Reexamination of *Culex pipiens* Hybridization Zone in the Eastern United States by Ribosomal DNA-Based Single Nucleotide Polymorphism Markers

Shaoming Huang,* Goudarz Molaei, and Theodore G. Andreadis

*Center for Vector Biology & Zoonotic Diseases, Connecticut Agricultural Experiment Station, New Haven, Connecticut*

**Abstract.** Mosquitoes in the *Culex pipiens* complex are important vectors of several disease-causing pathogens, including West Nile virus. In North America, the complex consists of *C. pipiens pipiens* form *pipiens*, *C. pipiens p. molestus*, *C. pipiens quinquefasciatus*, and their hybrids that exhibit substantial diversity in physiology, behavior, and geographic range. Hybridization among these mosquitoes is of concern because of potential implications for disease transmission. Currently, several morphological and molecular markers exist for differentiating members of the *C. pipiens* complex; however, these markers have specific limitations. We report here two highly reliable ribosomal DNA-based single nucleotide polymorphism (SNP) markers, CxpG2T and CxpA2d, for detecting *C. pipiens* complex mosquitoes containing *C. p. quinquefasciatus* alleles. Both CxpG2T and CxpA2d contain one allele that is present in all members of the *C. pipiens* complex, and the other allele is specific to *C. p. quinquefasciatus*. Testing of field populations from the eastern United States showed that these two SNP markers are capable of identifying a south to north gradient of *C. p. quinquefasciatus* and hybrids. The northern limit of detection of *C. p. quinquefasciatus* alleles in this study was in Fort Totten, NY (40.79°N), whereas the southern boundary was determined between Atlanta, GA (33.81°N) and Gainesville, FL (29.64°N). CxpG2T and CxpA2d were more accurate than the ACE-2 marker, and they may conceivably provide comparable resolution with microsatellite markers for detecting *C. p. quinquefasciatus* alleles.

**INTRODUCTION**

Understanding population structure of arthropod vectors is critical for evaluating their roles in disease transmission and developing more effective control strategies. In North America, the *Culex pipiens* complex consists of *C. pipiens pipiens* form *pipiens*, *C. p. p. molestus*, *C. p. quinquefasciatus*, and their hybrids. Members of this species complex are principle vectors of several disease-causing pathogens, including West Nile virus (WNV), and exhibit substantial diversity in physiology, behavior, and geographic range of distribution.

Hybridization among members of the *C. pipiens* complex is of great concern because of the potential implications for disease transmission. It seems that limited hybridization occurs between *C. pipiens* f. *pipiens* and *C. p. p. molestus*. In contrast, *C. p. p. f. molestus* and *C. p. quinquefasciatus* frequently hybridize within a sympatric zone between 36°N and 39°N latitudes in North America. Hybrid populations of *C. p. p. f. molestus* and *C. p. quinquefasciatus* have been traditionally identified by the relative positions of the dorsal (D) and ventral (V) arms (DV/D) of male phallosome as a morphological marker. Recently, Sanogo and others identified hybrid individuals beyond 39°N in Champaign, IL, using DV/D ratio, whereas Kothera and others defined a much wider hybridization zone in the central United States by adopting molecular markers HotAce.2 and microsatellites. Molecular markers (i.e., ACE-2/HotAce.2, CQ11, and microsatellites) have proven to be more accurate for detecting hybrids of *C. p. p. molestus* and *C. p. quinquefasciatus* than the morphological marker DV/D. Microsatellite markers provide more accurate information but are relatively costly and specialized equipment. ACE-2/HotAce.2 and CQ11 are less costly but not as accurate as microsatellites, and CQ11 is too polymorphic to be used as a diagnostic marker.

