Title

Identification and assessment of single nucleotide polymorphisms (SNPs) between *Culex* complex mosquitoes.

Authors

David S. Kang, Cheolho Sim*

Affiliations

Department of Biology, Baylor University, Waco, TX 76798, USA

* Corresponding author (cheolho_sim@baylor.edu).

Introduction

Culex pipiens complex mosquitoes are vectors for many human pathogens such as West Nile encephalitis, Rift Valley Fever, and Lymphatic filariasis (Diamond 2009; Lai *et al.* 2000; Meegan *et al.* 1980; Monath 1988). Including the *Culex pipiens* form pipiens and *Culex pipiens* form molestus biotypes, the *Culex pipiens* complex mosquitoes are an urban vector, with global distribution (Reusken *et al.* 2010; Shaikevich & Vinogradova 2004; Vinogradova 2000). Despite sharing similar physiology, *Culex pipiens* complex mosquitoes each possess unique eco-physiological adaptations for survival of environmental stresses such as cold winters and divergent mating behaviors tailored to under or above ground habitats (Table 1)(Barr 1957; Clements 1992; Harbach *et al.* 1984; Spielman 1967). Isolation and disruption of the genetic bases of these differences in vector competence, geographical distribution, and behavioral/reproductive traits is crucial to the control these disease vectors. Development of genetic markers across the vector genome is the first step towards these ultimate goals.

Table 1. Divergent eco-physiological traits between two biotypes of *Culex pipiens* complex mosquitoes.

Trait / Biotype	Breeding Site	Mating Pattern	Host-feeding Preference	Vitellogenesis	Overwintering
Pipiens	Epigeous (Above ground)	Eurygamous (Open spaces)	Ornithophilic (Birds)	Anautogenous (Requires bloodmeal)	Winter diapausing
Molestus	Hypogeous (Underground)	Stenogamous (Enclosed spaces)	Mammalophillic (Mammals)	Autogenous (First oviposition does not require bloodmeal)	No diapause

As the most common form of genetic variation, single nucleotide polymorphisms (SNPs) are preeminent molecular markers in genetic high-resolution mapping and population genetics studies (Berger *et al.* 2001; Black *et al.* 2001; Venter *et al.* 2001; Wang *et al.* 1998). The abundance of SNPs allows for a higher number of evenly spaced informative markers, a potential advantage over previously employed microsatellite and RFLP markers (Bourguet *et al.* 1998; Ewing *et al.* 1998; Kothera *et al.* 2010). Based on the divergent life strategies of *Culex pipiens* complex mosquitoes, and the high SNP frequencies found in other members of Culicidae we hypothesized the two *Culex pipiens* complex biotypes will yield an abundance number of SNP markers for use in genetic study.

Here, we report a set of 28 genes with informative SNP markers characterized from two lab strain biotypes: *Culex pipiens* form pipiens and *Culex pipiens* form molestus. Genes were selected to be evenly spaced across the reference *Culex pipiens quinquefasciatus* physical map, and were adapted from previous genetic mapping studies (Arensburger *et al.* 2010; Mori *et al.* 1999). Genetic polymorphisms in both coding and noncoding regions are documented in detail with regard to nucleotide diversity, coding bias and gene function. Combined with the characterization of SNPs across the *Culex pipiens* complex genomes, the physical map will allow quantitative trait loci (QTL) analysis: the identification of which loci contribute to polygenic phenotypic variation (Severson *et al.* 2001).

Data access

NCBI SNP Database (dbSNP accession nos. ss947844444 - ss947844519).

Annotated consensus sequences data, gene ontologies, genetic map locations, supercontig information, and validation results may be accessed at the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.9fg6h

Meta-information

Sequencing center – DNA Analysis Facility on Science Hill at Yale University (New Haven, CT, USA).

Platform and model - Applied Biosystems Genetic Analyzer.

Design description - In this study we characterized a set of SNP markers between two biotypes of the *Culex pipiens* complex, *Culex pipiens* form molestus and *Culex pipiens* form pipiens, for use in a high-resolution genetic mapping and population genetics. DNA pooled from 10 female specimens of each biotype were sequenced and analyzed for variation in 28 genes. The informative SNP markers were successfully identified and assessed from both *C. pipiens* biotypes. We expect that novel SNPs characterized in this study would be useful for genetic studies to elucidate the genetic basis of diverged eco-physiological traits between the two biotypes of the *C. pipiens* complex.

