analysis of the cDNA and expression during vitellogenesis. *Insect Biochemistry and Molecular Biology* 25: 19-27.

Chen J-S, Raikhel AS. 1996. Subunit cleavage of mosquito pro-vitellogenin by a subtilisin-like convertase. *Proceedings of the National Academy of Science USA* 93: 6186-6190.

Chen J-S, Cho W-L, Raikhel AS. 1994. Analysis of mosquito vitellogenin cDNA, similarity with vertebrate phosvitins and arthropod serum proteins. *Journal of Molecular Biology* 237: 641-647.

Chen J-S, Sappington TW, Raikhel AS. 1996. Extensive sequence conservation among insects, nematodes, and vertebrate vitellogenins reveals ancient common ancestry. *Journal of Molecular Evolution* 44: 440-451.

Chevallier A, Garel JP. 1979. Studies on tRNA adaptation, tRNA turnover, precursor tRNA and tRNA gene distribution in *Bombyx mori* by using two-dimensional polyacrylamide gel electrophoresis. *Biochimie* 61: 245-262.

Deitsch KW, Raikhel AS. 1993. Cloning and analysis of the locus for mosquito vitellogenic carboxypeptidase. *Insect Molecular Biology* 2: 205-213.

Donnell DM. 2004. Vitellogenin of the parasitoid wasp, *Encarsia formosa* (Hymenoptera : Aphelinidae): gene organization and differential use by members of the genus. *Insect Biochemistry and Molecular Biology* 34: 951-961.

Edwards MJ, Severson DW, Hagedorn HH. 1998. Vitelline envelope genes of the yellow fever mosquito, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* 28: 915-925.

Emilsson V, Kurland CG. 1990. Growth rate dependence of transfer RNA abundance in *Escherichia coli*. *EMBO Journal* 9: 4359-4366.

Gemmill RM, Hamblin M, Glaser RL, Racioppi JV, Marx JL, White BN, Calvo JM, Wolfner MF, Hagedorn HH. 1986. Isolation of mosquito vitellogenin genes and induction of expression by 20-hydroxyecdysone. *Insect Biochemistry* 16: 761-774.

Germond JE, Brownleudi M, Walker P, Debony E, Wahli W. 1984. Sequence homologies within the 5'end region of the 4 estrogen-controlled vitellogenin genes in *Xenopus*. *Experientia* 40: 624-624.

Gordadze AV, Korochkina SE, Zakharkin SO, Norton AL, Benes H. 1999. Molecular cloning and expression of two hexamerin cDNAs from the mosquito, *Aedes aegypti*. *Insect Molecular Biology* 8: 55-66.

Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research* 10: 7055-7074.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185: 862-864.

Hagedorn HH, Maddison DR, Tu Z. 1998. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. *Advances in Insect Physiology* 27: 335-384.

Hamblin MT, Marx JL, Wolfner MF, Hagedorn HH. 1987. The vitellogenin gene family of *Aedes aegypti*. *Memorias do Instituto Oswaldo Cruz, Rio de Janeiro* 82 III: 109-114.

Haunerland NH. 1996. Insect storage proteins: Gene families and receptors. *Insect Biochemistry and Molecular Biology* 26: 755-765.

Hirai M, Watanabe D, Kiyota A, Chinzei Y. 1998. Nucleotide sequence of vitellogenin mRNA in the bean bug, *Riptortus clavatus*: analysis of processing in the fat body and ovary. *Insect Biochemistry and Molecular Biology* 28: 537-547.

Holt RA, Subramanian GM, Halpern , 2002. The genome sequence of the malaria mosquito, *Anopheles gambiae*. *Science* 298: 129-149.

Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology* 151: 389-409.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution* 2: 13-34.

Jones CJ. 1993. Larval growth rates and adult reproduction of *Toxorhynchites splendens* (Diptera: Culicidae) with restricted dietary intake. *Environmental Entomology* 22: 174-182.

Kageyama Y, Kinoshita T, Umesono Y, Hatakeyama M, Oishi K. 1994. Cloning of cDNA for vitellogenin of *Athalia rosae* (Hymenoptera) and characterization of the vitellogenin gene expression. *Insect Biochemistry and Molecular Biology* 24: 599-605.

Kanduc D, Grazia di Corcia M, Lucchese A, Natale C. 1997. Enhanced expression of initiator TRNA(Met) in human gastric and colorectal carcinoma. *Biochemistry and Molecular Biology International* 43: 1323-1329.

Kapitskaya M, Wang S, Cress DE, Dhadialla TS, Raikhel AS. 1996. The mosquito ultraspiracle homologue, a partner of ecdysteroid receptor heterodimer: Cloning and characterization of isoforms expressed during vitellogenesis. *Molecular and Cellular Endocrinology* 121: 119-132.

Kokoza VA, Raikhel AS. 1997. Ovarian- and somatic-specific transcripts of the mosquito clathrin heavy chain gene generated by alternative 5'-exon splicing and polyadenylation. *Journal of Biological Chemistry* 272: 1164-1170.

Krakauer DC, Nowak MA. 1999. Evolutionary preservation of redundant duplicated genes. *Seminars in Cell and Developmental Biology* 10: 555-559.

Kurland CG. 1991. Codon bias and gene expression. *FEBS Letters* 285: 165-169.

Lanclos KD, Hamilton TH. 1975. Translation of hormone-induced messenger RNA inamphibian oocytes: I. Induction by estrogen of messenger RNA encoded for vitellogenic protein in the liver of the male African clawed toad (*Xenopus laevis*). *Proceedings of the National Academy of Science USA* 72: 3934-3938.

Liljenstrom H, von Heijne G. 1987. Translation rate modification by preferential codon usage: intragenic position effects. *Journal of Theoretical Biology* 124: 43-55.

Lee JM, Hatakeyama M, Oishi K. 2000. A simple and rapid method for cloning insect vitellogenin cDNAs. *Insect Biochemistry and Molecular Biology* 30: 189-194.

Lin Y, Hamblin MT, Edwards MJ, Barillas-Mury C, Kanost MR, Knipple DC, Wolfner MF, Hagedorn HH. 1993. Structure, expression and hormonal control of genes from the mosquito, *Aedes aegypti*, which encode proteins similar to the vitelline membrane proteins of *Drosophila melanogaster*. *Developmental Biology* 155: 558-568.

Linley JR. 1987. Diel rhythm and lifetime course of oviposition in *Toxorhynchites amboinensis* (Diptera: Culicidae). *Journal of Medical Entomology* 24: 99-105.

Logan SK, Wensink PC. 1990. Ovarian follicle cell enhancers from the *Drosophila* yolk protein genes: differential segments of one enhancer have differrent cell-type speicificities that interact to give normal expression. *Genes and Development* 4: 613-620.

Maddison DL, Maddison WP. 1999. *MacClade*. Version 4.0b. Sinauer Associates.

Martin D, Piulachs MD, Raikhel AS. 2001. A novel GATA factor transcriptionally represses yolk protein precursor genes in the mosquito *Aedes aegypti* via interaction with the CtBP corepressor. *Molecular and Cellular Biology* 21: 164-174.

Maenpaa PH, Bernfield MR. 1975. Subcellular redistribution of seryl-transfer RNA during estrogen-induced phosvitin synthesis and specificity of the estrogen effect. *Biochemistry* 14: 4820-4826.

Magnarelli LA. 1978. Bionomics of the salt-marsh mosquito, *Aedes cantator* (Diptera: Culicidae). *Environmental Entomology* 7: 512-517.

Martin D, Piulachs MD, Comas D, Belles X. 1998. Isolation and sequence of a partial vitellogenin cDNA from the cockroach, *Blattella germanica* (L.) (Dictyoptera, Blattellidae), and characterization of the vitellogenin gene expression. *Archives of Insect Biochemistry and Physiology* 38: 137-146.

Miller BR, Crabtree MB, Savage HM. 1997. Phylogenetic relationships of the Culicomorpha inferred from 18S and 5.8S ribosomal DNA sequences (Diptera: Nematocera). *Insect Molecular Biology* 6: 105-114.

Moriyama EN, Powell JR. 1997. Codon usage bias and tRNA abundance in . *Drosophila*. *Journal of Molecular Evolution* 45: 514-523.

Nardelli D, Gerber-Huber S, Schip FD, Gruber M, Geert AB, Wahli W. 1987. Vertebratae and nematode genes coding for yolk proteins are derived from a common ancestor. *Biochemistry* 26: 6397-6402.

Nose Y, Lee JM, Ueno T, Hatakeyama M, Oishi K. 1997. Cloning of cDNA for vitellogenin of the parasitoid wasp, *Pimpla nipponica* (Hymenoptera: Apocrita: Ichneumonidae): vitellogenin primary structure and evolutionary considerations. *Insect Biochemistry and Molecular Biology* 27: 1047-1056.

Peden J. 1997. *CodonW*, Version 1.3. Available at http://www.molbiol.ox.ac.uk/cu/codonW.html.

Piulachs MD, Guidugli KR, Barchuk AR, Cruz J, Simoes ZL, Belles X. 2003. The vitellogenin of the honey bee, *Apis mellifera*: structural analysis of the cDNA and expression studies. *Insect Biochemistry and Molecular Biology*. 33: 459-465.

Powell JR, Moriyama EN. 1997. Evolution of codon usage bias in . *Drosophila*. *Proceedings of the National Academy of Science USA* 94: 7784-7790.

Raag R, Appelt K, Xuong NH, Banaszak L. 1988. Structure of the lamprey yolk lipid protein complex lipovitellin phosvitin at 2.8-a resolution. *Journal of Molecular Biology*. 200: 553-569.

Racioppi JV, Gemmill RM, Kogan PH, Calvo JM, Hagedorn HH. 1986. Expression and regulation of vitellogenin messenger RNA in the mosquito, *Aedes aegypti*. *Insect Biochemistry* 16: 255-262.

Reinert JF, Harbach RE, Kitching IJ. 2004. Phylogeny and classification of Aedini (Diptera: Culicidae), based on morphological characters of all life stages. *Zoological Journal of the Linnaean Society* 142: 289-368.

Romans P, Tu Z, Ke Z, Hagedorn HH. 1995. Analysis of a vitellogenin gene of the mosquito, *Aedes aegypti* and comparisons to vitellogenins from other organisms. *Insect Biochemistry and Molecular Biology* 25: 939-958.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Springs Harbor, New York: Cold Springs Harbor Laboratory.

