COMBINING GENETIC DIVERSITY AND SPATIO-TEMPORAL DATA TO CHARACTERIZE THE SPATIAL ECOLOGY OF ANTHRAX ACROSS MULTIPLE SCALES

By

JOCELYN MULLINS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013

To Eva and Sebastian

ACKNOWLEDGMENTS

As I write I am not only on the brink of finishing this work, but also close to moving from my home of the last 8 years, almost as long as I have been in graduate school, and so I am thinking about everyone in my village who made these years possible. Of course, no graduate degree would be begun, let alone finished, without the committee, department and faculty mentors along the way. I owe many, many thanks to my chair, Dr. Jason Blackburn. Anthrax was a dim vet school memory when Dr. Blackburn offered me a job and a graduate position during a chance meeting in the hall, but I quickly became hooked. For the humor, the patience, the spell checking, the white board sessions when it would all become clear – thanks.

Dr. Bob Cook was first my mentor in the epidemiology department and became my minor advisor on this committee. My work with Dr. Cook exposed me to areas of public health and epidemiology I would have never ventured into otherwise, and I appreciate that breadth of experience greatly. Dr. Cook kept me focused on the public health aspects of this dissertation. For these years of support and mentorship I owe a great debt of gratitude.

Dr. Michael Binford took a veterinarian in the geography department in stride, and his wealth of knowledge of, experience in and insights about spatial and ecological processes helped me immensely, with entertaining digressions along the way. Dr. Liang Mao got me thinking about anthrax transmission from a spatial perspective, and has provided fresh perspectives on the work. Dr. Matthew Van Ert provided essential expertise in genetic analysis of *Bacillus anthracis* and with a great deal of patience managed to teach this veterinarian turned ecologist a thing or two about genetics.

Thanks are owed to the many people who gathered the data and samples in the field and genotyped the collections for the niche modeling work, including the ranch and wildlife managers in West Texas and Montana, and our colleagues in Kazakhstan and Italy. Special thank go to Dr. Martin Hugh Jones for his insights into anthrax ecology and extensive knowledge of the disease system. Thanks also go to Dr. Ted Hadfield for his laboratory and genotyping expertise.

The village is larger than the university. My family has always been behind me. I thank them all. I particularly want to acknowledge my grandmother Margaret and my mother. I can sincerely say I would not be writing this without them. And Bryan – it took me long enough, but I did it. Then there are the friends and colleagues who have become my extended family, those who have been sounding boards, cheerleaders, last minute babysitters, providers of comic relief, drivers for my children, and always ready to jump in when I needed something, thank you all. Particularly Joy, my therapist, running partner, relocation specialist, travel agent and best friend, we'll have long runs and family dinners again soon, New England style. Most importantly, at the center of it all, are my children, Eva and Sebastian. They are my biggest critics and my greatest fans and the reason my world goes around. Almost done, guys.

This dissertation would not have been possible without the funding sources behind the work. The laboratory work in Kazakhstan was funded by US Defense Threat Reduction Agency through the Cooperative Biological Engagement Program in Kazakhstan. Support was also granted by Emerging Pathogens Institute at the University of Florida in the form of a seed grant, laboratory space and technical support. Chapters 5 and 6 were supported by a National Science Foundation Doctoral

Dissertation Research Improvement Grant: “Linking Population Genetics of Bacillus anthracis with Spatio-Temporal Patterns of Anthrax Outbreaks.” The final stages of the work and writing were made possible by a University of Florida Graduate School Dissertation Scholarship Award.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

COMBINING GENETIC DIVERSITY AND SPATIO-TEMPORAL DATA TO CHARACTERIZE THE SPATIAL ECOLOGY OF ANTHRAX ACROSS MULTIPLE SCALES

By

Jocelyn Mullins

August 2013

Chair: Jason Blackburn
Major: Geography

This dissertation explores the relationships between genetic diversity and ecology for *Bacillus anthracis*, the causative agent of anthrax. The work addresses two major questions about these relationships, namely, the possibility of ecological differences between genetic lineages and the use of genetic information to understand outbreak dynamics. First, ecological niche modeling using the Genetic Algorithm for Rule-Set Production (GARP) was used to compare the ecological niche of a single genetic sub-lineage of the pathogen to that of a broader collection. Sub-lineage specific niche models were then tested for transferability to novel landscapes on which the same genetic grouping was found. Second, high resolution genetic analyses were used in conjunction with spatial analyses to explore potential transmission pathways within a single, large, anthrax outbreak. High resolution genotyping was then applied to a larger collection of anthrax isolates from multiple outbreaks. These outbreaks involved diverse host groups, from single host species to combinations of wildlife and livestock, and three major genetic lineages. Results of the ecological niche modeling support ecological differences between sub-lineages of *B. anthracis* and suggest that predictive

models of anthrax persistence can be refined by characterizing the local population's genetic structure and using the dominant genotype for niche modeling. Transferred models appear to be negatively affected by either genetic-ecological differentiation at a different level of genetic analysis then employed in this study, or by methodological restraints on transferability. The results of the spatial analyses suggest that anthrax transmission events occur within a limited spatial area, and, with high resolution genetic analysis, that areas of high transmission correspond to higher genetic diversity during an outbreak. In addition, there is evidence that wildlife and domesticated animals share either a common source of *B. anthracis* infection or are in close contact to facilitate transmission during an outbreak, emphasizing the importance of anthrax surveillance in wildlife in areas where livestock is also managed. This work includes the first genetically informed ecological niche models for *B. anthracis* and the first effort to combine multiple levels of genetic analysis with spatial analyses to understand anthrax outbreak dynamics. The results should guide approaches to future additional investigations of anthrax ecology.

CHAPTER 1
INTRODUCTION: THE ROSETTA STONE OF ANTHRAX

Introduction to the Dissertation

This dissertation investigates the multi-scale spatial relationships between genetics and ecology for the pathogen *Bacillus anthracis*. This gram positive, spore forming bacterium is responsible worldwide for anthrax. The pathogen primarily affects herbivorous mammals, but can also infect a wide variety of species including humans. The disease is well known because of its rich history and for its potential as a bioterrorism agent. Anthrax remains, however, a risk to wildlife, agricultural and human health in both developed and developing parts of the world. The current paradigm for understanding anthrax outbreaks is that animals are first exposed to an infective dose of spores from the environment, although the mechanisms through which these spores become accessible in natural systems are unclear. In endemic areas cases may occur sporadically in the form of single animals in a season, or less commonly many animals are infected in a spatially limited outbreak or larger epizootic over a period of weeks. Whether animals in an outbreak are all infected from the same environmental source or the infection is transmitted from an index case to subsequent animals also remains elusive. Likewise, the pathways through which subsequent animals are infected within an outbreak are poorly understood. This body of work seeks to explore factors that are potentially part of the answers to these unknowns. Because anthrax is a complex disease system, to understand its ecology it is necessary to draw from multiple disciplines. This introductory chapter serves to lay out the foundation for the dissertation by defining and describing the several disciplines which contribute to the work, and furthermore integrate these into a framework explore anthrax's unanswered

questions. Such a framework should have applications in other disease systems as well.

Chapter 2 reports the result of the first ecological niche model (ENM) developed for a single genetic sub-lineage of *Bacillus anthracis* and compares the results to those of a model created using outbreak data representing multiple genotypes. This study seeks to understand whether this sub-lineage of the pathogen demonstrates a unique ecology. The occurrence data used in the chapter span multiple years and species and considers pathogen ecology over a relatively large spatial extent. Chapter 3 then uses previously genotyped isolates from three countries, the United States, Kazakhstan and Italy, and tests whether ENMs built using the same ecological variables as Chapter 2 can be used to create models in each country and predict anthrax occurrences in the other countries. Successful predictions would be valuable for use in targeting veterinary surveillance and human health interventions, but it is unknown if the isolate collections used for analysis are genotyped at an appropriate resolution to make accurate broad scale ecological models. Chapter 4 builds on the results of the previous chapter and investigates methods to improve the ability of ENMs to predict anthrax in other landscapes. The chapter explores methodological issues with transferred models in an effort to understand the best way to utilize this potentially powerful tool.

Moving to a finer spatial and genetic resolution, Chapter 5 will examine the genetic diversity and spatio-temporal patterns of a single outbreak to illuminate potential transmission mechanisms and evaluates the use of high resolution genetic markers in tracing the course of an outbreak. It has been hypothesized that such markers will be valuable in understanding transmission mechanisms, but this has not yet been tested in

a single outbreak. Chapter 6 describes the genetic diversity of *B.anthracis* from recent wildlife and mixed wildlife/livestock outbreaks from Texas and Montana. This study will also examine associations between genotype and vertebrate host, frequency of reported cases and spatio-temporal trends.

Approaches to the Study of Anthrax

Anthrax is a zoonotic disease, meaning it is transmissible from animals to humans, although unlike most pathogens, anthrax is not typically transmitted until after the animal is dead. Anthrax is also a saprozoonosis, meaning its life cycle involves both a mammalian host and an abiotic reservoir site, namely soil in the case of anthrax. Studies of zoonoses are necessarily complex, as there are at a minimum three players in the disease system: the pathogen, the non-human host and a human host, all interacting dynamically. Anthrax ecology involves a much larger cast of characters: the pathogen, soil, climate, vertebrate hosts, carnivores and scavengers, insects, and humans who serve both as hosts and mechanical vectors. As anyone who has managed a large team of individuals all moving in different directions can attest, this makes understanding the disease challenging at best. A number of methodologies contribute tools, techniques, processes and concepts to studies of infectious disease, including those of anthrax ecology. There is a good deal of blending and overlap of these approaches and so they are here briefly defined and clarified, and their individual contributions and interactions made apparent.

Epidemiology is defined as the study of patterns, causes and effects of disease in defined populations, or, the distribution and determinates of health related states or events in specified populations. Moreover, It is the basic science of disease prevention (Last 2000). Epidemiology is concerned with determining risk factors for a disease in

order to utilize the information to design prevention and control strategies. The overarching purpose of the science is to control disease in populations, but ultimately findings also inform clinical decision making at the individual level. A sub-discipline of epidemiology, spatial epidemiology recognizes that disease is spatially heterogeneous and seeks to identify the factors responsible. Methods used by spatial epidemiologists include visualization (mapping), exploration and modeling. Specific exploration tools used include cluster detection, pattern analysis and regression methods (Collinge and Ray 2006). Ostfeld and Glass (2005) note that spatial epidemiology "has arisen as the principal scientific discipline devoted to understanding the causes and consequences of spatial heterogeneity in infectious diseases, particularly in zoonoses" and other authors include the use of geo-referenced data in their definitions (Elliott and Wartenberg 2004). The term spatial epidemiology has been applied primarily to investigations of infectious diseases, and, within that realm, primarily to studies of zoonotic and vector borne diseases. It is worth mentioning here that when applied to purely non-human populations the term epizootiology is more appropriate than epidemiology, as is epizootic in place of epidemic to describe an occurrence of cases in excess of that expected. In the specific case of anthrax, the term outbreak is used to denote a spatially and temporally localized increase in cases (which may be as few as one case) and epizootic is used when the disease occurs over a larger region or temporal window.

The discipline of ecology studies of interactions between living organisms and their biotic and abiotic environments (Odum 1959). The definition has been adapted to describe human ecology (the relationships between humans and their environment), social ecology (relationships between individuals, social groups and their environment),

political ecology (relationships between political, economic and social factors with environmental issues and changes), and human ecology of disease (ways in which behavior in its cultural and socio-economic contexts interacts with environment to affect disease) (Meade and Emch 2010). Spatial ecology, similar to spatial epidemiology, is a logical extension of the broadly defined discipline of ecology. Spatial ecology focuses on identifying and understanding spatial patterns of ecological phenomena, and landscape ecology is specifically concerned with understanding the relationships between the biological requirements of a species and where on the landscape ecological conditions and processes satisfy those requirements and the spatial scale at which these processes operate (Collinge and Ray 2006). Each of these sub-disciplines has a role in understanding disease processes, including those of anthrax.

Disease is, at its simplest, an interaction between a living organism (the pathogen) and the biotic and abiotic environments, so it is not surprising that the toolsets and methodologies of ecology have been used to study disease. Disease ecology is a term that was originally used by geographers to differentiate studies of health service delivery from those concerned with other disease studies, including diffusion patterns, mapping and multi-factorial analyses (Meade and Emch 2010). In this context, disease ecology included relationships of humans with the biological, physical and socio-cultural environments in models of disease, but rarely included actual environmental interactions or ecological processes. Today disease ecology is more closely aligned with traditional ecology. The field emphasizes the underlying processes through which populations interact with each other to influence the incidence, spatial distribution, and timing of diseases. While epidemiology focuses on

characteristics of specific diseases, this sub-discipline is more concerned with broader processes shaping disease occurrence and transmission. However, both disease ecology and spatial epidemiology have a historical context in Pavlovsky's concept of natural disease nidality and the early medical geographer Jacques May's factors, the overlap of which define zones in which humans may contact pathogens. Similar to the ideas of May, disease ecology views the dynamics of infectious disease as occurring within the overlap of the niches of the component populations in time and space. Disease ecology and epidemiology have different approaches and focus, but inform one another. Given the complexity of zoonotic disease systems, both disciplines are required to achieve a thorough knowledge of specific risk factors for and characteristics of a disease on the one hand, and of the individual ecologies of all the players of the system on the other. These are combined in order to understand how, where and when the disease is acquired and how it is best controlled. Collinge and Ray (2006) used the term community ecology to describe the species interactions underlying disease dynamics and emphasized the importance of the interface between epidemiology and community ecology. Spatial epidemiology and spatial ecology utilize geographic information systems (GIS) which is a powerful tool with which to explore spatial patterns and analyses.

The advent of molecular typing techniques reveals a new level of spatial complexity in the study of infectious disease. Genetics can be used to understand pathogen transmission, epidemiology and the physical or environmental drivers of disease (Archie et al. 2009). We now know that many bacterial pathogens have a spatial genetic structure (Smith et al. 2000, Girard et al. 2004, French et al. 2005,

Achtman 2008, Kugeler et al. 2009, Nakazawa et al. 2010, Pepperell et al. 2011), and genetic diversity has been linked to epidemiologic and ecologic patterns (Smith et al. 2000, Girard et al. 2004, Kugeler et al. 2009, Nakazawa et al. 2010). With evidence that genotype may play a substantial role in the processes underlying disease distribution in space and time, genetic analysis is proving to be a powerful tool with which to understand spatial patterns of disease. Two approaches can be used to understand the relationships between geography, ecology and genetics. Phylogeography, at its simplest, describes the geographic distribution of molecular lineages to understand patterns of evolution. This is usually undertaken at spatial macro-scales and across discretely defined populations. (Avise 1998, Kidd and Ritchie 2006, Knowles 2009, Hickerson et al. 2010). Landscape genetics borrows from the tradition of landscape ecology and molecular population genetics to study disease patterns operating at fine (meso to micro) scales (Manel et al. 2003, Storfer et al. 2006). Landscape genetics does not require an a priori definition of the population. In addition, this methodology seeks to understand movement of individuals, or pathogens, through analysis of spatial and temporal genetic patterns.

Anthrax

To study a disease system effectively one must be familiar with the biology of the pathogen as well as all the known and potential players, determinates and confounders in the disease system. Only in this way can the researcher ask relevant questions which result in applicable findings. Infections caused by *Bacillus anthracis* have been documented since ancient times (Sternbach 2003). It was one of the earliest bacteria to be studied, and has a special place in public health history as the first microorganism proven to cause a disease (Bardell 2002). First identified by the German physician

Robert Koch, the discovery of the anthrax bacilli led to the well-known Henle-Koch postulates (Morens 2003), used to establish disease causality and memorized by all beginning epidemiology students. Today, most citizens of developed countries know anthrax only as a biological weapon threat and are surprised to discover that not only is anthrax endemic throughout the modern world, it is a very real veterinary and human public health threat in many countries, perhaps including their own.

A Day in the Life of *Bacillus anthracis*

Bacillus anthracis is an aerobic or facultatitvely anaerobic Gram-positive rod which can be thought of as having a two parts to its life cycle. The bacteria are metabolically active, or vegetative, and divide in the mammalian host during the infective portion of the life cycle. After the host dies the cells form spores when exposed to nutrient-deficient environments outside the carcass. These spores are resistant to environmental conditions and can survive, metabolically dormant, in the soil for some time. The spore interacts with the abiotic environment through the exposporium, the biological 'shell' surrounding the spore. The degree to which spores are able to maintain viable is influenced by soil and environmental conditions. Associations of anthrax occurrence with humidity, alkaline soils, high soil moisture, high organic content, and calcerous soils (Van Ness 1956, 1971, Dragon and Rennie 1995) have been qualitatively described and supported by recent modeling efforts (Blackburn et al. 2007) as well as by physiological evidence (Williams et al. 2013). The chemistry of the spore and the its interactions with environmental conditions are well summarized in Hugh-Jones and Blackburn (2009). In brief, alkaline environments appear to buffer the spore against calcium loss and increase longevity (Dragon and Rennie 1995, Hugh-Jones and Blackburn 2009). Alkaline soils will also cause the exosporium to be

negatively charged and more able to adhere to particles in the soil, making persistence and subsequent uptake by a grazing animal more likely (Hugh-Jones and Blackburn 2009). Because of the high natural buoyancy of the hydrophobic spores (Dragon and Rennie 1995) they are able to float to and adhere to grasses as flood water recedes. In vitro work showed that the exposporium is not only essential to hydrophobicity and adherence to soil material, but also that spores were significantly more likely to adhere to soils having higher organic material and calcium content (Williams et al. 2013). Recently, experimental evidence has demonstrated the ability of B. antrhracis to replicate in rhizosperes of grasses and amoebas (Saile and Koehler 2006, Schuch and Fischetti 2009, Dey et al. 2012). These results have yet to be confirmed in the natural environment, but present intriguing possibilities for the how we think about ecology of anthrax.

Assuming, for the moment, that spores remain metabolically inactive in the soil or adhered to grass or leaves, herbivorous animals are exposed through ingestion, inhalation, or through breaks in the skin. Most exposures among wildlife and livestock takes place through ingestion of spores on vegetation during grazing or browsing, but inhalation exposure has also been documented (Beyer and Turnbull 2009). It is possible that ingested spores can gain entry through micro-abrasions in the oro-pharyngeal or esophageal mucosa or elsewhere within the gastrointestinal tract. Incubation in ruminants is approximately 3 to 5 days, depending on the size of the spore inoculum, the species and the route of exposure. Having gained access to the host, spores germinate and rapidly divide in the host's phagocytic cells. Virulent strains secrete a three-protein toxin which is speculated to be the cause of death of the host in

susceptible species, while bacterial septicemia is fatal in species more resistant to the toxin (Hugh-Jones and de Vos 2002). Together, protective antigen, edema factor and lethal factor result in the typical fulminant clinical signs associated with anthrax. Meanwhile, capsular antigens protect the bacterium from phagocytosis by the host immune system.

The route of exposure determines the clinical form of anthrax, namely, gastrointestinal, inhalational (fulminant) and cutaneous. The course of illness is determined by a combination of the size of the spore inoculum, host susceptibility, and immunological state of the host and is described as peracute (extremely fast), acute (fast) or chronic. Ruminant species most commonly experience the most severe peracute form and rarely survive (Dragon and Rennie 1995). These animals exhibit staggering, fever, excitation and then recumbency prior to death. There may also be cyanosis, dyspnea, severe submandibular swelling and hemorrhage, and dying animals may seek shade or water because of pyrexia. Death occurs rapidly in susceptible species and dead animals have a characteristic “saw horse” appearance. After death, vegetative cells are spilled into the environment through leakage of edema and blood and when the carcass is opened during scavenging. Sporulation occurs primarily in the soil, although this process depends on ambient temperatures being above 9 to 12 degrees (Beyer and Turnbull 2009). In an intact carcass vegetative cells will lose to competition from anaerobic bacteria during the decomposition process. Exclusion of carnivores, however, does not appear to significantly reduce ground contamination (Bellan et al. 2013b). A recent study in which carcasses were protected from scavengers found no difference in spore counts in the soil up to 4 days after death as

compared to carcasses left accessible (Bellan et al. 2013b). The authors noted the presence of carcass fluids on the soil other than terminal hemorrhage and speculate that decomposition may result in small tears in the skin that allow for leakage of fluids or air to enter the carcass and facilitate sporulation.

How it Plays Out: Epizootiology, Epidemiology, Prevention and Control

Epizootiology

All vertebrates can be infected with *B. anthracis*, but herbivores and ruminants in particular are most susceptible to disease. Carnivores and omnivores may have serological evidence of exposure and clinical illness has been documented in these species, but overall appear to have relatively low morbidity and mortality (Creel et al. 1995, Lembo et al.). Sero-prevalence among carnivores may be high in areas with high levels of transmission and in areas of high anthrax activity sero-convert at a young age (Hampson et al. 2011). Measuring sero-prevalence in carnivores and scavenger species, therefore, provides clues to the frequency of disease among the herbivores in an area and they are useful as sentinel species (Turnbull et al. 1992, Hampson et al. 2011, Lembo et al.). Even among susceptible herbivores, however, serological studies indicate exposure without significant morbidity or mortality can occur (Lembo et al. 2011) . In the Serengeti ecosystem, for example, wildebeest and buffalo were found to be seropositive, indicating they were infected and either did not become ill or recovered. In contrast, no seropositive zebra were identified, suggesting the disease in this species was uniformly fatal (Lembo et al.).

Exactly how animals become infected and how and why an epizootic is initiated and perpetuated is still speculation (Beyer and Turnbull 2009) Outbreaks may be initiated by increase in exposure due to increased availability of spores, changes in

animal behavior, changes in host susceptibility, or variations in concentrations of infective *B. anthracis* in the environment as a result of a life cycle outside of the host (Turner et al. 2013). There is similarly limited understanding of disease dynamics within an outbreak. A number of consistent observations provide clues or conundrums, depending on one's point of view. Outbreaks tend to be spatially localized (Hugh-Jones and de Vos 2002, Hugh-Jones and Blackburn 2009), perhaps indicating a focal area either suitable for the pathogen to be available to the host or to be transmitted, or transmission pathways are spatially limited. The species affected varies within and among outbreaks in a given landscape (Hugh-Jones and Blackburn 2009, Hampson et al. 2011, Lembo et al. , Turner et al. 2013). For example, within the Etosha National Park, peak anthrax mortality occurs in the dry season for zebra and springbok, but in the wet season among elephants (Turner et al. 2013) whereas in the Serengeti, Tanzania, grazers such as zebra, wildebeest, buffalo and livestock are most affected after droughts and browsers following heavy rains (Hampson et al. 2011). Additionally, the seasonality of epizootics varies between ecosystems. Anthrax is a dry season disease in some African ecosystems (Prins and Weyerhaeuser 1987, Turnbull et al. 1991, Lindeque and Turnbull 1994) and in North America (Hugh-Jones and de Vos 2002, Hugh-Jones and Blackburn 2009, Epp et al. 2010a), and a wet season disease in other African areas. In dry season ecosystems, and notably in North America, reports of large epizootics describe the onset of cases occurring during a period of hot weather that followed a season of higher than average rain (Gates et al. 1995, Dragon et al. 1999, Hugh-Jones and de Vos 2002, Ndiva Mongoh et al. 2008, Epp et al. 2010b).

Several hypotheses have been proposed for this pattern. The period of wet weather may float the buoyant spores and subsequent runoff distributed the spores which adhere to vegetation as the waters evaporate or recede. Second, hot dry weather may alter the immune status of hosts through increased insect burden, overcrowding of animals around resources and heat stress (Gainer and Saunders 1989). Dry vegetation may abrade oral and gastrointestinal mucosa of herbivores, facilitating entry of spores (Beyer and Turnbull 2009). Most outbreaks report no difference in gender among animals killed by anthrax, except among Canadian bison where males are disproportionately affected. These bison bulls tend to gather in wallows, a behavior which may explain greater burden of disease among males (Gates et al. 1995). Increased soil ingestion because of either flooding or short dry grass may increase exposure to spores, although neither soil ingestion nor nutrition status alone could explain the seasonality of epizootics in three species in Etosha National Park, Namibia, suggesting that other ecological, pathogen or host factors act to drive outbreaks (Turner et al. 2013).

Unfortunately, models of other livestock or zoonotic diseases have little application to anthrax, primarily because animals are not capable of transmitting the disease until after death. Furniss and Hahn (1981) attempted to use deterministic models to determine if an outbreak was driven by environmental contamination (contamination hypothesis) or by susceptibles contacting a fresh carcass (classic hypothesis). A threshold population for perpetuation of the outbreak was identified under the contamination hypothesis, but neither model fit the from the actual outbreak data well, leading the researchers to suggest that either not all animals in the population

were susceptible, not all spores were infectious, or the spatial pattern of contamination was more complex than modeled. The authors propose that for unknown reasons some carcass sites might remain infectious longer. A subsequent paper later identified a threshold under the environmental hypothesis (Hahn and Furniss 1983). It was noted that in these models prompt removal of carcasses can slow outbreak (Friedman and Yakubu 2012). Recently the Hahn and Furniss models were extended to include birth, death and migration to model anthrax epizootics in migratory populations (Friedman and Yakubu 2012).

There are a few more players in anthrax ecology that should be introduced. Spatio-temporal characteristics of outbreaks and experimental evidence suggest that flies play a role in anthrax ecology and transmission (Hugh-Jones and Blackburn 2009). Necrophagous flies, also known as blowflies, feed on carcasses of animals that have died from anthrax and then carry vegetative cells or spores to nearby vegetation where they vomit or defecate, leaving contaminated leaves for browsers to consume (Hugh-Jones and Blackburn 2009, Blackburn 2010). Biting flies, or hematophagous flies, may be capable of mechanically transmitting infections from bacteremic animals or carcass tissues to healthy ones, and such transmission may in theory occur over larger distances (Hugh-Jones and Blackburn 2009). The clinical characteristics of an outbreak affecting a large proportion of horses in Italy, specifically massive local subcutaneous edema consistent with an injection inoculum, further suggests that hematophagous flies play a role in perpetuating epizootics (Palazzo et al. 2012). In addition to flies, there are the scavengers. Vultures are often the first on the scene to consume tissues that are full of vegetative cells and upon opening the carcass begin the sporulation process.

Spores have been isolated from vulture feces in the vicinity of anthrax infected carcasses (Lindeque and Turnbull 1994), but whether vultures play a role in perpetuating outbreaks is unknown. Similarly, spores have been found in mammalian feces (Lindeque and Turnbull 1994). In ecosystems where mammalian scavengers are active they will open carcasses to consume infected tissues and furthermore can separate the carcass, enlarging the spatial area of potential contamination. And finally, but by no means least importantly, the genetics of *B. anthracis* are likely another key component to the puzzle of anthrax. In Kruger National Park, South Africa, the coarsely defined A and B lineage appeared to have different tolerances for soil pH and calcium level, with the A occupying a wider range of values than the B group, which was also more spatially restricted (Smith et al. 2000). Pathogen genetics may be one of the unknown factors behind spatial heterogeneity and host-pathogen-ecology interactions.

Epidemiology

Clinical anthrax in humans not associated with bioterrorism occurs mainly in two groups of human individuals: agricultural workers and industrial workers. Veterinarians, agricultural workers and farmers can be exposed to infected carcasses during necropsy, butchering and slaughtering (Steele and Helvig 1953, Woods et al. 2004, Akbayram et al. 2010, Gombe et al. 2010, Özden et al.). These exposures are typically cutaneous and treated effectively with antibiotics, but gastrointestinal exposure can also occur in this population (Gombe et al. 2010). In pastoral societies infection is associated with butchering, slaughtering and eating undercooked or raw meat and the epidemiology may be dictated in part by cultural practices. In Kazakhstan, Zimbabwe, West Bengal and Turkey most adult livestock associated infections were cutaneous resulting from butchering and slaughtering (Woods et al. 2004, Karahocagil et al. 2008, Gombe et al.

2010, Özden et al.). In contrast most reported cases in Zambia were gastrointestinal (Siamudaala et al. 2006), although this finding may be a result of only the sickest individuals seeking medical care and being reported to public health authorities. Cutaneous infection was also associated with having cuts on the hands when animals were handled (Woods et al. 2004, Chirundu et al. 2009). Human anthrax in Haiti is primarily cutaneous and associated with butchering, however, it is somewhat unique in that children are at equal risk as adults, and lesions on the face are common (Peck and Fitzgerald 2007). Researchers note that it is a common practice in Haiti to blow into small holes in the skin to facilitate skinning animals (Peck and Fitzgerald 2007), and tanning of hides often involves holding on end of the hide in the mouth. Industrial exposures occur primarily among workers processing hides and hair of infected animals. These exposures are typically inhalational or cutaneous and exposure in some individuals may result in sub-clinical infection (Bales et al. 2002).

A review of anthrax cases in the United States showed that cutaneous exposure was most commonly reported among butchers, ranchers and veterinarians (Bales et al. 2002). In contrast, inhalation exposure was more common in textile mill workers. A weak immune system was speculated to predispose some individuals to illness following these occupational exposures. Outside of these two main groups of at risk people, cases of gastrointestinal, inhalation and cutaneous anthrax have been reported over the past decade in the United States resulting from exposures to spores in the animal hides on imported drums (Anaraki et al. 2008, Guh et al. 2010, Lamothe 2011), and outbreaks of anthrax resulting from contaminated heroin have been reported in populations of intravenous drug users (Price et al. 2012). These populations represent

the ‘normal’ epidemiological presentation of anthrax among humans in very specific at risk populations for contrast with the high numbers of inhalational anthrax that would occur in a bioterrorism events.

Prevention and control

It cannot be over-emphasized that successful control of anthrax is dependent on active surveillance and pro-active vaccination. Unfortunately anthrax is prone to under-reporting, under-diagnosis and misdiagnosis. Wildlife deaths in remote areas can remain un-detected, and the appearance of the sick animal or carcass may mimic other diseases such as clostridial infection (blackleg). There may be economic and social pressures to not disclose livestock deaths and the high value of cattle in rural pastoralist communities means that these animals may still be butchered and consumed even when there is an awareness of anthrax (Peck and Fitzgerald 2007, Chikerema et al. 2013). In fact, knowledge that cutaneous anthrax is easily treatable may encourage butchering of potentially infected carcasses (Chikerema et al. 2013). Environmental and herd health management appear to play a role in anthrax epizootics in domesticated animals. Flooding in pastures, wet pasture, short grass and density of animals were associated with increased risk of cases in a multi-farm outbreak in Saskatchewan (Epp et al. 2010b). Preventive annual vaccination or timely vaccine administration after the start of an outbreak is important for control in domestic animals (Ndiva Mongoh et al. 2008, Epp et al. 2010b, Chikerema et al. 2012). In the Saskatchewan outbreak, a delay in vaccination of more than one week from the first cases resulted in an increase in risk of a farm experiencing cases (Epp et al. 2010b), and in a North Dakota outbreak, vaccination administered to a herd prior to the outbreak was protective as compared to vaccination during the outbreak (Ndiva Mongoh et al.

2008). Further evidence of anthrax incidence being related to vaccine policy comes from Zimbabwe (Pfukenyi et al. 2011), where over multiple years reports of anthrax infection fluctuated over a roughly 20 year period, with increases coinciding with periods of decreased vaccination due to war or government policy, and decreases in periods of active vaccination (Chikerema et al. 2012).

Unfortunately, vaccination still occurs mostly in reaction to case reports (Blackburn et al. 2007). It is particularly difficult to vaccinate livestock in rural areas with pastoral cultures because settlements and animals are spread over large, difficult to access areas where resources for both vaccine and personnel are scarce. Political turmoil, as seen in Zimbabwe, reduces resources and access to rural areas. In addition, because of the reactive nature of vaccination campaigns, some populations may believe the vaccine actually causes outbreaks (Siamudaala et al. 2006). Effective control is also dependent on quarantine, adequate disposal by burning, and good husbandry, which all may be difficult in resource poor communities; for example, lack of firewood and competing economic pressures interfere with burning carcasses (Siamudaala et al. 2006).

What Lies Beneath: Uncovering the Genetics of *Bacillus anthracis*.

Bacillus anthracis is metabolically active during the short period between infecting and killing the host, and since it does its job of killing the host extremely well there may be as few as only 20-40 doublings occur during the infection cycle (Keim et al. 2004). Depending on the local ecology, infections of suitable hosts may occur only once to a few times per year or less. Minimal to no replication is currently understood to occur while the organism is in the spore stage, although there is a suggestion that the bacteria may be capable of replicating in rhizospheres of grass plants (Saile and

Koehler 2006) and amoeba (Dey et al. 2012). Although the amount of replication that occurs in the environment and the importance of this to genetic diversity is unclear, the relatively limited opportunity for replication in mammalian hosts means the genome of *Bacillus anthracis* is highly conserved relative to other bacterial species. This homogeneity confounded efforts to understand the genetic diversity of the organism until within the past 15 years, when researchers identified portions of the genome that could be utilized for strain typing. The efforts to find methods for strain typing *B. anthracis* were spurred by the 2001 anthrax attacks when the need for molecular trace-back was essential to determining the source of the spore preparation contained in the famous letters. Molecular trace-back refers to using the genetic signature of the pathogen isolated from a case or outbreak to match that strain to a source.

Long distance movement of Bacillus anthracis is facilitated by humans via transportation of contaminated animal products, or in nomadic populations, possibly via infected animals. Introductions to new environments will often involve the one or few strains that had infected the animal or animals. Therefore, regional population patterns of genetic diversity are probably strongly influenced by Founder's effects (population bottlenecks) and genetic drift, which may explain the heterogeneous spatial distribution of genetic lineages. While mutation rate is probably the most important factor affecting genetic diversity in B. anthracis (Keim et al. 2004), how diversity is affected by selection within hosts or in the environment is unknown.

Genetic Techniques Used For *Bacillus anthracis*

Genetic analyses are performed to describing an isolate (the bacteria isolated from an animal or location), for molecular trace-back (using genetic characteristics to identify the source of an outbreak) and to describe and compare the genetic

characteristics of populations. They may also be used to enhance understanding of a disease's epidemiology and ecology. Three primary analytical systems are currently used to characterize individual isolates of *B. anthracis*. They are based on three sets of genetic markers: single nucleotide polymorphisms (SNPs), variable number tadem repeats (VNTRs) and single nucleotide repeats (SNRs). These three systems capitalize on diversity at specific places on the genome, or loci, that are prone to mutations which result in varying alleles (variants on a locus). Some alleles may appear at random in the population and are perpetuated without selection because they do not affect the function or appearance of the organism. Other alleles may increase or decrease the ability of the individual to survive and reproduce or replicate, in which case the allele will appear more or less frequently in the population, respectively. This selection process can occur in the host or in the environment. The usefulness of each system to the study anthrax ecology depends on the spatial-temporal scale being examined – or, in other words, on what question is being asked (Keim et al. 2004).

Single nucleotide polymorphisms. Single Nucleotide Polymorphisms (SNPs) are generated by substitutions of one nucleotide character for another during replication. They are relatively rare in the *B. anthracis* genome, such that the mutation rate is given as 10 -10 (Keim et al. 2004). Four allele states are possible at each locus and the maximum diversity value is 0.75 for each marker. Because SNPs are slowly evolving, mutations which likely occurred once in phylogenetic history are very unlikely to mutate to novel or ancestral states (eg, they exhibit low homoplasy), although recent findings suggest homoplasy in these markers can be detected (Antwerpen et al. 2011). For these reasons SNPs are used to define deep relationships and branches in the

phylogenetic tree. While multiple SNPs can be found on a phylogenetic branch, a small number is utilized to define major groupings (Van Ert et al. 2007a). This set of canonical SNPs ("canSNPs") was developed specifically to place a given isolate on a phylogenetic branch.

SNPs are useful for understanding genetic diversity at large spatial and temporal scales, such as global or continental historical collections and for placing isolates into the global context as defined by Van Ert et al. (Van Ert et al. 2007a) or to trace isolates to a broad geographic area. For example, a set of SNPs differentiates the epidemiologically important Ames lineage from others groups, and a specific SNP on the pXO1 plasmid can differentiate the China Ames strain from those from Texas (Van Ert et al. 2007b). The genetic scale of SNPs is too coarse to detect diversity in smaller geographic areas, although when diversity is found at such a spatial scale interesting questions about multiple introductions and the history of anthrax in that region may arise (Garofolo et al. 2011). The genetic diversity of *B. anthracis* on a global scale was mapped using the canSNPs (Van Ert et al. 2007a). In this study it appeared that the A lineage was the most common and widespread, although there were distinct regional differences in predominant sub-groups. SNPs have since been used to characterize the population genetics in North America (Kenefic et al. 2009), China (Simonson et al. 2009), Italy (Garofolo et al. 2011) and Kazakhstan (Aikembayev et al. 2010). Three SNP types were identified in northern Italy, including the trans-Eurasian, an A sub-group (005/006) primarily found in Africa, and B.Br.CNEVA which is found in contiguous parts of Europe (Garofolo et al. 2011).

SNPs have been used to describe the genetic diversity of the Brucella genus, *Yersinia pestis* and *Fransicella tularensis*, which are zoonotic bacterial pathogens causing brucellosis, plague and tularemia, respectively. It useful to compare findings between studies of these pathogens and those of *B. anthracis* are because all have spatially heterogenous genetic population structures, conserved genomes and complex local ecologies. Phylogenetic relationships within the Brucella genus and *F. tularensis* have been described using SNPs (Foster et al. 2009, Vogler et al. 2009). SNP markers were combined with ten years of epidemiological to investigate plague in Madagascar, and in that study genetic and epidemiologic patterns were interpreted as being consistent with pathogen transmission events occurring from rural to urban areas, which failed to result in establishment of the pathogen in urban areas (Vogler et al. 2013).

Multiple locus variable number tandem repeat analysis. Analysis of amplified fragment length polymorphisms (AFLPs), a variety of mutable loci, useful for differentiating strains of *B. anthracis* led to the discovery that the variations in fragment lengths were frequently a result of variable number tandem repeats (VNTRs) (Henderson et al. 1994, Harrell et al. 1995, Andersen et al. 1996, Jackson et al. 1997, Keim et al. 1997). Variable number tandem repeats are short, repetitive nucleotide sequences which can vary in repetition number. VNTRs are more diverse than SNPs because more allele states are possible and in addition they have a higher rate of mutation, on the order of less than 10 -5 to over 10-4 (Keim et al. 2004). Because less diverse loci provide information on deeper evolutionary relationships and more diverse loci reveal more recent evolutionary change, this wide range enables selection of collections of markers of various diversity values to highlight deeper and more recent

phylogenetic differences. The analysis of these sets of VNTRs is termed multiple-locus VNTR Analysis (MLVA). MLVA typing systems for *B. anthracis* having 8 (Keim et al. 2000), 15 (Keim et al. 2004), 25 (Lista et al. 2006) and 31 (Beyer et al. 2012) markers are currently in use. MLVA typing has been used for epidemiological trace back for anthrax outbreaks (Keim et al. 2001, Hoffmaster et al. 2002, Cheung et al. 2005, Marston et al. 2011, Fasanella et al. 2012), to characterize and map diversity in multiple regions, countries and ecosystems (Table 1-1), for demonstrating genetic-ecological associations discussed above (Smith et al. 2000), and for ecological niche modeling (Chapter 2). At least one *B. anthracis* VNTRs is associated with phenotypic differences (Sylvestre et al. 2003), although no VNTRs or MLVA types have of yet been found to have epidemiological significance.

Although VNTRs, like SNPs, are unique to each organism, a review of MLVA use in other pathogens is again useful. A 16 marker MLVA system was developed to speciate and further characterize members of the Brucella genus (Huynh et al. 2008), and this system was then used to link isolates and genetic groups to specific geographic areas to show that strains clustering together phylogenetically also tended to have common geography (Al Dahouk et al. 2007). Tumaremia is a widely distributed disease that can be transmitted through ticks and contact with infected animals or tissues. A six marker system identified two major genetic groups of F. tularensis, of which Biovar A was more common and diverse in North America than the B group, and the groups appeared to overlap geographically and temporally (Farlow et al. 2001). The A group was further subdivided using a 25 marker system (Johansson et al. 2004) and a later study describing the phylogeography of F. tularensis in the United States

showed that subgroups of A tended to be located in distinct parts of the continent (Farlow et al. 2005). Furthermore, these genetic subgroups were found to have different epidemiological characteristics which may be related to differences in transmission pathways such as vector versus lagomorph exposure (Staples et al. 2006). Tying the different spatial and epidemiological associations found in North America together, ecological niche models indicate that the major genetic groups of F. tularensis have different ecological preferences (Nakazawa et al. 2010).

MLVA has also been used to make inferences about tularemia disease ecology. For example, multiple F. tularensis clades were identified in an outbreak in Utah mediated by deer flies (Petersen et al. 2008). The authors of the study suggested that the outbreak was driven by ecological dynamics which created an ideal environment for the pathogen, as opposed to a single index case followed by transmission to subsequent cases. Other areas, however, show different genetic spatial patterns. Analysis of the disease in Sweden showed clustering of genotypes often in very small areas which were stable within and between seasons (Svensson et al. 2009) and while in France one of four major MLVA groups was widely distributed, the three less common groups appeared to have no geographic pattern at a regional scale (Vogler et al. 2011a). These findings all suggest that the ecology of tularemia, like anthrax, is closely tied to micro-scale landscape features and that population genetics may be essential to understanding the disease.

Plague, also like anthrax, is an ancient disease that remains active in foci around the world. MLVA was used to characterize the landscape level population structure of Y. pestis during an outbreak in prairie dogs (Girard et al. 2004). The spatio-temporal

patterns were interpreted to indicate that plague, in this ecosystem, underwent rapid expansion when introduced to a susceptible population and then a slower persistence phase. The genetic pattern also suggested that barriers to prairie dog dispersal, such as forested habitat, between populations were associated with greater genetic differentiation in Y. pestis in these groups. The same pattern was also observed at the global scale, this time using SNPs (Morelli et al. 2010). Interestingly, a similar pattern of expansion and low level endemnicity was described in an analysis of the dispersal of M. tuberculosis in Canada (Pepperell et al. 2011). MLVA and SNPs have been used to describe the phylogeography of plague in Madagascar (Vogler et al. 2011b). The ecological differences of lineages of Y. pestis have not yet been explored.

Single nucleotide repeats. Single Nucleotide Repeats (SNRs) are a class of VNTRs made up of sequences of repeated nucleotides which have very high mutation rates. The discoveries of two systems, each using four markers, are described (Stratilo et al. 2006, Kenefic et al. 2008b). The four loci described by Kenefic et al. (2008b), which have diversity values of 0.80 – 0.94, has practical benefits over the system described by Stratilo (2006) as well as possibly greater resolving power as demonstrated in an Italian outbreak (Garofolo et al. 2010). Because of the high potential for homoplasy, SNRs cannot be used to discriminate among diverse SNP or MLVA defined isolates. For the same reason, they should not be used to compare strains of the same MLVA type found in different geographic areas unless there is reason to suspect a temporally close epidemiological link.

The greatest potential utility of SNRs is to discriminate between isolates in an outbreak scenario in which strains are known to be identical on SNP and MLVA (Keim

et al. 2004). Using these mutable markers it may be possible to generate or confirm (or refute) hypotheses about mechanisms of pathogen movement across the landscape. Diversity in SNRs in a population should correlate with transmission levels such that high diversity can be interpreted to indicate a high level of anthrax activity. SNR differences have been described within single outbreaks (Kenefic et al. 2008a, Garofolo et al. 2010, Fasanella et al. 2012). Kenefic et al. (2008a) found six SNR types in 47 isolates from a South Dakota outbreak. Of these, 30 were a single SNR genotype noted as SGT-1. The other five isolates differed by a single base pair and appeared to be randomly distributed in both space and time. The authors interpreted the SNR diversity and random spatial distribution as being a result of multiple past outbreaks. A similar pattern was found in the analysis of 53 isolates from an epidemic in Italy (Garofolo et al. 2010). The most common genotype in the Italian epidemic, SD-1, appeared to give rise to three additional genotypes which each differed by 1 base pair from SD-1. A fifth genotype differed from SD-1 at three loci; the authors speculate that this strain may have resulted from a past outbreak whereas the more closely related genotypes were random events. Demonstrating that SNR diversity can result from an introduction of a single genotype, Fasanella et al. (Fasanella et al. 2012) found two SNR genotypes at two sampling sites during an outbreak in Bangladesh which was strongly suspected to be a result of contaminated feed.

Although the spatial patterns of genotypes in an outbreak have been depicted, SNRs have not yet been used to model the movement of *B. anthracis* in time and space within an outbreak or across multiple outbreaks in an epizootic (Kenefic et al. 2008a, Fasanella et al. 2010, Fasanella et al. 2012, Stratilo and Bader 2012). For SNRs to be

useful in studying transmission mechanisms it is assumed that a new SNR genotype evolves as a single event when an animal is infected and the bacilli is actively replicating. The locus (or loci) mutates during passage through the animal and the new genotype is then shed into the environment to infect a subsequent animal. That SNR type could be followed through the outbreak. It appears, however, that SNR analysis may in fact complicate epidemiological and ecological studies (Beyer et al. 2012, Stratilo and Bader 2012). As an example, analysis of SNR types of single MLVA-31 genotyped isolates from multiple years in Namibia demonstrates that SNR diversity needs to be considered in the context of the epidemiological situation (Beyer et al. 2012). Specifically, identical SNR types could trace the infection of four cheetahs to a single source, but diverse SNR genotypes on a single farm did not alter the conclusions drawn from MLVA analysis and epidemiological evidence. The finding of MLVA and SNR diversity in multiple soil samples from single carcass sites (Stratilo and Bader 2012) further complicates the SNR picture and suggests that the interpretation of SNR diversity will be informative only at fine scales or at the individual level. More work is needed to understand how to utilize these rapidly evolving markers.

It is important to understand that mutlilocus genotypes are only meaningful when compared to other individuals or populations. This is in contrast to ecological or epidemiological data, which are meaningful independently of other values. In addition, comparisons of genotypes within and between populations is limited by what has been termed “discovery bias” (Pearson et al. 2004). Phylogenetic trees, which graphically illustrate genetic relationships, are built using isolates available for genotyping at the time of the analysis, and subsequent studies attempt to place new isolates into this

framework. However, if the entire diversity of the population is not represented in the original set of isolates then the phylogenetic tree will be biased towards the genotypes present, and subsequent genotypes which were not represented in the original tree will be misplaced or interpreted out of context (Pearson et al. 2004, Pearman et al. 2008). It is therefore equally important to capture the full global diversity of *B. anthracis* as well as local diversity in order to answer questions about spatial and genetic patterns.

Evolutionary scale versus spatial scale

The ultimate utility of molecular markers, in any species, depends on mutation rates and the population being examined (Keim et al. 2004, Biek et al. 2007, Archie et al. 2009). There can be too much of a good thing, in that high levels of genetic diversity can obscure phylogenetic patterns and therefore spatial and epidemiological patterns as well (Keim et al. 2004, Beyer et al. 2012). To provide structure to the interpretation of genetic diversity in *B. anthracis*, a nested hierarchical approach has been proposed. Progressive hierarchical resolving assays using nucleic acids (PHRANA) has been proposed as a means of reducing the diversity of a population in a stepwise manner using the three systems described above (Keim et al. 2004). This scheme has been interpreted to mean that high SNR diversity among identical SNP and MLVA defined groups of isolates results when strains from multiple past outbreaks are active (Kenefic et al. 2008a). It would follow that higher than expected SNR diversity indicates higher sporadic disease activity in non-epidemic seasons than recorded, or that mortalities during an epidemic are more numerous than documented. Ecologically, it would also follow that if strains from multiple past outbreaks are active in a single epidemic it is less likely that a single index case triggers an outbreak, but instead suggests infections with spores of different genotypes are independent events facilitated

by ecological dynamics, as theorized in a study of *F. tularensis* dynamics (Petersen et al. 2008).

