UNIVERSITY OF CENTRAL OKLAHOMA
Edmond, Oklahoma
Jackson College of Graduate Studies

Genetic variation at major histocompatibility complex loci in *Neotoma albigula* (white-throated woodrat): Potential clues to the evolution of North American arenaviruses

A THESIS SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements

For the degree of
MASTER OF SCIENCE IN BIOLOGY

By
Sarah G. Hoss
Edmond, Oklahoma
2012

Genetic variation at major histocompatibility complex loci in *Neotoma albigula* (white-throated woodrat): Potential clues to the evolution of North American arenaviruses

A THESIS
APPROVED FOR THE DEPARTMENT OF
BIOLOGY

April 2012

By _____

Michelle Haynie, PhD
Committee Chairperson

_____

Jenna Hellack, PhD
Committee Member

_____

Robert Brennan, PhD
Committee Member

Acknowledgements

I would first like to thank my advisor, Dr. Michelle Haynie, my mentor and my friend. Thank you for fostering my talents, encouraging my dreams, and teaching me how to be a scientist. Next, I want to thank my committee members, Dr. Jenna Hellack and Dr. Bob Brennan, for their incredible support and guidance throughout this entire process. Thank you to Dr. Charles Fulhorst for providing the viral data and to the Museum of Texas Tech University for providing the tissue samples. Thank you to the UCO Office of Research and Grants for providing the funding for this research. I would also like to thank the following people who expanded my knowledge and gladly offered their assistance: Dr. Bill Caire, Dr. Martha Burford, Dr. Mel Vaughan, James Creecy, Linda Luna, Sharon LaFave, and Will Unsell. A special thanks goes out to my labmates, Amanda Eisemann, Shey Ramsey, Ethan Rowell, and Winnifred Pipkin. Finally, I want to thank my family, especially Susan and Ben Gill, Patricia and Greg Hoss, and Nancy and Vern Lowrey, for their incredible support and celebration in my small victories along the way. And, to my wonderful husband, Paul, who made this journey possible and remained by my side to its completion.

Table of Contents

List of Figures

Figure Page

List of Tables

Table Page

ABSTRACT OF THESIS
University of Central Oklahoma
Edmond, Oklahoma

NAME: Sarah G. Hoss

TITLE OF THESIS: Genetic variation at major histocompatibility complex loci in *Neotoma albigula* (white-throated woodrat): Potential clues to the evolution of North American arenaviruses

DIRECTOR OF THESIS: Dr. Michelle Haynie

PAGES: 70

ABSTRACT:

The major histocompatibility complex (*Mhc*) is an important component of vertebrate immune systems, and as such, genetic analysis of *Mhc* loci have been shown to provide information on susceptibility to viral infection. *Neotoma albigula* (white-throated woodrat) are naturally associated with at least 3 arenaviruses, and are considered the putative host species for all 3 viruses. Strains of arenaviruses can cause severe febrile diseases in humans and are considered a serious health risk where human-rodent interactions are common. The association of *N. albigula* with at least 3 viruses suggests a significant co-evolutionary history between the host and virus. In this study, *Mhc* loci were screened to detect genetic variation among *N. albigula* populations in Arizona. This was done to test the prediction that

levels of genetic variation at *Mhc* loci in *N. albigula* will correlate with specific arenaviral strains. The results have indicated moderate genetic variation among groups but almost no correlation with viral data.

CHAPTER 1
INTRODUCTION

Immunity: Immunity is necessary for organisms to resist or fight encounters with pathogens, such as bacteria and viruses, which cause disease. Host response to pathogens can be highly variable (Oldstone 1975). Immune responses must be able to distinguish "self" (cells and tissues that make up the organism) from "nonself" (foreign material that is not part of the organism). Any foreign material that elicits an immune response is known as an antigen. Antibodies, also known as immunoglobulins, are Y-shaped glycoproteins found on B-cells in serum and in mucosal secretions that recognize and bind to a specific portion of an antigen, called an epitope (Willey et al. 2009).

There are 2 types of immunity, innate and adaptive. Innate immunity is the "first line of defense" against pathogenic invasion. It involves physical barriers, such as skin and mucus membranes, and other general mechanisms that act upon immediate exposure. In innate immunity, each encounter with a pathogen is treated the same, regardless of the type of pathogen or extent of exposure. Leukocytes, specifically granulocytes, are involved in nonspecific immune responses. These responses include inflammation caused by allergies and infections triggered by parasitic worms and fungi. They also actively engulf bacteria via phagocytosis (Willey et al. 2009).

Adaptive immunity complements and intensifies the innate immune response. Adaptive immunity involves specialized cells that discriminate between self and nonself as well as neutralize foreign particles and infected cells. An important component of this system is the ability to recognize specific antigens upon a re-encounter of pathogens and other foreign materials and respond accordingly. The adaptive immune system either destroys the foreign invaders or renders them harmless so they cannot cause disease.

Lymphocytes are the important cells of the adaptive immune system. Examples of lymphocytes include B cells, involved in antibody-mediated immunity, and T cells, involved in cell-mediated immunity. Monocytes are cells that "bridge the gap" between the innate and adaptive immune systems. Monocytes include macrophages, the largest phagocytes that engulf infected and foreign cells, and dendritic cells, which process and present antigens to lymphocytes (Willey et al. 2009).

Adaptive immunity is conferred either naturally or artificially. Natural immunity is a result of normal life experience and is acquired actively, as a physiological response to infection, or passively, as an infant receives antibodies through his mother's breast milk. Artificial immunity is a result of medical treatment, specifically applied to humans or through human intervention by vaccination. Vaccinations allow the body to actively respond to a pathogen in small amounts, as an inactive form, or of reduced virulence. Artificial immunity is passively conferred through such treatments as plasma therapy, in which the patient receives antibodies that were made by another person or animal (Willey et al. 2009).

Adaptive immunity is either antibody-mediated or cell-mediated. Antibodies are found within body fluids or on B cells. The presence or absence of antibodies found in the blood serum of an individual is known as serostatus. In antibody-mediated (humoral) immunity, the antigens bound by antibodies are neutralized or "tagged" for later destruction by other cells. Cell-mediated immunity is directed by T cells that attack infected cells, causing them to break down or recruiting other cell-directed responses, such as phagocytosis. T cells have surface proteins, known as cluster of differentiation (CD) molecules, which play a role in intercellular communication (Willey et al. 2009). Examples include CD8 and CD4,

both important in antigen recognition. $CD4^+$ T cells are necessary in order for $CD8^+$ T cells and antibody-mediated immunity to be effective (Kotturi et al. 2010).

***Major histocompatibility complex*:** The major histocompatibility complex (*Mhc*) plays an important role in adaptive immunity in vertebrates (Klein 1987). In humans, it is found on the short arm of Chromosome 6 and is referred to as the human leukocyte antigen (HLA) region (Fernando et al. 2008). It is located on chromosome 17 in the house mouse, *Mus domesticus* (Meagher and Potts 1997). Historically, the *Mhc* extends 3.6 megabases and has been divided into three subregions: telomeric class I genes, centromeric class II genes, and class III genes (Fernando et al. 2008). The *Mhc* codes for cell surface glycoproteins. These glycoproteins bind to antigens derived from pathogens and present them for T cell recognition and appropriate immune response (Sommer 2005).

Mhc class I genes are expressed in all nucleated body cells. The *Mhc* class I molecules are found on the surface of these cells and identify them as "self." These molecules are important in identifying potential donors for organ transplants. *Mhc* class I molecules also bind to "foreign" peptide fragments that result from digested proteins in the cytoplasm. These proteins originate from intracellular pathogens (Klein 1986) or carcinogenic mutations. The peptide fragments are sent to the endoplasmic reticulum for processing and then associate with the newly synthesized *Mhc* class I molecule. Both the peptide and molecule are anchored in the plasma membrane. The peptide, now considered an antigen, is presented to a group of T cells, specifically $CD8^+$ or cytotoxic T cells. These T cells have a specific receptor for *Mhc* class I molecules and the presentation of the antigen activates them to kill the infected cell by inducing apoptosis (programmed cell-death).

Mhc class II genes are expressed only in T cells, B cells, macrophages, and dendritic cells. The *Mhc* class II molecules are found on the surface of these cells and are ultimately used to communicate with $CD4^+$ T cells to elicit an immune response involving the other cells. The *Mhc* class II molecules bind to peptide fragments digested from extracellular pathogens (Klein 1986), such as bacteria, viruses, and toxins, which were brought into the cell via endocytosis. Dengjel et al. (2005) concluded that *Mhc* class II molecule presentation of these antigens was derived from autophagy, the digestion mechanism involving a cell's own components. The peptide fragments associate with a pre-existing *Mhc* class II molecule and are sent to the plasma membrane. The antigen is then presented to $CD4^+$ T cells. The immune response that follows involves an increase in the number of $CD4^+$ T cells and the development of memory T cells which can react to the same antigen at a later time. For example, Janssen et al. (2003) found $CD8^+$ T cells primed in the presence of $CD4^+$ T cells were able to respond efficiently upon re-encountering an antigen. In addition, the recruitment of other immune cells occurs to neutralize the pathogen or increase other immune responses, such as inflammation.

Mhc class III genes code for secreted proteins that have immune functions. However, they are not used to discriminate between self and nonself, nor are they related to class I or class II molecules. Therefore, they are not essential to this research.

Arenavirus: Arenaviruses (family Arenaviridae) are ambisense, negative, single-stranded RNA viruses. They have a genome composed of 2 segments enclosed by a lipid envelope. They are spherical and have an average diameter of 110-130 nm. The name arenavirus is derived from the Latin root "arena" meaning "sandy," referring to the presence of nonfunctional ribosomes packaged within the finished virus particle when viewed in cross-

section (Figure 1). Their 11 kb genome consists of small (S) and large (L) segments, both of which contain positive and negative strands (Figure 2). The S RNA segment codes for two viral glycoproteins and a nucleoprotein. The L RNA segment encodes the RNA-dependent RNA polymerase needed for transcription.

Arenaviruses are Class V viruses (Baltimore Classification System). They have a non-lytic life cycle and cause both chronic and acute infections. In chronic infections, on-going viral replication and continual immune responses occur whereas in acute infections, the imbalance between viral replication and immune responses either eliminates the virus or the host dies (Oldstone 1975). Arenaviruses replicate in the host nucleus but do not incorporate into the host genome. In order to persist in the host, they continuously replicate and fully express viral proteins. They enter the host organism via inhalation or ingestion of infected secretion droplets. Once inhaled, arenaviruses have been known to enter the lymph tissues and spread throughout the body (Peters 2002). During the first 2 days of incubation, the virus attaches to the epithelium of mucus membranes. The virus enters the host cell via receptor-mediated endocytosis, causing the host cell's plasma membrane to bulge inward. Once inside the cell, the virus is contained within a vesicle known as an endosome. A hydrogen ion pump mechanism located within endosomes acidifies its internal environment (Galloway et al. 1983). The acidification of the endosome is the first step in degrading whatever materials are brought into the cell via endocytosis. The decreased pH within the endosome activates the release of the viral nucleocapsid into the host cell cytoplasm. The nucleocapsid consists of the ambisense RNA genome and the RNA-dependent RNA polymerase. The positive strands of the RNA segments are translated first. The negative strands of the RNA segments are used as a template to create positive strands, which are then

translated. These positive-strand RNA molecules are then used as a template to replicate the viral genome. The newly synthesized viruses acquire their envelope by exiting the host cell via budding.

The genus Arenavirus comprises 2 serocomplexes, based on their antigenic properties (Salvato et al. 2005), Lymphocytic Choriomeningitis-Lassa (Old World) and Tacaribe (New World). The Tacaribe serocomplex currently includes these North American viruses: Bear Canyon virus, Big Brushy Tank virus, Catarina virus, Skinner Tank virus, Tamiami virus, Tonto Creek virus, Whitewater Arroyo virus, Pichinde virus, and Real de Catorce virus (Table 1; Cajimat et al. 2011; Milazzo et al. 2008). Pichinde virus is the only Tacaribe complex virus that does not cause diseases in humans (Liang et al. 2009). Charrel and de Lamballerie (2009) suggest these viruses are derived from a common ancestor due to the recombinant nature of the S RNA segment. The segmented genome allows for reassortment between 2 viruses that co-infect a host cell. The combinations of L/S genomes can be used to trace lineages of phenotypes and origins of the RNA segments (de la Torre 2008).

Human immune responses to arenaviruses: Human infection by or with arenaviruses usually occurs by contact with infected rodents or with infectious rodent excreta (Bowen et al. 1996), or inhalation of infectious aerosols (Carballal et al. 1988). Contact with infected rodents is common in areas in which human populations are closely associated with rodent populations. In Venezuela, work on Guanarito virus suggests that human infection rates may be higher in areas of high human population density, such as in Guanare and Guanarito, than in outlying rural areas (Weaver et al. 2000). In Mexico, human consumption of woodrats is common in rural highland regions (Inizan et al. 2010), possibly leading to incidences of human viral infections.

Some arenaviruses, such as Junin, Machupo, and Guanarito viruses, can cause severe hemorrhagic fever in humans (Peters 2002). Symptoms of hemorrhagic fever illnesses include fever, fatigue, dizziness, abdominal pain, and muscle aches. In severe cases, patients may have internal bleeding, nervous system malfunction, shock, coma, and convulsions. Treatment for some hemorrhagic illnesses involves plasma transfusions and anti-viral drugs, such as Ribavirin (Enria and Maiztegui 1994). A vaccine has been developed for Argentine hemorrhagic fever, caused by Junin virus (Maiztegui et al. 1998), but not yet for other arenaviruses. Hemorrhagic fevers are rare in North America; however, Whitewater Arroyo virus may be responsible for 3 fatal cases of infection in California (CDC 2000). Due to the pathogenicity of arenaviruses as well as the lack of vaccines and antiviral treatments available, these viruses have been identified as Class A pathogens and are considered a possible bioterrorism threat (Kotturi et al. 2010).

To fight arenavirus infections in humans, an understanding of immune response mechanisms must be considered. Antibody-mediated immunity plays an important role in responding to New World arenavirus infections (Kotturi et al. 2010). Janssen et al. (2003) injected mice with lymphocytic choriomeningitis virus (LCMV) and monitored the corresponding immune response. Mice that had an absence of $CD4^+$ T cells upon initial infection developed a depleted number of $CD8^+$ T cells, and therefore a reduced overall immune response. Maiztegui et al. (1979) used immune plasma to treat patients infected with Argentine hemorrhagic fever in a double blind trial. They found that immune plasma treatment significantly reduces mortality when given in the first 8 days of infection and this is currently the only treatment for patients infected with this disease. T cell responses may also work in conjunction with antibody-mediated responses to arenaviruses (Kotturi et al.

