SIGNATURE PAGE

This thesis for honors recognition has been approved for the

Department of Life & Environmental Sciences.

Jennifer Vanmeter — Director — 4/25/17 — Date

Grant Hurst — Reader — 4/27/17 — Date

Kyle Tu — Reader — 5/1/17 — Date

Dan Lockman

Genetic Differentiation Among *Culex tarsalis* Populations in Montana

Dan Lockman

Department of Life and Environmental Sciences, Carroll College

Helena, MT

May 2017

Table of Contents

List of Figures

List of Tables

Abstract

Undergraduate researchers at Carroll College have been working to detect West Nile virus in the mosquito *Culex tarsalis*, the primary vector for the disease in the state of Montana. The purpose of this research is to use West Nile positive detections along with ecological data to create a risk model for the state of Montana that could predict the likelihood of a West Nile outbreak. The research presented in this paper attempts to further the goal of the risk model by determining the degree of genetic differentiation among different mosquito populations in Montana. Determining the genetic relatedness of different populations may help to predict how West Nile virus spreads across the state. One hundred twenty four individuals from eleven different locations in Montana were analyzed using ISSR analysis. Analyses included population pairwise F_{ST} values, percent polymorphic loci, and sources of variation between populations, all of which were calculated using analysis of molecular variance (AMOVA). Results suggest that while different populations of *Cx. tarsalis* seem to be genetically different, there is no evidence of genetic structure across the Continental Divide. Therefore, it appears that the Continental Divide does not act as a barrier to gene flow. It could be that the Continental Divide does not prevent the movement of mosquitoes from eastern Montana to western Montana. However, due to the high degree of genetic variability within populations, it seems more likely that this method of analysis yielded too much variation for genetic structure to be detected on a statewide level.

Introduction

West Nile virus was initially isolated in 1937 in the African country of Uganda, and has been found historically in Africa, Asia, and Southwest Asia (Hayes 2001). A member of the genus *Flavovirus*, West Nile virus was first introduced into the United States via New York in 1999 (Davis 2004). Subsequently the epidemic has spread from New York to the West Coast; this continent-wide spread of the virus took only three years, appearing in most states by 2002 (CDC 2002). The virus is primarily transferred by mosquitoes, and among the mosquito population, *Culex tarsalis* has demonstrated the greatest efficiency as a vector of West Nile virus (Goddard 2002; Barker 2009). As the virus has become a health concern across the western United States, it has become apparent that more should be known about the population dynamics of the mosquitoes that carry the virus.

West Nile virus was first detected in the state of Montana in 2002. Since then, efforts have been undertaken to develop geographic models to determine the risk of West Nile virus infection in the state (Hokit et al., 2013). These models were primarily created using mosquito infection rate data acquired via real time PCR analysis of *Cx. tarsalis* populations, and suggest that infection rates of West Nile virus are significantly higher in the eastern portion of the state (Hokit et al., 2013). These results suggest that there is some barrier to *Cx. tarsalis* gene flow in to western Montana. The Continental Divide acts as a geographic barrier, preventing the movement of mosquitoes, and, by extension, the movement of the virus (Barker et. al., 2009). The effectiveness of this barrier has been attributed to mountain ecosystems that vary greatly in temperature as the elevation rises, which directly affects the replication cycle of the virus (Barker et. al., 2009). Furthermore, the Tehachapi Mountains in Southern California act as a barrier to gene flow among native populations of *Cx. tarsalis* (Gimnig et al., 1999). As a result, the native

populations on either side of the Tehachapi Mountains have a greater degree of genetic variation than would be expected based on seasonal variation. (Gimnig et al., 1999). Additionally, Eisen et al. (2008) showed that the abundance of *Cx. tarsalis* populations decreased significantly as elevation rose above 1600 meters; again, this decrease may have been due primarily to lower temperatures and thicker forest cover. If a comprehensive geographic risk model for the state of Montana is to be developed, it would be helpful to further investigate what factors are important to the distribution and gene flow of *Cx. tarsalis* across the state. Specifically, I wanted to examine how the Continental Divide might change infection rate dynamics by separating and possibly isolating *Cx. tarsalis* populations in Montana.