However, PHRANA has not been put into a geographic context other than at a global to regional scale, and until recently was not used to investigate a single outbreak. Recent studies have complicated the interpretation of the hierarchical framework by documenting multiple MLVA and SNR genotypes from soil samples around single carcasses (Stratilo and Bader 2012) and within spatially and temporally limited outbreaks with detailed epidemiological information (Beyer et al. 2012). This former study compared MLVA and SNR genotypes from an outbreak affecting two geographically distinct populations located approximately 150 km apart. A southern population had a single MLVA genotype, while the northern population showed two MLVA genotypes, that of the southern population and a unique genotype. Two SNR types occurred in each the MLVA defined groups, one of which was isolated from both geographic populations. The temporal sequence of infections or mutational events in this outbreak was not known, but the work suggests that the geographic and temporal relationship of isolates not only must be evaluated carefully, but must be interpreted at finer scales than previously considered when using PHRANA approach. SNR diversity may develop within an outbreak more rapidly and consistently than previously recognized, and the epidemiological uses of this diversity are not yet understood. PHRANA is likely best considered as a conceptual framework for incorporation of genetic analysis, not a as an independent and unidirectional process. Spatial and epidemiological studies must conceptualize from the design stage how genetic data can

be used to answer the question and what genetic markers best address the question. Most importantly, marker diversity must be placed into a spatio-temporal context.

What Goes On Above: Geographic Techniques to Link Genetic Analysis and Spatial Patterns

Statistical Techniques

The first step in spatial studies of disease is visualization of the data (Collinge and Ray 2006). The simple act of mapping data using GIS is a powerful visual tool for planning further analyses. Maps illustrate and communicate data, hypotheses and conclusions. Patterns in the data can be visually described by simple point maps. Increasing the complexity of maps accordingly increases the power of a map to generate hypotheses and communicate ideas. Mapping has been used in multiple studies to illustrate phylogeographic patterns. The spatial distribution of genetic lineages has been illustrated for *F. tularensis* (Johansson et al. 2004, Farlow et al. 2005, Staples et al. 2006), Y*ersinia pestis* (Vogler et al. 2011b), as well as *B. anthracis* (Table 1-1;(Fasanella et al. 2005, Van Ert et al. 2007a, Simonson et al. 2009, Aikembayev et al. 2010, Fasanella et al. 2010). Such work provides visual confirmation that, at the spatial and genetic scales mapped, genetic branches of these pathogens are heterogeneously distributed and allows researchers to develop hypotheses for further investigation. The patterns and associations observed, however, are dependent on the genetic markers and spatial scales used.

A basic, yet essential, question is whether there is a relationship between genetic relatedness and spatial or temporal distance. The Mantel test was one of the first statistical methods employed to evaluate spatial patterns of genetic diversity. The Originally developed to detect spatio-temporal clustering of cancer cases, the technique

in its original context measured the relatedness of two matrices generated by calculating the pairwise temporal and spatial distances between pairs of individuals (Mantel 1967). In his original paper describing the test, the author correctly predicted that his statistic would have a variety of uses in epidemiologic studies. The Mantel test, in the context of this discussion, measures the relationship between genetic relatedness and geographic or ecological distance among pairs of individuals and though the use of the correlogram, a graphical representation of the test, can identify isolation by distance patterns as may occur via founders effect and genetic drift (Manel et al. 2003). An adaptation, the partial Mantel test, allows for control of a third variable. Rice, Martinez-Meyer, Peterson (2003) used Mantel tests to measure relationships between ecological distance and geographic distance of genetically dissimilar groups of *Aphelocoma* jays. A partial Mantel test was used to evaluate the relationship between genetic and ecological distance among lineages of *Aneidus flavipunctatus*, the black salamander, while controlling for geographic distance. Rissler (2007). Finding that that this relationship was significant, the authors divided the analysis by geographically defined groups of lineages to determine whether a single lineage or group of lineages drove the relationship. In another use of the test, the correlation between phylogenetic similarity and ecological envelopes of closely related species of Cuban anoles were compared, and it was determined that there was not a relationship between phylogenetics and niche similarity (Knouft et al. 2006). Mantel was also to compare genetic similarity and spatial distance among *Escherischia coli* isolates from domestic bovines (Davis et al. 2003).

Mantel test results can also provide clues to the processes driving pathogen transmission by testing spatial patterns of genetic diversity. For example, Carrel et al. (2010) used the Mantel test and a variation, the multiple regression using Mantel (MRM) to show that the genetic diversity of highly pathogenic (H1N1) avian influenza virus in Vietnam was related to geographic distance while controlling for time. Using a Mantel correlogram the authors demostrated that H1N1 diversity was lower than expected at small spatial lags and higher than expected at large spatial lags. The authors' interpretation was that isolation by distance had occurred in geographically distant areas as opposed to a gradual evolution of the virus as it spread over Vietnam. The Mantel test was used in a similar manner to evaluate the spatial structure of *Penicillium marnefei* isolates in Thailand (Fisher et al. 2005). In this study isolates from humans and bamboo rats were tested separately to show different spatial patterns within the two species, which indicated the pathogen undergoes different patterns of dispersal in the two species. A lack of a relationship genetic lineages of *F. tularensis* and spatial distance suggested that the lineages could have had separate reservoirs and with overlapping ecologies (Farlow et al. 2001). A strength of the Mantel test is that it can be applied at various genetic and geographic distances and therefore can be used to evaluate and explore relationships at multiple scales. A limitation of the Mantel test is an assumption that the relationships between the matrices are linear, although the test accommodates data transformations to account for lack of linearity. Finally, the test statistic, r, is sensitive to sample size in that larger samples tend to increase r.

The human eye is adept at finding clusters in data, even when no statistical deviation from randomness is present. Cluster analysis identifies significant patterns in

the spatial data, differentiating areas that deviate from random by either being more clustered or more dispersed. Varieties of statistics are used for cluster detection and can detect spatial, temporal and spatial-temporal clustering. In general, cluster detection methods can be divided in to global and local methods, and methods can be separated into those that detect clustering in aggregated or point data. Global cluster detection identifies the presence of data points or areas that deviate from random with a single test statistic, but does not identify where clustering occurs. Local techniques identify the locations of clustered points or areas as well as areas which are more dispersed than random. In the study of anthrax in Kruger National Park, South Africa, Smith et al. (2000) used Bernoulli cluster analysis to show different geographic distribution of two lineages of *B. anthracis*. Evaluation of disease clusters in an ecological context must be undertaken with caution, as clustering may be a function of a heterogeneous host population distribution, or unknown population factors, as opposed to a landscape feature creating environmental “hotspots” for pathogen persistence. When population data is available, methods that use the underlying population at risk for analysis overcome this limitation .

Ecological Niche Modeling

Ecological niche modeling (ENM) has evolved as a suite of methods with which to characterize the niche of a species by identifying non-random associations between known occurrences of a species and environmental variables. The niche can then be represented geographically by identifying areas having combinations of ecological conditions described by the model. Grinnell defined the ecological niche of a species as the environmental conditions in which the species can maintain without immigration (Grinnell 1917). Later, Hutchison expanded on this definition by showing that a species’

ecological niche consists of an n-dimensional hyperspace in which populations can maintain (Hutchinson 1957). This is considered the fundamental niche, or the theoretical space in which the species can persist, and was later divided into biotic and abiotic components. Within the hypervolume of the fundamental niche is the realized niche, which reflects the subset of conditions actually occupied by the species. The more ecologically limited realized niche is a consequence of biotic interactions such as competition or predation, geographic barriers to dispersal, and intrinsic limitations of the species to dispersal or reproduction. ENMs are typically concerned with abiotic, or climatic, variables. Discussion of modeling techniques and statistical evaluation of niche differences is beyond the scope of this paper (see Elith 2006, Soberon and Peterson 2005), but in general ENM uses locality data and environmental data layers in a correlative algorithm to identify geographic areas that are ecologically similar to those in which the species is known to occur. From these models the suite of environmental conditions associated with the species' occurrence can be derived.

ENM can be used to understand ecological differences between species, populations or genetic groups of a taxa {Blackburn, 2007 #187;Fisher, 2005 #107;Gonzalez, 2011 #353;Mullins, 2011 #366;Nakazawa, 2010 #21} (references for pathogens; lots of others could be used here). The ecological envelopes predicted by ENMs have been evaluated for niche differences using the Mantel test (Rice et al. 2003), PCA (Broennimann et al. 2007), multivariate analysis of variance **(**MANOVA) (Nakazawa et al. 2010) and combinations of these. Principle component analysis (PCA) is a data reduction technique that summarizes variation in a data set into new variables, or principal components. PCA has been used widely in the literature

investigating spatial patterns of evolution and has been incorporated at different stages of analysis, including to define genetic groups for modeling geographic distribution (Fisher et al. 2005) and to reduce environmental dataset niche modeling. Ecological envelopes generated by ecological niche modeling can be reduced using PCA to compare the space occupied by native and invasive populations of species (Broennimann et al. 2007) (Ron 2005)and to compare the niche of closely related lineages (Rissler and Apodaca 2007). While PCA is valuable as a way to overcome highly collinear datasets and reduce variance, the results can be difficult to interpret, explain and apply, particularly in a public health setting.

It has been argued that to be useful models should be transferable, or, in other words, the ecological niche model can identify suitable areas for the species in an geographic area or time other than the one in which the model was built (Jiménez-Valverde et al. 2009, Peterson et al. 2011). Transferred models have been used to evaluate whether species or populations have different ecological niches by testing how well a model built using one population can predict the other on the same or different spatial extents (Broennimann et al. 2007). However, the process of transferring models appears to have a number of methodological shortcomings (Peterson and Nakazawa 2008, Jiménez-Valverde et al. 2009, Rödder et al. 2009, Rödder and Lötters 2010). The failure of transferred models to predict known occurrences has been interpreted as niche differentiation (Broennimann et al. 2007), but other possibilities are that the modeled populations' niche may be more labile than the predicted populations', or genetically distinct sub-populations occupy unique ecological niches (Peterson and Holt 2003). Alternatively, the source of ecological data and the suites of variables used for

modeling appear to affect the spatial predictions made by transferred models more so than predictions on the training landscape (Peterson and Nakazawa 2008, Rödder and Lötters 2010). Suggestions for controlling for the effect of variable selection include evaluation of multicollinearity (Graham 2003), use of a Mantel test to evaluate the similarity of environmental data between native and novel landscapes (Jiménez-Valverde et al. 2009), and careful attention to the relevance of a variable to the species' biology (Rödder and Lötters 2010). These suggestions have not been tested, however, and more needs to be done to understand the methodology and limitations of transferred models. Transferring models should not be confused with extrapolation, which refers to projecting models into ecological values not included in the model. The latter also remains a challenge for ENM about which the user must be aware.

Efforts to understand the ecology of infectious disease may benefit greatly from ENM experiments. Models can be used to generate and test hypotheses about disease transmission and ecology (Peterson 2006) as well as to predict the spatial distribution of disease risk or disease vectors (Peterson et al. 2002b). Characterization of the ecological envelope of a pathogen species can provide clues about the dynamics of disease hosts, vectors and reservoirs. For example, overlapping ENMs were used to match epidemiologically important reservoirs and vectors for Chagas' disease (Peterson et al. 2002b), and by comparing the ecological niche of plague occurrences to that of the hosts it was found that disease foci had a unique "plague niche" common across all host species (Maher et al. 2010). Ecological niche modeling has also used to predict areas at risk for pathogen presence, including plague (Parmenter et al. 1999, Adjemian

et al. 2006, Maher et al. 2010, Giles et al. 2011, MacMillan et al. 2012), tularemia (Nakazawa et al. 2010) and anthrax (Blackburn et al. 2007).

Ecological niche modeling also has important potential contributions to the study of spatial patterns of evolution (Rissler and Apodaca 2007) (Pearman et al. 2008). Genetic analysis provides a method with which to divide species into biologically relevant groups for building more refined ENMs (Gonzalez et al. 2011). In turn, finding differences in niche between sub-populations may reveal cryptic genetic patterns. Genetic lineages of pathogens have been shown to have differences in ecological niche (Fisher et al. 2005, Nakazawa et al. 2010), spatial distribution (Fisher et al. 2005) as well as epidemiology (Kugeler et al. 2009). Similar to phylogeographic patterns, however, the results of niche modeling experiments comparing genetic groups may appear more or less similar depending on the phylogenetic level of analysis used (Pearman et al. 2008). Authors should explicitly discuss the implications of phylogenetic level, spatial scale and temporal scale on ecological niche modeling experiments incorporating genetic analysis. The optimal scales at which species are ideally modeled will likely vary according to both the ecology of the species and the hypothesis being tested.

The soil borne nature of *B. anthracis* makes it ideal for using ENMs to predict areas of potential pathogen persistence. In areas with recorded epidemiological data, ENM can be used to focus public health prevention and control measures to areas of both high likelihood of persistence and high epidemiological activity. Furthermore, niche models can test for genetic-ecological associations and to characterize the ecology of the pathogen on a specific landscape. The nature of the dispersal of *B.*

anthracis would have subjected the pathogen to Founder's effect in new environments; environments at the extreme of the isolate's niche could result niche shift because of labile niche characteristics or genetic evolution and adaptation (Holt et al. 2005), causing the ecology of anthrax to vary. Niche models can aid in prioritizing surveillance efforts in regions where past public health infrastructure does not allow for adequate disease recording or where disease has not been reported. Outbreak investigation also benefits from genetically informed ENMs. For example, in an area not predicted to support the presence of *B. anthracis*, either at all or of the outbreak strain, the source of the outbreak is likely to be a result of a food or human mediated introduction, as opposed to a natural focus (Fasanella et al. 2012).

The ecological niceh of anthrax has been characterized in two regions. Blackburn et al. (2007) modeled the niche of *B. anthracis* in the United States using the genetic algorithm for rule set production (GARP) and predicted areas having a high likelihood of pathogen presence. Furthermore, a model built in the United States was able to successfully predict anthrax occurrences in Mexico when projected onto that country Blackburn (2010). Later, Joyner et al. (2010) modeled the niche of the pathogen in Kazakhstan under current and future climate scenarios. Using current climate conditions models containing measures of the normalized difference vegetation index (NDVI) performed better that those without the measures (Joyner 2010). Chapters 2, 3 and 4 of this dissertation expand on these previous efforts. Specifically, the ecological niche of a single, successful sub-lineage of *B. anthracis* is modeled in Chapter 2,and Chapters 3 and 4 test whether models this sub-lineage can predict occurrences in novel landscapes.

Statistical Phylogeography

Phylogeography attempts to understand the processes that underlie the distribution of genetic variation within and among closely related species or populations (Avise 1998, Knowles 2009). Phylogeography also seeks to understand the evolutionary history of a species by viewing the genealogical tree in a spatial context, although the analyses are often more focused on the genetic analysis while the importance of geography is less emphasized (Kidd and Ritchie 2006). Phylogeographic techniques have been combined with human population movement patterns to study HIV evolution and dispersal (Gray et al. 2009) and to understand the dispersal of *Mycobacterium tuberculosis* by the fur trade (Pepperell et al. 2011). Phylogeography is valuable for understanding spatial patterns of evolution, but it does have some caveats. For example, most statistical phylogeographic techniques assume a homogenously distributed population, which is rarely true in natural populations (Kidd and Ritchie 2006). In addition, of importance to the epidemiologist seeking to understand disease transmission is that phylogeographic techniques use population level data, as opposed to individual level data, and therefore do not explicitly model fine scale spatial trends.

Landscape Genetics

Landscape genetics provides tools with which to gain a fine scale picture of where a pathogen is moving on the landscape, and perhaps why (Archie et al. 2009). While phylogeography combines biogeography and phylogenetics to investigate diversity patterns among populations at broad temporal and spatial scales, landscape genetics explicitly addresses Individual level and fine spatial and temporal scales (Manel et al. 2003, Storfer et al. 2006). Statistical phylogeography and landscape

genetics also differ in that the former is concerned with historical processes that shaped the current genetic variation, while the latter focuses on the effects of the contemporary landscape on genetic patters. Manel et al. (2003) described landscape genetics as an approach that studies the interactions between landscape features and microevolutionary processes. Quantifying these relationships between landscape variables and genetic variation combines landscape ecology, spatial statistics and population genetics (Storfer et al. 2006). In the past decade the theory, methodologies and applications of landscape genetics have increased exponentially, and have been applied to identifying barriers to pathogen movement (Biek et al. 2007, Blanchong et al. 2008, Cullingham et al. 2011), evolutionary mechanisms by which a pathogen spreads (Archie et al. 2009)and movement corridors (Epps et al. 2007).

Spatial studies of anthrax

The use of spatial analyses other than ecological niche modeling for studying anthrax ecology and transmission has been limited. Three published deterministic models have been used to understand transmission, although space was not explicitly parameterized (Furniss and Hahn 1981, Hahn and Furniss 1983, Gainer and Saunders 1989). Cluster analysis was used to demonstrate spatial clustering of anthrax outbreaks by host species (Aikembayev et al. 2010). Epp et al. (2010a) analyzed a 2006 outbreak in Saskatchewan with farm level point data and municipal level aggregated data. Using case farms and control farms, the authors showed spatiotemporal clustering at both levels of analysis and generated a model of background relative risk in the area during the outbreak. In addition, the analysis included a velocity vector map using a trend surface analysis to visualize the spatial pattern of spread during the outbreak. Three separate directions of movement were

identified which corresponded to the three significant spatio-temporal clusters. Although no conclusions were made from this portion of the analysis, the authors speculated that the clusters may have been influenced by different ecological conditions or insect vector populations. Seeking to understand the sporadic and often species-specific nature of outbreaks in Etosha National Park, Namibia, Hampson et al. (2011) used general linear mixed models to identify drivers of outbreaks in the ecosystem. They found that certain pairs of species tended to co-occur in outbreaks and that both the spatial heterogeneity of outbreaks and seropositivity among carnivores was positively associated with soil alkalinity. In buffalo, carcasses tended to be located closer to river bottoms than if deaths had occurred in random locations. Whether this is associated with exposure in these locations versus water seeking behavior during the septicemic phase is unclear. To date, few studies have incorporated genetic analysis of *B. anthracis* into spatial studies. Smith et al. (2000) employed cluster analysis was used to show the localized distributions of MLVA-8 defined genetic lineages in Kruger National Park, South Africa. No studies have used higher resolution genotyping systems for spatial analyse, nor have outbreak data been incorporated into such studies.

Expected Significance

Combining information about pathogen genetics, ecology and life history allows a greater understanding of disease dynamics. This greater understanding informs public health measures and may allow predictive models with which to predict the disease. To begin, genetic analyses allows more refined ecological niche models of *B. anthracis* persistence, which in turn allows better predictions of where the pathogen can persist based on the niche of the local population. In addition, genetic analyses may reveal patterns of pathogen movement which explain transmission dynamics within and

between outbreaks. This dissertation addresses unanswered questions about the use of molecular techniques in ecological studies of *B. anthracis*. Specifically, ecological niche modeling will be used to evaluate whether incorporation of genetic information improves spatial predictions of presence. Such models can be combined with historical epidemiological data and incorporated into surveillance schemes to better detect cases, guide vaccine delivery and stop outbreaks. Further considering the use of genetic analysis to understand anthrax transmission, this dissertation explores the hypothesis that SNR analysis will inform our understanding of outbreak epidemiology. Finally, this work will apply the three typing systems in a nested hierarchical fashion to progressively describe population genetic diversity in the context of characterizing pathogen dynamics. This work will provide guidance for future applications of these markers to epidemiological studies of anthrax and contribute to predictive models of anthrax occurrence and spread.

Table 1-1. Studies utilizing MLVA 8, 15, 25, and 31 typing systems and SNR analysis

Paper	Area	SNP	MLVA system	SNR	Mapped
Smith et al. 1999	Kruger National Park, SA	-	MLVA-8	-	(in Smith et al 2000)
Keim et al. 2000	Global	-	MLVA-8	-	Genotypes linked to geography on tree
Fouet et al. 2002	France	-	MLVA-8	-	Y
Gierczynski et al. 2004	Poland	-	MLVA-8	-	N
Ryu et al. 2005	Korea	Y	MLVA-8	Y	Y
Fasanella et al. 2005	Italy	-	MLVA-8	-	By region
Merabishvili et al. 2006	Georgia	-	MLVA-8	-	Location listed in a table
Maho et al. 2006	Chad	-	MLVA-8	-	Y
Lista et al. 2006	Italy	-	MLVA-25	-	Genotypes linked to geography on tree
Van Ert et al. 2007	Global	Y	MLVA-15	-	By region
Kenefic et al. 2008	South Dakota, US	Y	MLVA-15	Y	Y
Simonson et al. 2009	China	Y	MLVA-15	-	By region
Fasanella et al. 2010	Italy (Basilicata)	Y	MLVA-25	Y	Y
Aikenbayev et al. 2010	Kazakhstan	Y	MLVA-8	-	Y
Garofolo et al 2010	Italy (single outbreak)	-	MLVA-25	S-R	Y
Antwerpen et al. 2011	Bulgaria	Y	MLVA-15	-	Genotypes linked to geography on tree
Garofolo et al. 2011	Italy (Northern)	V	MLVA-15	-	Y
Durmaz et al. 2012	Turkey	-	MLVA-25	-	Y
Fasanella et al. 2012	Bangledesh	Y	MLVA-15	Y	Y
Ortalatli et al. 2012	Turkey	-	MLVA-25	-	Y
Beyer et al. 2012	Namibia	Y	MLVA-31	Y	Y
Hwa Jung et al. 2012	Korea	Y	MLVA-8	-	N

CHAPTER 2
ECOLOGICAL NICHE MODELLING OF THE *BACILLUS ANTHRACIS* A1.A SUB-LINEAGE IN KAZAKHSTAN

Introduction

Anthrax is a disease of wildlife, livestock and humans that remains a public health problem throughout the world. *Bacillus anthracis*, the causative agent of anthrax, is a soil-borne, spore-forming bacterium which persists in soil for long periods of time under appropriate conditions(Hugh-Jones and Blackburn 2009). Certain soil parameters, including pH, organic content and calcium, may be associated with spore survival (Van Ness 1956, 1971, Dragon and Rennie 1995, Smith et al. 2000, Hugh-Jones and Blackburn 2009). Anthrax outbreaks among livestock and wildlife result from exposure to these spores and are possibly influenced by climatic and physiological events (Parkinson et al. 2003, Epp et al. 2010b). In endemic areas human cases of anthrax primarily result from contact with infected livestock during slaughter or butchering (Bales et al. 2002, Woods et al. 2004) and control of livestock disease through vaccination and active surveillance of livestock and wildlife is essential for preventing human disease (Blackburn et al. 2007). However, widespread active surveillance is costly and vaccination of every animal is not feasible. It is far more practical to focus these efforts on areas of high risk. To identify these it is necessary to improve our understanding of the ecology of *B. anthracis* through which animal infection occurs.

The ecology of a pathogen such as *B. anthracis* can be explored using similar tools as those used for species distribution modelling and conservation planning. For example, ecological niche modelling (ENM) has been used to predict the potential ecological and geographic distribution of pathogens based on outbreak

locations(Peterson et al. 2004, Ron 2005, Blackburn et al. 2007, Joyner et al. 2010, Nakazawa et al. 2010), presence of disease vectors(Costa et al. 2002, Peterson and Shaw 2003, Adjemian et al. 2006) and disease reservoirs(Peterson et al. 2002c). The ecological niche of a pathogen, as for other types of species, is conceptualized as the N-dimensional hypervolume of ecological parameters within which the species can be maintained without immigration (Grinnell 1917, Hutchinson 1957). Various approaches to ENM identify non-random associations between a species' locality data and environmental parameters. Ecological niche modelling experiments of *B. anthracis* are particularly useful considering the potential associations between spore survival and ecological conditions(Van Ness 1971, Hugh-Jones and Blackburn 2009). Results can be used as a proxy for disease risk and integrated into focused surveillance strategies for wildlife and livestock in endemic areas and into vaccination strategies that target at risk herds before and during outbreak events(Blackburn et al. 2007).

Recently, studies of disease ecology have combined molecular genotyping techniques and ecological niche modelling to provide evidence that genetic lineages of a pathogen can have different environmental associations and potential geographic distributions (Fisher et al. 2005, Nakazawa et al. 2010). In general *B. anthracis* has relatively limited global diversity. However, multiple locus variable number tandem repeat analysis (MLVA) systems for *B. anthracis* can differentiate strains into distinct lineages and sub-lineages (Keim et al. 2000, Lista et al. 2006, Van Ert et al. 2007a). Analyses of a global collection of *B. anthracis* isolates suggests that the A lineage is globally distributed, while other lineages (B and C) are geographically restricted. These findings may be explained by adaptive differences, some of which carry fitness costs

that limit abundance and distribution of certain lineages or sub-lineages (Smith et al. 2000, Van Ert et al. 2007a). The ecological niche of *B. anthracis* has been modelled in the United States and Kazakhstan using locations of reported outbreaks (Blackburn et al. 2007, Blackburn 2010, Joyner 2010, Joyner et al. 2010). A stated limitation of these experiments was that the outbreak data potentially included multiple strains of *B. anthracis* (Blackburn et al. 2007, Joyner et al. 2010). If lineages of *B. anthracis* do exhibit niche specialization and unique geographic distributions, then it is plausible that current outbreak based ecological niche models are biased toward a dominant strain in a particular landscape. It would then follow that single lineage models may better predict presence of the pathogen at local scales and increase the value of public health measures (Blackburn et al. 2007).

Kazakhstan is situated in Central Asia, a region with some of the highest reported human anthrax incidence and mortality rates in the world (Cherkasskiy 1999, Hugh-Jones 1999). The majority of human anthrax cases in Kazakhstan are related to exposure to infected livestock or handling of products derived from infected livestock (Woods et al. 2004). In rural areas of Kazakhstan veterinary care and surveillance programs are limited by the country's large land mass and widely distributed rural populations. Vaccination of livestock occurs mainly in response to detected outbreaks. In countries such as Kazakhstan, prioritizing areas for vaccination and surveillance are necessary for disease control. Our group recently created a multi-variate ecological niche model to characterize the broad environmental conditions that support *B. anthracis* across Kazakhstan (Joyner 2010, Joyner et al. 2010). In a parallel effort, Aikembayev et al. used an eight marker MLVA typing system (MLVA-8) to describe the

diversity of *B. anthracis* within Kazakhstan from 88 archival strains(Keim et al. 2000, Aikembayev et al. 2010).

In this study, we first expanded on the previous outbreak based modelling experiment by adding four soil variables (pH, calcium levels, organic content and baseline water saturation) to the original set of environmental variables. Despite literature suggesting a strong relationship between soil characteristics such as high calcium levels and alkaline pH and spore persistence, the influence of available soil variables on *B. anthracis* ENM predictions has not been comparatively examined. We next used these twelve environmental variables and the collection of MLVA-8 genotyped samples to create an A1.a sub-lineage specific ecological niche model for Kazakhstan.

Methods

Anthrax Occurrence Data

Presence points for outbreak-based models were taken from the database created by Joyner et.al. (Joyner 2010). Briefly, historical records were used to construct a database of 3,947 anthrax outbreaks reported in Kazakhstan between 1937 and 2006. The data were sequentially filtered to create a dataset containing the latitude and longitude of outbreaks in cattle, sheep and goats between 1960 and 2000. The final dataset contained 258 spatially unique points, meaning that only one outbreak point occurred in each 8km2 pixel. An 8km2 resolution was chosen because outbreaks were mapped to the nearest village and some outbreaks occurred greater than 1km from the village coordinates. This data set is hereafter referred to as the full outbreak dataset.

A second dataset was constructed using outbreak isolates genotyped using canSNP and MLVA-8 by Aikimbayev et.al.(Aikembayev et al. 2010). Isolates were

grouped into the A1.a (n=78), A3.b (n=6) and A4 (n=4) lineage using unweighted pair group method with arithmetic mean (UPGMA) cluster analysis and the 89 B. anthracis genotypes identified by Keim et.al.(Keim et al. 2000). This was filtered to contain only spatially unique points at a resolution of 8km2 resulting in 42 spatially unique points, of which 39 were A1.a, two were A4 and 1 was A3b. Locality data are mapped in Figure 2-1. Only the A1.a sub-lineage had an adequate number of spatially unique points for modelling. The A1.a locations were geographically biased towards the south-eastern portion of the country. To reduce the problem of largely unsampled areas being considered as absence points, we created a polygon encompassing southeast Kazakhstan using latitude 48N and 60E as the northern and western boundaries, respectively, and the country boundaries in the south and east. The boundaries of this southern polygon were derived from examination of the locations of the A1.a isolates and from the different northern and southern ecological associations noted by Joyner (Joyner 2010). The southern polygon was used to clip the A1.a locality points in ArcMap and this set of southern A1.a locations was used for the sub-lineage experiment. The same procedure was used to create a southern outbreak dataset.

Environmental Data

We used six environmental coverages downloaded from the WorldClim website (www.worlclimorg)(Hijmans et al. 2005). The WorldClim variables are calculated from interpolation of monthly temperature and precipitation measurements recorded at stations located worldwide between 1961 and 2000. Monthly values are transformed by WorldClim into 19 bioclimatic variable grids that describe annual trends, seasonality and potentially limiting ecological parameters such as temperature of the coldest and warmest months. Two satellite-derived environmental variables describing temperature

and vegetation measures were obtained from the Trypanosomiasis and Land Use in Africa (TALA) research group (Oxford, United Kingdom) (Hay et al. 2006).

We added four soil variables to the set of eight environmental layers used in the previous study to test the influence of soil parameters on the outbreak model. Soil variables were derived from the Harmonized World Soil Database and were available at 1km2 resolution (FAO/IIASA/ISRIC/ISSCAS/JRC 2009). All coverages were re-sampled to 8km2 and clipped to the boundaries of Kazakhstan in ArcView 3.3. (Environmental Systems Research institute, Redlands, CA). An identical set of coverages were clipped to the southern polygon. The final set of coverages is given in Table 2-1.

Ecological Niche Modeling

This study used the Genetic Algorithm for Rule-Set Prediction (GARP) to perform the ecological niche modelling(Anderson et al. 2003). Models were developed in Desktop GARP v.1.1.3, which gives the user the option to write out the rule sets for each model. Briefly, GARP is a presence only modelling technique that detects non-random associations between species localities and specific environmental variables. Through an iterative process, relationships are expressed as a series of logic statements, or rules, of which there are four types: (1) logit – based on logistic regression; (2) atomic – single value for a given variable that predicts presence; (3) range - a range of values of a given variable that predicts presence; and (4) negated range – a range of values outside of which presence is predicted. Each individual GARP model is a set of 50 rules that are randomly generated, tested and modified. The user sets a maximum number of models to be created in a single experiment. A best subsets procedure within GARP then selects a set of optimal models based on

user defined omission and commission criteria(Anderson et al. 2003). The algorithm is a two-step process, where first relationships are defined in variable space through a random walk and then applied to the geographic landscape where those conditions are met(Blackburn 2010). GARP therefore has the benefit of being able to project rule sets onto the environmental layers of a different landscape and has been shown to be robust in this application (Peterson et al. 2007, Blackburn 2010).

Model Building and Evaluation

To test the effect of soils on the outbreak experiment we used the full outbreak dataset as locality points and the twelve environmental variables described in Table 2-1. The 258 spatially unique points were randomly divided into an 85% (n=218) training set used for model building and a 15% (n=39) testing set for model evaluation. The 32 southern A1.a points were divided into an 80% (n=26) training set and a 20%(n=6) testing set in order to maximize points available for testing (Peterson et al. 2002a, Stockwell and Peterson 2002, McNyset 2005). The A1.a training set was input into GARP with the set of environmental coverages clipped to the southern polygon for model development. Rules from this southern A1.a experiment were projected onto the entire landscape of Kazakhstan. In order to test the robustness of the A1.a model projections given the relatively small sample size and issues of transferability, two experiments using the southern outbreak data were performed. The first utilized all 142 southern outbreak points and is referred to as the large southern outbreak experiment. For the second, 32 points were randomly selected from the southern outbreak dataset (small southern outbreak experiment). For these experiments an 85%/15% and 80%/20%, respectively, external data split was performed and the experiments conducted as for the sub-lineage. The four experiments are summarized in Table 2_2.

For all niche modelling experiments, we specified 200 models with a maximum of 1,000 iterations and a convergence limit of 0.01. The training data were input into GARP with a 50% training/50% testing internal data partition. The best subset procedure selected the best 20 models under a 10% hard omission threshold and a 50% commission threshold. The resulting ten best subset models were imported in ArcGIS and summated using the raster calculator function of the Spatial Analyst extension. This created a single cumulative raster file of model agreement for B. anthracis presence ranging from 0 (all models predict absence) to 10 (all models predict presence). The more models that predict presence for a given pixel, the higher the likelihood that the pixel can support *B. anthracis*.

Rule types from the ten best models of the outbreak-soil experiment and the A1.a sub-lineage experiment were extracted with a python script (K.M. McNyset, US NOAA) and summarized to illustrate the relative numbers of each rule type. Dominant rules, or the subset of rules that together predict over 90% of the landscape, for each model were identified.. We extracted the minimum and maximum values of range rules using the python script. When logit rules were identified, we extracted the range of values across the pixels predicted by that rule using the"Extract Values to Points" routine of the Spatial Analyst extension in ArcMap. Median minimum and maximum values for each variable were calculated in SAS (SAS 9.2, Cary, N.C.) and plotted as a bar graph. Differences in median and maximum values between experiments were assessed using Wilcoxon-Mann_Whitneytest in SAS.

Predictive performance of the best subset models was evaluated with an area under the curve (AUC) in a receiver operating characteristic (ROC) analysis using the

independent test data withheld from the original datasets(Wiley et al. 2003). For the projected models, testing points included presence points outside the southern polygon in addition to points withheld from within the southern polygon. Values of AUC, which range from 0.5 (no different from random) to 1 (a perfect model), are derived from measures of sensitivity (absence of omission error) and specificity (absence of commission error). The calculated value is compared to that of a random model using a z-test. In addition, measures of omission and commission were calculated using the summed ten best subset models. Total and average omission values evaluate how well GARP predicts the presence of known locality points not included in the model building data. Total and average commission is the percent of pixels predicted as presence by the summated model and the average of this value for all ten best subset models, respectively. Large variation between the two measures of commission suggests substantial variation between the proportions of the landscape predicted present by each of the ten best subset models (McNyset 2005).

Environmental values of 5,000 randomly chosen points from areas predicted by all 10 of the best subset models were extracted using the “Extract Values to Points” routine of the Spatial Analyst extension in ArcMap. Values of each environmental variable at each presence point and at 5,000 random points representing the total available environmental space (background) were similarly extracted. Environmental variables appearing to be limiting factors for prediction of B. anthracis were chosen based on the rule set evaluation (Figure 2-5) and visualized in 2-dimensional ecological space against the background of available environmental conditions using R 2.1.1 (http://www.R-project.org.).

Results

Accuracy Metrics

Ecological niche modelling was performed using Genetic Algorithm for Rule-Set Prediction (GARP). Four experiments were run (outbreak-soil, A1.a sub-lineage, small southern outbreak and large southern outbreak) and are summarized in Table 2-2. All modelling processes reached convergence of accuracy (0.01) prior to reaching the maximum iteration setting (=1,000). The outbreak-soil model had an Area Under the Curve (AUC) of 0.7188 and was significantly different from a random model. Total omission of the outbreak-soil model was 2.6% and average omission was 9.9%, indicating that 97.4% of the testing points were predicted by at least one best subset model and 89.1% were predicted by all models. The AUC of the A1.a clade model was 0.6964 and was also significantly different from a random model and had a total and average omission of 0 and 13.1, respectively. Both the large and small outbreak models had AUCs significantly different than random. Accuracy metrics for all models are shown in Table 2-3.

Predicted Distributions of *B. anthracis*

Locations used for input into GARP are shown in Figure 2-1. Based on areas of agreement of a minimum of six of the best subset models, the outbreak-soil experiment predicted B. anthracis across much of northern Kazakhstan and in a narrow band of the southeast. The interior of the country, which is primarily arid, was not predicted to be suitable for the pathogen. The results are similar to those of the experiment without the soil variables with respect to the geographic extent of areas of six or more best subset model agreement (Figure 2-2). The outbreak-soil experiment expanded two areas in the north which had lower model agreement. The A1.a sub-lineage experiment predicts

a more extensive geographic distribution than that of the outbreak experiment, including areas in the northern interior and western portions of the country (Figure 2-3). The northern pockets of less suitable geographic areas seen in the outbreak-soil experiment were predicted to be unsuitable based on agreement of six or more best subset models. The overall extents of the geographic predictions of the two experiments were more similar in the south than in the north. The large and small southern outbreak experiments both predicted similar geographic extents as the outbreak-soil experiment (Figure 2-4). All three projected experiments (A1.a sub-lineage, large southern outbreak and small southern outbreak) were run ten additional times using random external data splits. The subsets of A1.a sub-lineage and small southern outbreak experiments showed greater degrees of spatial heterogeneity than did the large southern outbreak experiment set. (Appendix A).

Each GARP model is composed of 50 if-then type rules (logic, range, negated range and atomic) which predict the presence or absence of the species at each pixel. Rule types for the ten best subset models from the outbreak-soil and A1.a sub-lineage experiments were extracted and are summarized in Table 2-4. Just over half of the outbreak-soil experiment rules were logit and no atomic rules were included, whereas Range rules made up over 60% of the A1.a sub-lineage experiment rule types and this experiment included four atomic rules in the best subsets. Between 6 and 13 rules defined greater than 90% of areas predicted to be suitable for B. anthracis for each of the best subsets. Of the 95 rules which predicted the majority of the landscape in the outbreak-soil experiment, the majority (83%) were presence rules and of these 62% were range rules. The A1.a sub-lineage experiment had 99 total rules predict the

majority of the landscape; all but one of these was a presence rule and 73% of the presence rules were range rules. The environmental tolerances described by the dominant rules suggest that mean NDVI, altitude, mean temperature, minimum soil calcium and minimum soil organic content are limiting variables for B. anthracis in Kazakhstan (Figure 2-5). Median minimum values of mean NDVI, NDVI amplitude, annual precipitation, dry month precipitation, wet month precipitation, mean temperature, altitude and soil organic content are significantly different between the A1.a sub-lineage and the outbreak-soil experiment using the Wilcox-Mann-Whitney test at a 95% significance level. Median maximum values of NDVI amplitude, mean temperature, dry month precipitation, altitude, soil base saturation and soil organic content differ between the two experiments.

Values of the limiting variables were extracted from areas of ten best subset model agreement and plotted in two dimensional variable space. The A1.a sub-lineage experiment showed a broader ecological envelope than the outbreak-soil experiment based on areas of ten best subset model agreement, despite the smaller geographic area predicted by agreement of all ten models (Figure 2-6). The outbreak and A1.a sub-lineage experiments resulted in similar ranges of soil organic content values, while the range of minimum pH values occupied by the A1.a sub-lineage model is lower than that of the outbreak-soil model. The two A4 locations, which are distant from each other geographically, are found within a narrow range of mean NDVI and mean temperature, but occupy nearly opposite ends of the range of precipitation values. Finally, the A3.b location was associated with ecological conditions towards the outer boundaries of the ecological envelope predicted by the outbreak-soil experiment.

Discussion

This study assesses the addition of soil variables to a previously developed ecological niche model for *Bacillus anthracis* and is the first known to model the ecological and geographic distribution of a single sub-lineage of *B. anthracis*. Inclusion of available soil variables into our anthrax outbreak model resulted in subtle changes in the likelihood of the pathogen in areas of northern Kazakhstan, but did not substantially change the extent of geographic predictions or results of rule set analyses [26]. The areas predicted as less suitable by the outbreak-soil model correspond to regions of locally different values for all four soil variables (Appendix A). However, it is not known whether these areas represent a unique ecological region or if measurement in these areas was affected by error or bias. The low minimum soil calcium association found in the rule set analyses contrasts with previous literature suggesting that *B. anthracis* spore persistence is associated with high soil calcium levels (Smith et al. 2000, Himsworth 2008).The results of our rule set analysis, however, are not directly comparable to previous work in that different units of measurement and sampling techniques were used. In addition, the soils data available for this study had a relatively coarse resolution of 1km and were further aggregated to 8km to correspond to the outbreak data. As a result, fine resolution relationships between soil and anthrax occurrences would likely be missed in this experiment. Improved resolution of soil and outbreak data, such as exact carcass locations, are likely necessary to characterize the role of soil parameters in promoting anthrax spore persistence (Van Ness 1967, Smith et al. 2000) and for better understanding the spatio-temporal dynamics and ecology of local outbreaks (Blackburn 2006).

The A1.a sub-lineage experiment predicted a more extensive geographic area of anthrax presence than did the outbreak experiment. This is most pronounced in the northern and central portions of the country. When dominant rules from both experiments were extracted and analysed the median minimum values of most variables were significantly different. Furthermore, in two dimensional variable space, the A1.a sub-lineage was associated with a larger ecological envelope than the outbreak-soil data; in the case of minimum soil pH the areas predicted to be most suitable for *B. anthracis* based on the A1.a sub-lineage experiment had a lower range of values than did areas predicted by the outbreak-soil model. These findings illustrate that analyses of dominant rule sets alone should be interpreted with some caution. The variable ranges derived from the dominant rule sets summarize approximately only one-fifth (or fewer) of the total number of rules generated by GARP in the ten best subsets. The values extracted from predicted areas on the landscape, however, are derived from all 500 rules contained in the 10 best subsests and represent the spectrum of complex interactions between variables and the landscape. Because geographic areas of model agreement can be thought of as representative of all sub-sampled regions or populations (McNyset 2005), the finding of a larger ecological envelope for the A1.a sub-lineage experiment lends support to the hypothesis that the A1.a sub-lineage of *B. anthracis* may have broad environmental tolerances that influence its broad geographic distribution(Smith et al. 2000, Van Ert et al. 2007a).

The A lineage is more widely distributed globally than other subtypes, perhaps reflecting a greater level of fitness as compared to other lineages(Van Ert et al. 2007a). This finding has been shown on a local scale as well. Isolates of the A lineage in

Kruger National Park, South Africa, as defined by MLVA-8 typing, were more diffusely distributed and showed a distinctly different spatial cluster pattern than those of the B lineage. Furthermore, the B lineage isolates occupied a narrow range of available ecological conditions within those occupied by the A lineage isolates(Smith et al. 2000). In the experiments reported here, the A1.a sub-lineage locations were associated with lower pH values than the outbreak locations and this could provisionally support the findings from Kruger National Park (Smith et al. 2000). It follows that that a *B. anthracis* lineage and/or sub-lineage other than A1.a may predominate in the northern regions of Kazakhstan and is driving the narrower geographic and ecological prediction of the outbreak-soil experiment. Genotyping of additional isolates from northern Kazakhstan is necessary to evaluate this hypothesis. Locations of outbreaks of the A4 lineage in geographical and ecological space suggests that this genotype may also have a relatively broad distribution, which is consistent with the A4 sub-lineage being found across the Middle East and China (Simonson et al. 2009, Aikembayev et al. 2010). Although we cannot make inferences regarding the environmental affinities of the single A3.b isolate, we note it is located on the far eastern border of Kazakhstan in an area predicted to be unsuitable for anthrax by the outbreak-soil experiment. The A3.b sub-lineage has been isolated from geographically limited areas, most notably northern China and Texas, and the presence of this genotype is likely a result of historical trade routes (Van Ert et al. 2007a, Kenefic et al. 2008c, Simonson et al. 2009).

The genetic diversity of *B. anthracis* isolates in southern Kazakhstan is not surprising given the location of this area along the historic Silk Road (Aikembayev et al. 2010), but this diversity also implies that this region is supportive of spore persistence.

Associations between genotypes of *B. anthracis*, environment, virulence or host species have not yet been fully explored and it is unknown whether genotype influences epidemiological characteristics of outbreaks. Understanding these relationships will improve our understanding of anthrax disease ecology, help focus surveillance and efficiently direct proactive vaccination. Furthermore, this knowledge can help distinguish between naturally occurring outbreaks, contamination and potential bioterrorism, and greatly enhance epidemiological trace back (tracing outbreaks to source) efforts during outbreaks.

Few studies using GARP have quantified and examined rule types. GARP begins creating rule sets by choosing the first rule type at random and successful rules are carried forward into subsequent rule sets. Thus, the first rule type chosen at random will often predominate in the final rule sets. Blackburn and Joyner et. al. both presented summaries and distribution of the dominant rule types predicting *B. anthracis* in the continental United States and Kazakhstan, respectively (Blackburn 2006, Joyner 2010). Blackburn noted a predominance of range rules among a total of 63 rules predicting greater than 90% of the landscape (Blackburn 2006). Joyner et al divided Kazakhstan into northern and southern halves and modelled the two sections separately (Joyner et al. 2010). A greater percentage of range rules described the northern half of the country, whereas logit rules predominantly described the southern half of the country. Here, logit rules dominated in the out-break soil experiment and range rules in the A1.a sub-lineage experiment. Further work is required to tease out whether dominant rule types result from the stochastic nature of GARP or are related to complex interactions between the organism and environmental variables. Additional

future work should also explore how rule sets and ecological values found over predicted areas of the landscape can be used to enhance our understanding of the ecology of an organism.

Several authors have noted that ENM-based predictions of species with widespread distributions show reduced model accuracy which can be improved by dividing species or the range into sub-units (Peterson 2001, Stockwell and Peterson 2002, Rice et al. 2003, McNyset 2005, Hernandez et al. 2006). One explanation for the apparent poor accuracy of models of widespread species is the use of AUC. The AUC is sensitive to the area predicted to be suitable for a species relative to the total land area analyzed(Wiley et al. 2003, McNyset 2005). Other considerations include non-uniformity of presence locations (geographical bias), and biological factors such as local ecological adaptations and genetic diversity (Peterson 2001, Stockwell and Peterson 2002, Rice et al. 2003, McNyset 2005). For example, modelling of *Francisella tularensis* genotypes in the US yielded overlapping, yet different, geographic predictions and ecological associations (Nakazawa et al. 2010). Interestingly, this difference was apparent at intermediate, as opposed to coarse, phylogenetic levels. Similarly, Fisher et. al. showed that three genotype categories of the broadly distributed pathogenic fungus *Penicillium marneffei* correlated with environmental heterogeneity across Vietnam (Fisher et al. 2005). Using GARP, the genotypes classes were predicted to occupy three non-overlapping geographic areas. As a consequence of *B. anthracis* being a widely distributed species, models of anthrax outbreaks are subject to similar limitations in accuracy. Genotype specific models may therefore have improved accuracy and predictive power, and should be explored at multiple phylogenetic levels.

