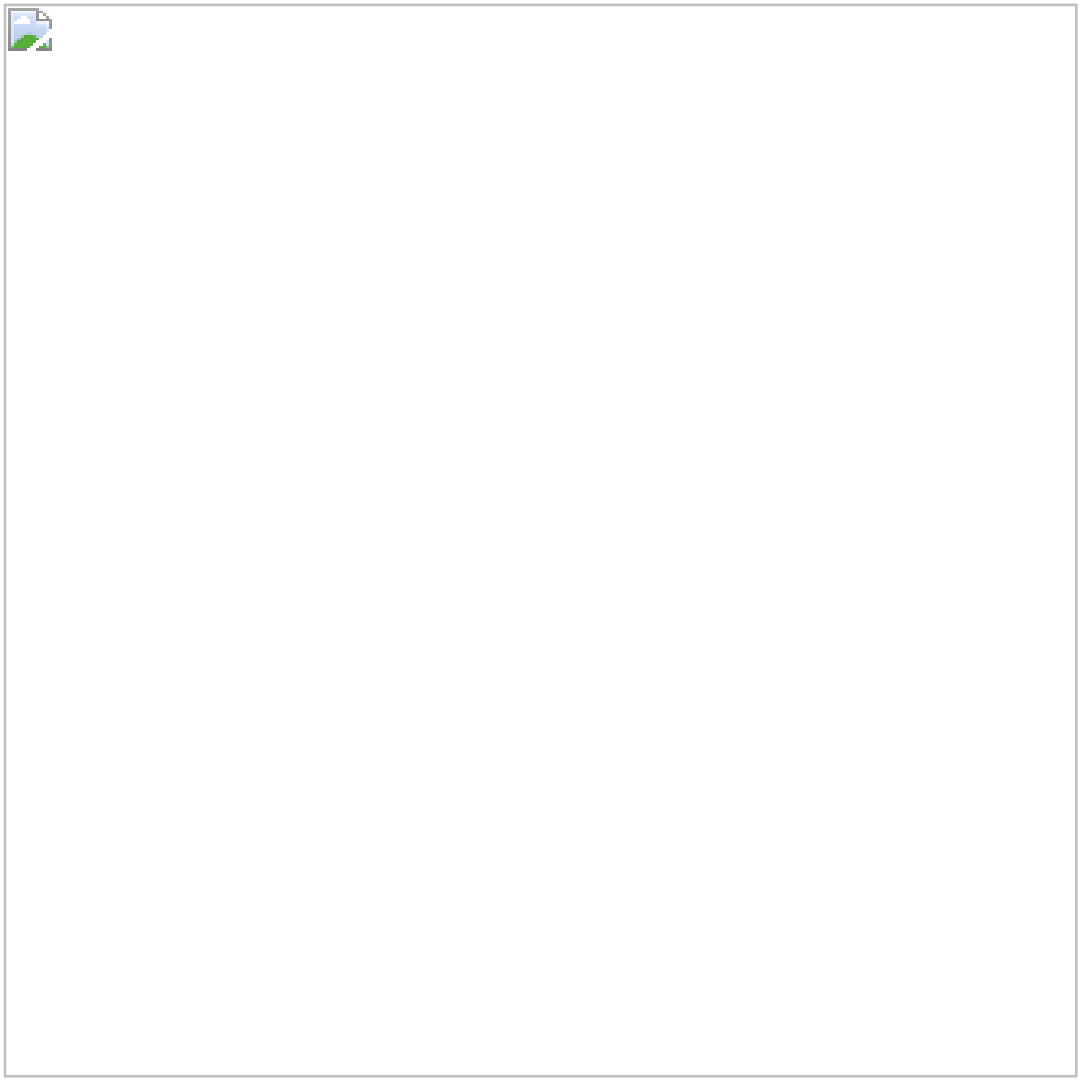

Genomic and Bioinformatic Analysis of NADPH-Cytochrome P450 Reductase in Anopheles stephensi (Diptera: Culicidae)

Authors: Suwanchaichinda, C., and Brattsten, L. B.

Source: Journal of Insect Science, 14(165) : 1-8

Published By: Entomological Society of America

URL: https://doi.org/10.1093/jisesa/ieu027

RESEARCH

Genomic and Bioinformatic Analysis of NADPH-Cytochrome P450 Reductase in *Anopheles stephensi* (Diptera: Culicidae)

C. Suwanchaichinda[1,2] and L. B. Brattsten[1]

[1]Department of Entomology, Rutgers University, New Brunswick, NJ 08901
[2]Corresponding author, e-mail: csuwanch@hotmail.com

Subject Editor: Michael Kanost

J. Insect Sci. 14(165): 2014; DOI: 10.1093/jisesa/ieu027

ABSTRACT. The cytochrome P450 monooxygenase (P450) enzyme system is a major mechanism of xenobiotic biotransformation. The nicotinamide adenine dinucleotide phosphate (NADPH)-cytochrome P450 reductase (CPR) is required for transfer of electrons from NADPH to P450. One CPR gene was identified in the genome of the malaria-transmitting mosquito *Anopheles stephensi* Liston (Diptera: Culicidae). The gene encodes a polypeptide containing highly conserved flavin mononucleotide-, flavin adenine dinucleotide-, and NADPH-binding domains, a unique characteristic of the reductase. Phylogenetic analysis revealed that the *A. stephensi* and other known mosquito CPRs belong to a monophyletic group distinctly separated from other insects in the same order, Diptera. Amino acid residues of CPRs involved in binding of P450 and cytochrome *c* are conserved between *A. stephensi* and the Norway rat *Rattus norvegicus* Berkenhout (Rodentia: Muridae). However, gene structure particularly within the coding region is evidently different between the two organisms. Such difference might arise during the evolution process as also seen in the difference of P450 families and isoforms found in these organisms. CPR in the mosquito *A. stephensi* is expected to be active and serve as an essential component of the P450 system.

Key Words: binding domain, gene structure, phylogenetic tree, sequence analysis

The mosquito *Anopheles stephensi* Liston (Diptera: Culicidae) is one of the major vectors that transmit the human malaria parasite *Plasmodium falciparum* Welch. Geographical distribution of the mosquito ranges from the Middle East to the Indian subcontinent (Kiszewski et al. 2004). Various types of insecticides have been used in controlling populations of mosquito vectors. Insecticide resistance or changes in insecticide susceptibility have been documented in *A. stephensi* (Ganesh et al. 2003, Enayati and Ladonni 2006, Tiwari et al. 2010, Tikar et al. 2011, Shetty et al. 2012).

The cytochrome P450 monooxygenase (P450) system is well known as a major mechanism of xenobiotic biotransformation. Involvement of specific P450 isoforms in insecticide resistance has been characterized in several insects, including mosquitoes (Nikou et al. 2003, Shen et al. 2003, Djouaka et al. 2008, Yang and Liu 2011, Bariami et al. 2012). As part of the redox reaction, a nicotinamide adenine dinucleotide phosphate (NADPH)-cytochrome P450 reductase (CPR) is needed in transferring electrons to P450s. CPR is a flavoprotein containing binding domains for flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), and NADPH, as determined by X-ray crystallography of rat CPR (Wang et al. 1997). NADPH serves as electron donor, whereas FAD and FMN serve as cofactors in the pathway of electron transfer. In eukaryotic microsomes, both P450s and CPRs are integral membrane proteins. Although there is such an extensive diversity of P450 isoforms identified in animals (Nelson 2011), each animal's genome generally carries only one CPR gene (Porter et al. 1990). Thus, interaction between the two enzymes in the microsomal environment is probably in the ratio of multiple P450s to one CPR. Feyereisen (2005, 2012) suggests a ratio of 6–18 P450s to 1 reductase in insects. Lycett et al. (2006) showed that silencing the CPR gene could impact insecticide susceptibility in the mosquito *Anopheles gambiae*.

