Composite Linkage Map and Enhanced Genome Map for Culex pipiens Complex Mosquitoes

PAUL V. HICKNER, AKIO MORI, DAVE D. CHADEE, AND DAVID W. SEVERSON

From the Eck Institute for Global Health and Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556 (Hickner; Mori, and Severson); and the Department of Life Sciences, University of the West Indies, St. Augustine, Trinidad; West Indies (Chadee).

Address correspondence to David W. Severson at the address above, or e-mail: severson.1@nd.edu.

Data deposited at Dryad: http://dx.doi.org/10.5061/dryad.rd465

Abstract

We report here the development of 65 novel microsatellite loci and construction of a composite genetic linkage map for Culex pipiens complex mosquitoes. Microsatellites were identified by in silico screening of the Culex quinquefasciatus genome assembly. Cross-species utility of 73 microsatellites for population studies in C. pipiens sensu stricto and C. quinquefasciatus was evaluated by genotyping a subset of samples collected in Indiana, United States, and Point Fortin, Trinidad. Allele frequencies of 67 microsatellites were within Hardy–Weinberg expectations in both population subsets. A composite linkage map was constructed based on restriction fragment length polymorphism and microsatellite polymorphisms in 12 independent F1 intercross mapping populations. The composite map consists of 61 marker loci totaling 183.9 cM distributed across the 3 linkage groups. These loci cover 29.5, 88.8, and 65.6 cM on chromosomes I–III, respectively, and allow for assignment of 10.4% of the genome assembly and 13.5% of the protein coding genes to chromosome position. Our results suggest that these microsatellites will be useful for mapping and population studies of 2 pervasive species in the C. pipiens complex. Moreover, the composite map presented here will serve as a basis for the construction of high-resolution genetic and physical maps, as well as detection of quantitative trait loci to aid in the investigation of complex genetic traits influencing phenotypes of interest.

Key words: composite linkage map, Culex quinquefasciatus, genome assembly, house mosquito, SSR, supercontigs

The Culex pipiens complex of mosquitoes includes some of the most geographically widespread and medically important vectors of human disease. They are key vectors of West Nile and St. Louis encephalitis viruses, as well as Wuchereria bancrofti, the parasite responsible for most periodic lymphatic filariasis (Day 2001; Turell et al. 2001; Fonseca et al. 2004; Michalski et al. 2010). Members of this species complex can be found on every continent except Antarctica (Vinogradova 2000), making them among the most pervasive mosquitoes known. Culex pipiens sensu stricto (L.), the northern house mosquito, inhabits temperate regions where females are able to survive the adverse conditions associated with winter by entering a programmed developmental arrest (i.e., reproductive diapause) (Vinogradova 2000). In contrast, Culex quinquefasciatus (Say), the southern house mosquito, inhabits most tropical and subtropical regions of the world and is genetically unable to enter diapause. Most members of this species complex live above ground where females must take a blood meal, usually from birds, prior to vitellogenesis and oviposition (anautogenous). However, the “molestus” form of C. pipiens lives predominantly underground and does not require a blood meal to lay an initial batch of eggs (autogenous). When they do blood feed, they tend to feed more readily on mammals (Byrne and Nichols 1999). These physiological and behavioral traits exemplify the ability of the C. pipiens complex to adapt to diverse environments. A greater understanding of the molecular underpinnings of these and other complex traits associated with their vectorial capacity could provide additional tools to help prevent the transmission of pathogens by these important disease vectors.