Analysis type - gDNA.

Run date - Summer 2014.

Library

Strategy - Sanger Chain-Termination Sequencing.

Taxon - *Culex pipiens* complex (molestus and pipiens biotypes).

Sex - Females.

Tissue - Whole body genomic DNA.

Location - Both biotypes are lab maintained colonies. They were established from the following areas:

- *Culex pipiens* form pipiens - Columbus, Ohio (Robich & Denlinger 2005). Provided by Dr. David Denlinger at The Ohio State University.
- *Culex pipiens* form molestus - Chicago, Illinois (Mutebi & Savage 2009). Provided by Dr. Linda Kothera at the Centers for Disease Control and Prevention Division of Vector Borne Infectious Diseases at Fort Collins, Colorado.

Sample handling - All mosquitoes were immediately subjected to homogenization and gDNA extraction immediately after being knocked out for 3 minutes in a -20^{o}C freezer.

Selection - DNA sequences of microsatellite or RFLP Loci from the previous genetic studies were used to identify the contigs in the genome database of the *Culex pipiens quinquefasciatus* (Johannesburg strain) physical map (genome size 579 Mb, 18,883 predicted genes) (http://cquinquefasciatus.vectorbase.org/), a sister species of *Culex pipiens,* as these sites have been proven reliable and their selection allows for comparison of past and future studies (Arensburger *et al.* 2010; Mori *et al.* 1999). The contigs containing the markers were identified, and then thirty-six candidate genes were selected near the microsatellite or RFLP locations. Primers, ranging between 300 and 600 bp, were developed using Primer3 (http://frodo.wi.mit.edu/primer3/) spanning the nested introns of the candidate genes.

Layout – Forward and reverse sequencing.

DNA extraction - Genomic DNA for each form of mosquito was extracted from 10 pooled females with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the

manufacturer's protocol and then tested for purity on a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE).

PCR amplification - gDNA pooled from 10 female *Culex pipiens* form molestus or 10 female *Culex pipiens* form pipiens was then amplified by PCR. PCR was performed on a T100 thermal cycler (Bio-Rad, Hercules) with 100 ng of genomic DNA in a final volume of 50 µl containing 0.5 µl *Taq* polymerase (Qiagen), 5 µl 10X buffer, 1.5 µl of 10 mM dNTPs, and 5 pmoles of primer. Amplification cycles consisted of 94°C for 2 minutes; 40 cycles of 94°C for 30 seconds, 47-53°C for 30 seconds, 72°C for 45 seconds; and a final extension of 72 for 5 minutes. Annealing temperatures were optimized for each primer pair, and PCR products were visualized on a 1.2% agarose gel with a 100 bp molecular weight ladder (Invitrogen, Carlsbad, CA).

Sequencing - Samples were next submitted to the DNA Analysis Facility on Science Hill at Yale University (New Haven, CT, USA).

Processing

SNP calling - DNA sequences from PCR products were then inspected and aligned using CLC Main Workbench v6 (CLC bio). Sequence quality was assessed and sequences were cropped in regions without clear consensus between forward and reverse sequences. In both coding and non-coding regions SNPs were classified as Type I, II, III or IV. Codon positions of SNPs in coding regions were further identified, and then the ratios of transversions to transitions and those of synonymous and nonsynonymous substitutions were tested for coding bias. DNAsp v5.10.01 was then utilized to determine nucleotide diversity, Ks, and Ka (Librado & Rozas 2009).

Validation - SNPs discovered in this study were validated by resequencing with reverse primers and Amplifluor SNPs Genotyping System (Millipore, Billerica, Massachusetts) using of gDNA from 10 new pooled female mosquitoes. Briefly, primers were designed with a unique hairpin loop at the 5' end with a quencher preventing the fluorophore reporter from fluorescing. Upon specific SNP annealing, polymerases on the complementary strand open the hairpin allowing for either FAM or JOE fluorescence. Analyses were performed on a Rotor-Gene Q real-time thermal cycler (Qiagen), with reactions containing 10 ng genomic DNA, 0.5 µl 20X Amplifluor SNP FAM Primer, 0.5 µl 20X Amplifluor SNP Joe Primer, 0.5 µl 20X specific primer mix (containing 0.5 µM Green Forward Primer, 0.5 µM Red Forward Primer, 7.5 µM Common Reverse Primer), 1.0 µl 10X Reaction Mix S-Plus buffer, 0.8 µl 2.5 mM dNTPs and 0.1 µl Titanium *Taq* DNA Polymerase (Clontech). Amplification cycles consisted of 96°C for 4 minutes; 18 cycles of 96°C for 10 seconds, 53-58°C for 5 seconds, 72°C for 10 seconds; 22 cycles of 96°C for 10 seconds, 53-58°C for 20 seconds, 72°C for 40 seconds and a final extension at 72°C for 3 minutes. Annealing temperatures were optimized for each primer pair, and fluorescence was monitored as per manufacturer protocol.