To date, only a few mosquito CPRs have been identified and deposited in the NCBI GenBank database: *Aedes sollicitans* (D. Sun and L.B.B., unpublished data), *Aedes aegypti* (Nene et al. 2007), *Anopheles funestus* (Matambo et al. 2010), *A. gambiae* (Nikou et al. 2003), *Anopheles minimus* (Kaewpa et al. 2007), and *Culex quinquefasciatus* (Arensburger et al. 2010). By using *A. gambiae* microarray platforms, a few possible P450 transcripts were identified in *A. stephensi* (Vontas et al. 2007). A small nucleotide sequence, possibly representing a fragment of a CPR gene, was obtained from *A. stephensi* by multilocus DNA sequencing as part of a phylogenetic study among different anopheline mosquitoes (Dixit et al. 2010). However, a complete sequence of the *A. stephensi* CPR gene has not, to date, been determined.

In this study, a complete coding sequence of a CPR gene was identified by comparative genomic analysis of the *A. stephensi* genome; the whole genome shotgun sequence of *A. stephensi* was originally deposited in the NCBI GenBank database by Hall and Jiang, Virginia Tech University, in 2012. The deduced amino acid sequence was analyzed by using bioinformatic and phylogenetic analyses. Conserved residues were further characterized in relation to rat and known mosquito CPRs. This study was undertaken because of our long-standing interest in the role of the P450 enzyme system in the responses of insects, especially mosquitoes, to insecticides applied in an attempt to control their populations and also in their responses to allelochemical substances in their ecological habitats (Suwanchaichinda and Brattsten 2002). (Note: the nucleotide sequence of the *A. stephensi* CPR transcript is available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession number TPA:BK008720.)

Materials and Methods

Genomic Analysis. To screen the NCBI nonredundant nucleotide database for any possible CPR gene identified in *A. stephensi*, the nucleotide sequence of the *A. gambiae* CPR gene was used as a query in a BlastN search. The *A. gambiae* CPR nucleotide and amino acid sequences were then used in BlastN and TBlastN searches, respectively, against the genome sequences in the NCBI database. An identified genome contig was compared with the *A. funestus* CPR gene to reconfirm its identity. Additionally, the *Ae. sollicitans* CPR cDNA, which was cloned and sequenced in our laboratory, was used as an inferential evidence for the annotation of the *A. stephensi* CPR gene.

Sequences obtained as results of both nucleotide and genome sequence database searches were subsequently used for further analysis.

Sequence Characterization. An identified genome contig containing a possible *A. stephensi* CPR gene was manually analyzed by comparing it with previously identified mosquito CPR genes to obtain a complete coding region. The intron–exon organization (from initiation codon to stop codon) and the sizes of exons of the *A. stephensi* CPR gene were determined according to CPR genes in the genome releases of *A. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus* available in VectorBase (http://www.vectorbase.org). The CPR gene of the Norway rat (*Rattus norvegicus*) was used as a representative of vertebrates for comparison. Information on the intron–exon organization of the rat CPR gene was obtained from Ensemble (http://www.ensembl.org).

A hypothetical translation of the complete open reading frame of the *A. stephensi* CPR transcript was performed to obtain a deduced amino acid sequence. The translated protein was then used in a BlastP search against the NCBI nonredundant protein database to confirm the *A. stephensi* CPR identity compared with other known CPRs in invertebrates and vertebrates. The isoelectric point (pI) and molecular mass of the *A. stephensi* CPR were calculated by a proteomics tool (Gasteiger et al. 2005). A transmembrane spanning region was predicted by TMpred (http://embnet.vital-it.ch/software/TMPRED_form.html) and TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM).

In addition to putative conserved domains identified as part of the BlastP search, the identification of binding domains (for FMN, FAD, and NADPH) and catalytic residues were predicted by Pfam 26.0 (http://pfam.sanger.ac.uk). A multiple sequence alignment between *A. stephensi* CPR and other mosquito CPRs was performed by ClustalW (Larkin et al. 2007). Amino acid residues involved in binding to cytochrome P450s and cytochrome *c* were predicted based on the rat CPR previously identified by Shen et al. (1989) and Shen and Kasper (1995). ClustalW was also used to create a multiple sequence alignment between mosquito and rat CPRs.

Phylogenetic Analysis. The deduced amino acid sequence of *A. stephensi* CPR was used in a BlastP search for known homologous CPRs deposited in the GenBank. Subsequently, CPRs representing vertebrates and arthropods with complete open reading frames were selected for phylogenetic analysis. The CPR amino acid sequences were first analyzed using T-Coffee multiple sequence alignment with default parameters (Notredame et al. 2000) to determine variable and conserved regions. A region of low homology at the N-termini was manually removed, and the remaining portion of the sequences was realigned using ClustalW embedded in the MEGA 4.0 software (Tamura et al. 2007). The created alignment file was then used for construction of a neighbor-joining phylogenetic tree. Gaps were eliminated from the algorithmic analysis, and 1,000 replications were performed as a bootstrap test of phylogeny. A cut-off value of 50% was set for the bootstrap consensus tree.

Results

Genomic and Phylogenetic Analysis. The BlastN search, using the nucleotide sequence of the *A. gambiae* CPR gene against the genome sequences in the NCBI GenBank database, identified a specific contig (GenBank ALPR01027092) of the *A. stephensi* genome. This contig was confirmed to contain one CPR gene by comparative genomic analysis. A small fragment of nucleotide sequence (GenBank HM171646) was acquired as a result of a BlastN search against the nonredundant nucleotide sequence. This partial sequence was incorporated in this study to correct a genomic sequencing error. The corrected genomic sequence was then compared with CPR genes previously identified in other mosquito species, which yielded a complete coding region of the *A. stephensi* CPR gene from the initiation codon (ATG) to the stop codon (TAA). The existence and expression of the gene were confirmed by the presence of an expressed sequence tag (EST) (Patil et al. 2009).

