J Mol Evol (2009) 68:403–413
DOI 10.1007/s00239-009-9216-7

Patterns of Variation in the Inhibitor of Apoptosis 1 Gene of *Aedes triseriatus*, a Transovarial Vector of La Crosse Virus

Eric T. Beck · Saul Lozano Fuentes · David A. Geske · Carol D. Blair · Barry J. Beaty · William C. Black IV

Received: 24 July 2008 / Accepted: 19 February 2009 / Published online: 24 March 2009

Abstract *Aedes triseriatus* mosquitoes transovarially transmit (TOT) La Crosse virus (LACV) to their offspring with minimal damage to infected ovaries. *Ae. triseriatus* inhibitor of apoptosis 1 (*AtIAP1*) is a candidate gene conditioning the ability to vertically transmit LACV. *AtIAP1* was amplified and sequenced in adult mosquitoes reared from field-collected eggs. Sequence analysis showed that *AtIAP1* has much higher levels of genetic diversity than genes found in other mosquitoes. Despite this large amount of diversity, strong purifying selection of polymorphisms located in the Baculovirus inhibitor of apoptosis repeat (BIR) domains and, to a lesser extent, in the 5′ untranslated region seems to indicate that these portions of *AtIAP1* are the most important. These results indicate that the 5′UTR plays an important role in transcription and translation and that the BIR domains are important functional domains in the protein. Single nucleotide polymorphisms (SNPs) were compared between LACV-positive and -negative mosquitoes to test for associations between segregating sites and the ability to be transovarially infected with LACV. Initial results indicated that five SNPs were associated with TOT of LACV; however, these results were not replicable with larger sample sizes.

Keywords *Aedes* triseriatus · Association mapping · Heated oligonucleotide ligation assay · Inhibitor of apoptosis

Introduction

The Eastern treehole mosquito, *Aedes (Ochlerotatus) triseriatus* (Say), is the primary vector of La Crosse virus (LACV), the leading cause of pediatric arboviral encephalitis in the Unites States (Watts et al. 1972). An important part of the LACV transmission cycle in the field involves ovarian infection in female mosquito and subsequent transovarial and transtadial transmission of the virus to her adult offspring, which are then infected and capable of transmission. Transovarial transmission (TOT) is also an important part of LACV overwintering in temperate climates (Watts et al. 1973, 1974, 1975; Beaty and Thompson 1975; McGaw et al. 1998). TOT refractory and permissive strains of *Ae. triseriatus* have been selected (Graham et al. 1999), and three quantitative trait loci have been mapped and shown to contribute additively to a female mosquito's ability to transovarially transmit LACV (Graham et al. 2003).

For LACV to be transmitted transovarially, the virus must infect but not disrupt ovarian tissues. The LACV s-segment encodes a small nonstructural protein (NSs)

Electronic supplementary material The online version of this article (doi:10.1007/s00239-009-9216-7) contains supplementary material, which is available to authorized users.

E. T. Beck · S. Lozano Fuentes · C. D. Blair · B. J. Beaty · W. C. Black IV (✉)
Arthropod-Borne and Infectious Diseases Laboratory, Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA
e-mail: wcb4@colostate.edu

D. A. Geske
Division of Vector Control, La Crosse County Health Department, La Crosse, WI 54601, USA

Present Address:
E. T. Beck
Department of Pediatric Infectious Disease, Medical College of Wisconsin, Milwaukee, WI 53201, USA

similar to the *Drosophila* pro-apoptotic protein, Reaper (Colon-Ramos et al. 2003). In mammalian cells and tissues, NSs expression or LACV infection may promote apoptosis. In contrast, LACV-induced apoptosis has not been detected in LACV-infected mosquito tissues. A candidate protein that may suppress apoptosis in infected tissues is the *Ae. triseriatus* inhibitor of apoptosis protein 1 (AtIAP1) (Blitvich et al. 2002), which is an ortholog of the well-characterized *Drosophila* inhibitor of apoptosis 1 (DIAP1). DIAP1 ubiquitinates the apical caspase Dronc to stop activation of downstream caspases that would eventually lead to apoptosis (Palaga and Osborne 2002). For apoptosis to occur, Reaper, Hid, Grim, and Sickle proteins must bind at their IAP-binding motifs to the Baculovirus inhibitor of apoptosis repeat (BIR) domains of DIAP1 (Bergmann et al. 2003). This binding blocks the ability of DIAP1 to inactivate Dronc, and the apoptotic cascade begins (Wang et al. 1999; Chai et al. 2000; Liu et al. 2000; Wu et al. 2000). AtIAP1 may act in a similar fashion to DIAP1 to counter the potential apoptotic effect of LACV in mosquitoes.

Previous observations concerning *AtIAP1* have also led us to consider it a candidate gene affecting LACV TOT. LACV is known to scavenge the 5′ methylated guanine cap as well as the adjacent oligonucleotide from host mRNAs to prime transcription of viral mRNAs (Beaty et al. 2000). Dobie et al. (1997) found that LACV predominantly scavenged the cap from an mRNA similar to AtIAP1 in a persistently infected *Ae. albopictus* larval cell line and in *Ae. triseriatus* eggs emerging from diapause (Dobie et al. 1997; Borucki et al. 2002).

The biology of the LACV TOT system provides a unique opportunity to exploit association mapping to determine if specific *AtIAP1* genotypes condition efficient TOT and overwintering. *Ae. triseriatus* eggs were collected from oviposition sites throughout southwestern Wisconsin, southeastern Minnesota, and northeastern Iowa. These were hatched, reared to adults, tested for LACV infection and then separated into TOT+ (infected) and TOT− (uninfected) groups. The *AtIAP1* gene of individual mosquitoes from both groups was amplified by polymerase chain reaction (PCR) and sequenced. The purpose of this study was to determine whether specific polymorphisms in the *AtIAP1* gene condition whether an *Ae. triseriatus* mosquito will become transovarially infected with LACV (in eggs being laid by an infected female mosquito). An association between specific polymorphisms and increased TOT potential would allow mosquito control agencies to focus more effort on controlling *Ae. triseriatus* populations that contain these polymorphisms in a large number of individuals. While testing this hypothesis, several additional genetic analyses were performed on the *AtIAP1* sequence.

Materials and Methods

Mosquito Collection and DNA Extraction

Ae. triseriatus eggs were collected by the La Crosse County Health Department from LACV endemic areas in southwestern Wisconsin, southeastern Minnesota, and northeastern Iowa where La Crosse encephalitis cases have been reported. The eggs were collected from June through August 2004 in cans that were painted black, half filled with tap water, and lined with seed germination article as an oviposition substrate. Five traps were used at each site and placed at or slightly higher than ground level. The egg liners were collected after 10 days and sent to Colorado State University where the eggs were hatched and reared to adults. Adults were killed and assayed for LACV using an immunofluorescence assay (Beaty and Thompson 1975). DNA was extracted from the thorax of each mosquito using the salt-extraction method (Black and DuTeau 1997), dissolved in 200 µl Tris-ethylenediaminetetraacetic acid (EDTA) buffer (10 mM Tris and 1 mM EDTA; pH 8.0), and stored at −70°C.