In this study, we have developed two ribosomal DNA (rDNA)-based single nucleotide polymorphism (SNP) markers, denoted CxpG2T and CxpA2d, that are reliable and easy to use for detecting *C. p. quinquefasciatus*. rDNA has reliably been used as a source of genetic variation in phylogenetic studies. The internal transcribed spacer 1 (ITS1) and ITS2 regions of rDNA have been extensively explored for species differentiation in a variety of organisms because of their low to moderate rate of polymorphism. However, earlier attempts in using ITS1 and ITS2 to differentiate members of the *C. pipiens* complex were unsuccessful because of insufficient variation.

In this study, we examined the intergenic spacer region (IGS), which exhibits higher rates of variation, contains sub-repeats and important promoter, enhancer, and binding sites for protein factors regulating rDNA transcription, and may play an important role in adaptation of populations to local environments. This region has been successfully used in developing molecular markers for other complex mosquito species such as malaria vectors in the *Anopheles gambiae* species complex.

**MATERIALS AND METHODS**

**Mosquito sampling and identification.** Populations of *C. pipiens* complex mosquitoes were sampled from 10 different geographic locations (Figure 1). Female mosquitoes were primarily sampled using CO₂-baited Centers for Disease Control and Prevention (CDC) miniature light traps or gravid traps baited with hay infusion. Mosquitoes from New York City (NYC) were sampled with hand-held aspirators from above-ground hibernacula located at Fort Totten in the borough of Queens. An underground population of *C. p. p. molestus* was sampled from sewer catch basins located at 91st Street in the Borough of Manhattan, NYC using a battery-powered modified CDC backpack aspirator (John W. Hock Co., Gainesville, FL). This population of
Cx. p. pipiens f. molestus has been found to be a pure Cx. p. pipiens f. molestus population in an earlier study. All mosquito specimens from Cambridge, MA, Hartford, CT, Fort Totten, NY, Manhattan, NYC, Trenton, NJ, and Houston, TX, have also been analyzed by microsatellite markers in the previous study. Mosquitoes were transported to the laboratory alive in coolers (4–8°C) with ice packs, inactivated on dry ice, or preserved in 95–100% ethanol. Mosquitoes were first identified based on morphological characters as members of the Cx. pipiens complex and subsequently, were subjected to a species-specific polymerase chain reaction (PCR) tests based on rDNA for confirmation. Identified specimens were either processed immediately for genomic DNA extraction or stored at −80°C.

Genomic DNA extraction. The thorax of each female mosquito was used for genomic DNA extraction to avoid cross-contamination from sperm stored in the spermatheca. The thorax was homogenized with a disposable microtube pestle (USA Scientific, Enfield, CT) in a 1.5-mL tube containing 180 μL phosphate buffer saline (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4, 1.47 mM KH2PO4) and subjected to genomic DNA extraction using DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA) according to the manufacturer’s recommended protocol for insect tissues. Isolated DNA from each mosquito was reconstituted in 50 μL AE buffer (10 mM Tris·Cl, 0.5 mM ethylenediaminetetraacetic acid [EDTA], pH 9.0; Qiagen) and stored at −20°C for later use.

Amplification, cloning, and sequencing of full-length IGS. Eukaryotic rDNA is arranged in tandemly repeated units (Figure 2) containing the genes for the 18S, ITS1, 5.8S, ITS2, and 28S (in the order of 5′ end to 3′ end) separated by spacers. The sequence flanking the 3′ end of one array and the 5′ end of the next array is known as IGS. A pair of universal primers Cxp28SF and Cxp18SR located at the 3′ end of the 28S and the 5′ end of the 18S, respectively, were designed to amplify the full-length IGS region (Table 1 and Figure 2).

The sequences of the conserved regions used for alignments and design of universal primers were obtained from GenBank. Sequences for the 28S were from insects (GenBank accession numbers M21017 for Drosophila melanogaster, L22060 for Aedes albopictus, and L78065 for An. albimanus), tick (AF291874), copepod (DQ180347), nematodes (AY580057 and AJ512837), and flatworm (EF095264). Sequences for the 18S were from insects (M21017 for D. melanogaster, EU057177 for Helicoverpa assulta, AY216700 for Aphanlus gossypii, and EU179519 for Anastrepha ludens), frog (X04025), bird (TURRR18S), scorpion (AF005442), human (X03205), and rat (X01117).