The effect of the Continental Divide as a means of preventing gene flow among Montanan populations of *Cx. tarsalis* was investigated when Fiocchi (2016) used *cytochrome c oxidase I* (*COI*) mitochondrial gene sequences to determine the degree of genetic differentiation among Montanan populations of *Cx tarsalis*. No significant degree of differentiation was found and Fiocchi (2016) suggested that this was due to the *COI* sequence being uninformative for the scope of his study. He went on to further suggest that because *COI* is such an important gene for the survival of the organism, it may be too highly conserved to be informative when asking questions about statewide genetic differentiation (Fiocchi 2016). It seems, therefore, that a less conserved sequence may be needed in order to answer the question of genetic differentiation.

To determine the genetic diversity among *Cx. tarsalis* populations from different locations across the state, inter-simple sequence repeats, or ISSRs, were analyzed in the present study. ISSRs are regions between microsatellite loci within the genome of the organism, and they have been used previously with a high degree of success to determine variation in several aphid species as well as the mosquito species *Aedes aegypti* (Abbott 2001). Several ISSR primer

sequences have been developed specifically for use in *Cx. tarsalis*, and these are the primers that were used in this study (Venkatesan 2008). I hypothesize that Montanan populations of *Cx. tarsalis* from the west side of the Continental Divide will be genetically distinguishable from populations on the east side of the Continental Divide. Documenting this genetic diversity may show that gene flow of *Cx. tarsalis* across the state is inhibited by the Divide.

Cx. tarsalis populations from eleven different sites from across the state of Montana were analyzed for this study. Three of these sites are on the west side of the Continental Divide, and range as far north as the Flathead Reservation and as far south as Helmville. The other eight sites are on the east side of the Continental Divide and range as far north as Havre and as far south as the Billings area. *Cx. tarsalis* populations were analyzed via PCR and gel electrophoresis banding patterns. The banding patterns were further analyzed to determine allele frequencies. The PCR protocol of Venkatesan (2008) was used. For continuity, all sampled populations were analyzed using only ISSR primer 2.

Materials & Methods

Mosquitoes were captured from various sites from across Montana between the years 2009 and 2012 by former students of Carroll College. The mosquitoes were baited and caught using CDC miniature light traps developed by J.W. Hock. The mosquitoes were then frozen at either -20°C for 3-4 hours, or -80°C for at least one hour. Both freezing protocols were equally as effective and were used interchangeably based on time constraints. The mosquitoes were then sorted to isolate *Cx. tarsalis*. *Cx. tarsalis* then had their genetic material extracted according to the method of Black and DuTeau (1997). The DNA was stored at -20°C in the Wiegand Undergraduate Research Lab at Carroll College.

I chose to analyze *Cx. tarsalis* individuals from eleven different sites in Montana (**Table 1**); three of the sites were from the west side of the Continental Divide and eight of the sites were from the east side of the Continental Divide (**Figure 1**). These sites were chosen in order to re-evaluate the work of Fiocchi (2016). The goal was to determine if using ISSR analysis instead of *COI* analysis would yield a more informative result.

Table 1. *Cx. tarsalis* Site location, date collected and individuals per site

Sample Number	County (East/West of CD)	Location	Number of Individuals	Date Collected
32	Lewis & Clark (E)	Canyon Ferry	6	2009
42	Powell (W)	Helmville	9	2009
459	Broadwater (E)	Toston	19	7/5/2011
485	Lake (W)	Ninepipe	7	7/11/2011
492	Blaine (E)	Harlem	6	7/12/2011
510	Missoula (W)	Fort Missoula	13	7/18/2011
573	Custer (E)	Miles City	9	8/6/2011
582	Sheridan (E)	Homestead	14	7/29/2011
620	Hill (E)	Havre	8	8/16/2011
699	Yellowstone (E)	Bundy Bridge	19	7/19/2012
700	Yellowstone (E)	Gritty Stone	14	7/19/2012