It is worthwhile to evaluate modelling limitations which could potentially explain differences between experiments. Despite studies showing GARP to be robust when predicting new landscapes (Ron 2005, Peterson et al. 2007, Blackburn 2010), our use of a projected modelling strategy may wrongly predict geographic and ecological distribution given the large geographic area in question. However, projected models created using southern outbreak points show similar geographic predictions as the outbreak-soil experiment, supporting that the broader geographic A1.a sub-lineage prediction is not simply an artefact of the modelling technique. That the large and small southern outbreak experiments showed a lesser degree of model agreement than the outbreak-soil experiment and that the large outbreak subsetting procedure had improved spatial homogeneity over the small southern outbreaks subsetting indicates that projected models may be sensitive to issues of sample size and clustering. Issues relating to sample size also appear to be important in our study. The predicted geographic distribution of *B. anthracis* varied among the 10 random data splits for both the Al.a sub-lineage and small southern outbreak experiments. In contrast, the same random data splitting procedure performed with the large southern outbreak and the outbreak-soil experiments showed spatial homogeneity among models(Joyner 2010). Although Stockwell and Peterson and Hernandez found that as few as 10 presence points are adequate for accurate GARP models(Stockwell and Peterson 2002, Hernandez et al. 2006), additional work has shown that certain species or geographic scenarios are more sensitive to sample size than others and that model accuracy can be sensitive to both sample size and extent of the species' range (Peterson et al. 2002a). Here, any effect of sample size is likely exaggerated by projection onto a large

geographic area and the relatively limited resolution the data (Peterson et al. 2002a, Hernandez et al. 2006).

Some additional limitations apply to our findings. The A1.a sub-lineage isolate collection was derived from multiple species, including livestock and humans, and from soil samples. The outbreak data, however, were derived from livestock only. The impact of this on model comparisons is unknown since associations between host, genotype and environment are as yet unexplored. We are limiting our modelled area by political boundaries as opposed to biogeographic limits. Finally, the genotyped isolate collection is geographically biased towards the southeast portion of Kazakhstan and spans a relatively long period of time(Aikembayev et al. 2010). Sampling bias is common in niche models using historical collections and can create artificial patterns in the data, although GARP is arguably less sensitive than other modelling algorithms to spatial bias (Stockwell and Peters 1999, Peterson 2001, Araujo and Guisan 2006, Blackburn et al. 2007). Future genotyping of additional isolates from under-sampled portions of the country, particularly the northern oblasts, will be essential to better characterize the genetic diversity and ecology of anthrax in Kazakhstan, allowing the construction of more refined predictive models.

The inclusion of available soil variables resulted in subtle changes in the predicted geographic distribution of anthrax in Kazakhstan, but the experiment is limited by the nature of available soil variables. Standardized soil variables and finer resolution data will be essential to characterizing the importance of soil parameters in *B. anthracis* persistence. The A1.a sub-lineage experiment showed a larger geographic and ecological distribution than the outbreak based experiment. Understanding genetic-

environmental associations will be essential to accurate modelling of anthrax for use in disease prevention and control in Kazakhstan.

Table 2-1. Environmental coverages used for GARP models.

Environmental Variable (unit)	Name	Source
Elevation (m)	Altitude	WorldClim*
Annual Temperature Range (^{0}C)	BIO7	WorldClim
Annual Mean Temperature (^{0}C)	BIO1	WorldClim
Precipitation of Driest Month (mm)	BIO14	WorldClim
Precipitation of Wettest Month (mm)	BIO13	WorldClim
Annual Precipitation (mm)	BIO12	WorldClim
NDVI₮ Amplitude (no units)	wd1014a1	TALA†
Mean NDVI (no units)	wd1014a0	TALA
Soil pH (-log(H^+))		HWSD‡
Topsoil Calcium (% weight)		HWSD
Topsoil Organic Content (% weight)		HWSD
Subsoil Base Saturation (%)		HWSD

*(www.worlclimorg)
† Trypanosomiasis and Land Use in Africa (TALA) research group
‡Harmonized World Soil Database

Table 2-2. Summary of experiments

Experiment	External Data Split (%Training/%Testing)	Area used for Model Building	Locality Data
Outbreak-Soil	85/15	All Kazakhstan	All spatially unique livestock outbreaks
A1.a Sub-lineage	80/20	Southern Polygon	Spatially unique A1.a isolates in southern polygon
Small Southern Outbreak	85/20	Southern Polygon	Random sub-set of spatially unique livestock outbreaks in southern polygon
Large Southern Outbreak	80/15	Southern Polygon	All spatially unique livestock outbreaks in southern polygon

Table 2-3. Sample sizes and accuracy metrics for GARP model building and evaluation

Metric	Outbreak-Soil	A1.a sub-lineage	Model Large Southern	Small Southern Outbreak
N to build models*	218	26	113	26
N to test models	39	13	145	147
Total Omission	2.6	0.0	0.0	0.0
Average Omission	9.9	13.1	19.1	15.5
Total Commission	32.7	19.18	12.74	17
Average	58.4	66.11	49.26	56.35
AUC†	0.7188	0.6964	0.7401	0.7386
SE	0.0466	0.0817	0.2410	0.04
Z	90.94	4.4449	16.3284	16.6241

*N was divided into 50%training/50% testing for each model iteration
†AUC = area under the curve

Table 2-4. Rules types from ten best models of the outbreak-soil and A1.a sub-lineage experiments. Values shown are number of rule types in the rule set (column %).

	Outbreak-Soil Rule Set										
Rule Type	2	6	13	21	24	25	29	39	48	51	Total
Logit	17(34)	32(64)	34(68)	25(50)	22(44)	26(52)	36(72)	35(70)	25(50)	27(54)	279(55.8)
Negated Range	2(4)	1(2)	0(0)	0(0)	2(4)	2(4)	0(0)	1(2)	7(14)	7(14)	22(4.4)
Range	31(62)	17(34)	16(32)	25(50)	26(52)	22(44)	14(28)	14(28)	18(36)	16(32)	199(39.8)

	A1.a Sub-lineage Rule Set										
Rule Type	1	10	21	40	49	51	54	76	91	93	Total
Atomic	1(2)	2(4)	1(2)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	4(0.8)
Logit	15(30)	5(10)	22(44)	20(40)	30(60)	2(4)	7(14)	24(44)	22(44)	9(18)	156(31.2)
Negated Range	10(20)	0(0)	4(8)	2(4)	0(0)	0(0)	0(0)	7(14)	0(0)	0(0)	23(5.6)
Range	24(48)	43(86)	23(46)	28(56)	20(40)	48(96)	43(86)	19(38)	28(56)	41(82)	317(63.4)

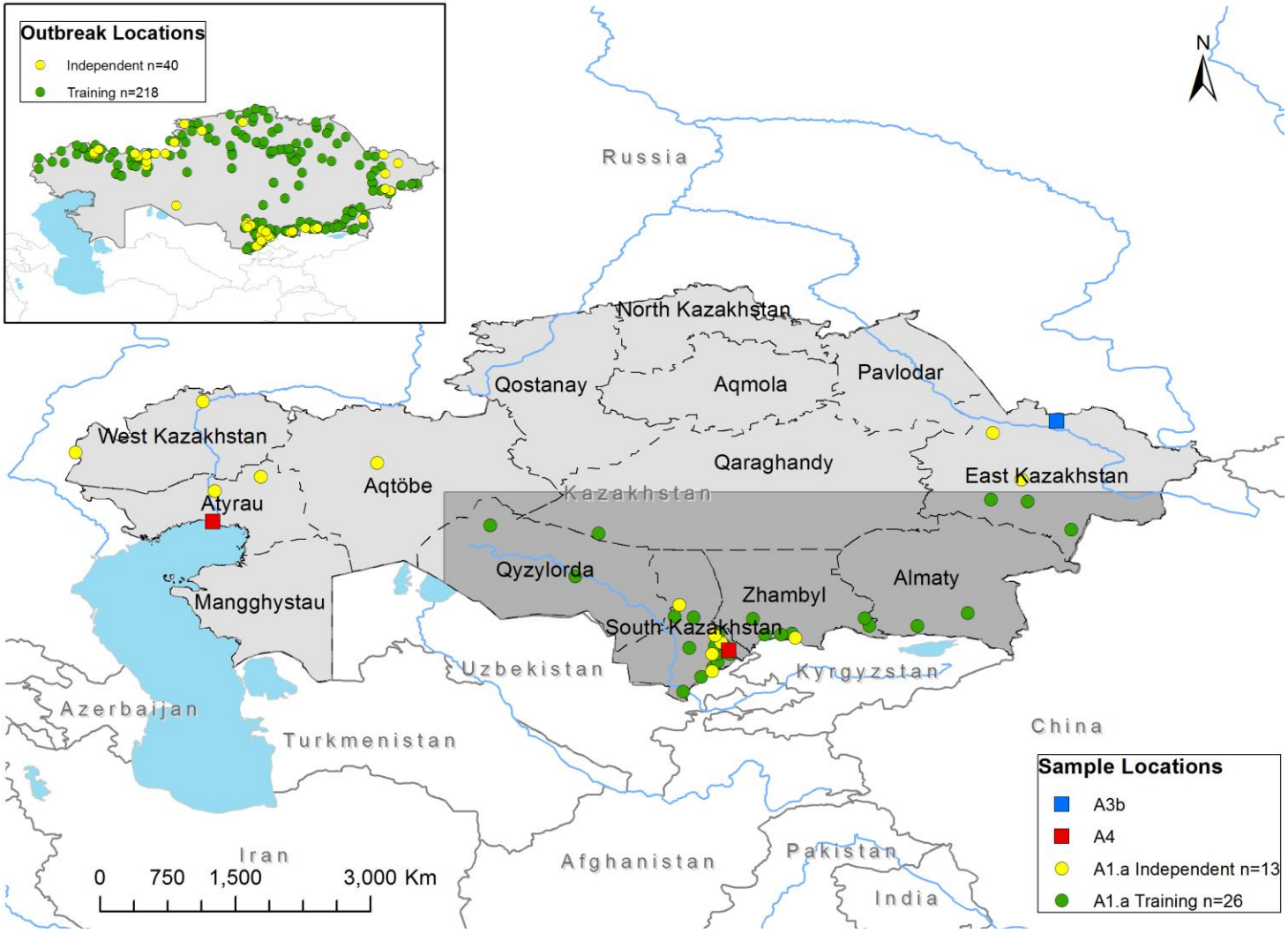

Figure 2-1. Map of Kazakhstan with anthrax locality data and locations of genotyped isolates. A) Points used for training and testing of sublineage model. B) Locations of outbreaks used for the full outbreak model. Shaded area indicates southern polygon used for creating projected models.

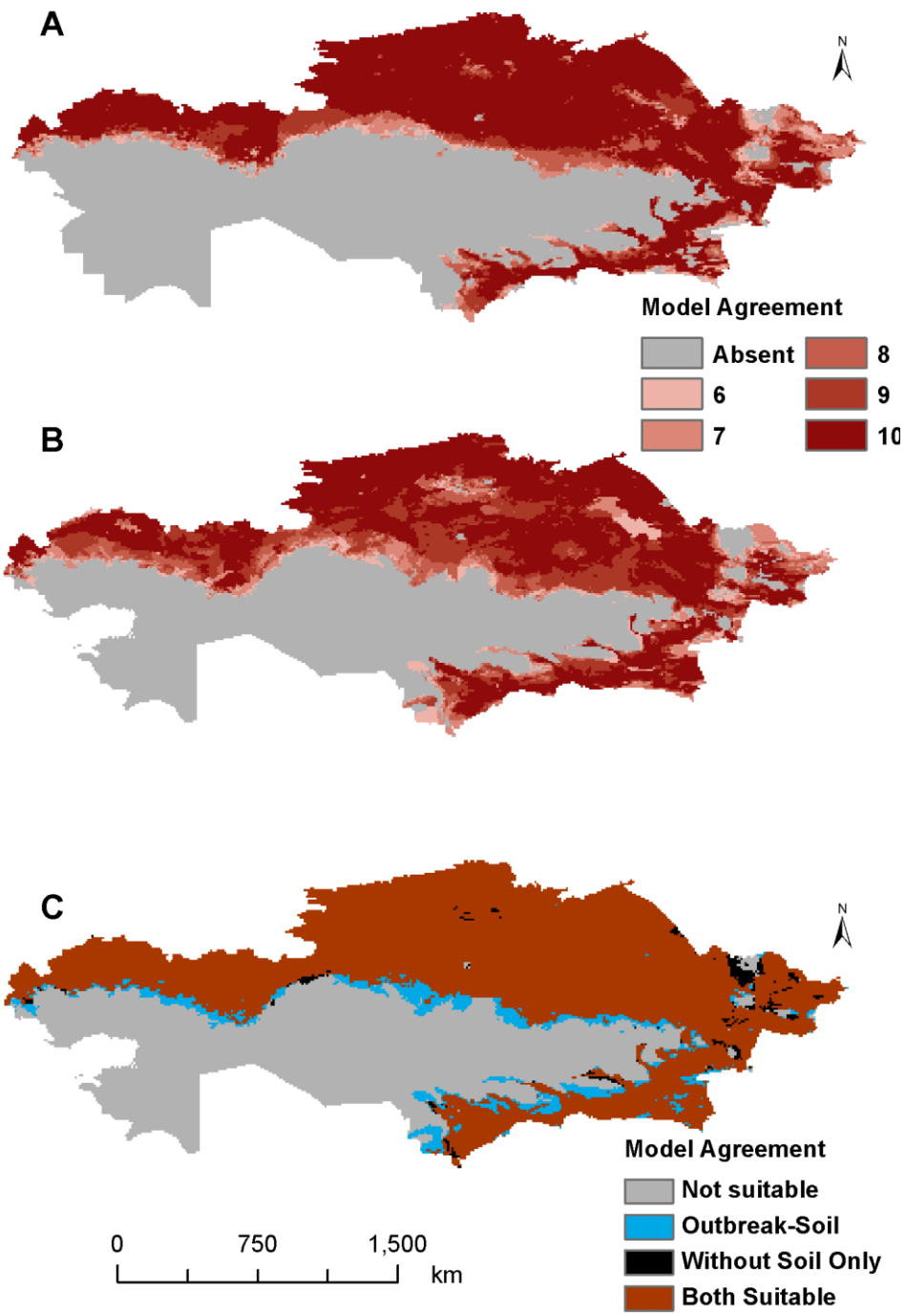

Figure 2-2. Predicted distribution of Bacillus anthracis in Kazakhstan based on outbreak data with and without soil variables. A) Outbreak experiment (excluding soil variables), B) Outbreak-soil experiment (including soil variables), C) Differences between distribution predicted by the two experiments.

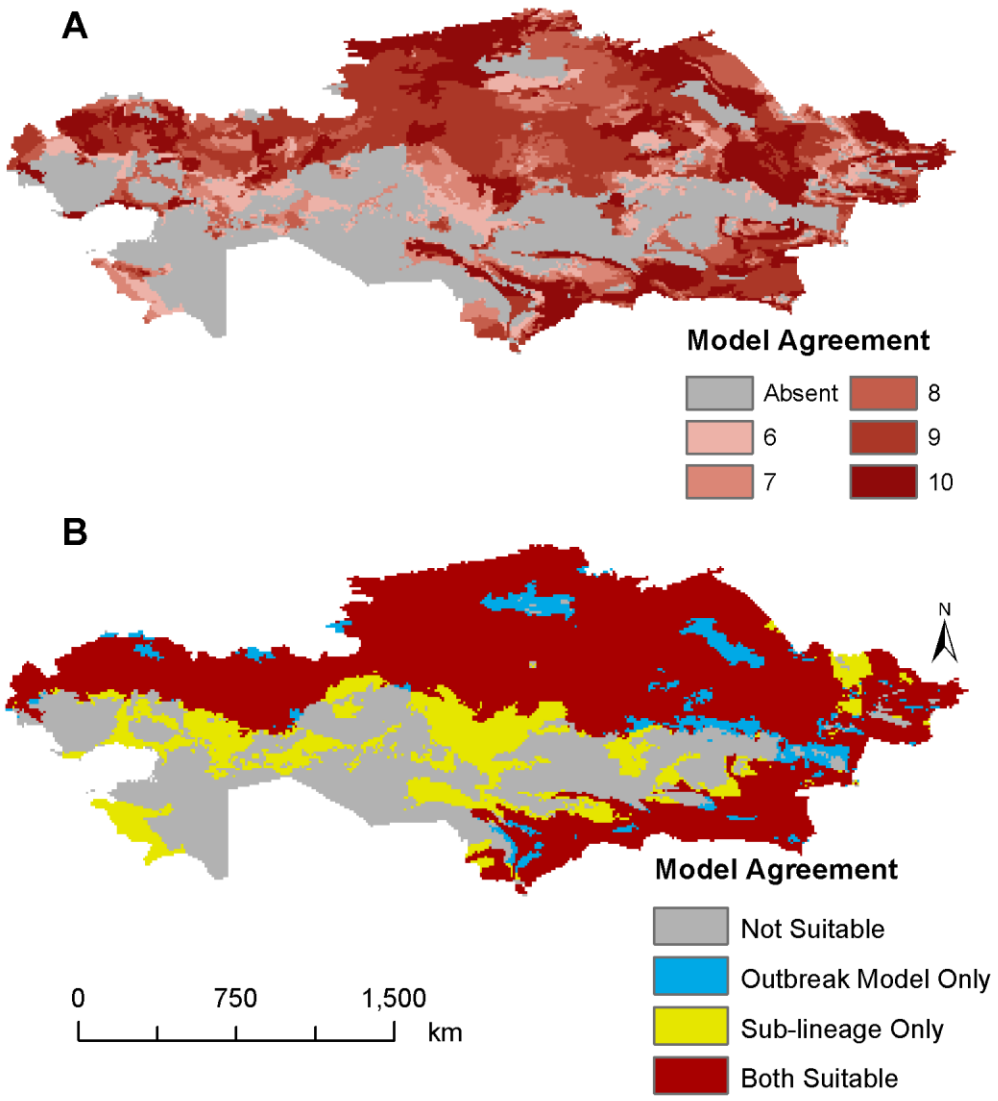

Figure 2-3. Comparison of predicted geographic distribution of B. anthracis. A) distribution of *B. anthracis* predicted by the sub-lineage experiment, B) difference between predicted distribution of the sub-lineage and the outbreak-soil experiments.

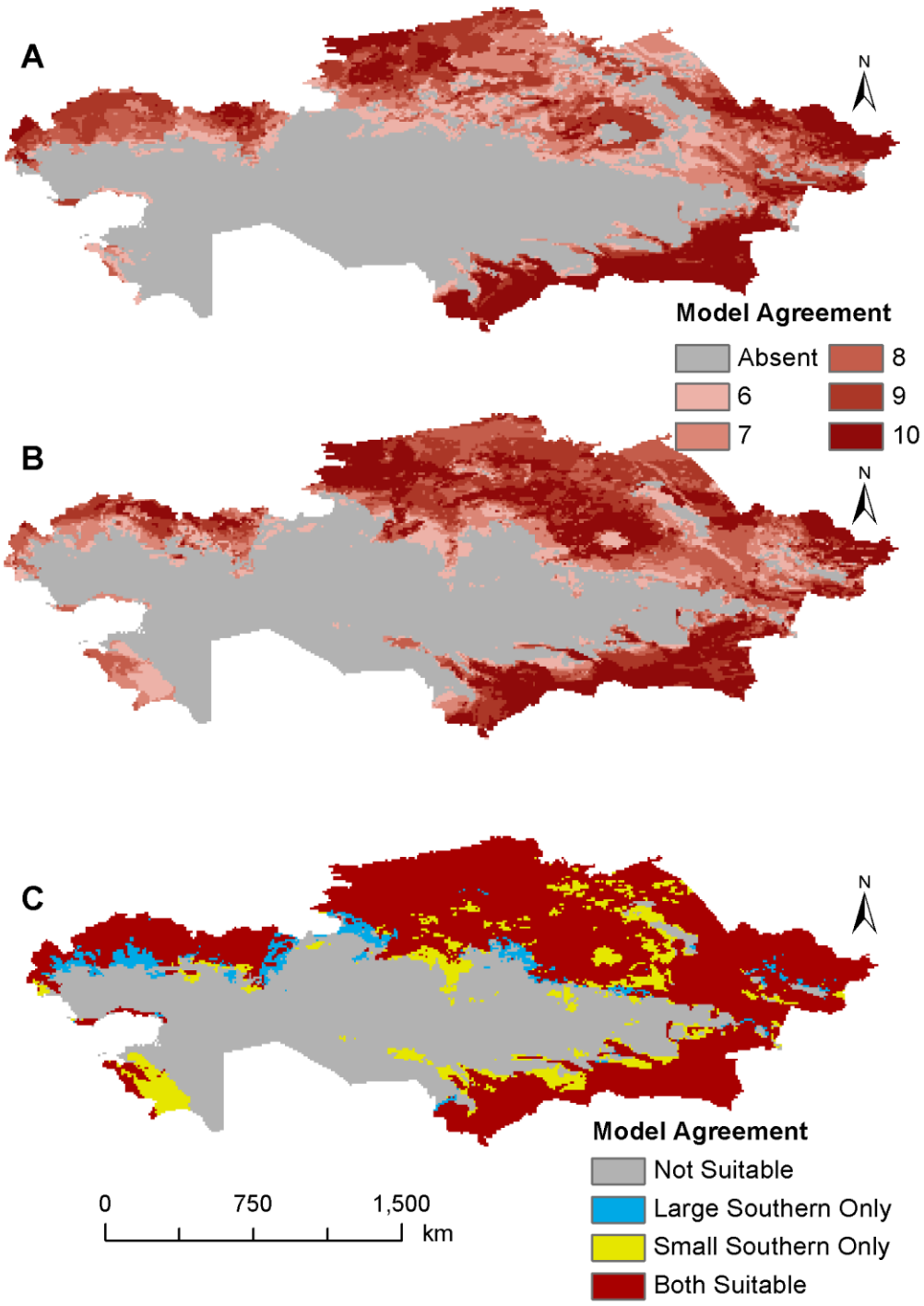

Figure 2-4. Predicted geographic distribution of *B. anthracis* based on A) large southern outbreak experiment and B) small southern outbreak experiment. C) Difference between predicted geographic distributions.

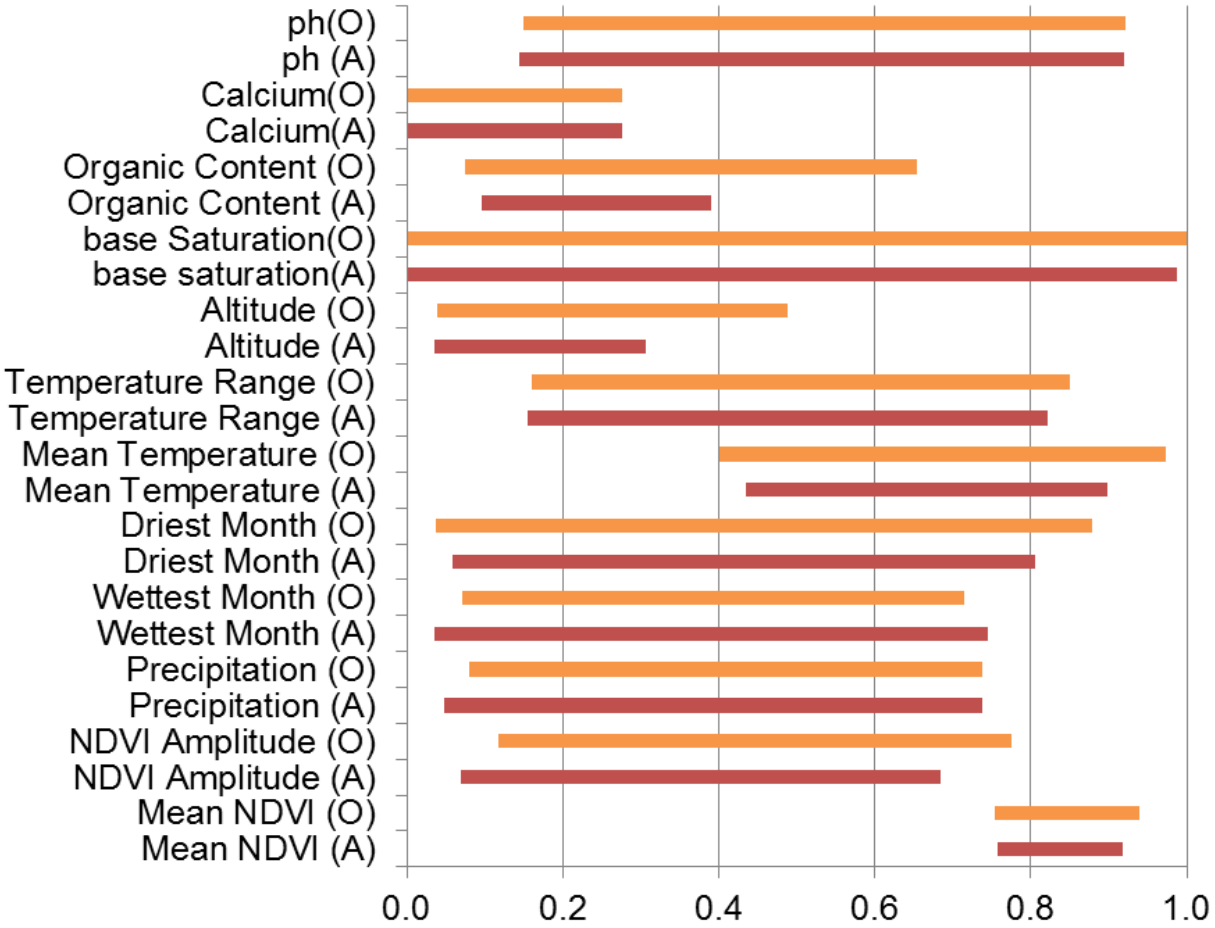

Figure 2-5. Median ranges of environmental variables predicting *B. anthracis* presence by the outbreak-soil experiment. O=Outbreak-Soil Experiment; A=A1.a sub-lineage experiment

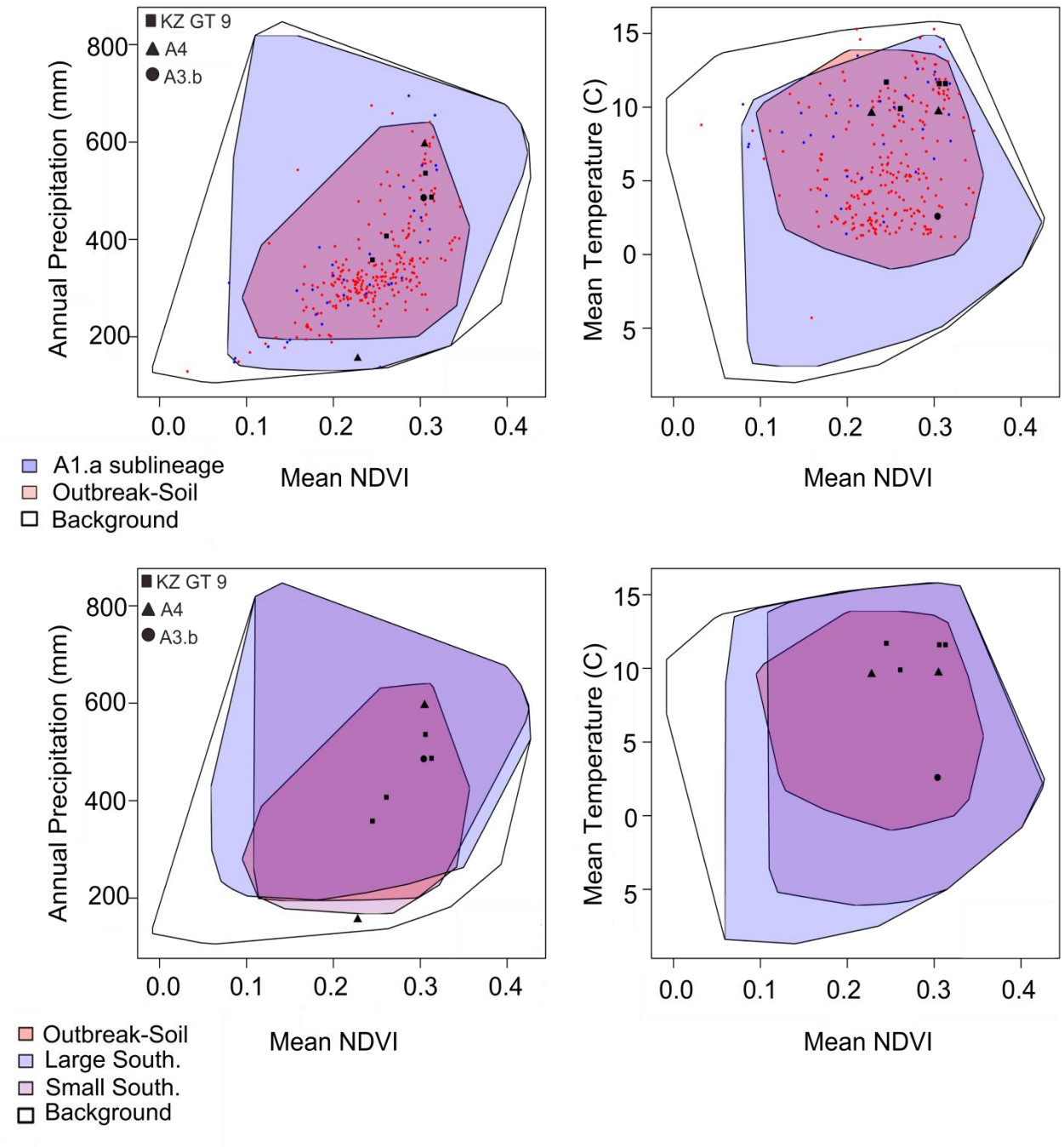

Figure 2-6. Predicted distribution of *B. anthracis* in ecological space based on areas of ten best subset model agreement. Red points = outbreak locations, blue points = A1.a isolate locations.

CHAPTER 3
ECOLOGICAL NICHE MODELING OF *BACILLUS ANTHRACIS* ON THREE CONTINENTS: EVIDENCE FOR GENETIC-ECOLOGICAL DIVERGENCE?

Introduction

Bacillus anthracis is a soil-borne, spore forming bacteria and the causative agent of anthrax in wildlife, livestock and humans worldwide. Metabolically dormant *B. anthracis* spores can persist in landscapes with suitable soil and ecological characteristics for long periods of time (Hugh-Jones and Blackburn 2009) such that years may pass between outbreaks. More recent evidence suggests *B. anthracis* may have an active soil life cycle (Schuch and Fischetti 2009). Despite a long recorded history of anthrax (Sternbach 2003), the environmental and epidemiological catalysts for epizootics are poorly understood. Control of the disease in livestock and humans is best achieved through annual vaccination of livestock and adequate surveillance to identify outbreaks early in the epidemic course (Blackburn et al. 2007). In wildlife populations, control is limited to active surveillance and proper disposal of carcasses (Blackburn et al. 2010a). The economic conditions in many anthrax endemic areas, however, combined with expansive rural geographies are such that early recognition of outbreaks and proactive distribution of vaccine is challenging, but could be facilitated by active, targeted efforts in areas of high likelihood of occurrence. Ecological niche models of *B. anthracis* identify geographic areas suitable for pathogen persistence, and these areas should be considered priorities for surveillance and vaccination. In particular, areas predicted to be supportive of the pathogen which also have a history of outbreak clusters should be targeted for active control (Kracalik et al. 2012).

Ecological niche models are frequently used to predict the potential distributions of species in ecologic and geographic space. A variety of correlative algorithms are

available, all of which aim to identify non-random relationships between known location data for the species and environmental variables and then identify geographic areas of predicted presence. Models developed on a native landscape and determined to accurately predict the species' presence in the native range can be projected, or transferred, onto novel geographical areas. Transferred models have been used to predict suitable areas for invasive species (Peterson 2003, Mukherjee et al. 2012), to predict the effects of climate change on species' range (Nakazawa et al. 2007, Joyner et al. 2010), and to test niche conservatism hypotheses(Broennimann et al. 2007, Mukherjee et al. 2011). Ecological niche modeling has been incorporated into phylogeographic studies by evaluating whether genetically defined sub-populations are associated with divergence in ecological niche and geographic distribution, as in Chapter 2 (Rice et al. 2003, Fisher et al. 2005, Rissler and Apodaca 2007, Nakazawa et al. 2010). Combining ENM and phylogeography is particularly informative for studies of globally distributed pathogens which have intricate, often little understood, interactions with spatially heterogeneous abiotic and biotic factors which may be linked to genetic variation. From a methodological perspective, widely distributed species tend to result in less accurate ecological niche models (McNyset 2005), and models for such species can be improved by dividing the larger population into biologically meaningful sub-populations such as those defined by genetic analysis (Fisher et al. 2005, Gonzalez et al. 2011).

The phylogeography of *B. anthracis* has been described at the global level (Van Ert et al. 2007a) and in multiple regions (Smith et al. 2000, Fasanella et al. 2005, Simonson et al. 2009, Aikembayev et al. 2010) using a combination of genetic markers

including single nucleotide polymorphisms (SNPs) (Van Ert et al. 2007a) and multiple locus variable number tandem repeats (MLVA) (Keim et al. 2000, Lista et al. 2006, Van Ert et al. 2007a). Although the geographic distribution of genetic lineages as defined by SNP and MLVA analysis exhibits heterogeneity, globally successful lineages can dominate strain collections in countries on separate continents. For example, the MLVA defined A1.a is widely distributed throughout North America and Eurasia (Van Ert et al. 2007a); this sublineage was further defined using SNP analysis into the trans-Eurasian ancestor (TEA) and the related Western North American (WNA) group (Van Ert et al. 2007a).

The possibility that genetic variation in bacterial pathogens is associated with spatial and ecological divergence is supported by studies demonstrating unique epidemiologic characteristics or ecological affinities among genetic groups of pathogens within a geographic region (Fisher et al. 2005, Kugeler et al. 2009, Nakazawa et al. 2010). In the case of *B. anthracis*, a study in Kruger National Park, South Africa (Smith et al. 2000) revealed intriguing genetic-ecological associations for the pathogen and a differential distribution of genotypes based on soil characteristics. In this study, Smith et al.(2000) detected spatial and ecological differences between genotypes representing the evolutionarily distinct *B. anthracis* A and B branches within the relatively narrow geographic boundaries of Kruger National Park. More recently, In Chapter 2 I suggested that *B. anthracis* genotypes belonging to A1.a sublineage in Kazakhstan were associated with a broader ecological space than the larger population containing multiple A cluster sublineages, further supporting the importance of genetic information in building relevant ENMs for the species .

In the countries of United States, Italy and Kazakhstan, strains belonging to the A1.a sublineage are ecologically established and dominate strain collections (Keim et al. 2000, Fasanella et al. 2005, Lista et al. 2006, Van Ert et al. 2007a, Aikembayev et al. 2010), providing a unique opportunity to study the dynamics of this successful group across diverse landscapes. In the United States, ecological niche modeling of anthrax outbreak locations predicted pathogen persistence primarily along a narrow corridor running from southwest Texas northward through the Dakotas (Blackburn et al. 2007). In that landscape the A1.a sublineage (SNP group WNA) is dominant, although small areas appear to support other sublineages, including A3.b and A4. The genetic diversity of *B. anthracis* and geographic distribution of anthrax outbreaks in Italy was described by Fasanella et al. (2005), Lista et al.(2006) and Van Ert et al. (2007a), and confirmed by Garofolo et al. (2011). The dominant genotypes in Italy fell into the MLVA A1.a group and TEA SNP group. Isolates belonging to the A1.a/TEA sublineage also predominate in the southern portion of Kazakhstan (Aikembayev et al. 2010), where the ecological niche and predicted geographic distribution has been modeled (Joyner 2010). I developed ecological niche models of this globally successful *B. anthracis* sublineage in the United States, Italy and Kazakhstan. Models were reciprocally transferred to determine if pathogen presence could be accurately predicted on novel landscapes.

Methods

Anthrax Occurrence Data and Environmental Data

Geographic information system (GIS) databases of *B. anthracis* isolates belonging to the MLVA-defined A1.a sublineage and SNP defined TEA/WNA from Italy, the United States (U.S.) and Kazakhstan were used to support these analyses. All

isolates were derived from collections previously genotyped using MLVA typing systems containing the set of eight markers (MLVA-8) described in Keim et al. (2000) and SNP analysis (Van Ert et al. 2007a). The U.S. isolates were derived from Kenefic et al. (2009), Blackburn et al. (2007) and new strains from recent field collections in western Montana and Texas (Blackburn, unpublished data). The isolates from Italy were first reported in Fasanella et al. (Fasanella et al. 2005) and mapped for this study. Several isolates from Italy were mapped only to the nearest province and, therefore, I removed the isolates from the analysis. Kazakh data were the same as described in Chapter 2. Isolates were independently grouped into genetic lineages using the unweighted pair group method with arithmetic mean (UPGMA) cluster analysis or canSNP group assignments (Fasanella et al. 2005, Van Ert et al. 2007a, Aikembayev et al. 2010). Remaining occurrence data were reduced to spatially unique points at an 8 km^2 resolution to accommodate the spatial uncertainty of the data (Joyner et al. 2010). Each occurrence dataset was randomly divided into an 80% training set for model building and a 20% dataset for testing the native projection. Figure 3-1 shows the occurrence point distributions for each of the countries used in this analysis.

Grids representing six bioclimatic variables (Table 3-1) and altitude were downloaded from WorldClim (www.worldclim.org) and two satellite-derived environmental variables describing measures of vegetation were obtained from the Trypanosomiasis and Land Use in Africa (TALA) research group (Oxford, United Kingdom; Table 3-1) (Hijmans et al. 2005, Hay et al. 2006). WorldClim bioclimatic grids (Bioclim) may be more biologically meaningful than annual mean, maximum and minimum values because the manipulation of monthly data results in variables that

represent annual trends and seasonality as well as extremes in environmental conditions that limit a species' range. All grids were resampled to 8 km^2 and clipped to country boundaries using ArcView 3.3 with the GARP datasets extension (Environmental Systems Research institute, Redlands, California, USA). The eight variables in the environmental dataset were chosen based on parameters reported to reflect persistence of *B. anthracis* in soils and used in previous studies {Blackburn, 2007 #187;Joyner, 2010 #192;Joyner, 2010 #18;Mullins, 2011 #366}. The background extent of each country was determined by political boundaries. These extents were chosen because reporting and control of anthrax is conducted within political boundaries.

Model Development

This study employed the Genetic Algorithm for Rule-Set Prediction (GARP) (Stockwell and Peters 1999) with a best subset procedure to perform the ecological niche modelling experiments (Anderson et al. 2003). Briefly, the GARP approach uses a two-step procedure in which sets of rules are developed iteratively to predict presence or absence in variable space using presence only input data and background data. These rule sets are combinations of if/then statements derived from either variable ranges or logistic regression functions. This process is a random walk and develops multiple models; the best subset procedure aids in the selection of optimal models based on user-defined thresholds of omission and commission. Optimal rule sets are then projected onto the landscape. Training data were input into GARP with a 50% training/50% testing internal data partition. For all experiments, I specified 200 models with a maximum of 1,000 iterations and a convergence limit of 0.01. The 10 best subset models were selected using a 10% hard omission threshold and a 50% commission

threshold. These output models were imported into ArcMap 10 (Environmental Systems Research institute, Redlands, California, USA) and summated to generate a single cumulative raster file of model agreement for *B. anthracis* presence. Grid cell values thus ranged from 0 (all models predict absence) to 10 (all models predict presence). Native models were trained in each of the three countries, then rule sets were projected onto the two other landscapes (for example, the model trained in the U.S. was projected onto Kazakhstan and Italy). As described in Chapter 2 , Kazakh models were trained with a subset of southern A1.a isolates and projected onto the entire Kazakh landscape, the U.S. and Italy.

Model Evaluation

Predictive performance of the best model subset was evaluated with an area under the curve (AUC) in a receiver operating characteristic (ROC) analysis using withheld independent test data for native models and all data for projected models. Values of AUC approaching one indicate a well-performing model while an AUC equal to 0.5 indicates the model performs no better than random and are tested statistically with a z-score (Z) and standard error (SE) estimates. AUCs were interpreted in conjunction with measures of omission and commission calculated using the summated 10 best models (Lobo et al. 2008, Peterson et al. 2008). Total commission is the percent of pixels which are predicted as presence by areas of 10 model agreement. Average commission is the average area predicted as presence by all subset models. The greater the difference between the 2 measures of commission, the greater the spatial heterogeneity among the 10 best subset models (McNyset 2005). Two measure of omission were also calculated. Total omission was calculated as the total number of test points falling into areas predicted as absence by all 10 models. Summed area

omission (SAO) was calculated as the omission error of areas of 10 model agreement. Models with small SAO values are desirable because complete model agreement represents the most conservative threshold with which to predict areas of presence. Models which most robustly predict areas with a high likelihood of pathogen persistence facilitate implementation of cost effective, focused public health measures such as surveillance and pre-emptive vaccination.

Results

All experiments reached convergence of accuracy prior to the maximum 1,000 iterations. The AUC scores for native projections all performed significantly better than random, and native models each had zero total omission and low average omission (Table 3-2). The native U.S. model predicted *B. anthracis* distributed in a north-south corridor in the center of the country (Figure 3-2). This band widens as it moves from southwest Texas northward into South and North Dakota and eastern Montana. Areas of the interior northwest are also predicted. The native Italian model predicted large sections of the southeastern mainland as well as the islands of Sardinia and Sicily. Areas of high likelihood also include coastal regions in central Italy and two relatively isolated regions of the northeastern and central north portions of the country. The Italian model accurately predicted areas of provinces in the northeast where A1.a strains have been documented, but were excluded from the analysis because of imprecise GIS data. In Kazakhstan, the native projection strongly predicted presence along the mountainous region of the southern portion of the country and a broad area in the north, while predicting the interior of the country with low model agreement.

Models projected to novel landscapes performed poorly. Transferred models which performed better than random were poor, and some transferred models were

more dispersed than random; measures of omission and commission were consistent with under-prediction by the Italian and U.S. models and over-prediction by the Kazakh model. (Figure 3-2, Table 3-2). The native U.S. model when transferred to Italy predicted similar geographic areas of Italy as the native Italian model, but with very low model agreement. The transferred model failed to predict the area in southern Basilicata which has experienced anthrax outbreaks and gave low agreement for known endemic areas. The native U.S. model transferred to Kazakhstan broadly predicted, with low model agreement, almost the entire landscape. Small areas in the southern and northern regions were predicted with higher model agreement. These areas were also predicted by the native Kazakh model, whereas in the northeast corner of the country, the U.S. and Kazakh models were very different.

The model trained in Italy transferred to Kazakhstan predicted a highly endemic area of southern Kazakhstan, although with very low model agreement. Projected onto the U.S., the model trained in Italy did highlight some areas predicted by the native trained model, namely western Texas, western Oklahoma and Colorado, but again with low model agreement. When transferred, the models trained in Kazakhstan over-predicted all of Italy and the majority of the landscape in the U.S.

Discussion

Evidence increasingly suggests phylogenetic analysis provides a meaningful way to subdivide *B. anthracis*, and other species, for more accurate niche modeling (Smith et al. 2000, Fisher et al. 2005, Rissler and Apodaca 2007, Gonzalez et al. 2011). Furthermore, an ability to apply models developed in a landscape with known anthrax locations to one in which anthrax is not reported or is poorly recognized would be of great value in predicting potential areas of emerging disease. This study tested whether

ecological niche models of the *B. anthracis* MLVA-8 defined A1.a sublineage can be used to predict the distribution of the same sublineage on novel landscapes. The native models reasonably predicted the occurrence points in these experiments, and furthermore the native Italian model predicted a region in Italy where the A1.a sublineage has been isolated, but due to lack of spatial data was not included in the occurrence dataset. Transferred models, however, failed to accurately predict documented anthrax occurrence points and tended to over-predict or under-predict presence on the non-native landscapes. Where transferred models were successful in predicting areas of known persistence, model agreement tended to be low. Both over-prediction and under-prediction of transferred models are considered failures, although these failures have different practical consequences for public health applications. When areas suitable for pathogen presence are not predicted by a model, the failure to incorporate this area into surveillance programs will result in cases of disease being overlooked. Over-prediction by a model, on the other hand, hinders epidemiologic investigations of cases by misallocating resources to monitor regions that are erroneously predicted to support the pathogen.

The failure of transferred models to accurately and consistently predict known occurrences presents an interpretive challenge. The lack of transferability may have resulted from methodological shortcomings of transferring models, or could reflect genetic-ecological divergence of the pathogen (Fisher et al. 2005, Broennimann et al. 2007, Rissler and Apodaca 2007, Peterson and Nakazawa 2008, Rödder and Lötters 2010). Evidence supporting genetic-ecological divergence was demonstrated in Kruger National Park, South Africa, where the A lineage had broader ecological tolerances than

the B lineage within the park (Smith et al. 2000). These results suggest that this ecological divergence in *B. anthracis* strains may also occur in a widely distributed sublineage. It is possible that successful genetic groups with broader tolerances are more likely to become established across ecological extremes, and the local population would then differentiate to develop a unique genetic signature associated with the local ecology (Holt et al. 2005). This process of regional-scale differentiation leading to spatially structured genetic variation has been described among other introduced species or pathogens (Fisher et al. 2005, Broennimann et al. 2007, Kenefic et al. 2009, Carrel et al. 2010, Nakazawa et al. 2010). In contrast, *B. anthracis* lineages with more limited tolerances, such as the B lineage and, although its ecological associations have not been characterized, the geographically limited C lineage, would be restricted ecologically and geographically. That Italian and U.S. models under predicted pathogen occurrence in Kazakhstan, whereas Kazakh models overpredicted a large extent of Italy and the U.S., may indicate that Kazakh strains have adapted to a broader ecological envelope. Although more complete genomic data are required to substantiate this hypothesis, this may reflect an earlier introduction of the sublineage into Kazakhstan than in Italy or North America.

Evidence for genetic differentiation within the classically defined MLVA A1.a sublineage includes the separation of this sublineage into the distinct TEA and WNA SNP groups (Van Ert et al. 2007a). More recent SNP analysis of the WNA lineage (Kenefic et al. 2009) suggests a significant evolutionary divergence between the North American (U.S.) and Eurasian (Italy and Kazakhstan) strains. The evolutionary divergence between Eurasian strains in Kazakhstan and Italy is not as well defined.

However, a 15-marker VNTR analysis of the Eurasian group reported by Van Ert et al. (2007a) suggests a high level of genetic diversity exists within the 'TEA' SNP group (A.Br.008/009). More recent VNTR based analyses based on 25 markers indicates that the Kazakh 'A1.a' strains, as previously defined using the 8-marker system, exhibit a considerable degree of genetic divergence (Sytnik, unpublished data) from European strains. Although this VNTR based diversity is intriguing, more comprehensive analysis is required to more accurately measure the evolutionary relationships between these populations, particularly considering that discovery bias inherently limits resolution in 'canonical' SNP data (Pearson et al. 2004). Despite the limitations of the genetic data used in the present analysis, the finding that native models were successful while transferred models failed to predict anthrax occurrence points suggests that niche specialization may have occurred within this broadly distributed sublineage.