PCR amplification products from three individuals of each pure population of Cx. p. pipiens f. pipiens (Hartford, CT), Cx. p. pipiens f. molestus (Manhattan, NYC), and Cx. p. quinquefasciatus (Houston, TX) were cloned for full-length IGS identification. The GeneAmp XL PCR Kit (Applied Biosystems, Foster City, CA) that contains rTth DNA polymerase, XL for increased fidelity, and long products was used for PCR amplifications. The 50-μL PCR reactions contained 0.2 μL (~1 ng) genomic DNA, 1× XL Buffer II (Applied Biosystems), 200 μM each dNTP, 0.2 μM each primer, 1.1 mM Mg (OAc)2, and 0.2 U rTth DNA polymerase.
1.5 units *Tth* DNA Polymerase XL (Applied Biosystems). PCR reactions were initially denatured at 94°C for 1 minute followed by 16 cycles of amplification at 94°C for 15 seconds and 68°C for 3 minutes and an additional 14 cycles of amplification at 94°C for 15 seconds and 68°C for 3 minutes with 15-second increments per cycle; they were finally extended at 72°C for 10 minutes.

PCR products were analyzed using a 0.8% low-melting agarose gel where bands corresponding to 1.5–5 kb were excised and purified using S.N.A.P. UV Free Gel Purification Kit (Invitrogen, Carlsbad, CA). Purified DNA was cloned into pCR-XL-TOPO vector and subsequently transformed into One Shot TOP10 chemically competent *Escherichia coli* cells using TOPO XL PCR Cloning Kit (Invitrogen). *E. coli* cells were spread onto Luria–Bertani (LB) agar plates containing 50 μg/mL kanamycin and cultured overnight. For each clone, five well-isolated colonies were randomly selected and cultured overnight in LB liquid medium containing 50 μg/mL kanamycin. Plasmids from all bacterial colonies were isolated using the QIAprep Miniprep kit (Qiagen, Santa Clara, CA). Plasmid DNAs were then subjected to EcoRI restriction assay to ensure that the cloned fragment was the correct size. Finally, the cloned full-length IGS was sequenced by primer walking.

Sequences were assembled using SeqMan II, Lasergene 7.0 (DNASTAR Inc., Madison, WI). All sequences were verified as rDNA fragments, and putative boundaries were annotated by searching against GenBank and *Cx. pipiens* genome database (Broad Institute, Cambridge, MA) using nucleotide blast program. Putative full-length IGS sequences were submitted to GenBank (accession numbers GU911305 to GU911348). Finally, all sequences were aligned in SeqMan II to identify SNP loci.

**Design of SNP genotyping by restriction assay.** An SNP containing a G–T transversion, designated CxpG2T, was located downstream of IGS (Figure 2). Another SNP containing A to deletion (hereafter called −) mutation, designated CxpA2d, was identified at the 3′ end of the 28S (Figure 2).

For CxpG2T, allele G is specific to *Cx. p. quinquefasciatus* and T is present in all members of the *Cx. pipiens* complex, indicating that the genotypes of *Cx. p. pipiens* f. pipiens and *Cx. p. p. molestus* are −, −, A−, and A−, respectively. To detect allele A, primer set CxpA2dF and CxpA2dEcoRI (Table 1) amplifying a 158-bp fragment was designed so that the reverse primer