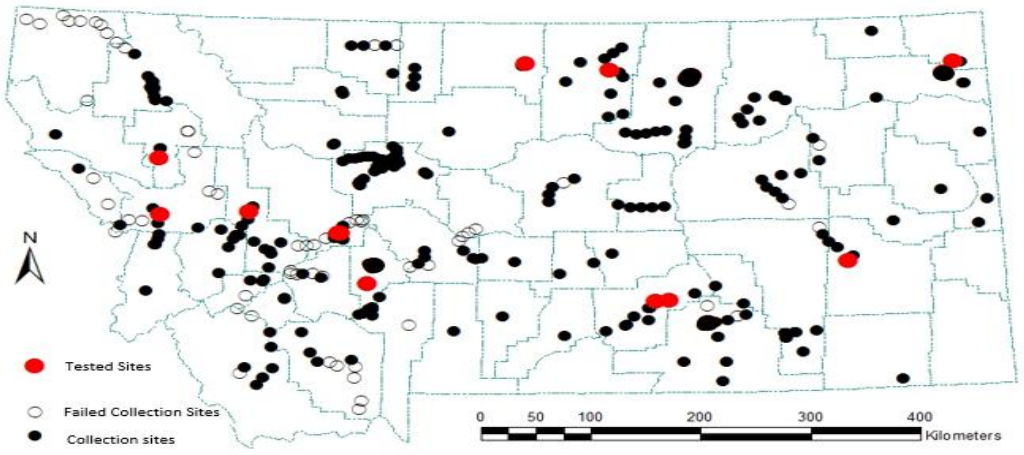

Figure 1. Map of collection sites in Montana. The Continental Divide separates the three westernmost tested sited from the other eight tested sites east of the Continental Divide.

The DNA from each individual was then amplified via polymerase-chain reaction (PCR) (Venkatesan 2008; **Table 2**). A master mix was made for each reaction, and 24 µL of this master mix was added to PCR tubes, along with 1 µL of template DNA. The samples were then placed in a thermocycler and subjected to the conditions outlined by Venkatesan (2008; **Table 3**). In every reaction, at least one negative control was included; in some cases, up to three negative controls were included. More negative controls were added in later PCRs to help differentiate between random tube contamination and reagent contamination. These negative controls contained 24 µL of master mix and 1 µL of PCR-certified sterile water and were included to make sure that the PCRs were not contaminated with non-target DNA.

Table 2. Adapted reagents/sample and original reagents/sample from Venkatesan (2008)

Reagent	**Adapted**	**Original**
10x Buffer	2.5 µL	2.5 µL
$MgCl_2$	1.75 µL	2.0 µL
dNTPs	0.5 µL	0.5 µL
Cx. tarsalis ISSR primer 2 (10µM)	2.0 µL	2.0 µL
PCR-certified water	16.25 µL	16.0 µL
Taq DNA polymerase	1.0 µL	1.0 µL

Table 3. Venkatesan (2008) thermocycler protocol, ISSR primer 2

Step 1	94°C, 2:00 minutes
Step 2	94°C, 0:30 seconds
Step 3	68°C, 0:30 seconds, -0.7°C per cycle
Step 4	72°C, 1:00 minutes
Step 5	Go to step 2, 12X
Step 6	94°C, 0:30 seconds
Step 7	55°C, 0:30 seconds
Step 8	72°C, 1:00 minutes
Step 9	Go to step 6, 35X
Step 10	72°C, 10:00 minutes
Step 11	Infinite hold at 8°C

The ISSR primer (**Table 4**) was ordered at a concentration of 100 µM and was diluted to 10 µM for use in PCR. Samples were visualized using gel electrophoresis. Each gel was made using 120 mL of 1x TBE buffer and 3 g of molecular grade agarose to create a 2.5% gel. These components were added together and heated until the agarose had completely dissolved in the buffer solution. Five µL of SYBR safe stain was then added to the melted agarose gel before being poured and allowed to cool. Five µL of blue loading dye was then added to each 25 µL sample before pipetting 25 µL of sample plus dye into each well of the solidified gel. The gels were run in 1x TBE buffer at 120 V for approximately 70 minutes. Each gel was then imaged via UV radiation in a gel hood. Pictures were saved for future analysis of banding patterns.

Each gel image (**Figure 2**) was then analyzed to determine the identity of each allele that was amplified by the ISSR primer in each individual. Allele identity was determined using the size of each band present, measured in kilobases (kb). Each *Cx. tarsalis* individual was then characterized by the presence or absence of a band at every allele location. These alleles are inherited as dominant markers, and therefore heterozygotes could not be determined.