2010). Argentine hemorrhagic fever patients had a decrease in the number of circulating T cells (Vallejos et al. 1989) and an indication of lymphoid necrosis in fatal cases. This provides evidence that infection with this virus causes some immunosuppressive activity (Peters 2002). In another study, neutralizing antibodies were found in rhesus macaques several weeks after a live-attenuated vaccination was administered to combat Argentine hemorrhagic fever (McKee et al. 1993). This delay in the detection of antibodies could be the result of T cell involvement in initial infection, following an expanded immune response.

Using knowledge of the immune response to arenaviruses, Kotturi et al. (2010) made the initial step for the development of a general arenavirus vaccine. They identified arenavirus epitopes associated with specific $CD4^+$ T cells in the hope of creating a multivalent vaccine that would protect against several strains of arenaviruses. Their initial results indicated that a general vaccine could be developed as several identified epitopes from multiple arenaviruses had the capacity to bind to multiple *Mhc* alleles.

Neotoma albigula: *Neotoma albigula,* the white-throated woodrat (Figure 3), is a member of the Cricetidae family. It is a medium-sized rodent with a brown dorsum, and hairs on the throat and pectoral areas are white to the base (Macêdo and Mares 1988). Edwards et al. (2001) divided *N. albigula* into 2 distinct species: *N. leucodon* (white-toothed woodrat) and *N. albigula*. They determined the geographic range of *N. leucodon* to include areas east of the Rio Grande in the southwestern United States and Rio Conchos in central Mexico (Figure 4). The geographic range of *N. albigula* includes arid regions west of the Rio Grande and the Rio Conchos (Edwards et al. 2001; Macêdo and Mares 1988). White-throated woodrats are found in a wide variety of habitats in Arizona and are abundant in areas associated with prickly-pear cactus (*Opuntia*) and cholla (*Opuntia*), and in pinyon-juniper

woodlands (Hoffmeister 1986; Vorhies and Taylor 1940). They construct houses, known as middens, in and around yucca, cacti, and rocks that are made of available material, typically sticks, cow chips, bones, discarded cans, and plant material (Hoffmeister 1986). The middens have an average diameter of 8 feet and are located mostly above ground. Nests within the middens are located below ground and have an average diameter of 8 inches (Vorhies and Taylor 1940). The nests serve as both a daytime retreat and a place to raise young (Hoffmeister 1986). Tissue samples used for this research were collected from *N. albigula* in Arizona inhabiting the following biomes: juniper-pinyon woodland, montane conifer forest, Sonoran Desert scrub - Arizona upland, Mohave Desert scrub, semi-desert scrub grassland, juniper-pinyon chaparral woodland, Sonoran Desert scrub - lower Colorado, and a citrus orchard (Abbott et al. 2004). Notable flora found at each biome is summarized in Table 2.

Neotoma albigula is primarily nocturnal and active through most of the year (Macêdo and Mares 1988). Breeding occurs throughout the year, but slows in late summer and early fall (Hoffmeister 1986). They are not social beyond mating and rearing of young (Feldman 1935), and adults typically show aggression, through fighting and foot-thumbing, towards each other when placed in close proximity. However, some adults have been noted to live together amicably in captivity, perhaps indicating similar behavior in the field (Feldman 1935). Females have 1 to 3 young per litter and can have multiple litters per breeding season (Baker 1956). Gestation in these woodrats lasts 37 or 38 days. Females lactate 20 to 25 days after parturition and weaning occurs 62 to 72 days after birth. Young become independent of their mother by the time she gives birth to the next litter (Richardson 1943).

***Arenavirus/Woodrat system*:** Arenaviruses are associated with rodents and some rodent-transmitted diseases in humans, and are adapted to specific rodent hosts (Mims 1981), including woodrats. In 1996, Fulhorst et al. (1996) isolated an arenavirus from *N. albigula* captured at Whitewater Arroyo, New Mexico. Genetic analysis of this arenavirus showed it was closely related to Tamiami virus, a Tacaribe serocomplex arenavirus associated with *Sigmodon hispidus* (hispid cotton rat; Calisher et al. 1970), also of the rodent family Cricetidae. Since this time, several other Tacaribe serocomplex arenaviruses have been identified and associated with *Neotoma* species inhabiting Texas, Oklahoma, Colorado, Utah, New Mexico, Arizona, California, and Mexico. A list of infected species by state is provided in Table 1. Kosoy et al. (1996) found antibodies to either Tamiami virus or Pichinde virus in 5 *Neotoma* species inhabiting Arizona, Colorado, New Mexico, California, and Utah; Fulhorst et al. (2001a) recovered 23 viral isolates from 11 *Neotoma* species taken from New Mexico, Oklahoma, Texas, and Utah; Calisher et al. (2001) reported Whitewater Arroyo virus prevalence in 3 species of *Neotoma* in southeastern Colorado; 2 *Neotoma* species in western Oklahoma also tested positive for Whitewater Arroyo virus (Nisbett et al. 2001); Abbott et al. (2004) detected arenavirus antibodies in 118 woodrats from 3 species in Arizona; and Milazzo et al. (2010) found 3 antibody-positive *Neotoma* species located in New Mexico, Texas, and Mexico.

Calisher et al. (2001) suggested that arenavirus infections within a white-throated woodrat population are maintained by long-lived, persistently infected adults. They also suggested that infection does not affect the behavior of the woodrats. Depending on the arenavirus, the mode of transmission, and age at the time of infection, the host may be persistently infected regardless of the presence or absence of antibodies (Charrel and de

Lamballerie 2009). A study of arenavirus infections in mice indicated that individuals persistently infected with LCMV showed immune tolerance to the virus (Mims 1981).

Transmission of arenaviruses among woodrats can occur horizontally through rodent-rodent encounters. Aggressive behaviors, including fighting and biting, when woodrats are in close proximity may lead to viral transmission. Agnostic encounters in times of high population density may also result in increased viral transmission (Calisher et al. 2001).

Viral transmission can also occur vertically with mothers passing the virus to their offspring in utero. Arenaviruses can pass easily into the uterus of pregnant rodent hosts (Peters 2002) and the mother can pass the infection on to the fetus via the placenta. Neonatal infection leads to continual and harmless viral carriage. In the early development of young woodrats, the immune system may recognize and tolerate foreign material as "self," resulting in maintenance of that infection throughout life (Mims 1981). Juveniles may also have delayed susceptibility due to protection by maternal antibodies or due to innate behaviors that keep them from adult contact (Calisher et al. 2001). An immune tolerant animal has circulating antibodies that react to a pathogen but do not neutralize the pathogen. If the same pathogen infects an adult with a fully functional immune system, rapid death can result due to the nature of the immune-mediated disease (Mims 1981). Fulhorst et al. (2001b) inoculated healthy woodrats with prepared Whitewater Arroyo viral stock. The inoculated individuals representing different age groups (newborn, juvenile, and adult) and the duration of infection as well as the persistence of viral shedding was determined. They found chronic infection persisted in woodrats inoculated shortly after birth and the newborn woodrats were able to transmit the infection horizontally to their mother. Fulhorst et al. (2001b) also found that viral shedding in infected newborn woodrats was persistent, whereas viral shedding in

infected adults was transient. In a natural population, Abbott et al. (2004) found a lack of association between antibody prevalence and the size, gender, and/or wounding of woodrats. This suggests that white-throated woodrats are infected at a young age, and that vertical transmission is important in arenaviral transmission.

***Infectious disease and genetic diversity*:** An immune response to a pathogen can lead to a co-evolutionary "molecular arms race" favoring the diversification of *Mhc* genes (Potts et al. 1994). The host-virus relationship may be a consequence of a long-term evolutionary affiliation (Cajimat et al. 2007, 2008; Fulhorst et al. 2002a, b; Milazzo et al. 2008). In their phylogenetic analysis of Tacaribe complex arenaviruses, Bowen et al. (1996) found that some viral lineages paralleled the taxonomic evolution of their host rodent. However, they did not find a correlation between the arenavirus phylogeny and the geographic distribution of viral strains. Weaver et al. (2000) found significantly higher nucleotide variation in Gaunarito virus isolates taken from rodent hosts as compared to those isolates taken from infected humans. They recommended studying the population genetics of the rodent hosts to determine if certain genotypes coevolved with certain host lineages.

The genetic diversity of *Mhc* genes may result from pathogen-driven selection, such as caused by infection with arenaviral strains. The *Mhc* contains multiple alleles due to high mutation rates and recombination. The *Mhc* in the human genome is a prime example of linkage disequilibrium, a nonrandom association between genotypes of 2 or more different loci. Selection on multilocus genotypes can create linkage disequilibrium. Some selection mechanisms, including kin recognition and mating preferences, could favor high levels of variation and an increase in the number of rare alleles within the *Mhc* (Gruen and Weissman 1997). Other selective pressures involving infectious disease can alter allelic frequencies of

loci that offer resistance to the disease, as well as those linked polymorphic loci that may not be involved in disease resistance (O'Brien and Evermann 1988).

High levels of variation at *Mhc* loci convey an adaptive advantage, resulting from an increased ability to resist disease (O'Brien and Evermann 1988; Richman et al. 2001, 2003). Viruses evolve faster than their hosts (Gojobori and Yokoyama 1987) and those that bud from cell surfaces, such as arenaviruses, may incorporate *Mhc* antigens on their surface coat (Oldstone 1975). According to O'Brien and Evermann (1988), some viruses can evade host immune responses by interfering with the expression of *Mhc* molecules. Infected cells which do not present antigens correctly are not targeted by T cells and are therefore not neutralized. Host populations with extremely polymorphic *Mhc* loci would be able to counter the viral mechanism of *Mhc* molecule interference by displaying a wide variety of genetic responses (O'Brien and Evermann 1988). Oldstone (1975) studied the extent of LCMV infection among *Mhc* recombinant mice. He found susceptibility of LCMV greatest among individuals with a certain allele, providing evidence that immune responses to this virus were genetically controlled. On the other hand, lack of variation at Mhc loci may lead to an increase in disease susceptibility (Hughes 1991; O'Brien and Evermann 1988).

Trans-species polymorphism has played a role in *Mhc* evolution, making *Mhc* useful in describing phylogenetic relationships among related species (Klein et al. 1998). Among vertebrates, *Mhc* class II genes are among the most polymorphic coding loci (Klein and Figuero 1986) and can be used to screen for genetic variation within and among populations. Pfau et al. (1999) used a *Mhc* class II gene to screen for genetic diversity among populations of *S. hispidus*. They found a great deal of polymorphism at the *Mhc DQ-A* exon 2 locus (Crew and Bates 1996). Van Den Bussche et al. (2002) examined genetic variation at a *Mhc*

class II locus in populations of white-tailed deer, *Odocoileus virginianus*, inhabiting Iowa, New York, Oklahoma, and Tennessee. They identified high gene and nucleotide diversity at the locus within and among the populations. Hedrick et al. (2000) screened a *Mhc* class II gene for genetic variation in a small population of Arabian oryx, *Oryx leucoryx*. The species experienced a bottleneck in the 1960s and the researchers wanted to determine how genetic variation related to disease susceptibility. They found low genetic variation overall but were able to identify 3 divergent alleles in the *Mhc* class II gene (Hedrick et al. 2000).

Overviews and Objectives: In this study, I am screening *Mhc* class II loci in *N. albigula* for genetic variation. Any variation found may result from selective pressures of immune responses to pathogens. The population of *N. albigula* in Arizona is associated notably with Whitewater Arroyo virus, Big Brushy Tank virus, and Tonto Creek virus (Milazzo et al. 2008). Typically, in host-virus relationships, 1 viral strain is associated with 1 host species. The association of multiple viral strains with 1 host may be significant. The explanation for the association of *N. albigula* with at least 3 arenaviruses lies in the genetic diversity of the arenavirus, the population genetics of the host, or both. Genetic variation of the arenavirus strains is currently under study by researchers at the University of Texas Medical Branch (e.g., Cajimat et al. 2011). I examined the population genetics of the host species, white-throated woodrats. I chose *Mhc* class II loci for this study because of their specific involvement with immune responses to arenaviruses. *Mhc* class II genes code for molecules that bind arenavirus epitopes. These epitopes are recognized by $CD4^+$ T cells, which then prime $CD8^+$ T cells and activate antibody-mediated immune responses. Additionally, *Mhc* class II loci have been shown to be highly polymorphic in rodents of the family Cricetidae (Crew and Bates 1996; Pfau et al. 1999).

In this study, I 1) determinee levels of genetic variation (number of alleles, allele frequency, and heterozygosity) at *Mhc* loci in *N. albigula* populations; 2) determined if *Mhc* markers could be used to detect genetic subunits within *N. albigula* populations; and 3) determined if there was a correlation between viral infection and *Mhc* alleles in *N. albigula* populations.

Hypothesis: I predicted that levels of genetic variation at *Mhc* loci in *N. albigula* were correlated with viral strains, such that specific alleles were associated with different strains of arenaviruses.

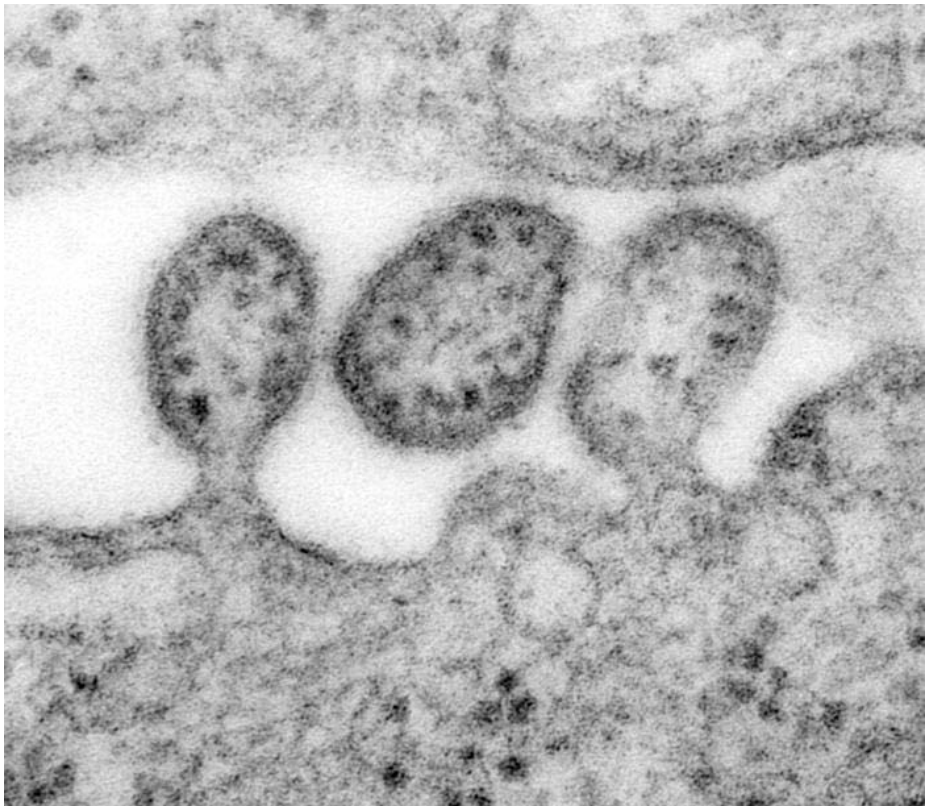

Figure 1. A highly magnified transmission electrograph of arenavirus particles. The dark spots within the viruses are nonfunctional ribosomes, giving them a “sandy” appearance (courtesy of the Center for Disease Control and Prevention: C. S. Goldsmith and D. Auperin, taken from http://phil.cdc.gov/PHIL_Images/8699/8699_lores.jpg.).