PCR and DNA Sequencing

The *AtIAP1* gene was amplified from each sample using three overlapping primer sets (Table 1 and Fig. 1). IAP1F and IAP2R amplified the region from the beginning of the 5′ UTR to the middle of the first BIR domain. IAP3F and IAP4R were designed to overlap the region amplified by IAP1F and IAP2R. These primers amplify a region beginning at the first BIR domain to near the end of the second BIR domain. IAP5F and IAP6R amplified a region from the middle of the second BIR domain to the 3′ polyadenylation site of the gene and thus overlapped the domain amplified by IAP3F and IAP4R (Fig. 1). PCR was completed with the following thermocycling parameters: 1 min at 95°C, 1 min at 51°C, and 2 min at 72°C; this was repeated 35 times. Products were separated on a 1%

Table 1 Primer sequences used for PCR amplification of the *AtIAP1* gene

Primer	Sequence (5′–3′)	Optimal annealing temperature (°C)
IAP0F	ACCATAAATGCATCTTCCAC	51
IAP1F	GGACCAAGAGTAGACGAAGAG	51
IAP2R	GCCCGACATAGTAAAAGC	51
IAP3F	GGACGGTTTTGTTCATCA	51
IAP4R	TACCACATGGCATGCTGT	51
IAP5F	CCTCAAGGATTGGGAAGC	51
IAP6R	CCAAAAACGATCACCTTTATTTTA	51

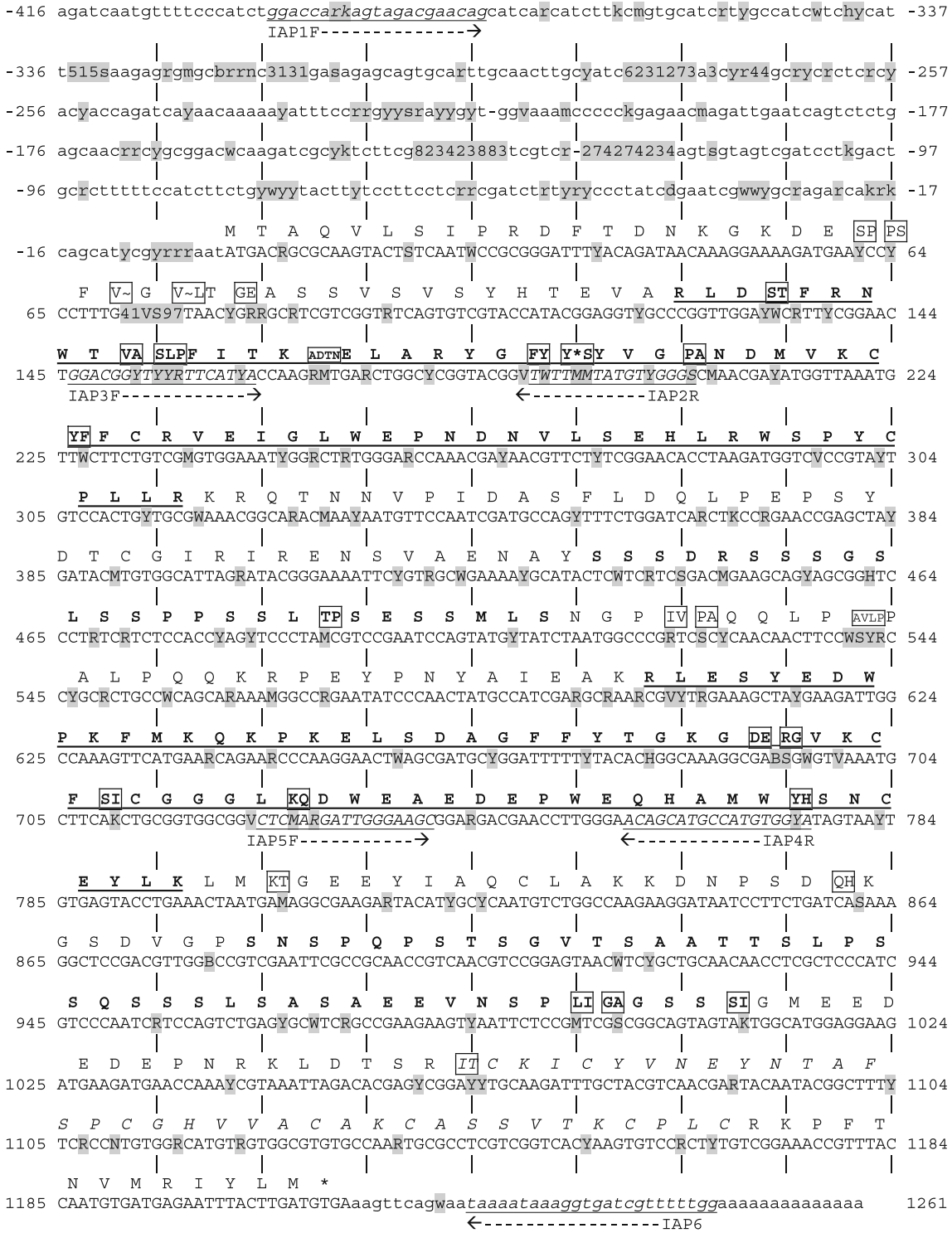

Fig. 1 *Ae. triseriatus* inhibitor of apoptosis 1 gene. Nucleotides in the 5′ UTR are labeled with negative numbers. The 3′ UTR begins at nucleotide 1,213. BIR domain amino acids appear in bold underline (nucleotides 124–318 and 601–798); amino acids in the serine-rich domain appear in bold (nucleotides 436–513 and 883–1,011); and amino acids in the zinc ring finger motif are in italics (nucleotides 1,063–1,170) (Blitvich et al. 2002). Polymorphic nucleotides are highlighted in gray and listed using the following code: R = A or G; K = G or T; M = C or A; Y = C or T; W = A or T; S = G or C; H = A, C, or T; B = C, G, or T; V = A, G, or C; N = A, G, C, or T; 1 = A or –; 2 = C or –; 3 = G or –; 4 = T or –; 5 = G, A, or –; 6 = A, T, or –; 7 = G, C, or –; 8 = C, T, or –; and 9 = A, G, C, or –. Boxed amino acids represent positions where nonsynonymous substitutions occur (amino acids are listed according to proportion with the greatest at the left to the least at the right). Primer sequences are italicized and underlined. Potential QTNs conditioning transovarial infection with LACV, as determined by complete sequence analysis with PGtheta, are found at positions −361, −268, 542, 555, and 570. ∼ = frameshift mutation. * = stop codon

