An Evaluation of Population Differentiation in *Culex tarsalis* in Montana Using *COI* Sequences

Jake Fiocchi

DEPARTMENT OF LIFE AND ENVIRONMENTAL SCIENCES

SIGNATURE PAGE

This thesis for honors recognition has been approved for the

Department of Life and Environmental Sciences.

Jennifer Glowienka — Director — 4/29/16 — Date

Brandon A. Hedges — Reader — 4/29/16 — Date

W M Nye — Reader — 4/29/16 — Date

ABSTRACT: The Infectious Disease Ecology project at Carroll College has established protocols to successfully detect West Nile virus in *Culex tarsalis*, a species of mosquito that acts as a vector for this disease. The detection data combined with ecological factors have been used to produce a risk model capable of predicting where West Nile positive specimens will be detected annually. This specific project aimed to determine the relatedness of *C. tarsalis* from various regions across the state of Montana as this knowledge could aid in determining how West Nile virus infections may spread across the state. *Cytochrome c oxidase I* (*COI*) gene sequences were analyzed from specimens from eleven locations across the state. Analyses included population pairwise distances and F_{ST} calculations, an analysis of molecular variance, and a neighbor-joining phylogenetic analysis. The results of these analyses suggest the Montana populations of *Culex tarsalis* are not genetically differentiated from one another. This means either that the populations truly are panmictic or that the sequence used is not informative for this question. Due to the low sequence variation among individuals, it seems more likely that this sequence is not an informative one for this study.

INTRODUCTION: Mosquitoes are common vectors of many diseases, including the West Nile virus (WNV), and the most prevalent carrier of WNV is *Culex tarsalis* (Barker et al. 2009). WNV was first documented in the state of Montana during the summer of 2002. Since that time, there have been over 550 documented cases in humans, over 300 documented cases in horses, and 12 WNV associated human deaths in the state of Montana (Opper 2016). As such, it is important that an ecological, biological, and

population based understanding of these carrier mosquitoes is established as this understanding may be key to aiding in the prevention of this disease.

Models that incorporate temperature variability, density of virulent competent birds, and *C. tarsalis* niche assessments have been created and used to predict where WNV positive mosquitoes will be detected in Montana annually (Hokit et al. 2013). However, the need to understand the relationship that exists between individual mosquito populations from various regions within the state is also important as this information may allow for the determination of where viral infections may spread. Genetic analyses have previously been used to determine migration and expansion of populations of *C. tarsalis* in the western United States (Venkatesan et al. 2007). Microsatellite markers have demonstrated that populations can be separated by both temporal and spatial boundaries (Gimnig et al. 1999; Venkatesan et al. 2007; Venkatesan et al. 2009). For example, the comparison of *C. tarsalis* populations' gene flow and West Nile virus' dispersal in the western United States helped demonstrate *C. tarsalis's* role as a vector (Venkatesan and Rasgon 2010). Unfortunately, all attempts made to understand gene flow in Montana with the utilization of microsatellite analysis in the Carroll College laboratories have not been informative to date (Pers. Comm. J. Glowienka).

The present research used the mitochondrial *cytochrome c oxidase subunit I* (*COI*) gene to compare genetic variation among populations of *C. tarsalis* in Montana. This gene has been selected due to its identity as a strong candidate for comparison in many invertebrates (Folmer et al. 1994). In populations of the Sonoran Desert, analysis of this gene revealed distinct formation of separate clades of *Culex* genera, and, in Chicago and New York, *COI* was presented as a sequence to distinguish between aboveground

and underground populations of *Culex pipiens*, however, the results did not distinguish these populations (Pfeiler et al. 2013; Kothera et al. 2010). I hypothesize that it will be possible to determine the connectedness of *C. tarsalis* from various regions across the state and distinguish populations that are relatively reproductively isolated. This may give insight into whether or not WNV is likely to transfer between the tested regions due to mosquito movement. These results could be incorporated into the risk model assessment presented to the state.

MATERIALS AND METHODS: Mosquitoes were collected from various regions of the state of Montana. The collection sites were determined by Dr. Grant Hokit on the basis of the nearby plant life, the existence of a water source, and the historical success of the collection site (**Figure 1**). Eleven sites were used in this study with 137 specimens (**Table 1**). These sites are marked in red on the collection map.