Sappington TW, Raikhel AS. 1998. Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochemistry and Molecular Biology* 28: 277-300.

Sappington TW, Kokoza VA, Cho W-L, Raikhel AS. 1996. Molecular characterization of the mosquito vitellogenin receptor reveals unexpected high homology to the *Drosophila* yolk protein receptor. *Proceedings of the National Academy of Science USA* 93: 8934-8939.

Savage HM. 2005. Classification of Mosquitoes in Tribe Aedini (Diptera: Culicidae): Paraphylyphobia, and Classification Versus Cladistic Analysis. *Journal of Medical Entomology* 42: 923-927.

Sharrock WJ, Rosenwasser TA, Gould J, Knott J, Hussey D, Gorden JI, Banaszak L. 1992. Sequence of lamprey vitellogenin. Implications for lipovitellin crystal structure. *Journal of Molecular Biology* 226: 903-907.

Shields DC, Sharp PM, Higgins DG, Wright F. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Molecular Biology and Evolution* 5: 704-716.

Shimeld SM. 1999. Gene function, gene networks and the fate of duplicated genes. *Seminars in Cell and Developmental Biology* 10: 549-553.

Silva R, Fischer AH, Burch JBE. 1989. The major and minor chicken vitellogenin genes are each adjacent to partially deleted pseudogene copies of the other. *Molecular and Cellular Biology* 9: 3557-3562.

Spieth J, Denison K, Kirtland S, Cane J, Blumenthal T. 1985. The *C. elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. *Nucleic Acids Research* 13: 5283-5295.

Spieth J, Nettleton M, Zuckeraprison E, Lea K, Blumenthal T. 1991. Vitellogenin motifs conserved in nematodes and vertebrates. *Journal of Molecular Evolution* 32: 429-438.

Sreenagesh MV, Jayabaskaran C. 1996. Light- and cytokinin-induced changes in the levels of leucine and tyrosine isoaccepting tRNA species and modified nucleotide contents of total tRNA in cucumber seedlings. *Biochemistry Molecular Biology International* 38: 1255-1264.

Sun G, Zhu J, Li C, Tu Z, Raikhel AS. 2002. Two isoforms of the early E74 gene, an Ets transcription factor homologue, are implicated in the ecdysteroid hierarchy governing vitellogenesis of the mosquito, *Aedes aegypti*. *Molecular and Cellular Endocrinology* 190: 147-157.

Swofford, DL. 2000. *PAUP. Phylogenetic Analysis Using Parsimony*. Version 4. Sinauer Associates.

Tata JR. 1988. Regulation of expression of *Xenopus* vitellogenin genes. *Developmental Biology* 5: 241-265.

Thompson JR, Banaszak LJ. 2002. Lipid-protein interactions in lipovitellin. *Biochemistry* 41: 9398-9409.

Tikasingh ES, Martinez R. 1991. Laboratory observations on the fecundity and longevity of *Toxorhynchites moctezuma* (Dyar and Knab) (Diptera: Culicidae). *Journal of the Florida Mosquito Control Assocication* 62: 8-11.

Trewitt PM, Heilmann LJ, Degrugillier SS, Kumaran AK. 1992. The boll weevil vitellogenin gene: nucleotide sequence, structure, and evolutionary relationship to nematode and vertebrate vitellogenin genes. *Journal of Molecular Evolution* 34: 478-492.

Tsang WS, Quackenbush LS, Chow BKC, Tiu SHK, He JG, Chan SM. 2003. Organization of the shrimp vitellogenin gene: evidence of multiple genes and tissue specific expression by the ovary and hepatopancreas. *Gene* 303: 99-109.

Tu Z. 1999. Genomic and evolutionary analysis of Feilai, a diverse family of highly reiterated SINEs in the yellow fever mosquito, *Aedes aegypti*. *Molecular Biology and Evolution* 16: 760-772.

Tufail M, Takeda M. 2002. Vitellogenin of the cockroach, *Leucophaea maderae*: nucleotide sequence, structure and analysis of processing in the fat body and oocytes. *Insect Biochemistry and Molecular Biology* 32: 1469-1476.

Tufail M, Lee JM, Hatakeyama M, Oishi K, Takeda M. 2000. Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera), and its structural and expression analyses. *Archives of Insect Biochemistry and Physiology* 45: 37-46.

Wahli W. 1988. Evolution and expression of vitellogenin genes. *Trends in Genetics* 4: 227-232.

Wheeler DE, Buck NA. 1996. A role for storage proteins in autogenous reproduction in *Aedes atropalpus*. *Journal of Insect Physiology* 42: 961-966.

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

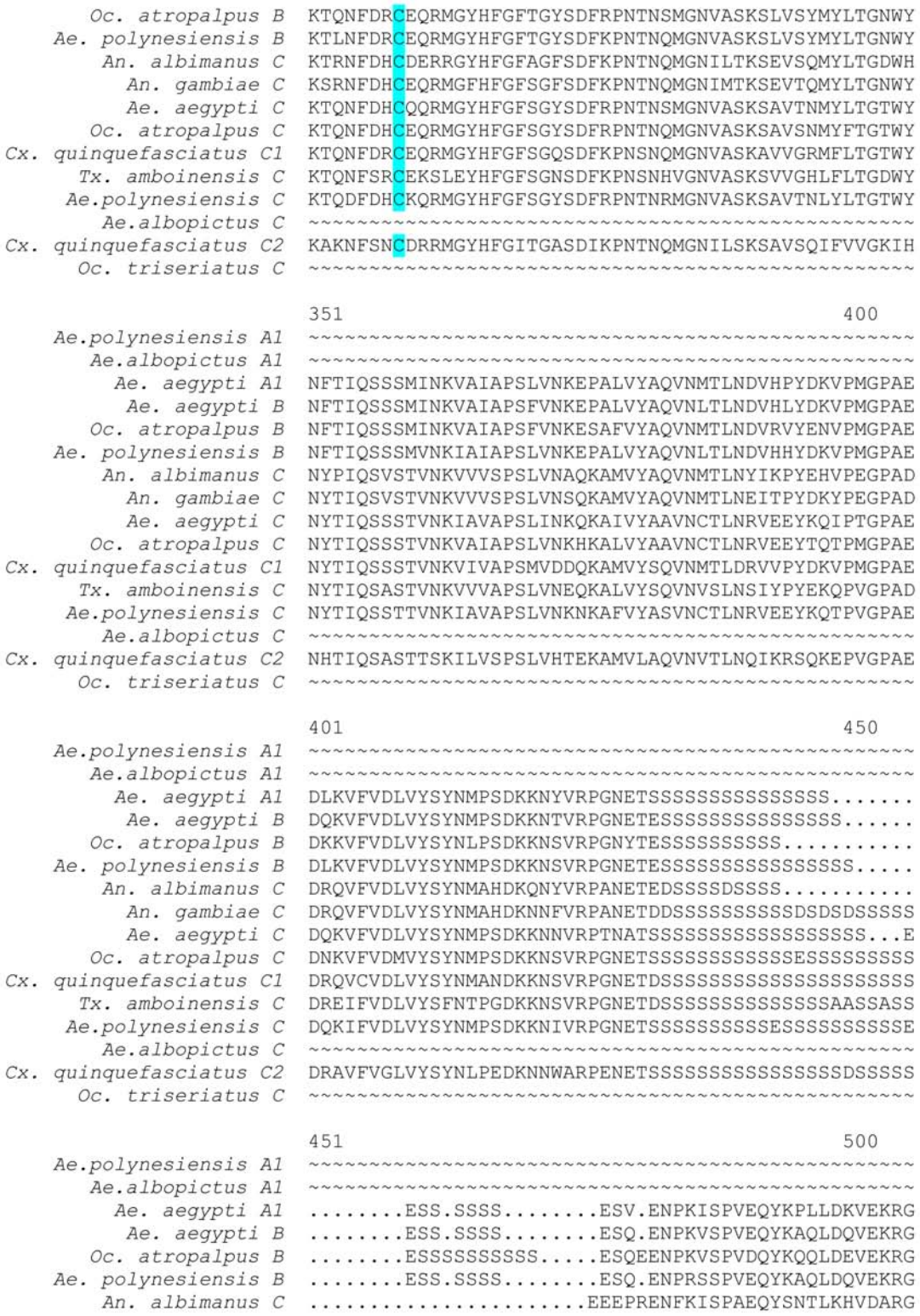
Yano K, Sakurai MT, Watabe S, Izumi S, Tomino S. 1994. Structure and expression of mRNA for vitellogenin in *Bombyx mori*. *Biochimica et Biophysica Acta* 1218: 1-10.