agarose gel containing Tris–acetate–EDTA buffer (40 mM Tris-acetate and 1 mM EDTA; pH 8.3). DNA bands were excised and purified using the Qiaquick Gel Extraction Kit (Qiagen, Valencia, CA). Gel-extracted PCR products were sequenced using both PCR primers at Colorado State University's Macromolecular Resources (Fort Collins, CO). Twenty-two samples had an insertion/deletion polymorphism (one to three bases in length and primarily located in the 5′UTR) in one allele. This is problematic for direct sequencing because PCR amplification of the genome creates two alleles that exist in different reading frames. These samples were cloned into the pCR2.1-Topo vector (Invitrogen, Carlsbad, CA), and the inserts of several clones were sequenced to ensure that data was obtained for each allele. Each PCR product (or plasmid) was sequenced with both forward and reverse PCR primers. Sequence trace files were aligned using Seqman II version 5.01 (DNAstar, Madison, WI). After alignment, the chromatogram ends were trimmed so that only high-quality sequence information was used. In addition, chromatogram alignments were visually analyzed and when there was a discrepancy between forward and reverse sequences, the results were corrected to reflect those indicated by the better-quality sequence. Generally, when this occurred, one of the chromatograms showed a well-defined peak, whereas the other sequence had some anomaly (e.g., high background or a random signal spike). The complete *AtIAP1* gene sequence was ascertained by assembling the three overlapping pieces. Some sequences were missing $\leq$50 base pairs from either end of the gene. This missing data were treated as unknown and were ignored in subsequent analyses. Near-complete sequences were determined for 45 LACV+ and 46 LACV− mosquitoes. Genotypes were recorded using the coding scheme in PGenome (Gorrochotegui-Escalante et al. 2005).

Analysis of Sequence Variability

The computer program DnaSP 4.10 (Rozas et al. 2003) was used for several genetic analyses of *AtIAP1* sequences from 91 individuals (45 LACV+ and 46 LACV−). These sequences were used to estimate nucleotide diversity (π) (Nei 1987), the SD of π (Nei 1987), and the F^* test for neutrality (Fu and Li 1993). In addition, polymorphisms in the coding region of *AtIAP1* were used to calculate the ratio of nonsynonymous-to-synonymous polymorphisms (k_A/k_S) and the transition-to-transversion ratio. Finally, the program compared the level of intraspecific synonymous and nonsynonymous polymorphisms within the *AtIAP1* gene to interspecific polymorphisms between *AtIAP1* and the *Ae. aegypti* and *Ae. albopictus IAP1* genes (*AeIAP1* and *AaIAP1*, respectively) (GenBank accession nos. DQ993355 and AF488809, respectively) using the McDonald-Kreitman test. Neutral-evolution theory predicts that the ratio of synonymous-to-nonsynonymous polymorphisms within one species will be identical in mean to the ratio seen between similar species.

Linkage Disequilibrium Analysis

Linkage disequilibrium among all pairs of segregating sites was analyzed using the program PGLD (Gorrochotegui-Escalante et al. 2005) to calculate Ohta's five D-statistics (Ohta 1982a, b). The disequilibrium between two segregating sites (D^2_{ST}) was estimated, and χ^2 analysis of this result was performed. After applying Bonferroni's correction, a half matrix of the results of pairwise comparisons was plotted. In addition, D^2_{ST} was regressed on the number of nucleotides between segregating sites. A priori sites that are closer together are expected to be in greater disequilibrium than sites that are farther apart. The significance of this regression using Mantel's test was assessed (Mantel 1967).

Analysis of Genotype Frequencies

Wright's F_{IS} summarizes the relation between observed and expected heterozygotes at each segregating site (Wright 1965):

$$F_{IS} = 1 - \left(\frac{H_{o(i)}}{H_{e(i)}}\right),$$

where $H_{o(i)}$ and $H_{e(i)}$, respectively, are the observed and expected frequencies of heterozygotes containing nucleotide i at a segregating site. Weir and Cockerham's f is an estimator of F_{IS} that is unbiased by small or unequal sample sizes (Weir and Cockerham 1984) and is calculated as:

$$f = \frac{b}{b+c},$$

where

$$b = \frac{1}{2(\bar{n}-1)}\left(\sum_y n_y H_{e(iy)} - \frac{(2\bar{n}-1)}{2\bar{n}}\sum_y n_y H_{o(iy)}\right);$$

$$c = \frac{\sum_y n_y H_{o(iy)}}{2\bar{n}};$$

$\bar{n} = \sum_y n_y$/number of collections; $H_{e(iy)} = 1 - \sum_i p_i^2$; p_i is the frequency of nucleotide i at a segregating site; and n_y is the size of collection y.

Association Mapping Based on Allele and Genotype Frequencies

The sequenced *AtIAP1* genes from 91 individual mosquitoes were analyzed using PGTheta (Gorrochotegui-

Table 2 Reporter and detector oligonucleotides used in genotyping *AtIAP1* segregating sites −361, −208, 542, 555, and 570

Oligonucleotide	Sequence (5′–3′)	Optimal ligation temperature (°C)
IAP-361dtcA	Biotin-TGATGATGGCRAYGATGCACT	60
IAP-361dtcC	Biotin-TGATGATGGCRAYGATGCACG	
IAP-361rpt	PO4-GAAAGATGCTGATGCTGT-fluorescein	
IAP-208dtcA	Biotin-TCAATCTTGTTCTCCGGGGGT	51
IAP-208dtcC	Biotin-TCAATCTTGTTCTCCGGGGGG	
IAP-208rpt	PO4-TTTTCCGCRGTTCAGCTC-fluorescein	
IAP542dtcC	Biotin-TTTCTGCTGWGGCAGCGCGGGCG	58
IAP542dtcT	Biotin-TTTCTGCTGWGGCAGCGCGGGCA	
IAP542rpt	PO4-CAGGAAGTTGTTGGGGGA-fluorescein	
IAP555dtcA	Biotin-ATTCYGGCCGTTTCTGCTGT	58
IAP555dtcT	Biotin-ATTCYGGCCGTTTCTGCTGA	
IAP555rpt	PO4-GGCAGCGCGGGCRCAGGA-fluorescein	
IAP570dtcA	Biotin-ATGGCATAGTTGGGATATTCT	58
IAP570dtcG	Biotin-ATGGCATAGTTGGGATATTCC	
IAP570rpt	PO4-GGCCGTTTCTGCTGWGGC-fluorescein	

