Microsatellite Characterization of Subspecies and Their Hybrids in *Culex pipiens* Complex (Diptera: Culicidae) Mosquitoes Along a North–South Transect in the Central United States

LINDA KOTHERA,1,2 ERIN M. ZIMMERMAN,1 CHRISTOPHER M. RICHARDS,3 AND HARRY M. SAVAGE1


ABSTRACT Mosquitoes in the *Culex pipiens* complex, *Cx. p. pipiens* L. and *Cx. p. quinquefasciatus* Say are morphologically similar and important vectors of West Nile and St. Louis Encephalitis viruses in the United States. The subspecies differ with respect to overwintering strategies, with *Cx. p. pipiens* entering diapause in response to winter conditions and *Cx. p. quinquefasciatus* lacking this ability, yet they hybridize when found in sympatry. Specimens (*n* = 646) were collected using gravid traps set along a transect from New Orleans, LA, to Chicago, IL. Microsatellite markers were used to: genetically characterize subspecies and hybrids, determine the degree and extent of hybridization to better define the hybrid zone, and examine the efficacy of hybrid detection between microsatellites and a single-gene assay based on the acetylcholinesterase2 gene (HotAce.2). The results support the presence of two distinct genetic entities, with a broad, stable hybrid zone in between. Admixture analyses classified >40% of individuals as hybrids. Allelic richness was markedly different at the northern and southern ends of the transect, and there was a significant isolation by distance effect. The hybrid zone seems to be wider and extends further to the south than previous work indicated, and as a result, we propose new boundaries compared with those indicated by a previous study. Microsatellites detected more hybrids than the HotAce.2 assay, although the latter assay remains useful as a screening tool. We suggest that the parental subspecies and the hybrid zone are maintained by heterosis combined with selection for diapause at northern latitudes.

KEY WORDS *Culex pipiens* complex, *Culex quinquefasciatus*, hybridization, microsatellites, transect

Closely related taxa with subspecific status may show significant ecological differences. North American subspecies of the *Culex pipiens* complex, *Cx. p. pipiens* L. or the northern house mosquito and *Culex pipiens quinquefasciatus* Say or the southern house mosquito are important vectors of West Nile (WN) and Saint Louis Encephalitis (SLE) viruses (family Flaviridae, genus Flavivirus) and show different behavioral traits, morphological characteristics, and geographic distributions (Barr 1957, Wilton and Smith 1985, Savage et al. 2006). However, these subspecies hybridize readily in the middle latitudes of the United States (Barr 1957), the offspring of these crosses are fertile (Rozeboom 1951), and stable hybrid populations seem to exist (Tabachnick and Powell 1983, Cornel et al. 2003, Savage et al. 2008). At the ends of their respective geographic ranges, the two subspecies are most reliably distinguished by morphological differences in male genitalia and by the ability/ inability to enter a winter resting phase known as diapause (Wilton and Smith 1985, Miller et al. 1996). Diapause by adult *Cx. p. pipiens* females infected with arboviruses provides an overwintering mechanism for West Nile and other viruses in northern latitudes (Eldridge 1968, Nasci et al. 2001, Anderson and Main 2006).

Sundararaman (1949) proposed a morphological method of distinguishing *Cx. p. pipiens* from *Cx. p. quinquefasciatus* by comparing the genitalia of male mosquitoes, where measurements of the distance among the dorsal and ventral arms of the phallosome produce taxon-specific ratios, known as DV/D ratios. Barr (1957) measured DV/D ratios of specimens in North America and determined that pure *Cx. p. pipiens* populations were found above 39° north latitude, and pure *Cx. p. quinquefasciatus* populations were found below 36° north latitude, with a hybrid zone consisting of both subspecies and hybrids located between 36° and 39° north latitude. Subsequent studies based on both morphological and genetic data have suggested that DV/D ratios are reliable indicators of subspecific identify only in northern and southern areas where relatively pure populations of *Cx. p. pipiens* or *Cx. p. quinquefasciatus* are present. In middle latitudes where hybridization is extensive, DV/D ratios seem weakly correlated with the genetic identification of...

