

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

landmarks, it also helped illustrate the congruence among chromosome arms between the *C. pipiens* and *A. aegypti* genomes. Consequently, we were able to utilize data from previous mapping studies in *C. pipiens* (Mori et al. 1999, 2007) and a composite linkage map for *A. aegypti* (Severson et al. 2002) to develop microsatellites broadly distributed across all 3 chromosomes in *C. pipiens sensu lato*. We used these microsatellites to analyze 6 independent F_1 intercross mapping populations, and we then used the JoinMap program to integrate these data with 6 independent F_1 intercross mapping populations analyzed using RFLP markers to construct a composite linkage map for *C. pipiens* complex mosquitoes. This also allowed for assignment of associated genome assembly supercontigs to chromosome positions.

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 *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), Primer3 plus (Untergasser et al. 2007), or OligoCalc (Kibbe 2007). To detect potential nontarget 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 *C. pipiens* or *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 3.1, 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 *shaggy* (CPIJ006114), *period* (CPIJ007193), and *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 (*C. quinquefasciatus*); Shasta and South Bend (*C. pipiens*); Gose (*Culex pipiens pallens*); and Shinkura (*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 μ L 0.2 M NaOH and incubated at 75°C for 10 min. Thereafter, 28.8 μ L 1 M Tris, pH

8.0 and 91.2 μ L sterile double-distilled water (ddH₂O) were added to the mixture bringing the volume to 200 μ L. Sterile ddH₂O was added to produce a final volume of 800 μ L for adult females and 500 μ L for adult males. One microliter of the DNA preparation was used in each PCR. PCR amplification was performed in 25 μ L reactions in 96-well PCR plates (Dot Scientific Inc. Burton, MI, USA). Each reaction contained 1 $\times$ Taq buffer (50 mM KCl, 10 mM Tris pH 9.0, 0.1% Triton X), 1.5 mM MgCl₂, 200 mM dNTPs, 5 pmoles of each primer, 1 unit of Taq DNA polymerase, and 1 μ L 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 v.4.0 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 = *C. quinquefasciatus* genome, 134 = supercontig 3.134, AC = AC repeat, 1 = first microsatellite developed with this motif from this supercontig, b = alternate primer sequence of a previously published microsatellite). To ensure that the microsatellites were truly novel, BLASTn analysis of previously published *C. pipiens* and *C. quinquefasciatus* microsatellite sequences was performed.

Cross-species utility of these microsatellites for population studies was evaluated by genotyping ~30 individual *C. pipiens* s.s. collected from Indianapolis, Indiana, USA, and *C. quinquefasciatus* collected from Point Fortin, Trinidad. The Indianapolis (IN) samples were a subset of samples collected in 2008 and previously determined to be in Hardy–Weinberg equilibrium (HWE) based on allele frequencies of 12 microsatellite loci (Hickner et al. 2010). The Point Fortin (PF) samples were a subset from samples collected during 2010 using the same method described for the IN samples. This population had not been previously characterized. 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 *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 F_0 and F_1 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 F_1 females used to produce each mapping population varied from 1 to 3. The F_0 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 F_2 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

Table 1 Summary of F_1 intercross mapping populations used for construction of the composite linkage map

F_0 laboratory strains		No. of F_2 scored	No. of markers	Marker type	Citation
♀	♂				
South Bend ^a	Johannesburg ^b	94	34	Microsatellite	—
Shinkura ^c	Boane ^b	98	20	Microsatellite	—
Chicago ^c	Boane	95	21	Microsatellite	—
Gose ^d	Johannesburg	95	20	Microsatellite	—
Trinidad ^b	Shasta ^a	95	29	Microsatellite	—
Shinkura	Chicago	95	28	Microsatellite	—
Shinkura	Vero Beach ^b	92	24	RFLP	—
Nile Delta ^c	Iowa ^d	191	13	RFLP	Mori et al. (1999)
Akoh ^c	Iowa	123	6	RFLP	Mori et al. (1999)
Akoh	Iowa	192	10	RFLP	Mori et al. (1999)
South Bend	Vero Beach	96	8	RFLP	Mori et al. (2007)
Gose	Vero Beach	192	9	RFLP	Mori et al. (2007)

See [Supplementary Table 1](#) online for additional information on the parental (F_0) strains.

^a *Culex pipiens*.

^b *Culex quinquefasciatus*.

^c *Culex pipiens* form molestus.

^d *Culex pipiens pallens*.