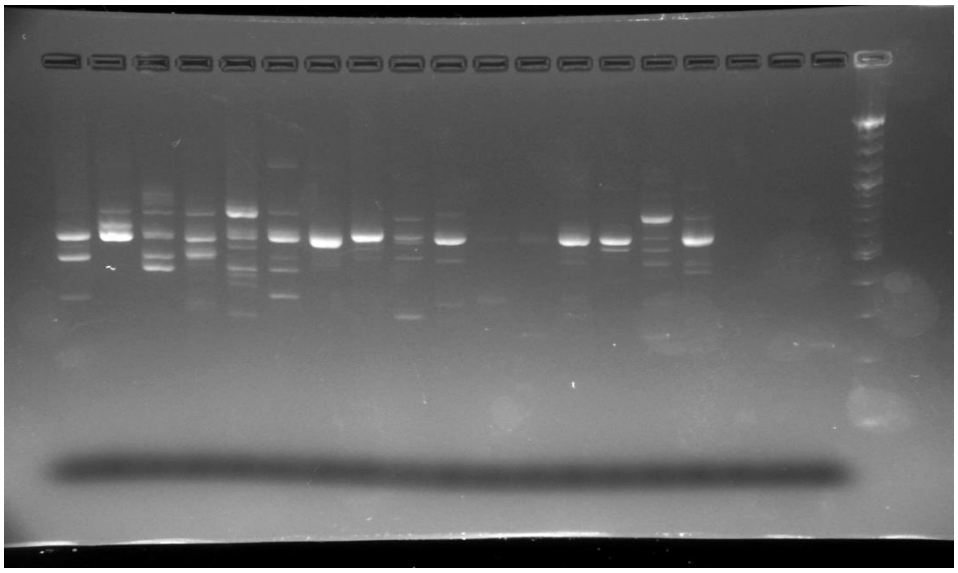

Figure 2. Agarose gel showing the banding patterns of each individual. Each well represents one individual and each band represents an amplified allele. 3 kb ladder shown on right side of gel.

Table 4. Primer sequence.	
Primer	Sequence
CTissr2	5' ACACACACACACACTG 3'

The allele assignments for each individual were organized in a matrix in Microsoft Excel. This data was then formatted to be analyzed in Arlequin software version 3.5.2 (Excoffier and Lischer, 2010) using the haplotype analysis function. *Cx. tarsalis* populations were first examined for genetic diversity at the intra-population level using percent polymorphic loci, or the number of polymorphic sites present out of the number of total loci. Populations were then compared to one another and examined for genetic similarity between each population using pairwise genetic distances and AMOVA (Analysis of Molecular Variance; Excoffier et al., 1992). Finally, populations were grouped together depending on which side of the Continental Divide they were collected on. The two groups were then compared using AMOVA to determine genetic diversity across the Continental Divide.

Results

Intra-population genetic diversity, sites tested, and number of individuals per site are displayed in **Table 5**. Percent polymorphic loci generally increased as population size increased, although there were several exceptions. The highest percentage of polymorphic loci, a measure of increasing genetic diversity within a single population, was found at site 699, which had 78% polymorphic loci and 19 individuals (**Table 5**). The lowest percent polymorphic loci was found at site 573, which had 9 individuals (**Table 5**).

Table 5. Sites tested, individuals per site, and percent polymorphic loci.

Site Number	Number of individuals	Number of Polymorphic Loci	% Polymorphic Loci
42	9	11/27	41%
492	6	12/27	44%
510	13	15/27	56%
582	14	18/27	67%
32	6	15/27	56%
459	19	18/27	67%
620	8	15/27	56%
699	19	21/27	78%
700	14	19/27	70%
485	7	10/27	37%
573	9	8/27	30%

All pairwise population F_{ST} values were significant with the exceptions of the 42-485, 42-459, 42-573, 510-492, and 485-573 pairings (**Table 6**). The highest pairwise F_{ST} value was found in the 492-573 pairing and the lowest pairwise F_{ST} value, of the values that were significant, was found in the 510-699 pairing **(Table 6**).