Mori et al. (1999) were the first to report a detailed genetic linkage map of C. pipiens. They utilized restriction fragment length polymorphism (RFLP) markers using random complementary DNAs (cDNAs) from the yellow fever mosquito Aedes aegypti to construct a genetic linkage map. Of the 22 cDNA clones used, 21 were mapped to scaffolds (supercontigs) in the C. quinquefasciatus genome assembly (Arensburger et al. 2010). This not only provided genetic
Materials and Methods

Microsatellite Identification and Characterization

Microsatellites were identified as previously described in Hickner et al. (2010). Briefly, loci were identified by screening the \textit{C. quinquefasciatus} genome assembly (CpipJ1) for di- and trinucleotide repeats using Tandem Repeats Finder (Benson 1999) or by pasting a supercontig nucleotide sequence into a Word document and using the “find” function to search for specific motifs. To help eliminate multiple-copy genome sequences, a 400–600 nucleotide region containing a target microsatellite sequence was subjected to BLASTn analysis against the CpipJ1 assembly. Primers were designed to produce polymerase chain reaction (PCR) amplicons ranging from 110 to 350 bp using Primer3 (Rozen and Skaletsky 2000), Primer3plus (Untergasser et al. 2007), or OligoCalc (Kibbe 2007). To detect potential non-target amplification and to determine if a given microsatellite was within a protein coding region, primer sequences were subjected to BLASTn analysis against the CpipJ1 assembly and the CpipJ1.2 gene build at VectorBase (Lawson et al. 2009).

Most of the supercontigs screened for microsatellites were chosen because their chromosome position had been previously determined based on RFLP mapping in \textit{C. pipiens} or \textit{A. aegypti} (Mori et al. 1999, 2007; Severson et al. 2002). However, we also chose several supercontigs based on their size and their having genes of interest. Supercontigs 3, 2, and 3.4 were screened because they are among the largest genomic scaffolds in the assembly. Supercontigs 3.119, 3.163, and 3.1500 were chosen in association with ongoing independent studies because they contained the genes \textit{shaggy} (CPIJ006614), \textit{period} (CPIJ007193), and \textit{cryptochrome 2} (CPIJ018859), respectively.

Primers were then tested in PCRs using total genomic DNA from 1 individual from each of 7 laboratory colonies: Boane, Johannesburg, and Trinidad (\textit{C. quinquefasciatus}); Shasta and South Bend (\textit{C. pipiens}); Gose (\textit{Culex pipiens pallens}); and Shinkura (\textit{C. pipiens form molestus}) (see Supplementary Table S1 online). DNA was extracted using a simple alkaline extraction method adapted from Rudbeck and Dissing (1998). Mosquitoes were homogenized in 80 µL 0.2M NaOH and incubated at 75°C for 10 min. Thereafter, 28.8 µL 1M Tris, pH 8.0 and 91.2 µL sterile double-distilled water (ddH$_2$O) were added to the mixture bringing the volume to 200 µL. Sterile ddH$_2$O was added to produce a final volume of 800 µL for adult females and 500 µL for adult males. One microliter of the DNA preparation was used in each PCR. PCR amplification was performed in 25 µL reactions in 96-well PCR plates (Dot Scientific Inc. Burton, MI, USA). Each reaction contained 1 × Taq buffer (50 mM KCl, 10 mM Tris pH 9.0, 0.1% Triton X-1), 1.5 mM MgCl$_2$, 200 mM dNTPs, 5 µmol of each primer, 1 unit of Taq DNA polymerase, and 1 µL of genomic DNA preparation (~20 ng). Thermal cycling was performed using Mastercycler® thermocyclers (Eppendorf AG, Hamburg, Germany) under the following conditions: initial denaturation for 5 min at 94°C followed by 30 cycles of denaturation for 1 min at 94°C, annealing for 1 min at 60°C, extension for 2 min at 72°C, and a final extension at 72°C for 10 min. Fluorophore-labeled (6-FAM®, HEX®, NED®) forward primers were prepared for microsatellites that amplified well in individuals from all 7 colonies based on UV visualization on ethidium bromide–stained 2% agarose gels.