From a modeling perspective, however, the effect of variable selection (Araujo and Guisan 2006, Peterson and Nakazawa 2008, Jiménez-Valverde et al. 2009, Rödder and Lötters 2010) and background (Austin 2007, Elith et al. 2010, Barve et al. 2011) on transferred projections must be considered as potential methodological limitations when interpreting the results of this study. The variable set used in the current experiments was chosen based on variables used in Chapter 2 and considered to be limiting factors in the persistence of *B. anthracis* based on previous niche modeling efforts performed within single landscapes (Blackburn et al. 2007, Joyner 2010). In studies exploring the effects of variable selection on transferred ecological niche modeling experiments, changes in the dimensionality (Peterson and Nakazawa 2008, Rödder et al. 2009) and source (Peterson and Nakazawa 2008) of environmental datasets resulted in different

geographic predictions, and these difference were more pronounced in the novel landscapes than in the native ones. Variable selection can limit transferability because limiting factors for the species vary geographically (Austin 2007, Rödder and Lötters 2010), and, in addition, interactions among ecological variables may differ across landscapes, which would alter the relationship of more distal variables with the pathogen (Elith et al. 2010). Ecological variables limiting the spatial distribution of *B. anthracis* may also vary with genetic lineage, reflecting niche specialization suggested by the work of Smith et al. (2000)in Kruger National Park, South Africa. The impact of variable selection, as well as that of different techniques for selection of environmental variable sets, on these results will be evaluated in additional experiments.

A second methodological consideration is that of the extent used for background. I have defined the background in this study according to political boundaries instead of using other suggested techniques such as minimum convex polygons (Rödder and Lötters 2010), global ranges (Broennimann and Guisan 2008), or the accessible extent (Elith et al. 2010, Barve et al. 2011) because *B. anthracis* dispersal to novel areas is primarily anthropogenic and therefore not constrained by natural physical boundaries or biological limits to movement. Therefore the potential area of dispersal, as currently understood, is limited by control and prevention measures which follow political boundaries. Similarly, the nature of surveillance for, diagnosis of and reporting of anthrax outbreaks is dependent on policies which fall within such limits. Despite this practical consideration, however, I recognize that within the political boundaries used for defining the extent for these models, considering the large geographic extents and latitudinal differences between the countries, the ranges of environmental variables is

likely differ between landscapes. Hence, concerns about extrapolation are significant and should be addressed in additional experiments.

While I suggest the results presented here may reflect niche differentiation within this sublineage of *B. anthracis,* while considering the potential effects of methodological problems, it is important to also consider the source of isolate/occurrence data and the effects of control efforts on the disease in each country (Blackburn et al. 2007). It is possible that the broad native predictions within Kazakhstan reflect an insufficient capacity to maintain widespread vaccination to reduce the overall burden of anthrax (Woods et al. 2004). In the case of the U.S., models likely reflect areas where the disease persists after decades of widespread vaccination efforts have resulted in a clear reduction in overall outbreak numbers and an accompanying contraction in the spatial extent of disease (Blackburn 2006). Differences in historical control, alone or in synergy with genetic and ecological factors, could explain some portion of the over- and under-prediction observed in the transference of models.

In this study I have used measures of omission, commission and AUC as accuracy metrics to evaluate native models and their transferability. The AUC measure has limitations when evaluating ecological niche models (Lobo et al. 2008, Peterson et al. 2008). Two described limitations are that AUC is influenced by the geographic extent of the landscape being studied and the AUC uses the entire ROC plot which would render comparisons between modeling platforms unreliable. Here, however, models are evaluated over the same extent (ie, a native U.S. model and a projection transferred to the U.S.) and all using the GARP modeling platform. The optimum weighting of omission and commission, although equally weighted in the AUC, varies

depending on the purpose of the experiment. In this study ENM was used to predict the potential presence of a pathogen and infer subsequent disease risk to inform public health policy. Errors of commission, or over-prediction, will result in excess expenditures for surveillance and interventions and mislead trace-back efforts, whereas errors of omission, or under-prediction, could result in sustained transmission (Peterson et al. 2011). Viewed in this context, optimal weighting of omission and commission will depend on the relative costs of surveillance and disease. The summed area omission (SAO) allows for refinement of models with the goal being that areas of total model agreement can be used as a conservative threshold for predicted presence. By calculating omission only in areas of complete model agreement, this conservative evaluation of model performance maximizes predictive value without incorporating potentially large geographic areas of low model agreement and therefore lower risk. Expensive surveillance and prevention programs can then be effectively targeted to highest risk areas.

Previous ecologic niche models described conditions favorable for *B. anthracis* outbreaks based on collections of isolates from multiple genetic lineages that were likely biased towards a dominant subset of genotypes (Blackburn et al. 2007, Joyner 2010) and the potential for using genetic analysis to improve models was subsequently demonstrated in Chapter 2. . These findings suggest genetic-ecological divergence exists among geographically dispersed populations of *B. anthracis* from the MLVA-8 defined A1.a sublineage. Some caveats apply, however. More comprehensive and higher resolution genomic data is required to better characterize the genetic differences between these populations. In addition, the methodological problems discussed here

must be explored in order to refine native models and to evaluate whether variable selection methods will enhance the transferability of models. Until our understanding of the genetic-ecological dynamics of *B anthracis* is better developed, I suggest that *B. anthracis* is best modeled on a country or regional level and with consideration of the genetic diversity of the population. Moving forward, understanding *B. anthracis* genetic-ecological associations on the landscape will result in construction of ecological niche models that are sensitive to genotype and region and are more successful in predicting outbreaks.

Table 3-1. Environmental variables used to develop ecological niche models.

Environmental Variable (unit)	Name	Source
Elevation (m)	Altitude	WorldClim†
Annual Temperature Range (^{0}C)	BIO7	WorldClim
Annual Mean Temperature (^{0}C)	BIO1	WorldClim
Precipitation of Driest Month (mm)	BIO14	WorldClim
Precipitation of Wettest Month (mm)	BIO13	WorldClim
Annual Precipitation (mm)	BIO12	WorldClim
NDVI Amplitude (no units)	wd1014a1	TALA‡
Mean NDVI (no units)	wd1014a0	TALA

†(www.worlclimorg) (Hijmans et al. 2005)

‡Trypanosomiasis and Land Use in Africa (TALA) research group (Oxford, United Kingdom) (Hay et al. 2006)

Table 3-2. Sample sizes and accuracy metrics for all native models and projections. Native projection are under the curve (AUC) scores are shown in bold for comparison. SE = standard error, Z = z-score, SAO = summed area omission.

Training Landscape	United States			Italy			Kazakhstan †			
Projection	Native	IT	KZ	Native	KZ	US	Native	KZ	IT	US
Training	48	-	-	28	-	-	24	-	-	-
Testing	12	35	39	7	39	60	8	15	35	60
AUC	**0.93**	0.51	0.48	**0.84**	0.56	0.43	**0.90**	0.71	0.45	0.48
SE	0.05	0.05	0.05	0.09	0.05	0.04	0.09	0.08	0.05	0.04
Z	5.84‡	7.72‡	5.99‡	3.86‡	97.91‡	16.42‡	3.13‡	4.52‡	15.93‡	7.71‡
Total Omission	0	62.1	13.2	0	86.8	96.4	0	0	0	0
Average Omission	6.7	71	51	1.4	74.8	82.2	0	19.2	10	17.7
Total Commission	8.38	0	0.84	27.86	0	0	19.77	13.17	9.04	28.67
Average Commission	22.88	7.71	43.54	50.21	0.05	3.17	46.3	54.28	90.9	83.35
SAO	16.67	-	89.47	14.29	-	100	0	46.15	100	62.50

†Native training and testing are for southern training area only
‡statistically significant value

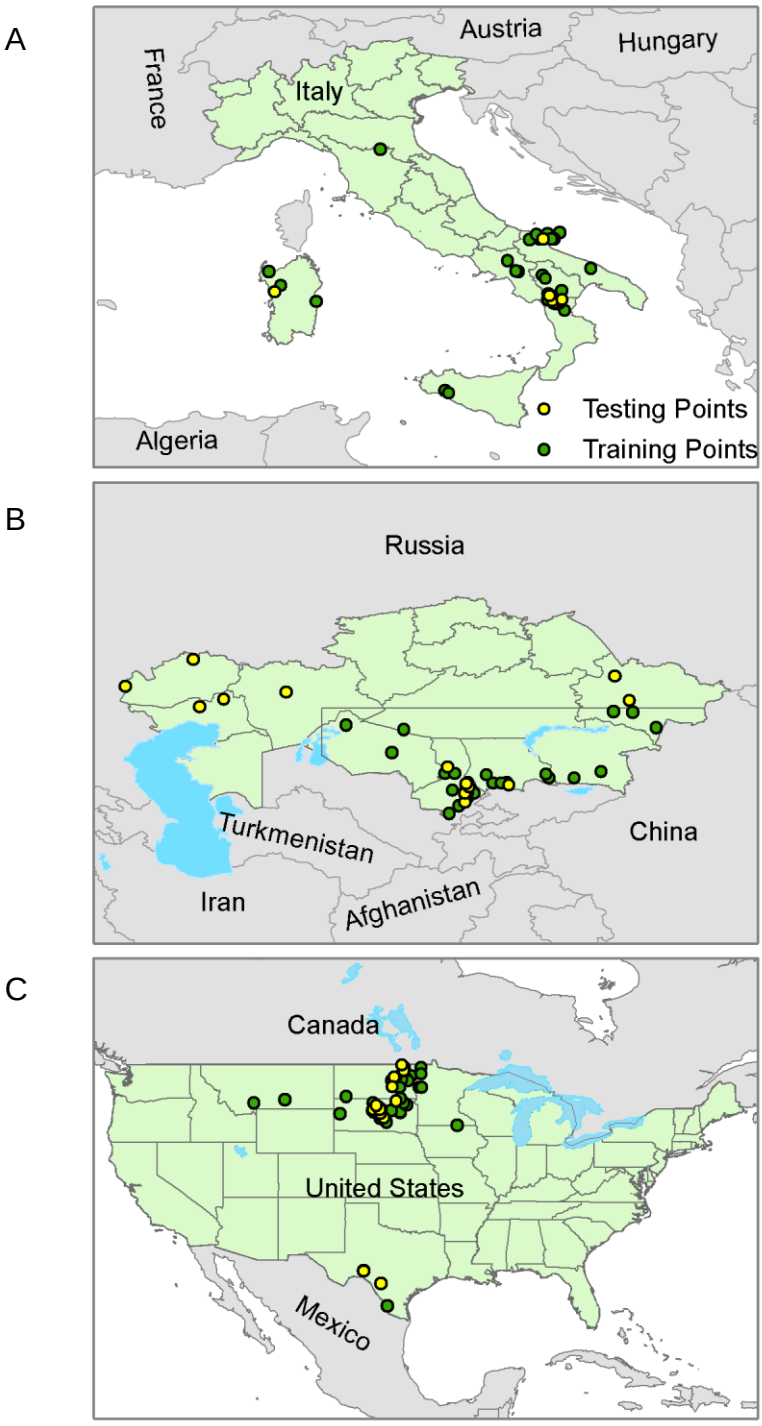

Figure 3-1. Geographic distribution of the training and testing points used for ecological niche model building and evaluation in A) Italy, B) Kazakhstan and C) United States

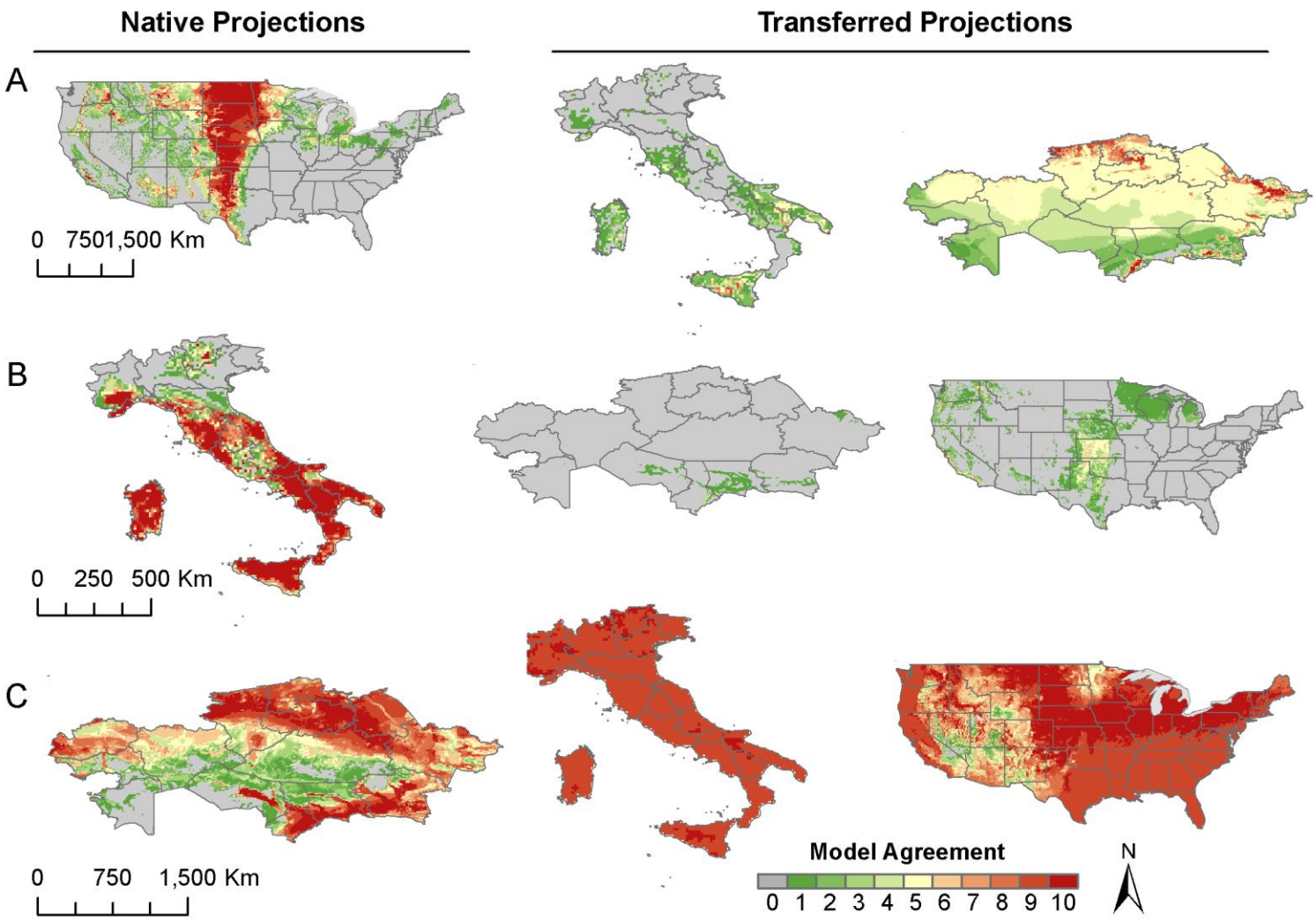

Figure 3-2. Predicted distribution of *Bacillus anthracis* by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

CHAPTER 4
ENVIRONMENTAL VARIABLE SELECTION AND TRANSFERRED ECOLOGICAL NICHE MODELS OF THE *BACILLUS ANTHRACIS* A1.A SUBLINEAGE

Introduction

The geographic distributions of many infectious diseases that affect animals and humans are determined by environmental, biotic and landscape variables. These factors drive outbreaks in endemic areas and influence where diseases will be able to emerge or re-emerge. Understanding the ecological drivers of each component of a disease system is essential to prediction and control of emerging and re-emerging zoonotic and vector borne diseases. GIS- based ecological niche models (ENMs) have been widely used to predict the potential distribution of species, including pathogen species (Peterson et al. 2002a, McNyset 2005, Blackburn et al. 2007, Nakazawa et al. 2010). ENMs are correlative techniques which find statistically significant associations between geo-referenced occurrence points and environmental variables producing presence/absence estimates. The models then assign a likelihood of occurrence to each grid cell in the study area as either a binary presence/absence or probability of presence to generate a prediction of the spatial distribution of the species. Such models are based on ecological niche theory as conceptualized by Grinnell (Grinnell 1917) and expanded by Hutchinson (Hutchinson 1957), in which the niche is defined by the n-dimensional hyperspace in which a species can maintain without immigration. Most applications of ENM are concerned with abiotic variables, including climate variables, which are not modified by the species and which operate on relatively coarse geographic scales (Soberón 2007).

According to some authors, ideal models are sufficiently general that they can predict a species' presence in novel geographic areas or time periods (Randin et al.

2006, Peterson et al. 2007, Jiménez-Valverde et al. 2009). This process of creating models and transferring those models in space and time has been applied to invasive species (Peterson 2003, Broennimann et al. 2007), climate change (Nakazawa et al. 2007, Blackburn 2010, Joyner et al. 2010), and infectious diseases (Peterson et al. 2002b, Blackburn et al. 2007, Maher et al. 2010, Nakazawa et al. 2010). An important potential use of transferred ENMs as applied to pathogens is predicting potential pathogen presence, as a proxy for disease risk, in geographic areas where disease is absent or unreported (Alexander et al. 2012). Successfully transferred models can be used to inform public health surveillance systems and epidemic planning (Blackburn et al. 2007). To develop generalized models with high transferability, models can be tested with known occurrences in spatial extents outside of the training area. Success of the models should be based on the accuracy of predictions in both the native landscape and the novel landscape as well as the similarity of native and transferred predictions in both areas (Randin et al. 2006). Typically, when ENMs are transferred in space or time, it is assumed that the species' climatic niche is conserved and the range of the study species is in equilibrium with environmental variables (Wiens and Graham 2005). Some evidence, however, has suggested that differences in ecological niche space occurs between populations of a species (Broennimann et al. 2007, Pearman et al. 2008, Nakazawa et al. 2010) that correspond to non-equivalent spatial predictions. In response, the methodological robustness of transferring models has been examined to better understand whether evidence for niche shift results from shortcomings in the modeling process. Out of such examinations came a number of studies considering the background, or spatial extent, used for ENM (Broennimann and Guisan 2008, Anderson

and Raza 2010, Barve et al. 2011, Saupe et al. 2012) and the effect of species characteristics on transferability (Randin et al. 2006), while others demonstrated the effect that different variable sets have on model output (Peterson and Nakazawa 2008, Rödder et al. 2009).

Selection of biologically relevant variables for ENM has depended on in depth understanding of the biology and ecology of the organism under study as well as the availability of spatially explicit ecologically variables. Environmental datasets of different dimensions, sources and variable composition appear to exert a more profound influence on transferred models than on native projections (Peterson and Nakazawa 2008, Rödder et al. 2009). Despite the critical importance of these data in the modeling process there are few suggested methods with which to select environmental variables for machine learning algorithms (Araujo and Guisan 2006, Jiménez-Valverde et al. 2009) and these have not been extensively explored. For example, Jimenez-Valverde et al. (2009) suggested that the correlation structures of data sets in the native and novel landscape be evaluated for similarity using Mantel tests. Others have noted the importance of evaluating datasets for multicollinearity (Graham 2003), although methods for coping with this problem for ENM development remain unexplored. Variables used for modeling should, ideally, exert a direct effect on the species, as opposed to use of more distal variables which tend to be correlated with proximal predictors (Randin et al. 2006, Jiménez-Valverde et al. 2009). In addition, transferring models to areas with novel combinations of ecological variables force transferred models to extrapolate (Elith et al. 2010).

Alternatively, failure of ecological niche models to accurately predict known occurrences in a novel landscape may be a result of phylogenetically associated variation in ecological niche, or niche specialization among the populations as suggested in Chapter 2 and 3 of this dissertation (Pearman et al. 2008). In this scenario, transferred models will appear to fail if the phylogenetic structure of the population differs between landscapes and is not explicitly considered during the modeling process (Peterson and Holt 2003). There is evidence that phenotypic plasticity or actual niche shift occurs in invasive populations (Pearman et al. 2008, Kolbe et al. 2012) as well among lineages of species (Gonzalez) including widely distributed pathogens (Smith et al. 2000, Fisher et al. 2005, Nakazawa et al. 2010), and ENMs, along with genetic analysis and experimental work, provide tools with which to explore these phenomena. Modeling applications can test phylogeographic hypotheses, and in turn modeling results which suggest ecological divergence among populations indicate where meaningful phylogenetic branching exists (Pearman et al. 2008, Gonzalez et al. 2011) and therefore it is important to discern between modeling artifact and actual model failure.

In Chapter 3, I reported that ecological niche models of the globally successful *Bacillus anthracis* A1.a genetic sublineage performed well when developed in and projected on the native landscapes of the United States, Italy and Kazakhstan, but each model was unable to predict occurrences of the same sublineage when transferred to the landscapes of the other countries as shown in Chapter 3 . In that chapter, I proposed failure of the transferred models may be a result of niche differentiation at a different phylogenetic level than used for modeling. However, I also considered that

variable selection may have also been a factor in the failure of model transferability. The goal of this current study is to test the influence of variable selection techniques on the transferability of *B. anthracis* A1.a sublineage ENMs to determine if the performance of transferred models can be improved

Methods

Bacillus anthracis

Bacillus anthracis is a spore forming bacterium and the causative agent of anthrax in wildlife, livestock and humans worldwide. Spores may persist in soils for long periods of time, and the literature supports environmental limits on the survival of spores (Van Ness 1956, Dragon and Rennie 1995) and therefore the geographic distribution of naturally occurring disease. There is some evidence that genetic lineages of B. anthracis, characterized using single nucleotide polymorphism (SNP) (Van Ert et al. 2007a), multiple locus variable number tandem repeat (MLVA) (Keim et al. 2000, Lista et al. 2006, Van Ert et al. 2007a) analyses have distinct ecological tolerances (Smith et al. 2000).

Anthrax Occurrence Data and Environmental Data

Historical geographic information system (GIS) databases of B. anthracis isolates from the countries of United States, Kazakhstan and Italy were used. The isolates used for modeling were limited to those in defined A1.a sub-lineage as defined by the 8 marker MLVA (MLVA-8) system described in Keim et al. (2000). The U.S. isolates were derived from Kenefic et al. (2009), Blackburn et al. (2007) and new strains from recent field collections in western Montana and Texas (Blackburn, unpublished data). Isolates from Italy and Kazakhstan were described in Fasanella et al. (Fasanella et al. 2005) and Chapters 2 and 3. , respectively. All US isolates fell within the Western

North American (WNA) SNP defined group, while Italian and Kazakh isolates grouped within the SNP defined Trans-Eurasian (TEA) group (Van Ert et al. 2007a). Occurrence data were reduced to spatially unique points at an 8 km^2 resolution to accommodate the spatial uncertainty of the data (Joyner et al. 2010) and randomly divided into an 80% training set for model building and a 20% dataset for testing the native projection. Figure 1 shows the occurrence point distributions for each of the countries used in this analysis.

Grids representing 19 bioclimatic variables and altitude were downloaded from WorldClim (www.worldclim.org). In addition, two satellite-derived environmental variables describing vegetation measures were obtained from the Trypanosomiasis and Land Use in Africa (TALA) research group (Oxford, United Kingdom; Table 4-1) (Hijmans et al. 2005, Hay et al. 2006). WorldClim bioclimatic grids (Bioclim) are variables representing annual trends and seasonality as well as extremes in environmental conditions. All grids were resampled to 8 km^2 and clipped to country boundaries using ArcView 3.3 with the GARP datasets extension (Environmental Systems Research institute, ESRI, Redlands, CA). Two base sets of environmental layers were created. The first (8-variable) used eight variables based on parameters used in previous studies and Chapters of this dissertation (Blackburn 2010, Joyner 2010, Joyner et al. 2010, Mullins et al. 2011). A second environmental data set contained the complete set of 22 variables (22-variable). Altitude was eliminated from the 8-variable and 22-variable datasets to test whether inclusion of this more distal variable (Austin 2007) influenced the results (“8 variable no alt” and “22 variable no alt”).

Values of each variable were extracted from every grid cell of all three landscapes using the Extract Multiple Values to Points routine in the Spatial Analyst extension of ArcMap 10 (Environmental Systems Research Institute 2011. ArcGIS Desktop: Release 10. Redlands, CA). Correlation matrices for each set of variables were generated in R 2.14.1. Variables were considered highly correlated if the Pearson correlation coefficient (r) was equal to or greater than 0.8. Collinearity was present within both the 8-variable and 22-variable native environmental datasets and was reduced by elimination of variables in each correlated pair judged to be less relevant to the ecology of B. anthracis. Two datasets were created from the 8 variable dataset. Because annual precipitation (BIO12) was significantly correlated with both precipitation of the wettest month (BIO14) and precipitation of the driest month (BIO13), the 7-variable set was created by eliminating BIO12 and the 6-variable set eliminated BIO14 and BIO13. The 22 variable dataset was similarly reduced in the following manner. In Reduced Set 1, as few as possible variables were eliminated by retaining one variable from each correlated pair based on which variable was most ecologically relevant. Two additional reduced datasets intended to be similar in dimensionality to the 8-variable dataset were produced based on selection of potentially relevant, uncorrelated variables. Reduced Set 2 included measures of seasonality not contained in the 8 variable set, and BIO12. Reduced set 3 contained the measures of seasonality and BIO13 and BIO14.

Ranges of available values for each variable on each landscape were visualized using box plots created using XLStat (Addinsoft SARL, New York, NY), an add on software package for use with Microsoft Excel. Although all variable ranges were

different using the Kruskal Wallis one-way analysis of variance test in R 1.14 (data not shown), a background variable set was selected by choosing variables having visually similar ranges in the three landscapes (Figure 4-2). This technique was chosen to minimize the amount extrapolation required to project models onto novel areas. All datasets are summarized in Table 4-2.

Correlation structures of the datasets were compared between landscapes using a Mantel test in the vegan package of R 2.14 (Oksanen et al. 2013). The correlation structures of the 8 variable datasets were similar only for the Italy-Kazakhstan pairing (Table 4-3). All other variable sets were statistically similar on Mantel testing for each country pairing, although r values ranged from 0.53 to 0.95 (Table 4-3).

Model Development

Ecological niche modelling was carried out using the Genetic Algorithm for Rule-Set Prediction (GARP) with the best subset procedure (Stockwell and Peters 1999, Anderson et al. 2003). In GARP a two-step procedure develops rules predicting presence or absence in variable space and projects the rules onto the landscape. Because GARP is a random walk process, multiple potential models are produced and a best subset procedure selects optimal models based on user-defined thresholds of omission and commission. Training data were input into GARP with a 50% training/50% testing internal data partition. For all experiments, up to 200 models were created with a maximum of 1,000 iterations and a convergence limit of 0.01. The ten best subset models under a 10% hard omission threshold and a 50% commission threshold were retained. The output rasters were imported into ArcGIS and summated to generate a single cumulative map of model agreement for predicted B. anthracis presence ranging from 0 (all models predict absence) to 10 (all models predict

presence). Models were trained in each native landscape using each of the environmental datasets. GARP rule sets from each experiment were projected onto the other two landscapes. Kazakh models were trained with a subset of southern A1.a isolates as described in Chapter 2 and 3.

Model Evaluation

Predictive performance of the best model subset was evaluated with an area under the curve (AUC) in a receiver operating characteristic (ROC) analysis using withheld independent test data for native models and all data for projected models. Values of AUC approaching 1 indicate a well-performing model and an AUC equal to 0.5 indicates the model performs no better than random. We used the following classification scheme to describe model performance: an AUC of <0.7 was considered poor, values between 0.7 and 0.8 were fair, values between 0.8 and 0.9 were considered good, and values over 0.9 were considered very good (Swets 1988). An area under the curve analysis was chosen, despite its described limitations (Lobo et al. 2008, Peterson et al. 2008), because all experiments were performed with the same modeling platform in the same geographic extents. Therefore, AUCs in this study are more comparable than when models made with different algorithms or in different extents are compared. To address the valid concern that AUC equally weights specificity and sensitivity, (Lobo et al. 2008, Peterson et al. 2008) AUCs were interpreted in conjunction with measures of omission and commission calculated using the summated ten best models. Total commission is the percent of pixels which are predicted as presence by areas of ten model agreement. Average commission is the average area predicted as presence by all subset models. The greater the difference between the two measures of commission the greater the spatial heterogeneity among

the ten best subset models (McNyset 2005, Blackburn et al. 2007). Total omission was calculated as the total number of test points falling into areas predicted as absence by all ten models. Summed area omission (SAO) was calculated as the omission error of areas of ten model agreement.

Results

All native models performed significantly better than random based on AUC and the majority were in the fair to very good range (Table 4-4). Overall, models trained in south Kazakhstan performed better than those trained in all Kazakhstan and are shown here. Transferred projections performed poorly, no better than random or significantly more dispersed than random with the exception of the background set for which the Kazakh model projected to Italy and the US resulted in AUCs in the fair range. In general, the Kazakh models overpredicted Italy and the US, and Italian models underpredicted both other countries. In all figures, the predicted distribution of *Bacillus anthracis* in shown in native landscapes in the left hand column and model predictions on novel landscapes are shown in middle and right columns. The color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

The 8 variable dataset first shown in Chapter 3 is shown in Figure 4-1. In comparison, the highly dimensional 22 variable dataset produced narrow native predictions in the United States and Italy and those models transferred to the other two landscapes predicted presence in very small areas with low model agreement (Figure 4-3). In Kazakhstan, this dataset had a different effect, predicting broad areas of high likelihood of pathogen presence in the south of the country, and broadly predicted areas in the north with low model agreement. The transferred Kazakh 22 variable dataset

model predicted very large extents of Italy and the US. Models produced using the 22 variable no altitude and reduced set 1 dataset had similar spatial predictions (Appendix B).

Removal of altitude from the 8 variable set improved the projection of the Kazakh model to the US and extended the native US model predictions farther west; other projections were similar to the 8 variable set (Appendix B). The 6 variable dataset had more restricted native predictions than the 7 variable dataset in all countries and, when transferred, the 6 variable dataset resulted in narrower geographic predictions in the novel landscapes, although both US projections predict a small area in northeast Italy not predicted by the native model (7 variable model shown in Figure 4-4, Appendix B for 6 variable model). The native Italian 7 variable dataset resulted in one best subset models predicting endemic areas of southern Kazakhstan and the central US, whereas the Italian 6 variable dataset failed to predict any of the Kazakh landscape and only very small areas of the US. A similar trend is seen in the reduced set 2 and 3 datasets, which mimic the 6 variable set and 7 variable set with the addition of measures of seasonality. The addition of measures of seasonality to the 6 variable set (reduced set 2; Appendix B) increased the extent of the spatial predictions in the native US and Kazakh models, in the Kazakh projection to the US, and in the US projection to Kazakhstan. A similar pattern resulted from the addition of measures of seasonality to the 7 variable dataset (reduced set 3; Figure 4-5). The Italian reduced set 3 model transferred to the US shifted spatial predictions of presence to the east.

The background dataset native projections had broad predictions in the US, markedly homogenous predictions in Italy, and resembled the 8 variable projection in

Kazkahstan (Figure 4-6). When projected, the US model predicted areas of Italy in a similar spatial pattern to the native model, although model agreement on the mainland of Italy was low. The US projection to Kazakhstan was also spatially similar to the native Kazakh model, but the projection had higher model agreement in north Kazakhstan. The Italian background model again underpredicted the US and Kazakhstan. The Kazakh model, however, predicted Italy with fair accuracy and considerable homogeneity of the best subset models. The Kazakh to US projection also had fair performance as measured by AUC, and although the best subset models tended to predict the north central and northwest portions of the country, the overall predictions followed the native model.

Discussion

Previous research has suggested that the transferability of ecological niche models (ENM) is affected by variable selection (Peterson and Nakazawa 2008, Rödder et al. 2009), spatial extent or background of the landscape available for modeling (Elith et al. 2010, Barve et al. 2011), and the species being modeled (Randin et al. 2006, Guisan et al. 2007). Although approaches to these methodological uncertainties have been proposed, none have been demonstrated to be superior across all modeling applications. This study employed variable selection techniques to evaluate the influence of variable sets on transferred ecological niche models of *Bacillus anthracis* created in GARP with the goal of optimizing predictions of anthrax occurrences in novel landscapes. We found that despite attention to creating structurally similar and ecologically relevant datasets, models were unable to robustly predict occurrences when transferred to novel landscapes. Native models reasonably predicted the presence of occurrences in our experiments, with 90% having AUC values over 0.7,

although the spatial distribution of predicted presence varied. Transferred models, however, tended to grossly over-predict or under-predict presence on the novel landscapes.

Not surprisingly, the variable sets with highest numbers of variables, as well as a high degree of multi-collinearity, produced restricted geographic predictions in several models, including the native ranges of Italy and the US and in the projections of the Italian and US models. The US models, however, did predict small areas of Kazakhstan also predicted by the native Kazakh model, indicating a climatic signature common to those locations. The Kazakh 22 variable model, in contrast, overpredicted Italy and the US with the exception of the Dakotas area in the US that was consistently predicted by native US models. Multicollinearity within ecological datasets negatively impacts regression models in ecological studies (Graham 2003) and will affect GARP models because of the inclusion of logistic rules. I therefore generated datasets which eliminated multicollinearity by choosing, based on understanding of the biology of *B. anthracis*, the most relevant variable in each correlated pair. The resulting models had subtle changes to native and transferred projections as compared to ones generated using the original 8 and 22 variable sets. There were consistencies in the reduced models based on the variables included from collinear pairs. For example, projections of US and Italian models created with the reduced 3 and 7 variable sets, both containing BIO13 and BIO14 instead of BIO12, were similar, as were those created with two variable sets containing BIO12. In addition to multicollinearity causing difficulties in model creation, differences in the nature of correlation structures of the data between landscapes may negatively influence the performance of transferred models (Jiménez-

Valverde et al. 2009). In these experiments, only the 8 variable environmental dataset had dissimilar correlation structures for country pairings based on Mantel testing, and subsequent data sets which eliminated multicollinearity were similar between all country pairings. Increasing the similarity of correlations structures based on *a priori* understanding of the organism did not improve the transferability of models in this set of experiments, but selection of variables did alter the spatial predictions made by transferred models. In addition to problems caused by multicollinearity in highly dimensional datasets, inclusion of distal variables will tend to produce restricted spatial predictions. Although such ecological factors do not exert a strong influence on the species, and in fact the species can occupy a large range of the ecological axis represented by a distal variable, the available gradient in a spatial extent may truncate the response of the model. Transferability is thus limited because models fail to capture the true ecological requirement of the species (Randin et al. 2006) and instead reflect incidental associations. Redundant data in large datasets is of particular concern in machine learning methods because the algorithm can focus on a random portion of such data; the problem is likely exacerbated by small occurrence datasets. Under these conditions models may fit the data in the training extent well, but are unable to predict new occurrences in a novel context (Dietterich 1995). Future work should address the utility of statistical techniques to cope with multicollinearity of ecological data used for machine learning or genetic algorithms and the influence of these techniques on transferability to novel areas and time periods.

The question of spatial extent used for background in presence only modeling has been more extensively examined. Proposed techniques for accounting for

background conditions include incorporation of both native and invaded ranges to equilibrate the ecological background (Broennimann and Guisan 2008), creating an imposed spatial or ecological limit (Anderson and Raza 2010), or selection of variables with similar gradients across landscapes (Randin et al. 2006). It has been suggested that the spatial extent should be confined to an relatively narrow area in which the species has been found or is able to access, because inclusion of inaccessible areas, as may occur when geo-political boundaries are used, or inclusion of all accessible areas under the assumption of environmental equilibrium, will result in overfitting (Elith et al. 2010, Barve et al. 2011), although the importance of extent of the training area has been recently questioned (Saupe et al. 2012). *Bacillus anthracis* represents a challenge to the concept of accessible space. The pathogen is hypothesized to have evolved in Africa and the Middle East and subsequently transported worldwide. Recent outbreaks, however, remind us that the movement of *B. anthracis* is not simply a historical problem (Fasanella et al. 2012). The organism continues to be translocated from one invaded range to another in modern times. The implication of this process is that *B. anthracis* is unlikely to be at equilibrium with the environment, as it may not have had an opportunity to be introduced to an entire spatial extent or to all ecological conditions suitable for establishment (Holt et al. 2005). Selecting historical and future accessible areas, and, in the absence of genetic trace-back, the area from which a population might have been sourced is problematic and potentially arbitrary. In this unique context, political or administrative boundaries can be appropriate for delimiting background geography for models of anthrax persistence because of the difficulties in

defining accessible space and, additionally, because the reporting and control of anthrax takes place within such boundaries.

In general, model transferability was asymmetrical; for example, the US background model reasonably predicted anthrax occurrences in Italy, but the Italian background model failed to predict occurrences in the US. Different available ranges of predictor variables among the landscapes are a potential cause of both differences in transferability between variable sets and of asymmetry of reciprocal predictions (Randin et al. 2006). Dissimilar climatic spaces require the model to extrapolate to new combinations of environmental conditions (Fitzpatrick and Hargrove 2009, Elith et al. 2010). I selected variables having similar gradients across all three landscapes as a method of controlling for background climatic space and reducing the need for the model to extrapolate into novel ecological contexts. The resulting US and Kazakh background models performed reasonably well when transferred, suggesting that equivalence of the available ecological space among landscapes is an important factor in model transferability, while the Italian native background model still underpredicted the US and Kazakhstan, although this was the only Italian model to predict any of the Kazakh landscape with a high level of model agreement. Interestingly, the US and Kazakh background models, when transferred to Italy, predicted areas in northern Italy where the SNP defined A.Br 005/006 sub-group, previously only seen in Africa, and isolates of the B lineage have been genotyped (Garofolo et al. 2011).

That models consistently fail to reciprocally predict occurrences with acceptable measures of omission and commission across all environmental datasets lends support to the three regional groups being ecologically divergent at the level of genetic analysis

used (Mullins et al. 2013). Similar to questions of spatial scale and resolution, niche divergence can be evident, or appear absent, at different genetic scales (Pearman et al. 2008), and here it appears that a level of ecological divergence occurred at different genetic scale than used for modeling. The A1.a group appears able to establish and persist in a relatively wide variety of ecological conditions, and at regional or local scales populations may further develop important genetic and ecological differences (Holt et al. 2005). The A1.a sub-lineage isolates in the US, which fall into the WNA SNP group found almost exclusively in North America, exhibit considerable genetic divergence from the trans-Eurasian (TEA) SNP defined strains identified in Kazakhstan and Italy. The evolutionary divergence between Eurasian strains in Kazakhstan and Italy is not as well defined, but is likely more significant than suggested by can-SNP or MLVA-8 typing (Sytnik, unpublished data). Better characterization of the genetic population structure of *B. anthracis* at multiple genetic and spatial scales will improve models predicting areas at risk for pathogen persistence. In addition, the epidemiologic implications of genetic diversity in *B. anthracis* are not currently understood and genetic analysis could provide insight into host, spatial and temporal specificity of outbreaks.

There are additional considerations, other than modeling artifact or population genetics, which could influence the transferability of these models which should be noted. ENMs, by definition, are based on climatic variables and reflect the abiotic hypervolume in which a species may maintain. A species' distribution is also constrained by historical limits to dispersal and survival as well as biotic interactions, which if different between landscapes will affect the transferability of ENMs (Randin et al. 2006). Human mediated effects on the historical and current distribution of anthrax

create spatial and temporal biases in occurrence data and present important limitations to this study and similar studies of pathogen ecology. For example, outbreaks are more likely to be detected and recorded in areas of higher surveillance or awareness, including locations having a long history of the disease, and where veterinary and public health resources are available. Similarly, the distribution of prevention and control efforts, such as vaccination distribution or carcass disposal, are spatially heterogeneous. Spatial differences in surveillance and control may limit the utility of transferred models of pathogens by biasing models to areas of less historical control.

Biotic variables typically operate at more local scales than abiotic predictors (Soberón 2007) and are challenging to incorporate into ENMs, although increasingly the need for this is being recognized. The resolution of data used for creating ecological niche models is thus an additional and related consideration (Guisan and Thuiller 2005). The coarse resolution of climatic variables may miss micro-scale variations that are locally important to *B. anthracis*, such that ENMs of this species are effectively created using distal variables. In general, it is suggested that sessile organisms or those only locally mobile may be predicted by finer resolution models with more precise matching of occurrences to predictor values (Guisan and Thuiller 2005). Also related to the scale of analysis, anthrax outbreak records from the three countries span periods of up to 50 years. Temporal aggregation of occurrence data and climatic data also means that associations between outbreaks and temporal climatic extremes will be lost. It is also possible that some recorded outbreaks represent food related or industrial outbreaks or which, in effect, mimic sink populations, which may reduce transferability of models (Blackburn et al. 2007), although the GARP method appears less sensitive to

outlier data than other ENM algorithms. These results underscore previous suggestions that *B. anthracis* ecology is best understood at local scales (Blackburn 2006). Finally, the small sample size of genotyped isolates from each landscape is a limitation to this study. Smaller sample sizes can result in overfit and unstable models and may account for some of the variation in predictions. Genetic analysis of outbreaks can assist in determining whether an outbreak resulted from an anthropogenic introduction or from a local environmental reservoir, and increased use of molecular trace-back in outbreaks will allow ENMs for anthrax and similar pathogens to be refined by discriminating between outbreaks arising from endemic pathogen populations and those that represent sink populations. Such efforts can and should also increase the sample size of genetically similar isolates for geographic analysis and ecological niche modeling.

Ecological niche models of *B. anthracis* predict the geographic space in which the pathogen is likely to persist and therefore where anthrax is likely to occur (Blackburn 2010). These predictions can be used to focus surveillance efforts and vaccine distribution in areas of high likelihood for pathogen persistence (Blackburn et al. 2007, Alexander et al. 2012) or discover previously unknown or unreported foci of persistence. However, transferred models have a number of potential methodological problems, of which predictor selection is substantial. Here, neither resolving collinearity within environmental data sets nor increasing the similarity of the correlation structures of the variables drastically improved performance of the transferred models, as measured by AUC, omission and commission in combination, over the original set of variables selected as being most ecologically relevant to the species. The variable set chosen based on similarity of ranges performed the best of all experiments. While additional

work is certainly needed to provide guidelines for selecting environmental predictors for experiments incorporating transferred models for anthrax, careful consideration of phylogenetic differences and phylogenetic scale should be undertaken also. For *B. anthracis*, I conclude that ENMs and other efforts to model anthrax ecology be undertaken at local scales and that increased efforts to understand population genetics of this important pathogen will improve our understanding of anthrax ecology.

Table 4-1. Environmental variables used to develop ecological niche models

Environmental Variable (unit)	Name	Source
Elevation (m)	Altitude	WorldClim*
Annual Mean Temperature (^{0}C)	BIO1	WorldClim
Mean Diurnal Range	BIO2	WorldClim
Isothermality (BIO2/BIO7)*100	BIO3	WorldClim
Temperature Seasonality	BIO4	WorldClim
Maximum Temperature Of Warmest	BIO5	WorldClim
Minimum Temperature Of Warmest	BIO6	WorldClim
Annual Temperature Range (^{0}C)	BIO7	WorldClim
Mean Temperature Of Wettest Quarter	BIO8	WorldClim
Mean Temperature Of Driest Quarter	BIO9	WorldClim
Mean Temperature Of Warmest	BIO10	WorldClim
Mean Temperature Of Coldest Quarter	BIO11	WorldClim
Annual Precipitation (Mm)	BIO12	WorldClim
Precipitation Of Wettest Month (Mm)	BIO13	WorldClim
Precipitation Of Driest Month (Mm)	BIO14	WorldClim
Precipitation Seasonality	BIO15	WorldClim
Precipitation Of Wettest Quarter	BIO16	WorldClim
Precipitation Of Driest Quarter	BIO17	WorldClim
Precipitation Of Warmest Quarter	BIO18	WorldClim
Precipitation Of Coldest Quarter	BIO19	WorldClim
NDVI† Amplitude (no units)	wd1014a1	TALA†
Mean NDVI (no units)	wd1014a0	TALA

*(www.worlclimorg) [46]

† Normalized Difference Vegetation Index; Trypanosomiasis and Land Use in Africa (TALA) research group (Hay et al. 2006)

Table 4-2. Variable used in each experiment.

Variable	8 variable	22 variable	8 Variable No Altitude	22 Variable No Altitude	6 Variable	7 Variable	Reduced 1	Reduced 2	Reduced 3	Background
Altitude	X	X			X	X	X	X	X	X
Annual Mean Temperature	X	X	X	X	X	X	X	X	X	
Mean Diurnal Range		X		X			X			
Isothermality		X		X						
Temperature Seasonality		X		X			X	X	X	
Max Temp Warmest Month		X		X			X			X
Min Temp Coldest Month		X		X						
Annual Temperature Range	X	X	X	X	X	X	X	X		
Mean Temp Wettest Quarter		X		X			X			X
Mean Temp Driest Quarter		X		X			X			X
Mean Temp Warmest Quarter		X		X						X

Table 4-2. Continued.