The identified *A. stephensi* CPR gene contained an open reading frame of 2,040 bp, coding for a polypeptide of 679 amino acids (Fig. 1).

```
atggacgcccagacagagacggaaatgcccacgggcaacgtgagcgacgagcccttcctc  60
 M  D  A  Q  T  E  T  E  M  P  T  G  N  V  S  D  E  P  F  L  20
ggcccgttggacatcatcctgctcgtcagcctgctggccggcaccgcctggtacctgctg  120
 G  P  L  D  I  I  L  L  V  S  L  L  A  G  T  A  W  Y  L  L  40
aagggcaagaaaaaagaaaatcaagctagtcagttcaaatcctactcgatccaaccgacg  180
 K  G  K  K  K  E  N  Q  A  S  Q  F  K  S  Y  S  I  Q  P  T  60
accgtcaacacgatgacaatggtggagaactcgttcatcaagaagctgcaatcctccggt  240
 T  V  N  T  M  T  M  V  E  N  S  F  I  K  K  L  Q  S  S  G  80
cgccgtttggtggtgttttacggttcccaaacaggcacggcagaggaatttgccggtcgt  300
 R  R  L  V  V  F  Y  G  S  Q  T  G  T  A  E  E  F  A  G  R  100
ctggcgaaggaaggaatccgctatcagatgaagggcatggtggccgatccggaggagtgc  360
 L  A  K  E  G  I  R  Y  Q  M  K  G  M  V  A  D  P  E  E  C  120
aacatggaagaactgctgatgttgaaggacatcgacaaatcgttggctgtgttttgctta  420
 N  M  E  E  L  L  M  L  K  D  I  D  K  S  L  A  V  F  C  L  140
gccacgtacggcgagggtgacccgacggacaattgtatggagttttacgactggatacag  480
 A  T  Y  G  E  G  D  P  T  D  N  C  M  E  F  Y  D  W  I  Q  160
aacaacgatctggacatgaccggtttgaattacgcggtatttggcctcggcaataagacg  540
 N  N  D  L  D  M  T  G  L  N  Y  A  V  F  G  L  G  N  K  T  180
tacgagcattacaacaaggtcggcatctacgtggataagcggctcgaggagctcggtgcg  600
 Y  E  H  Y  N  K  V  G  I  Y  V  D  K  R  L  E  E  L  G  A  200
aaccgagtgtttgagctaggactcggtgacgatgatgcgaacattgaggactacttcatc  660
 N  R  V  F  E  L  G  L  G  D  D  D  A  N  I  E  D  Y  F  I  220
acgtggaaggaaaagttttggccaacggtgtgcgattacttcggcatcgaaagtacgggc  720
 T  W  K  E  K  F  W  P  T  V  C  D  Y  F  G  I  E  S  T  G  240
gaggatgtgctgatgcgccagtaccgtctgctggagcaaccggacgtgagcgccgaccgc  780
 E  D  V  L  M  R  Q  Y  R  L  L  E  Q  P  D  V  S  A  D  R  260
gtgtataccggcgaggtggcacggctccactcgctccaaacgcagcgtccaccgttcgat  840
 V  Y  T  G  E  V  A  R  L  H  S  L  Q  T  Q  R  P  P  F  D  280
gcaaagaacccgttcctggcaccgatcaaggtaaacccgggagctgcacaaagccggcggc  900
 A  K  N  P  F  L  A  P  I  K  V  N  R  E  L  H  K  A  G  G  300
cggtcctgcatgcacgtcgagttcgatatcgaaggttcgaagatgcggtacgaggcgggc  960
 R  S  C  M  H  V  E  F  D  I  E  G  S  K  M  R  Y  E  A  G  320
gaccatctcgccatgtacccggtaaacgatcgggatctggtcgagcggctcggcaagctg  1020
 D  H  L  A  M  Y  P  V  N  D  R  D  L  V  E  R  L  G  K  L  340
tgcaaagccgacctggagacggtcttctcgctcatcaacaccgacacggacagcagcaag  1080
 C  K  A  D  L  E  T  V  F  S  L  I  N  T  D  T  D  S  S  K  360
aagcatccgttcccgtgcccaaccacgtaccggaccgccctgacgcactacctggaaata  1140
 K  H  P  F  P  C  P  T  T  Y  R  T  A  L  T  H  Y  L  E  I  380
acggccctgccgcgtacccacatcctgaaggagttggccgaatactgctcggaggaaaag  1200
 T  A  L  P  R  T  H  I  L  K  E  L  A  E  Y  C  S  E  E  K  400
gacaaagagttcctgcggtttatctcgtcgaccgcaccggaaggcaaggtcaagtaccag  1260
 D  K  E  F  L  R  F  I  S  S  T  A  P  E  G  K  V  K  Y  Q  420
gaatgggtgcaggacagctgccgcaacattgtgcacgtgctggaggacatcccgtcctgt  1320
 E  W  V  Q  D  S  C  R  N  I  V  H  V  L  E  D  I  P  S  C  440
cagccaccgatcgatcacgtgtgcgagctgctgccccgtttgcaaccccgctactactcc  1380
 Q  P  P  I  D  H  V  C  E  L  L  P  R  L  Q  P  R  Y  Y  S  460
atctcctcctcgtccaagatccatccgacgacggtgcacgtgaccgcggtgctggtacgg  1440
 I  S  S  S  S  K  I  H  P  T  T  V  H  V  T  A  V  L  V  R  480
tacgaaacgaaaacgggccgcctgaacaagggcgtggcgacgaccttcctcgcggaaaag  1500
 Y  E  T  K  T  G  R  L  N  K  G  V  A  T  T  F  L  A  E  K  500
cacccgaacgatggggaaccgttaccgcgcgtgcccatcttcatacgcaagagccagttc  1560
 H  P  N  D  G  E  P  L  P  R  V  P  I  F  I  R  K  S  Q  F  520
cgcctgccaccgaaaccggaaacgcccgtgataatggtaggccccggtacggggttggca  1620
 R  L  P  P  K  P  E  T  P  V  I  M  V  G  P  G  T  G  L  A  540
ccgttccgtggcttcatccaggagcgtgacttctgcaagcaggagggtaaagacatcgga  1680
 P  F  R  G  F  I  Q  E  R  D  F  C  K  Q  E  G  K  D  I  G  560
cagacgacactctactttggctgccgcaaacggtccgaggactacatatacgaggacgag  1740
 Q  T  T  L  Y  F  G  C  R  K  R  S  E  D  Y  I  Y  E  D  E  580
ctggaagactactccaagcggggcatcatcaatctgcgcgtggcgttctcgcgcgaccag  1800
 L  E  D  Y  S  K  R  G  I  I  N  L  R  V  A  F  S  R  D  Q  600
gacaagaaggtgtacgtgacgcatctgctagagcaagactcggacctgatatggaacgtt  1860
 D  K  K  V  Y  V  T  H  L  L  E  Q  D  S  D  L  I  W  N  V  620
atcggcgaaaacaagggacacttttacatctgcggtgatgcaaaaaatatggccaccgat  1920
 I  G  E  N  K  G  H  F  Y  I  C  G  D  A  K  N  M  A  T  D  640
gtgcgaaacattctgctcaaggtcatacggtcgaagggtgggctaagcgaaaccgaagcc  1980
 V  R  N  I  L  L  K  V  I  R  S  K  G  G  L  S  E  T  E  A  660
cagcagtacatcaaaaagatggaagcacaaaagcgatactcggcggacgtttggagctaa  2040
 Q  Q  Y  I  K  K  M  E  A  Q  K  R  Y  S  A  D  V  W  S  *  679
```

Fig. 1. The predicted sequence of *A. stephensi* CPR transcript. The sequence extends from the initiation codon to the stop codon. The upper and lower lines represent nucleotide and deduced amino acid sequences, respectively. The one-letter code for each amino acid is aligned with the second nucleotide of each codon. The predicted transmembrane region is underlined.

The calculated molecular mass of the deduced amino acid sequence was 77.51 kDa, with a pI of 5.62. The CPR apparently did not contain an N-terminal signal peptide. Instead, a transmembrane helix was found at the N-terminal region, indicating that this was a possible anchored site to the endoplasmic reticulum. The transmembrane hydrophobic region was predicted to be from Phe^{19} to Leu^{40} (Fig. 1). The *A. stephensi* CPR showed the highest identity to CPRs in *A. funestus*, *A. gambiae*, and *A. minimus* and a lower identity to CPRs in *Culex* and *Aedes* mosquitoes (Table 1). Evolutionarily, the *A. stephensi* CPR showed higher percentages of identity with dipteran insects than with those in other arthropod groups and rat (Table 1).