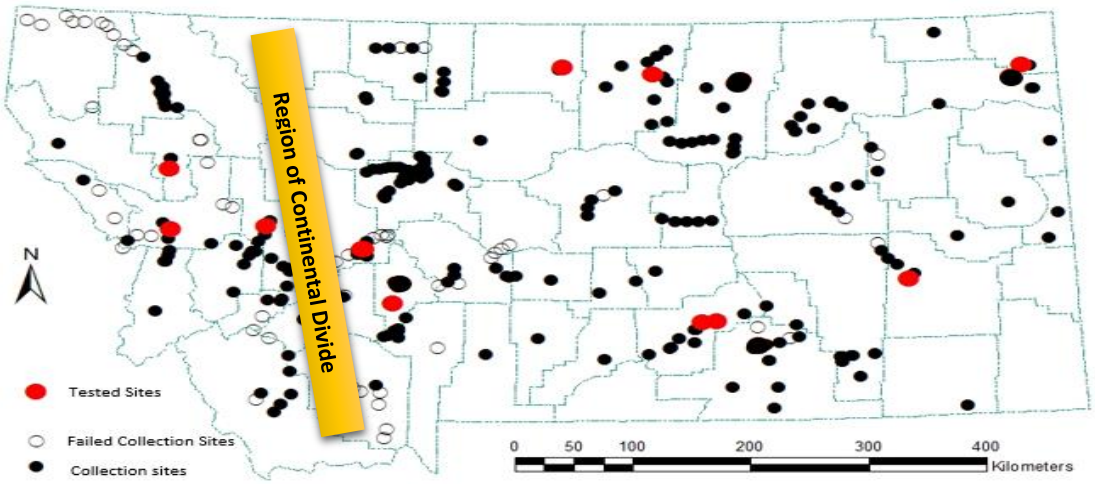

Figure 1: Sampling Collection Map.

Table 1: Test site identity, number of samples, and nucleotide diversity values.

Site County	Location	Sequences Obtained	Nucleotide Diversity
Lewis & Clark(L)	Canyon Ferry	8	0.002997
Powell(P)	Helmville	13	0.006171
Broadwater(B)	Toston	19	0.005957
Lake(A)	Ninepipe	6	0.004351
Blaine(N)	Harlem	6	0.005750
Missoula(M)	Fort Missoula	12	0.004945
Custer(C)	Miles City	9	0.004792
Sheriden(S)	Homestead	14	0.007728
Hill(H)	Havre	8	0.003164
Yellowstone(Y)	Bundy Bridge	19	0.004580
Yellowstone(E)	Gritty Stone	23	0.004496

The traps used to collect mosquitoes were CDC miniature light traps developed by J.W. Hock. The traps were baited with carbon dioxide from either a compressed tank or dry ice, set out at dusk, and any collected mosquitoes removed in the early morning. The mosquitoes were then placed in a -20℃ freezer for more than three hours to sacrifice the mosquitoes without damaging their DNA or RNA (Hokit et al. 2013). Upon removal

from the freezer, the trapped mosquitoes were sorted into non-target and target groups, which included *C. tarsalis*. Genetic material was then extracted from individual *C. tarsalis* from each collection site based on the method of Black and DuTeau (1997) which is summarized as follows: Individual sorted mosquitoes were transferred into 1.5 mL microcentrifuge tubes and homogenized in 25 µL of a grinding buffer containing 0.1 M NaCl, 0.2 M sucrose, 0.1 M Tris-HCl (pH 9.1), 0.05 M EDTA, and 0.05% SDS. After incubation for 30 minutes at 65˚C, 7 µL of 8 M KOAc solution was added to each sample; after which the homogenates were incubated on ice for 30 minutes to allow for the precipitation of the SDS from the original grinding buffer. The samples were centrifuged at 17,000 g for 15 minutes, and the supernatant was collected into a new 1.5 mL microcentrifuge tube. Each tube then had 100 microliters of pure ethanol added to the sample and was allowed to sit at room temperature until genetic material was precipitated. This precipitate was then retained as the pellet of another centrifugation, dried using a Speed Vac for 30-45 minutes, and resuspended in nuclease free water. The DNA samples were then used in a polymerase chain reaction adapted from Folmer et al. (1994) designated to amplify the *COI* gene sequence with the following specifications (**Table 2)**:

Table 2: Original master mix dictated by Folmer et al. (1994) and adapted master mix.