We tested individual microsatellite loci for Mendelian inheritance in one or more families, when possible. Alleles were scored using the ABI PRISM 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and ROX 400HD size standard. GeneMapper v4.0 software (Applied Biosystems, Foster City, CA, USA) was used to call the alleles, with subsequent visual verification of each sample. Different primers flanking the microsatellite were designed when null alleles were evident in a family. Each microsatellite was named based on the supercontig assembly where it was found, the repeat motif, and the primer version (C134AC1b: C$_{134}$AC$_1$). An exact test for HWE was performed following Guo and Thompson (1992) with 100 000 dememorization steps followed by 1 000 000 iterations using the program Arlequin (Excoffier et al. 2005). The program CONVert was used to format the input files for Arlequin (Glaubitz 2004).

Genetic Crosses

The laboratory strains used for generation of the mapping populations were isolated from field collections of \textit{C. quinquefasciatus} (Boane, Johannesburg, Trinidad, and Vero
Beach), *C. pipiens* s.s. (Shasta and South Bend), *C. pipiens pallens* (Gose), and *C. pipiens* form molestus (Chicago and Shinkura) (see Supplementary Table S1 online). Pairwise matings of the F₀ and F₁ generations were carried out using the following method: pupae were separated and placed in individual 500 mL cups containing ~250 mL of water until emergence, at which time ~5 females and 1 male were placed in a cheesecloth cage (30 × 20 × 20 cm) and allowed to mate. Approximately 7 days later the females were blood fed on rats, and each female was placed in a vial (22 × 46 mm) containing 1–2 cm of water until the eggs were laid. The number of F₁ females used to produce each mapping population varied from 1 to 3. The F₀ strains and the number of markers used for each mapping family are listed in Table 1.

**Linkage Mapping**

We constructed individual linkage maps and combined them using the Kosambi function (Kosambi 1943) in JoinMap 3.0 (Van Ooijen and Voorrips 2001). Microsatellites were anchored to RFLP or other microsatellite markers when they shared the same supercontig and they mapped to a common locus. The fixed order function in JoinMap 3.0 was used to maintain the order of the markers established in the individual maps. Because the mapping families were generated to investigate the genetic basis of female traits (diapause, autogeny, blood feeding), only F₂ females were genotyped and used for calculation of the composite map. Chromosomes were designated following Mori et al. (1999).

**Results**

**Microsatellite Identification and Characterization**

Sixty-five novel microsatellites were identified and characterized in this present study (see Supplementary Table S2 online). Mendelian inheritance of 5 microsatellites (C32TC1b, C32TGC1, C66TGT1, C111CAG1, and C550TGC1) was not tested in our laboratory colonies due to the absence of informative polymorphisms in the families. BLASTn analysis against the *CpipJ1.2* gene build revealed that 14 microsatellites (21.5%) map to protein coding genes, with 13 of those being within exons (Table 2). Two microsatellites (C660CTC1 and C660GTG1) were within different exons of a single gene (CPIJ014662). Analysis of the proteins comprising these microsatellites revealed that 10 of the 14 (71%) repeats represented runs of the amino acid glutamine. Nine of the 11 codons are CAG, with only 2 being the alternative codon CAA. One microsatellite (C111CAG1) is in the 5′ untranslated region of the protein CPIJ005853.

A total of 73 microsatellites were characterized in population subsets from IN and PF (see Supplementary Table S3 online). These included 65 microsatellites described here and 8 microsatellites described previously (Hickner et al. 2010). Of these, 2 were monomorphic in IN and 5 were monomorphic in PF. None were monomorphic in both populations. Six microsatellites had allele frequencies that deviated (P < 0.05) from Hardy–Weinberg expectations—or amplified poorly—in one of the population subsets tested. Five of these (C32TC1b, C38AC1, C66GT1, C309TGC1, and C446TC1) deviated in the IN population, whereas only 1 (C205GAC1) deviated in the PF population (see Supplementary Table S3 online). Each of these had heterozygote deficiencies suggesting the presence of null alleles at these loci. The total number of alleles identified for all markers was 364 in the IN population and 263 in the PF population. The number of alleles per locus ranged from 1 to 13 (mean = 5.1) in IN and from 1 to 11 (mean = 3.6) in PF.