Table 3 Genetic diversity in individual domains of the *AtIAP1* gene

Domain	Nucleotide position	Nucleotide diversity (π) (range)	θ/site	Average nucleotide differences (k)
5′ UTR	−416 to −1	0.014 (0.000–0.476)	0.04313	4.626
	1–123	0.008 (0.000–0.277)	0.02181	0.990
First BIR domain	124–318	0.010 (0.000–0.494)	0.03106	1.942
	319–435	0.005 (0.000–0.143)	0.02071	0.638
First serine-rich domain	436–513	0.013 (0.000–0.388)	0.02884	1.044
	514–600	0.024 (0.000–0.444)	0.03183	2.100
Second BIR domain	601–798	0.015 (0.000–0.600)	0.02196	2.923
	799–882	0.004 (0.000–0.124)	0.01442	0.326
Second serine-rich domain	883–1,011	0.007 (0.000–0.441)	0.01342	0.856
	1,012–1,062	0.001 (0.000–0.022)	0.00679	0.033
Zinc finger domain	1,063–1,170	0.018 (0.000–0.481)	0.02243	1.956
	1,171–1,212	0.000 (0.000–0.000)	0.00000	0.000
3′ UTR	1,213–1,261	0.000 (0.000–0.000)	0.00000	0.000
Whole gene	−416 to 1,261	0.011 (0.000–0.595)	0.02622	17.432

Escalante et al. 2005). This program compares nucleotide frequencies at segregating sites among, in this case, TOT+ and TOT− mosquitoes (download at http://www.evolcafe.com/popgen/download.html). At each segregating site, θ (Weir and Cockerham 1984) was estimated, and its consistency was assessed with 10,000 permutations according to the procedure of Doerge and Churchill (1996) as follows: The original data set was permuted by randomly assigning the genotype of one mosquito to another mosquito. After all genotypes were shuffled, θ was estimated between phenotypic groups and stored in memory. After 10,000 permutations, all θ were sorted. The 9,500th and 9,900th largest values, respectively, defined the 95% and 99% thresholds at each segregating site. Potential quantitative trait nucleotides (QTNs) were assigned when the original estimate of θ was >95% threshold of θ calculated by permutation. PGCon (Gorrochotegui-Escalante et al. 2005) is a program designed to perform contingency χ^2 analysis of genotypes as segregating sites with adjusted probabilities using Bonferroni's correction to determine if TOT rates are significantly different.

AtIAP1 Heated Oligonucleotide Ligation Assay

The heated oligonucleotide ligation assay (HOLA) (Lynd et al. 2005; Black et al. 2006) was used to determine the genotypes of 300 additional mosquitoes (150 TOT+ and 150 TOT−) at the 5 putative QTNs identified by PGTheta in the analysis of full sequences. These mosquitoes were collected from the same regions as the initial samples. PCR

on each sample used primers IAP0F (located 199-bp upstream of the sequence shown in Fig. 1) and IAP4R to amplify a portion of the *AtIAP1* gene that contained all 5 putative QTNs. HOLA reactions were conducted as described by Black et al. (2006) using the oligonucleotides listed in Table 2.

Genotypic results from HOLA analysis were combined with genotypic results from the previous sequence analysis to yield a total 195 TOT+ and 196 TOT− mosquitoes. Goodness-of-fit χ^2 analyses were performed to determine if any of these five putative QTNs could be used as indicators for susceptibility to transovarial infection with LACV.

Results

Analysis of Sequence Variability

The amplified *AtIAP1* sequence was 1,665 nucleotides in length with 416 bases in the 5′UTR, 1,212 bases in the coding region, and 37 bases in the 3′UTR (Fig. 1). In total, 113 segregating sites were found in the 5′UTR, 144 sites occurred in the coding region, and a single site was found in the 3′UTR, yielding a total of 258 segregating sites (supplemental Table 1). The overall nucleotide diversity (π) was 0.01133, and θ/site was estimated at 0.026 (Table 3). The average number of nucleotide differences (k) among pairs of mosquitoes was 17.432 (Table 3). π varied between 0.000 and 0.595 across the gene (Fig. 2 and Table 3), with π being approximately 13.5 times greater in synonymous versus nonsynonymous substitutions. DnaSP 4.10 does not estimate nucleotide diversity at aligned sites where any sequence is missing data, so the first 53 and the last 25 nucleotides were not included in this analysis. The coding region of *AtIAP1* (nucleotides 1 to 1,212) has a nonsynonymous-to-synonymous polymorphism (k_A/k_S) ratio of 0.235 and a transition-to-transversion ratio of 2.03 (Fig. 3).

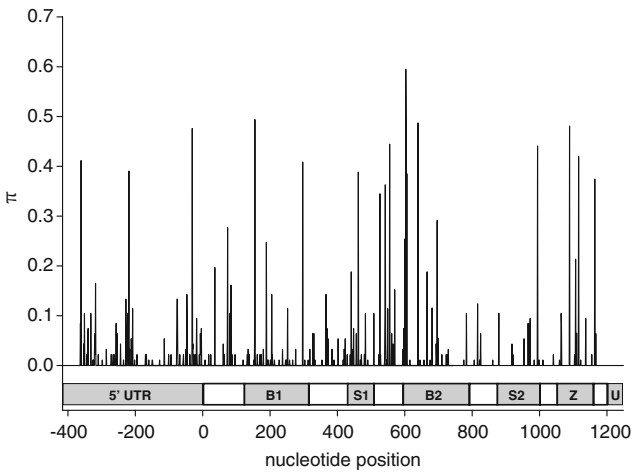

Fig. 2 Plot of nucleotide diversity (π) across *AtIAP1*. This figure shows the nucleotide diversity seen across the entire *AtIAP1* gene. The coding region begins at nucleotide 0 and ends at nucleotide 1,212. 5′UTR = 5′ untranslated region; B1 = first BIR domain; S1 = first serine-rich domain; B2 = second BIR domain; S2 = second serine-rich domain; Z = Zinc ring finger motif; and U = 3′ untranslated region

F_{IS} varies between −1 and 1; a positive value indicates an excess of homozygotes, and a negative value indicates an excess of heterozygotes. Segregating sites in the *AtIAP1* gene had a consistent excess of homozygotes. Of the 258 segregating sites, only 25 had a negative F_{is} (data not shown). The majority of these sites were found within the coding sequence and more specifically near the first serine-rich domain.

Fu and Li's F^* is a normalized comparison of all mutations (η) relative to the number of those appearing once ("singletons" = η_s). The underlying assumption of this test is that $F^* = 0$ ($\eta = \eta_s$) under neutrality: $F^* > 0$ ($\eta > \eta_s$) under balancing selection and $F^* < 0$ ($\eta < \eta_s$) under purifying selection. Our analysis (Fig. 4) shows that the majority of the gene has a negative F^* value; however, these values were only significant in a few small parts of the gene. In particular, the negative F^* values are significant in polymorphisms located in the BIR domains and to a lesser extent in the 5′UTR, suggesting that polymorphisms in these regions may compromise fitness and are rapidly eliminated through purifying selection. These observations are consistent with F_{IS} analysis because an excess of homozygotes is also indicative of purifying selection.