Variable	8 variable	22 variable	8 Variable No Altitude	22 Variable No Altitude	6 Variable	7 Variable	Reduced 1	Reduced 2	Reduced 3	Background
Mean Temp of Coldest Quarter		X		X						
Annual Precipitation	X	X	X	X	X		X	X		
Precipitation Wettest Month	X	X	X	X		X			X	
Precipitation of Direst Month	X	X	X	X		X			X	
Precipitation Seasonality		X		X			X	X	X	X
Precipitation Wettest Quarter		X		X						
Precipitation of Driest Quarter		X		X						
Precipitation Warmest Quarter		X		X			X			
Precipitation Coldest Quarter		X		X						
Mean NDVI	X	X	X	X	X		X	X	X	
NDVI Amplitude	X	X	X	X	X		X	X	X	X

Table 4-3. Results of Mantel testing for similarity of correlation structures across landscapes

Variable Set	US and Italy		Italy and Kazakhstan		US and Kazakhstan	
	r	p	r	p	*r*	p
8 variable	0.1728	0.195	0.7597	0.011	0.3293	0.108
22 variable	0.6587	0.001	0.8767	0.001	0.6636	0.001
8 variable no alt	0.9547	0.001	0.9337	0.001	0.9523	0.001
22 variable no alt	0.6624	0.001	0.8649	0.001	0.6699	0.001
6 variable	0.6921	0.005	0.9327	0.003	0.7088	0.006
7 variable	0.7307	0.005	0.9305	0.001	0.7607	0.002
Reduced 1	0.571	0.001	0.8177	0.001	0.5384	0.001
Reduced 2	0.718	0.002	0.8559	0.001	0.6451	0.005
Reduced 3	0.9233	0.001	0.8732	0.001	0.8251	0.001
Background	0.7679	0.002	0.7745	0.002	0.7927	0.004

Table 4-4. Accuracy metrics for all modeling experiments

Experiment	AUC	SE	Z	Total Omission	Average Omission	Total Commission	Ave Commission	SAO
8 Variable								
US								
Native	0.934	0.0498	5.8392	0	6.7	8.38	22.88	16.67
Italy	0.516	0.0544	7.7238	62.1	71	0	7.71	100
Kazakhstan	0.4767	0.0467	5.9889	13.2	51	0.84	43.54	81.58
Italy								
Native	0.8372	0.0942	3.861	0	1.4	27.86	50.21	0
Kazakhstan	0.5641	0.0483	97.9111	86.8	74.8	0	0.05	N/A
US	0.4346	0.0365	16.4161	96.4	82.2	0	3.17	100
Kazakhstan								
Native	0.9013	0.0918	3.1318	0	0	19.77	46.3	0
Kazakhstan	0.7418	0.0792	4.5203	0	19.2	13.17	54.28	66.67
Italy	0.4544	0.0538	15.926	0	10	9.04	90.9	0
US	0.4845	0.0382	7.7058	0	17.7	28.67	83.35	57.14
22 Variable								
US								
Native	0.9583	0.0404	7.265054	0	4.2	6.29	13.92	16.67
Italy	0.4777	0.053	24.27139	100	65.5	0	0.44	N/A
Kazakhstan	0.6477	0.0486	40.11205	68.4	74.4	0	0.47	N/A
Italy								
Native	0.8033	0.1003	3.567291	0	17.1	21.12	39.51	28.57
Kazakhstan	0.5	0.0469	-	100	73.7	0	0	N/A
US	0.4807	0.0381	36.71937	100	82.1	0	0.39	N/A
Kazakhstan								
Native	0.8779	0.1001	3.151893	0	0	24.46	49.31	0
Kazakhstan	0.692	0.0819	4.4846	7.7	41.8	8.71	24.95	53.84
Italy	0.5	0.0539	undef	0	0	100	100	0
US	0.1038	0.0125	3.0002	0	49.3	84.08	94.47	87.50

Table 4-4. Continued

Experiment	AUC	SE	Z	Total Omission	Average Omission	Total Commission	Ave Commission	SAO
US								
Native	0.8736	0.0655	5.314094	0	5	18	42.27	8.33
Italy	0.51	0.0542	6.913118	55.2	64.4	0.09	17.84	100
Kazakhstan	0.3692	0.0407	9.159517	0	14.2	81.2	95.49	42.11
Italy								
Native	0.8122	0.0989	3.593677	0	5.7	24.51	49.2	28.57
Kazakhstan	0.6595	0.0485	26.75657	63.2	79.7	0	0.5	N/A
US	0.535	0.0394	11.9949	60.7	84.2	0	4.36	N/A
Kazakhstan								
Native	0.9176	0.0851	3.121679	0	0	16.51	50.77	0
Kazakhstan	0.724	0.0804	4.4016	0	20.8	8.88	56.81	53.85
Italy	0.539	0.055	11.33669	0	1.4	78.65	97.73	13.79
US	0.6889	0.0395	10.43158	0	8.8	26.96	80.2	41.07
Reduced 3								
US								
Native	0.9075	0.0577	5.506267	0	2.5	14.54	35.28	16.67
Italy	0.5691	0.0547	7.242508	43.3	68.6	0.27	13.94	N/A
Kazakhstan	0.4957	0.0467	21.78076	0	0.8	93.01	99.21	N/A
Italy								
Native	0.8024	0.1005	3.556023	0	8.6	25.72	48.45	28.57
Kazakhstan	0.7044	0.0476	21.86479	52.6	79.8	0	0.8	N/A
US	0.7053	0.0392	12.75965	16.1	87.7	0.04	6.02	N/A
Kazakhstan								
Native	0.8318	0.1127	3.175586	0	0	33.67	65.3	0
Kazakhstan	0.6453	0.0831	3.9738	0	26.9	13.59	59.16	53.85
Italy	0.5	0.0539	undef	0	0	100	100	0
US	0.4427	0.0368	39.59178	0	1.4	97.39	99.23	14.29

Table 4-4. Continued

Experiment	AUC	SE	Z	Total Omission	Average Omission	Total Commission	Ave Commission	SAO
Background								
US								
Native	0.8907	0.0286	12.66	0	4.1	17.38	37.35	89.93
Italy	0.6264	0.0559	6.5489	17.2	66.1	0	20.19	N/A
Kazakhstan	0.5633	0.0483	6.2908	13.2	47.9	15.96	41.18	100
Italy								
Native	0.7662	0.1054	3.9328	40.51	50.47	0	1.4	14.29
Kazakhstan	0.5874	0.0493	18.7164	73	69.5	1.53	4.39	100
US	0.4525	0.0371	22.6264	100	82.1	0.27	2.23	91.89
Kazakhstan								
Native	0.8702	0.0341	2.9665	0	2.0	20.6	52.53	20.00
Kazakhstan	0.7445	0.1077	3.0132	14.3	6.7	7.18	43.33	28.57
Italy	0.7492	0.0528	7.6515	0	8.6	23.49	63.38	51.72
US	0.7845	0.0364	10.5564	3.6	23.6	5.9	42.94	96.43

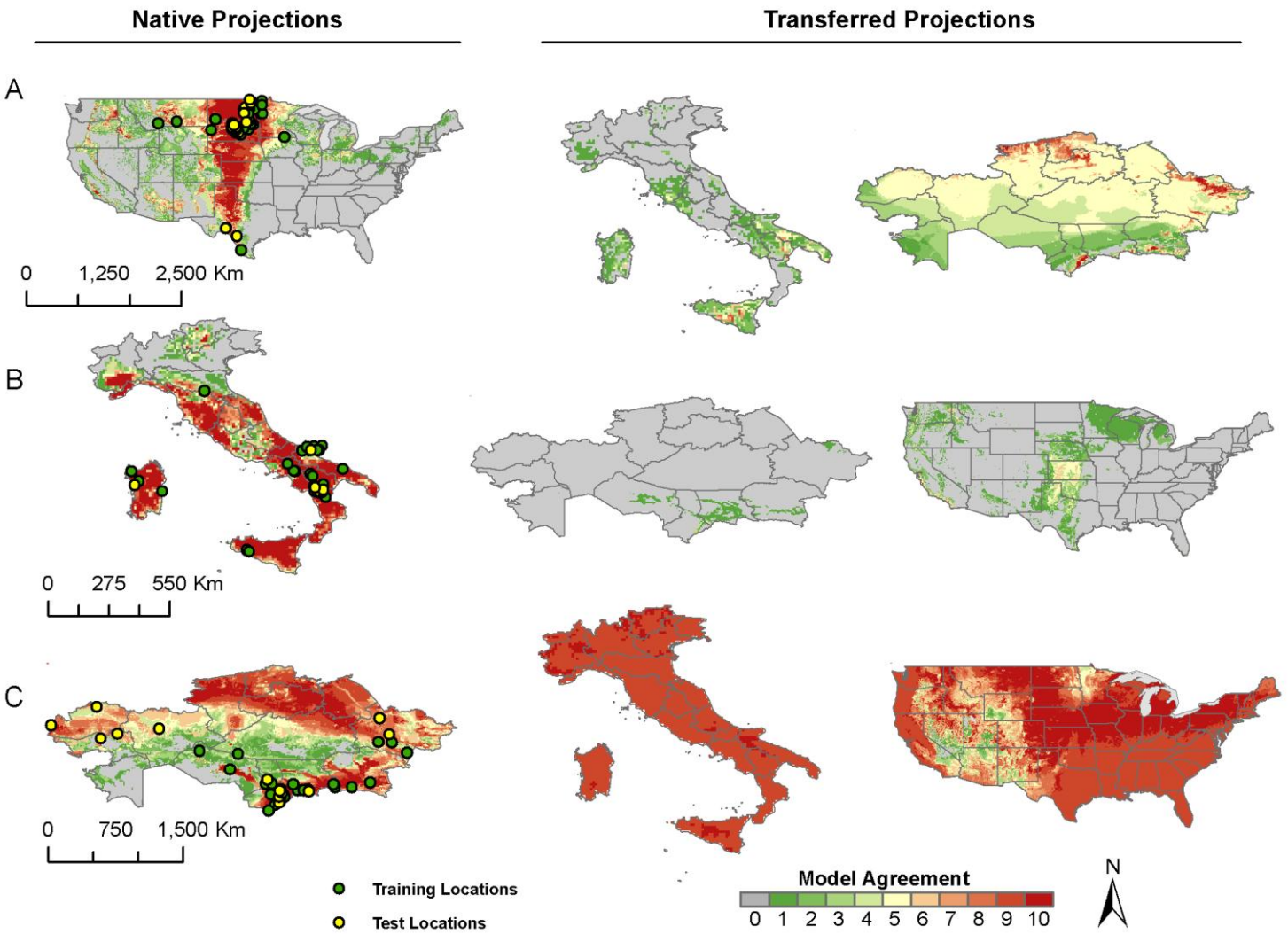

Figure 4-1. Geographic distribution of the training and testing points and results of the 8 variable model. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

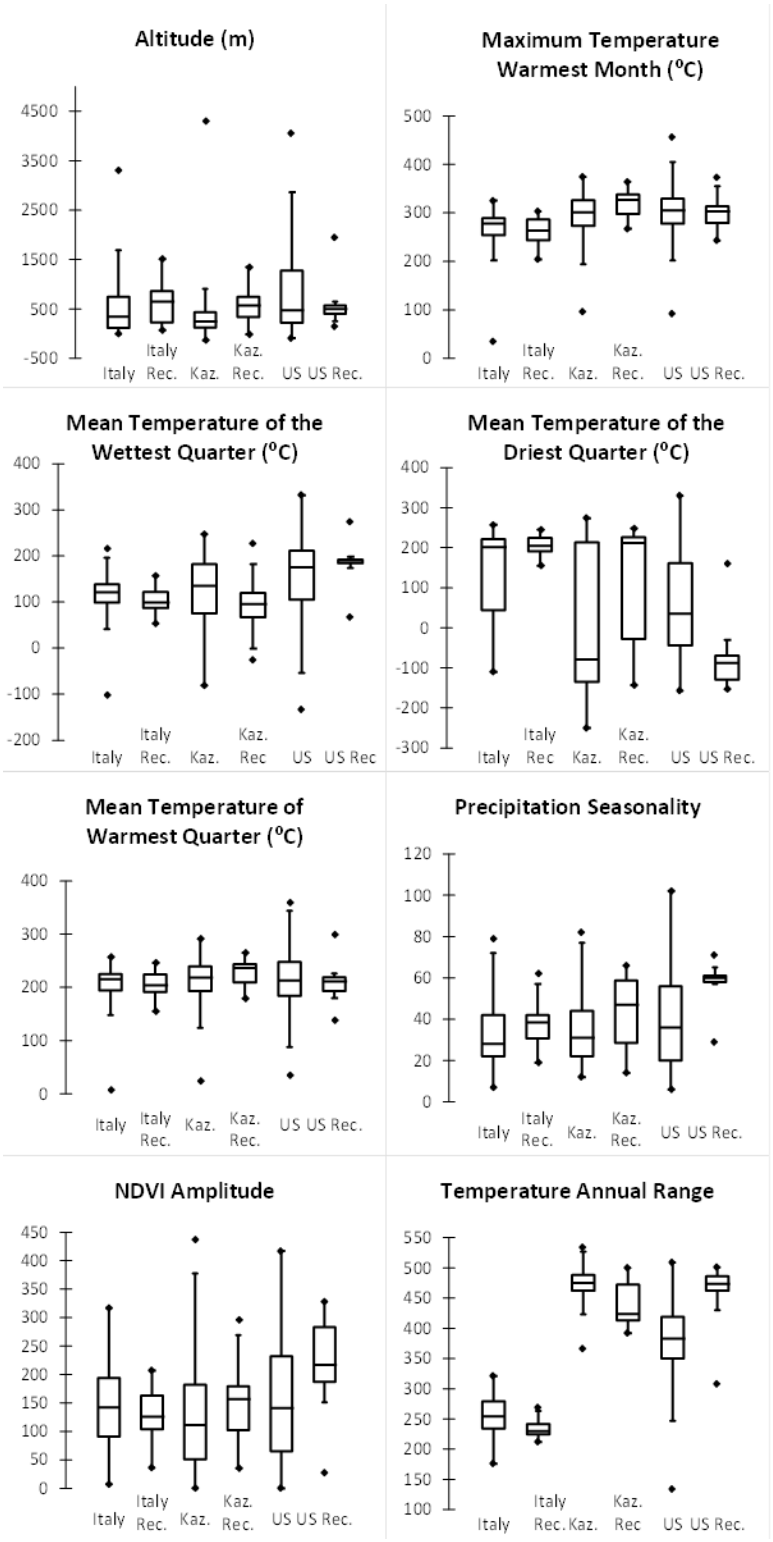

Figure 4-2. Boxplots of environmental variables illustrating comparative ranges of values. Shown are ranges of variables extract from every pixel on the landscape and from each cell having an anthrax record used for modeling (denoted by “rec.”). Variables included in the background dataset are shown; boxplots of annual mean temperature are shown to illustrate variable values that are non-overlapping across landscapes.

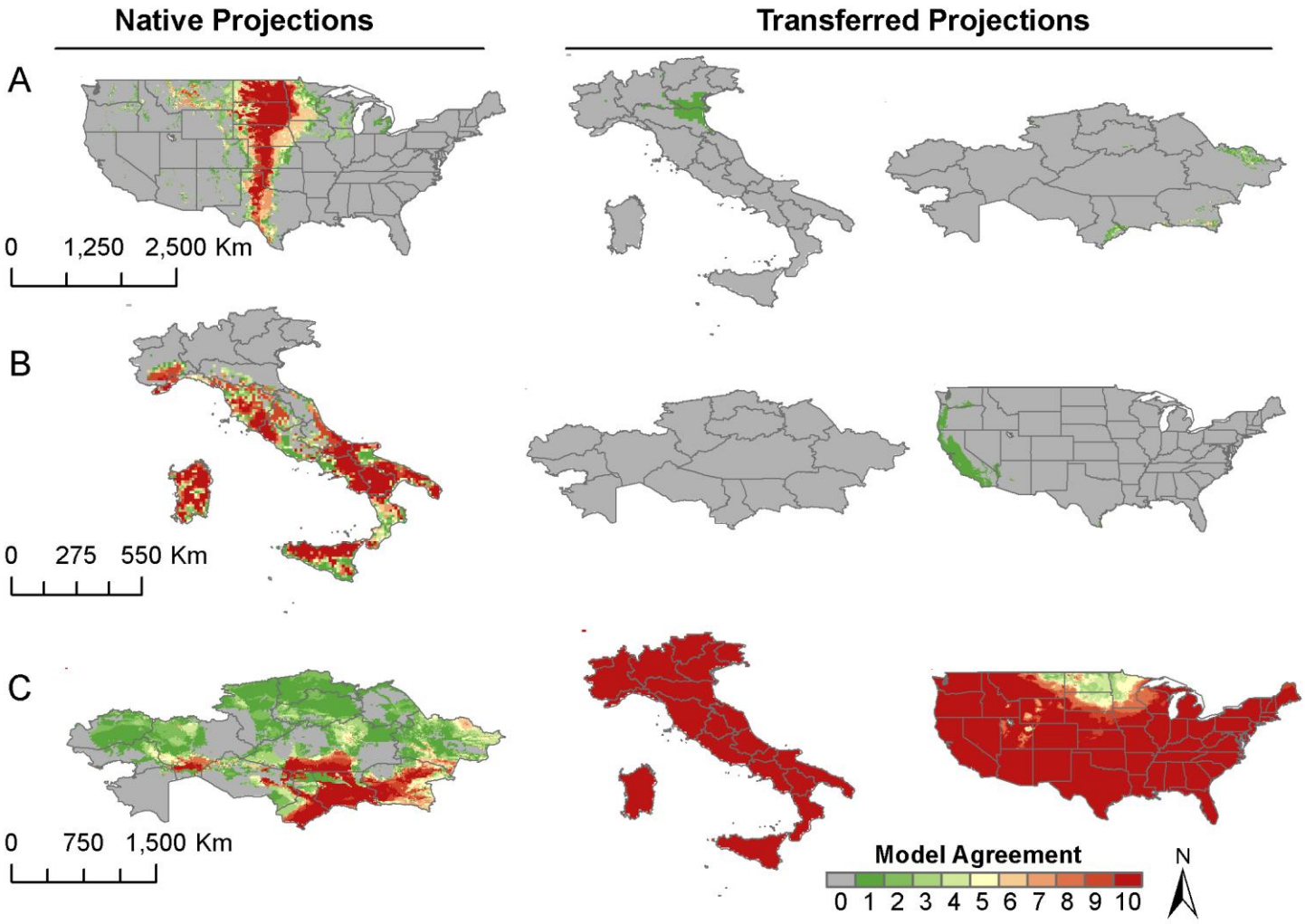

Figure 4-3. Results of the 22 variable experiment. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

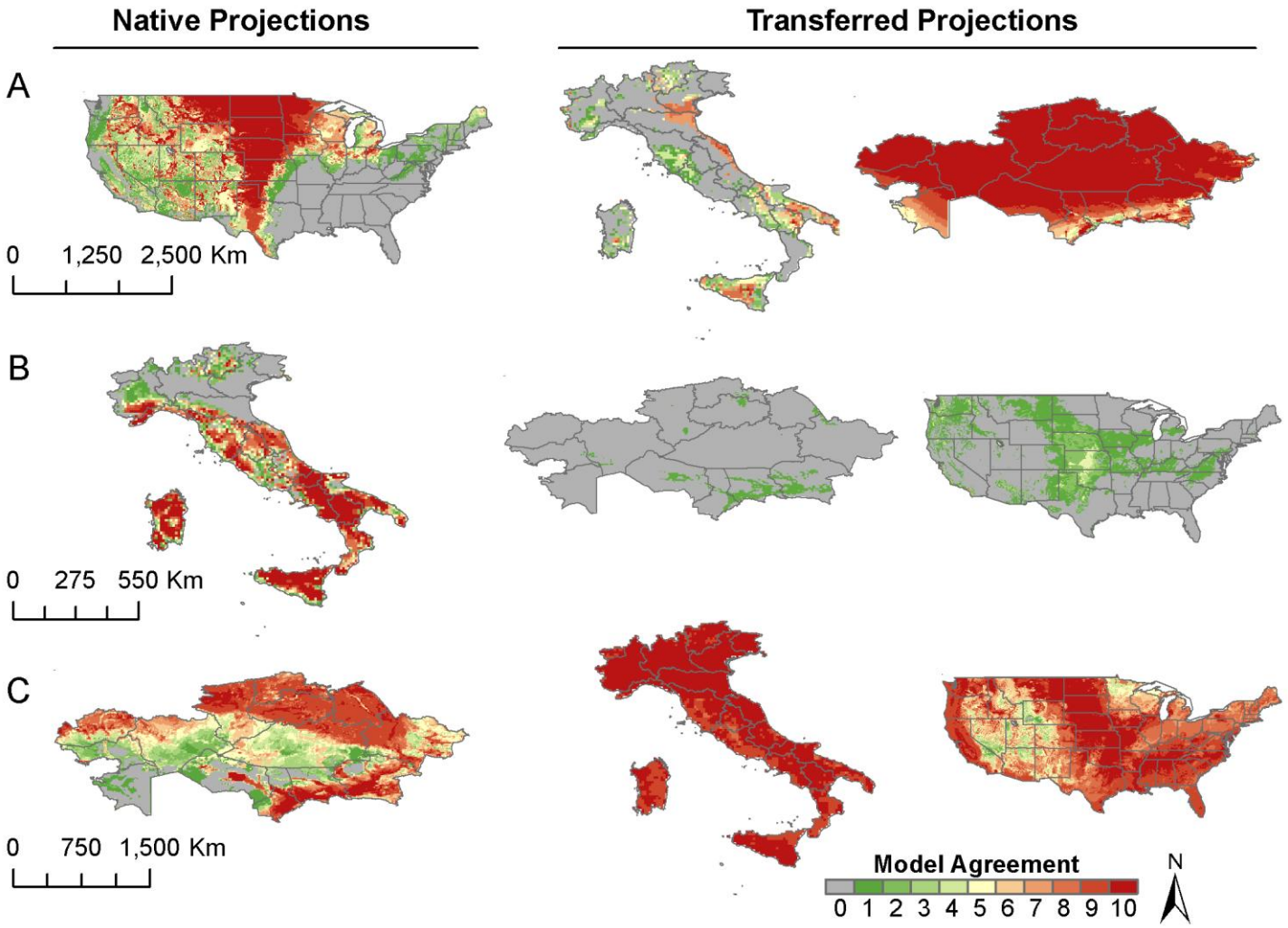

Figure 4-4. Results of the 7 variable experiment. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

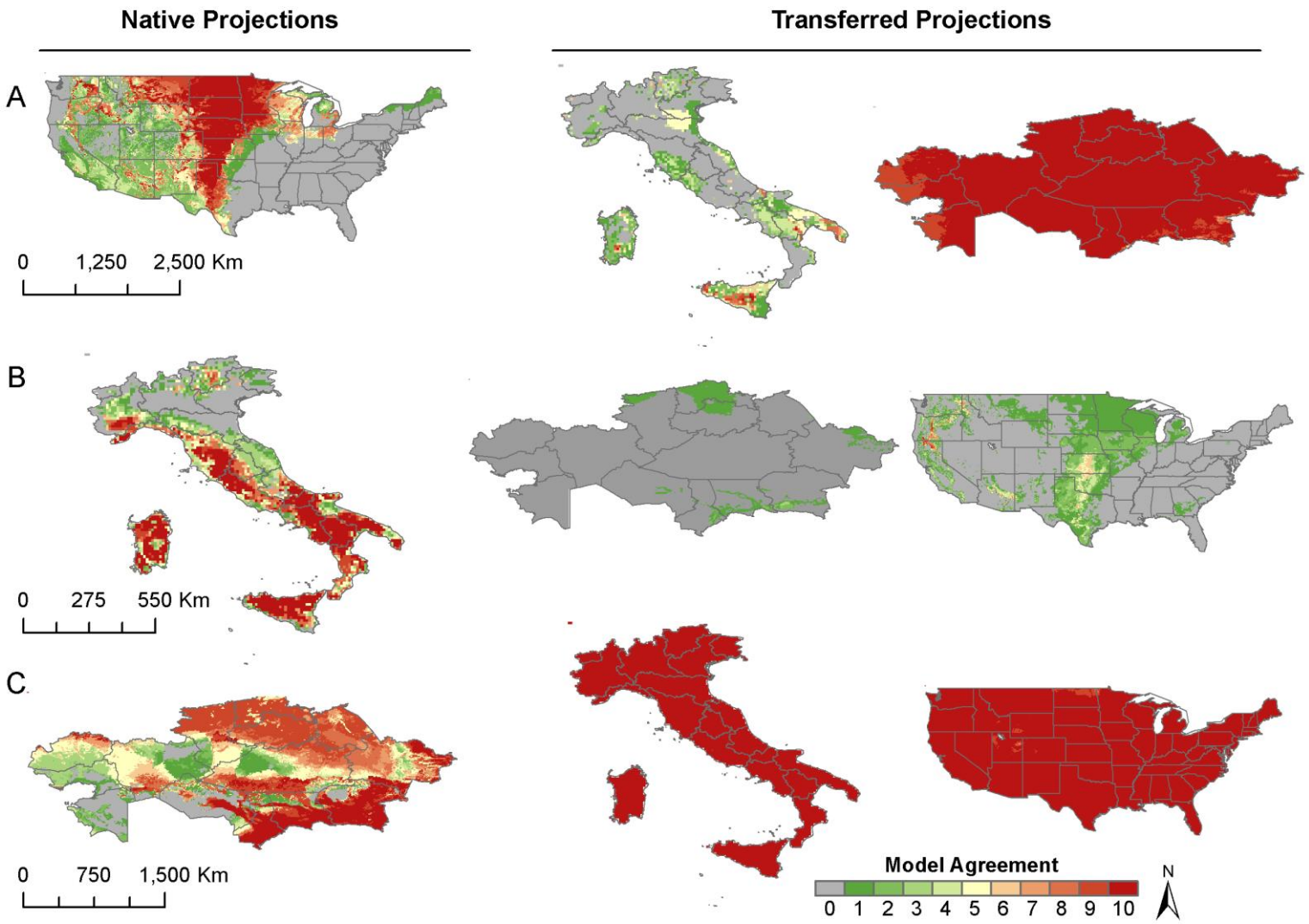

Figure 4-5. Results of the reduced 3 experiment. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

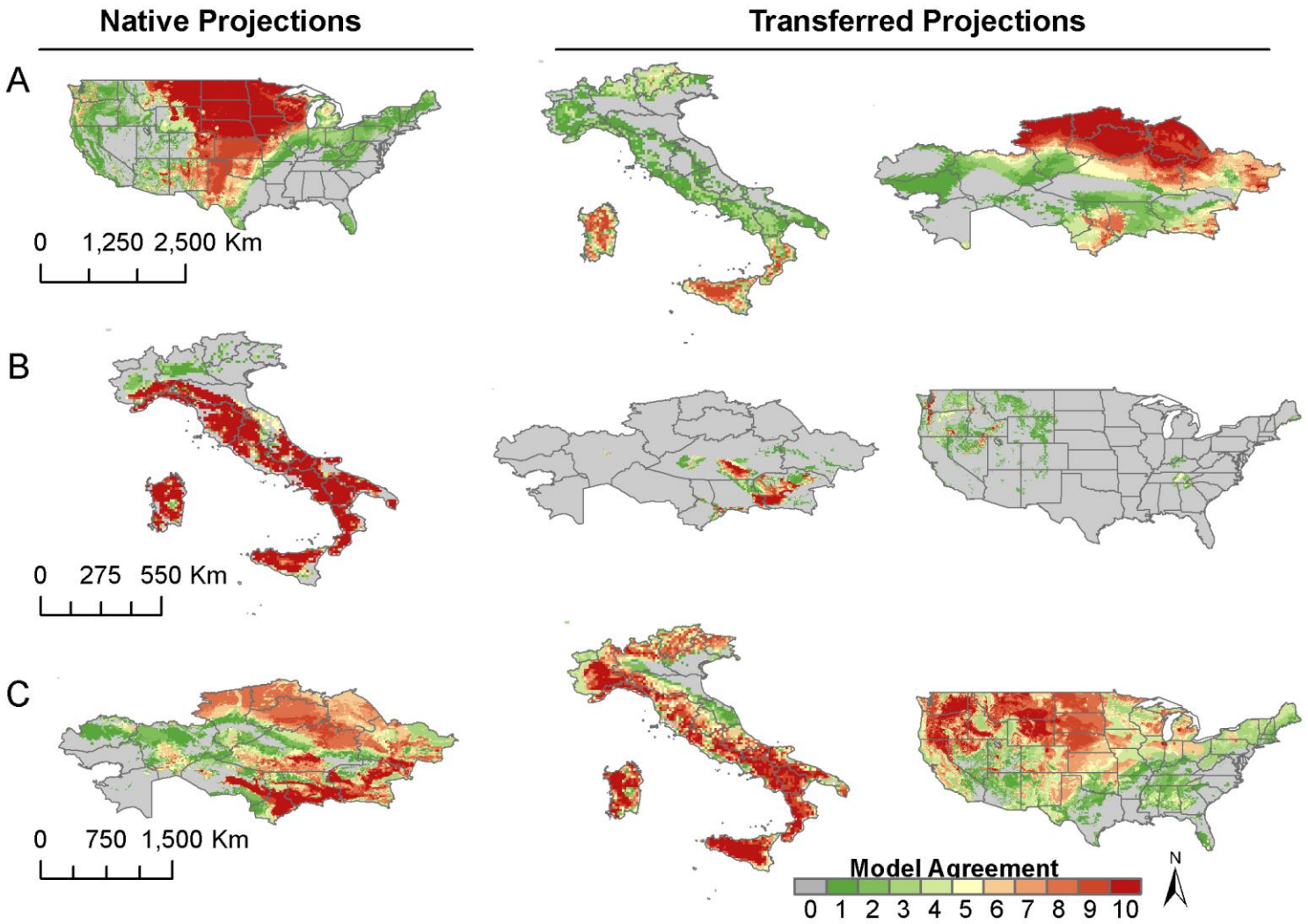

Figure 4-6. Results of the background experiment. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

CHAPTER 5
HIGH RESOLUTION GENOTYPING OF *BACILLUS ANTHRACIS* AND SPATIO-TEMPORAL PATTERNS OF AN ANTHRAX EPIZOOTIC IN WHITE-TAILED DEER, *ODOCOILEUS VIRGINANUS*

Introduction

Anthrax, caused by the spore forming gram positive bacterium Bacillus anthracis, is a disease of livestock and wildlife enzootic or re-emerging in discrete areas of the United States and an ongoing human and veterinary public health problem is many parts of the world. Within enzootic areas anthrax occurs sporadically in most years and epizootically in select years (Blackburn and Goodin 2013). Disease occurs most commonly in wild and domesticated herbivores, and humans primarily develop illness as a result of contact with infected tissues when slaughtering or butchering animals which have died from anthrax, and through consumption of undercooked meat (Woods et al. 2004, Gombe et al. 2010, Özden et al. 2012). Despite literature on soil characteristics of anthrax zones (Van Ness 1956, 1967, Dragon and Rennie 1995, Smith et al. 2000), empirical descriptions of anthrax occurrence in endemic areas (Turnbull et al. 1992, Dragon et al. 1999, Smith 1999, Turnbull and Shadomy 2010, Hampson et al. 2011, Turner et al. 2013) and anthrax as a bioterrorism agent (Fowler et al. 2005, Bravata et al. 2006, Zaric et al. 2008), unanswered questions remain about the mechanisms that initiate and perpetuate natural outbreaks. Anthrax ecology, including the timing of outbreaks, the species affected and the mechanisms driving the spread of the disease appears to be closely tied to the ecology of the landscape, the local climate, the assembly of host and vector species on the landscape and host behavior. As a result of this dependence on landscape level factors, it is increasingly apparent that anthrax dynamics should be studied within local contexts. Prevention and control of

anthrax is achievable in domestic livestock through vaccination, whereas in wildlife populations control of anthrax relies on prompt identification and disposal of carcasses (Blackburn et al. 2007). Knowledge of local anthrax patterns, therefore, can direct search efforts to facilitate carcass decontamination, which appears to lessen outbreak duration and intensity. This has direct economic benefits to wildlife populations in addition to protecting domestic animal and public health.

Bacillus anthracis has a two part life cycle consisting of an infective stage, during which vegetative cells replicate in the host, and a persistence phase in which spores maintain in soil. Certain soil and characteristics, such as high calcium content and alkalinity, may support spore persistence for long periods of time (Van Ness 1971, Dragon et al. 2005). More recent evidence also suggests that the pathogen may undergo periodic replication events in the environment (Schuch and Fischetti 2009, Schuch et al. 2010), though field evidence has not yet confirmed this. B. anthracis is transmitted through environmental pathways, as opposed to direct transmission from host to host. Herbivores are most likely infected through exposure to an infective dose of spores or, if successful replication does occur outside of mammalian hosts (Saile and Koehler 2006, Dey et al. 2012), vegetative cells from the environmental reservoir, although the potential of vegetative cells to survive and cause infection after entering an animal is not well characterized The primary mechanisms through which subsequent animals are exposed during an outbreak are poorly understood. Potential mechanisms through which animals are infected in an outbreak include contact with contamination in the environment around fresh carcasses or insect vectors, although simultaneous exposure to a common environmental source cannot be ruled out. The

spatially localized nature of outbreaks across most ecosystems (Hugh-Jones and de Vos 2002, Hampson et al. 2011) points to meso and micro scale ecological processes as being important to anthrax ecology. It has been suggested that the processes through which the pathogen is transmitted through the environment to a new host likely differs based on the feeding habits of the host during anthrax seasons (Hugh-Jones and de Vos 2002, Hampson et al. 2011). Grazers are likely to ingest spores or vegetative cells which are present on grasses or in soils due to runoff from rains or contamination from carcasses (Turner et al. 2013). Browsing species, in contrast, may consume vegetative cells or spores deposited on leaves by necrophagous flies, if present in the landscape (Blackburn et al. 2010b).

A number of anthrax investigations have mapped outbreaks at the individual carcass, farm or municipality level (Dragon et al. 1999, Kenefic et al. 2008a, Aikembayev et al. 2010, Epp et al. 2010a, Fasanella et al. 2010), and these studies provide valuable information about the spatial distribution of the disease. Fewer studies have explicitly integrated geospatial analyses, GIS and spatial statistics into anthrax outbreak studies. Space-time cluster analysis using a Bernoulli model demonstrated spatial and temporal clustering attributable to different genetic lineages in Kruger National Park, South Africa, and these lineages also appeared to have different ecological tolerances (Smith et al. 1999, Smith et al. 2000). In an effort to understand spatial patterns within a regional anthrax epizootic during a single summer in Saskatchewan, Canada, Epp et al. (Epp et al. 2010a) integrated a trend surface and vector velocity analysis to estimate the overall directionality of new farms reporting cases in relation to spatio-temporal clusters. At the farm level, each of three significant

spatio-temporal clusters was associated with a unique directional vector. The authors speculated that the timing and different directionalities of these clusters were the result of environmental conditions or insect vector populations which differed in both space and time. Moving to a longer time period, a dataset of outbreaks spanning a 40 year time period was analyzed using a kernel density estimation to show different historical distributions of anthrax in small and large ruminants in Kazakhstan (Aikembayev et al. 2010), and later a prospective cumulative sum technique (CUSUM) was used to demonstrate spatio-temporal clusters of livestock in the country (Kracalik et al. 2011). Both studies indicated common and persistent clustering in southern Kazakhstan, perhaps indicating ecological conditions well suited to pathogen persistence (Kracalik et al. 2011). More recently, environmental conditions associated with spatial clusters of disease identified using the Getis-Ord (Gi*) statistic and a multidirectional optimal ecotope algorithm (AMOEBA) were evaluated in Kazkahstan (Kracalik et al. 2012). This linkage of cluster analysis and environmental conditions supported associations between environmental conditions, pathogen persistence and disease occurrence, and such techniques provide guidance for targeting surveillance and vaccination delivery to areas identified as high risk.

While these spatial studies indicate that at regional levels there exists clustering and directionality of outbreaks, few individual level spatio-temporal studies have been undertaken to understand fine scale outbreak dynamics and none have linked such analyses to genetic diversity. Similar to the concept of epidemiological trace-back, in which genetic typing is used to locate a source of infection, the movement of specific genotypes across space and time can be traced over the course of an outbreak. This

gene flow is used to approximate movement or transmission of the pathogen. Combining genetic information with epidemiologic and spatio-temporal analyses could prove to be powerful for exploring fine scale anthrax dynamics and generate hypotheses about environmental transmission mechanisms. The molecular diversity of B. anthracis can be described using three primary sets of molecular markers. Single nucleotide polymorphisms (SNPs) place bacterial isolates into broad phylogenetic groupings (Van Ert et al. 2007a) and multiple locus variable number tandem repeat analysis (MLVA) provides more detailed assignments to genetic lineages and sub-lineages (Keim et al. 2000, Lista et al. 2006, Van Ert et al. 2007a). Within a single anthrax outbreak all isolates will often group into a single MLVA defined genotype. More highly mutable genetic markers, single nucleotide repeats (SNRs), may demonstrate diversity within a single outbreak, and even within a single carcass site (Stratilo et al. 2006, Kenefic et al. 2008a, Kenefic et al. 2008b, Fasanella et al. 2010). Because of this potential for rapidly evolving diversity, SNRs may be useful for tracing transmission pathways at fine spatial and temporal scales (Keim et al. 2004). A number of studies have mapped the distribution of MLVA and SNR genotypes at regional to fine scales within outbreaks (Kenefic et al. 2008a, Fasanella et al. 2010, Stratilo and Bader 2012), and one study has linked epidemiological data to SNR analysis over a multi-year period (Beyer et al. 2012). In the current study, spatio-temporal data from an outbreak in a single-host browser system in West Texas was analyzed to determine if spatial, temporal, or spatio-temporal trends could be discerned, which might suggest specific anthrax transmission mechanisms. We also evaluated results of a 25 marker MLVA (MLVA-25) and 4 marker SNR (SNR-4) analysis performed on isolates collected during

the epizootic to reveal additional information about the transmission of B. anthracis in this outbreak.

Methods

The Study Site

The study site is an approximately 7,406 hectare ranch located ~80 kilometers north of Del Rio, Texas along the Edwards Plateau (Figure 5-1A). The area is characterized by extensive arid grasslands and thorny shrubs growing in open stands. The eastern portion of the ranch is dominated by a branching lowland dry riverbed with dense vegetation, whereas the western portion of the range has a higher elevation, thinner soil and sparser vegetation. This ranch manages a herd of free range white-tailed deer, *Odocoileus virginanus,* and has an extensive history of anthrax (Hugh-Jones and Blackburn 2009), most recently including epizootics in 2001 and 2005, and sporadic cases in intermediate years (Blackburn and Goodin 2013). In all years cases occur exclusively on the central and eastern portions of the ranch. Since 2001, active surveillance has been conducted by ranch and research staff during the anthrax season which runs approximately from May to October. The 2005 epizootic had a mortality of at least 40 white-tailed deer, or a mortality estimate of ~3.7% (95% CI: 2.7 - 4.9%) of the herd (Blackburn 2006). The home ranges of white-tailed deer on the ranch were estimated by Blackburn (2006) using daily acquired VHF-telemetry signals of individual deer during the 2005 summer. Blackburn (2006) also intensively mapped biting fly densities during the 2005 epizootic summer, and confirmed that necrophagous flies associated with carcasses were positive for *B. anthracis* bacteria (Blackburn et al. 2010b). During the outbreak, carcasses were globally spatially clustered (Blackburn et al., in Review) and cases were concentrated to a single area in the central/eastern

region of the ranch despite a large population of deer distributed widely across the entire ranch.

Anthrax Occurrence Data

All data from this study were derived from an acute anthrax epizootic in the study site beginning early September 2005 and lasting approximately 3 weeks (Blackburn 2006). Direct searching from the ground and visualization of circling vultures, *Cathartes aura*, were used to locate carcasses. This is a commonly employed surveillance technique on this ranch as well as in other ecosystems (Turnbull et al. 1992, Bellan et al. 2013b). Once a carcass was located, location was recorded with handheld GPS units, gender determined and date of death and age were estimated when possible. Samples of tissue, small bone fragments and necrophagous flies were collected from carcasses. Soil samples were also obtained from carcass sites. Carcasses were then disposed of by burning, with a number of carcasses destroyed (burned) without sample collection, other than age, gender and GPS location. Anthrax was diagnosed using classical microbiological methods and polymerase chain reaction (PCR) as described in Blackburn et al. (2010b). Briefly, after heat shocking (30 minutes at 70^0C) or suspension in 100% ethanol, samples were concentrated by centrifugation. The pellet was re-suspended in water and plated onto sheep blood agar and ACTSBA media. Gram staining was used for presumptive identification of *B. anthracis* and suspect colonies were subcultured. Seventeen presumptive *Bacillus anthracis* isolates were recovered from diagnostic samples collected during the outbreak. One isolate was presumptively identified as *B. anthracis*, but failed to genotype as *B. anthracis* using MLVA-25 typing. Upon subsequent testing, it was negative on the *plcR* SNP assay (Easterday et al. 2005)and excluded from the dataset. Clinically suspect cases were

defined by clinical characteristics such as a typical saw-horse appearance to the carcass and evidence of shade seeking behavior, and positive cases are those which were laboratory confirmed. Cases that were culture negative in the laboratory were excluded from these analyses.

Spatio-temporal Analysis

The outbreak was mapped in space and time in ArcMap 10 (Environmental Systems Research Institute 2011. ArcGIS Desktop: Release 10. Redlands, CA:). Following Ward and Carpenter (2000), we employed a multiple procedure exploratory framework to detect spatial and spatio-temporal clustering. Blackburn et al. (In Review) confirmed global clustering during the outbreak using the average nearest neighbor index, which does not indicate the spatial scale of dependence; rather it confirms that the distance between carcasses were closer together than expected by random chance. Therefore, we used the distance-based Ripley's *K*-function to determine the distance at which spatial clustering of carcass locations occurred (Ripley 1977). The *K*-function is a second order analysis which identifies the spatial distance at which clustering occurs and is evaluated against a null distribution created using a Monte Carlo randomization (O'Brien et al. 1999). The point of maximum difference between observed and expected K value indicates the distance at which cases cluster. Data from radio-telemetry of white-tailed deer on the ranch during the outbreak season (Blackburn 2006) were used to evaluate relationships between space-time patterns of carcasses and deer behavior. Specifically, the average minimum, mean and maximum distances traveled by an animal between any two consecutive relocations was used to parameterize biologically relevant movement patterns of individual animals (Blackburn et al. In

Review Annals). All positive and suspect carcass locations were used for the *K*-function analysis, which was conducted in ArcMap 10.

Trend analysis and clustering techniques were applied to look for spatio-temporal patterns within the outbreak. First, second and third polynomial trend surfaces were fitted to the data in ArcMap using the geostatistical analyst extension to look for general trends in space and time. Trend analysis is a global surface fitting procedure that looks for larger scale, smooth changes in the data across the study area (Unwin 1975, Moore 1999, Berrang-Ford et al. 2006). This procedure expresses the independent variable, *z*, in this study date of death, as a polynomial function of the geographic coordinates, *x* and *y*. A regression model is fit by a least squares method such that the sum of the squared deviations from the trend surface is minimized. The trend function can be visualized as fit lines in three dimensions (*x*, *y* and *z*), or as an interpolated surface. This technique is sensitive to outliers and becomes unstable at the edges of the geographic extents of the data, and therefore is not used to extrapolate beyond the limits of an outbreak, but rather to understand global space time trends within spatial boundaries defined by known cases. The optimum polynomial function was evaluated using the root mean square error term and visualized in three dimensions and as an interpolated contour surface

The Mantel and Jacquez' k-nearest neighbor tests were used to test for global space-time clustering. The Mantel test evaluates the distance between events in time and space against a null hypothesis that the distances are independent (Mantel 1967). Specifically, this test was used to determine whether shorter distances between carcasses were associated with shorter temporal intervals between deaths. Jacquez' k-

nearest neighbor evaluates whether cases that are nearest neighbors in space are also nearest neighbors in time, under a null hypothesis the probability of two events being nearest neighbors in space is independent of the probability of them being nearest neighbors in time (Jacquez 1996a). The Jacques k-nearest neighbor differs from the Mantel test because it is more likely to detect non-linear relationships, and in addition it is independent of the density of events because it measures nearest neighbors rather than actual distances. The significance of both tests was also evaluated using a Monte Carlo randomization. The Mantel test and Jacquez's k-nearest neighbor tests were conducted using ClusterSeer software (ClusterSeer v2, BioMedware, Inc, Ann Arbor MI).

The direction test was used to determine whether carcass locations trended in any specific direction over the course of the outbreak (Jacquez 1996b). The direction test uses retrospective individual case data to calculate the average direction in which cases advance during an outbreak. The test outputs a directional vector for which east is 0 and north is 90 degrees (Figure 5-2), and a magnitude value of the vector ranging from 0 to 1. The magnitude, also called the angular concentration, represents the variance in the angles of the cases contributing to the vector. A magnitude approaching 1 indicates less variance in the angles and therefore a more concentrated directionality, whereas a magnitude of the vector approaching 0 indicates the outbreak pattern is more dispersed, suggesting that transmission occurs randomly or in a more diffuse pattern. Significance of the test is evaluated with a randomization procedure equivalent to holding the case locations fixed and varying their temporal occurrence and repeating the test to generate a null distribution against which the test data can be compared.

Chains of infection have at least two ends at the first and last case, but also can branch when cases occur at the same time. Cases may be connected in time in one of three time connection matrices. In a 'relative' time matrix, each case is connected to all of the subsequent cases and is best for understanding processes occurring over longer time spans. The 'adjacent' matrix connects cases only to their nearest neighbors in time and is best for processes occurring over shorter durations. The 'following' matrix connects each case only to the one immediately subsequent and is appropriate for tracing the average direction of transmission processes. Each of the time connection matrices was attempted to explore the scale of possible processes driving transmission. Lastly, the carcass locations were divided into two groups under the hypothesis that there were two separate epizootic events, and the direction tests carried out separately for each group. The direction test was carried out in ClusterSeer software.

Genetic Analysis

MLVA-25 genotyping. The 5' fluorescent-labeled MLVA-25 primers described by Lista et al. (2006) were synthesized. Multiplex polymerase chain reactions (PCR) were carried out using 0.5 uM each of the VNTR specific forward and reverse Universal tail primers, 1.0 ng of DNA template, 2.0 mM $MgCl_2$, 2.5 mM of each dNTP, and 1.25 U Platinum Taq Polymerase per reaction. Thermal cycling parameters were as per Lista et al. (2006). PCR products were diluted 1:125 in molecular grade water and 1.0 uL of the diluted multiplexes were mixed with 19.0 uL of a formamide/LIZ 1200 (Applied Biosystems, Foster City,CA, USA) size standard mixture and denatured.

SNR-4 genotyping. We individually amplified the four SNR loci described in Kenefic et al. (Kenefic et al. 2008b). The 10.0 µl PCR reactions were carried out with final concentrations of the following: 1.0 µL template DNA per reaction, 0.2 µM of each

of the four forward and reverse primers, 1X PCR buffer, 0.5 U per reaction *pfu* Polymerase (Agilent technologies, Wilmington DE), 3 mM $MgCl_2^*$, and 0.25 mM of each dNTP. The PCR products were pooled with a final dilution of HM-1, 2, and 6 at 1:20 and HM-13 at 1:10; 1.0 uL of the pooled products were mixed with 19.0 uL of a formamide/LIZ 500 (Applied Biosystems) size standard mixture and denatured. Fragment sizing for MLVA-25 and SNR-4 was performed on an ABI 3730 (Applied Biosystems) and VNTR sizes were determined using GeneMapper™ software (Applied Biosystems).

We examined genetic relationships between samples in the context of global representatives from Lista et al. (2006) using Unweighted Pair Group Method with Arithmetic Averages (UPGMA) cluster analysis. Matrix distances were calculated in PAUP 4.0 (Sinauer Associates, Inc., Sunderland, MA, USA) and imported into MEGA 5 (Tamura et al. 2007) in order to build phylogenetic trees based on MLVA-25. Data from SNR-4 analyses were evaluated using differences in basepair repeats between isolates grouping within MLVA-25 genotypes and UPGMA. Distance matrices were again generated in PAUP 4.0 and imported into MEGA 3.1 for tree building.

Results

Outbreak Description

Forty eight carcasses were identified during the epizootic period. Of these, 41 were geo-referenced and 33 were positive or suspect carcasses, of which 16 (48%) were positive and 17 (51%) were suspect. No demographic data, geographic coordinates or date found were available for 2 cases, both of which were considered suspect; 1 case record contained location data, but no date found. Locations of carcasses were overlain on a 10m digital elevation map (DEM) of the study area in

Figure 5-1B. The first positive carcasses were found September 6 and estimated to have died on September 4th, and the last positive or suspect carcass was found on September 28th. Three cases were estimated to have died on September 4th. These cases were dispersed spatially and are circled in Figure 5-1B. A date of death was determined for 20 carcasses. Figure 5-3 shows the epidemic curve of the outbreak by date carcasses were found (A; n=30) and by estimated dates of death (B; n=20). Of the 31 positive or suspect cases with gender data, 18 (58%) were male and 13 (42%) were female. All positive or suspect carcasses were located in the eastern/central region of the ranch.

Spatio-temporal analysis

The K-function analysis using 31 carcass locations, bins of 40m and a maximum distance of 1500m, which were chosen based on the minimum (40m) and maximum (1259m), respectively, daily movement of white-tailed deer as measured in 2005 (Blackburn et al. In Review) was run. Cases were clustered at distances of approximately 680m as determined by the maximum difference between the expected and observed K-value (Figure 5-4). Following calculations by Blackburn (2006), the home range of the nearest neighbor carcass in space frequently overlapped with the carcass locations even at the most conservative estimate of 350m, and increasingly as the home range estimate is made equivalent to the K-function results and the maximum home range estimates (Figure 5-5).