Although the above-mentioned studies are regionally informative, a full population genetic analysis of samples collected along a transect across the central United States is lacking. Transect studies are especially useful to define the limits of gene dispersal and can better characterize the degree of hybridization in areas of contact between taxa. In addition, it is possible that biologically important traits, for example, those associated with the ability to transmit WN and SLE viruses, or blood-meal host preferences (Savage et al. 2007), could be transferred between Cx. p. pipiens and Culex p. quinquefasciatus as well. We sampled specimens along a transect across the Mississippi River Basin from New Orleans, LA, to Chicago, IL. We selected this study area because of its history of intense SLE and WNV transmission to humans and because members to the Cx. p. pipiens complex are the primary vectors of these viruses in this area (Mitchell et al. 1980, Hayes et al. 2005).

Microsatellites are a popular choice for population and subspecies level studies, particularly to characterize hybridization, because they exhibit high levels of polymorphism that can be used to describe the genetic variation within populations and estimate the degree of genetic differentiation among populations (Selkoe and Toonen 2006). Several papers have described microsatellite loci that amplify in members of the Cx. p. pipiens complex (Fonseca et al. 1998, Keyghobadi et al. 2004, Smith et al. 2005, Edillo et al. 2007), and this study made use of published markers. Previous attempts to use genetic means to identify members of the Cx. p. pipiens complex have included subtractive hybridization (Crabtree et al. 1997), restriction fragment length polymorphisms (RFLPs) (Severini et al. 1998), and polymerase chain reaction (PCR) assays based on polymorphisms in the introns of the acetylcholinesterase 2 gene: ACE.2 (Aspen and Savage 2003) and HotAce.2 (Savage et al. 2007). As with any single-gene assay, first generation (F1) hybrids are readily detectable, but the frequency of later-generation hybrids may be underestimated because of recombination (i.e., when two F1’s cross, one half of the offspring will have markers for one or the other parent subspecies only).

The aims of this study were the following: (1) describe the genetic composition of Cx. p. pipiens, Cx. p. quinquefasciatus, and hybrids along a transect; (2) describe the degree of hybridization in admixed individuals; (3) use this information to better define the limits of the hybrid zone in this part of the United States; (4) establish baseline data for further studies of the genetic structure of the hybrid zone and the Cx. p. pipiens complex; and (5) compare the efficacy of the single gene HotAce.2 assay (Savage et al. 2007) as a diagnostic marker to detect hybrids with a microsatellite assay.

### Materials and Methods

**Sampling Methods and Description of Study Sites.** During August 2005, mosquitoes were collected along a 1,429-km transect from New Orleans, LA, to Chicago IL using gravid traps (Hausherr’s Machine Works, Toms River, NJ) baited with 4–5 liters of a horse manure/alfalfa hay infusion. Collections occurred over two or three trap-nights at each site. Trap nets were placed on dry ice, and specimens were transferred to labeled cryotubes (Nunc, Rochester, NY). Samples were stored on dry ice or held at −70°C until they were processed.

A total of 17 sites were sampled, and specimens were later consolidated to 14 sites. Examination of the data subsequent to their collection resulted in pooling data from three pairs of sites because of close geographic proximity (range, 0.6–7.1 km apart): the two southernmost sites in New Orleans, LA, the two northernmost sites in Chicago, IL, and the two sites located in Memphias, TN. Consolidation occurred after a comparison of pairwise FST values among sites confirmed that the pairs of sites in question were more similar to each other than to the adjacent site(s) (data not shown). The longitude and latitude values for consolidated sites were calculated from the midpoints along a straight line between respective pairs of sites. The data were then treated as coming from 14 sites, which were separated by an averaged distance of 112.4 km (range, 82.3–142.59 km). Geographic details of the 14 sites are given in Table 1.

**Initial morphological and genetic screening.** All specimens collected along the transect (n = 3,875) were examined morphologically and identified to species or lowest taxonomic unit as described previously by Savage et al. (2007). A random subsample of 50 Cx. p. pipiens complex adult female specimens was selected for microsatellite analysis from each site (n = 646), except for two sites, Effingham, IL and Kankakee, IL, which yielded only 31 and 15 Cx. p. pipiens complex mosquitoes, respectively.