Amino acid sequences were used for phylogenetic analysis. A neighbor-joining phylogenetic tree was constructed representing *A. stephensi* CPR along with previously characterized CPRs of

Table 1. Amino acid sequence identity of CPRs between *A. stephensi* and other selected organisms

GenBank accession no.	Species	Common name	Total score	Max identity (%)[a]
ABO77954	*A. funestus*	African malaria mosquito	1,392	97
AAO24765	*A. gambiae*	African malaria mosquito	1,386	97
ABL75156	*A. minimus*	Southeast Asian malaria mosquito	1,378	97
XP_001656715	*Ae. aegypti*	Yellow fever mosquito	1,273	88
ACL01092	*Ae. sollicitans*	Eastern saltmarsh mosquito	1,264	87
XP_001865801	*C. quinquefasciatus*	Southern house mosquito	1,243	87
NP_477158	*Drosophila melanogaster*	Fruit fly	1,120	78
ADD19306	*Glossina morsitans morsitans*	Tsetse fly	1,102	77
Q07994	*Musca domestica*	House fly	1,094	77
XP_971174	*Tribolium castaneum*	Red flour beetle	998	70
ADX95746	*Spodoptera exigua*	Beet armyworm	991	69
AAR26515	*Mamestra brassicae*	Cabbage moth	985	68
NP_001104834	*Bombyx mori*	Silkworm	978	68
EHJ63867	*Danaus plexippus*	Monarch butterfly	976	69
XP_002423980	*Pediculus humanus corporis*	Human body louse	972	67
AFD50507	*Cimex lectularius*	Bed bug	970	66
XP_002400171	*Ixodes scapularis*	Black-legged tick	816	59
AAA41683	*R. norvegicus*	Norway rat	792	55

[a]*E* value = 0.0.

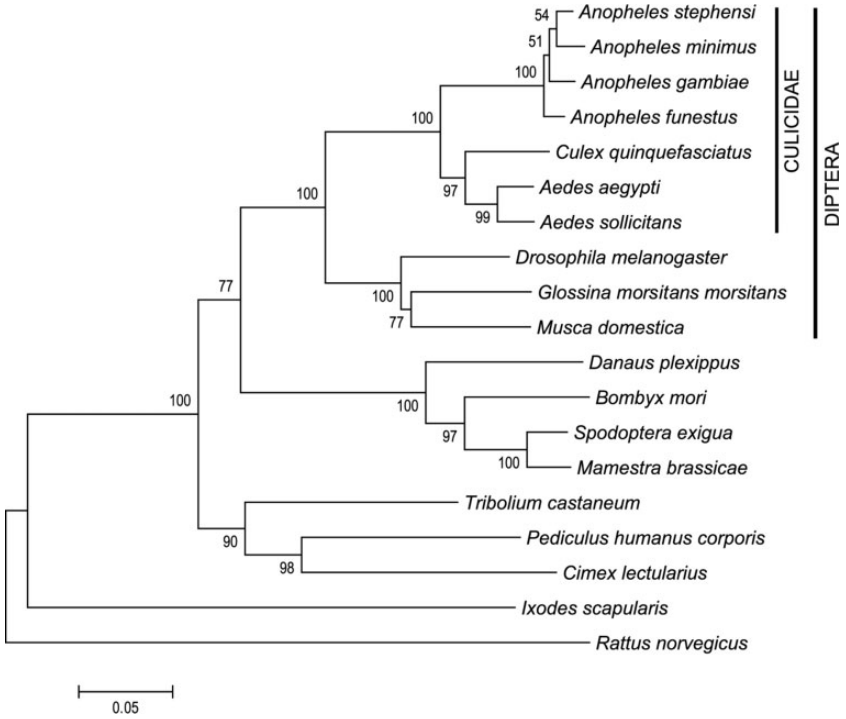

Fig. 2. A neighbor-joining phylogenetic tree of CPRs in *A. stephensi* and other selected organisms. Gaps were excluded ($n = 1{,}000$). The branch length represents the accumulation of amino acid changes. The scale indicates 0.05 amino acid change per site. The vertical bars indicate phylogenetic clusters.

mosquitoes and other selected arthropods. A rat CPR was used as an outgroup. As expected, all mosquito CPRs formed a cluster distinctly separated from other organisms (Fig. 2). In particular, CPRs of *A. stephensi* and other *Anopheles* mosquitoes were grouped together within a monophyletic group. The branch representing mosquito CPRs bifurcated from the same node as did the branch that represented other dipteran CPRs. This indicated that they possibly descended from a common ancestor.

The genomic DNA sequences of mosquito CPRs were analyzed for intron–exon organizations. Only species with publicly released genomes were included in this analysis. Based on the region of nucleotides from the initiation codon to the stop codon, seven exons were identified in the mosquito CPR genes (Fig. 3). The sizes of the exons of the *A. stephensi* CPR gene are generally similar to those of *A. gambiae*, *Ae. aegypti*, and *C. quinquefasciatus*. On the other hand, 15 exons were identified in the rat CPR within the coding region. The sizes of exons 3, 4, and 7 of the *A. stephensi* CPR were identical to exons 4, 5, and 15 of the rat CPR, respectively. There was only a slight difference in exon 1 between the two groups. The rat CPR exons 2 and 3 were comparable with the mosquito CPR exon 2, whereas the rat CPR exons 6–14 corresponded to the mosquito CPR exons 5 and 6. The last exon of the *A. stephensi* CPR appeared to be conserved, when compared with the last exon of the rat CPR (Fig. 4). CPRs in insects representing Diptera, Coleoptera, and Lepidoptera also contained some amino acid residues identical to those in *A. stephensi* and rat within this particular region.

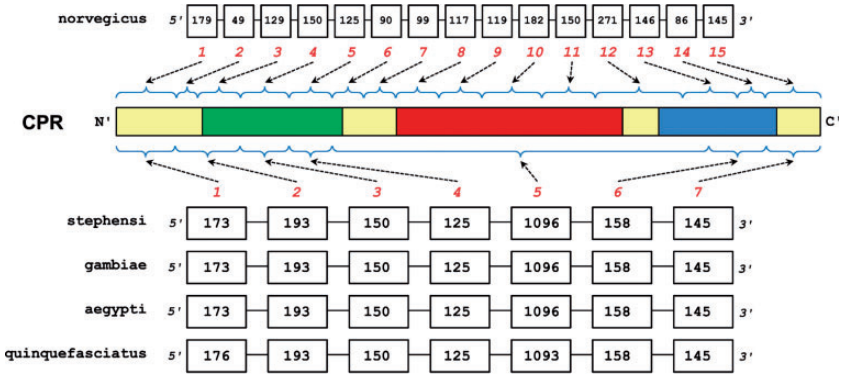

Fig. 3. Intron–exon organization of representative CPR genes. Only mosquitoes with their genomes completed and released to the public were analyzed for gene structure. Rat CPR was included for genetic comparison. The coding regions shown here start from the initiation codon to the stop codon. The square boxes represent exons, and the lines connecting the boxes represent introns. The sizes of both exons and introns are not drawn to scale. The nucleotide length of each exon is presented in each box. Sequential exon numbers (in red) are shown either under (in the case of rat CPR gene) or above (in the case of mosquito CPR genes) the corresponding exons. Binding domains for FMN, FAD, and NADPH are shown in green, red, and blue, respectively.