Figure 5-6A illustrates the locations of the 20 cases with data in space and time. In this figure, the height of the bar corresponds to the length of time elapsed between the start of the epizootic and the date of death. Figure 5-6B shows the results of second polynomial trend analysis, which minimized the root mean square error as

compared to first and third order polynomial functions. The trend analysis suggested new cases spread outward in space over the duration of the outbreak with the predominant overall trend towards the southeast and southwest, although it should be noted that three cases are recorded as having died on day one, two located towards the east of the outbreak area and one located more centrally on the ranch. Mantel testing did not indicate significant global space-time clustering (Mantel's r 0.1930, p=0.0680). The k-nearest neighbor test was not significant for global space-time clustering (test statistic=2.1630, p=0.2970).

The results of the relative matrix direction test using the 20 cases having date of death estimates was a vector of 349.41 degrees (Figure 5-2A), which indicates a roughly east – southeast advance of cases, although the magnitude of the vector, 0.0299, is low and not significantly different from the null hypothesis (p-value 0.7220). Using the adjacent and following time connection matrices produced similar results (Table 5-2). To evaluate the effect of sample size and data collection bias, the tests were repeated using all 30 cases with both geographic coordinates and at minimum a date found, assigning date found to date of death when the latter was unknown. The average angle of the relative matrix vector was then 347.51 with a significant concentration of 0.1406 (p=0.0010). Applying the adjacent and following matrices to this dataset changed the average vector angles to 36.99 (both adjacent and following matrices) with non-significant magnitudes of 0.6060 and 0.7600, respectively. However, if the group of 15 cases in the eastern portion of the outbreak (eastern outbreak group) is considered separately from the group of 5 located to the center of the ranch (central outbreak group), the results suggest a different set of patterns (Figure 5-2B). In the

eastern outbreak group, the relative direction test results in a vector of 27.66 degrees, or a east-northeast direction, with a magnitude of 0.1704 (p-value 0.0020) and the following and adjacent matrices have low magnitude southeast vectors. The central outbreak group has a relative matrix result of an east-southeast vector of moderate magnitude which approaches significance (321.31, 0.3298, p=0.0760), and here again use of tests of a more short term pattern shift the directionality to the west (Table 5-2).

Genetic data

Fourteen of the *B. anthracis* positive isolates were derived from carcasses and two were from soil samples taken adjacent to carcasses. All isolates grouped into the A4 (Vollum) cluster using the MLVA-25 Keim et al. (2000) nomenclature and the genotype 57 group using the Lista et al. (2006) groupings (Figure 5-7**)**. All fourteen isolates were identical on MLVA-25. SNR analysis of the 16 isolates yielded four subgenotypes (SGT; Table 5-3). The dominant SNR type in the outbreak was given the designation SGT-1 based on it being the most frequently identified genotype among the isolates available in this outbreak. The only carcass reported as dying on day 1 of the outbreak for which a genotyped sample was available was associated with SGT-1. Two different SNR genotypes were isolated from the carcass and adjacent soil at each of two locations. SGT-2 differed from SGT-1 by one base pair at HM-2 and was isolated from the soil adjacent to a carcass from which SGT-1 was isolated, and SGT-1 was isolated from soil adjacent to the carcass from which SGT-3 was identified. SGT -3, which differed from SGT-1 by loss of one base pair at HM-1, and SGT 4, which differed from SGT-1 at markers HM-1 and HM-2, were both isolated from animals that were found on September 14. Date of death, however, was not estimated for these two

cases. Figure 5-8 shows locations of all SGTs as well as the sampling location for carcass and soil isolates.

Discussion

The ecology of anthrax, like many zoonotic diseases and particularly those with environmental life stages and transmission, is complex and dependent on a multitude of factors. Characteristics of the primary host species, seasonal variations in host resource use and predation pressures, as examples, are likely to shape spatial and temporal variations in anthrax dynamics (Proffitt et al. 2010, Alexander et al. 2012). Such factors influence where and how on the landscape wildlife contacts the pathogen as well as the temporality of exposure, even if the presence of the pathogen has a more consistent spatial and temporal environmental signature. In addition, the presence (or absence) of potential insect vectors may also play a role in the ecology of anthrax. Likewise evidence suggests local pathogen population genetics can interact with soil characteristics (Smith et al. 2000) or climatic variables (Chapter 2). Because of this complexity, it is possible mechanisms driving anthrax infection and outbreak perpetuation that are unique to individual ecosystems. In this study a single outbreak in a restricted spatial area of West Texas in a three week time period was analyzed. This study had two goals, one, to explore spatial and spatio-temporal trends in the outbreak and two, to add high resolution genetic analysis to evaluate the genetic diversity of the pathogen relative to these spatial and spatio-temporal trends.

The locations of carcasses in the outbreak tended to be in the low lying dry riverbeds which are the primary landscape feature of the eastern portion of the ranch, despite the population of deer being distributed across the ranch (Blackburn et al. In Review). According to estimates of white-tailed deer home ranges in this ecosystem

derived during the outbreak period (Blackburn 2006), animals are likely to have died close to the place of infection. The K-function analysis indicated that carcasses tend to be clustered at a distance of approximately 680 meters, such that most carcasses locations were within the daily activity space of nearby deer which died of anthrax, particularly in the eastern portion of the outbreak, based on deer movement data form this study area (Blackburn et al., in Review). In this scenario, individuals have an opportunity to move near enough to a carcass to encounter freshly contaminated soil, browse contaminated by necropahgous flies, or biting flies which may have recently fed on bacteremic animals or carcass tissues. During the anthrax season (summer months), deer in Texas primarily consume leaves (browse) (Fulbright and Ortega-s 2006), making exposure through grasses and soils less likely than through contaminated browse. Areas of higher fly density, tended to occur in these riverbeds during the outbreak period (Blackburn 2006). Based on the analyses here and past ecological studies on this ranch, each route of exposure, namely direct contact with carcasses, contaminated browse or biting flies, is plausible.

An exception to this overlapping activity space/home range scenario is the case occurring on day 20 to the western most reach of the epizootic. The long spatial distance to this case may be explained by missed intermediate cases, higher than average movement of that animal or an insect vector, or independent exposure to an environmental source. In the low lying dry riverbeds, the soil is deeper and richer than in the higher elevations, and this deep soil presents conditions which are thought to promote and sustain *B. anthracis* persistence (Van Ness 1956, 1967, Dragon and Rennie 1995), perhaps providing multiple spatially dispersed opportunities for new

infections. However, it is important to consider the potential challenge of characterizing anthrax burden from mortality events given the high likelihood of missed carcasses. The rate of opportunistic carcass sighting and active surveillance relying on sighting of carcasses or scavenger species is likely to decay with the distance between the observers and the carcass sites (Bellan et al. 2013a). Recent modeling work indicates that in the Etosha National Park, Namibia, system approximately 3.8 times more zebra mortalities occurred than carcasses were sighted (Bellan et al. 2013a). While the parameters used for these models were specific to zebras in the Etosha ecosystem, the findings emphasize that mortality in an outbreak is likely to be higher than the numbers of carcasses identified. This is likely to hold true with deer outbreaks in Texas, as the landscape is vast and the size of deer carcasses small. The presence of three widely dispersed carcasses estimated to have died on the first known day of the outbreak suggests either that the outbreak began earlier in time and those carcasses were missed, or that conditions were favorable for animals to be infected with *B. anthracis* across the landscape, each with the potential to trigger multiple trajectories for cases during the epizootic.

The second polynomial trend analysis suggests one or more processes which expands the outbreak outward in all directions. Assuming individuals do not travel long distances during illness or as a result of the epizootic, as is suspected in the ecosystem discussed here (Blackburn 2006), transmission mechanisms driven by insects or contact with fresh carcass sites could potentially produce such a pattern. The global tests for spatio-temporal clustering did not indicate significant clustering in space and time. Using the available data with date of death and spatial coordinates, the Mantel's

r approached significance, suggesting that there was a tendency toward spatial-temporal clustering within the outbreak, although the Mantel test does not indicate specific areas in time or space or the scale at which clustering occurs. The Mantel test is limited, however, by sensitivity to sample size such that smaller samples tend to lower the *r* value as well as by the assumption of linearity. The k-nearest neighbor test should detect clustering with non-linear space-time relationships, and it also did not detect significant global spatio-temporal clustering. In the absence of significant clustering, another approach to detecting significant patterns in the outbreak in space and time was to assess the directionality of the outbreak. The relative time connection matrix direction test indicated that over the course of the entire outbreak the chain of infections trended to the east-southeast. Applying the test to all cases with geographic coordicates, however, suggested a shorter term a process trending to the east-north east. These directionalities roughly parallel the dry river beds common to the central and eastern portion of the ranch, and, furthermore, the measures of shorter term processes indicate a trend toward an area of high genetic diversity. An active process driving infection on a short temporal scale in that area is suggested by this combination of findings. I then chose to divide the carcass locations into two spatial groups because of the widely dispersed carcasses with death dates estimated on day 1 under the possibility that there were two separate outbreak events, such as suggested by Beyer et al. (2012) and Epp et al. (2010a). Here, the eastern outbreak and central outbreak groups showed different spatio-temporal directionalities. The long term direction of the infection chain in the eastern outbreak group was significantly to the east-northeast while the shorter term pattern was to the south, again following the low lying river beds,

whereas in the central outbreak group the longer term movement was to the southeast (approaching significance) and the shorter term to the west/southwest. This may suggest different mechanisms of transmission, perhaps influenced by micro-scale climate and ecology, variations in animal behavior in the two sections of the ranch, or variations in biting fly densities, and necrophagous fly contact with browse in these areas. Alternatively, the results may be biased by differential surveillance in these parts of the ranch or overall by missed carcasses (Bellan et al. 2013a). The different analyses shown here demonstrate that the results of the direction test are influenced by the time connection matrix used, yet this is rarely reported in the literature. I encourage researchers to report the matrix or matrices used and an interpretation in the context of the temporal scale represented by the matrix. The results of the direction test complement the trend analysis, with the direction of the vectors tending towards the same south to southeast movement, with some tendency to the southwest.

A key question in the analysis of this outbreak was whether the addition of genotyping to the spatial analyses would improve our ability to trace the course of the outbreak across the landscape by following movement of individual genotypes. All isolates were identified as a single MLVA-25 genotype, which is consistent with a single enzootic strain of *Bacillus anthracis* introduced into the area at some time in the past (Fasanella et al. 2012). The most common SNR genotype in the outbreak was SGT-1. The three other genotypes, SGTs 2, 3 and 4, represent polymorphisms between and within carcass sites. This high diversity within carcass sites, as well as overall diversity within an outbreak, has been noted previously (Kenefic et al. 2008a, Fasanella et al. 2010, Stratilo and Bader 2012). The animal at one of these sites from which less

common SGTs were isolated was determined to have died on day 8 of the outbreak, whereas a date of death was not available for either of the other two animals associated with genotypes other than SGT-1, although they were found mid-way through the epizootic period. In the context of the overall spatial and temporal course of the outbreak, two of the three sites with unique SGTs were in an area of high transmission activity in which a number of carcasses were in relatively close spatial and temporal proximity and, relative to the day one cases, were located along the average vector for long term movement of the chain of infection. The mutation steps, however, do not support a linear transmission of SGTs. SGT-4 differed from the dominant SGT-1 by 2 mutation steps in HM-1 and is most parsimoniously depicted as evolving from SGT-3. It appears, however, spatially removed from the area in which SGT-3 was isolated, suggesting either an overlooked chain of transmission, an independent exposure to a pathogen source with SGT-4 in the environment, or an independent set of mutation events.

The high diversity between and within carcass locations found here, taken with SNR genotypes appearing spatially and temporally dispersed in a larger epizootic (Fasanella et al. 2010), and previous evidence of polymorphic loci within carcass sites (Stratilo and Bader 2012), present challenges to interpreting genetic diversity in the context of a single outbreak or epizootic until a number of uncertainties can be resolved. SNR diversity is governed by a combination of sporadic and outbreak incidence rates, relative selection within the host and in the environment, the level of bacteremia (and thus the amount of replication in the host), and *in vivo* mutation rates. Without knowledge of these underlying parameters I cannot separate the effects of ecological

factors from those of sampling artifact and genetic drift (Beyer et al. 2012). To illustrate just one set of uncertainties, I do not know whether each unique SGT arose during the epizootic or if they were already present on the landscape as a result of high historical anthrax activity. If the latter, SGT-1 may represent a more fit genotype relative to less common SGTs. It also cannot be ruled out from the data available here whether multiple SNR genotypes occur at every carcass site, with SGT-1 being most common in this outbreak and therefore most often isolated, and, if so, whether this genotype dominates among the isolates tested because of selection pressures or simply because it was the genotype initiating the outbreak. In Namibia, genotyping of isolates collected over multiple outbreak years showed that a few MLVA defined genotypes were persistent across time, although one large outbreak was dominated by a MLVA genotype which was subsequently found only sporadically (Beyer et al. 2012). Although this finding represents a different scale of genetic analysis, it indicates that there may be temporal differences in the predominant genotypes in outbreaks in the same area. Genotyping of isolates from multiple years will determine whether SGT-1 is a persistent sub-genotype on this landscape. More comprehensive epidemiological data, field sampling and experimental work is required to clarify these gaps in understanding and generate testable hypotheses.

All analyses in this study are potentially limited by the small number of cases having complete data in relation to the number of cases suspected to have occurred during the outbreak. In many instances carcasses were burned before samples and data could be collected. Burned remains were consistently culture negative. While prompt and effective disposal is favorable in terms of disease control and is the priority

for ranch managers, there is a simultaneous need for increased efforts toward data and sample collection, in addition to continued efforts to identify sporadic and epizootic cases in endemic areas. Missed cases and missing data have significant epidemiological and analytical repercussions. Obviously, carcasses not promptly identified perpetuate the transmission of anthrax. Equally important, our ability to detect and employ genetic patterns in space and time is biased by incomplete epidemiologic data and biological samples. Our findings emphasize that samples should ideally be obtained and analyzed from every case in an outbreak in order to understand the fine scale dynamics of anthrax. Effort should be undertaken to systematically sample a carcass, surrounding soil, browse and insects to understand how to best utilize this SNR diversity. This can and must be done quickly and will ultimately require a cooperative effort between ranch managers, wildlife officials and veterinarians to develop strategies that maximize efficient use of time and personnel while also collecting valuable data and samples.

The results here suggest that close contact of animals with carcass sites is a key component of anthrax outbreak dynamics. This association should be evaluated in other ecosystems and used to guide future studies of anthrax ecology. Anthrax in managed wildlife populations results in a substantial economic burden by reducing income from wildlife hunting as well as through expenditures for surveillance, testing and disposal of carcasses. Additionally, the pathogen has the potential to spillover into livestock populations occupying the same or nearby land, resulting in serious financial, agricultural and public health consequences. Our results emphasize the need for thorough epidemiologic data collection and biological sampling of sporadic and

epizootic anthrax cases. Furthermore, genetic analysis should be performed for as many cases as feasible to augment spatial studies and descriptive analyses of anthrax. Although many questions about the dynamics of genetic diversity between and within outbreaks remain to be answered, these efforts will allow a better understand the relationship between genetic diversity and local anthrax ecology.

Table 5-1. Results of the direction tests using each matrix option. The results of the direction tests are illustrated graphically in Figure 5-2.

	Death Date			Date found for unknown death dates		
Matrix	Average Angle	Concentration	p-value	Average Angle	Concentration	p-value
Relative	349.41	0.0299	0.7220	347.51	0.1406	0.0010
Adjacent	331.37	0.2187	0.2700	36.99	0.1225	0.6060
Following	331.37	0.2187	0.5320	36.99	0.1225	0.7600
	Eastern Outbreak Group			Central Outbreak Group		
Matrix	Average Angle	Concentration	p-value	Average Angle	Concentration	p-value
Relative	27.6607	0.17038	0.0020	321.31	0.3298	0.0760
Adjacent	298.542	0.0986	0.7970	194.002	0.6656	0.0460
Following	298.542	0.0986	0.9010	194.002	0.66581	0.1890

Table 5-2. Allele sizes used to characterize sub-genotypes

	Allele size (number of base pairs at locus)				
Sub-genotype	HM1	HM2	HM6	HM13	No. Isolates
SGT1	84	[106]	90	118	14
SGT2	84	105	90	118	1
SGT3	[85]	105	90	118	1
SGT4	[83]	105	90	118	1

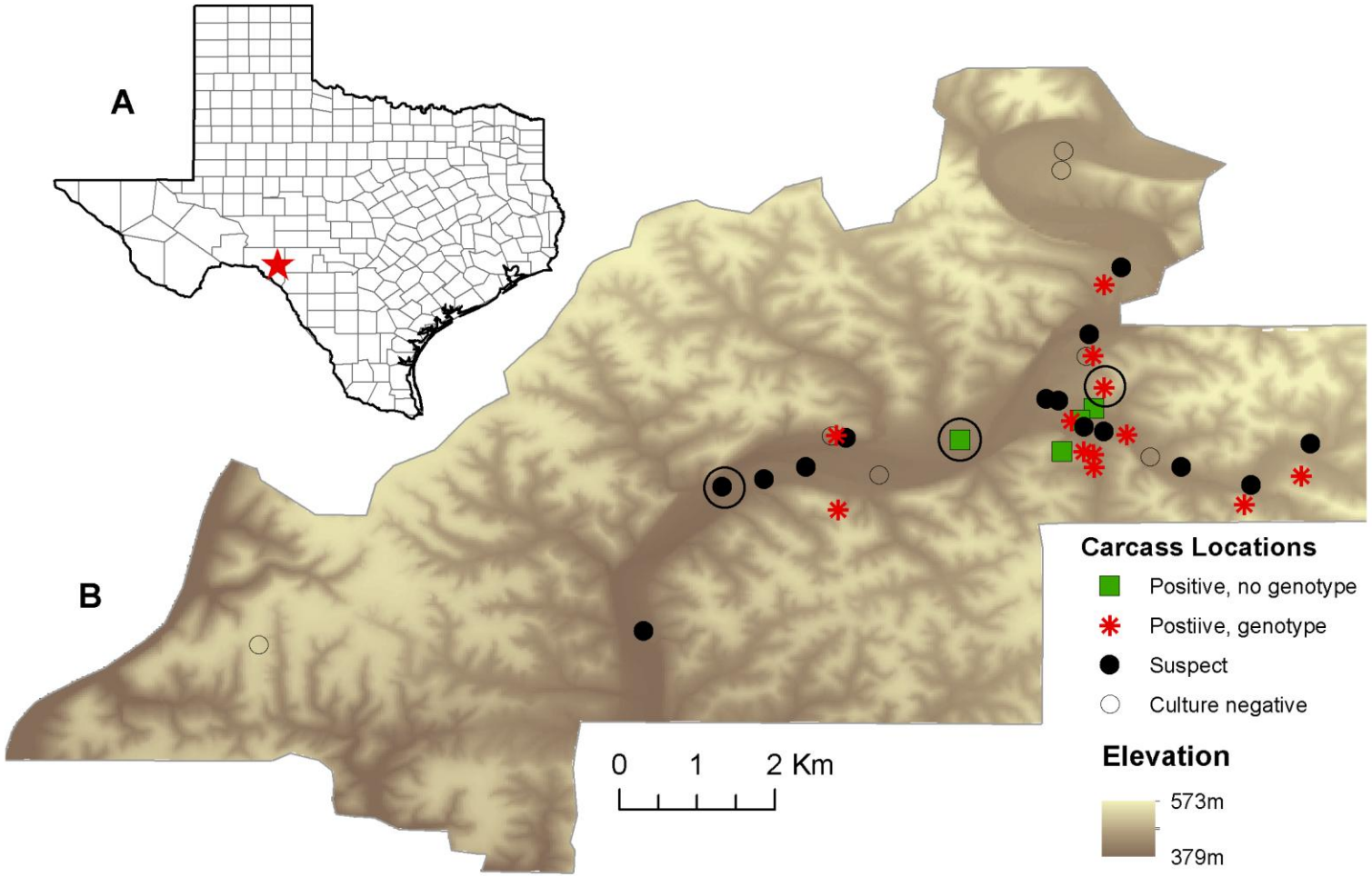

Figure 5-1. Location of study site and carcasses during the outbreak. A) Location of study site (red star) B) Locations of positive, negative and suspect carcasses shown over digital elevation map of the study area. Cases with dates of death on day 1 are circled.

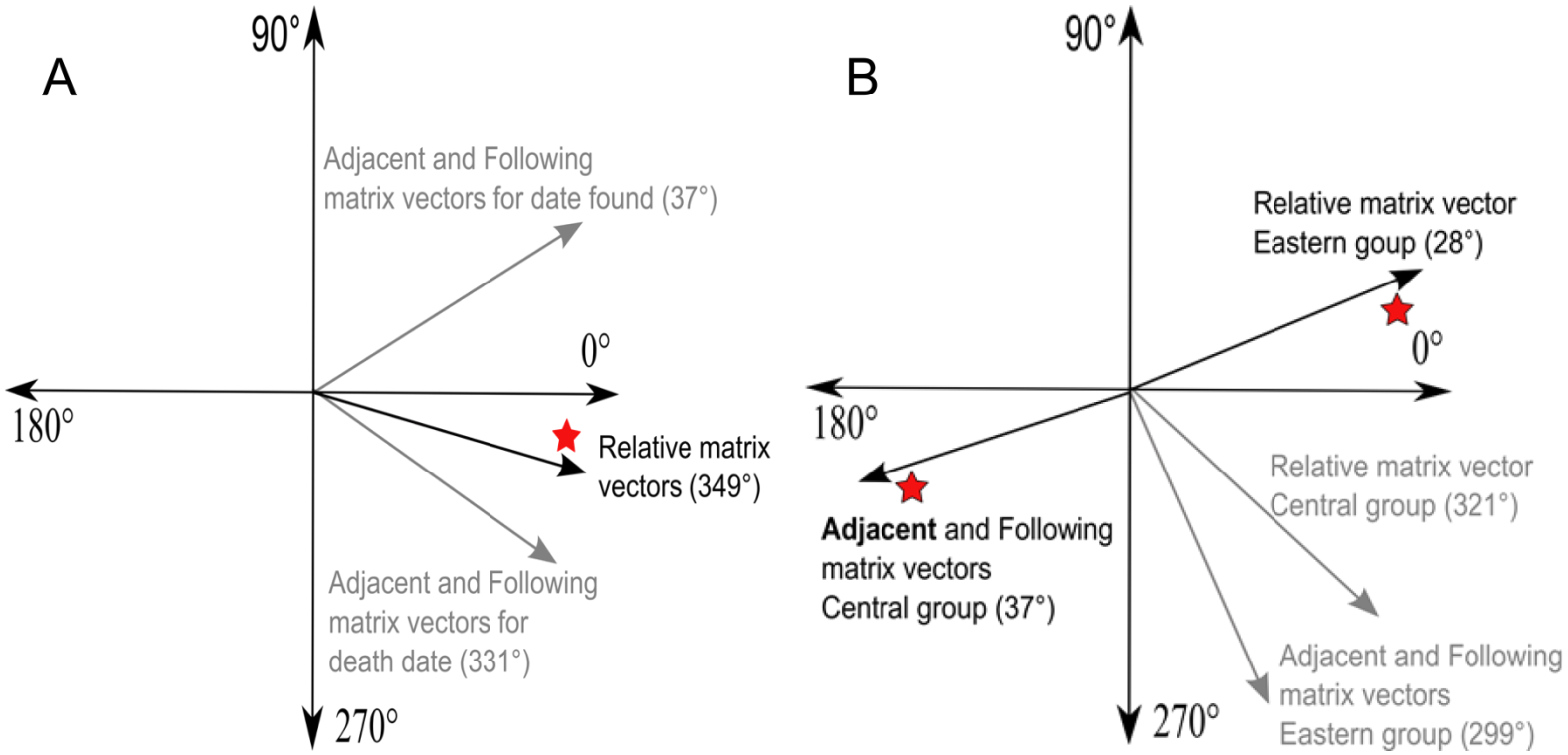

Figure 5-2. Spatial directions indicated by the direction test. East is at 0°, north is at 90°, west is 180° and south is 270°. A) using all carcass locations with dates of death; B) using all carcass locations with geographic coordinates and date found as proxy for date of death.

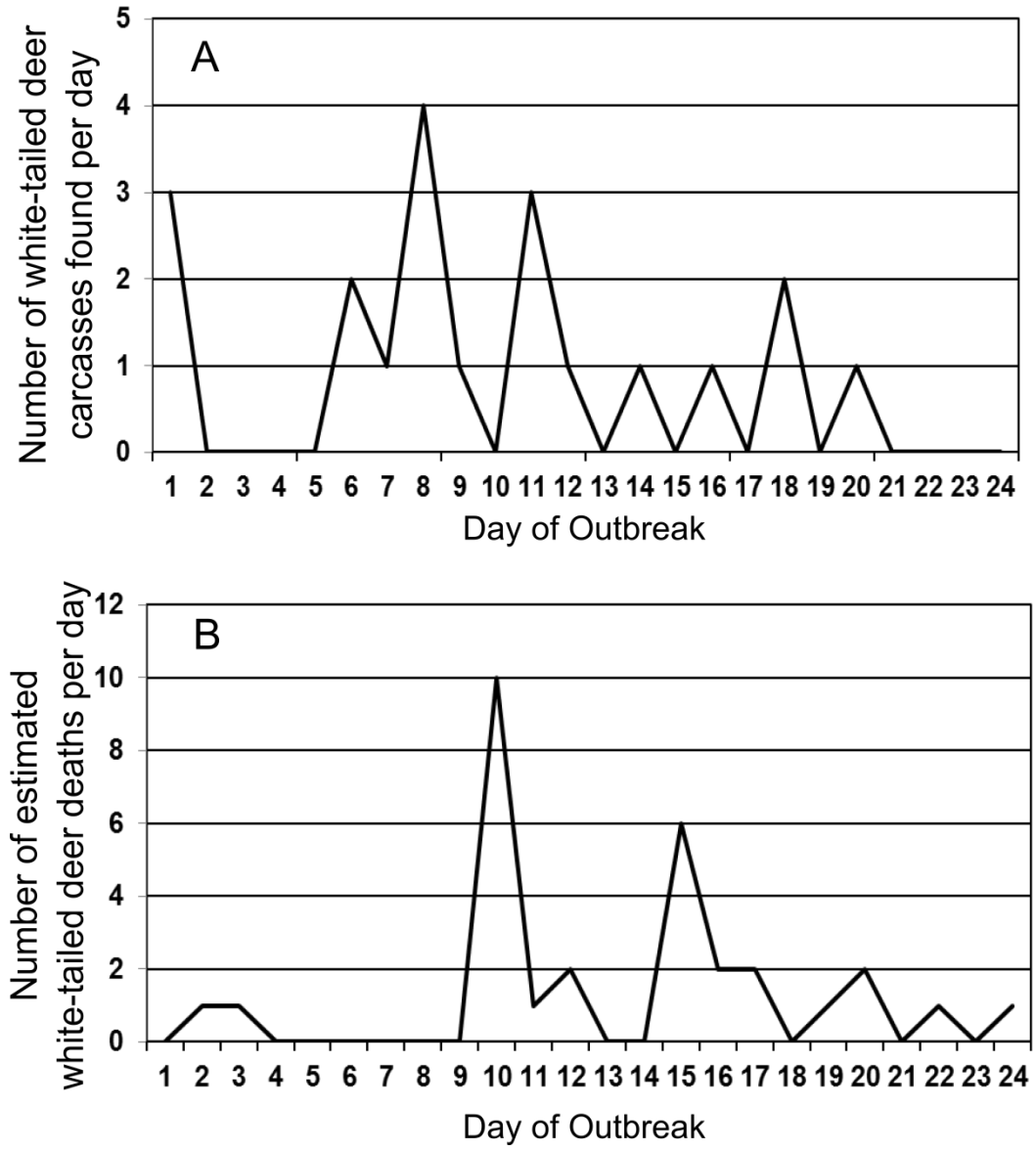

Figure 5-3. Epidemic curves. A) by date white tailed deer carcasses were sighted (n=30) and B) by estimated date of death of white tailed deer for which date of death was estimated (n=20).

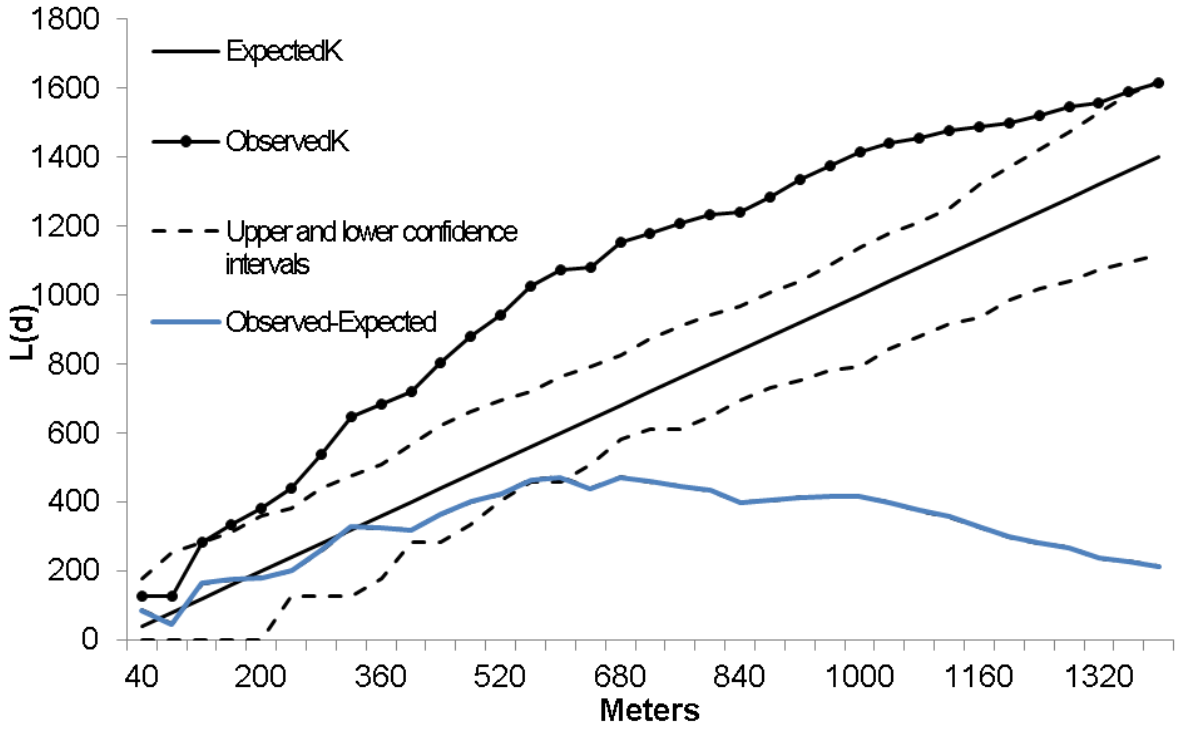

Figure 5-4. K-function analysis graphical output. The maximum difference between the observed and expected K occurred at 680m. The observed K function converges on the expected past 1250m.

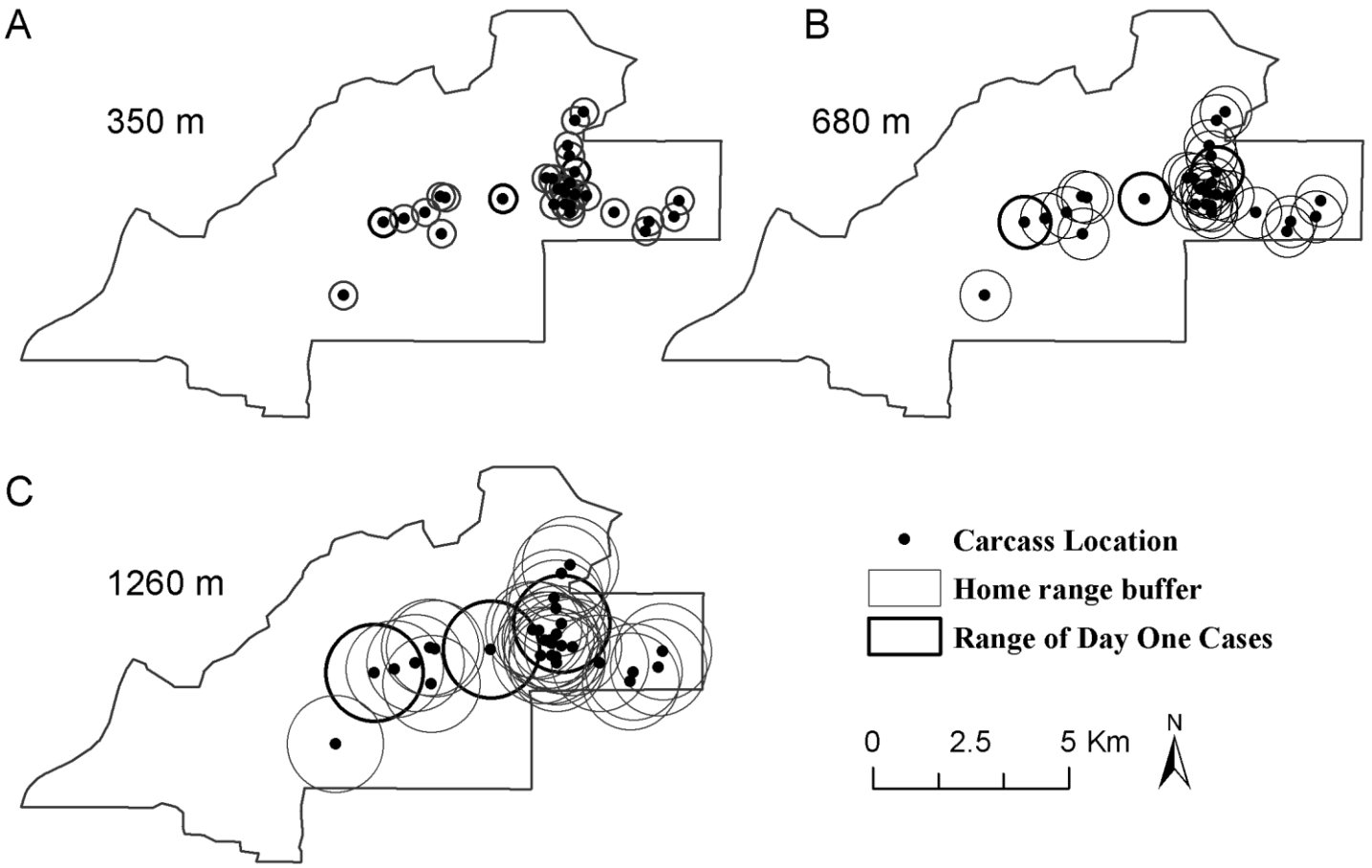

Figure 5-5. Map showing carcass locations and deer movement estimates. A) 360 meter, B) 680 meter, and C) 1259 meter buffers around each carcass location.

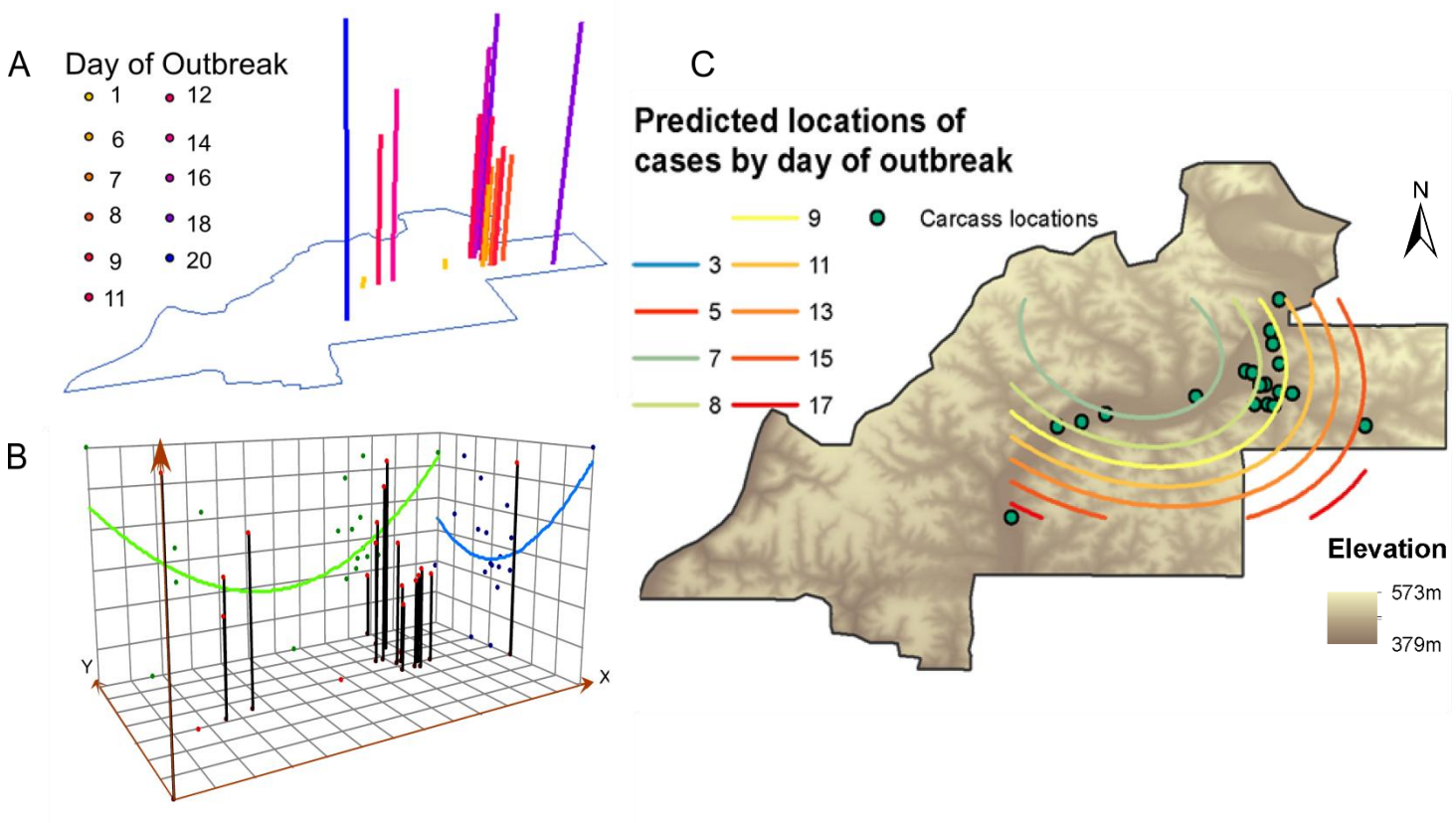

Figure 5-6. Three-dimensional representation of the outbreak and trend surface. A) 3-d plot of 20 carcass locations with geographic and temporal data. B) Second order polynomial trend analysis. East-west is on the x-axis, north-south is on the y-axis, and time is on the z-axis. C) Predicted carcass locations as interpolated by a trend-surface analysis model using all cases with geographic and temporal data.

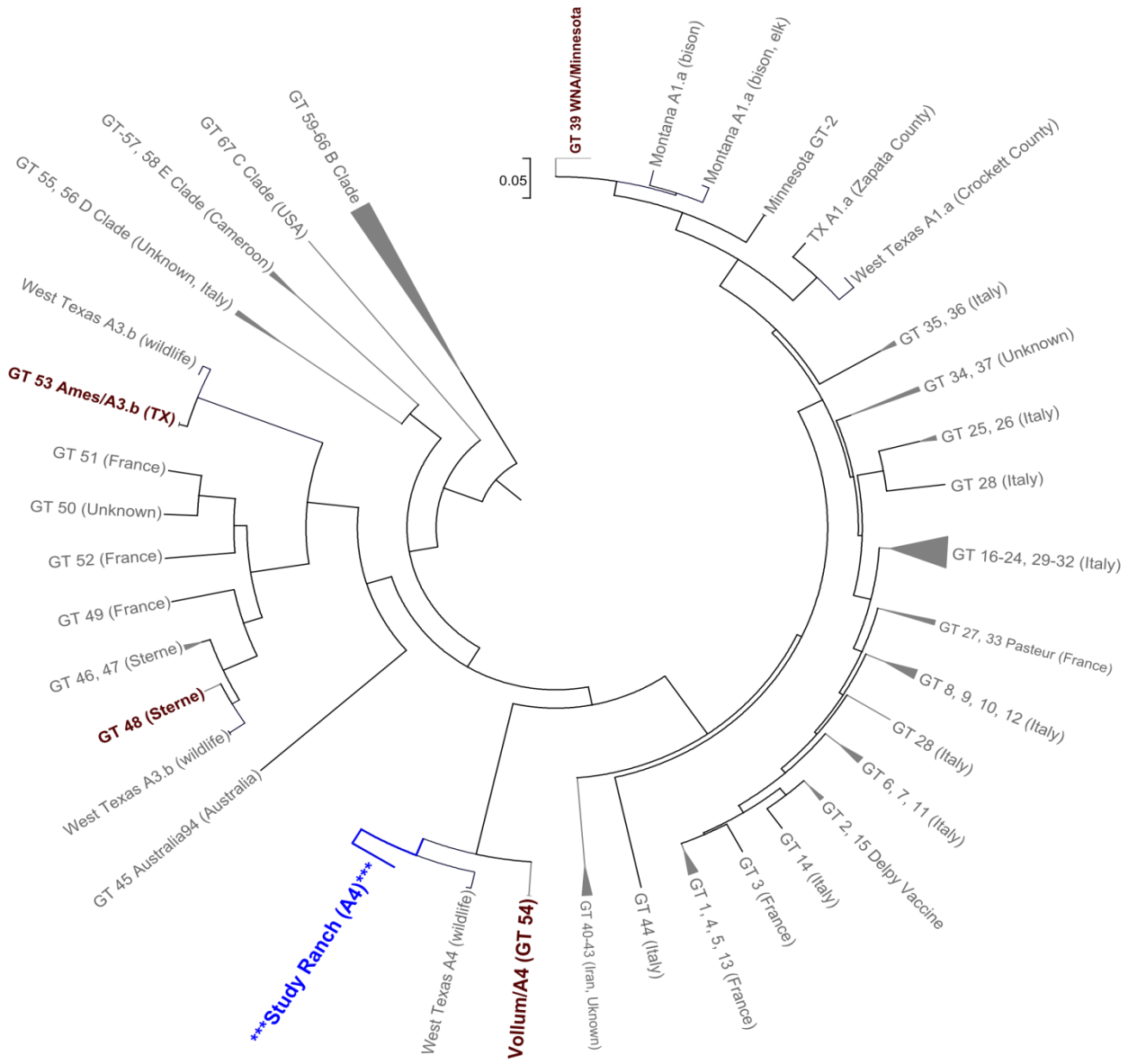

Figure 5-7. Phylogenetic tree showing MLVA-25 genotypes representing the global collection. Study site is indicated in blue. Reference strains (Ames, Vollum, Sterne, WNA) are indicated in dark red.

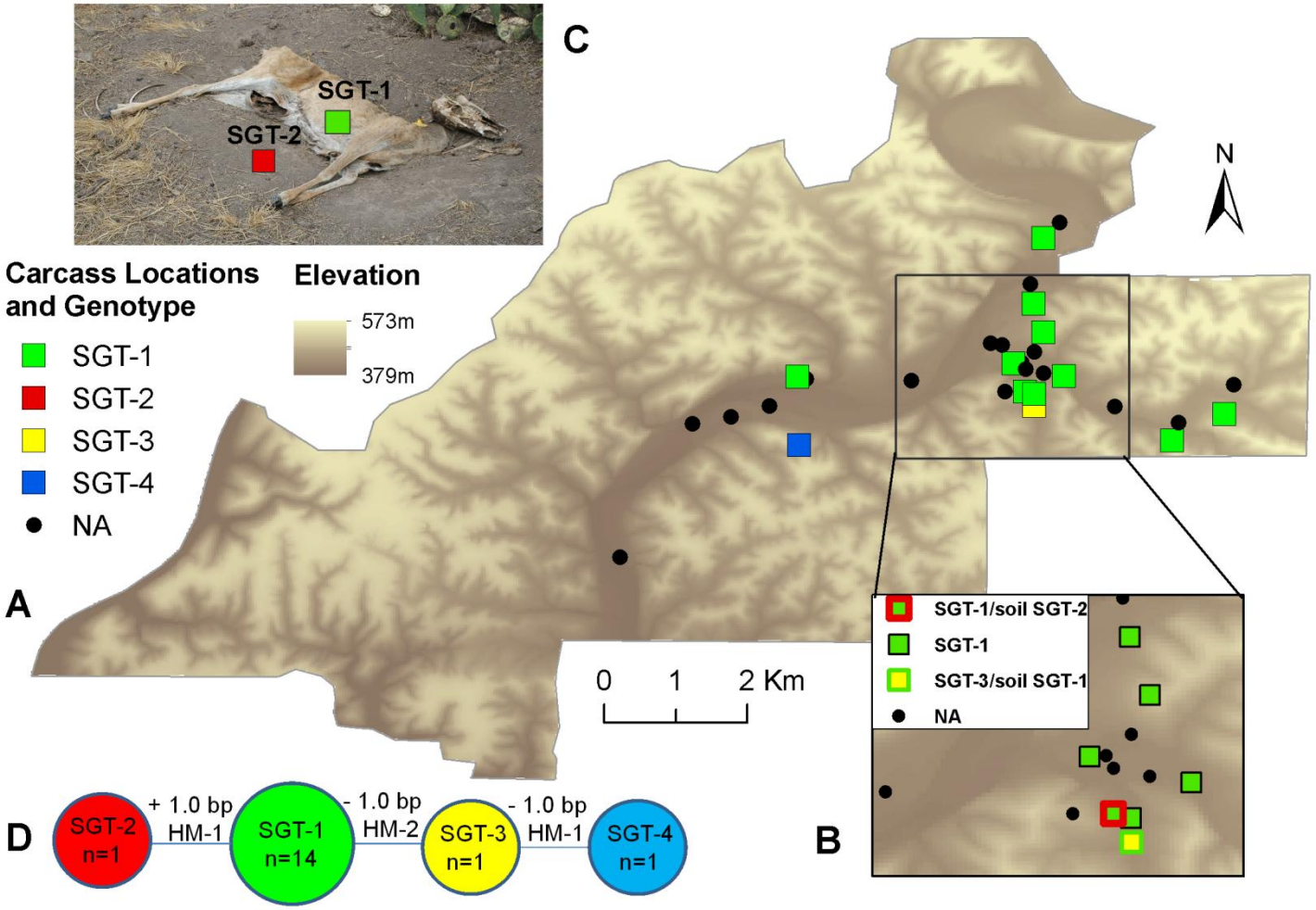

Figure 5-8. Results of SNR genotyping. A) Carcass locations and SNR genotypes (SGT). B) Detail of the high transmission area in which 2 SGTs were found at each of two carcass sites. C) A typical carcass found during the epizootic and sampling scheme through which diverse SGTs were collected. D) Genetic relationships among the 4 SGTs.