DNA was extracted from individual mosquitoes, and membership in the Cx. p. pipiens complex was confirmed with the internal transcribed spacer of ribosomal DNA (ITS) assay (Crabtree et al. 1995, Aspen et al. 2003). In addition, the HotAce.2 assay was used, and the level of agreement between this
assay and the full microsatellite analysis was examined. DNA extraction and initial genetic screening using the ITS and HotAce.2 assays followed the protocol by Savage et al. (2007), with the following changes: the entire body was used in the DNA extraction, ≈20 ng of DNA and 0.5 U HotStar Taq (Qiagen, Valencia, CA) were used, and the number of PCR cycles was increased to 40.

Selection of Microsatellite Markers and PCR and Sequencer Conditions. A panel of microsatellite loci was selected after conducting a literature search for markers that would be polymorphic in both subspecies and suitable for population genetic analysis (Table 2). The forward primer of each pair was fluorescently labeled, and primer pairs were multiplexed such that all loci could be run in two PCR reactions. Each reaction contained ≈20 ng of DNA, 1X PCR buffer, 2.5 mM MgCl₂, 200 μM of each dNTP (ABS, Foster City, CA), 0.5 U HotStar Taq polymerase (Qiagen, Valencia, CA), and primers in the concentrations listed in Table 2. PCR was carried out in 20-μl volumes in an MJ Research PTC-200 thermal cycler (Bio-Rad, Hercules, CA) with the following touch-down PCR program: 95°C (10 min) to activate the hot-start Taq, and then 96°C (5 min) as a denaturing step; 17 cycles of 94°C for 45 s, 62–54°C (decreasing 0.5°C cycle) for 45 s, and 72°C for 45 s; 13 cycles of 94°C for 45 s, 54°C for 45 s, and 72°C for 45 s; and 72°C (10 min). Amplification products were detected on a Beckman-Coulter (Fullerton, CA) CEQ8000 sequencer according to the manufacturer’s protocols and using a 400-bp size standard. The resulting electropherograms were analyzed with the Fragment Analysis Module software included with the CEQ8000 sequencer to generate a multilocus genotype for each specimen. Approximately 10% of specimens were analyzed twice, and the results from the first and second runs were identical.

Data Analysis. Convert (Glaubitz 2004) was used to format microsatellite data for analyses in Arlequin 3.1 (Excoffier et al. 2005) and Genepop 4.0 (Raymond and Rousset 1995). Levels of observed (Hₑ) and expected (Hₑₑ) heterozygosity using Nei’s unbiased estimate (Nei 1987) were calculated using Arlequin. Allelic richness was calculated using a sample size corrected method in FSTAT 2.9.3 (Goudet 1995). Hardy-Weinberg equilibrium (HWE) was evaluated with the probability test option of Genepop, which uses the Markov chain method of randomization to determine whether any locus-population pairs show an excess or deficit of heterozygotes. Characterization of departures from HWE was carried out using Weir and Cockerman’s (1984) estimate of Wright’s inbreeding coefficient, Fₛₑ. Genetic disequilibrium was assessed by permutation tests in FSTAT, which compared the 28 possible combinations of loci in each population.

An analysis of molecular variance (AMOVA) was conducted in Arlequin to obtain estimates of Weir and Cockerman’s (1984) F statistics, and to determine the manner in which genetic variance was partitioned in this system: within individuals, within sites, and among sites. Interpopulation Fₛₑ values were calculated in FSTAT and regressed, using the method of Rousset (1997), on a matrix of geographic distances in Genepop to ascertain whether there was an isolation by distance (IBD) effect. Mantel’s tests were conducted to determine the significance of the IBD effect.

The program Population Graphs (Dyer and Nason 2004) was used to graphically portray genetic relationships among populations. Population Graphs uses a technique similar to an AMOVA (Excoffier et al. 1992), in that it makes use of genetic distances between individuals, which are used to construct a network of nodes, the sizes of which are proportional to the within-population variance, and straight lines called edges, which are proportional to the among-population variance. The resulting network connects populations with statistically significant covariance while leaving populations with an independent covariance structure unconnected. The method used to construct a network is nonhierarchical, allows for reticulate relationships, and can elucidate patterns that may be difficult to detect when observing tabular data, such as those populations that serve as bridges to gene flow.