Results

Table 2. Distributions and diversities of single nucleotide polymorphisms in the *Culex pipiens* complex.

	Coding																	Non-Coding					
		Codon Polymorphic Position												Nucleotide Diversity					Polymorphism				
		Transition				Transversion				# Polymorphism Types													Nucleotide Diversity
Gene	L (bp)	1st	2nd	3rd	Total	1st	2nd	3rd	Total	Syn	Nonsyn	Indel	Total	π	πn	Ks	Ka	L (bp)	Ts	Tv	Indel	Total	π
CPIJ006671	89	0	0	1	1	0	0	0	0	1	0	0	1	0.0112	0.0000	0.0492	0.0000	0	0	0	0	0	0.0000
CPIJ003890	125	0	0	1	1	0	0	0	0	1	0	0	1	0.0080	0.0000	0.0319	0.0000	0	0	0	0	0	0.0000
CPIJ009089	150	1	0	4	5	0	0	1	1	6	0	0	6	0.0400	0.0000	0.1865	0.0000	0	0	0	0	0	0.0000
CPIJ002431	110	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	65	3	2	0	5	0.0769
CPIJ004343	161	0	0	4	4	0	0	0	0	4	0	0	4	0.0248	0.0000	0.1019	0.0000	0	0	0	0	0	0.0000
CPIJ004272	62	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	64	8	2	2	12	0.1563
CPIJ013307	150	1	0	0	1	0	0	1	1	2	0	0	2	0.0133	0.0000	0.0551	0.0000	0	0	0	0	0	0.0000
CPIJ003470	104	0	0	0	0	0	1	0	1	0	1	0	1	0.0096	0.0136	0.0000	0.0137	35	0	0	0	0	0.0000
CPIJ000470	167	0	0	1	1	0	0	2	2	3	0	0	3	0.0180	0.0000	0.0706	0.0000	20	0	0	0	0	0.0000
CPIJ007696	198	0	0	3	3	0	0	3	3	6	0	0	6	0.0303	0.0000	0.1191	0.0000	18	3	0	0	3	0.1667
CPIJ005613	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	63	3	0	0	3	0.0476
CPIJ006471	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	172	2	3	0	5	0.0291
CPIJ004396	132	1	0	2	3	0	0	0	0	2	1	0	3	0.0227	0.0101	0.0711	0.0101	0	0	0	0	0	0.0000
CPIJ005878	68	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	149	1	1	1	3	0.0134
CPIJ008264	54	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	64	0	0	1	1	0.0000
CPIJ008265	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	106	6	4	1	11	0.0943
CPIJ018569	123	0	0	0	0	0	0	1	1	1	0	0	1	0.0081	0.0000	0.0364	0.0000	0	0	0	0	0	0.0000
CPIJ000207	229	1	0	2	3	0	0	1	1	4	0	0	4	0.0175	0.0000	0.0772	0.0000	0	0	0	0	0	0.0000
CPIJ005652	136	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	63	2	1	0	3	0.0476
CPIJ008758	111	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	76	0	1	0	1	0.0132
CPIJ010827	98	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	181	1	1	2	4	0.0110
CPIJ008915	229	0	0	1	1	0	0	0	0	1	0	0	1	0.0044	0.0000	0.0248	0.0000	0	0	0	0	0	0.0000
CPIJ007044	306	0	0	3	3	0	0	1	1	4	0	0	4	0.0131	0.0000	0.0574	0.0000	0	0	0	0	0	0.0000
CPIJ004516	128	0	0	0	0	0	0	1	1	1	0	0	1	0.0078	0.0000	0.0332	0.0000	0	0	0	0	0	0.0000
CPIJ013966	170	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	12	1	0	0	1	0.0833
CPIJ008915	155	0	0	2	2	0	0	4	4	6	0	0	6	0.0387	0.0000	0.1837	0.0000	0	0	0	0	0	0.0000
CPIJ013141	24	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	82	0	1	0	1	0.0122
CPIJ008369	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	152	1	1	3	5	0.0132

L is length of amplicon; Syn, synonymous substitutions; Nonsyn, replacement substitutions; Ts, transitions; Tv, Transversions; π, nucleotide diversity; πn nonsynonymous nucleotide diversity; Ks, average nucleotide substitutions per synonymous site; Ka, average nucleotide substitutions per non-synonymous site.