Appendix 1

Alignment of the deduced amino acid sequences of mosquito vitellogenin genes. Conserved cysteine residues are highlighted. (...) indicate gaps inserted in the sequence. (~~~~) indicate missing sequences. Only partial sequences were obtained for *Ae. albopictus, Oc. triseriatus, and Ae.polynesiensis.*

```
                        1                                                50
    Ae.polynesiensis A1  MIAKLLILALAGLTVANQYEN.SFKGYK...N.GFDAG.YK...SYDT..
       Ae.albopictus A1  MLAKLLLLALAGLTVAYQYEN.SFKGYKNGYK.GYDAG.YK.GNGYDAEY
         Ae. aegypti A1  MLAKLLLLALAGLTAAYQYEN.SFKGYNPGYK.GYDAG.YK.GYGYDA..
          Ae. aegypti B  MLTKILLLALVGLTVAYQYESASFKGYKNGFE.GNYPG.YK.GYGYEA..
       Oc. atropalpus B  MIAKLLLLALVGLSAAYQYE....KGYKNGFERGAPSGEYK.NYPYEAGY
    Ae. polynesiensis B  MLTKVLLLALVGLTVAYQSESPYGKGNKNGFE.....GSYKYNYP.....
        An. albimanus C  MIAKLLLLTFVGLCTAYQ.............................
          An. gambiae C  MIAKLLLLTLVGLCTAYQ.............................
          Ae. aegypti C  MLVKLFLLALVGISTAYQ.............................
       Oc. atropalpus C  MIVKLFLLALVGISTAYQ.............................
Cx. quinquefasciatus C1  MFVKLLLLALVGISTAYQ.............................
      Tx. amboinensis C  MLVKLLLLGLVGSGTASQ.............................
     Ae.polynesiensis C  MLVKLLFLALVGISTAYQ.............................
        Ae.albopictus C  MLVKLLFLALVGISTAYQ.............................
Cx. quinquefasciatus C2  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Oc. triseriatus C  MIVKLFFLALVGISTAYQ.............................

                        51                                               100
    Ae.polynesiensis A1  GYKGYGYDAGYKYNNAGY~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  GYKGYGYDAGYKYNNAGYSYKSGFEYGYQNAYQAAFYKYRQ.NMT.DFEF
         Ae. aegypti A1  GYKGYGYDAGYKYNNQGYSYKNGFEYGYQNAYQAAFYKHRP.NVT.EFEF
          Ae. aegypti B  GYK.........YNNQGYSYKNGFEYGYQNAYQAAFYKHRQ.NVT.EFEF
       Oc. atropalpus B  GYKSYG.DKSF.YE..SYSYKNGFEYGYQNAYYGAFYQYRQ.NFTADFEY
    Ae. polynesiensis B  GYKGYGYDAGYKYNNAGYSYKNGFEYGYQNAYQAAFYKYRQ.NTT.EFQF
        An. albimanus C  ..................YSSEYEFPYS.............RPINET.GFEF
          An. gambiae C  ..................YSYEYEFPSS.............RPFNKT.GFEF
          Ae. aegypti C  ..................YSYRSEFPYG.............RPDNKT.GFEF
       Oc. atropalpus C  ..................YSYQNEFYYQ.............RPENKT.GFEF
Cx. quinquefasciatus C1  ..................YEYQNEFYNR.............RPENKT.GFEF
      Tx. amboinensis C  ..................YQYYHEFPYS.............RPENRT.GYEF
     Ae.polynesiensis C  ..................YSYQYKFPYG.............RPDIKN.GFEY
        Ae.albopictus C  ..................YSYQYKFPYG.............RPDIKN.GFEY
Cx. quinquefasciatus C2  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Oc. triseriatus C  ..................YSYQNEFYYD.............RPENKT.GFEY

                        101                                              150
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  SAWKPNYEYIYNVTSVTRTAL~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  SSWMPNYEYVYNVTSKTMTALAELDDQWTGVFTRAYLVIRPKSRDYVVAY
          Ae. aegypti B  SSWMPNREYVYNVTSKTMTALAELDDQWTGVYTRAYLVIRPKSRDYVVAY
       Oc. atropalpus B  SAWMPNYEYVYNVTSKTMTAMAELDDQYTGVFTRAYLVIRPKSPDYVVAY
    Ae. polynesiensis B  SSFMPNHEYVYNVTSVTRTALAELDDQWTGVYTRAILVIRPKSPDYVVAY
        An. albimanus C  GAWEANREYVYNVTTKTMTALPDLEDYWTGIVTHGYLVIRPKDHNYVVAY
          An. gambiae C  GAWEPNKEYVYNVTTKTMTALPDLEDYWTGIVTHGYLVIRPKDHNYVVAY
          Ae. aegypti C  GAWEPNRQYVYNVTSKTMTALPDLEDQWTGTFTRAYLVIRPKSPDYVFGY
       Oc. atropalpus C  GAWEPNRQYVYNVTSKTMTALPDLEDNWTGTVTRAYLVIRPKSPEYVVAY
Cx. quinquefasciatus C1  GAWEPNREYVYNVTSRTMTALADLADQWTGVITRARLIVRPKDPQYVVCY
      Tx. amboinensis C  GAWEPNREYVYNVTSKTMTALADLADQWTGFISRAYLVIRPKNPNYVVAY
     Ae.polynesiensis C  GAWEPNRQYVYNVTSKTMTALPDLEDQWTGTFTRAYLVFRPKSSEYVFGY
        Ae.albopictus C  GAWEPNRQYVYNVTSKTMTALPDLEDQWTGTYTRAYLVFRPKSSEYVFGY
Cx. quinquefasciatus C2  ~AWEPNQVYWYNVTTKTMTALPDVTEQWTGMLTRAKMVINPKSDGYVVGR
      Oc. triseriatus C  GPWEPNREYVYNVTSKTMTALP~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                        151                                              200
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
          Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
            Ae. aegypti A1  VKQPEYAVFNERLPHGYATKFY..HD.MFKFQPMPMSSKPFGIRYHKGAI
              Ae. aegypti B  VKQPEFAVFNERLPYGYATPFY..HD.MFKFQPMPMSSKPFGIRYHKGAI
           Oc. atropalpus B  VKQPEYTVFNERFPQGYATEFY..HD.YFKWVPLPMSSKPFGIRYHKGAI
       Ae. polynesiensis B  VKQPEYAVFNERLPYGYSNYYY.YQN.MLKFQPMPMSSKPFGIRYHKGAI
            An. albimanus C  IDRPTYAVFNEYLPRGYRTKLA.EFD..LKWQPMPFSSKPFGIYYNKGAV
              An. gambiae C  IDRPTYAAFNEYLPRGYRTELS.RFN..LKWQPMPFSSKPFGIYYNKGAV
              Ae. aegypti C  VKQPEYAVFNEYLPQGINTELS.HRS..LKWRPMPMSSKPIAIRYQKGTI
           Oc. atropalpus C  VKEPEYAVFNEYLPHGYNTELA.YHN..LKWRPMPMSSKPIALYYRKAAI
    Cx. quinquefasciatus C1  VKQAEYATFNEELPQGYRTNIWRELNQKLKWQPMPFSAKPFAIRYRKGAI
          Tx. amboinensis C  VKQPEYAVFNEHLHYGYNTEFQ.SSD..FKWQPMPLSTKPFGIVYHKGAI
         Ae.polynesiensis C  VKQPEYAVFNEYLPQGYSTDPL.NNN..FKWRPMPMSSKPFAIRYHNGVI
            Ae.albopictus C  VKQPEYAVFNEHLPQGYNTDPQ.YRN..FKWRPMPMSSKPFAIRYFNGVI
    Cx. quinquefasciatus C2  IDRAHYAQFNQYLADGHRSELS.DLK..LTWKPMPLSSKPFGIKYKKGAI
          Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                             201                                               250
       Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
            Ae. aegypti A1  KGLYVEKTIPNNEVNILKAWISQLQVDTRGANLMHSSKPIHPSKNEWNGH
              Ae. aegypti B  KGLYVEKTIPNNEVNILKAWISQLQVDTRGANLMHSSKPVHPSKNEWNGH
           Oc. atropalpus B  KGLYVEQDIPNNEVNIYKAWASQLQVDTRGANIMHSSKPAHPSKGEWNGY
       Ae. polynesiensis B  KGLYVEKTLPNNEVNILKAWISQLQVDTRGANLMHKSKPIHPTKDEWNGH
            An. albimanus C  KGFYVEKTVPNHEVNMLKAWVSQLQLDTQGAYVIKSEFNQFPENNTLTGV
              An. gambiae C  KGFYVEKTVPNHEVNMLKGWVSQLQLDTQGAYVIKSEFNQFPENNTLTGV
              Ae. aegypti C  KGFYVEQTVPNHEVNILKAWLSQFQLDTQGHHTYKSKYNQFPGNNSFTGV
           Oc. atropalpus C  KGFYVDQSVPNHEVNILKAWFSQFQLDTRGAHMVESKYNQFPGNNSFTGV
    Cx. quinquefasciatus C1  KGLYVEQTVPNHEVNILKAWASQLQLDTRGANVIKSKYNQFPENTTFTGV
          Tx. amboinensis C  KGLMVEQTVPNNEVNILKAWASQFQLDTYGANAIKSKYNQFPENNSFTGV
         Ae.polynesiensis C  KGFYVEKSVPNHEVNILKAWLSQFQLDTQGAHMYKSKYNQVPGNNSFTGM
            Ae.albopictus C  KGFYVEKVWPNHEVNILKAWLSQF~~~~~~~~~~~~~~~~~~~~~~~~~~~
    Cx. quinquefasciatus C2  KGLYVEKTVPNHEVNILKSWVSQLQLDLFGANLIKSKYNQLPENETANAV
          Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                             251                                               300
       Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
            Ae. aegypti A1  YKVMEPLVTGECETHYDVNLIPAYMIQAHKQWVPQGQLRG.EDDQFIQVT
              Ae. aegypti B  YKVMEPLVTGECETHYDVNLVPAYMIQAHKQWVPQGQLRS.QDDQFIQVT
           Oc. atropalpus B  YKVMEPLVTGECETSYDVNLVPHYLIQSHKQWVPQAQLRK.DDDYFIQVV
       Ae. polynesiensis B  YKVMEPLVTGECETHYDVNLIPAYLIQAHKEWVPLEQLRG.EDGQFIQVT
            An. albimanus C  FKTMEPSVTGECETLYDVNAVPEYFIQSHKEWVPQPQYYE.EDQHVFHVV
              An. gambiae C  YKTMEPSVTGECETLYDVNPVPEFHFQSHKEWVPQPQWLE.EDQHVFHVV
              Ae. aegypti C  YKVMEPVVTGKCKTLYDVSVVPPYMIQANKQWVPQPQLRE.EGQYFFQVV
           Oc. atropalpus C  YEVMEPVVTGECKTLYDVNVIPPYIVQANKHWVPQPQLRE.NDQYFFHVA
    Cx. quinquefasciatus C1  YKVMEPLVTGECETLYDVNVVPEHVIKGNKRFVPRPELRE.DNQYFVEVH