The model-based clustering program Structure (Pritchard et al. 2000) uses multilocus genotype data to infer individual membership in one or more clusters (K). Because it was unknown how the pro-

Table 2. Microsatellite loci used in this study

<table>
<thead>
<tr>
<th>Locus</th>
<th>N</th>
<th>Mult.</th>
<th>Size (bp)</th>
<th>5' label</th>
<th>Concentration (μM)</th>
<th>Ref.</th>
<th>Sp. allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>CxqTri4</td>
<td>5</td>
<td>1</td>
<td>114-126</td>
<td>D4</td>
<td>0.02</td>
<td>3</td>
<td>Cpp (3)</td>
</tr>
<tr>
<td>CxqGT4</td>
<td>8</td>
<td>1</td>
<td>137-159</td>
<td>D4</td>
<td>0.04</td>
<td>3</td>
<td>Cpq (1)</td>
</tr>
<tr>
<td>CxqCT56</td>
<td>19</td>
<td>1</td>
<td>247-286</td>
<td>D3</td>
<td>0.25</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CxqCT10</td>
<td>5</td>
<td>1</td>
<td>97-112</td>
<td>D3</td>
<td>0.10</td>
<td>4</td>
<td>Cpp (1)</td>
</tr>
<tr>
<td>CxqCAGI01</td>
<td>4</td>
<td>1</td>
<td>178-193</td>
<td>D3</td>
<td>0.10</td>
<td>4</td>
<td>Cpp (3)</td>
</tr>
<tr>
<td>CxqCT51</td>
<td>28</td>
<td>2</td>
<td>57-214</td>
<td>D4</td>
<td>0.10</td>
<td>2</td>
<td>Cpp (4)</td>
</tr>
<tr>
<td>CQ26</td>
<td>18</td>
<td>2</td>
<td>157-245</td>
<td>D3</td>
<td>0.15</td>
<td>1</td>
<td>Cpp (1)</td>
</tr>
<tr>
<td>CxqATCG9</td>
<td>15</td>
<td>2</td>
<td>212-267</td>
<td>D4</td>
<td>0.15</td>
<td>4</td>
<td>Cpq (2)</td>
</tr>
</tbody>
</table>

b Sp. allele, subspecies specific alleles. No. subspecies-specific alleles for each subspecies Cx. p. pipiens (Cpp) and Cx. p. quinquefasciatus (Cpq) (see Results).

N. no. alleles; Mult., multiplex group for PCR; 5' label, Beckman-Coulter fluorescent dye end labeling; Concentration, PCR primer concentration.
<table>
<thead>
<tr>
<th>Site</th>
<th>N_i</th>
<th>Locus name</th>
<th>Mean A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CQ26</td>
<td></td>
</tr>
<tr>
<td>New Orleans</td>
<td>50</td>
<td>CxqGT4</td>
<td>3.740</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CxqTri4</td>
<td>0.642</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CxqGT46</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CxqGT51</td>
<td>0.750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CxqATG9</td>
<td>0.679</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CxqCTG10</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CxqCAG101</td>
<td>0.667</td>
</tr>
</tbody>
</table>

Values in bold indicate significant departures from HWE (P < 0.05 after Bonferroni correction).

* NI: no. individuals sampled; Mean, average N_i; H_o, H_e, A: allelic richness averaged over all loci; N_o: no. alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; Mean N_o, mean no. alleles with SD per locus; θ, estimate of F stutter (Weir and Cockerham 1984).

**Table 3. Genetic diversity and differentiation measures for 14 sites in this study**

gram would interpret genetic differences among Cx. pipiens complex subspecies and their hybrids (i.e., whether hybrids would comprise a separate cluster). Structure was first used to determine the most likely number of clusters, with K ranging from 1 to 14 clusters. Each run had 30,000 burn-in steps and 100,000 data collection steps, used the admixture model with correlated allele frequencies, and was carried out without prior population information.