Table 2 Microsatellites within protein coding regions

Microsatellite	Repeat motif	Amino acid	Gene ID	Gene name/description
C36GT1	CAA	Glutamine	CPIJ002765	conserved hypothetical protein
C38GCT1	CAG	Glutamine	CPIJ003062	<i>cornetto</i>
C65CAG1	CAG	Glutamine	CPIJ004312	conserved hypothetical protein
C66TGT1	AAC	Asparagine	CPIJ004404	conserved hypothetical protein
C95GCA1	CAG	Glutamine	CPIJ005607	conserved hypothetical protein
C108CAG1	CAG	Glutamine	CPIJ005224	conserved hypothetical protein
C143TGC1	CAG	Glutamine	CPIJ007159	<i>forkhead box protein</i>
C186TGT1	CAA	Glutamine	CPIJ007969	conserved hypothetical protein
C309TGC1	CAG	Glutamine	CPIJ010840	<i>mitochondrial inner membrane translocase</i>
C550TGC1	CAG	Glutamine	CPIJ013972	<i>trithorax</i>
C660CTC1	GAG	Glutamic acid	CPIJ014662	<i>ubiquitin specific protease</i>
C660GTG1	CAC	Histidine	CPIJ014662	<i>ubiquitin specific protease</i>
C1500CAG1	CAG	Glutamine	CPIJ018859	<i>cryptochrome 2</i>

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).

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

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 25. 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 CX11/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.

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

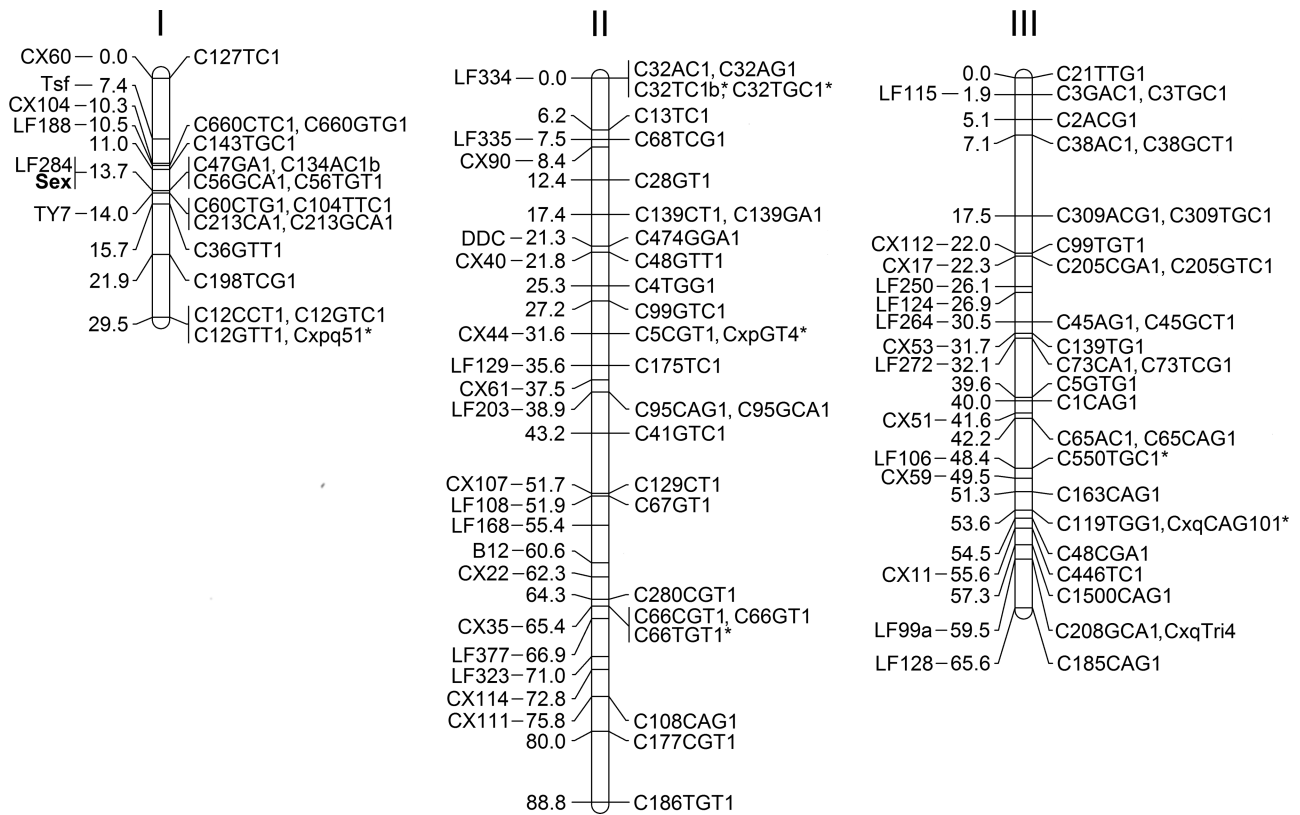


Figure 1. Composite RFLP and microsatellite linkage map for *Culex pipiens sensu lato* based on 12 independent F_1 intercross mapping populations. The cDNA clones used for RFLP mapping and the map distances (Kosambi centiMorgans, cM) are listed on the left side of the chromosomes—microsatellites are listed on the right. Markers with positions predicted based on mapping of other markers in the supercontig are indicated by an asterisk.