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Amplicon size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cxp28sF</td>
<td>5′-AGG GAA CTG GAG CTG GGT TTA G-3′</td>
<td>2.8 kb</td>
</tr>
<tr>
<td>Cxp18sR</td>
<td>5′-TGC CCT CCC TGG ATG TGG TAG C-3′</td>
<td></td>
</tr>
<tr>
<td>CxpG2TF</td>
<td>5′-AGG ACC CGC AGA GCA TGG CTA CCT TCG G-3′</td>
<td>239 bp</td>
</tr>
<tr>
<td>CxpG2T/ApaLI</td>
<td>5′-GTA CGA GAA AGC AGT GCG TCG CCC AAC AG-3′</td>
<td></td>
</tr>
<tr>
<td>CxpA2dF</td>
<td>5′-TTC GAT GTG GCG TCT TCC TAT CAT TGT G-3′</td>
<td>158 bp</td>
</tr>
<tr>
<td>CxpA2d/EcoRI</td>
<td>5′-TCC GTT AGA ATA GCG TGC TGT GAA-3′</td>
<td></td>
</tr>
</tbody>
</table>

F = forward primer; R = reverse primer.
(CxpA2dEcoRI) contained a T to A mismatch at the second position at the 3' end (Figure 4A). Introduction of this mismatch creates a recognition site GAATTC for the restriction enzyme EcoRI when A is present (Figure 4A). On EcoRI digestion, the amplicon would be cut into 131- and 27-bp fragments if only A is present and would be intact if only deletion is present (Figure 4B). Thus, one band containing a 158-bp fragment would be present for Cx. p. pipiens f. pipiens and Cx. p. pipiens f. molestus, and three bands containing 158, 131, and 27 bp, respectively, would be present for Cx. p. quinquefasciatus and its hybrids with Cx. p. pipiens f. pipiens (Figure 4B).

Both CxpG2T and CxpA2d contain one allele that is present in all members of the Cx. pipiens complex, whereas the other allele is specific to Cx. p. quinquefasciatus. This polarity determines that CxpG2T and CxpA2d exclusively detect individuals that contain Cx. p. quinquefasciatus genetic signature, but they do not differentiate pure Cx. p. quinquefasciatus and its hybrids with Cx. p. pipiens f. pipiens; this is because both contain the two alleles. These two SNP markers also do not differentiate Cx. p. pipiens f. pipiens, Cx. p. pipiens f. molestus, or their hybrids.

PCR reactions for both SNPs were performed in 20 μL reaction mix containing 0.4 μL genomic DNA, 2 μL 10× PCR buffer II (Applied Biosystems), 2 μL 25 mM MgCl₂, 0.4 μL 10 mM 2' deoxynucleoside 5' triphosphate (dNTP) mix, 0.4 μL 10 μg/μL bovine serum albumin (BSA), 0.2 μM each primer, and 0.5 units AmpliTaq Gold DNA polymerase (Applied Biosystems). PCR reactions were initially denatured at 95°C for 10 minutes followed by 30 cycles of amplification at 95°C for 15 seconds, 65°C for CxpG2T or 61°C for CxpA2d for 20 seconds, and 72°C for 20 seconds, and finally extended at 72°C for 7 minutes. ApaLI digestion assays were conducted in 30 μL containing 10 μL PCR product, 2 μL 10× buffer Tango (Fermentas, Ontario, Canada), and 10 units ApaLI (Alw44I) and incubated at 37°C for 4 hours. EcoRI digestion assays were carried out in 30 μL containing 10 μL PCR product, 2 μL 10× NEBuffer EcoRI (New England Biolabs, Ipswich, MA), 0.2 μL 10 μg/μL BSA, and 15 units EcoRI (New England Biolabs) and incubated at 37°C for 4 hours. Digested PCR products were visualized under ultraviolet (UV) light on ethidium bromide-stained 2.5% agarose gel.

**Field testing.** To validate the two SNP markers, 32 adult female mosquitoes for each population were subjected to PCR restriction assays. To ensure that the two SNPs were not created by polymerase errors and/or cloning artifacts, one specimen for each genotype was randomly selected from each field population and subjected to PCR amplification by using a proofreading enzyme, AccuPrim Pf/ DNA Polymerase (Invitrogen). PCR products were then directly sequenced to examine the presences of each allele. For comparison purposes, hybrid detection by ACE-2 marker was also performed in parallel on all populations.