The ΔK method of Evanno et al. (2005) was used to graphically portray the change in posterior probability of the data [L(K)] with successive numbers of clusters, such that L(K) = L(K + 1) - L(K), and ΔK is equal to the absolute value of L(K) averaged over a number of runs (we used 10) and divided by the SD of L(K). The probable number of clusters is associated with the largest ΔK value.

Structure was run to assign individuals to clusters, using individual proportion of membership (q) values from the stability analysis run (see below) with the highest log likelihood value. A classification threshold of q ≥ 0.90 was selected based on Vaha and Primmer (2006), who recommended using this value when the aim of the study is to correctly identify hybrids. Individuals with q values <0.90 in either cluster were taken to be admixed.

Finally, a similarity coefficient (Rosenberg et al. 2002) was calculated from the Structure output to evaluate the stability of individual assignments. We conducted 50 runs, each with 100,000 burn-in steps and 1,000,000 data collecting steps. Values of the similarity coefficient can range from zero to one, with values closer to one indicating that individuals are consistently placed into the same cluster.
The program NewHybrids (Anderson and Thompson 2002) was used to estimate each individual’s posterior probability of membership in six genotype frequency classes: parentals of each subspecies, first (F₁) and second (F₂) generation hybrids, and backcrosses to each subspecies. Classification followed the logic of Berthier et al. (2006), who classified individuals based on the q values in the pure parental categories. Thus, an individual was classified as a pure subspecies if q ≤ 0.90, F₁ if 0.50 < q < 0.90, and backcrossed if q < 0.50 for both parentals of each subspecies and backcrossed if 0.90 < q < 0.50 for one or the other subspecies. Test runs confirmed that results were not affected by the inclusion of species membership tags, so they were not included. Ten runs were conducted using Jeffreys-like priors, each with a burn-in period of 50,000 sweeps and a data collection period of at least 1,000,000 sweeps. Results were examined for uniformity, and final determinations of individual posterior probabilities of membership in each class were based on a representative run.

Classification of the subsampled specimens by the HotAce2 assay and microsatellite analysis was studied with a tabular comparison, as well as graphically, to assess the degree of agreement between assays among sites.

### Results

#### Genetic Diversity Within Sites

Specimens for this study were originally assayed with 11 microsatellite loci, but 3 loci, CQ46F3/R3, CxpGT9F2/R (Smith et al. 2005), and CxqGT108 (Edillo et al. 2007), were significantly out of HWE at almost all sites after Bonferroni correction (Rice 1989). This may have been because of errors during the amplification and visualization of fragments. Data for these three loci were removed from the data set, leaving a total of eight loci (Table 2), which were polymorphic at all collection sites, although one locus, CpqCTG10, had only 1 heterozygote and 49 homozygotes in the New Orleans, LA, population.

A total of 102 alleles were scored, and most alleles were found at several sites. There were several alleles (n = 9) that were more common in northern sites and absent at the southernmost site (New Orleans, LA) and also some (n = 5) that were found at the southern sites and absent at the northernmost site (Chicago, IL; Table 2). In addition, the number of alleles for a particular locus varied across sites, and allelic richness increased from south to north (Table 3). The average number of alleles per locus ranged from 3.00 ± 1.186 to 12.214 ± 3.068. Expected heterozygosity He was lowest on the southern and northern ends of the transect and higher toward the middle of the hybrid zone (mean = 0.627, range = 0.501–0.721). Sites generally conformed to HWE. Sites in the center of the hybrid zone tended to show a higher incidence of significant deviations from HWE (Bonferroni-adjusted α = 0.00045), but the patterns were not consistent over loci or sites (Table 3). Departures from HWE were positive, as quantified by FIS values, indicating an excess of homozygotes. Because nonequilibrium conditions were not confined to a particular locus or site, all loci were included in subsequent analyses. Several sites showed evidence of genetic disequilibrium, but these were not significant after Bonferroni correction (adjusted α = 0.00013). The loci were then treated as statistically independent.