          Tx. amboinensis C  YKVMEPQVTGHCETLYDVKVVPEYLVQANKQWVPQPQFRE.KNQYFLEIS
         Ae.polynesiensis C  YEVMEPVVTGKCKTLYDVSVIPPYMVQANKHWVPLPQLRE.EGQYFFQVV
            Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    Cx. quinquefasciatus C2  YKTMEPSVSGECETLYDVNVLPKYKIQSHDEWVPQPQYMQQDDEIF.EIV
          Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                             301                                               350
       Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
            Ae. aegypti A1  KTQNFDRCDQRMGYHFGFTGYSDFRPNTNQMGNVASKSLVSYMYLTGNWY
              Ae. aegypti B  KSLNFDRCEQRMGYHFGFTGYSDFKPNTNQMGNVASKSLVSYMYLTGNWY
```



```
          An. gambiae C  DSSSSSS.................EEEAENFKISTAEQYKKQAKEVERRG
          Ae. aegypti C  SSSSDSSSS...............ESQ.ENPKISPVYQYKAQLDEVEKRG
       Oc. atropalpus C  S.....................ESQEENPKVSPVYQYKEQLDEVEKRG
Cx. quinquefasciatus C1  SSSSESDSSSSSDSSSSSS.....EEQ.KNQKDNPKISPAEQHKQLLKDV
      Tx. amboinensis C  SSSSSSSS.....SEESNK........NDNPKISPAEQHQKKLNAAEKRA
     Ae.polynesiensis C  SSSSE.....................LRENPKVSPVYQYKTQLDEVEKRG
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  DSSSSSSSSSSSSESSSSSSSSSSEEKDNKKISPAEQHKDALKQVEKRE
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         501                                             550
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  N..RYRRDLNAIKEKKYYEAYKMDQYRLHRLNDTSSDSSSSDSSSSSSS.
          Ae. aegypti B  N..RFRRDLNAYKEKKYYEAYKMDQYRLSRKNDTSSDSSSSDDSSSSSSS
       Oc. atropalpus B  N..RNRRDLNAYKEKKYYEAYKMDQYRLSRKNDTSSDSSSSDDSSSSSSS
    Ae. polynesiensis B  N..RFRGDLNAYKEKKYYEAYKMDQYRLSRKNDTSSDSSSSDDSSSSSSS
        An. albimanus C  N..RNRRDLTAQKEKEYYESYKRDQHRLHKENDTSS...SSDDSSSSSSS
          An. gambiae C  N..RNRRDLNAFKEKQYYEAYKRDQYRLRKQNDTSSDSSSSDDSSSSSSS
          Ae. aegypti C  N..RNRRDLNAFKEKKYYEAYKLDQYRLSRKNDTSSDSSSSDDSSSSSSS
       Oc. atropalpus C  N..RNRRDLNAYKEKKYYEAYKMDQYRLNRANDTSSDSSSSDDSSSSSNS
Cx. quinquefasciatus C1  KNRRTRRDLNAQKEKKYYEAYKMDQYRLSRNNDTSSDSSSSDDSKSSSSS
      Tx. amboinensis C  N..RNRRALDSYKEKKYYEAYKKDQYRLRRTNDTSSDSSSSDSSSSSSS.
     Ae.polynesiensis C  N..RNRRDLNAYKEKKYYEAYKMDQYRLSRKNDTSSDSSSSDDSSSSSSS
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  RSTRNRRDLNAQKEKKYYEAYKMDQYRLSRDNDTSSDSSSSDDSSSSSSS
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         551                                             600
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  ....ESKEHRNGTSSYSSSSSSSSSSSSESSSYSSSSSSSSESYSISS.
          Ae. aegypti B  SSS.ESHERNNGTSSDSSSSSSSSSS..YSSSSSSSSSSSSESYSMSS.
       Oc. atropalpus B  ....ESHEQRHNSSDSSSSSSSSSSSS....DSSSS.....ESLSLSS.
    Ae. polynesiensis B  SSS.ESLEHNNSTSSHSSSSSSSSSSS....DSSSSSSSSSESYSMSS.
        An. albimanus C  SSESDEHEHLFSSSSES.........................DSDSLSS.
          An. gambiae C  SSSSESDEHDFYSSSES.........................DSDSLSSE
          Ae. aegypti C  SSSQESNERNNSSSSSSSSSSSSSS....QSYSSSSSS...ESYSLSS.
       Oc. atropalpus C  S...EFQPRGNSTSSSSSSSSSSSSSSSSSESDSSSS....ESDSLSS.
Cx. quinquefasciatus C1  SSSSESNENSKQSGSGSSSSSSSSSSSSSSQSPSSSSSS...ESDSLSS.
      Tx. amboinensis C  ....ESQ.................................SDSLSS.
     Ae.polynesiensis C  SSSQESNERNNSSSSSSSSSSSSSSS...QSYSSSSSS...ESYSLSS.
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  SSSSESREHRKNGTLADNSSSSSSSSSSSSDSKSSSSSSSSSSSESLDSS
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         601                                             650
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  EEYYYQPTPANFSYAPEAPFLPFFTGYKGYNIFYARNVDAIRSVGKLVEE
          Ae. aegypti B  EEYYYQPAPENFKYAPEAPFLPYFTGYKGYNIFYARNVDAQRSVGKLIEE
       Oc. atropalpus B  EEFYYQPAPENFNYAPEAPFLPYFTGYKGFNIYYARNVDATRSVSKLIDE
    Ae. polynesiensis B  EEYYYQPAPENFNFAPEAPFLPYFTGYKGYNVFFARNVDVSHSVGKLVEE
        An. albimanus C  EEYYYQPIPESLKDAPQTPFLPYFTGYKGYSVQYARNVDATHYVYKLAYE
          An. gambiae C  E..FYQPIPESMKDAPQTPFLPYFTGYKGYSVQYAHNVDASRYAYKLAYE
          Ae. aegypti C  EEFYYQPAPESFKDAPQAPFLPFFTGYKGYNIQYARNVDGQRSIYRLVQE
       Oc. atropalpus C  EEFYYQPAPESFKEAPQAPFLPFFTGYKGYNIQYARNVDGQRAIYKLVQE
```

```
Cx. quinquefasciatus C1  EEQYYQPMPESFKEAPQAPFLPFYTGYKGYTVQYARNVDGPRAVSKLVQE
      Tx. amboinensis C  EEAYYQPSPKSFKEAPQTPFLPFHTGYKGYSIHFAHNIDGPSVVSKLAHE
     Ae.polynesiensis C  EEFYYQPAPESFKDAPQAPFLPFFTGYKGYNIQYARNVDGQRSIYKLVQE
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  EE.YYQPSPKDLDNAPAAPLLPFATGYKGSSIQNARNVDAPRIVAQLVKN
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         651                                       700
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  IASDLENPSNLPKANTMSKFNILTRAIRAMGYEDIYELAQKYFVSQKERQ
          Ae. aegypti B  IAEDLQNPSTLPKANTMSKFNILTRAIRAMGHEEMYELAQKYFVSQKERQ
       Oc. atropalpus B  IASDLQNPSTLPKSNTLSKFNIVTRAIRAMNYQEIYDLAQKYFVSQKERQ
    Ae. polynesiensis B  IAEDLVNPSTLPKANTMSKFNILTRAIRAMGHEEMYALAQKYFVSLKERS
        An. albimanus C  IASELQEISQVPKSNTLNKFTILARVLRTMNYQDIYDVCQKLFVSQKERE
          An. gambiae C  IADELQEISQVPKSNTLNKFTILARVLRTMHYQDIYDVCQKLFVSQKERE
          Ae. aegypti C  AVDELQNPSTLPKSNTLSKFNIVSRIFRTMNYQDIYEVAQKYFVSQQERK
       Oc. atropalpus C  AAEELQNPSTLPKSNTLSKFNIVSRIFRTMNYQDIYDIAQKYFVSHKERK
Cx. quinquefasciatus C1  IADELQNPSNLPQSNTLNKFNILCRAIRTMDYKETYEVAQKFFVSQKERQ
      Tx. amboinensis C  IANDLQNPSALPQSNTLNKFNILAGIMRTMPYQELYDVAQKFFVSNKERK
     Ae.polynesiensis C  AVDELQNLSTLPKSNTLSKFNIVSRIFRTMSYQDIYDVAQKYFVSLQERK
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  IAEDFQNPSIIPKSNTLAQFNHLTRLIRTMDHQELYDCAQKLFVSEKERQ
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         701                                       750
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  VAQFSDKKFSKRVDAWVTLRDAVAEAGTPSAFKLIFDFIKEKKLRGYEAA
          Ae. aegypti B  DAKFNDKKFSKRLDAWVTLRDALAEAGTPSAFKLISEYIKEKKLRGVEAA
       Oc. atropalpus B  AAEFSDKKFSKRVDAWAAFRDAVAEAGTPSAFKLISQYIKEKKLRGHEAA
    Ae. polynesiensis B  EALFSDKKFSKRVDAWVTFRDAVAEAGTPAAFKLINEYIKEKKLRGYEAA
        An. albimanus C  EGSNHSESFAKKCDAWNTFRDALAQAGTPAAFKVIKELIEEKKLRGDEAA
          An. gambiae C  EGSNHSESFAKKVDAWNTFRDALAQAGTPPAFKVIKELIEEKKLRGNEAA
          Ae. aegypti C  EGNNKSEKFAKRVDAWIFIRDALAEAGTPPAFKVIKEFIEEKKLRGYEAA
       Oc. atropalpus C  EGDYNSEKFAKRVDCWNFIRDALAETGTPPAFKVIKEFIEEKKLRGSEAA
Cx. quinquefasciatus C1  EGDNHSEKFAKRCDAWNTFRDALAEAGTPPAFKVIKEFIEEKKLRGAEAA
      Tx. amboinensis C  EGDNHSQQFAKRVDAWNTFRDAIAEAGTPPAFQVIKEFIEQKKLQSFEAA
     Ae.polynesiensis C  EGNNHSEKLAKRLDAWYFVRDALAEAGTPPAFKVIKEFIEEKKLRGYEAA
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  QGDKHSEKFAIRCDAWNVFRDAMAEAGTPPVFKVIKQYIEEKKLRGIEAA
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         751                                       800
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  TVIASLAQSIRYPTEHLLHEFFLLVTSDVVLHQEYLNATALFAYSNFVNQ
          Ae. aegypti B  TVVASLAKSIRYPTEQLLHEFFLLVTSDAVQHQEYLNATALFAYSDFVNQ
       Oc. atropalpus B  SVIATLVQSTRYPTEQLLHEFFLLVTSSEVQHQDYLNVTALISYSDFVNQ
    Ae. polynesiensis B  TVVASLAKSIRYPTEKLLHEFFLLVTSDVVQHQEYLNVTALFAYTDFVNQ
        An. albimanus C  TVIATLPKTIRYPTQTIMHEYFLLVTSNTVQHQEYLNVTALMSYCDFLNR
          An. gambiae C  SVIATLPKTIRYPTETVMHEYFLLVTSNAVQHQEYLNTTALISYCDFLNR
          Ae. aegypti C  GVFSTLVNSIRYPTESLLHEFFLLATSDTVQHQEGLNTTALFTYSYFVNQ
       Oc. atropalpus C  SVFSAFAQSIRYPTESLLHEFFLLATSDTVQHQEGLNTTALFTYTYFVNQ
Cx. quinquefasciatus C1  SVIATLPQSIRYPTEPLMHEFFLLVTSDTVQHQDSLNVTALISFTHFVNQ
      Tx. amboinensis C  SVVAMLPQSIRFPTEALMHEFFLLVTSDAVQHQEHLNVTALISFTNFVNQ
     Ae.polynesiensis C  GVFSTLVGSIRYPTESLLHEFFLLATSDTVQHQEGLNTTALFAYSYFVNQ
```