**RESULTS**

We obtained 44 cloned full-length IGS sequences (accession numbers GU911305 to GU911348), all of which had approximately the same size of 1.8 kb because of the lack of 60 bp, ~200 bp, or other size subrepeats commonly found in other Dipteron insects. The IGS region in the present study contained many microsatellite sequences (simple sequence repeats) of mostly 2–4 bp. When all microsatellite sequences were removed, the IGS sequences shared greater than 90% identity (data not shown). There was greater sequence identity within each group of Cx. pipiens and Cx. f. molestus than between those of different groups (Figure 5). However, a few individuals from the Cx. p. pipiens f. pipiens group exhibited higher sequence identity to individuals from Cx. p. quinquefasciatus in the IGS region than those individuals within each group. Cx. p. pipiens f. molestus mosquitoes showed closer relatedness to Cx. p. pipiens f. pipiens mosquitoes in the IGS sequences than they did to Cx. p. quinquefasciatus (Figure 5). These full-length IGS sequences also contained several SNPs. By thorough screening of these SNPs, we identified two markers, designated CxpG2T and CxpA2d, that consistently detected the Cx. p. quinquefasciatus genetic signature among Cx. pipiens complex mosquitoes.

CxpG2T and CxpA2d successfully amplified corresponding sequences from all field-sampled specimens. The restriction banding patterns were consistent. Testing of all field populations enabled us to identify a south to north gradient of Cx. p. quinquefasciatus and hybrids. The southernmost and
northernmost ends contained pure *Cx. p. quinquefasciatus* and *Cx. p. p. pipiens* f. p. pipiens, respectively. Varying degrees of individuals with *Cx. p. quinquefasciatus* genetic signature were identified at both the northernmost and southernmost end of the hybridization zone (Figure 1). The northern detection limit of *Cx. p. quinquefasciatus* genetic signature by CxpG2T and CxpA2d was at Fort Totten, NY. CxpG2T showed greater resolution than CxpA2d in detecting individuals bearing *Cx. p. quinquefasciatus* alleles within the hybridization zone except in Trenton, NJ (Figure 1). Likewise, both CxpG2T and CxpA2d were more reliable than ACE-2 except in Fairfax, VA. The northern detection limit of *Cx. p. quinquefasciatus* genetic signature by ACE-2 was also much farther south (Fairfax, VA; 38.854°N) compared with Fort Totten, NY (40.796°N) as identified by CxpG2T and CxpA2d.

All mosquitoes sampled from the southernmost locations (Gainesville, FL and Houston, TX) contained genetic signatures of *Cx. p. quinquefasciatus*, which were examined by CxpG2T, CxpA2d, and ACE-2, suggesting that the southern boundary of the hybridization zone of *Cx. p. p. pipiens* f. p. pipiens and *Cx. p. quinquefasciatus* is likely located between Atlanta, GA and Gainesville FL, approximately 32°N. None of the markers used in the present study identified *Cx. p. quinquefasciatus* alleles in any individuals in the *Cx. p. p. pipiens* f. molestus population sampled from an underground habitat in Manhattan, NYC.

**DISCUSSION**

Understanding genetic variation and species distribution boundaries is vital for evaluating the respective role of local vector populations in disease transmission. Among molecular markers for population genetic studies, SNPs offer rapid and cost-efficient technology to resolve some of the outstanding questions regarding species differentiation and hybrid groups. Based on the rDNA IGS region, we developed two SNP markers, CxpG2T and CxpA2d, to evaluate the extent of introgression of *Cx. p. p. pipiens* f. p. pipiens and *Cx. p. quinquefasciatus* in *Cx. pipiens* complex and redefine the hybridization zone in the eastern United States that extends beyond the currently recognized boundaries. Although the core domains of the rDNA coding regions have been highly conserved, the IGS region intervening between the 28S and 18S rRNA subunits often displays sequence divergence and length variation among members of closely related species or among local populations of the same species, thus allowing differentiation of morphologically inseparable cryptic species and characterization of
SNP MARKERS FOR CULEX PIPIENS COMPLEX