#### Population Differentiation

The AMOVA results showed that most (77.88%) of the genetic variance in this system is within individuals (Table 4). The remaining variance was partitioned about equally within sites (10.3%) and among sites (11.8%). Permutation tests with the corresponding F-statistics (FRT = 0.221, FRS = 0.117, FST = 0.118) were all significant (P < 0.00001), indicating substantial genetic structure in this system. FST values per locus ranged from 0.043 for locus CpqCAG101 to 0.329 for locus CpqTri4 (Table 5).
3). Pairwise $F_{ST}$ values among sites ranged from 0.003 to 0.322, with most sites showing significant differentiation after correcting for multiple tests (Bonferroni adjusted $\alpha = 0.00055$; Table 5). The Mantel’s test to determine whether genetic differentiation increased with geographic distance (i.e., an IBD effect) was significant ($r = 0.733; P < 0.00001$). The results of the Population Graphs analysis show two subgraphs corresponding to the northern and southern sites, respectively (Fig. 2). The size of each circle corresponds to the within-site variation and shows the general pattern of greater allelic richness at northern sites, with the exception of Vaiden (Table 3). The southern cluster includes sites where *Cx. p. quinquefasciatus* specimens represent $\geq 10\%$ of specimens, whereas the northern cluster includes sites where $\geq 96\%$ of specimens are either *Cx. p. pipiens* or hybrids (Table 6). Memphis is unusual among sites in the northern cluster because nearly all specimens (49 of 50) were classified as hybrids, whereas the remaining northern sites contained both hybrids and *Cx. p. pipiens*.

**Admixture Analyses.** The optimal number of clusters defined by the greatest value of $\Delta K$ was associated with $K = 2$ (Fig. 3). This represents what Evanno et al. (2005) referred to as the “uppermost level of structure” in the data, and the information obtained by increasing the number of clusters drops sharply after grouping the data into two clusters. The analysis to quantify the stability of assignments of individuals to clusters showed that assignments were very stable across runs, and the similarity coefficient was high (0.978; $\alpha^2 = 0.020$). The similarity coefficient was lower when $K = 3$ (results not shown).

Cluster membership along the transect as determined by Structure showed a gradual change from *Cx. p. quinquefasciatus* in the south to *Cx. p. pipiens* in the north (Fig. 4). Using $q_i$ values to determine cluster membership, 170 (26%) individuals were classified as pure *Cx. p. quinquefasciatus*, 192 (30%) as pure *Cx. p. pipiens*, and 284 (44%) were admixed. For comparison, a representative run at $K = 3$ is also shown in Fig. 4 to provide a visual representation of where many hybrids were collected. The sites on the southern and northern ends of the transect had two (4%) and one (2%) admixed individual(s), respectively (Fig. 5). Aside from the two sites at each end of the transect, sites had at least 22% admixed individuals. One anomalous individual at the Effingham site was classified as pure *Cx.
Although all other *C. p. quinquefasciatus* individuals were confined to sites south of Sikeston, MO, NewHybrids yielded similar results with respect to the proportion of pure individuals, with 169 (26%) classified as pure *C. p. quinquefasciatus* and 187 (29%) as pure *C. p. pipiens*. NewHybrids found no F1 individuals. The remainder of individuals (n = 290) were classified as admixed, with 186 (29% of the total) classified as F2 hybrids, 49 (8%) as backcrossed to pure *C. p. quinquefasciatus*, and 55 (9%) as backcrossed to pure *C. p. pipiens*. When considered by site, similar trends were observed between the NewHybrids and Structure analyses, with the “admixed” proportion from Structure being allocated into different hybrid classes in NewHybrids (Fig. 6).

**Comparison of Methods to Classify Hybrids.** As expected, more admixed individuals were detected with microsatellites (44% of the microsatellite subsample) than with the HotAce.2 assay (30% overall; Table 6). In addition, the microsatellite analysis detected more hybrids toward the ends of the transect than were detected by the HotAce.2 assay (Table 6). In the microsatellite analysis, >10% of specimens were scored as hybrids at all sites except the two southernmost sites located in LA, and the northernmost site, Chicago. In contrast, the HotAce.2 assay scored >10% of specimens as hybrids only at the central six sites. In a direct graphic comparison of 646 specimens analyzed by both assays (Fig. 7), there was agreement between the two assays in the proportion of specimens classified as *C. p. pipiens*, *C. p. quinquefasciatus*, and hybrids, except for two sites: Batesville, MS, and Sikeston, MO.