Component	Original Master Mix (µL per sample)	Adapted Master Mix (µL per sample)
Nuclease Free Water	28.2	31
10X Standard Taq Reaction Buffer	5	5
25 mM $MgCl_2$	5	5
Forward Primer(LCO1490)	2.5	2.5
Reverse Primer (HCO2198)	2.5	2.5
10 mM dNTP Mix	5	1
Taq DNA Polymerase 5000 U/mL	0.8	1
DNA Sample	1	2

Each primer was obtained in a dehydrated form and subsequently brought to a concentration of 100 µM. Before use in PCR reactions, primers were diluted to a concentration of 10 µM, as specified in Folmer et al. (1994; **Table 3**).

Table 3: Primers used and sequences.

Primer	Sequence
LCO1490	GGTCAACAAATCATAAAGATATTGG
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA

Following the creation of the master mix and addition of the DNA samples, the PCR reaction samples were placed in a thermocycler following the protocol of Folmer et al. (1994; **Figure 2**).

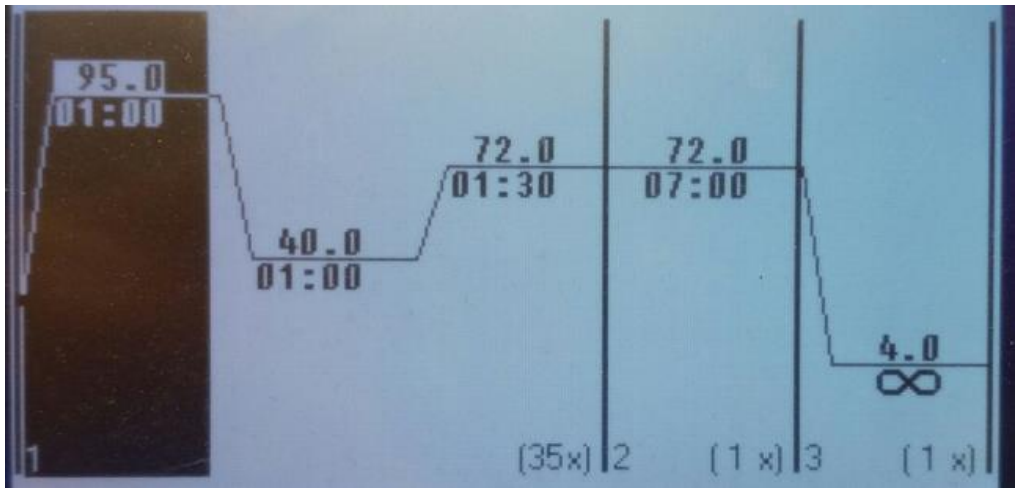

Figure 2: Folmer et al. (1994) thermocycler protocol. The x-axis represents time while the y-axis represents temperature in Celsius. The numbers above the line are the temperature settings for that period, while the numbers below represent the duration of the period. Vertical lines indicate sections where repeats occur, while the number in parenthesis indicates the number of repeats.

The reaction samples were then run out onto an agarose gel using electrophoresis to determine the success of the reaction. Each sample that was amplified was prepared for sequencing by administering 5 µL of ExoSap-It, manufactured by Affymetrix, Inc., then placing it in the thermocycler for the ExoSap-It protocol: holds samples at 37˚C for 15 minutes and then at 80˚C for 15 minutes (Pers. Comm. Glowienka). The resulting samples were then sent to Macrogen, Inc. in South Korea for sequencing. The subsequent sequences, forward (LCO) and reverse (HCO), for each sample were then used to create sample contiguous sequences in the CodonCode Aligner program (www.codoncode.com). These sequences were aligned using the ClustalX program (Thompson and Jeanmougin N.D). From ClustalX, the sequences were transferred to Arlequin 3.5 for Windows (Excoffier and Lischer 2010), where they were analyzed for nucleotide diversity per population, population pairwise F_{ST} values, and haplotype frequencies. An analysis of molecular variance (AMOVA) was also completed to determine if the East-West divide of Montana could be a barrier to gene flow (**Table 4**).