Figure 6. Dot plot matrices of IGS sequences of six Dipteran insect species revealing different sizes of subrepeats. Symmetric lines forming a square box along the central diagonal indicate presence of a subrepeat. The size of the square dictates the size of the subrepeat. Dot plots were generated using the Dotplot program with an 11-bp sliding window. Symmetric dots or lines along the diagonal represent the repeat number and length. GenBank accession numbers for IGS sequences used for the plot were AF004986 (Ae. aegypti), AGU10135 (An. gambiae), AF498237 (An. sinensis), M65063 (Ae. albopictus), L17048 (D. funebris), and GU911335 (Cx. p. quinquefasciatus).
by slave ships. Testing of these hypotheses could be a challenging task further complicated by multiple mosquito introductions because of increasing global trade and transportation. Recent global genetic analyses of Cx. p. quinquefasciatus in North America,\textsuperscript{32-34} and at least a second introduction of Cx. p. quinquefasciatus into Hawaii has been suggested.\textsuperscript{35,36} In our study, both CxpG2T and CxpA2d contain one allele that is present in all Cx. p. pipiens complex members, whereas the other allele is specific to pure Cx. p. quinquefasciatus, allowing for the detection of a south to north gradient penetration of Cx. p. quinquefasciatus ancestry. This polarity in CxpG2T and CxpA2d suggests that Cx. p. p. pipiens and Cx. p. quinquefasciatus may have been derived from a single ancestral population.

In conclusion, we found that IGS-based SNP markers CxpG2T and CxpA2d are reliable and cost-effective (minimal cost incurred by the additional restriction digestion step) for the identification of Cx. p. pipiens complex mosquitoes bearing Cx. p. quinquefasciatus genetic signature and reexamination of the hybridization zone. Our finding also indicates that the IGS region can be successfully used for the development of convenient markers such as SNPs, and in conjunction with other molecular markers (e.g., microsatellite markers), it can accommodate characterization of genetic variation, identification of morphologically identical incipient species, and understanding of the evolution of Cx. p. pipiens complex mosquitoes. Finally, with the release of the Cx. p. quinquefasciatus genome sequence,\textsuperscript{41,42} it is anticipated that additional studies will take advantage of the genome sequence to develop higher-resolution markers for examining genetic variation pertinent to population differentiation as well as associations with heterogeneity in disease transmission.

Received November 29, 2010. Accepted for publication June 8, 2011.

Acknowledgments: The authors thank Arv Farajollahi, Mercer County Mosquito Control for providing mosquitoes from Trenton, NJ; Dr. Jorge Arias, Fairfax County Disease Carrying Insects Program for sampling mosquitoes from Fairfax, VA; Dr. Charles Apperson, North Carolina State University for providing mosquitoes from Raleigh, NC; Dr. Rosemarie Kelly, Georgia Department of Health for providing mosquitoes from Atlanta, GA; and Dr. Daniel Kline, US Department of Agriculture, Agricultural Research Service, Mosquito and Fly Research Unit for sampling mosquitoes from Gainesville, FL. We are also grateful for the technical assistance of the Connecticut Agricultural Experiment Station support staff, John Sheppard, and Michael Thomas.

Financial support: Funding for this research was provided in part by Laboratory Capacity for Infectious Diseases Cooperative Agreement number U50/CCU68086-01 from the Centers for Disease Control and Prevention, US Department of Agriculture Specific Cooperative Agreement number 58-6615-1-218, and US Department of Agriculture-administered Hatch funds CONH0768 (to T.G.A.).

Authors’ addresses: Shaoming Huang, Goudarz Molaei, and Theodore G. Andreadis, Center for Vector Biology and Zoonotic Diseases, Connecticut Agricultural Experiment Station, New Haven, Connecticut, E-mails: Shuang@sjmosquito.org, Goudarz.Molaei@ct.gov, and Theodore.Andreadis@ct.gov.

REFERENCES