Table 3. Distribution of synonymous and non-synonymous substitutions between coding regions, non-coding regions in respect to transitions and transversions.

Coding Region Replacements	**Total**	**%**
Synonymous		
Transitions	28	60.9
Transversions	16	34.8
Nonsynonymous		
Transitions	1	2.2
Transversions	1	2.2
Nonsense	0	0.00
Missense	2	4.4
Total	46	

Noncoding Region Replacements	**Total**	**%**
Transitions	31	64.6
Transversions	17	35.4
Total	48	

Total Replacements	**Total**	**%**
Transitions	60	63.8
Transversion	34	36.2
Total	94	

Table 4. Polymorphism class and the degeneracy of the genetic code.

Polymorphism	Coding	Wobble Position	Fourfold degenerate	Non-coding
Transitions				
Class I (C/T or G/A)	28	24	4	31
Transversions				
Class II (C/A or G/T)	2	1	1	9
Class III (C/G)	12	12	12	1
Class IV (A/T)	2	2	2	7
Indels	0	0	0	10
Total	44	39	19	58

Figure 1. Distribution of single nucleotide polymorphisms for all regions examined between molestus and pipiens biotypes. Purine to purine substitutions and pyrimidine to pyrimidine substitutions are defined as transitions, while purines to pyrimidine mutations or vice versa are classified as transversions.

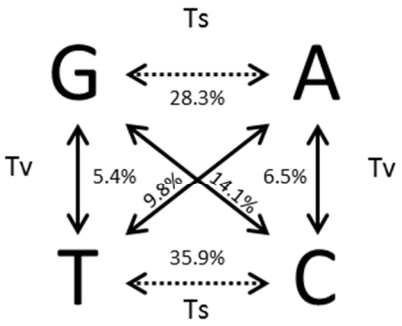
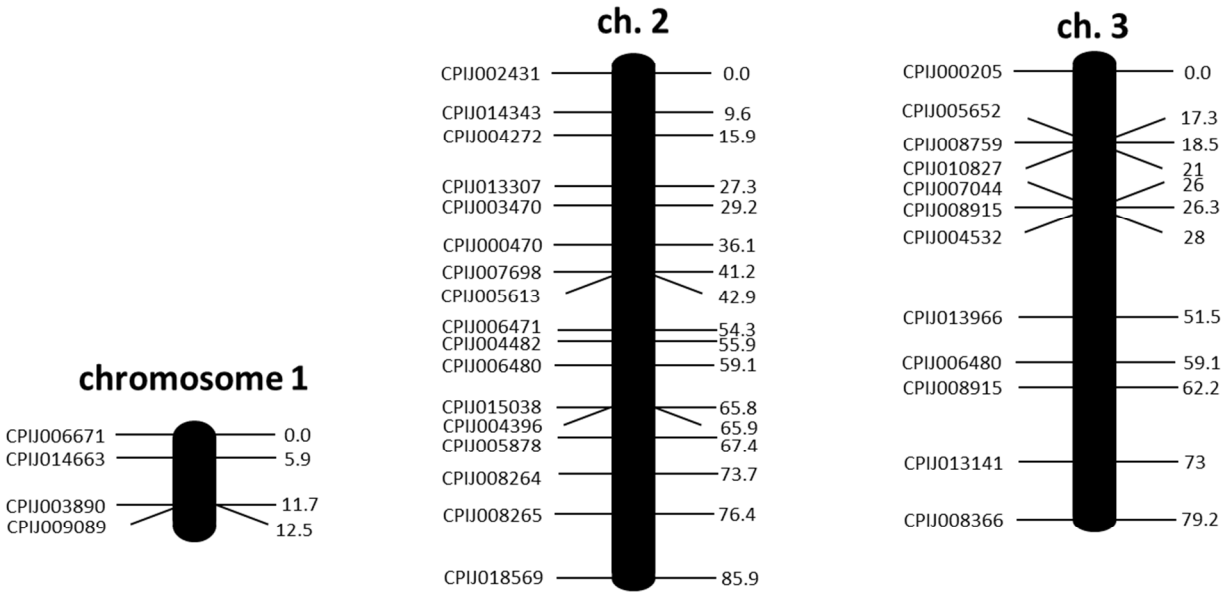


Figure 2. Integration of the genetic linkage map and the physical map. The genetic map and genetic distance (cM) is adapted from our study and previous genetic mapping analysis (Arensburger *et al.* 2010; Mori *et al.* 1999).