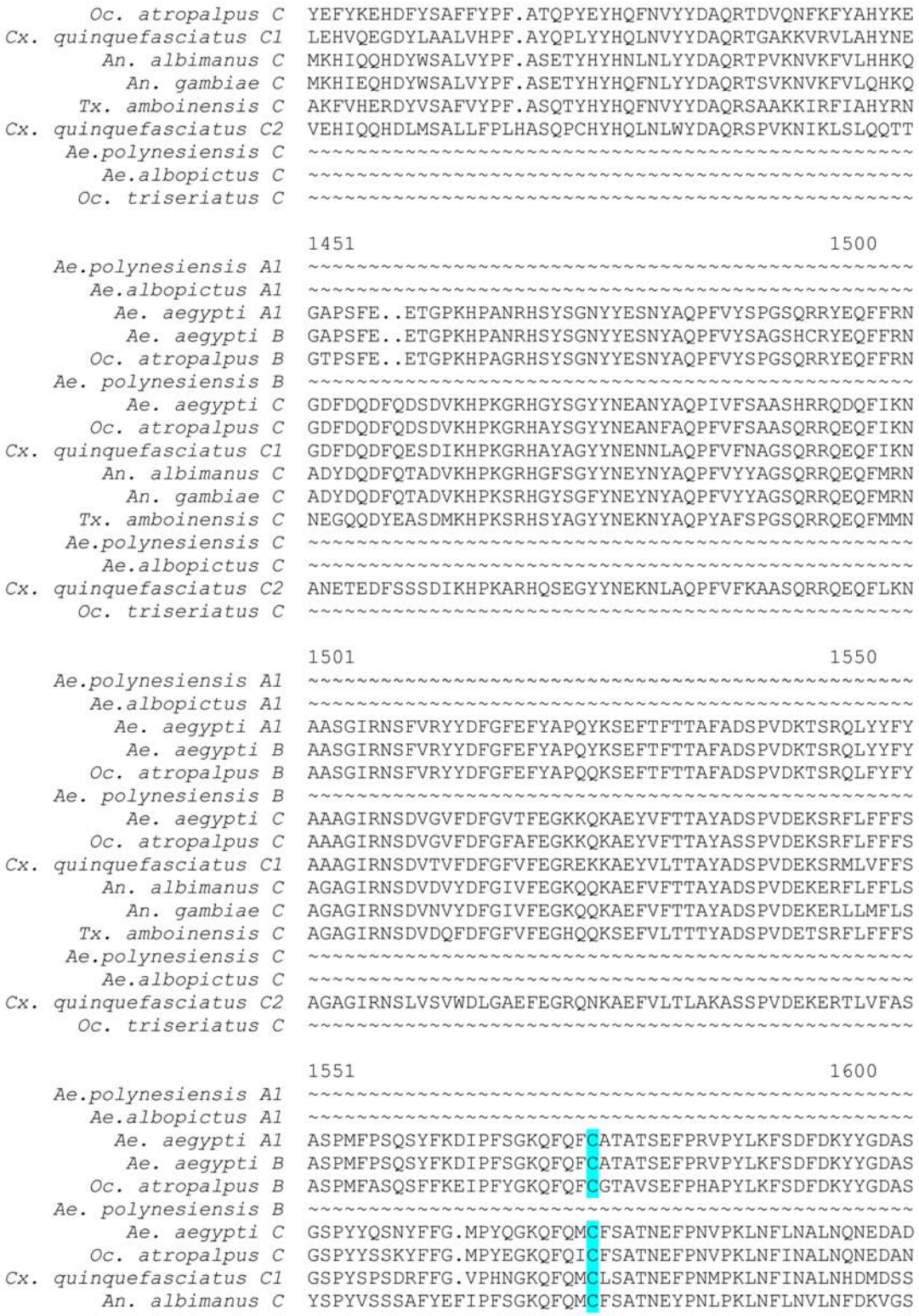
```
         Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  SVVATLPKRIRYPTETLMHEFFLLATSTAVQHQETLNATALIAFTDFLNR
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          801                                             850
   Ae.polynesiensis A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1   AHVSNRSAYNYYPVFSFGRLADADYKIIEHKIVPWFAHQLREAVNEGDSV
          Ae. aegypti B   AHVSNRSAYNYYPVFSFGRLADADYKIIEHKIVPFFAHQLREAVNQEDSV
       Oc. atropalpus B   AHVSNRSAYNYYPVFSFGRLADADYKIIEHKLVPWLAHNLREAVHRGDSV
   Ae. polynesiensis B   AHVSNRSAYNYYPVFSFGRLSDADYKIIEHK~~~~~~~~~~~~~~~~~~~
        An. albimanus C   AQVNNQSAYNYYPVHSFGRLADEDYKIVAHKVVPWLSHQLREAVKAGDSI
          An. gambiae C   AQVNNRSAYNYYPVYSFGRLADADYKIVAHKVVPWFAHQLREAVKAGDSV
          Ae. aegypti C   AHVNNRSAYNYYPVYSFGRFADADYKIVAHKIVPWFSHQLREAVNARDSV
       Oc. atropalpus C   AHVNNRSAYNYYPVYSFGRLADADYKIVAHKIVPWFSHQLREAINEGDSV
Cx. quinquefasciatus C1  AHVSNRSAYNYYPVHSFGRLADSDYKIVAHKIVPWLAHQLREAVNEGDSI
      Tx. amboinensis C   AHVNNRSAYNYYPVHSFGRLANADYPIVAHKIVPWFAHQLREAVNAENSI
    Ae.polynesiensis C   AHVNNRSAYNYYPVYSFGRFADADYKIVAHK~~~~~~~~~~~~~~~~~~~
         Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  AHVNNQSALNYYPVNSFGRLADSKYKIVAHKAVPWLAHQLREAVQEADSE
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          851                                             900
   Ae.polynesiensis A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1   KIQVYIRSLGNLGHPQILSVFEPYLEGTIQITDFQRLAIMVALDNLVIYY
          Ae. aegypti B   KTQVYIRALGNLGHPQILSVFEPYLEGTIKITDFQRLAIMVALDNLVIYY
       Oc. atropalpus B   KTQVYIRTLGNLGHPQILSVFEPYLEGKIPVTNYQRLAMIVSLDKLVVYY
   Ae. polynesiensis B   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        An. albimanus C   KVQVYIRCLGHLGHPEILNVFEPYLEGKIPVTHFQRLAMIVAFDRLVENY
          An. gambiae C   KVQVYIRCLGHLGHPEILNVFEPYLEGKIPVTHFQRLAFIVALDRLVENY
          Ae. aegypti C   KAQVYIRCLGNLGHPEILNVFEPYLEGKYQVSDYQRLAMVVAFDKLVENY
       Oc. atropalpus C   KAQVYIRCLGNLGHPEILNVFEPYMEGKYHVSDFQRLAMVAAFDQLVENY
Cx. quinquefasciatus C1  KIQVYIRALGNLGHPEILNVFEPYLEGKIPVSDFQRLCIVAGMDKLVENF
      Tx. amboinensis C   NIQVYVRALGNLGHPEILNVFEPYMEGKIRVSDFQRLAMVVALDRLAENY
    Ae.polynesiensis C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  RIQVYIRAIGNLGHPEILNVFEPYLEGKIPVTNFQRFAIVMSLDRLVENF
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          901                                             950
   Ae.polynesiensis A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1   PSLARSVLYRAYQNTADVHEVRCAAVHLLMRTDPPADMLQRMAEFTHHDP
          Ae. aegypti B   PSLARSVLYRAYQNTADVHEIRCAAVHLLMRTDPPADMLQRMAEYTHQEP
       Oc. atropalpus B   PELAQAVLFRAYQNTGDVHEIRCAAVHMLMRTDPSAQILQRMAEFTHYDP
   Ae. polynesiensis B   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        An. albimanus C   PRLARSVLFKVYQNTGDAHEVRCAAVYLLVRTKPPVYMLQRMAEQTHYDP
          An. gambiae C   PRLARSVLFKVYQNTGDAHEVRCAAVYLLIRTKPPVYMLQRMAEQTHYDP
          Ae. aegypti C   PHLARSILYKVYQNIGDIHQIRCAAVHMLMRANPPADMLQRMAEYTYYDP
       Oc. atropalpus C   PHLARSVLFRVYQNIGDVHQIRCAAVHLLMRANPPADMLQRMAEHTHYDP
Cx. quinquefasciatus C1  PKLARSVLFKVYQNTGDVHEVRCAAVHLLIRAEPQIEMMQRMAQQTNEEP
      Tx. amboinensis C   PKLARSVLFKVYQNTGDVHQIRCAAVHFLMRTEPPAEMFQRMAEHTHSDP
    Ae.polynesiensis C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  PKLARTVLYRVYQNNADVDEVRCAAAMLLMRTTPPVAMLQRMAEKTDENN
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                          951                                        1000
    Ae.polynesiensis A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1   RLYVRAAVKSAIETAALADDYDEDSKLALNAKAAINFLNPEDVSIQYSFN
          Ae. aegypti B   SRYVRYAVKSAIETAALADEYDNYSDLAVNAKAAVNFLYPEDSSVKYSIN
       Oc. atropalpus B   SQYVRAAVKSAIETATVADEYDNYSKLAENAKAAAPFLYPEDFSSQYSFN
    Ae. polynesiensis B   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        An. albimanus C   STYVRAAVKTALESASEADEFDDDDEFWQNAQAAIKHLNPRDFSLQYSGT
          An. gambiae C   STYVRAAVKTALESASEADEFDDDYEFSQNAQAAVKHLNPRDFSLQYSGT
          Ae. aegypti C   SRYVRAAVKSALESAAESYDYDYYNEFAENAKAAVKFLNPEDFSFQYSSS
       Oc. atropalpus C   SKYVRAAVKTALESAAEAYDYDQYNEFAENAKASIKILNPEDFSFQYSST
Cx. quinquefasciatus C1   SLYVRAAVKTALEYASQADEYEADRTFANNAKAAIKLLDPEVYGLQYSSN
      Tx. amboinensis C   SKFVRAAVKSAIESAAEADEYDNESELAENAKAALKFLNPEDFSLQYSSS
     Ae.polynesiensis C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2   SPQVSALVKSLIESAANTDEFDDDSELAQNARAAVKMLNPNEYGLQYSSA
      Oc. triseriatus C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          1001                                       1050
    Ae.polynesiensis A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1   HIRDYALENLELSYRLHYGEIASNDHRYPSGLFYHLRQNFGGFKKYTSFY
          Ae. aegypti B   HIRDYAMENLELTYRLHYGEIASNDHRYPSGMFYHMRQNFGGFKKYTSFY
       Oc. atropalpus B   HIRDYALENLEMTYRLHYGEIASGDHYYPNGVFYHLRQNFGGFKKYTSFY
    Ae. polynesiensis B   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C   YIRDYAFENQEMSYRMYYGQIAADDHVMPNGMFFQLRNNFGGYKKYSSSY
       Oc. atropalpus C   YIRDYAFENLNLAYRMYYGQVAADDHYVPNGMFFQLRNNFGGFKKYTTFY
Cx. quinquefasciatus C1   YLRDIALNNLEMAYRMYVGQIASDDHWLPNGMFVHLRKNLGGVKRHTTFW
        An. albimanus C   YLRDFAFKELELSYRLYFSQIASDDHYVPSGFFFHLRKNLGGLKRFSTFY
          An. gambiae C   FLRDFAFKELELSYRMYFSQIAADDHYVPSGFFFHLRKNMGGLKRFSTFY
      Tx. amboinensis C   YVQDYYMEQMELSFRAYIGQFASENHFVPHGFFYQLRKNMGGLKRYSTFY
     Ae.polynesiensis C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2   HFRQYAMKELDISYRLQAGQIASDNHPVPTGAWLHWHENLGGLKRLSSYH
      Oc. triseriatus C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          1051                                       1100
    Ae.polynesiensis A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1   YLVSSMEAFFDIFKKQYNTKYFADYYKSADYSTNYYNFDKYSKYYKQYYY
          Ae. aegypti B   YLVSSMEAFFDVFSKQYNTKYFADYYKSADYSTNFYNYDKYSKYYKQYFY
       Oc. atropalpus B   FLVSSMEAFFDVFNQQYSTKYFEDYYKSSDYSTNFYNYDKYFQYYKQNYY
    Ae. polynesiensis B   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C   YLFSSMEAFFDLVDKQCDRSYFKDDYKSSDYYYKYYKQFPNKKSEYFDKY
       Oc. atropalpus C   YLFSSMESFFDLIDKQYDSSYFQEDYKSADYYYKYYKKFPNKKSEYFDKY
Cx. quinquefasciatus C1   YLVSSMETFLDLVDKQYDSSNISEDYKSADYYYKYYQQFPEKKSEYFEKY
        An. albimanus C   YLVSSRN.FFDILDKQYDTYNKHEEYKSNDYYYNYYKQYPHLFKDYFSKY
          An. gambiae C   YLISSMETFFDLLDKQYDSYNKHQEYKSSDYYYKYYKQYPHLFKDYFSQY
      Tx. amboinensis C   YLFSNMESFFDLMDNQFDTSKLRNANK.............PASKSAKHHMY
     Ae.polynesiensis C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2   YIVSNMDALFDLLDNKVMTVEEQKKEWRQESRQSRAN.............
      Oc. triseriatus C   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          1101                                       1150
    Ae.polynesiensis A1   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

 ISSN: 1536-2442