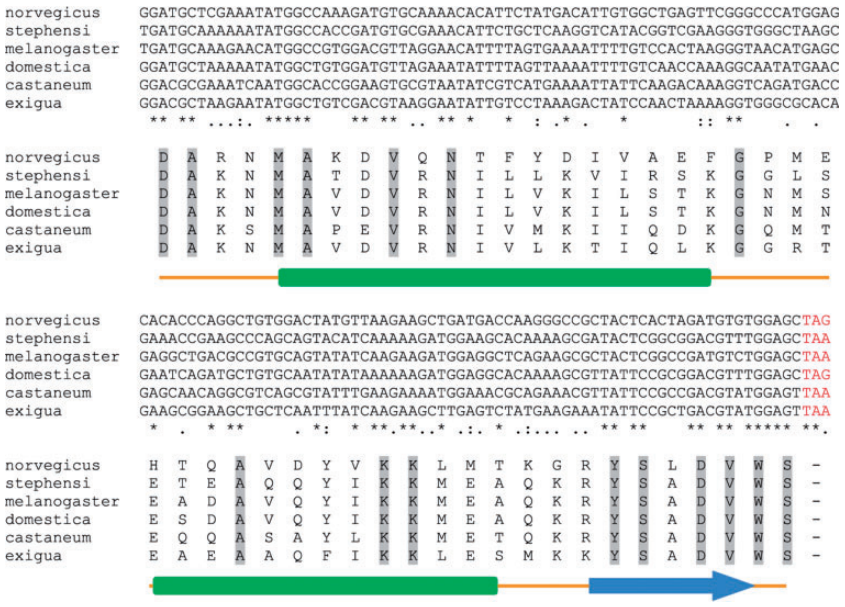

Fig. 4. Nucleotide and amino acid sequence comparisons of the last exons of the CPR coding regions between *A. stephensi* and other organisms. The one-letter code for each amino acid is aligned with the second nucleotide of each codon. The predicted secondary structures of the *A. stephensi* CPR are shown under the amino acid sequences. The green bar represents the alpha helix, and the blue arrow represents the beta sheet. Identical residues in amino acid sequences are highlighted. Organisms included in the analysis are *R. norvegicus* (representing a vertebrate); *Drosophila melanogaster* and *Musca domestica* (representing Diptera along with *A. stephensi*), *Tribolium castaneum* (representing Coleoptera), and *Spodoptera exigua* (representing Lepidoptera).

Amino Acid Sequence Analysis. A multiple sequence alignment among mosquito CPRs was carried out to determine conserved regions involved in ligand bindings. The *A. stephensi* CPR clearly contained all conserved FMN-, FAD-, and NADPH-binding domains (Fig. 5). Catalytic residues (Ser^{460}, Cys^{631}, Asp^{676}, and Trp^{678}) were identified in the mosquito CPR. These residues may be essential in the hydride transfer reaction as previously characterized in the rat CPR (Shen et al. 1999, Hubbard et al. 2001). The FMN-binding domain of the mosquito CPR contained two conserved tyrosine residues (Tyr^{143} and Tyr^{181}) that might be critical in FMN binding (Shen et al. 1989; Fig. 5). In addition, Phe residues at positions 86 and 219 were conserved in all mosquitoes except *A. minimus*, which instead contained Leu at both positions. Phe^{86} and Phe^{219} are involved in FMN binding (Sarapusit et al. 2008, 2010).

The FAD- and NADPH-binding domains carry conserved sequence motifs recognized within the ferredoxin reductase structural family. One of the most conserved motifs is Arg-x-Tyr-Ser(Thr), which is found in all members of the family (Dym and Eisenberg 2001). Indeed, ^{457}Arg-Tyr-Tyr-Ser^{460} was detected within the FAD-binding domain of the *A. stephensi* CPR. This motif was also conserved in *A. funestus*, *Ae. aegypti*, *Ae. sollicitans*, and *C. quinquefasciatus*. However, the *A. minimus* and *A. gambiae* CPRs contained Ser and His, respectively, instead of Tyr at the position 459. Another conserved sequence motif within this particular protein family is Met-x-x-x-Gly-Thr(Ser)-Gly(Ala)-Ile-x-Pro, identified in the NADPH-binding domain of all mosquito CPRs (Fig. 5). With one exception, the mosquito CPRs carried Leu instead of Ile in the motif.

One major function of microsomal CPRs is the electron transfer from electron donors to P450s. The interaction between the two

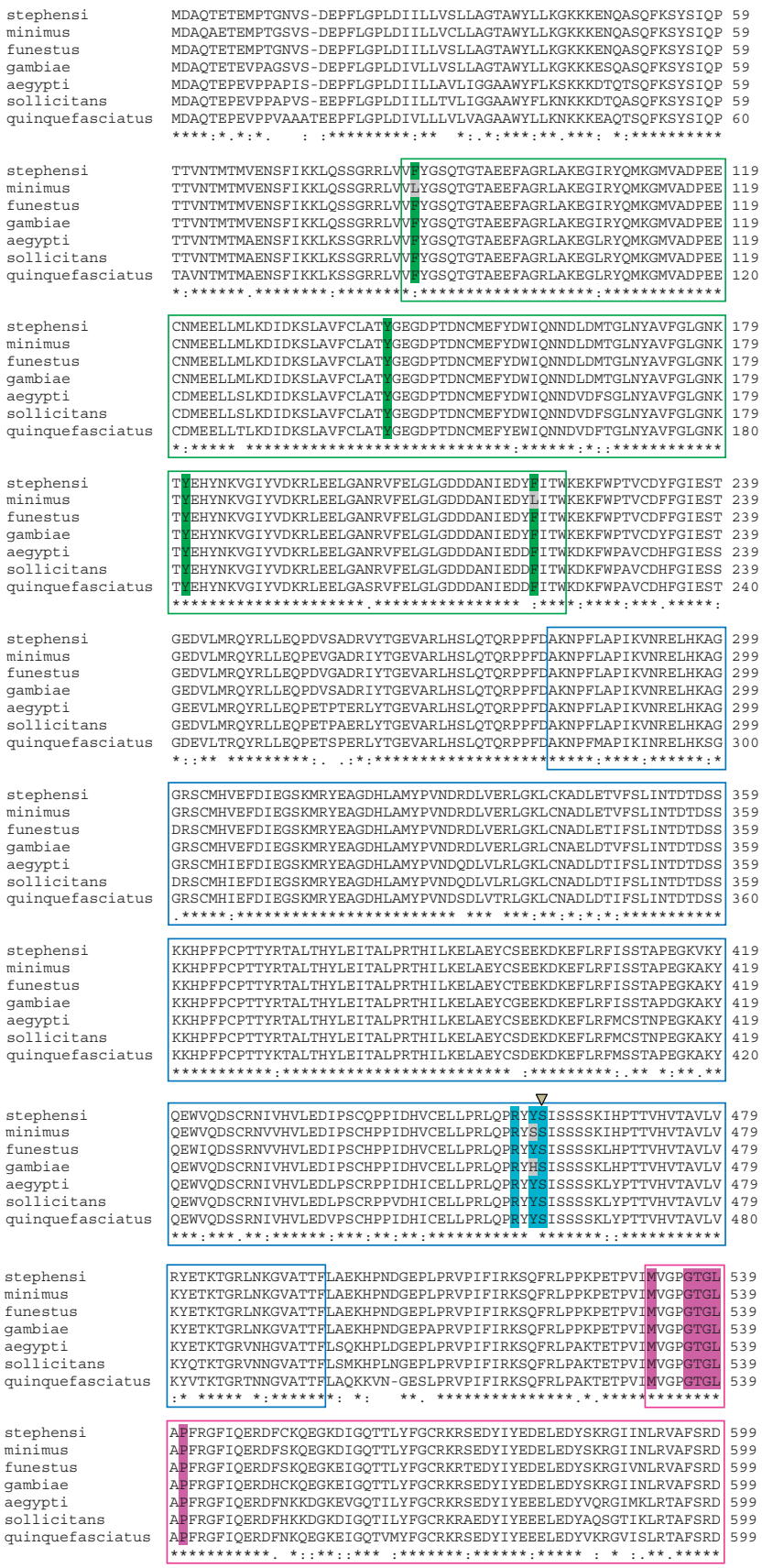

Fig. 5. Conserved regions for ligand bindings in the *A. stephensi* CPR. The multiple sequence alignment illustrates FMN-, FAD-, and NADPH-binding domains of mosquito CPRs according to Pfam. The binding domains are boxed (green, FMN; blue, FAD; purple, NADPH). Triangles above amino acids indicate catalytic residues involved in hydride transfer. Essential residues involved in ligand bindings are highlighted.