Table 3 Putative misassembled supercontigs identified by genetic mapping of multiple markers within a supercontig assembly

Supercontig	Marker	Map position	Marker	Map position
3.5	CX44, C5CGT1	2–31.6	C5GTG1	3–39.6
3.28	C28GT1	2–12.4	CX51	3–41.6
3.48	CX40, C48GTT1	2–21.8	C48CGA1	3–54.5
3.65	CX90	2–8.4	C65AC1, C65CAG1	3–42.2
3.99	C99GTC1	2–27.2	CX112, C99TGT1	3–22.0
3.134	C134AC1b	1–13.7	CX61	2–37.5
3.139	C139CT1, C139GA1	2–17.4	CX53, C139TG1	3–31.7
3.177	CX114	2–72.8	C177CGT1	2–80.0
3.309	LF250	3–26.1	C309ACG1, C309TCG1	3–17.5

Supercontigs with their corresponding markers and map positions are listed in each row.

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.

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_1 intercross mapping families ranging in size from 94 to 192 to construct a composite linkage map for *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

Supplementary material can be found at <http://www.jhered.oxfordjournals.org/>.

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 Dexter 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

- Abreu AA, Albaina A, Alpermann TJ, Apkenas VE, Bankhead-Dronnet S, Bergek S, Berumen ML, Cho C-H, Clobert J, Coulon A, et al. 2012. Permanent genetic resources added to molecular ecology resources database 1 October 2011–30 November 2011. *Mol Ecol Res*. 12:374–376.
- Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F, et al. 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science*. 330:86–88.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 3:e3376.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27:573–580.
- Byrne K, Nichols RA. 1999. *Culex pipiens* in London Underground tunnels: differentiation between surface and subterranean populations. *Heredity*. 82:7–15.
- Day JF. 2001. Predicting St. Louis encephalitis virus epidemics: lessons from recent, and not so recent, outbreaks. *Annu Rev Entomol*. 46:111–138.
- Edillo FE, Tript F, McAbee RD, Foppa IM, Lanzaro GC, Cornel AJ, Spielman A. 2007. A set of broadly applicable microsatellite markers for analyzing the structure of *Culex pipiens* (Diptera: Culicidae) populations. *J Med Entomol*. 44:145–149.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*. 1:47–50.
- Fonseca DM, Atkinson CT, Fleischer RC. 1998. Microsatellite primers for *Culex pipiens quinquefasciatus*, the vector of avian malaria in Hawaii. *Mol Ecol*. 7:1617–1619.
- Fonseca DM, Keyghobadi N, Malcolm CA, Mehmet C, Schaffner F, Mogi M, Fleischer RC, Wilkerson RC. 2004. Emerging vectors in the *Culex pipiens* complex. *Science*. 303:1535–1538.
- Glaubitz JC. 2004. CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol Ecol Notes*. 4:309–310.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 48:361–372.
- Hickner PV, Debruyne B, Lovin DD, Mori A, Behura SK, Pinger R, Severson DW. 2010. Genome-based microsatellite development in the *Culex pipiens* complex and comparative microsatellite frequency with *Aedes aegypti* and *Anopheles gambiae*. *PLoS One*. 5:e13062.
- Hickner PV, Debruyne B, Lovin DD, Mori A, Saski CA, Severson DW. 2011. Enhancing genome investigations in the mosquito *Culex quinquefasciatus* via BAC library construction and characterization. *BMC Res Notes*. 4:358.
- Katti MV, Ranjekar PK, Gupta VS. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol*. 18:1161–1167.
- Keyghobadi N, Matrone MA, Ebel GD, Kramer LD, Fonseca DM. 2004. Microsatellite loci from the northern house mosquito (*Culex pipiens*), a principal vector of West Nile virus in North America. *Mol Ecol Notes*. 4:20–22.
- Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res*. 35:W43–W46.
- Kosambi DD. 1943. The estimation of map distances from recombination values. *Ann Eugen*. 12:172–175.
- Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, et al. 2009. VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res*. 37:D583–D587.
- Michalski ML, Erickson SM, Bartholomay LC, Christensen BM. 2010. Midgut barrier imparts selective resistance to filarial worm infection in *Culex pipiens pipiens*. *PLoS Negl Trop Dis*. 4:e875.
- Mori A, Romero-Severson J, Severson DW. 2007. Genetic basis for reproductive diapause is correlated with life history traits within the *Culex pipiens* complex. *Insect Mol Biol*. 16:515–524.
- Mori A, Severson DW, Christensen BM. 1999. Comparative linkage maps for the mosquitoes (*Culex pipiens* and *Aedes aegypti*) based on common RFLP loci. *J Hered*. 90:160–164.
- Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. *Bioinformatics methods and protocols: methods in molecular biology*. Totowa (NJ): Humana. p. 365–386.
- Rudbeck L, Dissing J. 1998. Rapid, simple alkaline extraction of human genomic DNA from whole blood, buccal epithelial cells, semen and forensic stains for PCR. *Biotechniques*. 25:588–90, 592.
- Severson DW, Meece JK, Lovin DD, Saha G, Morlais I. 2002. Linkage map organization of expressed sequence tags and sequence tagged sites in the mosquito, *Aedes aegypti*. *Insect Mol Biol*. 11:371–378.
- Smith JL, Keyghobadi N, Matrone MA, Escher RL, Fonseca DM. 2005. Cross-species comparison of microsatellite loci in the *Culex pipiens* complex and beyond. *Mol Ecol Notes*. 5:697–700.
- Timoshevskiy VA, Severson DW, Debruyne BS, Black WC, Sharakhov IV, Sharakhova MV. 2013. An integrated linkage, chromosome, and genome map for the yellow fever mosquito *Aedes aegypti*. *PLoS Negl Trop Dis*. 7:e2052.
- Turell MJ, O’Guinn ML, Dohm DJ, Jones JW. 2001. Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus. *J Med Entomol*. 38:130–134.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res*. 35:W71–W74.