```
        Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti A1  SKDSEYYQKFYGQK.KDYYNDKE.PFKFTAPRIAKLLNIDAEEAEQLEGQ
           Ae. aegypti B  NKDNEYYQKYNGQK.KNFYSDKE.PSKFTASRIGKLLNIDAQEAEQLEGQ
         Oc. atropalpus B  NKNNQYYEKYYGKSKDSAYSEKE.PPKYTAARIAKLLNIDAEEAEQLEGQ
     Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
           Ae. aegypti C  YKSHGPQSDYYQKFAKAEYNDKE.PQKYSTTRIAKLLNIDPREAEELEGQ
        Oc. atropalpus C  YQFHGPQSDYYQKFSKSEYSDKE.PQKYSTTRIAKLLNIDPQEAEELEGQ
 Cx. quinquefasciatus C1  FKTHESQNEMNKKY..NEQNSKY.FQKYSTTRIAKLLNIDPQEAEEMEGQ
         An. albimanus C  SKNHKYQNDYFEQF..GNKNQED.FQKWSTTRIAKMLNIDPEEAEELEGQ
          An. gambiae C  NKNHKYQNDYYEQF..GNKNQEE.FQKWSTTRIAKLLNIDPEEAEELEGQ
       Tx. amboinensis C  MK...SEHGGYHSSSKSKYGDNE.QQNWSTSRIASLLNIDPEEAQELEGQ
      Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 Cx. quinquefasciatus C2  ..........EKQEKASKEQDNKAEQKWSTGRIAKLLNIDPEDVEQVEGQ
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          1151                                          1200
     Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti A1  LLFKLFNGYFFTAFDNQTIENLPHKMRHLFENLEDGYAFDVTKFYQQQDV
           Ae. aegypti B  LLFKLFNGYFFTAFDNQTIENIPRKVKHLFEDLEDGYAFDFTKFYQQQDV
        Oc. atropalpus B  FLFKLFNGYFFQAFDNKTIENIPSKLKHLFHDLEDGYSFDFTKFYQQQDV
     Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
           Ae. aegypti C  FLVKLFNGYHFYAFNNQTIENSPQYIKKLFRELEDGLNFNYTKFYQQEEA
        Oc. atropalpus C  FMMKLFNGYYFYAFDNQTIENAPRYLKEWFEDLENGWNFNLTKFYQQQEV
 Cx. quinquefasciatus C1  LLLPLFNGKYFAAFNNQTIENIPRKLKELCDDLEDGMSFNITKFYQYQPL
         An. albimanus C  FLFQIFNGERFFAFNNQTIEQFPSVAKKFFADFEDGYAYNITKFYQQQAV
           An. gambiae C  FMFQIFNGERFFAFNNQTIEQFPSLVKKYFEDFEDGFAYNMTKFYQQNVV
       Tx. amboinensis C  FMMPLFNGEYFYAFNNQTVEDVPRRARQFFSDLEDGVRKNFTKFYQQQVV
      Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 Cx. quinquefasciatus C2  LWLEIFNAPHLIAFDNNTIDELPRVIKKFMKDLEVNWSVNITKVYQQGLV
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          1201                                          1250
     Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti A1  VLAWPLATGFPFIYTLKAPTVFKFEVDASAKTHPQVYKMPAGHPETENDD
           Ae. aegypti B  VLAWPLATGFPFIYTLKVPTVVKFELDASAKTHPQVYKQPAGHPENENDD
        Oc. atropalpus B  VLAWPLATGFQFLYTLKTPTVFKFEFDANAKTFPQIHKQPTGHPENENDD
     Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
           Ae. aegypti C  SLAFPLATGFPFVYTLKTPTVFKFETEAKVKTYPRVHQKPTGHPENENDD
        Oc. atropalpus C  SLCFPLATGFPFVYTLKTPTVFKFETEARVKTYPSIHETPTGNPENENDD
 Cx. quinquefasciatus C1  TLAFPLASGLPFVYTLKNPTVVKIETEATIKTHPSIVKKPAGHPETENDD
         An. albimanus C  SIAFPLATGLPFTYTLKTPTMVKFEFETTATTYPSIFQTPTGYPEKEFDD
           An. gambiae C  TMAFPLATGLPFHYSLKTPTLMKFEFEASATTYPSIFKTPTGYPEKENDD
       Tx. amboinensis C  TLAFPLASGVPFVYSLKTPTVVKFETEATIKTYPNVANQPAGHPETENDD
      Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 Cx. quinquefasciatus C2  TVAMPLETGFPFTFSTQSPTLVKLEVDASAETKPNMARKPAGHPENGNDE
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          1251                                          1300
     Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti A1  FFYMPQSINGSVDVNLLYHRMVDAKVGFVTPFDHQRYIAGYQKKLHGYLP
           Ae. aegypti B  FFYMPQTLNGSMDVNLLYHRMVDAKVGFVTPFDHQRYIAGYQKKLHGYLP
```