**Discussion**

Levels of genetic diversity and differentiation, as well as results from the cluster analyses, are consistent with two distinct genetic entities contributing to a hybrid zone. For instance, the Structure analysis unambiguously supported the presence of two clusters and was not supportive of the hypothesis that hybrids constituted a distinct third entity. In addition, allelic richness was noticeably different between the ends of the transect and lowest at the southern end of the transect. Both Chicago and New Orleans are port cities, and because New Orleans maintains year-round activity, we expected greater allelic richness at that site. However, lower allelic richness is consistent with a previous study of *C. p. quinquefasciatus*, which attributed lower allelic richness in U.S. *C. p. quinquefasciatus* populations to their recent introduction, perhaps from East Africa or Asia (Fonseca et al. 2006).

Departures from HWE were most often found within the hybrid zone and less frequently toward the ends of the transect. Measures of genetic differentiation provide more support for the idea that two genetically distinct
subspecies are present on either end of the transect. Considerable genetic structuring exists among sites, as indicated by the results of the AMOVA and Population Graphs analyses. The latter analysis suggests the presence of two groups of sites: a southern group of six sites that is connected to a northern group of eight sites by Memphis, the site with the highest proportion of hybrids (Fig. 2; Table 6). Adjacent sites are usually connected, which is consistent with pairwise $F_{ST}$ values located on or close to the diagonal (Table 5) that were often not statistically significantly different. Wright (1978) suggested guidelines to qualify the extent of genetic differentiation in a system. He noted that $F_{ST}$ values <0.05 indicated little genetic differentiation, whereas values >0.05 indicated successively greater amounts of differentiation. The geographic distance associated with an $F_{ST}$ value of 0.05 is $\approx 300$ km (Fig. 1). Therefore, individuals sampled at distances of 300 km or greater would be expected to be part of genetically discernable groups. Despite this seemingly large distance, a significant positive relationship was observed between genetic differentiation and geographic distance along the transect (Fig. 1).

The threshold used to identify hybrids may have underestimated the proportion of hybrids and should

Fig. 4. Individual membership coefficients for Cx. pipiens complex mosquitoes ($q_i$ values) for $K = 2$ (top panel) and $K = 3$ (bottom panel), using the model-based clustering program Structure. Red, Cx. p. quinquefasciatus; yellow, Cx. p. pipiens; blue, hybrids. Population labels are below each panel and arranged from south to north. Admixed individuals were defined as having $q_i$ values <0.90 in both clusters when $K = 2$.

Fig. 5. Histograms created using output from the model-based clustering program Structure representing the proportion of individuals at each site classified as Cluster 1 (black, Cx. p. quinquefasciatus); Cluster 2 (white, Cx. p. pipiens), or admixed (gray; i.e., with a $q_i$ value <0.90).
be considered a somewhat conservative estimate (Vaha and Primmer 2006). Overall, there was a high percentage of hybrids detected, >40%, and the two programs used to classify hybrids, Structure and NewHybrids, yielded similar results. The genetic composition of sites along the transect shows a wide hybrid zone capped on either end by sites with high numbers of pure subspecies.

Interestingly, no F1 individuals were detected with the NewHybrids analysis. Other studies have reported this phenomenon in other taxa (Barilani et al. 2005, 2007; Berthier et al. 2006; Wallace 2006), and it is
consistent with the findings of this study given the large width of the hybrid zone. The chance of mating with a pure individual within the zone is less than the chance of mating with another hybrid. How the high degree of admixture observed in this study affects biological attributes such as pesticide resistance or vector competence is unknown, although preliminary data suggest hybrids are more efficient laboratory vectors of WNV than either parental subspecies (H.S., unpublished data).

The hybrid zone seems wider and extends further to the south than previously thought, and hybrids comprised >20% of individuals at 10 of the 14 sites. Of those 10 sites, 5 had hybrid frequencies >50%. Given the results of this study, we propose that the hybrid zone encompass an area bound by those sites comprised of >20% hybrids, which is inclusive of the sites from Brookhaven, MS to Champaign, IL. The proposed boundaries extend further north and significantly further south with respect to earlier work based on variation in
the DV/D ratio of male genitalia (Barr 1957) (Fig. 8). Also shown for comparison in Fig. 8 are the boundaries of the hybrid zone defined by the single gene HotAce.2 assay, with sites composed of >10% hybrids placed inside the hybrid zone (Table 6).