- Van Ooijen JW, Voorrips RE. 2001. JoinMap 3.0, Software for the calculation of genetic linkage maps. Wageningen, the Netherlands: Plant Research International.
- Vinogradova EB. 2000. *Culex pipiens pipiens* mosquitoes: taxonomy, distribution, ecology, physiology, genetics, applied importance and control. Moscow (Russia): Pensoft. p. 250.

Zoghbi HY, Orr HT. 2000. Glutamine repeats and neurodegeneration. *Annu Rev Neurosci.* 23:217–247.

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

Corresponding Editor: Adalgisa Caccone

Supplementary Table S1. Species and origins of the laboratory strains used in the study.

Strain	Species	Origin (year)	Source	Citation
Boane	<i>C. quinquefasciatus</i>	Boane, Mozambique	AJ Cornel	Cornel et al. 2003
Chicago	<i>C. pipiens</i> f. <i>molestus</i>	Chicago, IL, USA (2009)	ED Walker	Mutebi and Savage 2009
Akoh	<i>C. pipiens</i> f. <i>molestus</i>	Akoh, Japan	M Kamei	Mori et al. 1999
Gose	<i>C. pipiens pallens</i>	Gose, Japan	M Kamei	Mori et al. 2007
Iowa	<i>C. pipiens</i>	Ames, IA, USA	BM Christensen	Mori et al. 1999
Johannesburg	<i>C. quinquefasciatus</i>	Jonannesburg, South Africa	AJ Cornel	Arensburger et al. 2010
Nile Delta	<i>C. pipiens</i>	Egypt (1996)	H.Farid	Mori et al. 1999
Shasta	<i>C. pipiens</i>	Shasta, CA, USA	AJ Cornel	Cornel et al. 2003
Shinkura	<i>C. pipiens</i> f. <i>molestus</i>	Tokushima, Japan (1998)	A Mori	-
South Bend	<i>C. pipiens</i>	South Bend, IN, USA (2008)	A Mori	-
Trinidad	<i>C. quinquefasciatus</i>	Point Fortin, Trinidad (2010)	DD Chadee	-
Vero Beach	<i>C. quinquefasciatus</i>	Vero Beach, FL, USA (1997)	E Van Handel	Mori et al. 2007

Supplementary Table S2. Primer sequences and supercontig (SC) positions for microsatellites identified in the study. Position in SC is the first nucleotide on the 5' end of the forward primer.