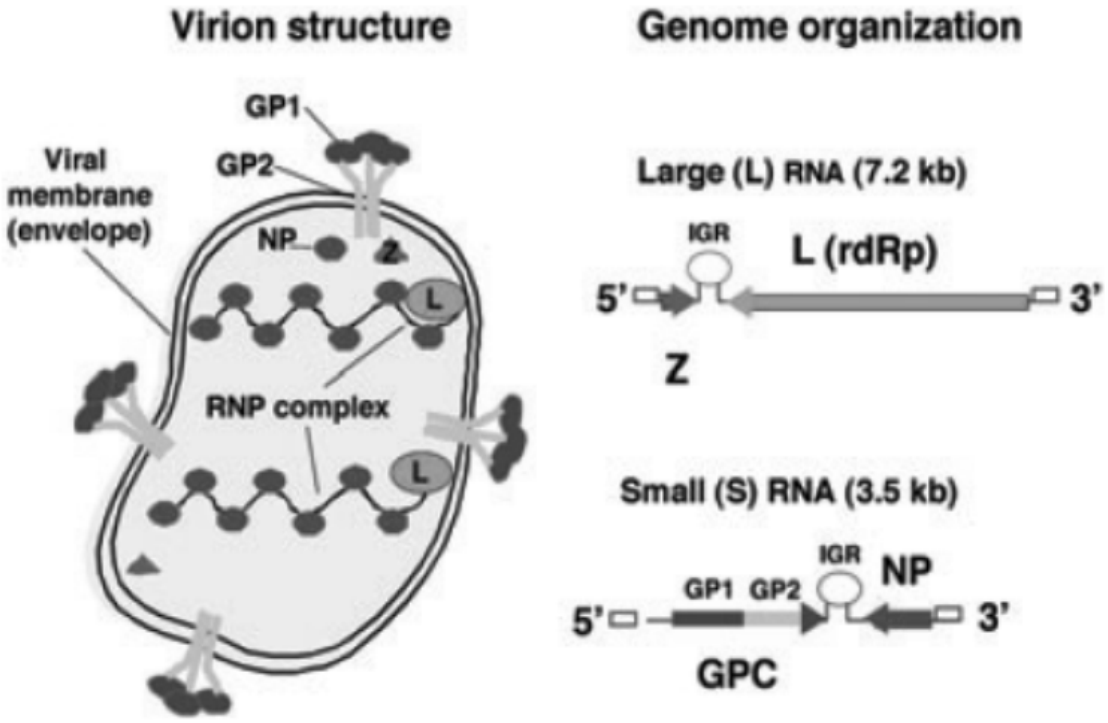

Figure 2. Depiction of an arenavirus (family Arenaviridae; taken from Rojek and Kunz 2008). Arenaviruses are spherical enveloped viruses. Important structures located within the virion include glycoproteins (GP1 and GP2), a nucleoprotein (NP), and a ribonucleoprotein (RNP) complex. The genome is organized into 2 segments, large and small, both with ambisense reading frames. The genome includes a stem-and-loop intergenic region (IGR) on each segment. Coding sequences within the genome include the glycoprotein precursor (GPC) region and the RNA-dependent RNA polymerase (rdRp) region.

Figure 3. *Neotoma albigula* (white-throated woodrat) (Seebyseeing.net).

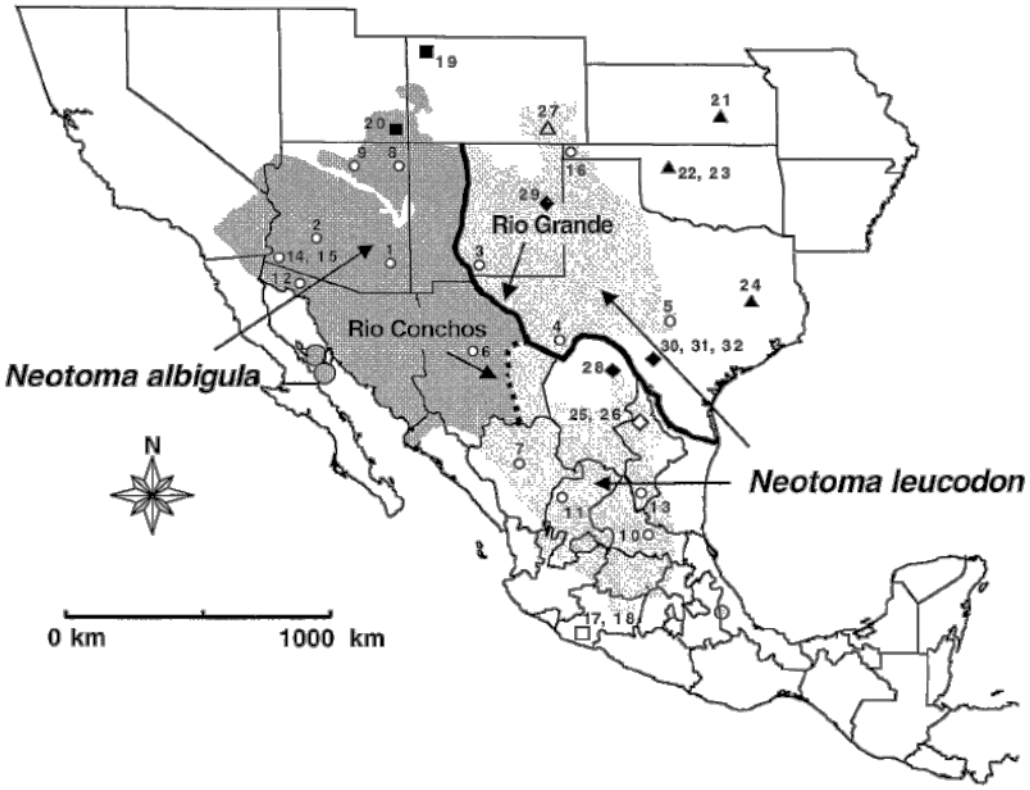

Figure 4. Geographic distribution of *Neotoma albigula* and its sister species *N. leucodon* (modified from Edwards et al. 2001).

Table 1. Woodrat (*Neotoma*) species infected with arenaviruses in the United States and Mexico. References refer to the original description of the virus. Current strain information provided by Cajimat et al. (2011). BCNV=Bear Canyon virus, BBTV=Big Brushy Tank virus, CTNV=Catarina virus, RCTV=Real de Catorce virus, SKTV=Skinner Tank virus, TAMV=Tamiami virus, TTCV=Tonto creek virus, WWAV=Whitewater Arroyo virus.

Species	Virus	Location/County	State	Reference
N. albigula	arenavirus	Cochise	Arizona	Kosoy et al. 1996
N. albigula	arenavirus	Maricopa	Arizona	Kosoy et al. 1996
N. albigula	arenavirus	Yavapai	Arizona	Kosoy et al. 1996
N. albigula	arenavirus	Apache	Arizona	Kosoy et al. 1996
N. albigula	arenavirus	Apache	Arizona	Abbott et al. 2004
N. albigula	arenavirus	Navajo	Arizona	Abbott et al. 2004
N. albigula	arenavirus	Yavapai	Arizona	Abbott et al. 2004
N. albigula	TTCV	Gila	Arizona	Abbott et al. 2004
N. albigula	BBTV	Graham	Arizona	Abbott et al. 2004
N. albigula	arenavirus	Hesperus	Colorado	Kosoy et al. 1996
N. albigula	WWAV	Las Animas	Colorado	Calisher et al. 2001
N. albigula	WWAV	McKinley	New Mexico	Fulhorst et al. 1996
N. albigula	arenavirus	Jones Ranch	New Mexico	Kosoy et al. 1996
N. albigula	arenavirus	Las Cruces	New Mexico	Kosoy et al. 1996
N. albigula	arenavirus	Otero	New Mexico	Kosoy et al. 1996
N. albigula	arenavirus	Socorro	New Mexico	Kosoy et al. 1996
N. albigula	WWAV	McKinley	New Mexico	Fulhorst et al. 2001a
N. albigula	WWAV	Cimarron	Oklahoma	Nisbett et al. 2001
N. cinerea	WWAV	San Juan	Utah	Fulhorst et al. 2001a
N. fuscipes	arenavirus	Ventura	California	Kosoy et al. 1996
N. fuscipes	arenavirus	San Diego	California	Kosoy et al. 1996
N. lepida	arenavirus	San Diego	California	Kosoy et al. 1996
N. leucodon	arenavirus	Nuevo León	Mexico	Milazzo et al. 2010
N. leucodon	RCTV	Catorce	Mexico	Cajimat et al. 2011
N. leucodon	arenavirus	San Luis Potosí	Mexico	Milazzo et al. 2010
N. leucodon	WWAV	Cimarron	Oklahoma	Fulhorst et al. 2001a
N. macrotis	BCNV	Riverside	California	Cajimat et al. 2011
N. macrotis	BCNV	Los Angeles	California	Cajimat et al. 2011
N. mexicana	arenavirus	Apache	Arizona	Kosoy et al. 1996
N. mexicana	arenavirus	Socorro	Arizona	Kosoy et al. 1996
N. mexicana	SKTV	Coconino	Arizona	Abbott et al. 2004
N. mexicana	WWAV	Las Animas	Colorado	Calisher et al. 2001
N. mexicana	WWAV	Larimer	Colorado	Cajimat et al. 2011
N. mexicana	arenavirus	Nuevo León	Mexico	Milazzo et al. 2010
N. mexicana	WWAV	Socorro	New Mexico	Fulhorst et al. 2001a
N. mexicana	arenavirus	Natural Bridges	Utah	Kosoy et al. 1996
N. mexicana	WWAV	San Juan	Utah	Fulhorst et al. 2001a
N. micropus	WWAV	Las Animas	Colorado	Calisher et al. 2001
N. micropus	arenavirus	San Fernando	Mexico	Milazzo et al. 2010
N. micropus	WWAV	Otero	New Mexico	Cajimat et al. 2011
N. micropus	arenavirus	Otero	New Mexico	Milazzo et al. 2010
N. micropus	WWAV	Harmon	Oklahoma	Nisbett et al. 2001
N. micropus	CTNV	Dimmit	Texas	Fulhorst et al. 2001a
N. micropus	CTNV	La Salle	Texas	Fulhorst et al. 2001a
N. micropus	arenavirus	Dickens	Texas	Milazzo et al. 2010
N. micropus	arenavirus	Motley	Texas	Milazzo et al. 2010
N. micropus	arenavirus	Ward	Texas	Milazzo et al. 2010
N. stephensi	arenavirus	Apache	Arizona	Kosoy et al. 1996
N. stephensi	arenavirus	Zuni	New Mexico	Kosoy et al. 1996

Table 2. Arizona biomes from which *Neotoma albigula* samples were taken for this study. Examples of plants found in each biome are included as notable genera.

Arizona biomes	**Notable genera**
Juniper-pinyon woodland	*Juniperus, Pinus*
Montane conifer forest	*Pinus, Quercus, Juglans, Populus*
Sonoran Desert scrub – Arizona upland	*Parkinsonia, Prospopis, Carnegiea, Optuntia*
Mohave Desert scrub	*Larrea*
Semi-desert scrub grassland	*Boutela, Muhlenbergia, Hilaria*
Juniper-pinyon chaparral woodland	*Juniperus, Pinus, Yucca*
Sonoran Desert scrub – lower Colorado	*Larrea, Prosopis, Olneya, Atriplex*
Citrus orchard	*Citrus*

CHAPTER 2
MATERIALS AND METHODS

Sampling: One thousand one hundred ninety-six *Neotoma albigula* individuals were collected from 32 localities throughout Arizona (Figure 5; Abbott et al. 2004). Specimens were euthanized and tissues were obtained, frozen, and deposited in the Natural Sciences Research Laboratory (NSRL), The Museum of Texas Tech, Lubbock, Texas. Voucher specimens were also prepared for all samples. Subsets of these samples have already been analyzed using mtDNA cytochrome *b* sequences (Milazzo et al. 2008) and microsatellite data (Haynie 2006), and are currently being analyzed using mtDNA control region sequences (Haynie et al., unpublished data). A subset of 163 individuals (approximately 5 from each locality, as available) was used in this study to be screened for *Mhc* loci variation (Table 3).

DNA Extraction: Genomic DNA was extracted from *N. albigula* liver samples using a DNeasy tissue extraction kit (Qiagen, Valencia, California). Tissue was cut into small pieces and placed in a 1.5mL microcentrifuge tube. Both 180µL Buffer ATL and 20µL proteinase K were added and mixed by vortexing. The 1.5mL tube was incubated at 56°C overnight. After incubation, the 1.5mL tube was vortexed for 15s and 200µL Buffer AL was added. The tube was vortexed and 200µL 100% ethanol was added. The tube contents were mixed thoroughly by vortexing. The mixture was pipetted into a DNeasy mini spin column in a 2mL collection tube. The spin column was centrifuged at 8000 rpm for 1 min. The flow through and collection tube were discarded. The spin column was placed in a new 2mL collection tube and 500µL Buffer AW1 was added. The spin column was centrifuged at 8000 rpm for 1 min. The flow through and collection tube were discarded. The spin column was placed in a new 2mL collection tube and 500µL Buffer AW2 was added. The spin column was centrifuged at 13200 rpm (the highest setting) for 3 min. The flow through and

collection tube were discarded. The spin column was transferred to a new 1.5mL microcentrifuge tube and 200µL Buffer AE was added for elution. The tube was incubated at room temperature for 1 min. The spin column was centrifuged at 8000 rpm for 1 min. The 1.5mL tube was labeled with: "A," TK number, DNA, *N. albigula*, and the date. The spin column was transferred to a new 1.5mL microcentrifuge tube, 200µL Buffer AE was added, and it was centrifuged at 8000 rpm for 1 min. The 1.5mL tube was labeled with: "B," TK number, DNA, *N. albigula*, and the date. The spin column was discarded and both 1.5mL tubes were stored in the appropriate freezer. The protocol was repeated for all 163 tissue samples. DNA extraction was verified by electrophoresis on a 1% agarose gel. Ethidium bromide was added to visualize the DNA, using a UV light.