```
       Oc. atropalpus B  FFYFPQVLNGSVDVNLLYHRMVDAKVGFITPFDHQRYIAGYQKKLQGYLP
    Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C  FIYWPKLLNGSIDVNLLYHRMSDAKVGFVTPFDHQRYIAGYQKKLQVNLP
       Oc. atropalpus C  FIYWPKVLNGSIDVNFLYHRMVDAKVGFVTPFDHQRYISGYQKKFYTHLP
Cx. quinquefasciatus C1  FIHVPRVLNGSVDANLVFHRLIDSKIGFITPFDHQRYISGVQKKIQVFLP
        An. albimanus C  FIRFPRWFNGSADVNVAYSRLVDAKVGFITPFDHQRYVAGYQKKMQGYLP
          An. gambiae C  FIHMPRWFNGSADVNMAYSRLVDAKVGFITPFDHQRYVAGYQKKFQGYLP
      Tx. amboinensis C  FIHWPQAINGSIDVNMLYHRLVDAKVGFITPFNHQRYIAGYHKKFQAYFP
     Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  HIHIPLFANVTADVNVVYSRLIDAKVGFNTPFDHQRYIAGFQKKHHIHAP
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         1301                                        1350
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  FNVELGLDFVKDEYEFEFKFLEPKDDHLLFHMSSWPYTGYKDITDMRPIA
          Ae. aegypti B  FTVDLGLDLVNDEYEFEFKFQEPKEDHLMFHMSSWPYTGYKDITDMRPIA
       Oc. atropalpus B  ISVELSMDFVNDEYEFEFEFQDPKEDHLLFHMSSWPYTGYKDITDMRPIA
    Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C  FSVDLSLNFETDEYEIELSPLEPKKDQLFFHLSSWPYTGYKDITEMRPIA
       Oc. atropalpus C  FSVDLSFDFESHEYEFDFSPLKPKEDQLLFHMSSWPYTGYKDITEMRPIA
Cx. quinquefasciatus C1  LSLEFNWDLENDQYETELEILDKKEDKLLFHLSAWPYTGFKDITDMRPIS
        An. albimanus C  FSFDFGFDFENNDYEVNVQPLEPHKDVLLFHLSSWPYTGYKDITDLRPMS
          An. gambiae C  FSFDFGFDFENNDFEVNVQPLEPKKDVLLFHMSSWPYTGYKDIADLRPMA
      Tx. amboinensis C  ASIEFSFDFQNDRYEFEFEPLQPKEDQLLIHLSSWPYTGYKDITDVRPIA
     Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  LRLEAQLDNAQNEYELNIQPLEPKKDILLAHISSWPYTAYKDITDIRPIA
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         1351                                        1400
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  ENPNAKIVHDDNQSTKTMEHTFGQDMTGVALRFHAKYDFDLIN...FQQF
          Ae. aegypti B  ENPNAKIVHDEDQTTKSLEHTFGQDMTGVALRFHAKYDFDMMN...FQKV
       Oc. atropalpus B  ESPNAKIVHDEDQTTKTFDYTFGQDMTGVALRFHAKYDDDFIN...FPVF
    Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C  DSPNVRIVHDDDQTTKSFEHTFGQDMTGFGLRFHAKYDQDFFN...FEQF
       Oc. atropalpus C  ENPNFKIVHDDDQSTKYFEHTFGPEMTGFALRCHAKYDQDFFNFFSFERF
Cx. quinquefasciatus C1  ENPNAKIIHANEHNTQTFQQTVGQDMFGFALRVHAKYDEDIID...VQSI
        An. albimanus C  EQPSVRVLHDRQQTTKSYEQTFGERLFGVAFRFQAKYDKDFID...YAYF
          An. gambiae C  EQPSVHILHDRAQTTKSFEQSFGQSLTGVALRFQAKYDKDFID...YAYL
      Tx. amboinensis C  ENANAKIVHNEQQTTKTFQHTFGQHFTGMALRFYGKYDADFFD...IGRI
     Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  ESPNAQILHDAVRRTKTVEGTIGQQLTGVALRYQAKYDKPALV...FGDI
      Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         1401                                        1450
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  WSLIQKNDFVSAVNYPF.AYQPYEYHQFNLFYDSQRTHAKSFKFFAYQKF
          Ae. aegypti B  FGMLKKHDLVSAVTYPF.AYQPYEYHQFNLYYDAQRTHAKSFKFFAYQKF
       Oc. atropalpus B  WNMLNKHDFVSAVNYPF.AYQPYHYHQFNLYYDSQRTHAKSFKFFAYHKF
    Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C  YDFYKQHDFYSAFFYPF.ATQPYEYHQFNVYYDAQRTDIEKIKFSAHYKE
```