Locus	Repeat motif	Predicted size (bp)	Forward primer/reverse primer (5'-3')	SC	SC size (bp)	Position in SC
C1CAG1	CAG ₆	264	GCATTCGACCCACACCTAGC TCGTGGCTCCCGTAGTAGGA	3.1	3,873,040	645,288
C2ACG1	ACG ₆	148	CGCACTTCAACGAACTCGAA ATGATCGTCGTTGGGACACA	3.2	2,744,360	1,646,608
C3GAC1	GAC ₇	299	GTAGTACGACACTTGGAGGAC CCGTGAAGGAGCTGGAGCA	3.3	2,758,190	2,527,292
C3TGC1	TGC ₆	247	GACGCGTACGGTGTAACGA AGCAGCACCAGGTAAGAAGTCA	3.3	2,758,190	1,813,235
C4TTG1	TTG ₆	136	CGAGCATAAACACCGGCTTC TCATCTTGCGGTGCTCACTT	3.4	2,511,003	1,045,226
C5CGT1	CGT ₇	202	GCGACCAACTCCACAGGA TATCGCAAGCAAGCAGTTCA	3.5	2,487,969	1,002,522
C5GTG1	GTG ₇	258	GTTCTGGATGTTCTGGCCGA GCTGATGTGAAGCAGGATACTC	3.5	2,487,969	482,818
C12CCT1	CCT ₅	180	ATGAACGGAACGGAATGGAG TTACGACCTTGGCACCTCTG	3.12	1,895,535	837,313
C12GTC1	GTC ₆	220	GGTCCTCCTGGTTCATGGAT CTTGGCTGTGTTGGTTCGAC	3.12	1,895,535	268,913
C12GTT1	GTT ₇	162	TCGGACTTCCATCATCATCAA CGGAGGACCATCGTCTCTTT	3.12	1,895,535	1,558,138
C13TC1	TC ₁₀	179	ACCCAGCAAGGAGTCAACCA CGGATGCTTACCGCATATCA	3.13	1,876,709	1,640,955
C21TTG1	TTG ₈	195	GCGTCTTATGTGCGCTAGGAA GAACGCGCTCCAATCAACTC	3.21	1,629,438	1,283,628
C28GT1	GT ₁₀	223	AGCGGAAACTGACCTATCCAA CGGCTGATGAGCTTTCGTGT	3.28	1,692,035	341,223
C32AG1	AG ₁₀	160	TCCAAAACGCGATTGTTACGA TGCACGATCCGAAGCTATCA	3.32	1,521,851	1,155,450
C32TC1b	TC ₁₄	198	CGCATGCAATCAAAGGAGGC ATCCAGCGAGTATGTGAGCAC	3.32	1,521,851	266,744
C32TGC1	TGC ₇	129	CGATTGTTGTTCCGACTGGT CTGGCCAGTAGAGGTGACAGAA	3.32	1,521,851	787,639
C36GTT1	GTT ₆	188	GAATCAGCCGGATCTTGGTC TGGGACTCGCAGATTCATCA	3.36	1,496,554	1,303,951
C38AC1	AC ₈	166	CGGTATGCGCTCATGATTTG GGTGTGGCTGGTTGACGAT	3.38	1,436,690	1,236,529
C38GCT1	GCT ₇	173	CATTCAGCACCTTCACGACA GCAGGATTCGTCTTTAGCC	3.38	1,436,690	581,278
C41GTC1	GTC ₇	167	TCCTCGTGAATGTGCAAGG TTCGTGCTGGAATCTCTTGG	3.41	1,406,373	626,448
C45AG1	AG ₈	172	GAACCTCGTGCGTACTGTCAA CGCACATCATCTGTCTGGATT	3.45	1,442,824	444,156
C45GCT1	GCT ₇	184	GCGTGTTTCAGACGCAATGGAA TCCCAGCCATTAGCCACTTTCA	3.45	1,442,824	1,118,880
C47GA1	GA ₉	236	GCTTGACCGAGTGTAACGAGA ATGTGCGTTGCCAACCATTA	3.47	1,403,514	1,313,771
C56GCA1	GCA ₇	207	GTTCTGCCCACCATAATTCACC CCAAATGTGCCCTTCCCTATG	3.56	1,226,699	973,937