It bears repeating that pure Cx. pipiens complex subspecies seem to persist at either end of the transect, suggesting the presence of a mechanism that prevents the two taxa from being completely introgressed. Although our study was not designed to address the mechanism that maintains the parental subspecies and the hybrid zone, we believe that environmental regulation of diapause and the selective advantage conferred by diapause in northern latitudes remains the most viable explanation for the observed distributions.

Diapause is induced by environmental cues experienced by the pupal stage: primarily shortened photoperiods and secondarily reduced water temperatures during the late summer and early fall in northern latitudes (Eldridge 1968, Sanburg and Larsen 1973, Spielman and Wong 1973). In North America, Cx. p. pipiens’s southern distribution is believed to be limited to latitudes where the spring photoperiod is long enough to break diapause (Spielman 2001). In contrast, Cx. p. quinquefasciatus does not enter winter diapause and the inability of Cx. p. quinquefasciatus and hybrid specimens with a significant Cx. p. quinquefasciatus ancestry to diapause when exposed to short day lengths preceding harsh winter conditions is believed to limit the northern distribution of this subspecies (Eldridge 1968, Wilton and Smith 1985). Laboratory crosses between pure strains show that the nondiapause state is dominant, with only 7% of F1 progeny displaying diapause (Wilton and Smith 1985). Indeed, various backcrosses show that ≈75% of the genome must be Cx. p. pipiens in origin before a significant number of specimens (43%) express diapause (Wilton and Smith 1985). Thus, there is likely a fitness cost associated with diapause that helps maintain pure subspecies at either end of the transect.

Such a phenomenon has been shown in studies conducted with Drosophila melanogaster, which have suggested that fitness trade-offs exist for forms that diapause versus those that do not (Harshman and Hoffmann 2000, Schmidt et al. 2005). Forms that do not diapause mature and reproduce earlier, have shorter life spans, and are less able to respond to environmental stress compared with individuals that diapause (Tatar et al. 2001). Members of the Cx. pipiens complex may experience similar trade-offs, which could limit the area of introgression between subspecies.

Recently, a study on quantitative trait loci (QTL) associated with laboratory crosses of Cx. p. pipiens and Cx. p. quinquefasciatus documented the presence of heterosis in hybrids, where some QTL heterozygote phenotypes for diapause-related life history traits were outside the range of either parent (Mori et al. 2007). The authors suggest that heterosis across the hybrid zone could allow populations to adapt to local environmental conditions. Thus, heterosis in middle latitudes in conjunction with strong selection for diapause in northern areas and selection against diapause in southern areas could allow for the maintenance of the subspecies identities in northern and southern areas and a broad, stable hybrid zone.

Differences in the proportion of hybrids detected between the microsatellite versus the HotAce.2 assay were most notable at the sites toward the edges of the hybrid zone (Table 6). Because the HotAce.2 marker is inherited in a simple Mendelian fashion, all F1 specimens, but only approximately one half of F2 hybrids, would show a hybrid genotype for this marker (Aspen and Savage 2003). Crosses among hybrids would alter these ratios, but in general, we expect the HotAce.2 assay to under-represent the true frequency of advanced-generation hybrids. Nevertheless, the results of the two assays generally agree (Fig. 7), and given the ease with which one can perform the HotAce.2 assay, its utility as a subspecies-diagnostic marker remains relevant.

Acknowledgments

We thank the CDC field and laboratory team for collection of specimens, morphological identification, and initial processing of specimens. Team members include L. Colton, E. Edwards, J. Gold, E. Gordon, H. Savage, and G. Sutherland. We thank B. Biggerstaff, CDC, for statistical consultation and for production of Fig. 7 and B. Eisen, CDC, for production of Fig. 8.

References Cited


March 2009  KOTHERA ET AL.: GENETICS OF Cx. p. pipiens COMPLEX ALONG A TRANSECT  247


Received 7 October 2008; accepted 21 December 2008.