The neutral mutation hypothesis was tested using *IAP1* sequences from *Ae. albopictus* and *Ae. aegypti*. The McDonald–Kreitman (MK) test measures whether there are differences in the ratio of synonymous-to-nonsynonymous polymorphisms within species compared with the ratio between species. The MK test (performed on the 1,227 sites in the coding sequence) showed no significant difference ($P = 0.38$) in the ratio of synonymous-to-nonsynonymous polymorphisms within *Ae. triseriatus* samples (206 synonymous to 56 nonsynonymous) and

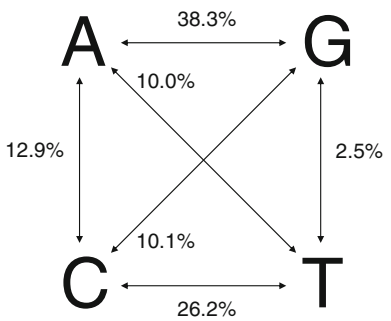

Fig. 3 Percentage of transitions and transversions in the coding sequence of the *AtIAP1* gene

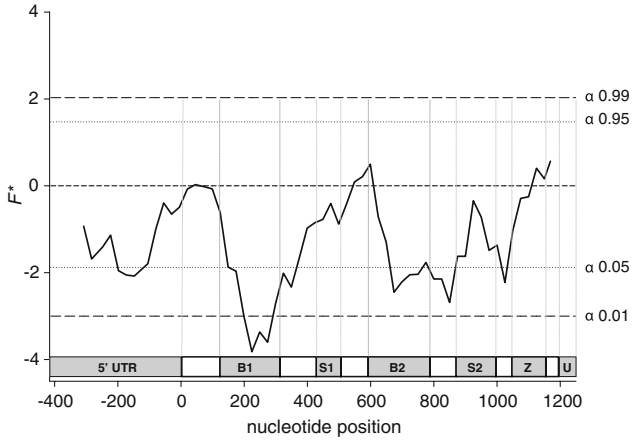

Fig. 4 Plot of Fu and Li's F^* across *AtIAP1*. A significant negative value indicates that polymorphic sites are under purifying selection. A significant positive value indicates that polymorphic sites are under balancing selection. The coding region begins at nucleotide 0 and ends at nucleotide 1,212. 5′UTR = 5′ untranslated region; B1 = first BIR domain; S1 = first serine-rich domain; B2 = second BIR domain; S2 = second serine-rich domain; Z = Zinc ring finger motif; U = 3′ untranslated region

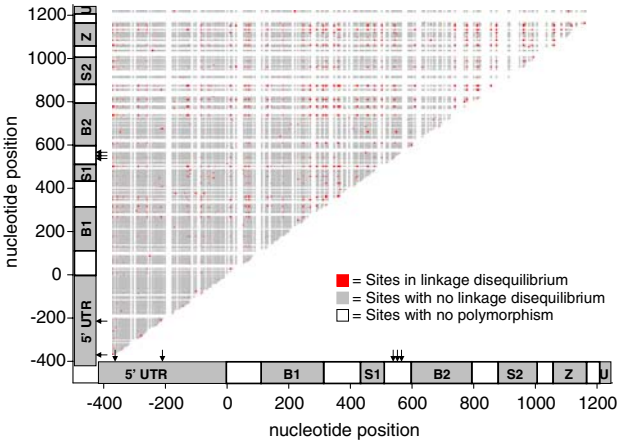

Fig. 5 A half-matrix showing linkage disequilibrium coefficient D^2_{ST} between *AtIAP1* segregating sites. 5′UTR = 5′ untranslated region; B1 = first BIR domain; S1 = first serine-rich domain; B2 = second BIR domain; S2 = second serine-rich domain; Z = Zinc ring finger motif; U = 3′ untranslated region. The coding region begins at nucleotide 0 and ends at nucleotide 1,212. Potential QTNs conditioning transovarial infection with LACV, as determined by complete sequence analysis with PGtheta, are found at positions −361, −268, 542, 555, and 570 and are marked with arrows

between *Ae. triseriatus*, *Ae. aegypti*, and *Ae. albopictus* samples (100 synonymous to 34 nonsynonymous). These results indicate that a general pattern of neutral evolution is responsible for the large amount of polymorphisms seen in the coding region of *AtIAP1*.

Linkage Disequilibrium Analysis

Polymorphisms occurring between the first BIR and serine-rich domains were more likely to be in disequilibrium among themselves than with other segregating sites (Fig. 5). There was little linkage disequilibrium among segregating sites in the 5′UTR. Regression analysis of D^2_{ST} on the number of nucleotides between segregating sites indicated that neither the y-intercept nor the slopes were significantly >0, meaning that the recombination frequency between polymorphisms occurring at nearby sites is similar to the recombination frequency between those polymorphisms that occur farther apart.

Association Mapping Based on Allele and Genotype Frequencies

PGTheta identified five sites (at nucleotides −361, −208, 542, 555, and 570) in which θs estimated from the original data set were >95% of permuted θs (Fig. 6). Two of these putative QTNs were found in the 5′UTR, and the other three were found between the first serine-rich and the second BIR domain. QTNs 555 and 570 encode synonymous substitutions. QTN 542 encodes a transition in the second codon (GCG ⇔ GTG), causing an A ⇔ V amino acid substitution. Because the two other positions in this codon are also polymorphic, leucine and proline are also possible amino acid substitutions. QTNs 542 and 555 were in disequilibrium. The LACV infection rate was significantly different ($P < 0.05$) among genotypes at QTNs −361, 555, and 570 but not at sites −208 and 542 (data not shown).

Association Mapping Based on Single-Nucleotide Polymorphisms

To further evaluate these five putative QTNs, HOLA was performed on an additional 150 LACV+ and 150 LACV− mosquitoes. These results were analyzed separately and then combined with the results from the fully sequenced mosquitoes. The transovarial infection rate was not statistically different among genotypes at any of the putative QTNs (Fig. 7).