References

Arensburger P, Megy K, Waterhouse RM*, et al.* (2010) Sequencing of Culex quinquefasciatus establishes a platform for mosquito vomparative henomics. *Science* **330**, 86-88.

Barr AR (1957) The distribution of Culex p. pipiens and C.P. quinquefasciatus in North America. *Am J Trop Med Hyg* **6**, 153-165.

Berger J, Suzuki T, Senti KA*, et al.* (2001) Genetic mapping with SNP markers in Drosophila. *Nature Genetics* **29**, 475-481.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology* **46**, 441-469.

Bourguet D, Fonseca D, Vourch G*, et al.* (1998) The acetylcholinesterase gene Ace: a diagnostic marker for the Pipiens and Quinquefasciatus forms of the Culex pipiens complex. *J Am Mosq Control Assoc* **14**, 390-396.

Clements AN (1992) *The biology of mosquitoes*, 1st edn. Chapman & Hall, London ; New York.

Diamond MS (2009) *West Nile encephalitis virus infection : viral pathogenesis and the host immune response* Springer, New York, NY.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175-185.

Harbach RE, Harrison BA, Gad AM (1984) Culex (Culex) Molestus Forskal (Diptera, Culicidae) - neotype designation, description, variation, and taxonomic Status. *Proceedings of the Entomological Society of Washington* **86**, 521-542.

Kothera L, Godsey M, Mutebi JP, Savage HM (2010) A comparison of aboveground and belowground populations of Culex pipiens (Diptera: Culicidae) mosquitoes in Chicago, Illinois, and New York City, New York, using microsatellites. *J Med Entomol* **47**, 805-813.

Lai CH, Tung KC, Ooi HK, Wang JS (2000) Competence of Aedes albopictus and Culex quinquefasciatus as vector of Dirofilaria immitis after blood meal with different microfilarial density. *Vet Parasitol* **90**, 231-237.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Meegan JM, Khalil GM, Hoogstraal H, Adham FK (1980) Experimental Transmission and Field Isolation Studies Implicating Culex-Pipiens as a Vector of Rift-Valley Fever Virus in Egypt. *American Journal of Tropical Medicine and Hygiene* **29**, 1405-1410.

Monath TP (1988) *The Arboviruses : epidemiology and ecology* CRC Press, Boca Raton, Fla.

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.

Mutebi JP, Savage HM (2009) Discovery of Culex pipiens pipiens form molestus in Chicago. *J Am Mosq Control Assoc* **25**, 500-503.

Reusken CBEM, de Vries A, Buijs J*, et al.* (2010) First evidence for presence of Culex pipiens biotype molestus in the Netherlands, and of hybrid biotype pipiens and molestus in northern Europe. *Journal of Vector Ecology* **35**, 210-212.

Robich RM, Denlinger DL (2005) Diapause in the mosquito Culex pipiens evokes a metabolic switch from blood feeding to sugar gluttony. *Proc Natl Acad Sci U S A* **102**, 15912-15917.

Severson DW, Brown SE, Knudson DL (2001) Genetic and physical mapping in mosquitoes: molecular approaches. *Annual Review of Entomology* **46**, 183-219.

Shaikevich EV, Vinogradova EB (2004) [Molecular genetic methods for the identification of the urban mosquito Culex pipiens pipiens F. molestus (Diptera, Culicidae)]. *Parazitologiia* **38**, 406-412.

Spielman A (1967) Population structure in the Culex pipiens complex of mosquitos. *Bull World Health Organ* **37**, 271-276.

Venter JC, Adams MD, Myers EW*, et al.* (2001) The sequence of the human genome. *Science* **291**, 1304-+.

Vinogradova AB (2000) *Culex Pipiens Pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control* Pensoft.

Wang DG, Fan JB, Siao CJ*, et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.