DNA Amplification: DNA was amplified using standard polymerase chain reaction (PCR) protocols. PCR is a technique in which copies of specific segments of DNA are exponentially created. Saiki et al. (1985) first applied PCR methods to β-globin gene sequences, coupled with oligomer restriction methods, to aid in the diagnosis of sickle-cell anemia. They found an estimate of 70-100 percent efficiency in the amplification of cloned β-globin sequences. As part of the PCR process, the strands of DNA are first separated at the appropriate melting temperature. Then, the temperature is lowered so the oligonucleotide primers can anneal to the specified segments of the complementary strands in the presence of a DNA polymerase. The temperature is then raised to extend the primers, creating replicated strands. The process of lowering and raising the temperature continues until the desired number of copies is created. Saiki et al. (1985) found that the PCR process amplified the specific sequence 200,000-fold after 20 cycles.

The following fluorescently labeled primers were used for PCR (Table 4): *Peromyscus leucopus* PeleDQAex2-R and PeleDQAex2-F (Crew and Bates 1996); *P. maniculatus* PEB1 and PEB2 (Richman et al. 2001); and *Mus musculus* MusE2Ab1 and MusE2Ab2 (She et al. 1991). All primers were shown to be variable in the species from which they were isolated. Pfau et al. (2001) successfully amplified the *Mhc* exon 2 locus of *S. hispidus* using the *P. leucopus* primers developed by Crew and Bates (1996). The Pele and PEB primers were chosen based on their isolation from species related to *N. albigula*. *Peromyscus maniculatus*, *P. leucopus*, and *N. albigula* are all members of the subfamily Neotominae and all are principal hosts of either hantavirus or arenavirus (Fulhorst et al. 2007). *Mus* primers were selected as a test marker in an attempt to obtain additional data. All 3 primers successfully amplified *N. albigula* DNA.

The reagents used in the PCR process included GoTaq Flexi DNA Polymerase (Promega Corporation, Madison, Wisconsin) and FailSafe PCR 2X PreMix B (Epicentre Biotechnologies, Madison, Wisconsin). A master mix was made using the following amounts of reagents: 1.25µL of each primer, 0.25µL DNA polymerase, and 7.75µL water. To each PCR tube the following was added: 10.5µL master mix, 12.5µL PreMix B, and 1µL of DNA for a total volume of 24µL. Pele primers were optimized using the following protocol: 94°C 2 min; 35 cycles (94°C 30 sec, 48°C 30 sec, 72°C 90 sec); 72°C 15 min; 15°C soak. PEB and *Mus* primers were optimized using the following protocol: 94°C 2 min; 35 cycles (94°C 30 sec, 58°C 30 sec, 72°C 90 sec); 72°C 15 min; 15°C soak. PCR product was verified by electrophoresis on a 1% agarose gel. Ethidium bromide was added to visualize the DNA under a UV light.

Genetic analysis: All amplicons were screened using capillary electrophoresis-based single-strand conformation polymorphism (SSCP) methodologies on an ABI3130 Genetic DNA Analyzer (Applied Biosystems Inc., Foster City, California). SSCP methodologies were originally used to distinguish 2 alleles at a given chromosome locus (Orita et al. 1989). Orita et al. (1989) found that a mobility shift difference in single-stranded DNA on a neutral polyacrylamide electrophoresis gel was due to single nucleotide polymorphisms (SNPs), also known as point mutations. The substituted nucleotide caused a conformational change in the DNA strand when re-annealed. This caused the single strands to move at a different rate through the gel. In homozygous individuals, 2 of the DNA strands re-annealed similarly and 2 bands were seen on the gel. In heterozygous individuals, all 4 DNA strands re-annealed in different ways and 4 bands were visualized on the gel (Figure 6). As compared to restriction fragment length polymorphism (RFLP) techniques, SSCP was found to be superior in detecting genetic polymorphisms (Orita et al. 1989).

More recently, SSCP analysis techniques were studied using capillary electrophoresis (CE) as opposed to gel electrophoresis. Arakawa et al. (1996) performed SSCP using CE on a Model 270A capillary electrophoresis system (Applied Biosystems Inc., Foster City, California) with a monitored on-column 260 nm array. Using amplified DNA fragments with a known single nucleotide mutation, they determined the sensitivity of SSCP by CE methods. They were able to detect a point mutation as a shift in the peak mobility of the separated single-strands. As the individual DNA strands of each sample re-annealed and folded, they migrated through the polymer in the capillary array. When the strands of DNA passed by the laser, blue or green wavelengths were emitted, corresponding to the forward or reverse fluorescently labeled primers, respectively. In homozygous individuals, 2 of the

DNA strands re-anneal similarly due to the same point mutation and migrated at the same speed. Therefore, in homozygous individuals 2 peaks are visualized on the computer screen: 1 blue and 1 green (Figure 7). In heterozygous individuals, all 4 DNA strands re-anneal in different ways, migrating at 4 separate speeds. Thus, in these individuals 2 blue peaks and 2 green peaks are portrayed (Figure 8).

Furthermore, Arakawa et al. (1996) studied how acrylamide polymer concentration, running temperature, and base size affect peak mobility. They found that high polymer concentration and low temperatures are preferred to maintain the DNA strand conformation. However, high temperatures could be used to detect longer fragments. Cai and Touitou (1993) found that reducing primer concentration during PCR could help in optimizing SSCP analysis.

In this study, SSCP was used to screen and find point mutations within a large number of DNA samples. DNA fragments within the PCR product were denatured using heat and maintained in the denatured state using formamide. This was followed by rapid chilling to prevent re-annealing of complementary strands. The separated DNA fragments were pulled through a capillary array filled with POP^{TM} Conformational Analysis Polymer (CAP; Applied Biosystems Inc.), a non-denaturing sieving medium. CAP was prepared using the following proportions (as recommended by Applied Biosystems Inc.): 55.6% CAP, 10% glycerol, 10% 10X EDTA Genetic Analyzer Buffer, and 24.4% water. The protocol used for capillary electrophoresis was as follows: oven temperature 35°C; polymer fill volume 4840 steps; current stability 5 amps; pre-run voltage 15kV; pre-run time 180s; injection voltage 1.2kV; injection time 12s; voltage number of steps 40nK; voltage step interval 15s; data delay time 1s; run voltage 15kV; run time 2000s. The mobility of the DNA

was compared to a size standard, GeneScan™ 500 ROX™ (Applied Biosystems Inc.). SSCP genotypes were scored using GeneMapper version 4.0 software (Applied Biosystems Inc.).

Statistical analyses: Analyses of genetic diversity, including number of alleles, allele frequency, heterozygosity, and polymorphic information content (a measure of the marker's useful in linkage analysis) were completed using CERVUS 3.0 (Kalinowski et al. 2007). CERVUS is a program often used to assign parentage in populations based on genotyping. It was originally developed to assign paternities using genetic data from a population of red deer (*Cervus elaphus*) from Rum, Scotland (Marshall et al. 1998). The current version of CERVUS incorporates a genotype error rate of 0.01, as well as corrected likelihood equations (Morrissey and Wilson 2005) to allow for a higher number of assigned paternities at a given confidence interval. For this study, CERVUS 3.0 was used to obtain standard summary statistics of genetic diversity based on allele information, rather than to assign parentage in a population.

Population genetic structure was analyzed using STRUCTURE 2.3.3 (Pritchard et al. 2000). STRUCTURE is a Bayesian analysis program developed to assign individuals to a population or determine how many populations or clusters best fit a range of individuals based on multilocus genotype data. The program's assumptions include Hardy-Weinberg equilibrium and marker linkage equilibrium. In Bayesian analysis, the observed data are treated as fixed, known quantities while the unknown parameters are treated as random variables (Martin 2005). STRUCTURE uses the Markov Chain Monte Carlo (MCMC) method to integrate parameter space and assign populations (Pritchard et al. 2000). The program is initially run with a set "burn-in period" in which the computer uses algorithms to explore

subsets of combinations of the data to "find" the best probability value. Once the designated burn-in period is complete, all previous data are disregarded and the computer continues with MCMC iterations. This is a trial-and-error process. The results obtained are dependent upon how long the burn-in period runs. Falush et al. (2003, 2007) extended the original models to include linked loci and correlated allele frequencies as well as null alleles and dominant markers, respectively. Hubisz et al. (2009) added new models to STRUCTURE to account for data sets in which there are few loci, few individuals, or little divergence among groups.

STRUCTURE 2.3.3 was used in this study to assess population structure of the woodrat samples from Arizona. Genetic diversity information obtained in this research would acquire better context if the number of populations was ascertained. The following parameters were set in the program: 163 individuals, 4 loci (designated as PEBB, PEBG, PeleB, PeleG), K= 1-5 (number of populations assumed) and RANDOMIZED was turned off. Additional parameters involved running the data at 3 burn-in periods: 1000, 100,000, and 1,000,000; with 1000, 10,000, and 100,000 MCMC iterations respectively after burn-in. The K value considered the "best fit" for these data was determined based on length of burn-period and the estimated Ln probability of data value (Ln P(D)). The Ln(P(D)) is an estimate of the ln probability of X|K, where X is the genotype and K is the number of populations characterized by a set of alleles at each locus (Pritchard et al. 2000).

Dr. Fulhorst of the University of Texas Medical Branch provided serostatus data, the presence or absence of arenavirus antibodies per individual, for a small subset of woodrats used in this study (Table 5). Most of the woodrats came from Hackberry Creek, a locality in which he found viral presence to be unusually high (Fulhorst, pers comm.). To determine if a relationship exists between the alleles identified by GeneMapper at each locus and the

serostatus of that individual, a linear correlation coefficient, r, was calculated (D'Agostino Sr. et al. 2006). In addition, a Phi Coefficient of Association was calculated to measure the relationship between the genotype and serostatus of the individual, both binary variables (D'Agostino Sr. et al. 2006).

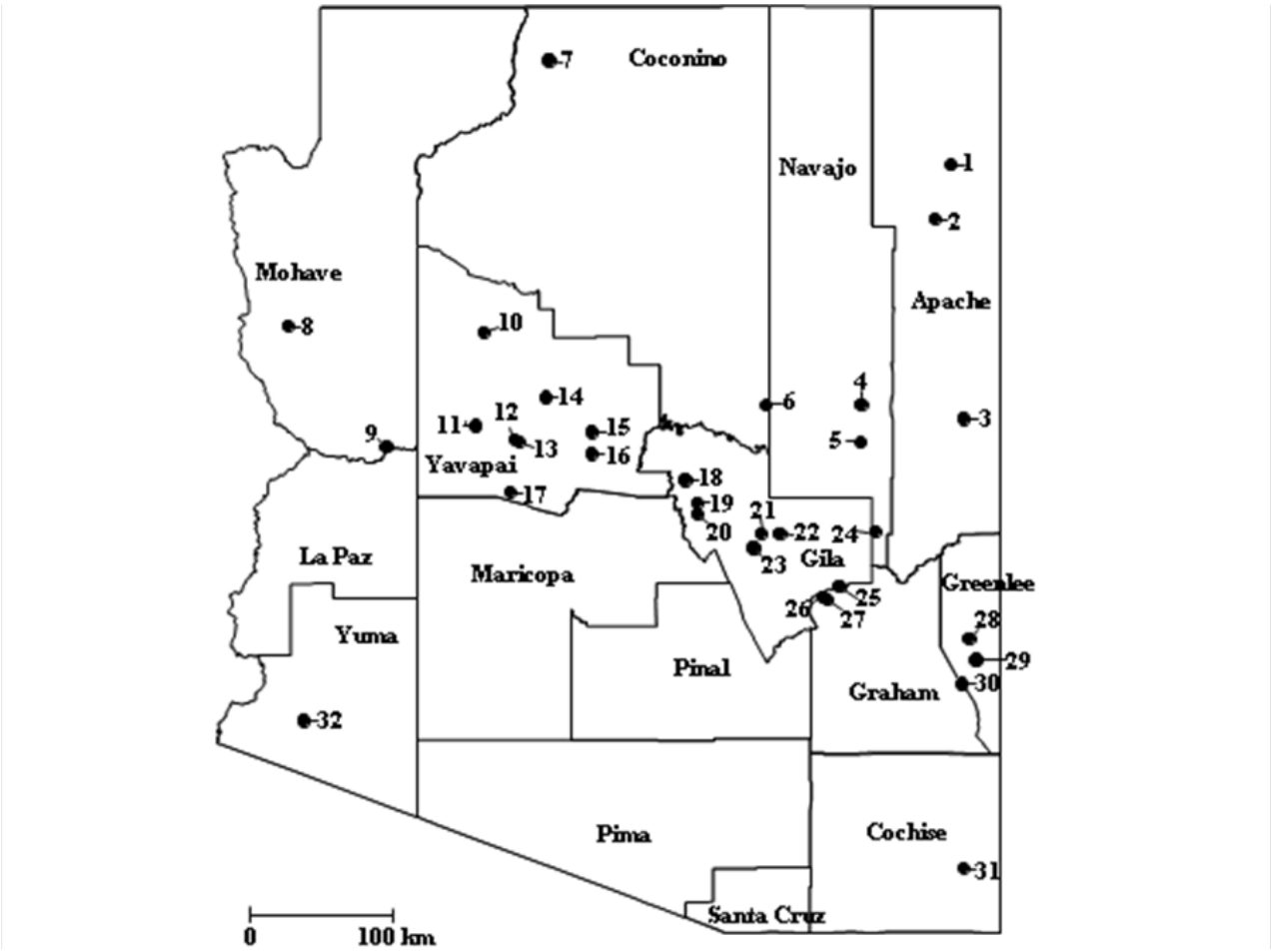

Figure 5. Map of Arizona showing collecting localities for *Neotoma albigula*. Site numbers correspond to data in Table 3 (modified from Abbott et al. 2004).