C56TGT1	TGT ₈	231	ATCTCGCTGCTGGCTAAAGTC	3.56	1,226,699	918,098
C60CTG1	CTG ₆	284	AAGGTCAACCCGAGTGTGC	3.60	1,176,631	404,126
C65CAG1	CAG ₆	173	CGGCCACTAGCGCACATAC	3.65	1,116,611	594,671
C66CGT1	CGT ₇	138	GCTGGTTGGAAAGGTGATCG	3.66	1,106,043	495,995
C66GT1	GT ₁₂	181	GCTAGCTTGGTTGGTACTGGAA	3.66	1,106,043	427,138
C66TGT1	TGT ₇	320	CGACTCGTTGGCAGGTGAA	3.66	1,106,043	483,175
C67GT1	GT ₉	127	CATGATCCGGCTTAGCGAGT	3.67	1,097,170	306,039
C68TCG1	TCG ₁₃	150	CTGCTGGGGACACCTTGAAC	3.68	1,113,402	253,604
C73CA1	CA ₉	150	CTGTTTGCAGAAATGCGTCGC	3.73	1,095,011	1,034,749
C73TCG1	TCG ₇	210	CCTCGAGTGGAGTGGTCAC	3.73	1,095,011	1,014,826
C95CAG1	CAG ₆	185	GGACTGGTTGTTGACGCTGT	3.95	956,384	317,088
C95GCA1	GCA ₆	177	GTAGTGGTGCCGGTGTTCATC	3.95	956,384	620,572
C99GTC1	GTC ₆	151	ATTTGACGTGTGTACGAATACGC	3.99	949,261	355,499
C104TTC1	TTC ₆	191	GTTCTGCTGTCTGCTGTCTGA	3.104	930,648	918,077
C108CAG1	CAG ₈	229	ATGCTCTCTGGCTGGGATTG	3.108	1,001,057	416,025
C111CAG1	CAG ₅	180	TTCGGAAGTCGTCCACACT	3.111	935,238	154,520
C119TGG1	TGG ₇	228	CAGCGCACACTCGAATCTTTC	3.119	924,848	102,604
C129CT1	CT ₇	118	CTACTGCTGGCGGCTGACTT	3.129	873,558	381,523
C134AC1b	AC ₇	197	TTGCACACTGGTTGGACCTG	3.134	821,072	30,425
C139CT1	CT ₉	128	CTCCTCCACAGACATCGACACA	3.139	823,831	744,116
C139GA1	GA ₈	117	CATCAGGAACAGGGCTACCA	3.139	823,831	714,439
C143TGC1	TGC ₆	206	CATCAACACCTGCCGACAC	3.143	815,811	391,127
C163CAG1	CAG ₇	163	TCTGCTTGCTGGACGTAGTCA	3.163	813,728	163,335
C175TC1	TC ₁₀	192	AGCGGAGGTAGTGGAGCATC	3.175	767,004	379,530
C177CGT1	CGT ₆	233	CCAATCGCAACCACAAGAAC	3.177	728,683	457,894
C185CAG1	CAG ₇	146	TCCGATGGTTTATCCGACTG	3.185	709,732	232,381
C186TGT1	TGT ₈	167	AGCAATGCAAAGCAGGTGAC	3.186	747,982	310,769
C198TCG1	TCG ₆	150	CCAGTCCCAGCAGCAGAATC	3.198	672,410	515,329
			CGTGCGCTGTTCAGTAAGC			
			AGACAACATGTGCCGCATTC			
			GGTTCGCCTCGACTATCACA			
			AAGTGGCGTGAAGTCTGCTG			
			ACCACAGTCCGTCCTCTTCC			
			TAGCCTCCCGTTGTTCAAGTCA			
			CAGCATTGTTGTCATCCATC			
			GGTGAAGGTCAGCCACTCA			
			ACAGCTGACTCTCGTCGACA			
			TCTGGGCGCGAGACTTCTA			
			GCACCCGACCCAAGTTCTC			
			GCCACCCGAGCACAATAACA			
			CCACGTGGCGATAGCATAATC			
			AGCGCATAAGCCTGGAAGTT			
			CAAGCGGGTTCTTCAAGTCA			
			ATGGTCCACGTGCAGAATCA			
			CCGGAATTCAAGAGCACCTC			
			TCCGATCCGCTACACATCCA			
			CGACCGATTGAGCCAGATCA			
			CGTCGTCGTACACACTCACA			
			CCTGGCACGTCAAGAATCAC			
			GGCCAAAGCCTACCCAAGTC			
			CGTCCACCAGAATCACATCCA			
			GTATCGGGGATCGGGTTGTT			
			CCCAAGCAATGACCCATGAT			
			TGGCTGTCAACATACGTGTCAT			
			TCGAGCATAGTGACCGGAAC			

C205GAC1	GAC ₅	135	CTCGTTGTTCCACCTTTGTC ACCCGTGTCGAGGTAAGCA	3.205	667,856	100,114
C205GTC1	GTC ₆	154	CGGTAACACTTGGCACAACC CCCGTCAACCTCGAAGAACT	3.205	667,856	282,319
C208GCA1	GCA ₆	124	CTCTAGCAGCCAGATGTGCA GGCCACCCAAACAACACTTC	3.208	649,753	278,321
C213CA1	CA ₈	144	CAGCAGTTAACAACATGCGATC GGAGCTTCTTCGTCAGTTGTCA	3.213	619,851	133,631
C213CGA1	CGA ₈	242	CATACGCCACACATCCTTGC CTCACCGCAGCAGGTATAACA	3.213	619,851	601,247
C280CGT1	CGT ₅	237	GCTGCACATCGTACCAACCA CCCAAGCCACCTACCTTCAA	3.280	510,693	190,861
C309ACG1	ACG ₅	143	CGCGAAGAGCAGCTGTAACCT GGGCCATCTCGTACTCCATC	3.309	504,204	388,483
C309TGC1	TGC ₅	184	CATCCACCGGCAACTGACTA GGCAAGTTCACCAACACGAA	3.309	504,204	473,636
C446TC1	TC ₁₂	178	CAGACGGTAATCCGGAATGAC CTTGAAGGAGAAGGTATCGTGC	3.446	375,653	303,054
C474GGA1	GGA ₇	192	ACGCGGTCATCCACTTGAA GAGGCGAGATATCACCACGA	3.474	368,624	209,783
C550TGC1	TGC ₆	247	GCTGCACGAGTATGGGTTGA ACGGTGACCAACTCCAGCAA	3.550	337,235	94,189
C660CTC1	CTC ₆	163	GTGCAACGACTCCTCCTCAA GAGACGGGTCGAGCAGAAC	3.660	289,487	210,132
C660GTG1	GTG ₅	201	CTGCTGGGACAAGCTCTGCT AACTGCTGCGGCATTTCATCT	3.660	289,487	200,646
C1500CAG1	CAG ₅	182	ATCAGCAACGTCAACCACCAC CGGTTTGATGTATGCGGAAC	3.1500	83,312	38,249