Discussion

The genetic analysis of *AtIAP1* showed 258 segregating sites. Of the 144 sites in the coding region, 37 were nonsynonymous. The overall nucleotide diversity (π) was 0.011, and θ/site was 0.026. Figure 2 indicates that variation is uniformly distributed across the gene. Compared with the *Ae. aegypti Early trypsin* gene, π was nearly equivalent at 0.012, but θ/site was 0.010, only one-third that found in *AtIAP1*. In the *Ae. aegypti abundant trypsin*

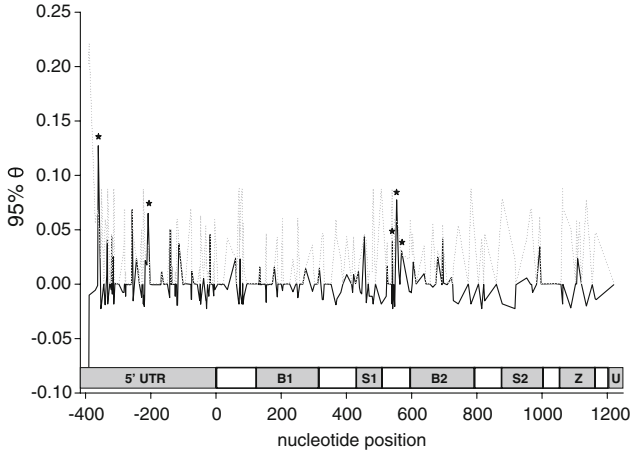

Fig. 6 Association mapping of *AtIAP1* comparing polymorphic nucleotide frequencies between TOT+ and TOT− mosquitoes using PGtheta. An asterisk indicates a position where the estimated θ (as shown by solid lines) is <95% of the permuted θ values (as shown by dotted lines). The coding region begins at nucleotide 0 and ends at nucleotide 1,212. 5′UTR = 5′ untranslated region; B1 = first BIR domain; S1 = first serine-rich domain; B2 = second BIR domain; S2 = second serine-rich domain; Z = Zinc ring finger motif; *U* = 3′ untranslated region

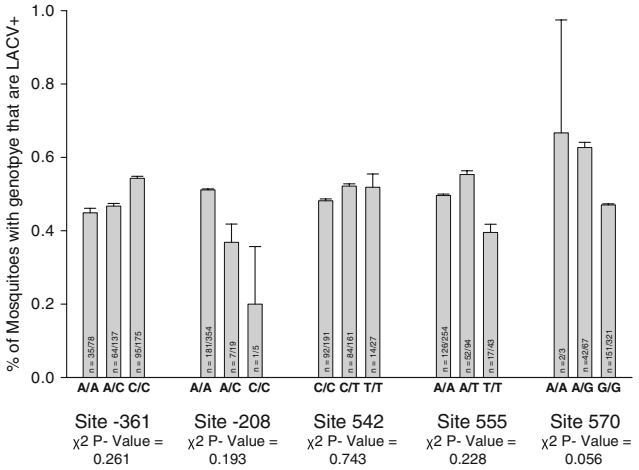

Fig. 7 Association mapping comparing genotype frequencies of LACV+ and LACV− samples at putative QTNs. Potential QTNs conditioning transovarial infection with LACV were identified by complete sequence analysis with PGtheta of 45 LACV+ and 46 LACV− mosquitoes. One hundred fifty LACV+ and 150 LACV− samples, which were previously unsequenced, were genotyped at potential QTNs using the HOLA. Results from samples genotyped by sequencing and by the HOLA were combined and graphed as percent of each genotype that was LACV+. A χ^2 goodness-of-fit analysis was used to determine if LACV infection rates differed between genotypes. The number on each bar represents the number of LACV+ individuals/number of individuals with the genotype. In all, 391 samples were genotyped by either sequence analysis or by the HOLA at each site; however, results from the HOLA analysis were occasionally inconclusive, which is why the number tested does not always add up to 391 at each site

gene, π was only slightly lower at 0.009, and the θ/per site (0.009) was again one-third that found in *AtIAP1*. Thus, the average diversity per site is one-third as large in *AtIAP1*.

Previous studies have indicated that these collections come from a large panmictic population with low levels of genetic drift. Neutral theory predicts that with a large effective population size, the large number of polymorphisms can be maintained only if positive directional selection is weak (Kimura 1983); otherwise the large numbers of singletons and the greater θ/site seen in *AtIAP1* are difficult to explain. This greater diversity could arise through relaxation of selection or through balancing or diversifying selection; however, the high proportion of synonymous substitutions is not consistent with diversifying or balancing selection. Furthermore, the overall negative values of *F** (Fig. 4) indicate that the numbers of singletons exceeded the overall numbers of shared polymorphisms, which is evidence of purifying selection. Significant negative *F** estimates occurred in polymorphisms located in the BIR domains. In *Drosophila*, BIR domains are targeted by proapoptotic caspases to neutralize DIAP1. Amino acid substitutions in this portion of the gene would compromise this function in AtIAP1 and thus probably the fitness of *Ae. triseriatus*.

Analysis of the *AtIAP1*-coding region showed that the ratio of nonsynonymous-to-synonymous polymorphisms (Ka/Ks) is 0.235. This ratio is higher than the Ka-to-Ks ratio seen in the mosquitoes *Anopheles funestus* (0.181) (Wondji et al. 2007), *An. gambiae* (0.192) (Morlais et al. 2004), and *Ae. aegypti* (.204) (Morlais and Severson 2003).The ratio is much higher compared with that of *Drosophila melanogaster* (0.115) (Moriyama and Powell 1996). A higher Ka-to-Ks ratio in *AtIAP1* could result from either more nonsynonymous polymorphisms or fewer synonymous polymorphisms than other mosquito genes. *AtIAP1* has, on average, 1 SNP/8.4 bp in the coding region, whereas *An. funestus*, *An. gambiae*, and *Homo sapiens* have 1 SNP/138, 125, and 1,000 base pairs, respectively (Wang et al. 1998; Morlais et al. 2004; Wondji et al. 2007). This means that greater nonsynonymous polymorphisms, rather than fewer synonymous polymorphisms, are probably responsible for the increased Ka-to-Ks ratio in *AtIAP1*.

An. funestus, *An. gambiae*, *Ae. aegypti*, and *D. melanogaster* all have lower genetic diversity per site (π) in their noncoding regions than in synonymous polymorphisms found in the coding region (Moriyama and Powell 1996; Morlais and Severson 2003; Morlais et al. 2004; Wondji et al. 2007). This trend is also seen in the *AtIAP1* gene, where π per site in the 5′UTR is 0.014, and π per site in synonymous sites is 0.035. This result indicates that the noncoding region is under greater purifying selection than the synonymous sites in the coding region, which makes sense if the 5′UTR plays an important role in transcription or translation of *AtIAP1*.

The BIR region of DIAP1 has been shown to be sufficient for preventing apoptosis, and the BIR domain alone is even more efficient at inhibition of apoptosis then the full-length DIAP1 protein (Hay et al. 1995; Vucic et al. 1998). Perhaps the reason for the high level of diversity seen in the *AtIAP1* gene is that the BIR domains are the only portion of the gene that have any selection pressure acting on them. This could explain why MK tests show a general pattern of neutral selection acting on the coding sequence, whereas Fu and Li's *F** analysis indicates that the *AtIAP1* gene is under relatively little purifying selection apart from the BIR domains.