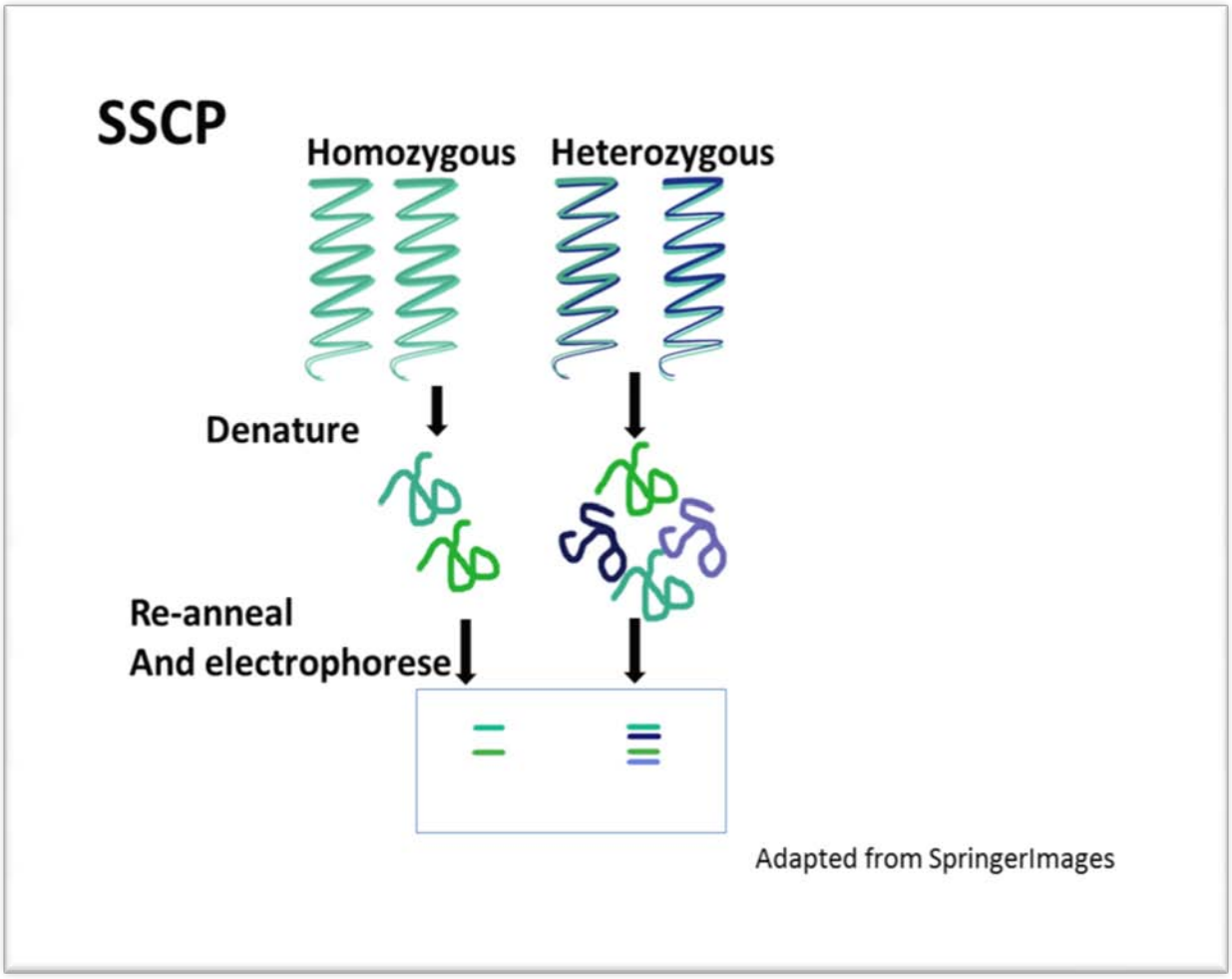

Figure 6. Depiction of single-strand conformational polymorphism analysis. The zig-zag lines represent DNA strands. The denatured strands re-anneal based on a point mutation. In homozygous individuals, both DNA strands contain the same point mutation, thus only 2 conformational changes are possible. In heterozygous individuals, all 4 single strands of DNA contain a different point mutation and so 4 conformational changes are possible. When run on an electrophoretic gel, a homozygous individual will be represented by 2 bands, whereas a heterozygous individual will be represented by 4 bands (adapted from Springerimages).

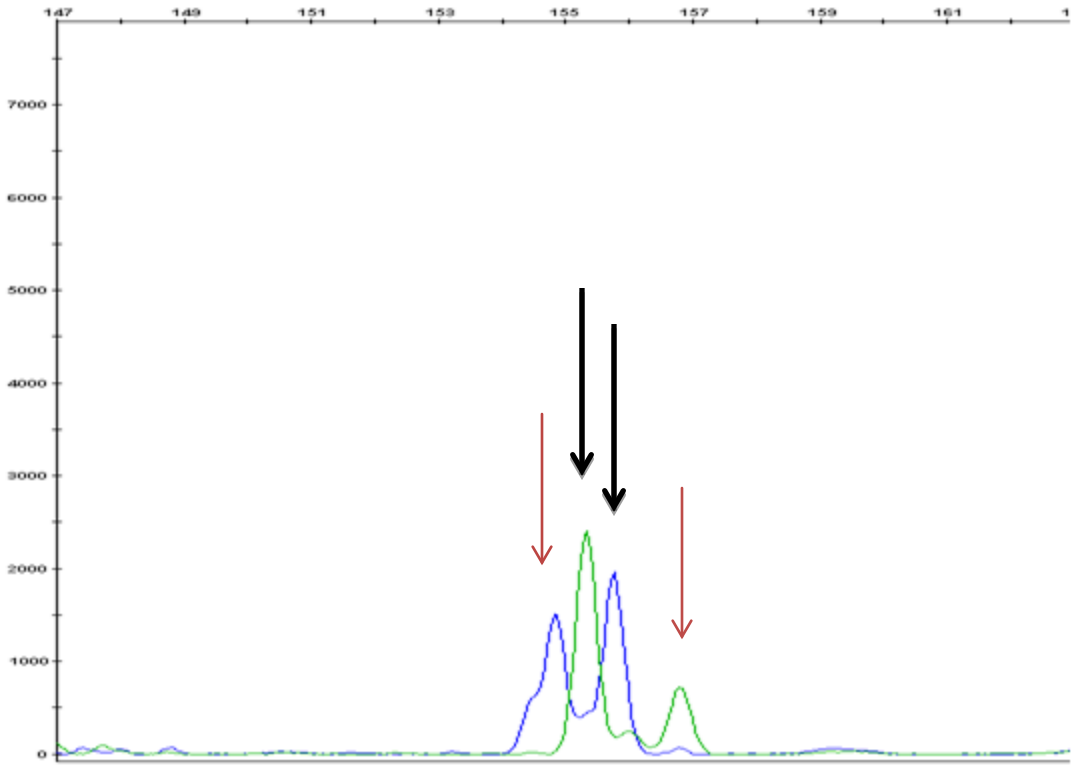

Figure 7. Homozygous individual represented by single strand conformational polymorphism analysis using capillary electrophoresis. The blue and green peaks correspond to the 2 bands visualized using gel electrophoresis (see Figure 6). Black arrows indicate alleles, red arrows indicate artifacts.

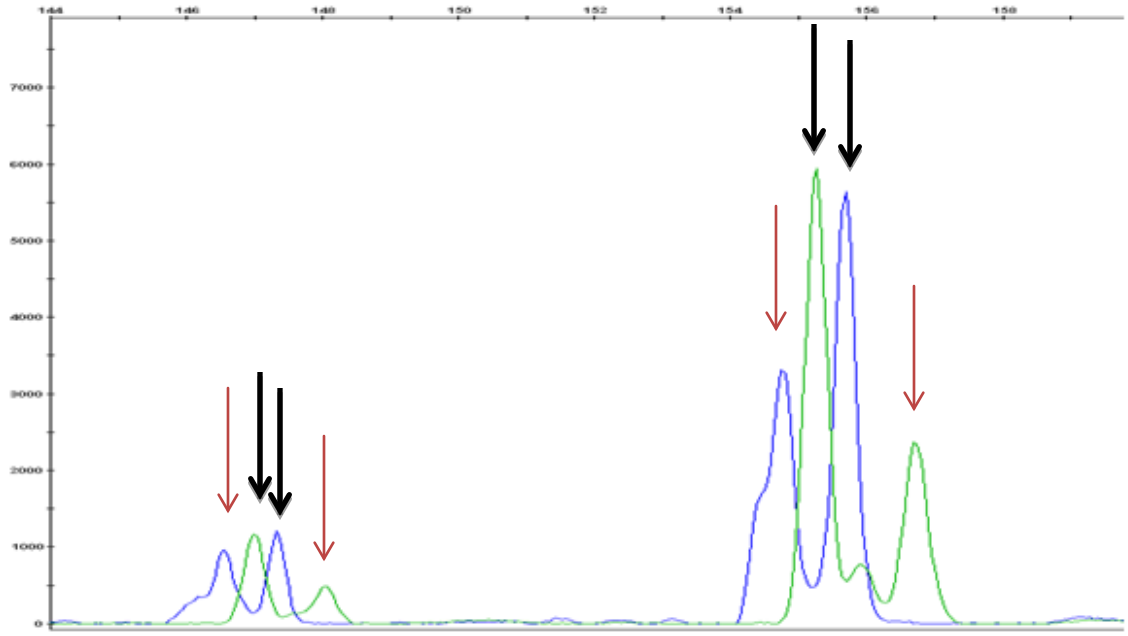

Figure 8. Heterozygous individual represented by single strand conformational polymorphism analysis using capillary electrophoresis. The blue and green peaks correspond to the 4 bands visualized using gel electrophoresis (see Figure 6). Black arrows indicate alleles, red arrows indicate artifacts.

Table 3. Locality data for 1,196 individuals collected from 32 localities in Arizona. For each locality, site number (Site, corresponding to Figure 5), specific locality name (Name), total number of individuals collected (N), and number of individuals included in this study (n) are provided.

Site	Name	N	n
1	AZ: Apache Co.; Three Turkey	1	1
2	AZ: Apache Co.; CDC	3	2
3	AZ: Apache Co.; Saint Johns	6	5
4	AZ: Navajo Co.; MVP Pig Farm	14	5
5	AZ: Navajo Co.; Lone Pine Reservoir	12	5
6	AZ: Navajo Co.; Trick Tank Draw	10	5
7	AZ: Coconino Co.; Snake Gulch	2	2
8	AZ: Mohave Co.; Oatman	15	5
9	AZ: Mohave Co.; Love Camp/Lake Alamo	93	5
10	AZ: Yavapai Co.; Pine Flat	65	5
11	AZ: Yavapai Co.; Hillside	123	5
12	AZ: Yavapai Co.; Wagner	46	2
13	AZ: Yavapai Co.; Hassayampa	80	5
14	AZ: Yavapai Co.; Granite Dells Ranch	36	5
15	AZ: Yavapai Co.; Sycamore Station	63	5
16	AZ: Yavapai Co.; Horseshoe Ranch	17	5
17	AZ: Yavapai Co.; Sayer Spring	132	5
18	AZ: Gila Co.; Barnhardt Trailhead	10	5
19	AZ: Gila Co.; Windmill Tank	85	5
20	AZ: Gila Co.; White Cow Mine	38	5
21	AZ: Gila Co.; Cherry Creek	15	5
22	AZ: Gila Co.; Gleason Flat	17	5
23	AZ: Gila Co.; Coon Creek	15	5
24	AZ: Gila Co.; Sierra Anchas Mountains	2	2
25	AZ: Graham Co.; Warm Springs	5	4
26	AZ: Graham Co.; Hackberry Creek	36	29
27	AZ: Graham Co.; Brushy Tank	30	5
28	AZ: Greenlee Co.; San Francisco River	11	5
29	AZ: Greenlee Co.; McDowell Road	16	5
30	AZ: Greenlee Co.; Black Hills	30	5
31	AZ: Cochise Co.; Chiracahua Mountains	20	1
32	AZ: Yuma Co.; Welton Citrus	148	5
	Total	1196	163

Table 4. Fluorescently labeled primers used for PCR in this study.

Primer	Species	Sequence	Dye Label	Citation
PeleDQAex2-F	*Peromyscus leucopus*	5’ ACAGCTGACCATGTTGGCGCCTA 3’	6-FAM (Blue)	Crew and Bates 1996
PeleDQAex2-R		5’ CACGTACCATTGGTAGCTGGGGTA 3’	HEX (Green)	
PEB1	*P. maniculatus*	5’ TTGGTCAGGGACCCCCAACCTCGGT 3’	6-FAM (Blue)	Richman et al. 2001
PEB2		5’ CAGGAGGTTGTGGTGGTCCAGGG 3’	HEX (Green)	
MusE2Ab1	*Mus musculus*	5’ CACGGCCCGCCGCGCTCCCGC 3’	6-FAM (Blue)	She et al. 1991
MusE2Ab2		5’ CGGGCTGACCGCGTCCGTCCGCAG 3’	HEX (Green)	

Table 5. Serostatus data corresponding to alleles and genotypes of a subset of *Neotoma albigula*. Each individual from whom tissues were collected are assigned a number as a record of deposition. The NSRL at Texas Tech University uses Tissue and Karyotype (TK) numbers for this purpose. The locality information is given for each individual (BT = Brushy Tank; HC = Hackberry Creek). The presence or absence of antibodies for an arenavirus (AVab) is given as a 1 or 0, respectively. Two loci ("Pele" and "PEB") were used in this statistical analysis. PEB and Pele alleles are designated with peak color and number. For instance PEBB1 refers to the first "blue" peak emitted during SSCP analysis. Genotypes (homo = homozygous; hetero = heterozygous) were assigned based on the base pair size of the blue peaks for each locus. An asterisk * indicates missing data. Genotypes listed with N/A were unable to be determined. Only those individuals with complete sets of data were used in the statistical analysis.

TK#	LOCALITY	AVAb	PeleB1	PeleB2	PeleG1	PeleG2	Genotype	PEBB1	PEBB2	PEBG1	PEBG2	Genotype
114521	BT (map 27)	1	291	291	291	291	Homo	147	158	147	158	Hetero
114522	BT (map 27)	0	288	288	288	288	Homo	147	156	147	158	Hetero
114523	BT (map 27)	0	292	292	292	292	Homo	147	156	147	155	Hetero
114524	BT (map 27)	1	*	*	*	*	N/A	156	156	156	156	Homo
114525	BT (map 27)	1	285	285	285	285	Homo	156	156	156	156	Homo
114558	HC (map 26)	1	288	288	288	288	Homo	147	158	147	159	Hetero
114559	HC (map 26)	1	284	284	284	284	Homo	147	156	147	157	Hetero
114561	HC (map 26)	1	288	288	288	288	Homo	147	156	147	158	Hetero
114562	HC (map 26)	1	*	*	*	*	N/A	147	156	147	155	Hetero
114563	HC (map 26)	1	*	*	*	*	Homo	147	158	147	158	Hetero
114564	HC (map 26)	1	*	*	*	*	Homo	147	158	147	158	Hetero
114565	HC (map 26)	1	*	*	*	*	Homo	147	154	147	155	Hetero
114566	HC (map 26)	1	288	288	288	288	Homo	147	156	147	158	Hetero
114567	HC (map 26)	0	*	*	*	*	N/A	147	158	147	158	Hetero
114568	HC (map 26)	1	*	*	*	*	N/A	*	*	*	*	N/A
114569	HC (map 26)	0	*	*	*	*	N/A	147	147	147	147	Homo
114571	HC (map 26)	1	*	*	*	*	N/A	147	158	147	158	Hetero
114572	HC (map 26)	1	*	*	*	*	N/A	147	156	147	155	Hetero
114573	HC (map 26)	1	*	*	*	*	N/A	147	156	147	155	Hetero
114574	HC (map 26)	1	292	292	292	292	Homo	147	154	147	155	Hetero
114575	HC (map 26)	1	*	*	*	*	N/A	147	156	147	155	Hetero
114576	HC (map 26)	1	*	*	*	*	N/A	147	147	147	147	Homo
114577	HC (map 26)	1	285	285	285	285	Homo	147	158	147	159	Hetero
114578	HC (map 26)	1	285	285	285	285	Homo	147	156	147	157	Hetero
114579	HC (map 26)	1	*	*	*	*	N/A	156	156	156	156	Homo
114581	HC (map 26)	1	*	*	*	*	N/A	147	156	147	155	Hetero
114582	HC (map 26)	1	285	285	285	285	Homo	147	158	147	158	Hetero
114584	HC (map 26)	1	*	*	*	*	N/A	147	158	147	158	Hetero
114585	HC (map 26)	1	282	282	282	282	Homo	147	158	147	158	Hetero
114586	HC (map 26)	1	286	286	286	286	Homo	147	156	147	155	Hetero
114589	HC (map 26)	1	287	287	287	287	Homo	147	156	147	155	Hetero

CHAPTER 3
RESULTS

Determining levels of genetic variation: The first objective of this study was to determine levels of genetic variation (number of alleles, allele frequency, and heterozygosity) at *Mhc* loci in *N. albigula* populations. Genetic variation was analyzed at the Pele and PEB loci only. Due to inconsistent and indistinguishable multiple peaks, the Mus locus is not included in statistical analyses.