```
        An. gambiae C  FSPYVSSSAFFEFIPFSGKQFQMCFSATNQYPNMPKLNFLNVLNFDKVGS
     Tx. amboinensis C  ASPWDASNSIFKDFPFSGQQFQVCLSGTNRFPNVPRLNFLNALNYGTNAS
    Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  ASPYIAVG.SKKQHQ.ACLSLTEKYPSVPMLNYITALQNDVTSEIDLELS
     Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                        1601                                             1650
   Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae. aegypti A1  QYFDFLYGESCQGGAHIAVKGKQKQTGKCREYLRFSDVAKACKEQMANGY
        Ae. aegypti B  QYFDFLYGESCQGGAHIAVKGKQKQTGKYREYLGFSDVAKACKEQMANGY
     Oc. atropalpus B  QYFDFLYGESCQGGAHISVKGKQKQTGKYREYLRFSDVAKACKKQMADGY
   Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti C  LKWELSYGEKCQGGAQVSIKGKLHQTDMYRHHLRTSRVGSTCKDQMDRGF
     Oc. atropalpus C  LKWEFSYGEKCQGGAQFSMKGKLHQSDEYRHRLRISDIGSTCKEQMDNGY
Cx. quinquefasciatus C1  LRWELSYGEKCQGGAQVSMKGKLKQSDEYRHKLRISEVGQRCKQQMDQGR
      An. albimanus C  MNWELSYGEKCQGGSHVSMKGKLAQSEYYRHYLRISEVGQYCKQQMDRGY
        An. gambiae C  MNWELAYGEKCQGGSHVSMKGKLIQSEPYRHFLRISEAGQSCKEQMDQGY
     Tx. amboinensis C  LNWELSYGEKCQGGAHISMRGQLEQTDYYRHYLRESHIGEACKQQMDKGY
    Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  FGEKCAGGAQVSVSGKLRQTDLWRETLRSSAIVKKCKNQMAEGYFALPEC
     Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                        1651                                             1700
   Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae. aegypti A1  YQFEECQQAIDQAYYYDFYDYAIEYKDVGSVAKNLTNKFYNYFQYAFYPY
        Ae. aegypti B  YQMEECQRAIDQAYYYDFYDYSFEYKDVGAVAKNISAKFYDYFQYAFYPY
     Oc. atropalpus B  YQFEECQQAIDQAYYYDFYDYSMEYKDVGAVAKNFSTKFYDYFQYAFYPY
   Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti C  YQLKECQNATRQASYFDQYFFKFDFKNFDSASQNLTYKFFNFFQQFAYPY
     Oc. atropalpus C  YQLKECQNATRQASYFDQYFFKFDYKNFDSYSQNMTYKFFNFFQQFAYPY
Cx. quinquefasciatus C1  FQLQDCQNATRQAGYLDQYKLDVEFKDVGSYARNWTLKVVDWVQHMTYPW
      An. albimanus C  FQLPACQNATRQAGYFDSYSFDFEYKDVSTYAKNMTYKFFDFARYMAFPY
        An. gambiae C  FQLRACQNATRQGGYFDQYSFNFEYKDVSNYAKNLTYQFFDYARYFTFPY
     Tx. amboinensis C  YQFAECQNATRQASYMDKYHFNFEYKDVGHVAKNFSYKLNDFFKFMTYPY
    Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  QNATRLASALDHYTFDIEFKEIPSSVRNMTNKALNWVQSAVITRWEEDCV
     Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                        1701                                             1750
   Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae. aegypti A1  FESNFFYHGKSNYIKAEFEFAPYGDYYNASFFGPSYAFQVQNYPVFNDYS
        Ae. aegypti B  FESNFFYRGKHNQIKAEFEFAPYGDYYNASFFGPSYAFQVKNYPIENDYS
     Oc. atropalpus B  FESNFFYRGKRTQIKAEFEFAPYGDYYNASFYGPSYAFQIQNYPAENDYS
   Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti C  YESNYFYKGKDNQAQFNFELAPYADCFNASFFGPEFAFKVENYPIYNYYY
     Oc. atropalpus C  YESNYFYKGKNNQAQFNFELAPYADYFNASFFGPEFAFKVDNYPIDNYYY
Cx. quinquefasciatus C1  FEPNHVHKGKSNKIEFEFEMSPYGDYLNASAFAPEYAFEIENYPVDSFWA
      An. albimanus C  YSEDFFYQGKHDQFKFDFQLAPYGDFYNASFYGPQYSFEVENYPIDNEYA
        An. gambiae C  WNEDYFFQGKHNQFQIDFQLAPYFDYYNASFYGSDRSFAIQNYPIESEYA
     Tx. amboinensis C  YQSNYLYKGKENKFRFSFDMTPNFDYFNASFYGPEYAFEVNNHPIEDYYL
    Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  SHKGKEGKAQLKIELSPRVSHINVTLATPNRKIEIENLPVENEWMKSLVL
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         1751                                          1800
   Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti A1  TYFPYFFKYTFFPRYQPYYMHRLPSHKPRNRPYYE.LSNYEQFAIFDRKP
         Ae. aegypti B  VYFPYFFKYTFFPRYQPYFLHRLPAHKPRNRPYYE.LSNFEQFAAFSRKP
      Oc. atropalpus B  TYYPYFFKYTFFPQYYPYFLHRLPSHKERSRHYYE.LSDYEQFAVFDRKP
   Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti C  RYFATV.....................HSDLSFFDRFATYAYRGQYH...
      Oc. atropalpus C  RYFASV.....................HSDLSFLDRFFTYAYRGQYT...
Cx. quinquefasciatus C1  RYFTAV.....................HTDLAWYERLGTYAYQGNYK...
       An. albimanus C  RYFFAI.....................HPELDYYERMLTYAYRGNYH...
        An. gambiae C  RYFFSV.....................HPDFDYYERMFNYAYRGNYH...
    Tx. amboinensis C  RYFASV.....................HANMDWYERLATYAYQGDFY...
    Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  V..........................HPDLAWNERLASYAYNGEMN...
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         1801                                          1850
   Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti A1  QYPSCSFSNDNFYTFDNKKYFYDMGECWHAVMYTVKPDYDFYAQQSHFYN
         Ae. aegypti B  QYPSCSFSKDYFYTFDDKKYFYDMGECWHAVLYTVKPDYDFYAQQSHFYN
      Oc. atropalpus B  QYPSCSFSSDYFYTFDDKKYFYDMGDCWHVVLHTVKPDYDFYAQHSHFYN
   Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti C  ..PSCSISSKYVATFDGKTYDYTMGDCWHVVLHTVKPDYEYYAYQSHFQN
      Oc. atropalpus C  ..PSCAVSSKYIATFDGKYYDYTMGDCWHVLLHTVKPDYEYYPYQSHFQN
Cx. quinquefasciatus C1  ..SSCAVSSKYVDTFDGRNYEYNMGECWHVVLHTVKPDYEYYAQQSHFQN
       An. albimanus C  ..PSCVVSNKFVNTFDGKTFDYELGSCWHVLLHTVKPDYYFYAQDSHFMN
        An. gambiae C  ..PSCAVSNKFVNTFDGKTYDYELGNCWHVVLHTVKPDYYFYGQDSHFMN
    Tx. amboinensis C  ..PSCAFSSKYVDTFDGRTYNYQFGDCWHALVYTVKPDYQYYAQQSHFRN
    Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  ..PSCVVAPKYVDTFDGRTYDYETGTCWHVAMHTVKPELEVSPEHSHFYA
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         1851                                          1900
   Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti A1  SDFEYKYKNGFEEYEQFAALARRGSDNQLYFKFLFG.....DNYIEVFPN
         Ae. aegypti B  SDFEYKYKNGFEEYEQFAALARRGSDNQLYFKFLFG.....DNYIEIFPN
      Oc. atropalpus B  SDFEYKYKNGFEEYEQFSALARRGADNQLFFKFLFG.....DNVIDVFPN
   Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti C  ADSEYRFKNGFYEDEQISVLGRSGPSNEMYFKVILGQYKQNDYNIDIIPN
      Oc. atropalpus C  SDYEYRFKNGFDEDQQISILGRSGPQNELYFKVILGQYKQNDYNIDIIPD
Cx. quinquefasciatus C1  ADTEYRFKNGFSEDEQISVLARSGPSQELYFKVILGQYKQNDYNIDILPK
       An. albimanus C  SDYEYNWKNGFSEDEQITILARHGEDNQLYLKAILGQYKQNDYNIDIIPH
        An. gambiae C  SDYEYNWKNGFGEDEQITILARHGEDNQLFLKAILGQYKQNDYNIDIIPH
    Tx. amboinensis C  SDSEYKFKNGFEAYEQFSLLARNGPQNEVYLKFVLGQANENDYIVDVVPN
    Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  SDVEQRWSNGFDEHEQITVLTRTVENNQQHLKVVLGQQEQWDYNIDIVPN
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                          1901                                             1950
   Ae.polynesiensis A1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
     Ae.albopictus A1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae. aegypti A1     NGGVPFVKYNGRPYDISKSNIAHFEYKEGYPSFPFFYAFAYPNKDLEVSF
        Ae. aegypti B     NADTPFVKFNGRPYDISKYNIAHFEYKEGYPSFPFFYAFAYPNKDLEVSF
     Oc. atropalpus B     NGGVPFVYFNDRPYDISQYNIAHFEAMEGYASFPFFYAFAYPNKDVEFSF
  Ae. polynesiensis B     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti C     GADLPKVYINGKPQQIHDKYAVELYTNDNGGEQPLFRCYALPGNELEINI
     Oc. atropalpus C     GSDLPKVYINGKPQQVHEKYAVELYTNDNGGDQPLFRCYALPGKELEISI
Cx. quinquefasciatus C1   GAELPEVFINGKPQQIHEKYAVELYTNDNGGDQPLFRCYALPGKEIEISI
      An. albimanus C     GHELPHVYINGKPQQIHEKYAIEMYTNDDGGDQPLIRVYALPGNELEISF
       An. gambiae C      GHELPMVYINGKPQQIHEKYAVEMYTNDDGGDQPLIRVYALPGNELEISF
    Tx. amboinensis C     GADLPQVYINGQPQQIHEKYAVEIESNDD.GEQPLFRAYALPGKELEFSY
    Ae.polynesiensis C    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2   GAQLPIVYVNDEPLQVHDKYTIPMYTADE.GEQPLVRVHALAGKELVVDI
    Oc. triseriatus C     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          1951                                             2000
   Ae.polynesiensis A1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
     Ae.albopictus A1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae. aegypti A1     FGGKLKFATDGYRARFFSDYSFYNNFVGLCGTNNGEYFDEFVTADQCYMR
        Ae. aegypti B     FGGKLKFATDGYRARFFSDYSFYNNFVGLCGTNDGEYFSDFVTPDQCYMR
     Oc. atropalpus B     FGEKLKFATDGYRARFFSDYSYYNNFVGLCGTNNGESYDEFVTPDQCYMR
  Ae. polynesiensis B     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti C     RNDGLKIVHDGYRARFFADQSYYNNFAGLCGTNNGEYYDDFVTPDQCYMR
     Oc. atropalpus C     RDD~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C1   RDDDVKIVLDGQRARVFADQKYFDNFVGLCGTNNGELYDDYVTPDQCYMS
      An. albimanus C     RDDDIKIVFDGYRARVFADQSYYNNFVGLCGTNNGEGEDDFITPDQCVMR
       An. gambiae C      RDDDIKIVFDGYRARFFADQSYFNNFVGLCGTNNGEGEDDFITPDQCVMR
    Tx. amboinensis C     RDYDLTVVFDGYRARIYADQSYFNDFVGMCGTNNGEYNDDFVTPNHCYMS
    Ae.polynesiensis C    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2   RDGQIVIVCDGYRAQILTGQTFYDNTVGLCGTNNKQEEDDFITPQQCVMR
    Oc. triseriatus C     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          2001                                             2050
   Ae.polynesiensis A1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
     Ae.albopictus A1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae. aegypti A1     KPEFFAASYAITGQNCTGPAKAFNYAYQQKAKQECVKREVYYGDIIYNQE
        Ae. aegypti B     KPEFFAASYALTGQNCTGPAKAFNYAYQQKAKQECVKREVYFGDVIYNQE
     Oc. atropalpus B     KPEFFAASYALTGQNCTGPAKAFNFAYQHKAKEECVKKEVYYGNIIYEQE
  Ae. polynesiensis B     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae. aegypti C     KPEYFAASYAITGQNCTGPAKAFNYAYQQKAKEECVKREVFYGNVIYERD
     Oc. atropalpus C     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C1   QPEYFAASYALTGQNCSGPAKAFNIAYQQKAKEECVKEEVFYGNVISEQE
      An. albimanus C     KPEYFAGSYAIHGLNCSGPASAYYTEYHQKAQEHCVKPQYYFGNVISELE
       An. gambiae C      KPEYFAASYALTGMNCSGPAQAYFTEYHQKAQQHCVKPQYYFGNVISEQE
    Tx. amboinensis C     KPEYFAASYAITAQNCSGPAKAYNIAYQQESKARCVKHDVFYGNVISEQE
    Ae.polynesiensis C    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus C    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2   KPEYFAASWAVTGQNCTGPAKAFAIASQQKQKEACLKVEYMYGNVVSDVE
    Oc. triseriatus C     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                          2051                                             2100
   Ae.polynesiensis A1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  YYHPRYRYYNHNVE.ESSSSSSSSSDSSSSSSSSSEFSSLGRSGSSSS..
          Ae. aegypti B  YYHPRYRYYNHNVE.ESSSSSSSSSDSSSSSSSSSSSSSESKSKSSS...
       Oc. atropalpus B  FYHQRYRYYNHNVD.DLSSSSSSSSSDSSSSSSESKSNDSSSSSSS...
    Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C  FYRQRYRYY..NHNVEESDSSSSSSSSSSDSSSSSSDSSSSSSSESRSR
       Oc. atropalpus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C1  AGRKRYRYY..NHNVEDSSSSSSSSSSSSSSSESDSKSNSDSSSSSSSSS
        An. albimanus C  AGRSRYNYYYKDFDLSDSSSSE....SDESDSSSSSSSS...........
         An. gambiae C  AGRQRYNYYYKDFDLSDSSSSESSSSSDESDDSNSSSS............
      Tx. amboinensis C  AGRMRYRYNHK...IDSSDSSSSSSSSSSSSSSSSSSSSSSSSESNSGS.
     Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  AGRKRYRYYNHNVDSSSSSESDSSSSSSSSESSESNNKSDSSSSSSSSGS
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         2101                                          2150
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  ......SSSSEEQ.....KEFHPHKQEHSMKECPVQHQHQFFEQGDRICF
          Ae. aegypti B  ......SSSSEEQ.....KEYHPHKQSHSLKDCPVQHQHQFFQQGDKVCF
       Oc. atropalpus B  ......DSSSEEN.....KEYHAHKQEHTLKDCPVQHQHQFFQQGDQICF
    Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C  SHSGSSSSSSSEEN....KEYHPHQQQHSVKECEVKHQHQFFEQGSQICF
       Oc. atropalpus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C1  SEEEKITRNINKESKGTYKEFHADRQKHTLKECEIQHQHQIVEQGAKTCI
        An. albimanus C  EED..........HRSP.SEFFAEKQYFTEKECGIQHRVQYIEQGDKICF
          An. gambiae C  EE............RKPNREHFFEKQQYTEKECPVKYQAQYVEQGDKICF
      Tx. amboinensis C  SEH...............KEYHPHKQTQTLKECQIRQQQQMVEQADKICF
     Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  S.....SSSES.......KEHNPAKQKYSGKECDIVHQVQYVERGSEICF
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         2151                                          2200
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  SLRPLPVCHSKCAATEKISKYFDVHCFEKDSTQAKKYKSEIGRGYTPDFK
          Ae. aegypti B  TLRPLPACHSKCVATDKVSKYFDVHCFEKSSPQAEKYKSEIARGYNPDFK
       Oc. atropalpus B  TLRPLPACHSKCAPTDKISKYFDVHCFDKNSSKAQQYKSE~~~~~~~~~~
    Ae. polynesiensis B  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          Ae. aegypti C  SVRPLPACPSHCAATDKTPKYFDVHCRSASDPAAQLYKSQIQKGYSPDMK
       Oc. atropalpus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C1  TKRPLPVCPSHCVASNKLQKYVDVHCHDSDDQSVKLYKNQIAKGYNPDMS
        An. albimanus C  SIRPLPTCSSQCKAFDKIQKYVDVHCRDITDTAAQLFKQQIRKGVNPDMS
          An. gambiae C  TSRPLPTCASQCKATEKAPKYVDVHCRDATDSVAQLYKQQIRKGVNPDMS
      Tx. amboinensis C  TLRPLPACSSRCQPTEKTTKYVDVHCRDASDSVAQLYKKQISRGVNVDLS
     Ae.polynesiensis C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        Ae.albopictus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Cx. quinquefasciatus C2  SKRPLPSCNSRCKAIEVANKYFDYHCRPLEDSIAQMWKEQIRKGVNPDMS
       Oc. triseriatus C  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                         2201                2221
    Ae.polynesiensis A1  ~~~~~~~~~~~~~~~~~~~~~
       Ae.albopictus A1  ~~~~~~~~~~~~~~~~~~~~~
         Ae. aegypti A1  SFAPHKTYKFNYPKSCVYKAY
          Ae. aegypti B  SYASHKTFKFNYPKSCVYKAY
```