enzymes requires specific bindings that involve certain residues within the molecules. Acidic residue clusters ^{207}Asp-Asp-Asp^{209} and ^{213}Glu-Glu-Asp^{215} in rat CPR interact with cytochrome *c* or P450 (Shen and Kasper 1995). The *A. stephensi* CPR contained these conserved binding clusters as shown in the multiple sequence alignment in Fig. 6. Residues ^{210}Asp-Asp-Asp^{212} in the *A. stephensi* CPR appeared to be highly conserved in comparison with rat and other mosquitoes. However, the residue Glu at position 214 in the rat CPR was substituted with Asp in all mosquitoes. Substitution of Asp at position 215 in the rat CPR with Tyr was observed only in *Anopheles* but not in *Aedes* and *Culex* mosquitoes.

Discussion

Genome, proteome, and transcriptome have become valuable resources of information for studying mosquito vectors in combination with the availability of bioinformatic tools. Genomic analysis in this study identifies one CPR gene in the mosquito *A. stephensi*. The gene encodes a polypeptide containing 679 amino acid residues. The length of the translated protein is exactly equal to that of CPRs in other known mosquitoes previously characterized. The amino acid sequence of *A. stephensi* CPR shows the highest identity with the CPRs of *A. funestus* (Matambo et al. 2010), *A. gambiae* (Nikou et al. 2003), and *A. minimus* (Kaewpa et al. 2007). The identity of these mosquito CPRs is consistent with the results of phylogenetic analysis, showing that they belong to the same monophyletic group. It is likely that they perform similar physiological functions. CPRs of the other few known dipteran species are grouped together as another cluster. These two phylogenetic groups apparently bifurcate from the same node. Lepidoperan CPRs clearly form a separate cluster. CPRs of the bed bug (*Cimex lectularius* L.) and human body louse (*Pediculus humanus corporis* L.) fall into the same clade, which is consistent with a previous study (Zhu et al. 2012).

Since the completion of human and rat genome sequences, the number of genome sequence projects in insects has increased and promises to accelerate with the 5,000 insect genome project (Levine 2011). With the mosquito genomes available to date, genomic sequences of CPR genes of mosquitoes and vertebrates were analyzed for intron–exon organization. From the initiation codon to the stop codon, 7 and 15 coding exons were found in the mosquito and rat CPRs, respectively. This comparison reveals unique differences and similarities between the two groups. The sizes of exons 3, 4, and 7 of the *A. stephensi* and other mosquito CPRs are similar to exons 4, 5, and 15 of the rat CPR. However, there is an intron between exons 2 and 3 in the rat CPR, which does not exist in the region where the mosquito CPR exon 2 is located. There is an intron between exons 5 and 6 in the mosquito CPRs, which is not found in the region where the rat CPR exon 13 is located. Exons 3 and 4 of the mosquito CPRs and exons 4 and 5 of the rat CPR are highly conserved, likely because they represent the FMN-binding domain. Codons corresponding to the two Tyr residues involved in FMN binding (Shen et al. 1989) obviously reside within these two exons. In the case of the last exon of the mosquito and rat CPRs (exon 7 and exon 15, respectively), their similarity is consistent with their protein structure homology, particularly the sequence conservation at the C-terminal region (Fig. 4). Therefore, the similarities between the mosquito and rat CPRs are not restricted to their gene structures but extend to their protein structures as well. In addition, some amino acid residues in other insect CPRs as shown in Fig. 4 are also identical to those in *A. stephensi* and rat CPRs within this region. These findings provide an insight into the evolutionary relationship of gene organization and domain formation in insect and vertebrate CPRs. Although the completion of genome assembly of *A. stephensi* is being generated, chromosomal location of the CPR gene in the mosquito is still unknown without gene mapping. However, the CPR gene was mapped on chromosome X in *A. gambiae* (Holt et al. 2002). It is possible that the CPR gene in *A. stephensi* may also be located on the same chromosome.

Based on the study of rat CPR, Porter and Kasper (1986) propose that CPRs originated from the fusion of two ancestral genes coding for a flavodoxin and a ferredoxin reductase. CPRs are unique proteins in the way they contain FMN-, FAD-, and NADPH-binding domains, in addition to binding sites for cytochrome *c* and P450. There is also a connecting domain between the FMN- and FAD-binding domains (Wang et al. 1997). The *A. stephensi* CPR has ligand-binding sites similar to those of other mosquitoes, and the locations of these sites are consistent with the binding domains as previously described in other species (Porter and Kasper 1986, Koener et al. 1993). As shown in the multiple sequence alignment among mosquito CPRs (Fig. 5), the *A. minimus* CPR contains Leu instead of Phe residues at both 86 and 219 positions. Sarapusit et al. (2008, 2010) conducted kinetic studies of the wild-type and mutant *A. minimus* CPRs and found that substitutions of the Leu residues with Phe increased retention of the FMN cofactor and stability of the enzyme. In comparison with the wild-type CPR, single mutation (L86F or L219F) and double mutation (L86F/L219F) of the enzymes also enhanced deltamethrin degradation when the mutants were reconstituted with the *A. minimus* P450 CYP6AA3 (Sarapusit et al. 2010). Both Phe^{86} and Phe^{219} may be necessary for FMN binding and the FMN domain stabilization and thus aid the P450 in insecticide metabolism.

The highly conserved sequence motifs, Arg-x-Tyr-Ser(Thr) and Met-x-x-x-Gly-Thr(Ser)-Gly(Ala)-Ile-x-Pro, as parts of the FAD- and NADPH-binding domains, respectively, are present in the *A. stephensi* CPR. These sequence motifs are also conserved in other mosquito species. However, the residue at position 459 of the former motif in *A. minimus* and *A. gambiae* CPRs differs from that in other mosquito CPRs. In addition, Sarapusit et al. (2013) recently identified Cys^{427} of the *A. minimus* CPR as another residue possibly involved in FAD binding. This residue is conserved in *A. stephensi*, *A. gambiae*, *Ae. aegypti,* and *Ae. sollicitans* but not in *A. funestus* and *C. quinquefasciatus*. Further experiments will be required to elucidate the effects, if any, of these substitutions on the functions of the enzyme particularly in different organisms.