Supplementary Table S3. Allele size range, observed heterozygosity (H_O), expected heterozygosity (H_E) and HWE p-value (P-v) for 75 microsatellites in field samples collected from Indianapolis, IN, USA (*C. pipiens* s.s.), and Point Fortin, Trinidad (*C. quinquefasciatus*).

Locus	Indianapolis, IN, USA						Point Fortin, Trinidad					
	n	No. of alleles	Size range (bp)	H_O	H_E	P-v	n	No. of alleles	Size range (bp)	H_O	H_E	P-v
C1CAG1	32	4	258-267	0.53	0.64	0.53	30	3	255-261	0.30	0.32	0.22
C2ACG1	29	3	140-146	0.76	0.65	0.67	30	2	143-146	0.50	0.44	0.67
C3GAC1	31	7	286-297	0.84	0.81	0.45	30	4	289-297	0.37	0.40	0.18
C3TGC1	31	3	243-249	0.52	0.45	0.61	27	2	243-246	0.15	0.14	1.00
C4TGG1	29	1	126	monomorphic	n/a		30	3	126-134	0.50	0.52	0.84
C5CGT1	32	5	204-211	0.81	0.74	0.16	31	2	208-211	0.29	0.30	1.00
C5GTG1	32	3	254-260	0.28	0.25	1.00	31	3	257-263	0.55	0.53	0.64
C12CCT1	32	5	171-182	0.75	0.70	0.42	32	4	168-178	0.47	0.57	0.30
C12GTC1	31	6	208-223	0.77	0.78	0.40	32	3	211-220	0.31	0.35	0.14
C12GTT1	31	11	157-181	0.87	0.90	0.61	32	6	171-184	0.75	0.76	0.43
C13TC1	32	9	171-186	0.81	0.84	0.44	27	3	171-175	0.44	0.39	0.28
C21TTG1	32	5	187-196	0.78	0.77	0.87	25	2	190-196	0.20	0.18	1.00
C28GT1	32	9	206-229	0.88	0.82	0.75	32	7	212-229	0.75	0.72	0.32
C32AC1	31	9	168-182	0.71	0.77	0.74	32	5	172-178	0.75	0.63	0.07
C32AG1	32	5	150-166	0.72	0.72	0.21	32	5	150-166	0.78	0.70	0.20
C32TC1b		poor amplification—no test performed					30	8	181-201	0.87	0.80	0.91
C32TGC1	32	4	120-129	0.38	0.35	0.39	31	2	123-129	0.03	0.03	1.00
C36GTT1	32	5	183-198	0.38	0.33	1.00	32	5	186-202	0.72	0.67	0.79
C38AC1		poor amplification—no test performed					30	7	148-171	0.63	0.61	0.27
C38GCT1	28	2	166-169	0.14	0.14	1.00	31	3	166-172	0.58	0.48	0.39
C41GTC1	32	6	145-162	0.66	0.72	0.60	31	2	150-161	0.35	0.34	1.00
C45AG1	30	5	141-169	0.67	0.60	0.22	29	3	167-173	0.40	0.46	0.36
C45GCT1	32	3	175-181	0.50	0.56	0.73	31	3	175-187	0.58	0.57	0.53
C47GA1	32	2	234-236	0.53	0.48	0.71	31	4	232-238	0.42	0.52	0.10
C48CGA1	23*	5	120-135	0.48	0.48	0.46	33	2	132-135	0.67	0.50	0.07
C48GTT1	23*	5	307-334	0.74	0.65	0.13	30	3	328-337	0.67	0.66	0.65
C56GCA1	31	3	192-198	0.19	0.23	0.40	30	3	195-201	0.23	0.34	0.09
C56TGT1	29	1	213	monomorphic	n/a		32	4	220-236	0.69	0.62	0.71
C60CTG1	32	6	276-286	0.88	0.78	0.20	32	1	282	monomorphic	n/a	
C65AC1	23*	9	268-286	0.65	0.75	0.65	30	11	280-303	0.83	0.82	0.51
C65CAG1	32	5	163-175	0.69	0.60	0.39	32	3	166-172	0.56	0.56	0.93
C66CGT1	31	9	115-137	0.84	0.82	0.79	32	3	131-137	0.65	0.63	0.32
C66GT1	30	9	164-184	0.47	0.86	<0.01	32	10	165-187	0.72	0.80	0.46
C66TGT1	30	4	326-335	0.71	0.73	0.99	30	4	322-335	0.57	0.55	0.96
C67GT1	31	3	127-139	0.32	0.35	0.37	29	4	125-141	0.62	0.63	1.00