CHAPTER 6
BACILLUS ANTHRACIS GENETIC DIVERSITY IN NORTH AMERICA

Introduction

Anthrax is disease affecting wildlife, livestock and humans caused by the spore-forming bacterium *Bacillus anthracis*. The disease has a major economic impact worldwide as a result of the high mortality seen in both livestock and wildlife in outbreak events. Anthrax appears to be well controlled in livestock populations where vaccination is administered annually. However, in areas where vaccination programs are discontinued (Fasanella et al. 2010) or fail (Chikerema et al. 2012), or in wildlife populations in which vaccination is not possible (Blackburn et al. 2007), enzootic anthrax areas continue to experience annual sporadic cases and periodic outbreaks. Cases of sporadic and outbreak associated anthrax carry a significant human health risk, with the majority of human cases resulting from contact with or ingestion of animals which have died from anthrax or contact with contaminated animal products (Woods et al. 2004, Chirundu et al. 2009, Özden et al. 2012). The human health risk is not limited to enzootic areas, as evidenced by cases of anthrax resulting from imported hides for drums (Anaraki et al. 2008, Guh et al. 2010) and contaminated illicit drugs (Booth et al. 2010).

Bacillus anthracis is transmitted indirectly through the environment and has a life cycle consisting of a vegetative stage, in which the bacterium multiplies in the host, and a spore phase, during which it persists in the soil or environment. There is some evidence that a limited amount of spore germination and replication may occur in the environment (Schuch et al. 2010, Dey et al. 2012), although this has yet to be confirmed in natural conditions. Herbivores are exposed to spores in soil or on grasses or browse.

In outbreaks, which are typically spatially and temporally limited, it is unclear whether multiple cases are the result of simultaneous exposure of all animals to an environmental source, exposure to fresh carcass sites, or insect vector mediated pathogen spread. It is understood that certain environmental conditions favor long term pathogen persistence, including calcium rich and alkaline soils (Van Ness 1971, Dragon and Rennie 1995). Outbreaks are associated with climatic factors, and although the specific seasonality of anthrax outbreaks varies across ecosystems, within ecosystems characteristics of anthrax seasons are well recognized (Hugh-Jones and de Vos 2002, Hampson et al. 2011, Turner et al. 2013). This has led to hypotheses about host, behavioral and ecological triggers and drivers of outbreaks (Blackburn 2006, Hampson et al. 2011, Turner et al. 2013). Genetics also play a role in the ecology of anthrax, with some lineages appearing to tolerate wider ranges of environmental conditions than others (Chapter 2) (Smith et al. 2000), although how the population genetics of *B. anthracis* interacts with ecological processes within and between outbreaks is not known.

The molecular diversity of *B. anthracis* has been characterized at global, regional and local levels using combinations of three sets of molecular markers. Single nucleotide polymorphisms (SNPs) are evolutionarily stable markers which place bacterial isolates into broad phylogenetic groupings (Van Ert et al. 2007a). Multiple locus variable number tandem repeat analysis (MLVA) uses combinations of genetic loci with varying levels of diversity to provide more detailed assignments of isolates to genetic lineages and sub-lineages (Keim et al. 2000, Lista et al. 2006, Van Ert et al. 2007a). Within a single anthrax outbreak, and often across multiple outbreak years in a

geographically restricted area, all isolates will frequently group into a single MLVA defined genotype. More highly mutable genetic markers, single nucleotide repeats (SNRs), have been shown to exhibit diversity within single outbreaks in South Dakota (Kenefic et al. 2008a), Italy (Garofolo et al. 2010), Canadian bison (Stratilo and Bader 2012) and Namibia (Beyer et al. 2012), which have 100 year, multi century, 100 year and ancient histories of anthrax, respectively. Genetic analysis augments epidemiological and ecological studies of disease by enabling the tracking of pathogen strains within or between outbreaks (Smith et al. 2000, Girard et al. 2004, Staples et al. 2006, Nakazawa et al. 2010, Pepperell et al. 2011, Vogler et al. 2011b). Efforts to understand anthrax outbreak dynamics may be able to capitalize on rapidly evolving SNR markers to trace pathogen movement at fine spatial and temporal scales (Keim et al. 2004). Pathogen population diversity may also correlate with the frequency of infection and therefore could be used to better understand the burden of disease in a given ecosystem or management area. However, diversity in SNR types has yet to be confirmed across additional anthrax endemic areas, including those in which the disease is thought to be recently emerging or re-emerging.

In the United States anthrax is enzootic, emerging or re-emerging in a narrow band running from the upper Midwest south to western Texas (Blackburn et al. 2007). Although the exact route of introduction of *B. anthracis* into the United States is unclear, it was most likely a result of early colonists' activities including cattle trading, emptying of cargo after transatlantic voyages, industrial hide tanning, and bone meal production (Stein 1945). Reports of anthrax infection date back to the early 19th century (Stein 1945) and the infection was not uncommon in livestock and humans throughout the

United States until the adoption of vaccine for livestock in the 1950s (Stein and Van Ness 1955). In West Texas, which has a long history of anthrax, it was shown in Chapter 5 that sixteen isolates from a single large outbreak in white-tailed deer spanning a 3-week period had the same MLVA-25 genotype and four SNR sub-genotypes (SGTs). While the sample size in that study was insufficient to make conclusions about pathogen movement, it was shown that areas of highest diversity corresponded spatially and temporally with areas of high carcass density. Here, the SGTs of isolates derived from 10 anthrax outbreak sites, including additional isolates from multiple years from the site described in Chapter 5, were examined. These areas not only represent multiple SNP and MLVA-25 defined lineages, but also different host species and ecosystems. The goals were to describe the SNR diversity in each of the areas and determine whether SNR analysis will reveal diversity in these loci across a study sites and genetic lineages of *B. anthracis*. Confirming this will validate the consistent presence of SNR diversity in anthrax outbreaks and will justify efforts to collect genetic data for more rigorous and thorough analyses. Analysis of multiple years of the same site allows comparisons of genetic diversity across years to understand whether one genotype is consistently found at that site across years as well as to examine for longer term spatial patterns.

Methods

Study Sites

West Texas

The study area in West Texas is a sparsely populated region in which the major land-use is private ranching, with a major portion of landowner income generated by commercial deer and wildlife hunting. The region is made up of extensive arid

grasslands and stands of thorny shrubs. The area has a long history of anthrax (Blackburn et al. 2007), including annual sporadic cases and frequent larger outbreaks (Blackburn and Goodin 2013), and the pathogen population includes at least three lineages of *B. anthracis*: A1.a (Western North American, or WNA), A3b (Ames like) and A4 (Vollum like) based on SNP typing (Van Ert et al. 2007a). Study sites TX-1-9 are private ranches from which *B. anthracis* isolates from carcasses and carcass sites, geographic coordinates of carcass sites, and epidemiological data were available. The locations of and characteristics of TX 1-9 are described in Figure 6-1 and Table 6-1, respectively. Site TX-1 experienced a large multi-week epizootic in 2005 which is described in Chapter 5. Additional isolates from the site from 2004 and 2009 were analyzed in addition to isolates from large outbreaks in the other nearby sites. Samples from TX-8 included carcasses and water troughs during a 2009 outbreak and additional carcass sites that were sampled in 2010, one year after the animals died. The carcass locations were mapped in space and time in ArcMap 10 (Environmental Systems Research Institute 2011. ArcGIS Desktop: Release 10. Redlands, CA).

Montana

The study site in Montana (MT-1) is a 45,969 hectare privately owned ranch in Southwestern Montana. Consisting of a mixture of cropland and native dry rangeland for grazing, the property is managed as a working bison (Bison bison) fdgfproduction operation. There are also established wildlife populations on the ranch, including elk, mule deer, Shiras moose and pronghorn. A large anthrax outbreak affecting multiple host species occurred in the summer of 2008. Although there have been recent livestock outbreaks in eastern Montana (e.g. Summer 2005; Blackburn et al. 2007), the 2008 epizootic was the first known report in western Montana since the 1950s and the

first known outbreak from the immediate surrounding area. The outbreak included approximately 300 bison, over 30 mature elk, and two white tailed deer and was one of the first outbreaks of anthrax in free-ranging elk in the United States. *B anthracis* isolates from two bison carcasses were analyzed as well as multiple isolates from each of two elk carcasses. This outbreak was noteworthy in that exclusively bull elk were affected, and in the spatial differences in elk and bison carcasses. Bull elk carcasses were found primarily at higher elevations and on ridge tops, while the majority of bison cases occurred on lower elevation pastures. The outbreak had as an index case a bison from the managed population, and subsequently wild elk carcasses were found for which exact dates of death could not be determined.

Diagnostics and Genetic Analysis

Anthrax diagnostics

Anthrax was diagnosed using classical microbiological methods and polymerase chain reaction (PCR) as described in Blackburn et al. (2010b). Briefly, after heat shocking (30 minutes at 70^0C) or suspension in 100% ethanol samples were concentrated by centrifugation. The pellet was re-suspended in water and plated onto sheep blood agar and ACTSBA media. Gram staining was used for presumptive identification of *B. anthracis* and suspect colonies were subcultured.

Genetic analysis

MLVA-25 genotyping. The 5' fluorescent-labeled MLVA-25 primers described by Lista et al. (2006) were synthesized. Multiplex PCR reactions were carried out using 0.5 uM each of the VNTR specific forward and reverse Universal tail primers, 1.0 ng of DNA template, 2.0 mM $MgCl_2$, 2.5 mM of each dNTP, and 1.25 U Platinum Taq Polymerase per reaction. Thermal cycling parameters were as per Lista et al. (2006).

PCR products were diluted 1:125 in molecular grade water and 1.0 uL of the diluted multiplexes were mixed with 19.0 uL of a formamide/LIZ 1200 (Applied Biosystems, Foster City,CA, USA) size standard mixture and denatured.

SNR-4 genotyping. We individually amplified the four SNR loci described in Kenefic et al. (Kenefic et al. 2008b). The 10.0 µl PCR reactions were carried out with final concentrations of the following: 1.0 µL template DNA per reaction, 0.2 µM of each of the four forward and reverse primers, 1X PCR buffer, 0.5 U per reaction *pfu* Polymerase (Agilent technologies, Wilmington DE), 3 mM $MgCl_2^*$, and 0.25 mM of each dNTP. The PCR products were pooled with a final dilution of HM-1, 2, and 6 at 1:20 and HM-13 at 1:10; 1.0 uL of the pooled products were mixed with 19.0 uL of a formamide/LIZ 500 (Applied Biosystems) size standard mixture and denatured. Fragment sizing for MLVA-25 and SNR-4 was performed on an ABI 3730 (Applied Biosystems) and VNTR sizes were determined using GeneMapper™ software (Applied Biosystems).

Unweighted pair group method with arithmetic mean (UPGMA) cluster analysis was used to establish genetic relationships. Distance matrices were generated in PAUP 4.0 (Sinauer Associates, Inc., Sunderland, MA, USA) and imported into MEGA 3.1 (Kumar et al. 2004) for tree building.

Results

West Texas Study Area

A4 (Vollum)

All isolates from West Texas sites TX-1, TX-2, TX-3, TX-4, and TX-5 grouped into the classically defined A4 (Vollum like) cluster as described by Keim et al. (2000) nomenclature, and into genotype GT 57 using the Lista et al. (2006) terminology (Figure

6-2). The isolates were distinct from the pure Vollum strain. Thirty-three of 37 isolates from West Texas grouping in this lineage had identical MLVA-25 genotypes. Nine SNR sub-genotypes (SGTs) were identified among the 33 isolates in this group (Table 6-2, Figure 6-3) and were numbered continuing from the 2005 outbreak described in Chapter 5. SGT-1, which was the predominant SGT in that outbreak, was found exclusively in 2005 on that study site and was not identified in the 2004 or 2009 isolates from TX-1 or in neighboring sites. SGT-2, found in one isolate in 2005 on TX-1, was also identified in two isolates from 2004. Another isolate from the TX-1 2005 outbreak, SGT-4, also occurred in the site in 2009 and was identified in samples from three white-tailed deer in site TX-2. A novel SNR type, SGT-5, was isolated from one of the two 2009 TX-1 isolates. The distribution of the 2004, 2005 and 2009 carcass locations and genotypes from TX-1 are shown in Figure 6-4.

The remaining Vollum-like SGTs and associated sites are shown on Table 6-3. Thefghfgh SGT found in TX-5, in which domestic cattle, exotic (non-native) deer and white-tailed deer were affected, differed from the other SGTs in this MLVA-25 group at HM-6 and HM-13. The four isolates from the 2004 outbreak in TX-3, which differed on MLVA-25 from isolates at other sites at the pX0-1 allele, were the same on SNR analysis (Table 6-2). The SNR alleles were identical to those of SGT-6 from TX-2. The spatial distribution of carcasses and SGTs found in TX-2 is shown in Figure 6-5.

A3b (Ames) MLVA-25

Isolates from TX-6 and TX-7 were related to the A3.b (Ames) cluster as described by Keim et al. (2000), and to GT-53 as per the Lista et al. (2006) nomenclature. TX-6 isolates were most closely aligned with the original Ames strain, whereas the TX-7 isolates appear to be more closely related to the Sterne group

originally described in South Africa. Four SGTs were found in the eight isolates from the 2010 outbreak on TX-6 (Figure 6-6A; Table 6-2) and the spatial distribution of carcasses and SGTs from the site are shown in Figure 6-6B. The two isolates from TX-7 lacked the pX02 plasmid on MLVA-25 and therefore HM-1 and HM-13 were absent from the SNR analysis.

A1.a (WNA) MLVA-25

Isolates from TX-8 and 9 grouped into the A1.a (WNA) cluster as described by Keim et al. (2000). Within TX-8 there were two SGTs (Table 6-2). SGT-1 was identified in 3 animals sampled in the 2009 outbreak, and also from two water troughs sampled. A second SGT (SGT-2) was identified in one carcass sampled in 2009. Additional samples were obtained in 2010 from the soil around carcasses of animals which died in the 2009 outbreak. These isolates were also of the SGT-1 SNR type. MLVA-25 analysis of three *B. anthracis* isolates from TX-9 failed to yield several loci; however, the loci available were identical across all markers and consistent with an A1.a (WNA) strain. On SNR analysis all three isolates were identical, and different on two markers from the TX-8 SGT-1 (Table 6-2).

Montana A1.a (WNA) MLVA-25

Seven isolates were analyzed from the 2008 outbreak in MT-1. Three isolates were collected from one bull elk and 2 isolates from a second bull elk. One sample was available from each of two bison carcasses. All isolates fell into the Keim et al. (2000) defined A1.a cluster, and specifically the SNP defined WNA grouping. Two SGTs were found in each bull elk carcass: SGT-1 and SGT-2 in one, and SGT-1 and SGT-4 in the other. One sub-genotype, SGT-1, was isolated from both an elk and one of the two bison carcasses. SGT-3 was found only in the other bison (Figure 6-7; Table 6-2).

Discussion

The genetic lineages of *Bacillus anthracis* found in the United States are likely a result of multiple independent introductions which took place in the 19th century (Stein 1945, Van Ert et al. 2007a, Simonson et al. 2009). After being introduced into an area suitable for pathogen persistence, genetic drift or selection pressures in the environment or within the host, or a combination thereof, likely shaped the population genetics into the distribution found today. This resulting clonal population structure limits the use of SNP and MLVA markers for epidemiological analysis, but it is possible that more rapidly evolving SNR markers can provide more insight into local anthrax dynamics. The results of this analysis support previous findings of SNR diversity among *B. anthracis* isolates from naturally occurring outbreaks that are identical on SNP and MLVA typing. This diversity was found across the three *B. anthracis* lineages examined in the United States.

Anthrax outbreaks in Texas frequently involve multiple species within the spatial and temporal confines of the outbreak. Similarly, the 2008 outbreak in Western Montana involved at least three species (bison, elk and white tailed deer). In each multi-host outbreak analyzed here, SNR sub-genotypes (SGTs) were common to multiple host species. Specifically, all carcasses in TX-3, with a mixed cattle and white-tailed deer population, had the same SGT, and in TX-4, where cattle, exotic deer and white tailed deer were infected, all carcass sites yielded identical SGTs. Although the sample size of genotyped carcasses from the outbreak in Western Montana is small, one SGT was shared between both of the sampled elk and a bison, despite the spatial differentiation of carcass locations, as illustrated in Figure 6-7B. The sharing of identical SGTs between wildlife and livestock suggests a common source of infection or very

close epidemiological associations between the host populations. While this distinction needs to be addressed as part of the greater effort to understand anthrax ecology, it emphasizes the spillover potential of anthrax (Alexander et al. 2012) and the importance of both continued vaccination of livestock and monitoring of wildlife populations in close contact with livestock (Blackburn et al. 2007, Blackburn and Goodin 2013). It also points out a need to understand the movement patterns of both managed populations and wildlife, particularly with respect to patterns within the anthrax season.

Another notable aspect of the isolates from MT-1 was the high diversity within each of the two elk carcasses. In previous work, and in Chapter 5 of this work, multiple SGTs were isolated from the carcass and soil at the carcass site (Stratilo and Bader 2012), whereas here polymorphic markers were found in different parts of the carcass. These large (approximately ~400 kg) animals had been dead as long as three weeks, and it is possible that the large body volume and time between death and sporulation allowed for diversity to develop within the carcass, although it cannot be ruled out that the animal was circulating multiple SGTs before death. While more research and greater sampling is needed to confirm these results, this within carcass diversity challenges the *in vivo* mutation rates of SNRs. The MLVA-25 genotype of the Montana outbreak isolates was closely related to isolates from Minnesota, and, furthermore, a second MLVA-25 genotype was found in a bison for which DNA was not available for SNR typing, suggesting a more complex history of anthrax in this region than previously characterized. Because infection cycles are required to generate this level of diversity, it is likely that anthrax has been circulating in the area undetected.

This analysis expanded that of Chapter 5 by analyzing isolates from 2004 and 2009 sporadic cases on TX-1. The two samples from 2004 yielded SGT-2 which was found once in the 2005 outbreak, although one of the carcasses from which the 2004 isolates were obtained was located at some distance from where the SGT was isolated in 2005 (2.8 km) and in 2004 (2.7 km), yet close to a grouping of 2005 carcass sites from which only one isolate, SGT-1, was genotyped. The two 2009 carcasses yielded SGT-4, also isolated in the 2005 outbreak, and a novel SGT. The 2009 SGT-4 isolate was obtained from a carcass located 2.5 km from the 2005 SGT-4 location. It should be noted, however, that the 2009 SGT-4 isolate was derived from a carcass found very close to a carcass location in the 2005 outbreak from which a sample for genetic analysis was not available. It appears that focused areas on the landscape within TX-1 are supportive of sustained environmental transmission across years, and furthermore, these areas are associated with high diversity within these SNR markers. Curiously, the dominant SGT in the 2005 outbreak, SGT-1, was not isolated in other years or on nearby study sites. While this may be a result of sample size, it also indicates that SGT-1 could have been either selected for in the 2005 season or was an index genotype which then spread through the deer population through local transmission pathways.

Two other sites having large multi-week outbreaks of anthrax in white-tailed deer had similar levels of diversity as site TX-1 despite small sample sizes of isolates. TX-2 had three SGTs among six isolates, and TX-6 had four SGTs found among seven isolates. The significant differences between the A3.b (Ames-like) isolates from TX-6 and TX-7 are not surprising in the context of the large spatial difference between those

sites, which would have made possible divergent evolution after introduction of even identical strains. In these two white-tailed deer outbreaks, the spatial distribution of carcasses (Figures 6-5 and 6-6B) was similar to that seen in the large 2005 outbreak on TX-1 (Chapter 5). Given evidence of home ranges sizes in deer being similar across this region of Texas (Blackburn 2006), this lends additional support to the hypothesis of local transmission pathways and the importance of contact with carcass sites in anthrax ecology. Looking at diversity across sites, three A4 West Texas sites were located in close proximity to one another, namely, TX-1, TX-2 and TX-3. Subgenotype-4, which appeared in TX-1 in both 2005 and 2009, was present in half of the 2010 isolates from TX-2. The same sporadic strain from 2005 and 2009 in site TX-1 appearing as an epizootic strain in 2010 on TX-2 could have been due to convergent mutational processes, but we must also consider the possibility that the conditions in 2010 were optimum for SGT-4, or, similar to the above discussion of SGT-1, that local transmission mechanisms amplified this otherwise sporadic strain into an outbreak event. In Etosha National Park, Namibia, a much larger ecosystem, a few MLVA defined genotypes were shown to persist across outbreaks, although one outbreak year was dominated by a MLVA genotype otherwise found only sporadically (Beyer et al. 2012). Local transmission mechanisms, such as contact with infected carcasses or insect mediated transmission, could produce such a clonal outbreak, as could environmental selection.

The sampling of TX-8 provided a unique opportunity to evaluate environmental contamination during the outbreak and one year later. Spores isolated from the sediment of two water troughs yielded the same SGT as found in a cow during the outbreak. Vultures, *Cathartes aura*, are common scavengers of carcasses in this

ecosystem, and this species will commonly bathe in a nearby water body after feeding. Because no carcasses were found in these water troughs, and vultures were observed in the troughs (Blackburn, pers. comm.), it is likely that the source of this contamination was vultures depositing spores which had adhered to their feathers while feeding. It is not known if the spores in the troughs would constitute an effective source of infection for an herbivore, but this does indicate that avian scavengers could play a role in the movement of *B. anthracis* spores during an outbreak. Following the 2009 outbreak in TX-8 the affected pasture was fenced to prevent access by livestock the following season. In the summer of 2010 soil samples from 2009 carcass locations were positive for *B. anthracis* spores, and, furthermore, these were of the dominant SGT from the outbreak the previous year. This shows that the SNR genotype from an outbreak may again become available to animals in subsequent years.

One of the isolates from the 2010 outbreak on TX-2 had SNR allele sizes identical to those of the 2004 outbreak in TX-3, although the MLVA-25 genotype differed at the pXO1 locus. Whether this is represents a mutation in the VNTR and conservation of the SNR genotype, or convergent mutational processes creating identical SNR genotypes is unknown. It is possible that the in vivo mutation rates of certain VNTR alleles exceed that of the SNR alleles. This finding makes clear that SNRs must be interpreted not simply in the hierarchical context of SNP and MLVA genotypes (Keim et al. 2004), but also, and necessarily, in the context of spatial and temporal distribution.

The consistent finding of SNR diversity within outbreaks confirms the possibility of using these markers to better understand transmission pathways of *B. anthracis*. Furthermore, the application of SNR analysis here shows possible epidemiological links

between infections across species and across outbreak locations, and the spatial findings supports earlier findings implicating local transmission pathways. Here I have also presented evidence for possible movement of an SGT by avian scavengers and for sustained environmental contamination with an outbreak strain. More comprehensive and systematic sampling of carcass sites and the environment in both sporadic and outbreak seasons should yield important insights into anthrax transmission.

Table 6-1. Characteristics of study sites.

Site	Outbreak year	Host species	Estimated deaths	Anthrax history	MLVA-25 Cluster	Number of isolates
TX-1	2005, 2004, 2009	White-tailed deer	>48 (2005)	Enzootic	A4	15 (2005), 2(2004), 2(2009)
TX-2	2010	White-tailed deer, bison, exotic deer		Enzootic	A4	6
TX-3	2004	White-tailed deer, Domestic cattle		Enzootic	A4	4
TX-4	2010	White-tailed deer, Domestic cattle, exotic deer		Enzootic	A4	5
TX-5	2010	White-tailed deer		Enzootic	A4	1
TX-6	2010	White-tailed deer	51	Enzootic	A3.b	8
TX-7	2009	White-tailed deer		Enzootic	A3.b	2
TX-8	2009	Domestic cattle		Enzootic	A1.a	8
TX-9	2012	Sheep, cattle		Enzootic	A1.a	3
MT-1	2008	Elk, Bison, White tailed deer	>30 Elk, 300 Bison, 2 White-tailed deer	Emerging/re-emerging	A1.a	7

Table 6-2. Allele sizes of sub-genotypes (SGTs). STs are grouped by MLVA-25 genotype.

		Allele size (number base pairs at locus)				
MLVA-25	Subgenotype	HM1	HM2	HM6	HM13	No. Isolates
TX-1 A4	SGT-1	84	106	90	118	12
TX-1 A4	SGT-2	84	105	90	118	3
TX-1 A4	SGT-3	85	105	90	118	1
TX-1,2 A4	SGT-4	83	105	90	118	4
TX-1 A4	SGT-5	82	105	90	118	1
TX-2 A4	SGT-6	83	106	90	118	1
TX-2 A4	SGT-7	83	104	90	118	2
TX-4 A4	SGT-8	83	105	89	119	5
TX-5 A4	SGT-9	82	104	89	118	1
TX-3 A4	SGT-1	82	106	90	118	4
TX-6 A3.b	SGT-1	91	110	89	118	4
TX-6 A3.b	SGT-2	91	109	89	118	2
TX-6 A3.b	SGT-3	91	108	89	118	1
TX-6 A3.b	SGT-4	86	110	89	118	1
TX-7 A3.b	SGT-1	*	108	90	*	2
MT-A1.a	SGT-1	81	104	90	116	3
MT-A1.a	SGT-2	81	103	90	116	1
MT-A1.a	SGT-3	81	105	90	116	1
MT-A1.a	SGT-4	80	104	90	116	2
TX-8 A1.a	SGT-1	80	111	89	115	7
TX-8 A1.a	SGT-2	90	112	89	115	1
TX-9 A1.a	SGT-1	78	111	90	115	3

Table 6-3. Locations and host species of sub-genotypes from the A4 (Vollum) group in West Texas

MLVA	SNR Genotype	Site	Bovine	White tailed deer	Exotic Deer	soil	Total
A4	SGT-1	TX1		12		0	12
A4	SGT-2	TX1		2		1	3
A4	SGT-3	TX1		1			1
A4	SGT-4	TX1, TX2		5			5
A4	SGT-5	TX1		1			1
A4	SGT-6	TX2		1			1
A4	SGT-7	TX2		2			2
A4	SGT-8	TX4	3	1	1		5
A4	SGT-9	TX5		1			1
A4 TX-3	SGT-1	TX3	3	1			4

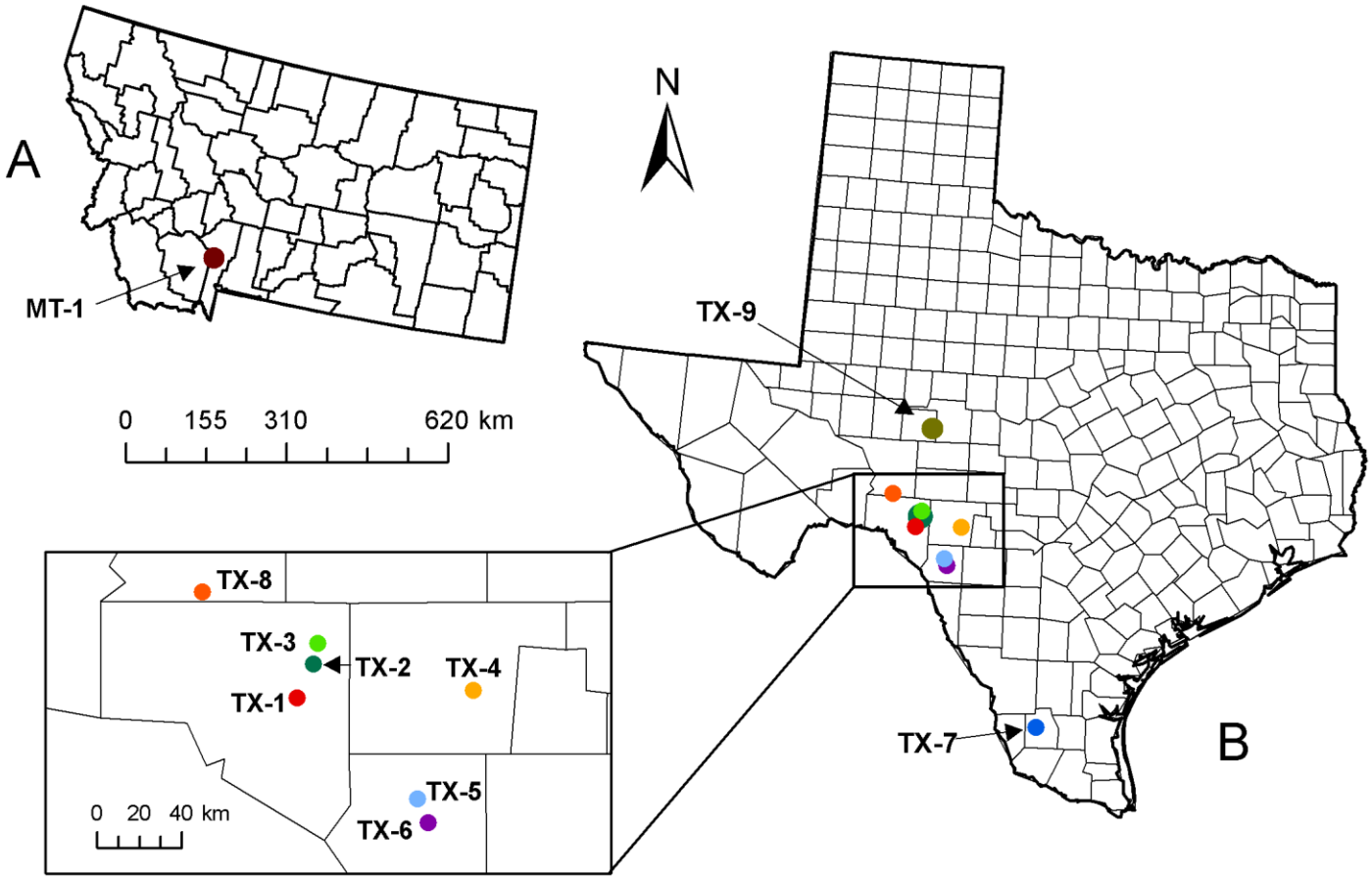

Figure 6-1. Locations of study sites. A) Montana and B) West Texas

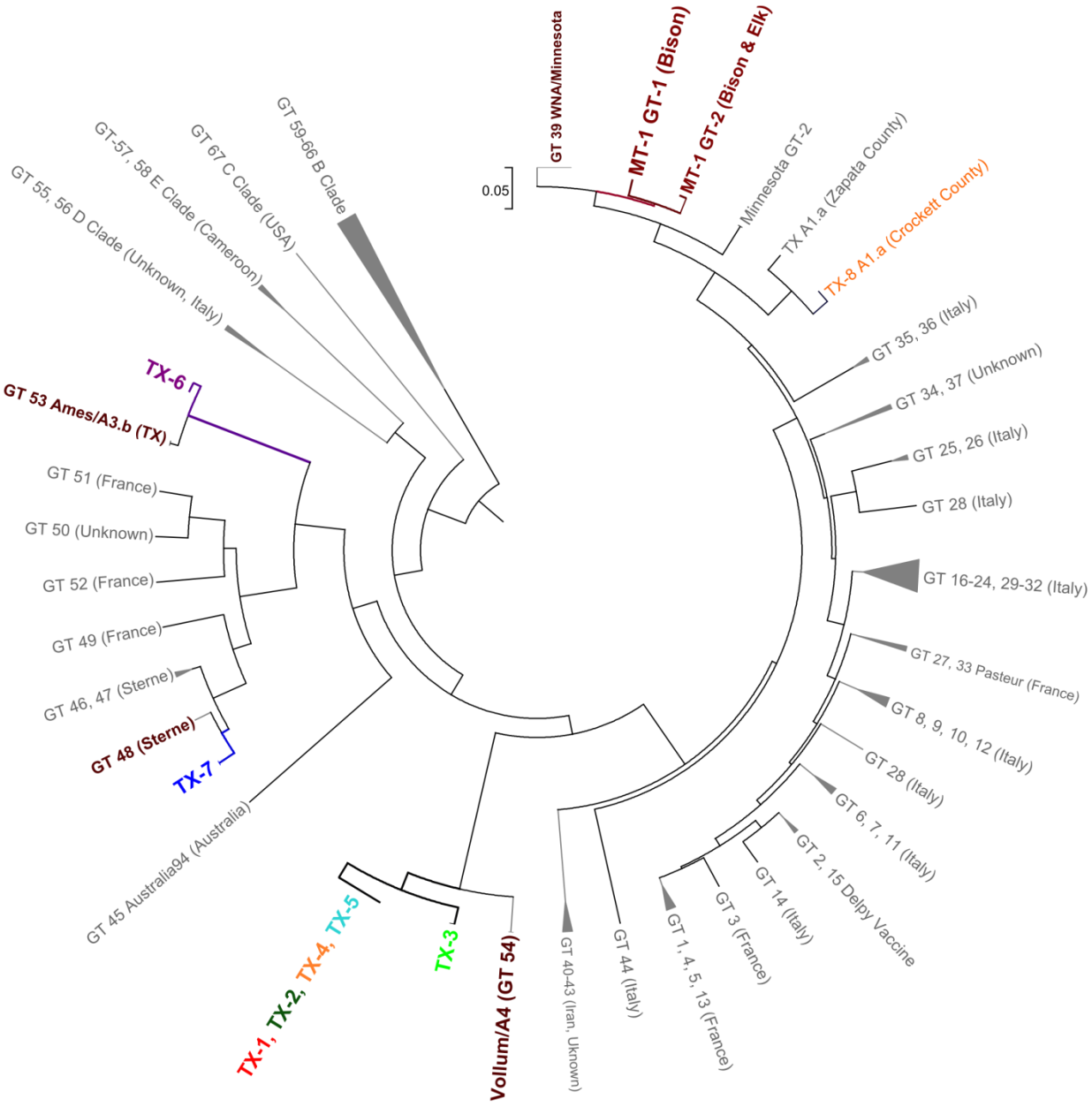

Figure 6-2. MLVA-25 phylogenetic analysis. Study sites are indicated in colors corresponding to Figure 6-1. Reference strains (Ames, Vollum, Sterne and A1.a WNA) are indicated in dark red text. Genotype designations (GT) as based on the Lista et al. (2006) nomenclature.

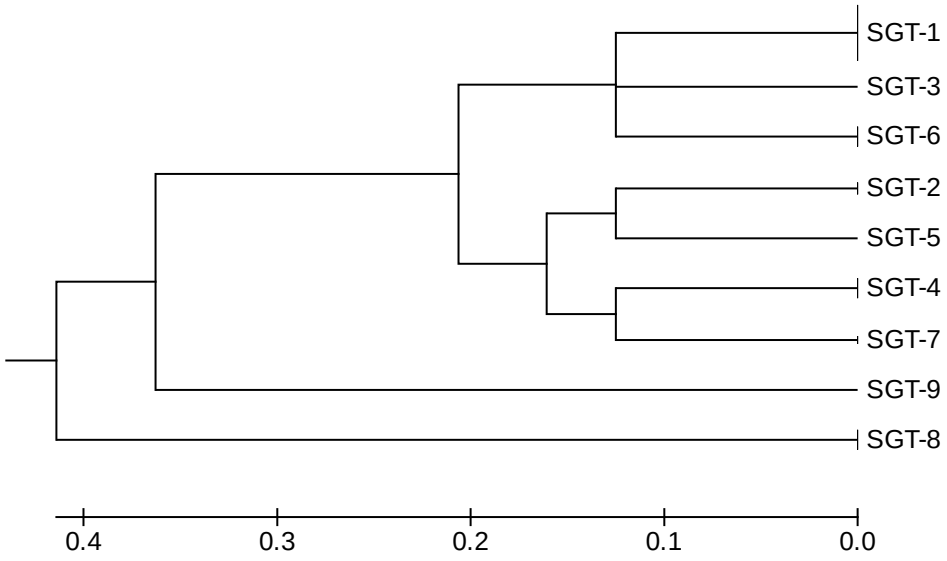

Figure 6-3. Compressed phylogenetic tree of subgenotypes (SGTs) from the A4 (Vollum-like) MLVA-25 group on sites TX-1,TX-2,TX-4 and TX-5.

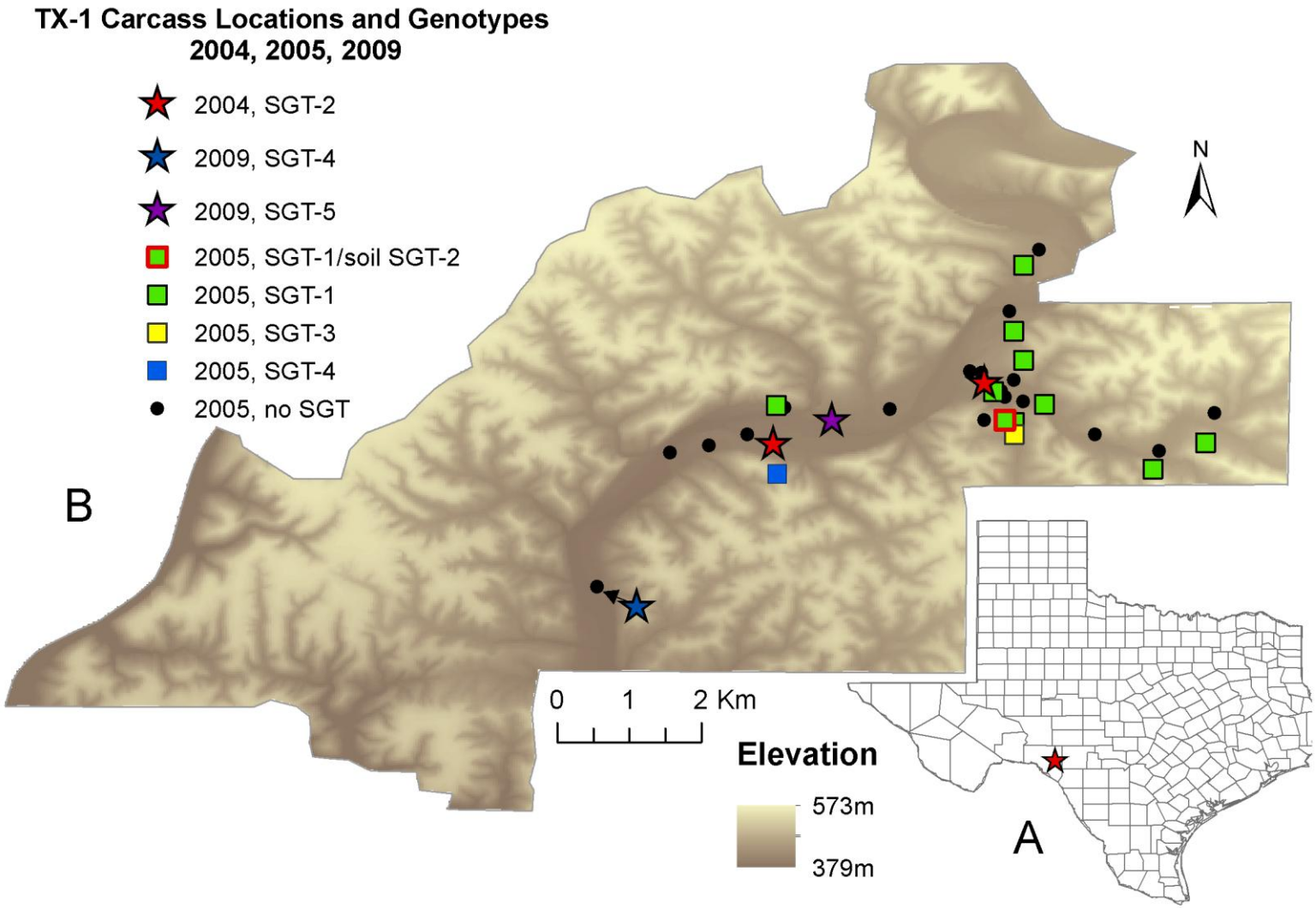

Figure 6-4. Locations of carcasses and subgenotypes in site TX-1 from 2004, 2005 and 2009. A) Location of study site TX-1. B) Locations of carcasses and subgenotypes from 2004, 2005 and 2009. The 2009 SGT-4 carcass, depicted with a blue star, overlaps with a 2005 carcass location from which no genetic material was obtained, and has been moved slightly for clarity

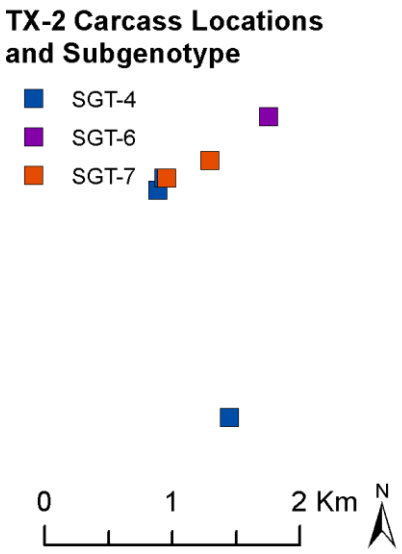

Figure 6-5. Locations of carcasses and subgenotypes (SGTs) in TX-2

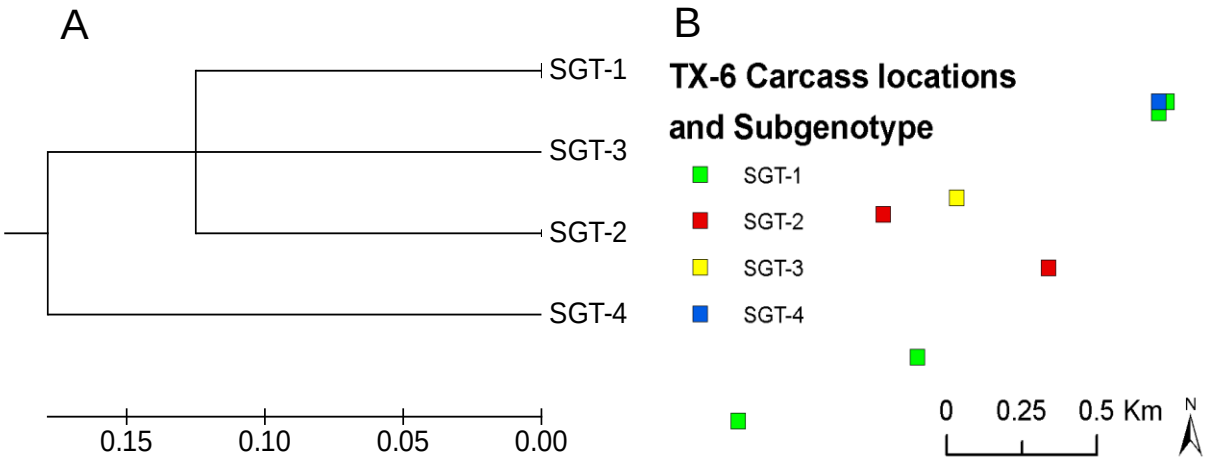

Figure 6-6. Phylogenetic tree, carcass and A3.b (Ames-like) subgenotype (SGT) locations in TX-6. A) Phylogenetic tree of relationships between SGTs B) Carcass and SGT locations.

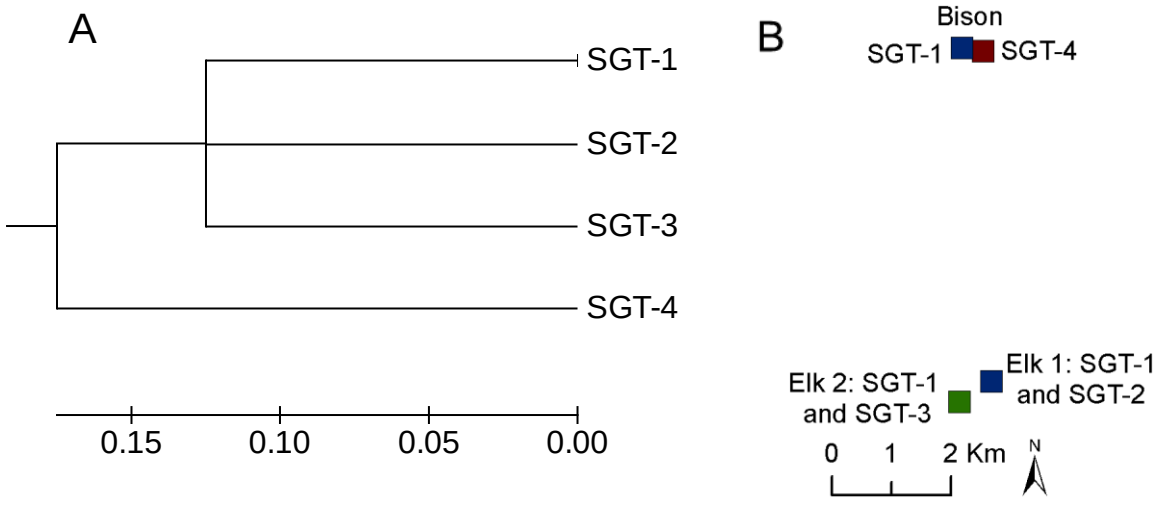

Figure 6-7. Phylogenetic tree, carcass and A1.a (WNA-like) subgenotype (SGT) locations in MT-1. A) Phylogenetic tree of relationships between SGTs B) Carcass and SGT locations.

CHAPTER 7
CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions and Contributions

This dissertation work set out to investigate the relationships between the genetic diversity and the ecology of *Bacillus anthracis*, the causative agent of anthrax. Specifically, the studies addressed previously unexplored questions about anthrax ecology and the role of population genetics in disease distribution and ecology as well as the utility of genetic information to inform our understanding of anthrax epizootiology and ecology. The work presented here applied theory from the literature to historical and emergent anthrax collections and considered how the results of these applications can be translated into strategies for disease control and prevention. As with many investigations of new ideas, the studies presented here pose more new questions than they answer, and the results provide a foundation for future efforts to link anthrax genetics and ecology.

In Chapter two I considered whether limiting the input to ecological niche models (ENM) to a single genetic grouping would allow more accurate and refined models to predict, geographically, where *B. anthracis* is likely to persist in Kazakhstan. The widespread A1.a lineage, which may be globally successful by virtue of an ability to tolerate a wider array of ecological conditions, appeared to occupy a wider ecological niche, and corresponding larger spatial distribution, than the larger dataset consisting of multiple genetic lineages. This was most true in the north of the country, and suggested that outbreaks in this part of Kazakhstan may be a result of a different genetic group. The implication of these results is that ecological niche models of B. anthracis are refined and better able to predict areas of pathogen persistence if built

with locally relevant isolates. It also supports evidence for genetic lineages of the pathogen having different ecological tolerances.

A major application of ENM to infectious disease is the ability to predict areas of disease risk in novel landscapes as a way of guiding surveillance and prevention efforts (Blackburn 2007). To that end, Chapter 3 tested whether collections of isolates of a single MLVA-8 defined lineage could predict the presence of the same lineage in other countries. The models failed to predict known anthrax occurrences in novel landscapes. Given evidence of finer scale genetic differentiation between these groupings I concluded that this level of genetic analysis was not appropriate for predicting *B. anthracis* persistence using transferred models. However, the methodological considerations that may limit transferability of ENMs are only beginning to be understood, and while a number of techniques were proposed in the literature to overcome these limitations they had not been tested systematically. Chapter 4 employs variable selection techniques to test their ability to improve the transferred models of *B. anthracis*. In that study I found that the best performing variable set was based on the similarity of variable values across all landscapes, suggesting that extrapolation into ecological space not part of the model building is a major limiting factor in transferred models.

While anthrax outbreaks have been analyzed spatially and the genetic diversity of large outbreaks characterized and mapped, no study had combined high resolution genetic analysis with spatial statistics. This combination should be a powerful method with which to explore unanswered questions about anthrax transmission dynamics. In Chapter 5, I found that areas of high transmission, as represented by a higher density of

cases in time and space, corresponded to areas of higher SNR diversity. In addition, the clustering of carcass locations at a distance within the estimated home range of neighboring cases suggests the importance of local transmission events in anthrax outbreaks. Chapter 6 describes the SNR diversity of 8 outbreak locations, including three major MLVA-25 genetic groups. This chapter confirms that SNR diversity within outbreaks is consistently present, and demonstrated that foci of anthrax persistence are present on the landscape. The sharing of SNR genotypes between domesticated host species and wildlife has critical management implications not only for anthrax, but for other pathogens shared between these species as well.