Interaction between the two protein components is required to transfer electrons to an electron acceptor by a CPR. Nisimoto (1986) originally specified two neighboring clusters consisting of acidic residues

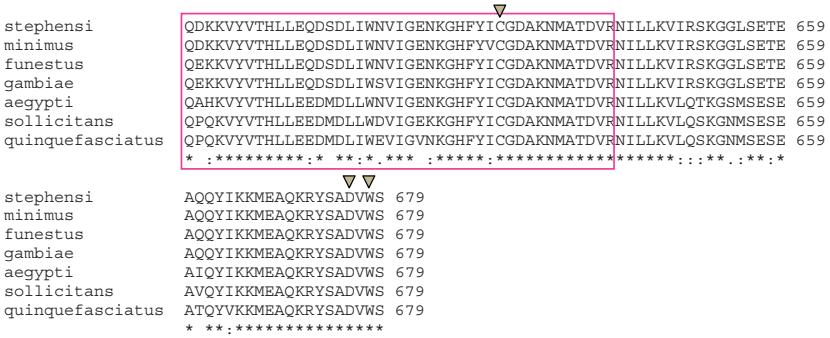

Fig. 5. Continued

Fig. 6. Conserved residues involved in cytochrome P450 or cytochrome *c* interactions. The multiple sequence alignment shows two clusters of acidic residues of mosquito CPRs in comparison with rat CPR. The clusters are boxed, and identical residues are highlighted.

(^{207}Asp-Asp-Asp^{209} and ^{213}Glu-Glu-Asp^{215}) that are important in the interaction with cytochrome *c*. In comparison with the rat enzyme, two acidic clusters (^{210}Asp-Asp-Asp^{212} and ^{216}Glu-Asp-Tyr^{218}) are found in the *A. stephensi* CPR. The first cluster is entirely identical in the two species, but there are two differences in the second cluster. Specifically, one of the two (Asp^{217}) is a conserved substitution that is found in the bacterium *Desulfovibrio vulgaris* Hildenborough flavodoxin (Dubourdieu and Fox 1977). A study based on site-directed mutagenesis in a rat CPR suggests that the first acidic cluster interacts primarily with P450s, whereas the second cluster interacts primarily with cytochrome *c* (Shen and Kasper 1995). It is likely that the *A. stephensi* CPR can perform electron transfers to P450s in the mosquito. Whether the *A. stephensi* enzyme can efficiently transfer electrons to cytochrome *c* will require further investigation. The assumption is supported by activity studies of recombinant CPRs in *A. minimus* (Kaewpa et al. 2007), *A. gambiae* (Lian et al. 2011), and *Ae. sollicitans* (C.S., unpublished data). Although there are similarities between mosquito and vertebrate CPRs, variations in terms of ligand-binding capacities and enzyme kinetics between the two groups are likely (Lian et al. 2011).

Similar to P450s, microsomal CPRs are typically present as integral proteins anchored on the membrane of the endoplasmic reticulum. The anchoring segment of CPRs is a type I signal-anchor sequence with N′-lumen and C′-cytoplasm topology (Kida et al. 1998). The membrane anchoring is also critical for the reductase function. Solubilized reductase without the N′-hydrophobic segment seems to lose its ability to interact with P450s (Black and Coon 1982). As expected, in the case of the *A. stephensi* CPR, the protein does not have a secretion signal peptide but does carry an N′-hydrophobic portion predicted to be a transmembrane region. This suggests that the mosquito CPR can function as part of the electron transfer system to terminal electron acceptors like P450s, which also reside in the microsomal membrane.

In conclusion, analysis of the *A. stephensi* genome reveals a gene encoding for a CPR. The enzyme is highly conserved among mosquitoes and also comparatively similar to CPR enzymes in other organisms. These findings facilitate the understanding of CPR evolution and the role the enzyme plays in electron transfer to its electron acceptor including the P450 family of enzymes, which are often of critical importance in the metabolism of xenobiotic compounds. The results of this study facilitate further elucidation of the mosquito *A. stephensi* CPR functions, including their expected involvement in xenobiotic metabolism.

Acknowledgments

This study was supported in part by Hatch funds from the New Jersey Agricultural Experiment Station.

References Cited

Arensburger, P., K. Megy, R. M. Waterhouse, J. Abrudan, P. Amedeo, B. Antelo, L. Bartholomay, S. Bidwell, E. Caler, F. Camara, et al. 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science 330: 86–88.

Bariami, V., C. M. Jones, R. Poupardin, J. Vontas, and H. Ranson. 2012. Gene amplification, ABC transporters and cytochrome P450s: unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti*. PLoS Negl. Trop. Dis. 6: e1692.

Black, S. D., and M. J. Coon. 1982. Structural features of liver microsomal NADPH-cytochrome P450 reductase: hydrophobic domain, hydrophilic domain, and connecting region. J. Biol. Chem. 257: 5929–5938.

Dixit, J., H. Srivastava, M. Sharma, M. K. Das, O. P. Singh, K. Raghavendra, N. Nanda, A. P. Dash, D. N. Saksena, and A. Das. 2010. Phylogenetic inference of Indian malaria vectors from multilocus DNA sequences. Infect. Genet. Evol. 10: 755–763.

Djouaka, R. F., A. A. Bakare, O. N. Coulibaly, M. C. Akogbeto, H. Ranson, J. Hemingway, and C. Strode. 2008. Expression of the cytochrome P450s, CYP6P3 and CYP6M2 are significantly elevated in multiple pyrethroid resistant populations of *Anopheles gambiae* s.s. from Southern Benin and Nigeria. BMC Genomics 9: 538.

Dubourdieu, M., and J. L. Fox. 1977. Amino acid sequence of *Desulfovibrio vulgaris* flavodoxin. J. Biol. Chem. 252: 1453–1463.

Dym, O., and D. Eisenberg. 2001. Sequence-structure analysis of FAD-containing proteins. Protein Sci. 10: 1712–1728.

Enayati, A. A., and H. Ladonni. 2006. Biochemical assay baseline data of permethrin resistance in *Anopheles stephensi* (Diptera, Culicidae) from Iran. Pakistan J. Biol. Sci. 9: 1265–1270.

Feyereisen, R. 2005. Insect cytochrome P450, pp. 1–77. *In* **L. I. Gilbert, K. Iatrou, and S. S. Gill** (eds.), Comprehensive molecular insect science. Elsevier, Oxford, United Kingdom.

Feyereisen, R. 2012. Insect CYP genes and P450 enzymes, pp. 236–316. *In* **L. I. Gilbert** (ed.), Insect molecular biology and biochemistry. Academic Press, London, United Kingdom.

Ganesh, K. N., J. Urmila, and V. A. Vijayan. 2003. Pyrethroid susceptibility & enzyme activity in two malaria vectors, *Anopheles stephensi* (Liston) & *A. culicifacies* (Giles) from Mysore, India. Indian J. Med. Res. 117: 30–38.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. 2005. Protein identification and analysis tools on the ExPASy server, pp. 571–607. *In* **J. M. Walker** (ed.), The proteomics protocols handbook. Humana Press, Totowa, NJ.

Holt, R. A., G. M. Subramanian, A. Halpern, G. G. Sutton, R. Charlab, D. R. Nusskern, P. Wincker, A. G. Clark, J. M. Ribeiro, R. Wides, et al. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298: 129–149.

Hubbard, P. A., A. L. Shen, R. Paschke, C. B. Kasper, and J. J. Kim. 2001. NADPH-cytochrome P450 oxidoreductase. Structural basis for hydride and electron transfer. J. Biol. Chem. 276: 29163–29170.

Kaewpa, D., S. Boonsuepsakul, and P. Rongnoparut. 2007. Functional expression of mosquito NADPH-cytochrome P450 reductase in *Escherichia coli*. J. Econ. Entomol. 100: 946–953.