Ninety-five individuals were analyzed at the Pele locus in CERVUS (Table 6). Ten alleles were detected, migrating in a range from 282 to 292 base pairs. The allele with the highest frequency was 288 (0.2947). All individuals were found to be homozygous, although the expected heterozygosity was 0.842. The polymorphic information content (PIC) value was 0.8197.

One hundred fifty-four individuals were analyzed at the PEB locus in CERVUS (Table 7). Eight alleles were detected, migrating in a range from 146 to 163 base pairs. The alleles with the highest frequency were 130 and 135, occurring at 0.4221 and 0.4383, respectively. The observed heterozygosity was 0.936, greater than the expected heterozygosity of 0.685. The PIC value for this locus was 0.5504.

Determine if Mhc markers can detect genetic subunits within populations: The second objective of this study was to determine if *Mhc* markers can be used to detect genetic subunits within *N. albigula* populations. The results from STRUCTURE revealed the possibility of 2 distinct clusters (K=2) from the woodrat genetic data (Figure 9). The criteria for determining the number of clusters included the Ln(P(D)) value of -1761.7 (Table 8) and the burn-in length of 100,000 with 10,000 Markov Chain Monte Carlo repetitions. Individuals were assigned to 1 cluster or the other if the probability was greater than 94%. In

addition, posterior probabilities (Pr(X|K) were manually calculated for K=1-5 to verify the probability in assigning 2 clusters (Table 8). The posterior probability calculated for K=2 was 0.9975.

The mean F_{ST} value was 0.1058, indicating moderate genetic differentiation between clusters. The average expected heterozygosity between individuals of cluster 1 was 0.7581 and 0.7746 for cluster 2. Figure 10 depicts the estimated fraction of clusters found at each locality in Arizona. However, numerous individuals within localities could not be assigned to 1 cluster or the other.

Determining a correlation between viral infection and Mhc alleles: The third objective of this study was to determine if there was a correlation between viral infection and *Mhc* alleles in *N. albigula* populations. For the subset of individuals for which serostatus data was provided, a correlation coefficient was calculated to determine if a linear correlation existed between specific alleles at each locus and individuals who tested either antibody positive or negative for an arenavirus. The correlation coefficient, r, ranges from -1 to +1. A value of -1 indicates a strong negative linear correlation and a value of +1 indicates a strong positive linear correlation. A value of 0 for r indicates no correlation between the variables. For consistency with CERVUS analysis only the "blue" alleles at the Pele and PEB loci were considered. For the Pele locus, there was a negative correlation (r = -0.4052) between the designated alleles and serostatus. For the PEB locus, there was a positive but very small correlation (r = 0.1307, r = 0.2496) between the specific alleles and serostatus (Table 9a).

An analysis was also completed to determine if a nominal correlation existed between the genotype of the individual at each locus and antibody positivity or negativity. For this test, a Phi coefficient was calculated using a 2x2 contingency table. The range of a

calculated Phi coefficient is similar to a correlation coefficient. A value of -1 indicates a strong negative correlation, a value of +1 indicates a strong positive correlation and a value of 0 indicates no correlation. At the PEB locus, the Phi coefficient was +0.09, an indication of almost no correlation between the variables. At the Pele locus, no Phi coefficient was able to be calculated, due to all individuals being homozygous. The results are summarized in Table 9b.

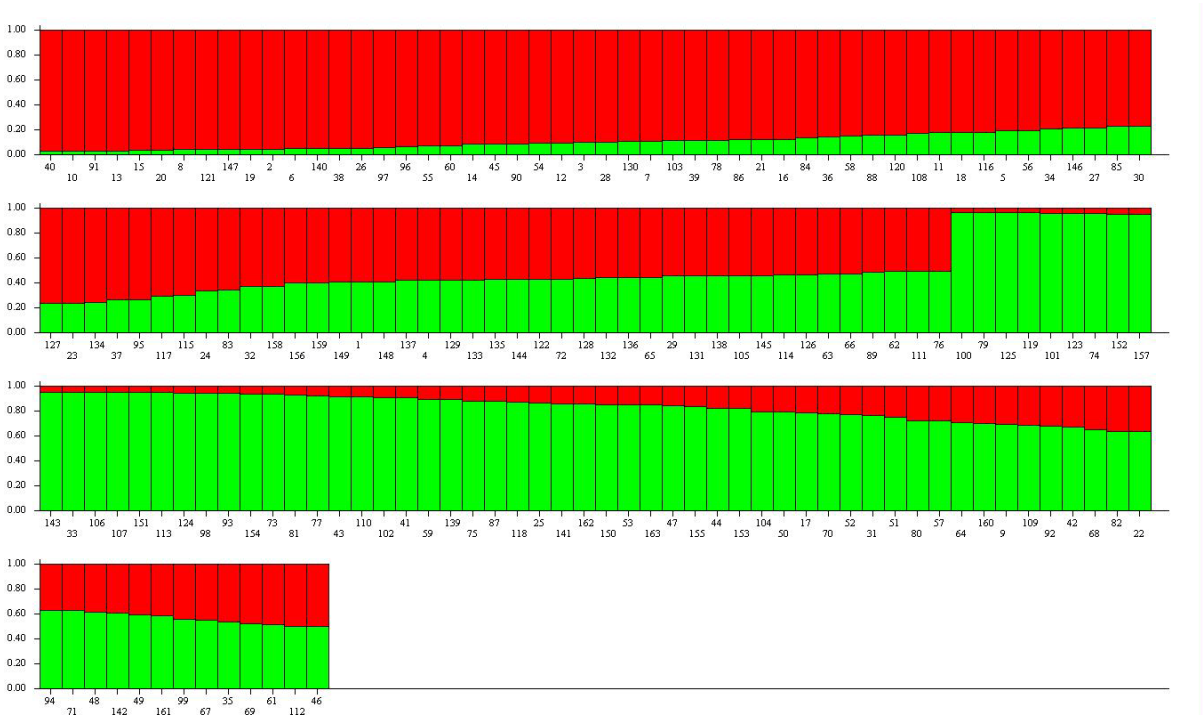

Figure 9. Based on "best-fit" results from STRUCTURE, the population of *Neotoma albigula* used in this study are clustered into 2 groups (K=2), shown as red and green. Individuals that contain a mixture of 40-60% of either cluster cannot be definitively assigned to a group.

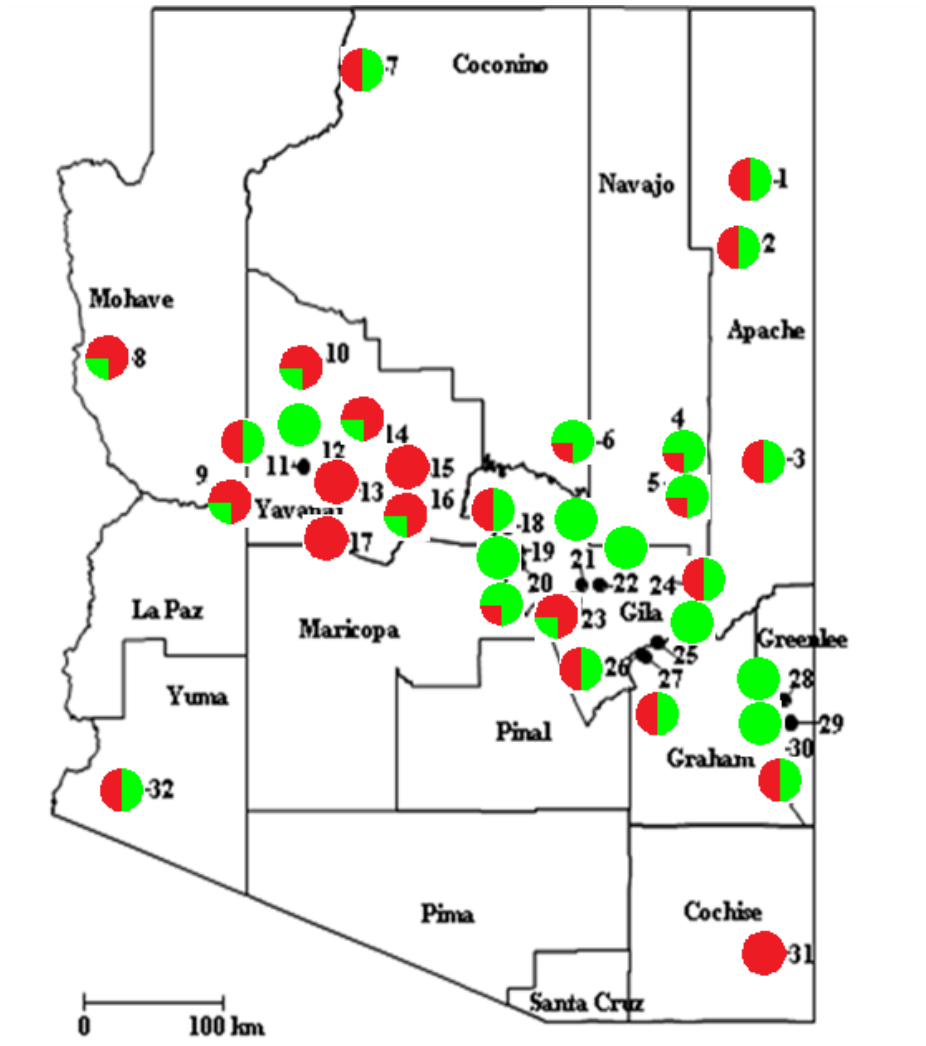

Figure 10. Based on "best-fit" results from STRUCTURE, the population of *Neotoma albigula* used in this study are clustered into 2 groups (K=2), shown as dark red and light green (see Figure 9). Individuals were categorized as belonging to the red cluster, green cluster, or a mixture of both. The number of individuals taken from each locality (see Table 3) was also considered in mapping the distribution of each cluster at all localities in Arizona (map modified from Abbot et al. 2004).

Table 6. CERVUS results at the Pele locus. The 10 designated alleles are listed by base pair length as well a count of how many copies were detected. The number of heterozygous and homozygous individuals and the frequency of each allele were also calculated. (N = the number of individuals analyzed; H_O = observed heterozygosity; H_E = expected heterozygosity; PIC = polymorphic information content)

CERVUS – Pele locus **N = 95** **Mean H_o = 0.000** **Mean H_E = 0.842** **Mean PIC = 0.8197**				
Allele	Count	Heterozygous	Homozygous	Frequency
282	14	0	7	0.0737
283	4	0	2	0.0211
284	6	0	3	0.0316
285	28	0	14	0.1474
286	2	0	1	0.0105
287	12	0	6	0.0632
288	56	0	28	0.2947
290	24	0	12	0.1263
291	20	0	10	0.1053
292	24	0	12	0.1263

Table 7. CERVUS results at the PEB locus. The 8 designated alleles are listed by base pair length as well a count of how many copies were detected. The number of heterozygous and homozygous individuals and the frequency of each allele were also calculated. (N = the number of individuals analyzed; H_O = observed heterozygosity; H_E = expected heterozygosity; PIC = polymorphic information content)

CERVUS – PEB locus **N = 154** **Mean H_o = 0.936** **Mean H_E = 0.685** **Mean PIC = 0.5504**				
Allele	Count	Heterozygous	Homozygous	Frequency
146	6	6	0	0.0195
147	130	118	6	0.4221
154	15	15	0	0.0487
156	135	87	24	0.4383
157	2	2	0	0.0065
158	18	18	0	0.0584
161	1	1	0	0.0032
163	1	1	0	0.0032

Table 8. Posterior probabilities (Pr(X|K) calculated for K=1-5 based on Ln(P(D)) values computed by STRUCTURE using allelic data (Pritchard et al. 2010).

K	Ln(P(D))	Pr(X|K)
1	-1791.1	0.1092
2	-1761.7	0.9975
3	-1767.6	0.2689
4	-1811.5	0.0842
5	-1789.6	0.1522

Table 9a. Summary of correlation statistics involving allelic data and serostatus data. The alleles used in the analysis were the "blue" peaks at each locus. One blue peak was used for the Pele locus because all individuals were homozygous. A correlation coefficient for each allele was calculated to detemine if a relationship existed with the presence or absense of antibodies for an arenavirus. A weak, negative relationship was found between the Pele blue alleles and serostatus data. A weak, positive relationship was found between the PEB blue alleles and serostatus data.

Alleles	Correlation coefficient
PeleB	-0.4052
PEBB1	+0.1307
PEBB2	+0.2496

Table 9b. Summary of correlation statistics involving genotypic data and serostatus data. Because both sets of variables, genotypes (homozygous and heterozygous) and serostatus (presence or absence of antibodies) were dichotomous, a Phi coefficient was calculated to determine if a relationship existed between them. No value could be determined at the Pele locus because all individuals were homozygous. A very weak correlation was found between the genotypes at the "PEB" locus and serostatus.

Genotypes	Phi coefficient
Pele	N/A
PEB	+0.09

CHAPTER 4
DISCUSSION

In this study, *Mhc* loci were used as genetic markers to screen for variation in *Neotoma albigula* to determine the coevolutionary association of this species with at least 3 arenaviruses. The goals of this study were to 1) determine levels of genetic variation (number of alleles, allele frequency, and heterozygosity) at *Mhc* loci in *N. albigula* populations; 2) determine if *Mhc* markers could be used to detect genetic subunits within *N. albigula* populations; and 3) determine if there was a correlation between viral infection and *Mhc* alleles in *N. albigula* populations. The first 2 goals produced positive results. *Mhc* markers detected levels of genetic variation and genetic subunits. The third goal produced negative results. Based on correlation calculations, it cannot be concluded if a relationship exists between *Mhc* variation and viral infection.

The overall hypothesis of this study was not supported because a correlation between the levels of genetic variation at *Mhc* loci in *N. albigula* and viral strains could not be established. There are some possible reasons that the hypothesis might still be accepted even though not supported by my data: 1) optimization issues involving the SSCP analysis methodology, 2) small sample sizes, and 3) insufficient viral data availability.