C68TCG1	32	7	135-153	0.59	0.71	0.31	31	3	144-150	0.52	0.57	0.14
C73CA1	31	6	139-155	0.68	0.77	0.08	32	3	139-143	0.56	0.55	0.78
C73TCG1	32	4	205-213	0.47	0.50	0.50	32	2	210-213	0.09	0.09	1.00
C95CAG1	30	4	176-185	0.60	0.70	0.26	32	2	179-185	0.06	0.06	1.00
C95GCA1	30	3	167-176	0.57	0.60	0.39	30	1	176	monomorphic	n/a	
C99GTC1	29	4	140-155	0.73	0.67	0.13	30	2	149-152	0.47	0.51	0.72
C99TGT1	23*	8	184-223	0.74	0.82	0.15	33	4	211-220	0.61	0.54	0.65
C104TTC1	28	5	186-196	0.79	0.73	0.62	32	2	189-192	0.16	0.15	1.00
C108CAG1	32	6	216-237	0.75	0.73	0.27	32	3	225-240	0.50	0.53	0.12
C111CAG1	31	3	173-179	0.52	0.66	0.37	32	3	173-179	0.53	0.56	0.80
C119TGG1	31	4	223-233	0.48	0.66	0.13	32	3	223-233	0.47	0.51	0.33
C127TC1	23*	8	102-120	0.83	0.82	0.90	33	5	102-116	0.85	0.76	0.38
C129CT1	31	6	116-128	0.65	0.70	0.60	32	1	122	monomorphic	n/a	
C134AC1b	32	4	191-201	0.56	0.45	0.50	32	3	193-197	0.50	0.55	0.34
C139CT1	32	6	170-186	0.78	0.79	0.70	31	7	172-199	0.67	0.83	0.11
C139GA1	31	8	106-117	0.94	0.83	0.91	30	4	109-117	0.70	0.75	0.11
C139TG1	23*	13	198-237	0.70	0.88	0.14	32	4	153-167	0.69	0.65	0.12
C143TGC1	29	4	200-206	0.48	0.52	0.50	32	4	197-206	0.47	0.41	0.28
C163CAG1	29	6	150-165	0.69	0.70	0.35	29	4	153-165	0.59	0.64	0.71
C175TC1	31	8	179-201	0.71	0.81	0.22	30	6	182-193	0.67	0.71	0.14
C177CGT1	32	4	227-233	0.41	0.40	0.45	30	3	227-233	0.70	0.63	0.66
C185CAG1	32	4	140-149	0.41	0.39	1.00	31	5	125-149	0.71	0.60	0.62
C186TGT1	32	4	159-168	0.50	0.60	0.46	29	2	165-168	0.48	0.46	1.00
C198TCG1	32	4	135-150	0.63	0.70	0.89	32	4	138-150	0.78	0.70	0.90
C205GAC1	31	7	122-137	0.61	0.72	0.35	31	4	123-137	0.48	0.66	0.04
C205GTC1	29	5	144-157	0.52	0.63	0.35	30	1	154	monomorphic	n/a	
C208GCA1	32	4	115-127	0.41	0.44	0.14	31	5	115-127	0.61	0.61	0.99
C213CA1	31	4	134-140	0.84	0.67	0.08	30	7	140-152	0.81	0.79	0.12
C213CGA1	31	4	139-153	0.55	0.50	0.44	30	3	144-156	0.50	0.58	0.11
C280CGT1	28	4	235-241	0.71	0.66	0.71	31	2	239-245	0.26	0.23	1.00
C309ACG1	31	3	124-141	0.45	0.41	1.00	32	3	124-144	0.34	0.33	0.71
C309TGC1	24	3	172-181	0.29	0.60	<0.01	32	3	175-181	0.41	0.37	0.84
C446TC1	31	7	162-176	0.55	0.78	<0.01	27	3	172-178	0.15	0.18	0.19
C474GGA1	32	3	181-194	0.59	0.65	0.79	30	4	181-194	0.70	0.64	0.34
C550TGC1	32	3	245-251	0.22	0.20	1.00	28	3	248-254	0.29	0.26	1.00
C660CTC1	32	3	145-163	0.47	0.66	0.13	32	1	160	monomorphic	n/a	
C660GTG1	31	3	196-205	0.58	0.52	0.59	32	3	199-205	0.56	0.60	0.82
C1500CAG1	29	7	151-184	0.72	0.75	0.17	30	2	175-178	0.13	0.18	0.24

*These data are from Indianapolis collection site 3 in Hickner et al. (2010).