Out of all the genetic diversity found between and among populations, the majority of variation was found within populations. There was a small amount of genetic variation among the populations within each group east or west of the Continental Divide, and there was no variation among groups (**Table 7**). These results show that all of the genetic variation found among the eleven populations tested was due to differences within each population and among populations within each group (**Table 7**). There was no significant genetic structure found between the two groups on opposite sides of the Continental Divide.

Table 6. Population pairwise F_{ST} values. All F_{ST} p-values were significant except for those marked *, which were insignificant.

Site	42	510	485	492	582	32	459	573	620	699	700
42	0.00000										
510	0.21671	0.00000									
485	0.03302*	0.30802	0.00000								
492	0.29388	0.09237*	0.37901	0.00000							
582	0.10656	0.22601	0.19222	0.26757	0.00000						
32	0.17945	0.10098	0.25436	0.12239	0.19902	0.00000					
459	0.04560*	0.13339	0.13913	0.22226	0.11441	0.19497	0.00000				
573	0.00382*	0.35036	-0.02056*	0.43898	0.20422	0.33230	0.12796	0.00000			
620	0.29328	0.23110	0.36575	0.23471	0.22054	0.23453	0.23384	0.39169	0.00000		
699	0.13510	0.05313	0.20071	0.07414	0.15713	0.12185	0.08353	0.23768	0.21039	0.00000	
700	0.07023	0.17617	0.12559	0.16364	0.06978	0.17268	0.13397	0.13303	0.24889	0.06671	0.00000

Table 7. **AMOVA analysis testing for genetic differentiation between populations of *Cx. tarsalis* in Montana.**

Source of Variation	Variation (%)	F_{ST}	p-value
Among populations within Eastern Montana and Western Montana	-2.79	-	-
Among populations regardless of orientation with respect to the Continental Divide	18.32	-	-
Within individual populations in Montana	84.47	0.15527	0.00000

Discussion

This study indicates that there is a high degree of genetic differentiation between individual populations of *Cx. tarsalis* in Montana. All but five of the F_{ST} pairings had p-values that were below the significance level of 0.05, which indicates that populations within the state are genetically distinguishable from one another. This is in contrast to the findings of Fiocchi (2016), who found that populations of *Cx. tarsalis* in the state were similar to the point of being identical. While the present research using ISSR analysis indicates that populations within the state are distinguishable, it also suggests that there is no barrier to gene flow from the western side of the state to the eastern side. AMOVA analysis of the source of genetic variation found that 100% of the variation was found either within each individual population or between populations regardless of orientation with respect to the Continental Divide. There was no variation found between groups (East and West) with respect to the Continental Divide. This refutes the hypothesis that populations of *Cx. tarsalis* from the eastern side of the Continental Divide will be genetically distinguishable from populations on the western side of the Continental Divide. These findings are in opposition to previous results from Ventakesan (2010) that indicate a barrier to gene flow between the eastern Rocky Mountains and the Great Plains

plateau and a study done by Gimnig et al. (1999) that found that the Tehachapi Mountains in California created a barrier to gene flow among local populations of *Cx. tarsalis*. A total of 12 populations were analyzed by Gimnig et al. (1999), a figure comparable to the 11 populations analyzed in this study.

The fact that previous studies using ISSR analysis (Venkatesan 2008; Venkatesan 2010; Gimnig et al., 1999) successfully showed differentiation across montane regions suggests that further analysis may be needed. The present study analyzed *Cx. tarsalis* populations using only one of the primers used by Venkatesan (2008), CTissr2. Venkatesan (2008) identified ten ISSR primer sequences that successfully identified genetic differentiation, and future analysis using a combination of several of these ISSR primers, rather than just one, may reveal more genetic structure when analyzing *Cx. tarsalis* populations within the state. A greater number of individuals and a more diverse set of sites may also be needed. The present study analyzed 124 individuals from 11 different sites, three of which were located west of the Continental Divide. All of these sites showed a high degree of within population genetic variability. It may be that the genetic variation between individuals is high enough that genetic structure is not evident at the statewide level.