**Linkage Mapping and Genome Assignment**

We were able to map 69 microsatellites and 39 cDNA clones to 63 genetic loci distributed across 3 linkage groups with
a combined length of 183.9 cM (Figure 1). The lengths of chromosomes I–III are 29.5, 88.8, and 65.6 cM, respectively. We assigned the map locations of 8 microsatellites (C32TC1b, C32TGC1, C66TGT1, C550TGC1, CxpGT4, Cxpq51, CxqCAG101, and CxqTri4) based on the position of other markers within the same supercontig assembly. The microsatellites CxpGT4 (Keyghobadi et al. 2004), Cxpq51 (Abreu et al. 2012), CxqCAG101 (Edillo et al. 2007), and CxqTri4 (Smith et al. 2005) were described previously and found within supercontigs mapped in our study. Because we did not genotype males in the F2 generations, the sex-determining locus was placed on the composite map based on its association with marker LF284 (Mori et al. 1999).

Linkage mapping suggested the presence of misassemblies in 9 supercontigs (Table 3). Misassemblies were evident when different markers from 1 supercontig mapped to disparate positions on the linkage map. Markers in 7 supercontigs (3.5, 3.28, 3.48, 3.65, 3.99, 3.134, and 3.139) mapped to locations on different chromosomes, whereas markers in 3.177 and 3.309 mapped to different positions on the same chromosome. BLASTn analysis of the cDNAs and microsatellites against the contigs suggested that the errors were made while assembling the contigs into supercontigs rather than assembling the reads into contigs. The map reported here assigns chromosome position to 10.4% (56.4 of 540.0 Mb) of the C. quinquefasciatus genome assembly and 13.5% (2541 of 18 883) of the protein coding genes (Lawson et al. 2009; Arensburger et al. 2010).

Table 2  Microsatellites within protein coding regions

<table>
<thead>
<tr>
<th>Microsatellite</th>
<th>Repeat motif</th>
<th>Amino acid</th>
<th>Gene ID</th>
<th>Gene name/description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C36GT1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ00765</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>C38GCT1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ003062</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>C65CAG1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ004312</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>C66TGT1</td>
<td>AA</td>
<td>Asparagine</td>
<td>CPIJ004404</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>C95GCA1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ005607</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>C108CAG1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ005224</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>C143TGC1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ007159</td>
<td>forkhead box protein</td>
</tr>
<tr>
<td>C186TGT1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ007969</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>C309TGC1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ10840</td>
<td>mitochondrial inner membrane translocase</td>
</tr>
<tr>
<td>C550TGC1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ13972</td>
<td>trithorax</td>
</tr>
<tr>
<td>C660CT1</td>
<td>GAG</td>
<td>Glutamine</td>
<td>CPIJ014662</td>
<td>ubiquitin specific protease</td>
</tr>
<tr>
<td>C660GCT1</td>
<td>CA</td>
<td>Histidine</td>
<td>CPIJ014662</td>
<td>ubiquitin specific protease</td>
</tr>
<tr>
<td>C1500CAG1</td>
<td>CA</td>
<td>Glutamine</td>
<td>CPIJ018859</td>
<td>cryptochrome 2</td>
</tr>
</tbody>
</table>

Despite the increased marker density in the current map, substantial gaps remain that preclude high-resolution genome analysis. Filling these gaps using the strategy employed here is limited by the fragmented state of the current genome assembly. A combination of high-resolution linkage mapping and physical mapping is necessary to substantially enhance the current assembly. Genotyping large mapping families or advanced intercross lines using methods such as restriction-site associated DNA tag sequencing (Baird et al. 2008) could potentially assign a large number of supercontigs to their respective chromosome positions. Physical mapping of A. aegypti metaphase chromosomes via in situ hybridization of BAC clones carrying genetic markers facilitated the construction of an integrated linkage, chromosome, and genome map (Timoshevskiy et al. 2013). A BAC library has been constructed that would make this a feasible approach for improving the C. quinquefasciatus genome assembly (Hickner et al. 2011).