PAUP* (Swofford 2003) was used to generate a neighbor-joining tree which was used to determine if individuals from the same population clustered together phylogenetically.

Table 4: AMOVA Groups

West	Powell	Lake	Missoula					
East	L & C	Broadwater	Blaine	Custer	Sheriden	Hill	Yellowstone-1	Yellowstone-2

RESULTS: Population level nucleotide diversity was low with an average value of 0.004 substitutions per site (**Table 3**). All population pairwise F_{ST} values were low, with most p-values much greater that the alpha level of significance of 0.05. The F_{ST} value for Yellowstone County Bundy Bridge site and Sheriden County site were significantly different. These values suggest that, overall, variation among the haplotype frequencies within each population is extremely similar (**Table 5**).

Table 5: Population pairwise F_{ST} values. County abbreviations are as follows: Lewis & Clark=L, Powell=P, Broadwater=B, Lake=A, Blaine=N, Missoula=M, Custer=C, Sheriden=S, Hill=H, Yellowstone (Bundy Bridge)=Y, Yellowstone (Gritty Stone)=E. P-values for all pairwise F_{ST} values were greater than 0.05 with the exception of the comparison marked with *.

	L	**P**	**B**	**A**	**N**	**M**	**C**	**S**	**H**	**Y**	**E**
L	0.00000										
P	-0.01675	0.00000									
B	-0.03157	-0.06041	0.00000								
A	-0.04563	-0.07361	-0.07010	0.00000							
N	-0.02546	-0.02565	-0.04566	-0.04962	0.00000						
M	-0.07904	-0.05405	-0.03926	-0.05931	-0.03189	0.00000					
C	-0.02922	-0.02103	-0.04458	-0.03121	-0.02889	-0.03854	0.00000				
S	0.00057	0.05242	0.02965	0.01615	0.02814	-0.04576	0.01199	0.00000			
H	-0.03374	0.12923	0.03005	0.06032	0.00188	-0.05521	0.05156	-0.06220	0.00000		
Y	-0.00578	0.00420	-0.01373	-0.00347	-0.00955	0.03195	-0.0673	0.06465*	0.07163	0.00000	
E	-0.01789	-0.01706	-0.02961	-0.04392	-0.01553	-0.04694	-0.02905	0.03712	0.02443	0.00850	0.00000

For the AMOVA, the groups were determined by their respective location in relation to the mountain range- either east or west of the Continental Divide (**Table 4**). The results of the AMOVA revealed that for this data set, 100% of variation occurred within populations with no variation detected among groups or among the populations within groups.

Among the 137 COI sequences used in this study, there were 388 invariable sites and 44 variable sites, of which 17 sites were phylogenetically informative while 27 sites were not. The unrooted neighbor joining tree shows that individuals from the same population cluster phylogenetically with individuals from other populations. In no case is a single clade comprised primarily of individuals from one population (**Figure 3**). On this tree, each of the branches from the main base represents a unique haplotype, totaling over thirty.