Optimization of SSCP analysis: The optimization of SSCP analysis was a particularly challenging aspect of this research study. There are some aspects of this methodology that may require some modifications prior to future research endeavors.

The recommended polymer for SSCP analysis using CE was conformational analysis polymer (CAP). Unlike other polymers, such as POP-4 and POP-7 (Applied Biosystems Inc.), CAP had to be diluted with glycerol, EDTA buffer, and water before use. Inconsistencies in the preparation of CAP with each new run on the genetic analyzer created

early difficulties in SSCP optimization. These issues were resolved by diluting all CAP at once and running as many samples as possible at one time.

To properly identify alleles, a size standard (or ladder) must be involved. The size standard used for this SSCP analysis was ROX 500. This size standard had been utilized in other methodologies, such as microsatellites genotyping. However, ROX 500 seems to be incompatible with the CAP used in the genetic analyzer. The size standard peaks had to be visually inspected and manually adjusted before GeneMapper was able to call alleles. Once this was completed, allele calling proceeded as per standard methodology.

Software capable of analyzing SSCP-analyzed loci was not currently available during this research study. This issue was resolved using available microsatellite-based software to identify and designate *Mhc* alleles. The physical differences between microsatellite loci and *Mhc* loci could have affected the way the allele peaks were identified, thus the researcher visually inspected the allele peaks for accuracy.

Ascertainment bias may have affected the data collection of this research, producing negative results. Ascertainment bias is any distortion in measuring the true frequency of a variable based on the way it was obtained. In this study the genotyping of individuals was based on conformational changes in DNA strands due to point mutations. The PCR products containing the point mutation were obtained through successful amplification by non-species specific primers. Similar loci found among different species may be of differing lengths or prevalence. These may explain the appearance of the allele peaks shown in GeneMapper.

The migration of DNA fragments may also be responsible for obtaining negative results. As the DNA fragments migrate through the capillary array, they are detected by a laser in the genetic analyzer at a certain point of time. The fragments are assigned a “size”

by the computer software based on the time they pass the laser. The DNA fragments are folded structures, not linear strands as they migrate through the array. The differences in migration of the DNA fragments are not based on size, which is what the software considers, but, rather how each is folded. Due to the high sensitivity of wavelength detection, a small variation in migration time due to folding could be responsible for the differences in alleles detected by the analysis software.

Allele calls: The identification of alleles and maintaining consistency among individuals was of significant interest in this study. When SSCP analysis was first being developed (Orita et al. 1989), DNA fragments were ran on a polyacrylamide electrophoretic gel. Alleles were identified as bands on the gel compared to a size standard. SSCP analysis using CE allows computer software to identify peaks of a certain size range as alleles. This technique is more sensitive than observing bands on a gel. The sensitivity of the laser may be causing a detection of a series of emitted wavelengths, resulting in multiple peaks per size range. The computer software analyzed these multiple peaks as multiple alleles per locus.

The presence of multiple alleles per locus called by GeneMapper may have distorted the actual number of alleles present at a given locus. Ten alleles were detected at the Pele locus and 8 alleles were detected at the PEB locus. Some alleles occurred at a notably higher frequency than others, so it is possible that those are "true" alleles. Evans et al. (2010) identified 1 to 2 *Mhc* class II alleles among individuals in their study of Chinook salmon populations. They also identified, based on nucleotide site variation, 17 distinct alleles. In this study, sequencing the alleles will provide insight as to which alleles at a given locus are "real."

As an interesting observation, there was also a noticeable difference in the amplitudes of the peaks (see Figure 8). Certain alleles have been known to be preferentially amplified during PCR. Walsh et al. (1992) provided several possible reasons for this. Two alleles in a heterozygous sample may denature at different temperatures. The allele which denatures less efficiently may not be detected, leading to errors in genotyping. Differences in the priming of synthesis during PCR could also lead to 1 allele being favorably amplified over the other. Walsh et al. (1992) also suggested that any amplification of microsatellites that flank the desired region will result in PCR products of varying lengths. In some circumstances, the shorter fragments are amplified more than the longer fragments. Finally, stochastic fluctuations (sampling error) may be responsible for what appears to be preferential amplification if the genome sample is very small and there is variation in the number of allele copies (Walsh et al. 1992).

Population genetics: Examining the genetics of a population can give a large amount of information regarding small scale evolutionary events, such as natural selection, genetic drift, gene flow, migration, or mutation rates. Population genetics can also be used to understand the relatedness between species (e.g., Haynie et al. 2007) or diversity within populations (e.g., Mendez-Harclerode et al. 2007). In this study, the genetics of the white-throated woodrat population were examined to determine if there was an explanation for the association of this species with at least 3 arenaviruses. The number of alleles, the frequency of those alleles, and levels of observed heterozygosity compared to expected heterozygosity are all indicative of the genetic make-up of a population.

Our data resulted in 10 and 8 alleles at the Pele and PEB loci, respectively. Some of the alleles occurred at a notably higher frequency than others. These common alleles at the

Pele locus included 285, 288, 290, and 292. At the PEB locus, the common alleles were 147 and 156. Some of the uncommon alleles may be unique, which could be confirmed with sequencing and further analysis. The existence of few unique alleles in these data indicates high gene flow, which supports the results produced by STRUCTURE in which only 2 clusters were found.

Genetic variation in a population is measured by the number of heterozygous individuals, regardless of the number of alleles found per locus. The PEB locus contained fewer alleles but a greater number of heterozygotes. This may imply that the PEB locus could be a significant *Mhc* marker for population studies.

There are other factors that explain the population genetic results of this study. Out of nearly 1200 white-throated woodrats collected throughout the state of Arizona, 163 were used in this study. Those 163 individuals represent 5 individuals, on average, taken from 32 different localities. Some localities included 30 individuals, whereas others only included 1 individual. The variation in the number of samples per locality, as well as the overall small sample size, could explain the number of alleles and frequency of those alleles seen at each locus.

Levels of heterozygosity indicate levels of variation in a population. The small sample size in this study could explain the large deviation in both the observed and expected levels of heterozygosity at both loci. The PEB locus had a larger observed value than expected while the Pele locus had a much smaller observed value than expected. Increasing the sample size per locality should yield more accurate results regarding levels of variation in the population. In addition, Clark et al. (2005) found that correcting for ascertainment bias in

sample data sets of the Human Genome Project reduced the discrepancies in the estimations of heterozygosity and population subdivisions.

For this study, the individual woodrats were considered to originate from 1 large population in Arizona. However, the existence of 2 or more subpopulations can give more insight regarding the genetic structure of the population as a whole. To detect population structure, the program STRUCTURE was used. STRUCTURE results indicated that the individuals formed 2 clusters, with several individuals containing mixtures of both clusters. When cluster information was plotted per locality on the map of Arizona (see Figure 10), there was a noticeable distinction between those samples found in the western part of the state as compared to the eastern part.

Four loci (2 "loci" for PEB and 2 "loci" for Pele) were used in the analysis and the statistics indicate that there is moderate genetic differentiation between the 2 clusters. Pritchard et al. (2000) found that there was an increase in the accuracy of assigning K populations in STRUCTURE using more loci in analysis. With only 4 loci used, there would be a discrepancy in our cluster estimates. Accuracy in K is also dependent on the number of individuals, the amount of admixture, and the differences in allele frequencies among populations. It should also be noted that the value of K, however, may not necessarily reflect clear biological significance (Pritchard et al. 2000).

Viral data: The viral data supplied by Dr. Fulhorst provided preliminary insight into applying correlation statistics to allelic and serostatus data. Correlation analysis was applied to the allelic and genotypic data for 18 individuals for the Pele locus and 30 individuals for the PEB locus. For the Pele alleles, the negative value can be interpreted as individuals with "smaller" alleles (allele calls less than or equal to 285) at this locus are more likely to have

antibodies than those who have “larger” alleles (allele calls greater than or equal to 288). The positive values calculated for the PEB alleles can be interpreted as individuals with “larger” alleles (allele calls greater than or equal to 156) are more likely to have antibodies than those who have “smaller” alleles (allele calls less than or equal to 147).

The small correlation coefficient value for both loci indicates a weak association between variables, regardless of the negative or positive value. The overlap in number of white-throated woodrats screened for genetic variation at *Mhc* loci and tested for the presence of antibodies is also currently limited. As more data are collected and become available, the statistical analysis applied to both variables should increase the accuracy of the results.

Future research: A majority of the time devoted in this study involved the optimization of SSCP analysis. Many aspects of this study can be elaborated upon in future endeavors. The first task will be to sequence the designated alleles of the Pele and PEB loci used in the final analysis. Nucleotide diversity will be analyzed, allowing the researcher to define which and how many alleles occur at each locus. Sequencing may also indicate the possible presence of null alleles at the Pele locus. Null alleles are the result of non-amplification by PCR due to mutations found within the priming region of the sequence, causing the primer to anneal improperly. The result is a deficit in the detection of heterozygotes in a population (Callen et al. 1993). This could explain the high levels of homozygotes at the Pele locus.

Another primer, MhcIIBe2 (Lenz et al. 2009) has become available and work has already begun in amplifying previously extracted *N. albigula* DNA. This primer may be amplifying 2 regions, providing more loci for SSCP analysis using capillary electrophoresis.

Eventual plans involving this primer will include sequencing the resulting loci as well as performing other statistical analyses with viral data, as it becomes available.

Meagher and Potts (1997) used analysis of microsatellites flanking and within the *Mhc* of *Mus domesticus* to discriminate among 9 *Mhc* haplotypes. Their success and the availability of microsatellite analysis software in GeneMapper provides support that similar techniques could be used for further analysis of *Mhc* loci from this study.

In addition to *Mhc* loci, research involving 7 microsatellite loci and the mitochondrial DNA D-loop of the white-throated woodrat population is currently ongoing. Microsatellite loci have been analyzed using STRUCTURE and 5 distinct groups have been assigned (Haynie et al., unpublished data). Mitochondrial D-loop sequences have been analyzed using MEGA software (Tamura et al. 2011) and 4 clades have been identified (Haynie et al., unpublished data). All 3 data sets (*Mhc*, microsatellite, and mitochondrial DNA) and an extended viral dataset will ultimately be combined for further analysis.

The viral data applied in statistical analysis indicated the presence or absence of antibodies against an arenavirus within the individual woodrat. This does not specify which arenavirus caused the infection, when the infection occurred, or the extent of the infection in the woodrat. There are limitations to serostatus tests and it is possible that individuals who tested negative for antibodies had actually been infected. Other rodents collected along with the woodrats have not yet been tested for antibodies (Fulhorst, pers comm.) and work will continue on transmission and infectivity of arenaviruses in this group of white-throated woodrats.

Distinguishing one virus from another is a point of discussion among members of the scientific community. Viruses are classified, among other things, by their genetic sequence,

morphology, and infectious capabilities. There is no "standard" when it comes to identifying viral species or strains, though 5% sequence divergence is typically used as a guideline relating to the taxonomic status of other organisms (Baker and Bradley 2006). Thus remains the questions of how many arenaviruses exist and how many of those are associated with white-throated woodrats.

The data and results generated from this study will be used in a collaborative research project with Texas Tech University and the University of Texas Medical Branch. The success of SSCP using CE methodologies in screening *Mhc* markers for variation could be applied to the almost 1200 *N. albigula* samples currently in storage at the NSRL. Screening genetic markers for variation would be a quick way to determine which individuals contain unique alleles to be sequenced. Previous findings involving the genetic variation of *N. albigula* as well as serostatus and viral information will be used in conjunction with this research to gain a better understanding of the co-evolution between *N. albigula* and arenaviruses.

CHAPTER 5
LITERATURE CITED

Abbott, K.D., Milazzo, M.L, Keith, J., Bradley, R.D., and Fulhorst, C.F. 2004. Epizootiology of arenaviral infections in the white-throated woodrat (Muridae: Sigmodontinae) and other woodrats in Arizona. Journal of Vector Ecology 29:355-364.

Arakawa, H., Nakashiro, S., Maeda, M., and Tsuji, A. 1996. Analysis of single-strand DNA conformational polymorphism by capillary electrophoresis. Journal of Chromatography A 722:359-368.

Baker, R.H. 1956. Mammals of Coahuila, Mexico. University of Kansas Publications of the Museum of Natural History 9:125-335.

Baker, R.J. and Bradley, R.D. 2006. Speciation in mammals and the genetic species concept. Journal of Mammalogy 87:643-662.

Bowen, M.D., Peters, C.J., and Nichol, S.T. 1996. The phylogeny of new world (Tacaribe complex) arenaviruses. Virology 219:285-290.

Cai, Q. and Touitou, I. 1993. Excess PCR primers may dramatically affect SSCP efficiency. Nucleic Acids Research 21:3909-3910.

Cajimat, M.N.B., Milazzo, M.L., Bradley, R.D., and Fulhorst, C.F. 2007. Catarina virus, an arenaviral strain principally associated with Neotoma micropus (southern plains woodrat) in Texas. American Journal of Tropical Medicine and Hygiene 77:732-736.

Cajimat, M.N.B., Milazzo, M.L, Borchert, J.N., Abbott, K.D., Bradley, R.D., and Fulhorst, C.F. 2008. Diversity among Tacaribe serocomplex viruses (Family Arenaviridae) naturally associated with the Mexican woodrat (Neotoma mexicana). Virus Research 133:211-217.

Cajimat, M.N.B., Milazzo, M.L., Haynie, M.L, Hanson, J.D., Bradley, R.D., and Fulhorst, C.F. 2011. Diversity and phylogenetic relationships among the North American Tacaribe serocomplex viruses (family *Arenaviridae*). Virology 421:87-95.

Calisher, C.H., Tzianabos, T., Lord, R.D., and Coleman, P.H. 1970. Tamiami virus, a new member of the Tacaribe group. American Journal of Tropical Medicine and Hygiene 19:520-526.

Calisher, C.H., Nabity, S., Root, J.J., Fulhorst, C.F., and Beaty, B.J. 2001. Transmission of an arenavirus in white-throated woodrats (*Neotoma albigula*), Southern Colorado, 1995-1999. Emerging Infectious Diseases 7:397-402.

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R. I., Mulley, J.C., and Sutherland, G.R. 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. American Journal of Human Genetics 52:922-927.

Carballal, G., Videla, C.M., and Merani, M.S. 1988. Epidemiology of Argentine hemorrhagic fever. European Journal of Epidemiology 4:259-274.

Centers for Disease Control and Prevention, 2000. Fatal illnesses associated with a New World arenavirus—California, 1999–2000. MMWR Morbidity and Mortality Weekly Report 49:709-711.

Charrel, R.N. and de Lamballerie, X. 2009. Zoonotic aspects of arenavirus infections. Veterinary Microbiology 140:213-220.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., and Nielsen. R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Research 15: 1496-1502.

Crew, M.D. and Bates, L.M. 1996. DNA sequence of the *Peromyscus leucopus* MHC class II gene *Aa* (*MhcPeleAa*). Immunogenetics 43:323-324.