As expected, most of the microsatellites are not within protein coding regions of the genome. However, 26% of the microsatellites with trinucleotide repeats were mapped to 9 supercontigs mapped to a 5.4 cM region on chromosome I that includes the sex-determining locus, confirming earlier results suggesting there is considerable reduction in recombination on chromosome I compared with chromosomes II and III, and chromosome I in A. aegypti (Mori et al. 1999). The length of chromosome II increased slightly from 85.9 to 88.8. This was due to the addition of 2 loci (C177CGT1 and C186TGT1) beyond CX111/C108CAG1 on the current map. The length of chromosome III decreased from 79.2 to 65.6 cM despite an increase in the number of loci from 15 to 23. This was due to differences largely in the 2-point recombination estimates between the same markers among the different mapping populations. The linear order of the markers reported in Arensburger et al. (2010) was maintained in the current map with the exception of LF99a/C208GCA1, which was initially placed above CX111/C446TC1 in the composite map. The order reported here (CX11/C446TC1—LF99a/C208GCA1—LF128/C185CAG1) represents the only data for these markers where the 3-point estimates were derived from a single mapping family.

Discussion

The number of loci on the C. pipiens linkage map has been increased from 38 to 63, therefore increasing marker density from one marker every 5.1 cM to one marker every 3.1 cM (Arensburger et al. 2010). The combined length increased marginally from 177.6 to 183.9 cM. The length of chromosome I increased from 12.5 to 29.5 cM. This was a modest gain considering that we added microsatellites from 8 supercontigs, and only 3 supercontigs (3.36, 3.198, 3.12) spanned beyond the markers previously mapped. Microsatellites on
exons, with most (11 of 14) representing repeats of glutamine. Katti et al. (2001) compared the frequencies and types of amino acid repeats in Drosophila melanogaster, Caenorhabditis elegans, and Saccharomyces cerevisiae and found that glutamine repeats make up the greatest proportion of amino acid repeats (≥7) in all 3 organisms. Moreover, D. melanogaster has a comparatively high number and proportion, with 1555 glutamine repeats out of a total of 2993 (Katti et al. 2001). Hyperexpansions of glutamine have been linked to some human diseases called polyglutamine diseases. These are predominantly neurodegenerative diseases and include Huntington’s disease and some spinocerebellar ataxias (Zoghbi and Orr 2000).

This present study provides the first composite linkage map for C. pipiens complex mosquitoes, which includes microsatellites and their corresponding genome scaffolds. This effort represents an ~41% increase in genome assignment to chromosome positions, for example, from 5.7% to 10.5% of the existing genome assembly (Arensburger et al. 2010). It also brings the total number of microsatellites developed for this species complex to over 100 (Fonseca et al. 2006).
1998; Keyghobadi et al. 2004; Edillo et al. 2007; Hickner et al. 2010; Smith et al. 2005; Abreu et al. 2012). In summary, we utilized data from 12 F\textsubscript{1} intercross mapping families ranging in size from 94 to 192 to construct a composite linkage map for \textit{C. pipiens} complex mosquitoes. The integration of microsatellites—a practical alternative to RFLP—should help facilitate additional mapping and population studies, as well as provide landmarks for enhancing the genome assembly in this important disease vector.

**Supplementary Material**


**Funding**

National Institutes of Health, National Institute of Allergy and Infectious Diseases, grant no. RO1-AI079125-A1.

**Acknowledgments**

We thank the field crew in Trinidad: Stephen “Billy” Deonarine, Doolam Lalla and Dester Maraj for their assistance in sample collections from Point Fortin, Trinidad, and Jim Erwin and Shawn Moore for their assistance in sample collections in Indianapolis, IN, USA. We also thank Ned Walker for providing us with the Chicago strain and Anton Cornel for providing us with the Johannesburg, Shasta and Boane strains.

**References**


Received February 23, 2013; First decision April 1, 2013; Accepted June 4, 2013

Corresponding Editor: Adalgisa Caccone