Several other possible reasons exist for the high level of nucleotide diversity observed in the *AtIAP1* gene. One is that we are actually sequencing the same gene from several *Ae. triseriatus* subpopulations, however, previous studies have shown that there are no barriers to gene flow in the upper Midwest and that the mosquito exists as one panmictic population (Beck et al. 2005). Another possibility is that we have actually collected two different species of mosquitoes. Again, this seems unlikely because *Ae. triseriatus* mosquitoes are relatively easy to identify compared with other mosquitoes present in the study area with one exception. *Ae. hendersoni* can be found in the same geographic area and is closely related (in fact, the two mosquitoes can interbreed to form viable hybrids) (Munstermann et al. 1982). However, few documented examples exist of interspecific hybrids from the field, most likely because these mosquitoes occupy different niches (Truman and Craig 1968; Grimstad et al. 1974). *Ae. hendersoni* mosquitoes feed and breed in tree canopies, whereas *Ae. triseriatus* mosquitoes remain relatively close to the ground (Copeland and Craig 1990). Because our samples were collected at ground level, it is unlikely that we collected *Ae. hendersoni* along with *Ae. triseriatus*. A final possibility is that several *IAP* paralogs are present and that this sequence information represents data from multiple genes in the same family. This possibility is not unrealistic because seven *IAP* genes have been identified in the principal vector of malaria, e.g., *An. gambiae* (Christophides et al. 2002). However, these genes can be distinguished from one another, and throughout the course of this study there was no information to indicate that multiple undistinguishable paralogs were sequenced. This possibility is difficult to test without the genome sequence of this mosquito. Therefore, although it is possible that we sampled multiple subpopulations, species, or IAP paralogs, these explanations seem less likely than the *AtIAP1* gene having high levels of diversity.

Five significant QTNs associated with TOT were detected in a sequence analysis of 91 mosquitoes. However, the subsequent prospective case control study failed to validate any of these 5 QTNs (Fig. 7). Interestingly, 19 of the 258 polymorphic sites observed during sequence analysis were not in Hardy-Weinberg equilibrium (this includes 4 of the potential QTNs with the fifth having a Hardy-Weinberg equilibrium probability of 0.0534) (supplemental Table 1). Although the sequence chromatograms show no obvious reason for doubting their validity, sequencing error is the most likely reason for this divergence from equilibrium. In addition to the positive F_{IS} values, these data indicates that sequence analysis likely misrepresented the genotype of several heterozygotes as homozygotes. This would indicate that the HOLA assay has a lower error rate than direct sequencing of PCR products, and that the follow-up case control study likely represents the truth about the relation of these SNPs and their effect on transovarial transmission. It is also a possibility that the initial sequencing study was performed with too few individuals, thus causing some of these sites to appear as QTNs when in fact they were not. Irrespective of the reasons, this study did not find an association between *AtIAP1* polymorphisms and TOT in field collected *Ae. triseriatus*.

Implicit assumptions in our study design decreased its power to detect valid QTNs. The design assumes that the *Ae. triseriatus* mothers of the offspring in this study were uniformly susceptible to infection with LACV and that all mothers were exposed to LACV. If true, then the only differences among reared TOT+ and TOT− adults would arise from genetic differences among mothers in the genes that condition TOT. The first assumption is probably valid; typically *Ae. triseriatus* are uniformly susceptible to oral or vertical infection with LACV (Woodring et al. 1998). The second assumption is obviously valid with TOT+ offspring but is probably false among TOT− offspring. The minimal field infection rate of *Ae. triseriatus* with LACV is 3.4 to 12.7/1,000, thus making it difficult to verify whether offspring are uninfected because their mothers were never exposed to LACV or because the mothers were genetically incapable of TOT (Clark et al. 1983). This assumption could have been eliminated by returning mosquitoes to the laboratory, uniformly exposing them to LACV with either oral or intrathoracic inoculation, and then collecting eggs and analyzing the resulting offspring. However, this is problematic because *Ae. triseriatus* is a difficult species to colonize directly from the field.

Problems arising from nonuniform exposure of a study group do not preclude association-mapping studies. Human genetic epidemiologists must deal with nonuniform exposure in identifying genetic factors that condition genetic susceptibility to heritable or infectious diseases. Nevertheless, nonuniform exposure does lower the power of association mapping to detect valid QTNs. Only QTNs with large effects on phenotype are likely to be detected.

This study has shown that polymorphisms in the *AtIAP1* gene likely do not have a significant effect on TOT of

LACV. However, these results show that there is much greater diversity seen in this gene compared with genes studied in *D. melanogaster*, *An. funestus*, *An. gambiae*, and *H. sapiens*. These results also indicate that although the *AtIAP1* gene is highly polymorphic, it is generally evolving according to neutral theory. However, the 5′UTR and the two BIR domains are under stronger purifying selection than the remainder of the coding sequence. For this reason, it is likely that the 5′UTR and the BIR domains play major roles in transcription and translation efficiency and protein function, respectively.

References

Beaty BJ, Thompson WH (1975) Emergence of La Crosse virus from endemic foci: fluorescent antibody studies of overwintered *Aedes triseriatus*. Am J Trop Med Hyg 24:685–691

Beaty BJ, Rayms-Keller A, Borucki MK, Blair CD (2000) La Crosse encephalitis virus and mosquitoes: a remarkable relationship. ASM News 66:349–357

Beck ET, Bosio CF, Geske DA, Blair CD, Beaty BJ, Black WCIV (2005) An analysis of gene flow among Midwestern populations of the mosquito *Ochlerotatus triseriatus*. Am J Trop Med Hyg 73:534–540

Bergmann A, Yang AY-P, Srivastava M (2003) Regulators of IAP function: coming to grips with the grim reaper. Curr Opin Cell Biol 15:717–724

Black WC IV, DuTeau NM (1997) RAPD-PCR and SSCP analysis for insect population genetic studies. In: Crampton J, Beard CB, Louis C (eds) The molecular biology of insect disease vectors: a methods manual. Chapman and Hall, New York, NY, pp 361–373

Black WC IV, Gorrochotegui-Escalante N, Duteau NM (2006) Heated oligonucleotide ligation assay (HOLA): an affordable single nucleotide polymorphism assay. J Med Entomol 43:238–247

Blitvich BJ, Blair CD, Kempf BJ, Hughes MT, Black WCIV, Mackie RS, Meredith CT, Beaty BJ, Rayms-Keller A (2002) Developmental- and tissue-specific expression of an inhibitor of apoptosis protein 1 homologue from *Aedes triseriatus* mosquitoes. Insect Mol Biol 11:431–442

Borucki MK, Kempf BJ, Blitvich BJ, Blair CD, Beaty BJ (2002) La Crosse virus: replication in vertebrate and invertebrate hosts. Microbes Infect 4:341–350