D'Agostino Sr., R.B., Sullivan, L.M., and Beiser, A.S. 2006. Introductory applied biostatistics. Thompson Brooks/Cole: Belmont, CA.

de la Torre, J.C. 2008. Reverse genetics approaches to combat pathogenic arenaviruses. Antiviral Research 80:239-250.

Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Müller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Hans-George, R., and Stevanovic, S. 2005. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proceedings of the National Academy of Science 102:7922-7927.

Edwards, C. W., Fulhorst, C.F., and Bradley, R.D. 2001. Molecular phylogenetics of the *Neotoma albigula* species group: further evidence of a paraphyletic assemblage. Journal of Mammalogy 82:267-279.

Enria, D. and Maiztegui, J.I. 1994. Antiviral treatment of Argentine hemorrhagic fever. Antiviral Research 23:23-31.

Evans, M.L., Neff, B.D., and Heath, D.D. 2010. MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*). Heredity 104: 449-459.

Falush, D., Stephens, M., and Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-1587.

Falush, D., Stephens, M., and Pritchard, J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7:574-578.

Feldman, H. W. 1935. Notes on two species of wood rats in captivity. Journal of Mammalogy 65:300-303.

Fernando, M.M.A, Stevens, C.R., Walsh, E.C., De Jager P.L., Goyette, P., Plenge, R.M., Vyse, T.J., and Rioux, J.D. 2008. Defining the Role of the MHC in Autoimmunity: A Review and Pooled Analysis. PLoS Genetics 4: e1000024.

Fulhorst C. F., M. D. Bowen, T. G. Ksiazek, P. E. Rollin, S. T. Nichol, M. Y. Kosoy, and Peters, C.J. 1996. Isolation and characterization of Whitewater Arroyo Virus, a novel North American arenavirus. Virology 224:114-120.

Fulhorst C. F., R. N. Charrel, S. C. Weaver, T. G. Ksiazek, R. D Bradley, M. L. Milazzo, R. B. Tesh, and Bowen, M.D. 2001a. Geographic distribution and genetic diversity of Whitewater Arroyo virus in southwestern United States. Emerging Infectious Diseases 7:403-407.

Fulhorst, C.F., Milazzo, M.L., Bradley, R.D., and Peppers, L.L. 2001b. Experimental infection of *Neotoma albigula* (Muridae) with Whitewater Arroyo virus (*Arenaviridae*). American Journal of Medicine and Hygiene 65:147-151.

Fulhorst, C. F., Bennett, S.G., Milazzo, M.L., Murray, Jr., H., Webb, J.P, Cajimat, M.N.B., and Bradley, R.D. 2002a. Bear Canyon virus: an arenavirus naturally associated with the California mouse (*Peromyscus californicus*). Journal of Emerging Infectious Diseases 8:717-721.

Fulhorst, C. F., Milazzo, M.L., Carroll, D.S., Charrel, R.N., and Bradley, R.D. 2002b. Natural host relationships and genetic diversity of Whitewater Arroyo Virus in southern Texas. American Journal of Tropical Medicine and Hygiene 67:114-118.

Fulhorst, C.F., Milazzo, M.L., Armstrong, L.R., Childs, J.E., Rollin, P.E., Khabbaz, R., Peters, C.J., and Ksiazek, T.G. 2007. Hantavirus and arenavirus antibodies in persons with occupational rodent exposure, North America. Emerging Infectious Diseases 13: 532-538.

Galloway, C.J., Dean, G.E., Marsh, M., Rudnick, G., and Mellman, I. 1983. Acidification of macrophage and fibroblast endocytic vesicles in vitro. Proceedings of the National Academy of Sciences 80:3334-3338.

Gojobori, T. and Yokoyama, S. 1987. Molecular evolutionary rates of oncogenes. Journal of Molecular Evolution 26:148-156.

Gruen, J.R and Weissman, S.M. 1997. Evolving views of the major histocompatibility complex. Blood 9: 4252-4265.

Haynie, M. L. 2006. Population genetics of four species of *Neotoma* from the southwestern United States. Ph.D. Dissertation. Texas Tech University, Lubbock.

Haynie, M.L., Fulhorst, C.F., Rood, M. Bennett, S.G., Hess, B.D., and Bradley, R.D. 2007. Genetic variation in multilocus microsatellite genotypes in two species of woodrats (*Neotoma macrotis* and *N. fuscipes*) from California. Journal of Mammalogy 88:745-758.

Hedrick, P.W., Parker, K.M., Gutiérrez-Espeleta, G.A., Rattink, A., and Lievers, K., 2000. Major histocompatibility complex variation in the Arabian oryx. Evolution 54: 2145-2151.

Hoffmeister, D. F. 1986. Mammals of Arizona. The University of Arizona Press and Arizona Game and Fish Department, Arizona.

Hubisz, M.J., Falush, D., Stephens, M., and Pritchard, J.K. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322-1332.

Hughes, A. L. 1991. MHC polymorphisms and the design of captive breeding programs. Conservation Biology 5:249-251.

Inizan, C.C., Cajimat, M.N.B., Milazzo, M.L., Barragán-Gomez, A., Bradley, R.D., and Fulhorst, C.F. 2010. Genetic evidence for a Tacaribe serocomplex virus, Mexico. Emerging Infectious Diseases 16: 1007-1010.

Janssen, E.M., Lemmens, E.E., Wolfe, T., Christen, U., von Herrath, M.G., Schoenberger, S.P. 2003. CD4+ T cells are required for secondary expansion and memory in CD8+ T lymphocytes. Nature 421:852-856.

Kalinowski, S.T., Taper, M.L., and Marshall, T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099-1106.

Klein, J. 1986. Natural history of the major histocompatibility complex. Wiley, New York, NY.

Klein, J. 1987. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Human Immunology 19:155-162.

Klein, J. and Figuero, F. 1986. Evolution of the major histocompatibility complex. Critical Reviews in Immunology 6:295-386.

Klein, J., Sato, A., Nagl, S., and O'hUigín, C. 1998. Molecular trans-species polymorphism. Annual Review of Ecology and Systematics 29:1-21.

Kosoy, M.Y., Ellito, L.H., Ksiaxek, T.G., Fulhorst, C.F., Rollin, P.E., Childs, J.E., Mills, J.N., Maupin, G.O., and Peters, C.J. 1996. Prevalence of antibodies to arenaviruses in rodents from the southern and western United States: evidence for an arenavirus associated with the genus *Neotoma*. American Journal of Tropical Medicine and Hygiene 54:570-575.

Kotturi, M.F., Bottenz, J., Maybeno, M., Sidney, J., Glenn, J., Bui, H., Oseroff, C. Crott, S., Peters, B., Grey, H., Altmann, D.M., Buchmeier, M.J., Sette, A. 2010. Polyfunctional $CD4^+$ T cell responses to a set of pathogenic arenaviruses provide broad population coverage. Immunome Research 6:4.

Lenz, T.L., Wells, K., Pfeiffer, M., and S. Sommer, S. 2009. Diverse MHC IIB allele repertoire increases parasite resistance and body condition in the Long-tailed giant rat (*Leopoldamys sabanus*). BMC Evolutionary Biology 9:269.

Liang, Y., Shuiyun, L., and Ly, H. 2009. Molecular determinants of Pichinde Virus infection of guinea pigs: a small animal model system for arenaviral hemorrhagic fevers. Annals of the New York Academy of Sciences 1171: E65–E74.

Macêdo, R. H.,and Mares, M.A. 1988. *Neotoma albigula*. Mammalian Species 310:1-7.

Maiztegui, J.I., Fernandez, N.J., de Damilano, A.J. 1979. Efficacy of immune plasma in treatment of Argentine haemorrhagic fever and association between treatment and a late neurological syndrome. Lancet 2:1216-1217.

Maiztegui, J.I., McKee, K.T., Barrera-Oro, J.G., Harrison, L.H., Gibbs, P.H., Feuillade, M.R., Enria, D.A., Briggiler, A.M., Levis, S.C., Ambrosio, A.M., Halsey, N.A., Peters, C.J., and the AHF Study Group. 1998. Protective efficacy of a live attenuated vaccine against Argentine hemorrhagic fever. Journal of Infectious Disease 177:277-283.

Marshall, T.C., Slate, J., Kruuk, L.E.B., and Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639-655.

Martin, A.D. 2005. Bayesian Analysis. In *The Oxford Handbook of Political Methodology*, pp. 494-512. Oxford University Press, New York.

Meagher, S. and Potts, W.K. 1997. A microsatellite-based MHC genotyping system for house mice (*Mus domesticus*). Hereditas 127:75-82.

Mendez-Harclerode, F.M., Strauss, R.E., Fulhorst, C.F., Milazzo, M.L., Ruthven III, D.C., and Bradley, R.D. 2007. Molecular evidence for high levels of intrapopulation genetic diversity in woodrats (*Neotoma micropus*). Journal of Mammalogy 88:360-370.

Milazzo, M.L., Cajimat, M.N., Haynie, M.L., Abbott, K.D., Bradley, R.D., and Fulhorst, C.F., 2008. Diversity among tacaribe serocomplex viruses (family Arenaviridae) naturally associated with the white-throated woodrat (*Neotoma albigula*) in the southwestern United States. Vector-Borne and Zoonotic Diseases 8:523–540.

Milazzo M.L., Barragán-Gomez, A., Hanson, J.D., Estrada-Franco, J.G., Arellano, E., González-Cózatl, F.X., Fernández-Salas, I., Ramirez-Aguilar, F., Rogers, D.S., Bradley, R.D., and Fulhorst, C.F. 2010. Antibodies to Tacaribe serocomplex viruses (family Arenaviridae, genus *Arenavirus*) in Cricetid rodents from New Mexico, Texas, and Mexico. Vector-Borne and Zoonotic Diseases 10:629-637.

McKee Jr., K.T., Oro, J.G., Kuehne, A.I., Spisso, J.A., and Mahlandt, B.G. 1993. Safety and immunogenicity of a live-attenuated Junin (Argentine hemorrhagic fever) vaccine in rhesus macaques. American Journal of Tropical Medicine and Hygiene 48:403-411.

Mims, C.A. 1981. Vertical transmission of viruses. Microbiological Reviews 45: 267-286.

Morrissey, M.B. and Wilson, A.J. 2005. The potential costs of accounting for genotypic errors in molecular parentage analyses. Molecular Ecology 14:4111-4121.

Nisbett, R. A., Caire, W., Stuart, M.D., Caddell, G.M., Crutcher, J.M., and Calisher, C.H. 2001. Serological survey of Oklahoma rodents: Evidence of the presence of a hantavirus and an arenavirus. Proceedings of the Oklahoma Academy of Science 81:53-66.

O'Brien S. J., and Evermann, J.F. 1988. Interactive influence of infectious disease and genetic diversity in natural populations. Trends in Ecology and Evolution 3:254-259.

Oldstone, M.B.A. 1975. Relationship between major histocompatibility antigens and disease: possible associations with human arenavirus diseases. Bulletin of the World Health Organization 52: 479-486.

Orita, M., Suzuki, Y., Sekiya, T., and Hayashi, K. 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874-879.

Peters, C. J. 2002. Human infection with arenaviruses in the Americas. Current Topics in Microbiology 262:65-74.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Prichard, J.K., Wen, X., and Falush, D. 2010 Documentation for STRUCTURE software: Version 2.3. Chicago: University of Chicago.

Pfau R.S., Van Den Bussche R.A., McBee K., and Lochmiller R.L. 1999. Allelic diversity at the Mhc-DQA locus in cotton rats (*Sigmodon hispidus*) and a comparison of DQA sequences within the family Muridae (Mammalia: Rodentia). Immunogenetics 49:886-893.

Pfau, R.S., Van Den Bussche, R.A., and McBee, K. 2001. Population genetics of the hispid cotton rat (*Sigmodon hispidus*): patterns of genetic diversity at the major histocompatibility complex. Molecular Ecology 10:1939-1945.

Potts, W.K., Manning, C.J., and Wakeland, E.K. 1994. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. Philosophical Transactions of the Royal Society of London, Series B 346:369-378.

Richardson, W.B. 1943. Wood rats (*Neotoma albigula*): their growth and development. Journal of Mammalogy 24:130-143.

Richman, A. D., Herrara, L.G., and Nash, D. 2001. MHC class II beta sequence diversity in the deer mouse (*Peromyscus maniculatus*): implications for models of balancing selection. Molecular Ecology 10:2765-2773.

Richman, A. D., Herrara, L.G., Nash, D., and Schierup, M.H. 2003. Relative roles of mutation and recombination in generating allelic polymorphisms at an MHC class II locus in *Peromyscus maniculatus*. Genetics Research 82:89-99.

Rojek, J.M. and Kunz. S. 2008. Cell entry by human pathogenic arenaviruses. Cellular Microbiology 10: 828-835.

Saiki, R.K., Sharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350-1354.

Salvato, M. S., Clegg, J.C., Buchmeier, M.J., Charrel, R.N., Gonzalez, J.P., Lukashevich, I.S., Peters, C.J., Rico-Hesse, R., and Romanowski, V. 2005. Family *Arenaviridae*. In: (eds Faquet C. M., Mayo M. A., Maniloff J., Desselberger U., Ball L. A.) *Virus taxonomy: eighth report of the International Committee on Taxonomy of Viruses*, pp. 725-733. Elsevier Academic Press, New York.

She, J.X., Boehme, S.A., Wang, T.W., Bonhomme, F., and Wakeland, E.K. 1991. Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. Proceedings of the National Academy of Sciences 88: 453-457.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology 2.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731-2739.

Vallejos, D.A., Ambrosio, A.M., Fueillade, M.R., and Maiztegui, J.I. 1989. Lymphocyte subsets alteration in patients with Argentine hemorrhagic fever. Journal of Medical Virology 27:160-163.

Van Den Bussche, R.A., Ross, T.G., and Hoofer, S.R. 2002. Genetic variation at a major histocompatibility locus within and among populations of white-tailed deer (*Odocoileus virginianus*). Journal of Mammalogy 83:31-39.

Vorhies, C. T. and Taylor, W.P. 1940. Life history and ecology of the white-throated woodrat, *Neotoma albigula albigula* Hartley, in relation to grazing in Arizona. University of Arizona, College of Agriculture Technical Bulletin 86:454-529.

Walsh, P.S., Erlich, H.A., and Higuichi, R. 1992. Preferential PCR amplification of alleles: mechanisms and solutions. Genome Research1:241-250.

Weaver, S. C., Salas, R.A., de Manzione, N., Fulhorst, C.F., Duno, G., Utrera, A., Mills, J.N., Ksiazek, T.G., and Tesh, R.B. 2000. Guanarito virus (Arenaviridae) isolates from endemic and outlying localities in Venezuela: sequence comparisons among and within strains isolated from Venezualan hemorrhagic fever patients and rodents. Virology 5266:189-195.

Willey, J.M., Sherwood, L.M., and Woolverton, C.J. 2009. Prescott's Principles of microbiology. McGraw-Hill, New York, NY.