Kida, Y., S. Ohgiya, K. Mihara, and M. Sakaguchi. 1998. Membrane topology of NADPH-cytochrome P450 reductase on the endoplasmic reticulum. Arch. Biochem. Biophys. 351: 175–179.

Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S. E. Sachs, and J. Sachs. 2004. A global index representing the stability of malaria transmission. Am. J. Trop. Med. Hyg. 70: 486–498.

Koener, J. F., F. A. Cariño, and R. Feyereisen. 1993. The cDNA and deduced protein sequence of house fly NADPH-cytochrome P450 reductase. Insect Biochem. Mol. Biol. 23: 439–447.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

Levine, R. 2011. i5k: the 5,000 insect genome project. Am. Entomol. 57: 110–113.

Lian, L. Y., P. Widdowson, L. A. McLaughlin, and M. J. Paine. 2011. Biochemical comparison of *Anopheles gambiae* and human NADPH P450 reductases reveals different 2'-5'-ADP and FMN binding traits. PLoS One 6: e20574.

Lycett, G. J., L. A. McLaughlin, H. Ranson, J. Hemingway, F. C. Kafatos, T. G. Loukeris, and M. J. I. Paine. 2006. *Anopheles gambiae* P450 reductase is highly expressed in oenocytes and *in vivo* knockdown increases permethrin susceptibility. Insect Mol. Biol. 15: 321–327.

Matambo, T. S., M. J. Paine, M. Coetzee, and L. L. Koekemoer. 2010. Sequence characterization of cytochrome P450 CYP6P9 in pyrethroid resistant and susceptible *Anopheles funestus* (Diptera: Culicidae). Genet. Mol. Res. 9: 554–564.

Nelson, D. R. 2011. Progress in tracing the evolutionary paths of cytochrome P450. Biochim. Biophys. Acta 1814: 14–18.

Nene, V., J. R. Wortman, D. Lawson, B. Haas, C. Kodira, Z. J. Tu, B. Loftus, Z. Xi, K. Megy, M. Grabherr, et al. 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316: 1718–1723.

Nikou, D., H. Ranson, and J. Hemingway. 2003. An adult-specific CYP6 P450 gene is overexpressed in a pyrethroid-resistant strain of the malaria vector, *Anopheles gambiae*. Gene 318: 91–102.

Nisimoto, Y. 1986. Localization of cytochrome c-binding domain on NADPH-cytochrome P450 reductase. J. Biol. Chem. 261: 14232–14239.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302: 205–217.

Patil, D. P., S. Atanur, D. P. Dhotre, D. Anantharam, V. S. Mahajan, S. A. Walujkar, R. K. Chandode, G. J. Kulkarni, P. S. Ghate, A. Srivastav, et al. 2009. Generation, annotation, and analysis of ESTs from midgut tissue of adult female *Anopheles stephensi* mosquitoes. BMC Genomics 10: 386.

Porter, T. D., and C. B. Kasper. 1986. NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. Biochemistry 25: 1682–1687.

Porter, T. D., T. W. Beck, and C. B. Kasper. 1990. NADPH-cytochrome P-450 oxidoreductase gene organization correlates with structural domains of the protein. Biochemistry 29: 9814–9818.

Sarapusit, S., C. Xia, I. Misra, P. Rongnoparut, and J. J. Kim. 2008. NADPH-cytochrome P450 oxidoreductase from the mosquito *Anopheles minimus*: kinetic studies and the influence of Leu86 and Leu219 on cofactor binding and protein stability. Arch. Biochem. Biophys. 477: 53–59.

Sarapusit, S., P. Lertkiatmongkol, P. Duangkaew, and P. Rongnoparut. 2013. Modeling of *Anopheles minimus* mosquito NADPH-cytochrome P450 oxidoreductase (CYPOR) and mutagenesis analysis. Int. J. Mol. Sci. 14: 1788–1801.

Sarapusit, S., S. Pethuan, and P. Rongnoparut. 2010. Mosquito NADPH-cytochrome P450 oxidoreductase: kinetics and role of phenylalanine amino acid substitutions at leu86 and leu219 in CYP6AA3-mediated deltamethrin metabolism. Arch. Insect Biochem. Physiol. 73: 232–244.

Shen, A. L., and C. B. Kasper. 1995. Role of acidic residues in the interaction of NADPH-cytochrome P450 oxidoreductase with cytochrome P450 and cytochrome c. J. Biol. Chem. 270: 27475–27480.

Shen, A. L., D. S. Sem, and C. B. Kasper. 1999. Mechanistic studies on the reductive half-reaction of NADPH-cytochrome P450 oxidoreductase. J. Biol. Chem. 274: 5391–5398.

Shen, A. L., T. D. Porter, T. E. Wilson, and C. B. Kasper. 1989. Structural analysis of the FMN binding domain of NADPH-cytochrome P450 oxidoreductase by site-directed mutagenesis. J. Biol. Chem. 264: 7584–7589.

Shen, B., H. Dong, H. Tian, L. Ma, X. Li, G. Wu, and C. Zhu. 2003. Cytochrome P450 genes expressed in the deltamethrin-susceptible and -resistant strains of *Culex pipiens pallens*. Pestic. Biochem. Physiol. 75: 19–26.

Shetty, V., D. Sanil, and N. J. Shetty. 2012. Insecticide susceptibility status in three medically important species of mosquitoes, *Anopheles stephensi*, *Aedes aegypti* and *Culex quinquefasciatus*, from Bruhat Bengaluru Mahanagara Palike, Karnataka, India. Pest Manag. Sci. 69: 257–267.

Suwanchaichinda, C., and L. B. Brattsten. 2002. Induction of microsomal cytochrome P450s by tire-leachate compounds, habitat components of *Aedes albopictus* mosquito larvae. Arch. Insect Biochem. Physiol. 49: 71–79.

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

Tikar, S. N., M. J. Mendki, A. K. Sharma, D. Sukumaran, V. Veer, S. Prakash, and B. D. Parashar. 2011. Resistance status of the malaria vector mosquitoes, *Anopheles stephensi* and *Anopheles subpictus* towards adulticides and larvicides in arid and semi-arid areas of India. J. Insect Sci. 11: 85.

Tiwari, S., S. K. Ghosh, V. P. Ojha, A. P. Dash, and K. Raghavendra. 2010. Reduced susceptibility to selected synthetic pyrethroids in urban malaria vector *Anopheles stephensi*: a case study in Mangalore city, South India. Malar. J. 9: 179.

Vontas, J., J. P. David, D. Nikou, J. Hemingway, G. K. Christophides, C. Louis, and H. Ranson. 2007. Transcriptional analysis of insecticide resistance in *Anopheles stephensi* using cross-species microarray hybridization. Insect Mol. Biol. 16: 315–324.

Wang, M., D. L. Roberts, R. Paschke, T. M. Shea, B. S. Masters, and J. J. Kim. 1997. Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. Proc. Natl Acad. Sci. U S A 94: 8411–8416.

Yang, T., and N. Liu. 2011. Genome analysis of cytochrome P450s and their expression profiles in insecticide resistant mosquitoes, *Culex quinquefasciatus*. PLoS One 6: e29418.

Zhu, F., S. Sams, T. Moural, K. F. Haynes, M. F. Potter, and S. R. Palli. 2012. RNA interference of NADPH-cytochrome P450 reductase results in reduced insecticide resistance in the bed bug, *Cimex lectularius*. PLoS One 7: e31037.

Received 1 February 2013; accepted 18 March 2013.