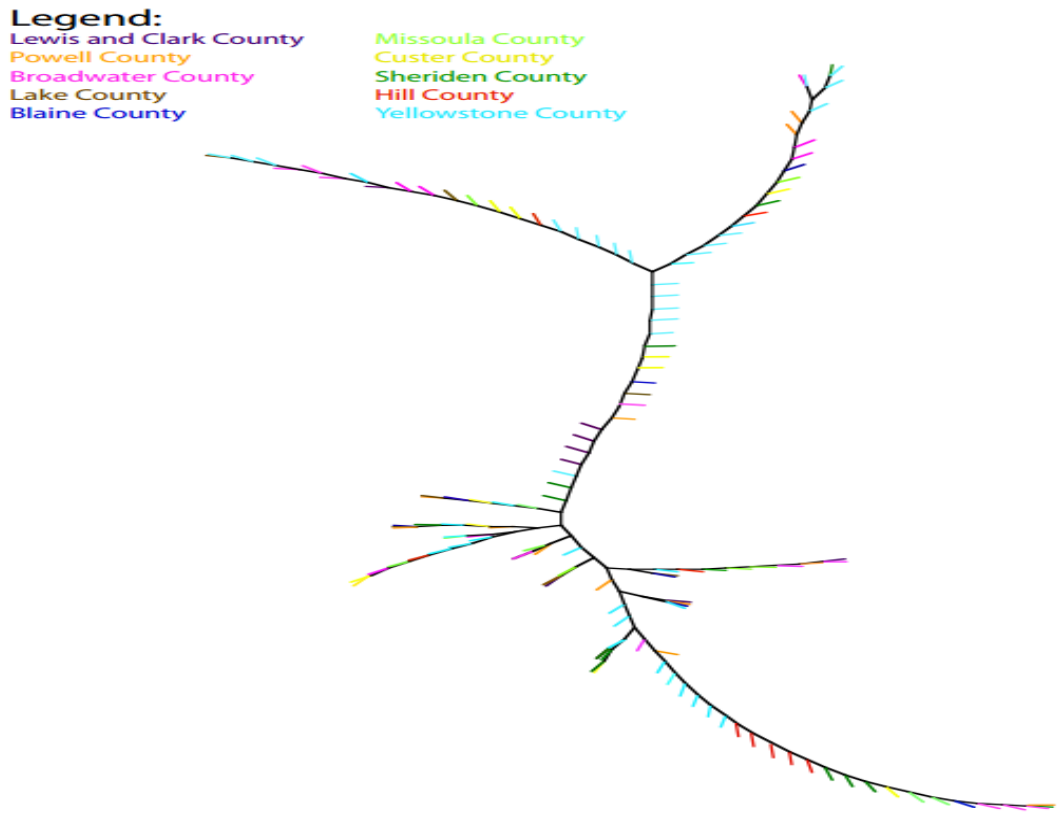

Figure 3: Unrooted Neighbor-Joining Tree Each branch represents a different individual, with population membership identified by color (see legend).

DISCUSSION: The results of this research indicate that there is very low diversity among *COI* sequences and that populations are not genetically differentiated from one another. Previous studies have shown *COI* nucleotide diversity values in *C. tarsalis* range from 0.005 to 0.0079 (Venkatesan et al. 2007). Only four of the eleven populations in this

study had values in this range while the remaining seven had lower values. The majority of F_{ST} values produced by this data set have associated p-values exceeding the 0.05 level of significance. This suggests that variation in sequences of the populations tested were similar to the point of being almost identical. These F_{ST} values indicate that the populations of *C. tarsalis* in the state of Montana demonstrate extensive panmixia. The AMOVA analysis indicates that most of the genetic variation occurs within the state population rather than within regions of the state, suggesting extensive gene flow among regions while simultaneously refuting the hypothesized border created by the Continental Divide. That being said, the fact that the test results show very little variation among groups or populations within groups is suspect as previous studies based on microsatellite analysis have found temporal and spatial separation of clades of *C. tarsalis* (Gimnig et al. 1997).

The results of this study suggest that either there is no distinction among the *C. tarsalis* populations across the state or the gene sequence used for this analysis, *COI,* is not an effective indicator sequence for population differentiation in this species. The phylogenetic tree (**Figure 3**) visualizes the lack of population separation effectively, and, given the low number of informative loci, it seems more likely that the sequence used was not an informative one for this population level analysis. *COI* might be a sequence better aimed for separation of species in a genus, as it has been demonstrated in the Sonoran Desert where it resolved 6 clades of separate *Culex* species rather than of individuals within a single species (Pfeiler et al. 2013). A similar study in New York and Chicago used multiple methods of analysis to differentiate *C. pipiens* populations above and belowground and only the use of *COI* suggested the aboveground and belowground

populations were the same (Kothera et al. 2010). This may indicate that, as *COI* is a functional gene, it may be too conserved for this level of analysis. Future research should look to determine the relatedness of *C. tarsalis* from various regions within the state using a different gene that is less conserved than *COI*, thus being more variable and, hopefully, informative loci can be incorporated. A possible candidate gene could be the mitochondrial *NADH Dehydrogenase Subunit 5* gene sequence as it has successfully formed clades in studies of the Asian tiger mosquito, *Aedes albopictus*, in Brazil and the United States (Birungi and Munstermann 2002). Microsatellite analysis could also be revisited as it has been used for informative comparisons on many occasions (Gimnig et al. 1999, Venkatesan et al. 2007, Venkatesan et al. 2009). Based on the results of this study, it is equally likely for WNV to spread to any region of the Montana. Therefore, it is not possible to predict WNV statewide movement.