Chai JJ, Du CY, Wu JW, Kyin S, Wang XD, Shi YG (2000) Structural and biochemical basis of apoptotic activation by Smac/DIABLO. Nature 406:855–862

Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G et al (2002) Immunity-related genes and gene families in *Anopheles gambiae*. Science 298:159–165

Clark GG, Pretula HL, Rohrer WH, Harroff RN, Jakubowski T (1983) Persistence of La Crosse virus (California encephalitis serogroup) in north-central Illinois. Am J Trop Med Hyg 32:175–184

Colon-Ramos DA, Irusta PM, Gan EC, Olson MR, Song J, Morimoto RI, Elliott RM, Lombard M, Hollingsworth R, Hardwick JM et al (2003) Inhibition of translation and induction of apoptosis by bunyaviral nonstructural proteins bearing sequence similarity to reaper. Mol Biol Cell 14:4162–4172

Copeland RS, Craig GB Jr (1990) Habitat segregation among treehole mosquitoes (Diptera: Culicidae) in the Great Lakes region of the United States. Ann Entomol Soc Am 83:1063–1073

Dobie DK, Blair CD, Chandler LJ, Rayms-Keller A, McGaw MM, Wasieloski LP, Beaty BJ (1997) Analysis of La Crosse virus S mRNA 5′ termini in infected mosquito cells and *Aedes triseriatus* mosquitoes. J Virol 71:4395–4399

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285–294

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Gorrochotegui-Escalante N, Lozano-Fuentes S, Bennett KE, Molina-Cruz A, Beaty BJ, Black WCIV (2005) Association mapping of segregating sites in the early trypsin gene and susceptibility to dengue-2 virus in the mosquito *Aedes aegypti*. Insect Biochem Mol Biol 35:771–788

Graham DH, Holmes JL, Higgs S, Beaty BJ, Black WCIV (1999) Selection of refractory and permissive strains of *Aedes triseriatus* (Diptera: Culicidae) for transovarial transmission of La Crosse virus. J Med Entomol 36:671–678

Graham DH, Holmes JL, Beaty BJ, Black WCIV (2003) Quantitative trait loci conditioning transovarial transmission of La Crosse virus in the eastern treehole mosquito, *Ochlerotatus triseriatus*. Insect Mol Biol 12:307–318

Grimstad PR, Garry CE, DeFoliart GR (1974) *Aedes hendersoni* and *Aedes triseriatus* (Diptera: Culicidae) in Wisconsin: Characterization of larvae, larval hybrids, and comparison of adult and hybrid mesoscutal patterns. Ann Entomol Soc Am 67:795–804

Hay BA, Wassarman DA, Rubin GM (1995) *Drosophila* homologs of baculovirus inhibitor of apoptosis proteins function to block cell death. Cell 83:1253–1262

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK

Liu ZH, Sun CH, Olejniczak ET, Meadows RP, Betz SF, Oost T, Herrmann J, Wu JC, Fesik SW (2000) Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. Nature 408:1004–1008

Lynd A, Ranson H, McCall PJ, Randle NP, Black WCIV, Walker ED, Donnelly MJ (2005) A simplified high-throughput method for pyrethroid knock-down resistance (kdr) detection in *Anopheles gambiae*. Malar J 4:16

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

McGaw MM, Chandler LJ, Wasieloski LP, Blair CD, Beaty BJ (1998) Effect of La Crosse virus infection on overwintering of *Aedes triseriatus*. Am J Trop Med Hyg 58:168–175

Moriyama E, Powell J (1996) Intraspecific nuclear DNA variation in *Drosophila*. Mol Biol Evol 13:261–277

Morlais I, Severson DW (2003) Intraspecific DNA variation in nuclear genes of the mosquito *Aedes aegypti*. Insect Mol Biol 12:631–639

Morlais I, Poncon N, Simard F, Cohuet A, Fontenille D (2004) Intaspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. Am J Trop Med Hyg 71:795–802

Munstermann LE, Taylor DB, Matthews TC (1982) Population genetics and speciation in the *Aedes triseriatus* group. In: Steiner WW (ed) Recent developments in the genetics of insect disease vectors. Stipes, Champaign, IL, pp 433–453

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Ohta T (1982a) Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc Natl Acad Sci USA 79:1940–1944

Ohta T (1982b) Linkage disequilibrium with the island model. Genetics 101:139–155

Palaga T, Osborne B (2002) The 3D's of apoptosis: death, degradation and DIAPs. Nat Cell Biol 4:E149–E151

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Truman JW, Craig GB Jr (1968) Hybridization between *Aedes hendersoni* and *Aedes triseriatus*. Ann Entomol Soc Am 61:1020–1025

Vucic D, Kaiser WJ, Miller LK (1998) Inhibitor of apoptosis proteins physically interact with and block apoptosis induced by *Drosophila* proteins HID and GRIM. Mol Cell Biol 18:3300–3309

Wang DG, Fan J-B, Siao C-J, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J et al (1998) Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. Science 280:1077–1082

Wang SL, Hawkins CJ, Yoo SJ, Muller HAJ, Hay BA (1999) The *Drosophila* caspase inhibitor DIAP1 is essential for cell survival and is negatively regulated by HID. Cell 98:453–463

Watts DM, Wright RE, Hanson RP, Defoliart GR, Morris CD (1972) Transmission of La Crosse virus (California Encephalitis Group) by mosquito *Aedes triseriatus*. J Med Entomol 9:125–127

Watts DM, Pantuwatana S, Defoliart GR, Yuill TM, Thompson WH (1973) Transovarial transmission of La Crosse virus (California Encephalitis Group) in the mosquito, *Aedes Triseriatus*. Science 182:1140–1141

Watts DM, Thompson WH, Yuill TM, Defoliart GR, Hanson RP (1974) Overwintering of La Crosse virus in *Aedes triseriatus*. Am J Trop Med Hyg 23:694–700

Watts DM, Pantuwatana S, Yuill TM, Defoliart GR, Thompson WH, Hanson RP (1975) Transovarial transmission of La Crosse virus in *Aedes triseriatus*. Ann NY Acad Sci 266:135–143

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Wondji C, Hemingway J, Ranson H (2007) Identification and analysis of single nucleotide polymorphisms (SNPs) in the mosquito *Anopheles funestus*, malaria vector. BMC Genomics 8:5

Woodring J, Chandler LJ, Oray CT, McGaw MM, Blair CD, Beaty BJ (1998) Short report: Diapause, transovarial transmission, and filial infection rates in geographic strains of La Crosse virus-infected *Aedes triseriatus*. Am J Trop Med Hyg 58:587–588

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Wu G, Chai JJ, Suber TL, Wu JW, Du CY, Wang XD, Shi YG (2000) Structural basis of IAP recognition by Smac/DIABLO. Nature 408:1008–1012