Additionally, it may be possible that the Continental Divide is not a significant enough barrier to gene flow in the state of Montana. Previous studies (Gimning et al., 1999; Eisen et al., 2008) have found that *Cx. tarsalis* populations decrease significantly in number past elevations of 1600 m. Marias Pass near Glacier National Park passes over the Continental Divide at an elevation of 1589 m. Of all the passes over the Continental Divide in Montana, Marias Pass is the only one below 1600 m. Several other passes along the Continental Divide that have elevations close to 1600 m include Rogers Pass (1710 m) near Lincoln, MT, MacDonald Pass

(1924 m) near Helena, MT, and Homestake Pass (1929 m) near Butte, MT. It may be possible for *Cx. tarsalis* to traverse these relatively low mountain passes and bypass any type of barrier to gene flow created by montane habitats.

In summary, individual *Cx. tarsalis* populations in Montana demonstrate a high degree of genetic variability, meaning that populations are genetically distinguishable from one another. When populations are grouped together and examined for genetic differentiation across the Continental Divide, the two groups appear genetically similar. This suggests that a barrier to *Cx. tarsalis* gene flow does not exist within the state of Montana.

Acknowledgements

I would like to thank my thesis advisor, Dr. Jennifer Glowienka, for guiding me through this research. I would like to thank Drs. Grant Hokit, Gerald Shields, and Kyle Strode for their assistance in the writing of this paper. The research reported in this study was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM103474, awarded to Dr. Grant Hokit. The contents of this paper are solely the responsibility of the author and do not represent the views of the National Institutes of Health.

Literature Cited

Abbot, P. 2001. "Individual and Population Variation in Invertebrates Revealed by Inter-simple Sequence Repeats (ISSRs)." *Journal of Insect Science* 1.8 (2001):n. pag. Web.

Barker CM, Bolling BG, Black WC, Moore CG, Eisen L. 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.

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.

Centers for Disease Control and Prevention West Nile virus activity–United States, September 5–11, and Texas, United States, January 1–September 9, 2002. MMWR Morb Mortal Wkly Rep. 2002;51:812–23

Davis CT. 2004. Genetic variation among temporally and geographically distinct West Nile virus isolates, United States, 2001, 2002, (vol 9, pg 1423, 2003). Emerging Infectious Diseases. 10(1):160-160.

Eisen L, Bolling BG, Blair CD, Beaty BJ, Moore CG. 2008. Mosquito species richness, composition, and abundance along habitat-climate-elevation gradients in the northern Colorado front range. Journal of Medical Entomology. 45(4):800-811.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among dna haplotypes - application to human mitochondrial-dna restriction data. Genetics. 131(2):479-491.

Excoffier, L. and H.E. L. Lischer. 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

Fiocchi, J. "An Evaluation of Population Differentiation in *Culex tarsalis* in Montana Using *COI* Sequences." Thesis. Carroll College. (2016). Print.

Gimnig JE, Reisen WK, Eldridge BF, Nixon KC, Schutz SJ. 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.

Goddard LB, Roth AE, Reisen WK, Scott TW. 2002. Vector competence of California mosquitoes for west nile virus. Emerging Infectious Diseases. 8(12):1385-1391.

Goddard LB, Roth AE, Reisen WK, Scott TW. 2003. Vertical transmission of West Nile virus by three California Culex (diptera : Culicidae) species. Journal of Medical Entomology. 40(6):743-746.

Hayes CG. 2001. West Nile virus: Uganda, 1937, to New York City, 1999. West Nile Virus: Detection, Surveillance, and Control. 951:25-37.

Hokit G, Alvey S, Geiger JMO, Johnson GD, Rolston MG, Kinsey DT, Bear NT. 2013. Using undergraduate researchers to build vector and West Nile virus surveillance capacity. International Journal of Environmental Research and Public Health. 10(8):3192-3202.

Venkatesan M, Rasgon JL. 2010. Population genetic data suggest a role for mosquito-mediated dispersal of west nile virus across the western united states. Molecular Ecology. 19(8):1573-1584.

Venkatesan, M. 2008. Population and Quantitative Genetics in the West Nile Virus Vector *Culex tarsalis*. Diss. Johns Hopkins U, 2008. Ann Arbor, MI: ProQuest LLC. (2009). Print.

Venkatesan M, Westbrook CJ, Hauer MC, Rasgon JL. 2007. Evidence for a population expansion in the West Nile virus vector Culex tarsalis. Mol Biol Evol. 24(5):1208-1218.