ACKNOWLEDGEMENTS: Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM103474. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. These funds were awarded to Dr. Grant Hokit. Research reported in this publication was also supported by the Howard Hughes Medical Institute under award number 52007534. This grant was also awarded to Dr. Grant Hokit. This project was completed under the guidance of Dr. Jennifer Glowienka and would not have been possible without the contributions of Drs. Grant Hokit and Sam Alvey, as well as many past and present Carroll College students (especially Meghan McKeown and Seth Dotson, without whom my work load would have been drastically more daunting) and the students and faculty

from our cooperators: Montana State University, Little Big Horn College, Chief Dull Knife College, and Aaniih Nakota College. I would also like to thank Drs. David Hitt, Brandon Sheafor, and Gerald Shields for their assistance in the writing and reviewing of this paper.

LITERATURE CITED:

Barker, C.M., B.G. Bolling, W.C. Black IV, C.G. Moore, L. Eisen. 2009. Mosquitoes and West Nile virus along a river corridor from prairie to montane habitats in eastern Colorado. Journal of Vector Ecology. 34(2): 276-293.

Birungi, J. and Munstermann, L.E. 2002. Genetic structure of *Aedes albopictus* (Diptera: culicidae) populations based on mitochondrial *ND5* sequences: evidence for an independent invasion into Brazil and United States. Annals of the Entomological Society of America. 95(1): 125-132.

Black, W.C., and DuTeau, N.M. 1997. RAPD-PCR and SSCP analysis for insect population genetic studies. pp.361-373 in *The Molecular Biology of Insect Disease Vectors: A Methods Manual*. Chapman & Hall, New York.

Excoffier, L., and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Folmer, O., M. Black, W. Hoeh, R. Lutz, R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology. 3(5): 294-299.

Gimnig, J.E., W.K. Reisen, B.F. Eldridge, K.C. Nixon, S.J. Schutz. 1999. Temporal and spatial genetic variation within and among populations of the mosquito *Culex tarsalis* (Diptera: culicidae) from California. Journal of Medical Entomology. 36(1): 23-29.

Hokit, G., S. Alvey, J.M.O. Geiger, G.D. Johnson, M.G. Rolston, D.T. Kinsey, N.T. Bear. 2013. Using undergraduate researchers to build vector and West Nile virus surveillance capacity. International Journal of Environmental Research and Public Health. 10: 3192-3202.

Kothera, L., M. Godsey, J. Mutebi, H.M. Savage. 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. Journal of Medical Entomology. 47(5): 805-813.

Opper, R. 2016. West Nile Virus (United States, Montana Department of Public Health and Human Services). http://dphhs.mt.gov/publichealth/cdepi/diseases/westnilevirus.aspx

Pfeiler, E., C.A. Flores-Lopez, J.G. Mada-Velez, J. Escalante-Verdugo, T.A. Markow. 2013. Genetic diversity and population genetics of mosquitoes (Diptera: Culicidae: Culex spp.) from the Sonoran Desert of North America. The Scientific World Journal. 2013: 1-11.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J., and Jeanmougin, F. N.D. Directory of /pub/ClustalX [Computer software]. Strasbourg: The Authors.

Venkatesan, M., C.J. Westbrook, M.C. Hauer, J.L. Rasgon. 2007. Evidence for a population expansion in the West Nile virus vector *Culex tarsalis*. Molecular Biology and Evolution. 24(5): 1208-1218.

Venkatesan, M., K.W. Broman, M. Sellers, J.L. Rasgon. 2009. An initial linkage map of the West Nile virus vector *Culex tarsalis*. Insect Molecular Biology. 18(4): 453-463.

Venkatesan, M. and Rasgon, J.L. 2010. Population genetic data suggest a role for mosquito-mediated dispersal of West Nile virus across the western United States. Molecular Ecology. 19(8): 1-12.