Future Directions

Ecological Niche Modeling and *Bacillus anthracis*

The work suggests important considerations for future niche modeling efforts for *B. anthracis*. First, given the diversity of anthrax isolates described here in North America, as well as diversity demonstrated in other parts of the world (Fouet et al. 2002, Lista et al. 2006, Garofolo et al. 2011) , regional models using coarsely defined lineages or multiple lineages may be unable to accurately predict pathogen persistence at relevant spatial scales. The ideal phylogenetic scale for generating *B. anthracis* niche models needs to be further explored, but is likely to be at the MLVA-25 level or comparable. This will require more extensive efforts to collect and genotype isolates from outbreaks in order to build datasets of adequate size for robust ENM building. Spatial scale is another consideration for ENMs to predict *B. anthracis* persistence. Anthrax outbreaks tend to be spatially limited, and in some areas, such as the study site of Chapter 5, focal areas of the landscape appear to have greater levels of pathogen persistence. Therefore, future modeling efforts should incorporate higher resolution

environmental data. As field evidence regarding the potential for B. anthracis to replicate in the environment grows, the addition of biotic factors to ecological niche models may be possible, and, furthermore, may be essential to fine scale predictions of pathogen persistence.

The spatial extent of ecological niche models for *B. anthracis* is an additional consideration that requires further consideration later work. There is considerable debate in the literature about the effects of extent on niche models, with some experts stating that the spatial extent of ENMs should be limited to areas in which the species is capable of dispersing (Soberon and Peterson 2005, Soberón and Nakamura 2009). The unanswered questions about anthrax transmission pathways in natural outbreaks and epizootics, combined with the potential for the pathogen to be carried long distances in feed and animal products, makes *B. anthracis* difficult to fit into a conventional framework for considering spatial extent. The choice of spatial extent should be explicitly discussed in the context of the question being asked. To understand where on a local landscape, such as a ranch or management area, *B. anthracis* may persist, a narrow spatial extent is reasonable and preferable. However, in efforts to understand broader trends, or areas of potential persistence, larger spatial extents are necessary. In that context, administrative boundaries, which to some extent limit dispersal and which also influence reporting, are defensible. A greater understanding of local and longer range transmission pathways is necessary to guide the choice of spatial extent in ENM studies of anthrax, although ultimately the study question will guide this choice.

The ENMs built in this dissertation used historical datasets spanning years to decades. In addition, although measures of annual fluctuation are used, climate data is temporally averaged. Therefore these ENMs cannot detect extreme weather events or seasonal climatic signatures associated with outbreaks. However, these historically based ENMs identify areas have value as are a foundation for future studies and as a guide for public health measures. Areas of high likelihood of pathogen persistence should guide intensive host and environmental sampling efforts for use in modeling fine scale anthrax spatio-temporal dynamics. Finally, surveillance is costly. Anthrax enzootic areas are often large and remote, and in many parts of the world occur where resources for surveillance are scarce. Areas identified as having a high likelihood of anthrax persistence as defined by multiple decades of occurrences can be used to identify areas for passive surveillance, and in conjunction with epidemiological data focus active surveillance and vaccination distribution (Kracalik et al. 2011).

High Resolution Genotyping and Anthrax Ecology

Chapters 5 and 6 explored the use of high resolution genotyping as an aid to understanding anthrax outbreak dynamics. A major limitation of these studies was the small sample sizes available for genotyping. Studies of other pathogens, including *Francisella tularensis* (Farlow et al. 2005, Nakazawa et al. 2010) and *Yersinia Pestis* (Girard et al. 2004, Vogler et al. 2011b), have clearly demonstrated the importance of genetic analysis *in* characterizing disease ecology. The work presented here suggest interesting possibilities and suggest hypotheses about transmission pathways within and between anthrax outbreaks, but my ability to make conclusions was limited by the relatively few cases with complete epidemiological and genetic information. The interesting patterns seen in the work here strongly emphasize the need for more

complete data collection and comprehensive genotyping efforts from outbreaks in wildlife and domestic animal populations. Moreover, the shared genotypes in wild animals and livestock underscore the importance of characterizing disease dynamics where those populations interact.

Increased collection of data will also allow for explorations of associations between host, pathogen and environmental conditions in a landscape genetics framework. The finding of a dominant genotype on TX-1 in 2005 which was not seen elsewhere in time or space, begs the question of whether climatic or host conditions selected for that strain on that landscape, whereas in 2010 SGT-4 was the dominant strain in a nearby area. These patterns have been reported elsewhere at different spatial (Garofolo et al. 2010) and temporal and genetic (Beyer et al. 2012) scales and warrant further investigation. A landscape genetics approach, which incorporates individual level genetic data to model gene flow and spatial patterns can be used with SNR data to explore ecological determinates of the genetic patterns suggested by this work. In addition, linear modeling approaches incorporating genetic, ecological and host characteristics can explore determinates of anthrax outbreaks. Future research will also shed light on the factors resulting in anthrax foci on the landscape.

The diversity described in Chapters 5 and 6 also illustrate the gaps in knowledge about the in vivo mutation rates of the genetic markers used for these analyses. Without this information we cannot estimate accurately whether two identical SNR types, for example, are epidemiologically linked or the result of convergent mutational processes. The effects of selection within the host, as well as in the environment, are

an equally important gap in understanding. Gaining this knowledge will be critical for fully utilizing high resolution genetic analyses to characterize anthrax ecology.

While this dissertation has posed more questions than it set out to address, it makes a clear case that genetic data is a key player in anthrax ecology at multiple scales. Future studies should incorporate systematic sampling of outbreaks to provide greater power to spatial and genetic analyses. These efforts will involve multiple disciplines, but should also actively involve the stakeholders – the public health workers, ranchers, small livestock owners and veterinarians who are most affected by anthrax around the world. The results of this work support the value of using of statistical approaches to link genetics of *B. anthracis* and ecological processes.

APPENDIX A
CHAPTER 2 SUPPLEMENTAL FIGURES

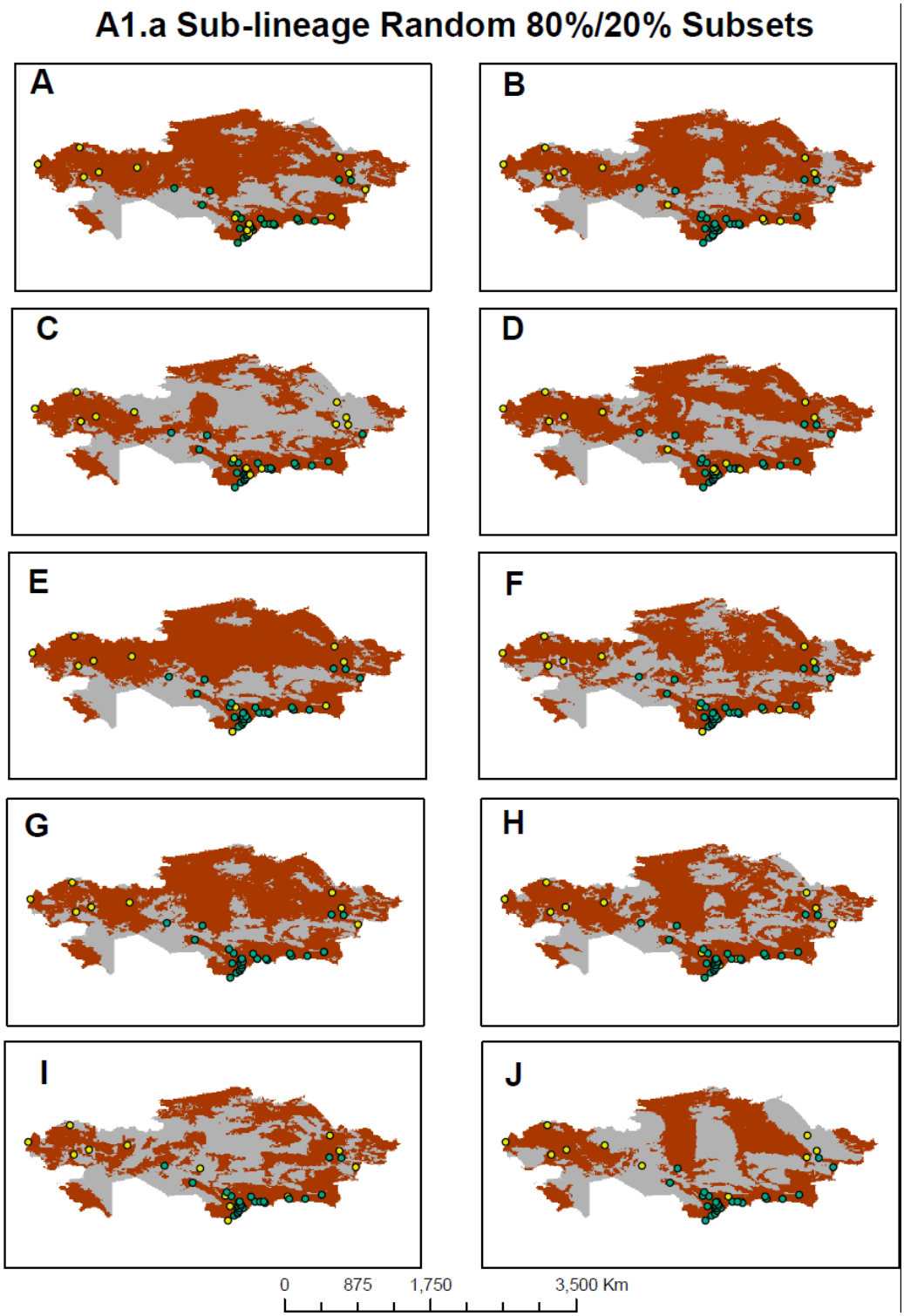

Figure A-1. A1.a sublineage random subsets. A through J are geographic predictions of models built with one of 10 random training and testing subsets. Gray areas are predicted by 0 to 5 models, red areas by 6 to 10 models.

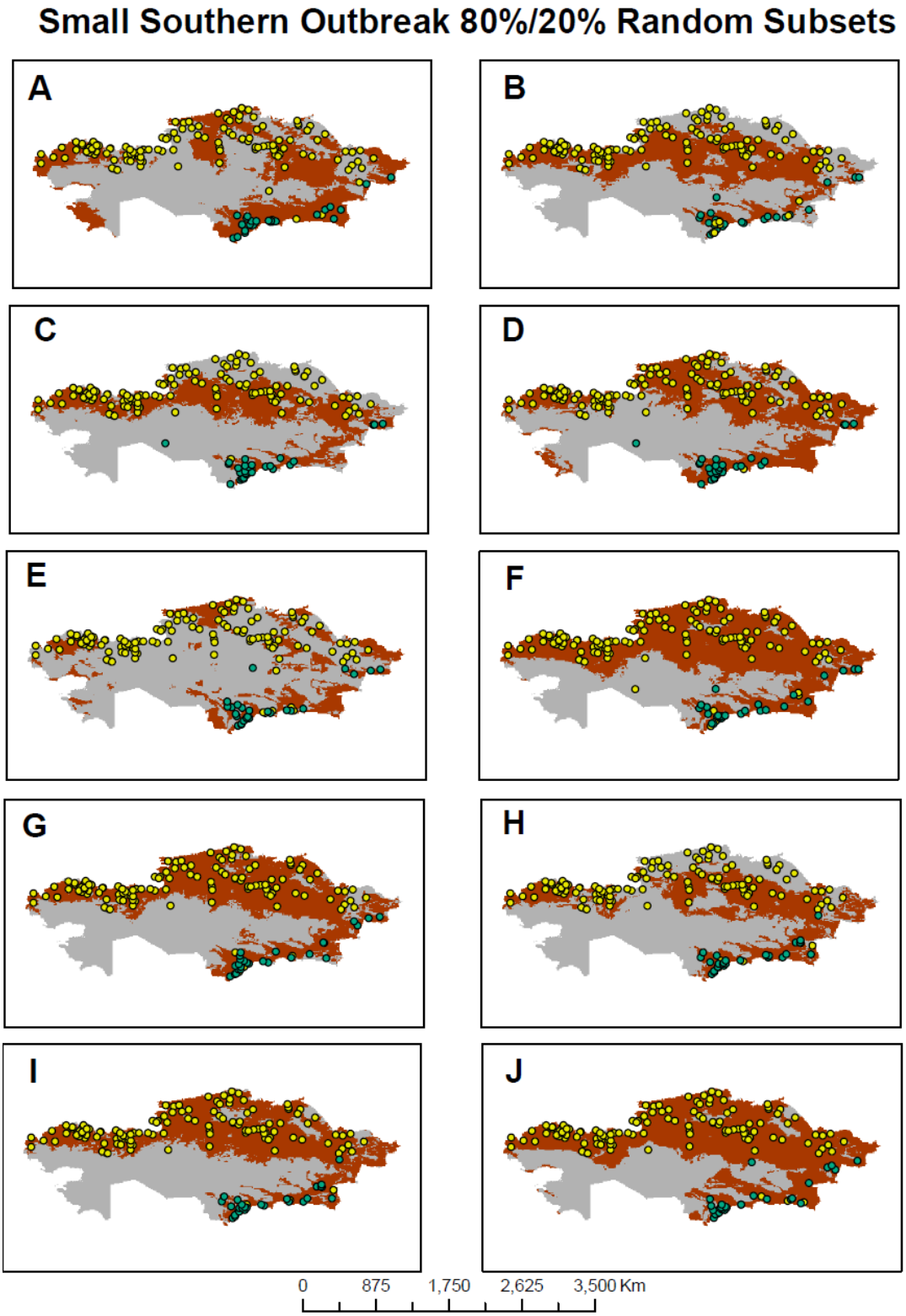

Figure A-2. Small southern outbreak random subsets. A through J are geographic predictions of models built with one of 10 random training and testing subsets.Gray areas are predicted by 0 to 5 models, red areas by 6 to 10 models.

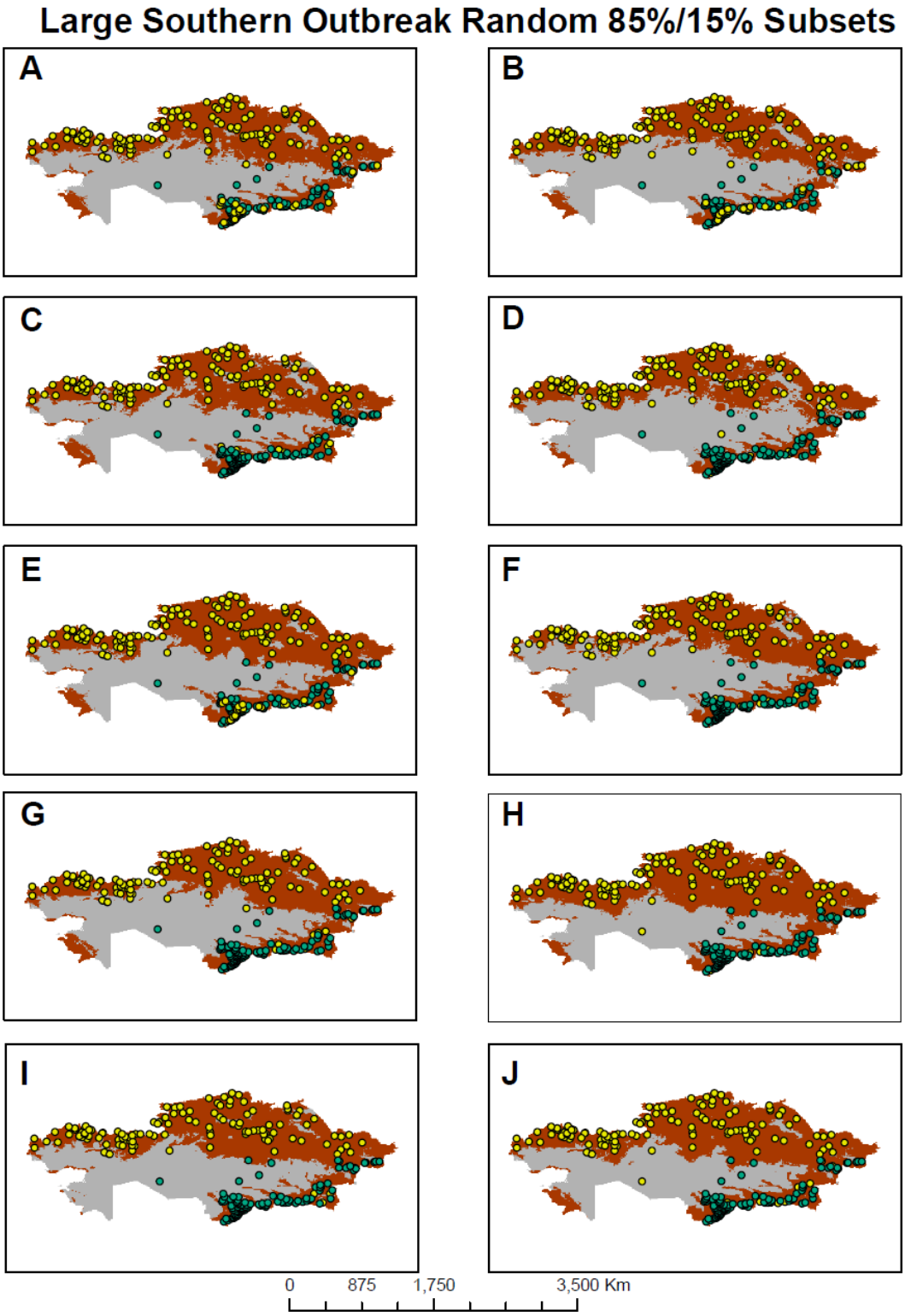

Figure A-3. Larger southern outbreak random subsets. A through J are geographic predictions of models built with one of 10 random training and testing subsets.Gray areas are predicted by 0 to 5 models, red areas by 6 to 10 models.

Table A-1. Accuracy metrics for A1.a sub-lineage experiments using random 80%/20% subsets (N to build models=26; N to test models=13). Model designations correspond to figures.

	Model									
Metric	A	B	C	D	E	F	G	H	I	J
Total Omission	0	0	0	0	0	0	0	0	0	8.3
Average Omission	14.6	15.4	28.3	10.8	7.5	17.5	17.7	21.7	32.5	19.7
Total Commission	22.58	18.69	11.38	24.06	36.26	10.17	10.43	11.73	12.93	15.99
Average Commission	65.2	63.77	48.23	63.88	71.93	59.93	67.49	59.74	56.99	58.23
AUC	0.6584	0.7034	0.6523	0.7527	0.6727	0.7364	0.6862	0.6881	0.5949	0.6255
SE	0.0829	0.0814	0.0864	0.0784	0.0859	0.0828	0.0821	0.0854	0.0866	0.0866
z-score	4.2181	4.4028	3.742	4.866	4.4394	4.3221	4.3662	4.004	3.3957	3.7671

Table A-2. Accuracy metrics for small southern outbreak experiments using random 80%/20% subsets (N to build models=26; N to test models=147). Model designations correspond to figures.

	Model									
Metric	A	B	C	D	E	F	G	H	I	J
Total Omission	0	0.08	9.4	0	35.9	3.1	3.1	0.8	0.8	1.6
Average Omission	40.3	34	31.8	29.5	67	14.6	19	39	20.3	16.8
Total Commission	16.08	6.48	13.06	10.32	6.13	31.77	16.95	6.13	15.99	28.5
Average	46.02	43.09	36.39	50.29	29.08	61.66	51.31	39.37	48.67	59.05
AUC	0.594	0.6789	0.7154	0.6675	0.5117	0.6323	0.7148	0.6855	0.7003	0.6483
SE	0.026	0.0264	0.027	0.0264	0.0258	0.0266	0.0259	0.0268	0.026	0.0264
z-score	12.38	14.343	14.756	13.918	12.218	14.450	15.126	13.952	15.092	14.612

Table A-3. Accuracy metrics for large southern outbreak experiments using random 85%/15% subsets (N to build models=113; N to test models=145). Model designations correspond to figures.

	Model									
Metric	A	B	C	D	E	F	G	H	I	J
Total Omission	0	0.7	0	0.7	0	0	0	0.7	0	1.4
Average Omission	20.5	16.5	22.3	21.5	9.6	18.2	16.1	5.1	15.6	16.2
Total Commission	11.7	12.47	12.25	7.49	30.7	17.28	16.2	36.28	15.53	19.59
Average	53.3	50.53	51.51	48.21	60.54	53.14	55.62	63.38	54.05	57.1
AUC	0.7036	0.7361	0.6902	0.7371	0.718	0.7027	0.7138	0.7171	0.7008	0.7
SE	0.0247	0.024	0.0249	0.0241	0.0245	0.0245	0.0245	0.0245	0.0246	0.024
z-score	15.713	16.835	15.313	16.413	17.126	15.953	16.232	18.363	16.136	16.04

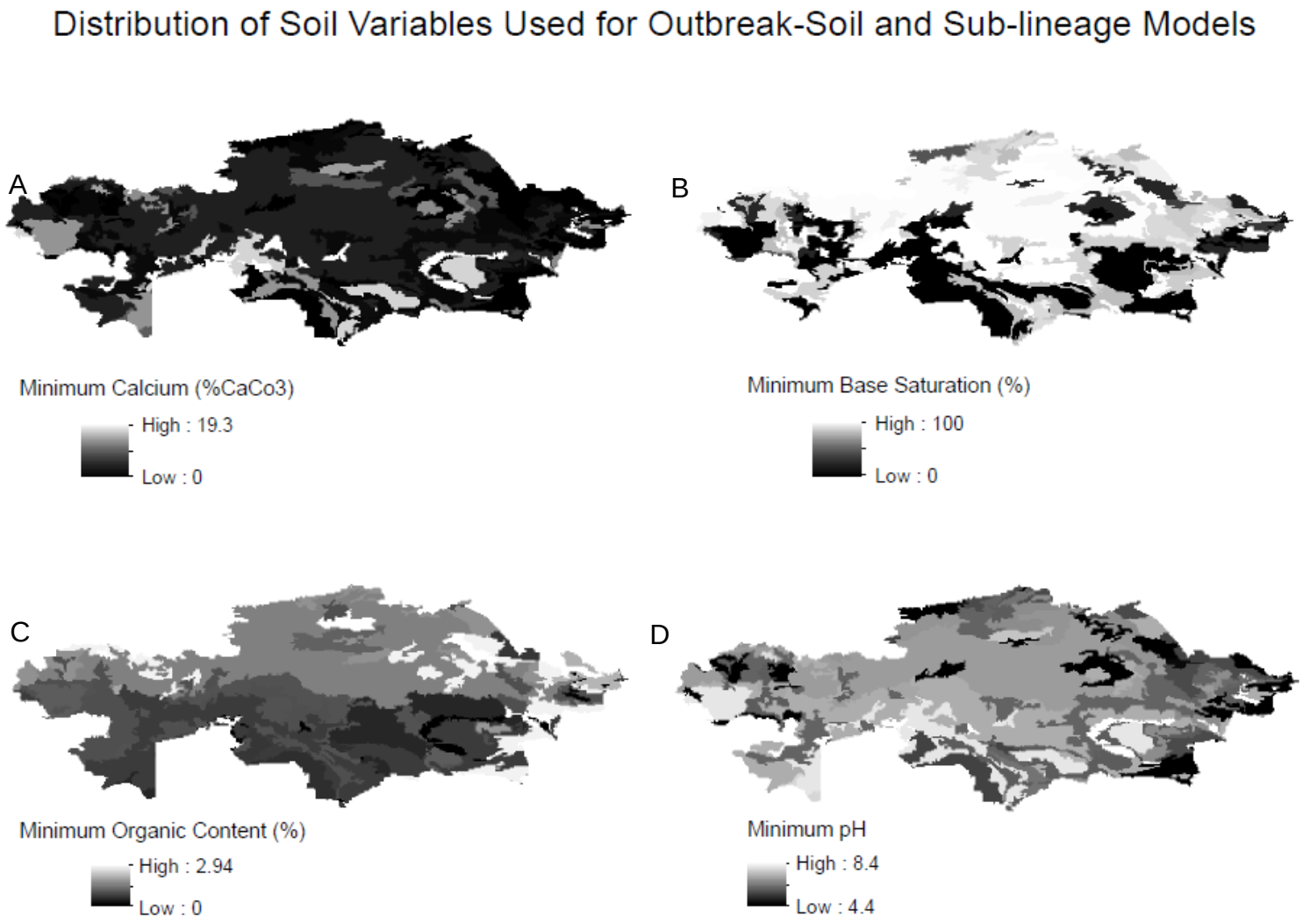

Figure A-4. Mapped values of the four soil variables. A) minimum soil calcium, B) minimum base saturation C) minimum soil organic content and D) minimum soil pH.

APPENDIX B
CHAPTER 4 SUPPLEMENTAL FIGURES

In all figures, the predicted distribution of *Bacillus anthracis* in shown in native landscapes in the left hand column and model predictions on novel landscapes are shown in middle and right columns. The color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

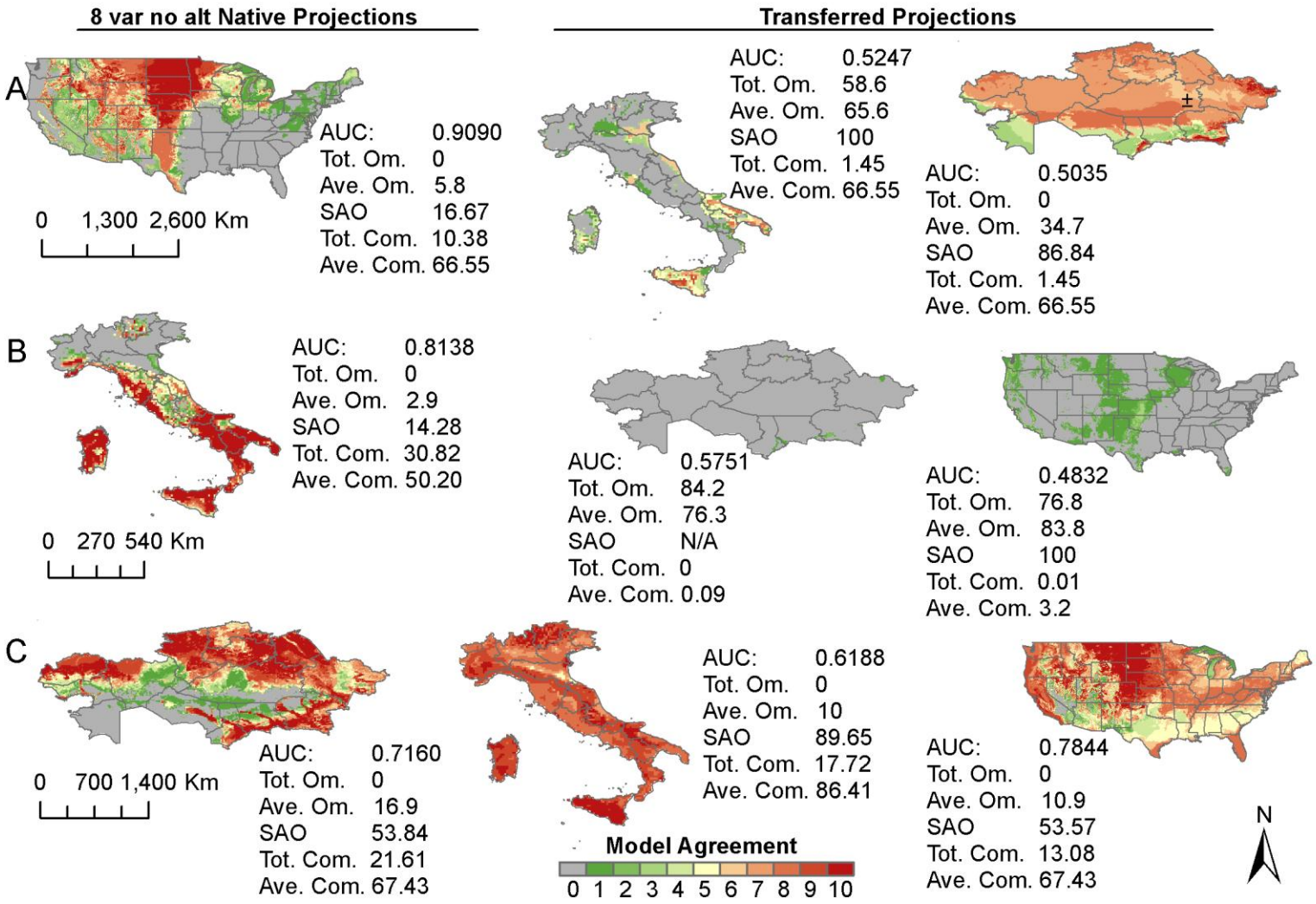

Figure B-1. Results of the 8 variable no altitude experiment. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

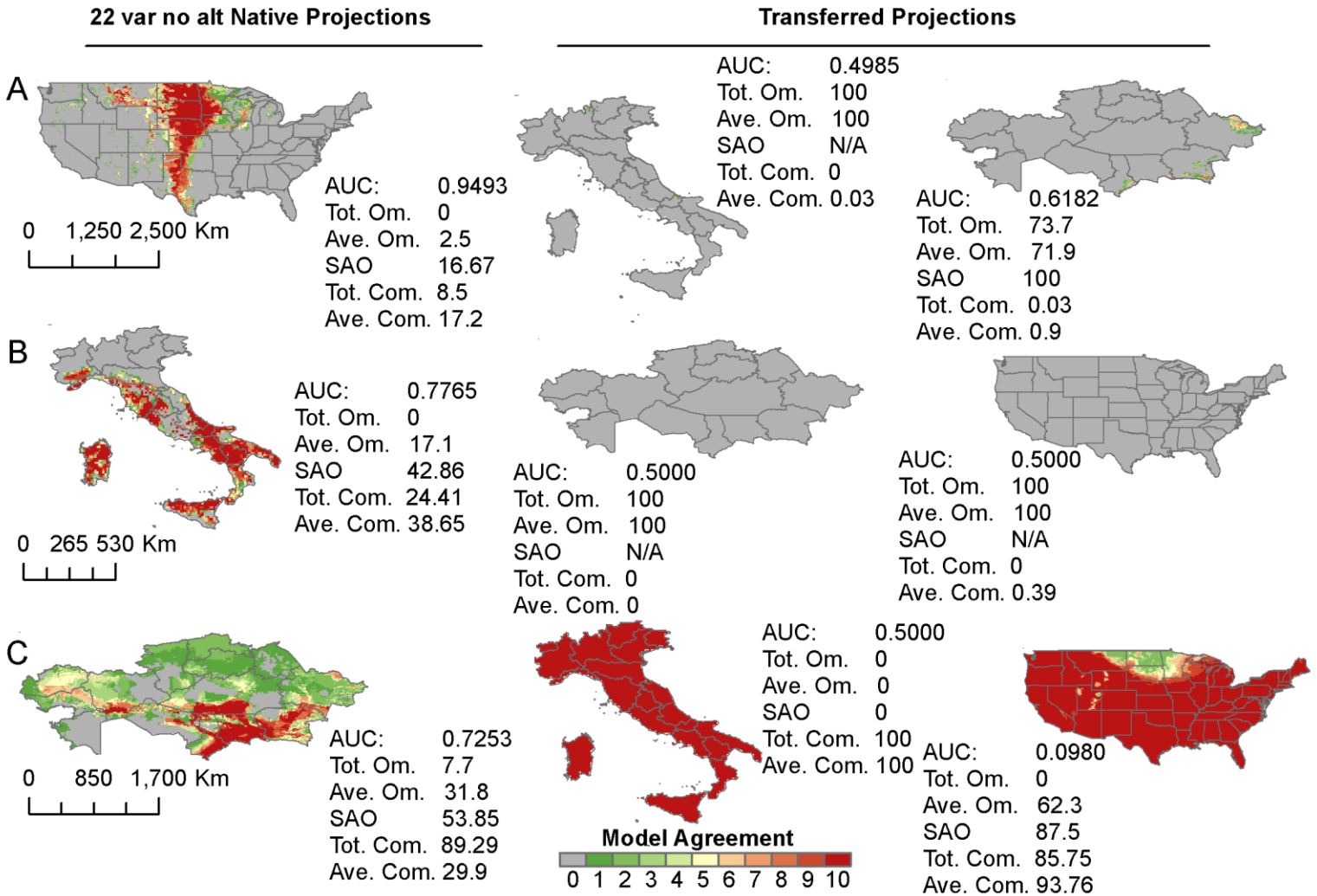

Figure B-2. Results of the 22 variable no altitude experiment. Predicted distribution of *Bacillus anthracis* by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

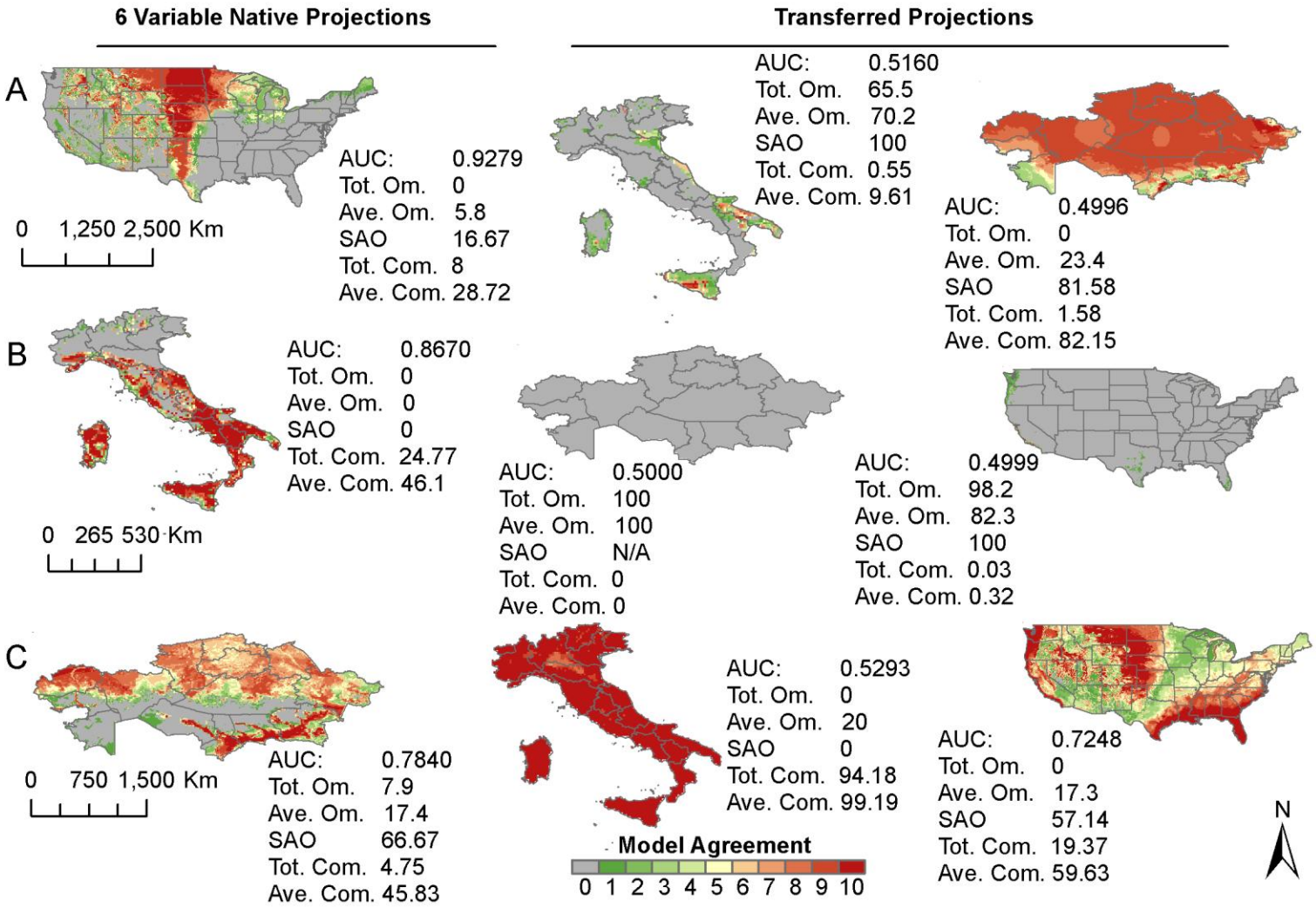

Figure B-3. Results of the 6 variable experiment. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

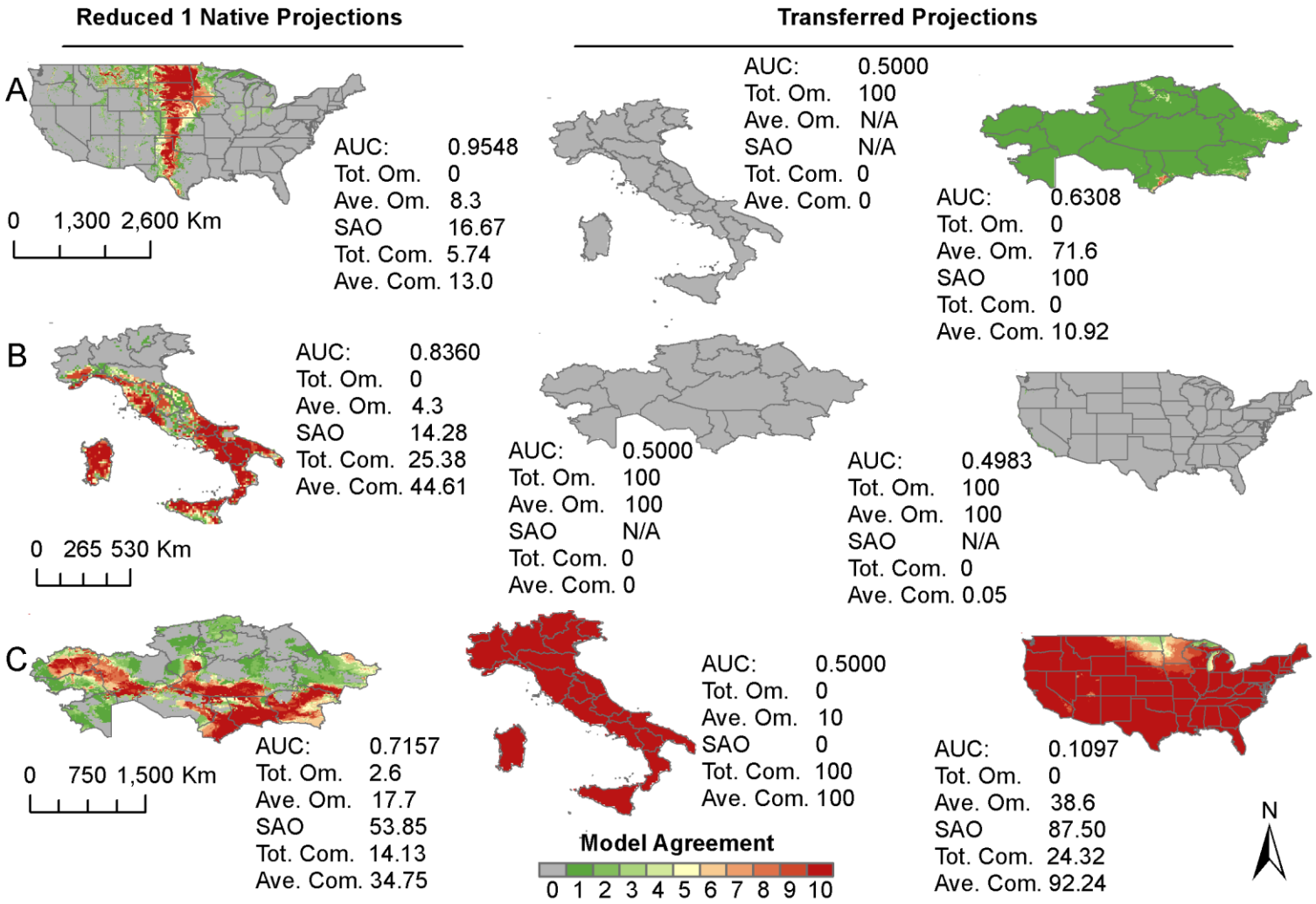

Figure B-4. Results of the reduced 1 experiment. Predicted distribution of Bacillus anthracis by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

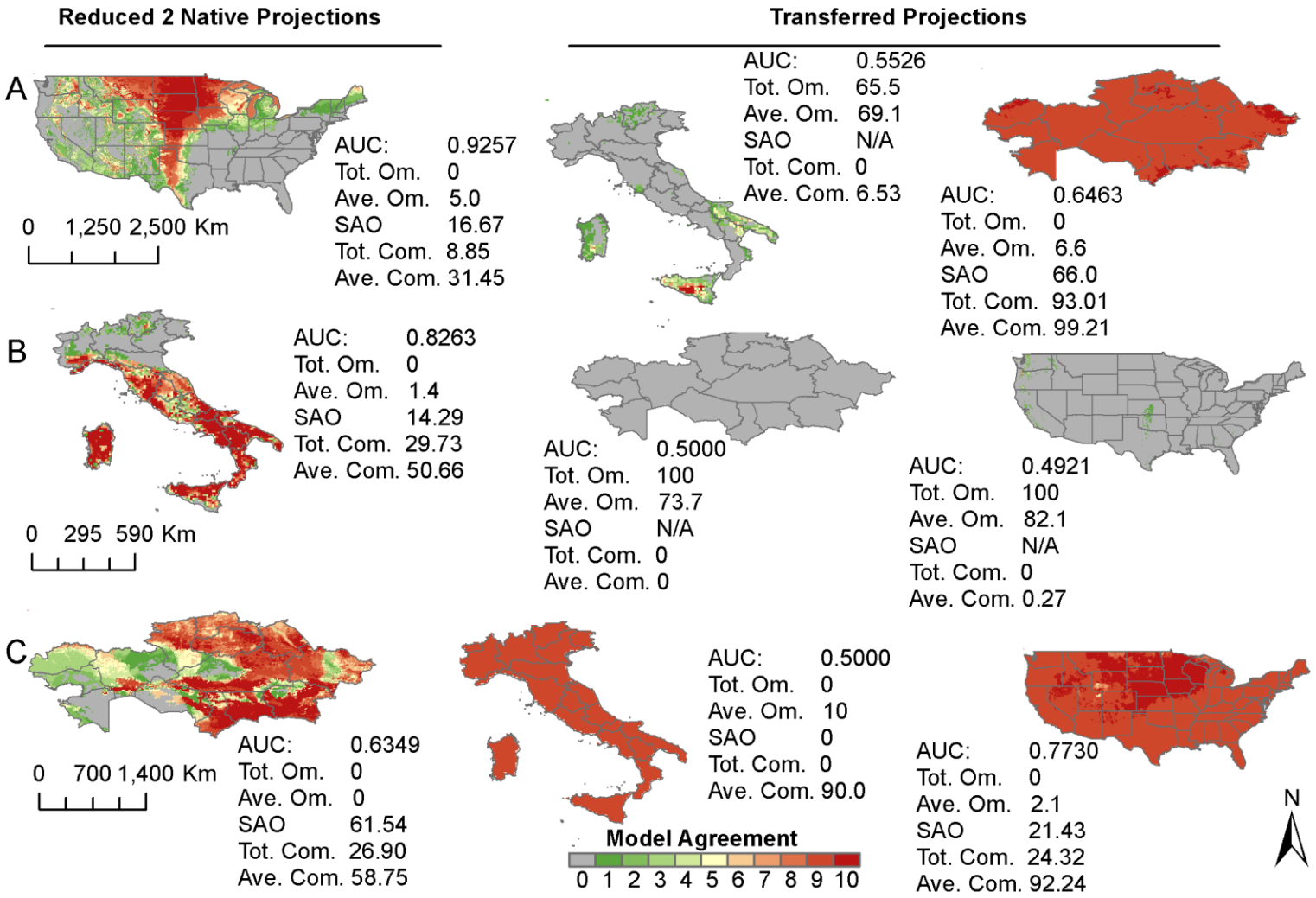

Figure B-5. Results of the reduced 2 experiment. Predicted distribution of *Bacillus anthracis* by native and transferred projections of models built in A) the United States, B) Italy and C) Kazakhstan. Color ramp indicates the level of model agreement from zero (no models predict presence) to ten (all models in the best subset predict presence).

LIST OF REFERENCES

Achtman, M. 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. 62:53-70.

Adjemian, J. C. Z., E. H. Girvetz, L. Beckett, and J. E. Foley. 2006. Analysis of Genetic Algorithm for Rule-Set Production (GARP) modeling approach for predicting distributions of fleas implicated as vectors of plague, Yersinia pestis, in California. Journal of Medical Entomology 43:93-103.

Aikembayev, A. M., L. Lukhnova, G. Temiraliyeva, T. Meka-Mechenko, Y. Pazylov, S. Zakaryan, G. Denissov, W. R. Easterday, M. N. Van Ert, P. Keim, S. C. Francesconi, J. K. Blackburn, M. Hugh-Jones, and T. Hadfield. 2010. Historical distribution and molecular diversity of Bacillus anthracis, Kazakhstan. Emerg Infect Dis 16:789-796.

Akbayram, S., M. Doğan, C. Akgün, E. Peker, M. Selçuk Bektaş, A. Kaya, H. Çaksen, and A. F. Öner. 2010. Clinical findings in children with cutaneous anthrax in eastern Turkey. Pediatric dermatology 27:600-606.

Al Dahouk, S., P. L. Flèche, K. Nöckler, I. Jacques, M. Grayon, H. C. Scholz, H. Tomaso, G. Vergnaud, and H. Neubauer. 2007. Evaluation of< i> Brucella</i> MLVA typing for human brucellosis. Journal of microbiological methods 69:137-145.

Alexander, K. A., B. L. Lewis, M. Marathe, S. Eubank, and J. K. Blackburn. 2012. Modeling of Wildlife-Associated Zoonoses: Applications and Caveats. Vector-Borne and Zoonotic Diseases 12:1005-1018.

Anaraki, S., S. Addiman, G. Nixon, D. Krahe, R. Ghosh, T. Brooks, G. Lloyd, R. SPENCE, A. Walsh, and B. MCCTOSKEY. 2008. Investigations and control measures following a case of inhalation anthrax in East London in a drum maker and drummer, October 2008. Euro surveillance 13:733-735.

Andersen, G. L., J. M. Simchock, and K. H. Wilson. 1996. Identification of a region of genetic variability among Bacillus anthracis strains and related species. Journal of bacteriology 178:377-384.

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling 162:211-232.

Anderson, R. P. and A. Raza. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. Journal of Biogeography 37:1378-1393.

Antwerpen, M., D. Ilin, E. Georgieva, H. Meyer, E. Savov, and D. Frangoulidis. 2011. MLVA and SNP analysis identified a unique genetic cluster in Bulgarian Bacillus anthracis strains. European Journal of Clinical Microbiology & Infectious Diseases 30:923-930.

Araujo, M. B. and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography 33:1677-1688.

Archie, E. A., G. Luikart, and V. O. Ezenwa. 2009. Infecting epidemiology with genetics: a new frontier in disease ecology. Trends in Ecology & Evolution 24:21-30.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological Modelling 200:1-19.

Avise, J. C. 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology 7:371-379.

Bales, M. E., A. L. Dannenberg, P. S. Brachman, A. F. Kaufmann, P. C. Klatsky, and D. A. Ashford. 2002. Epidemiologic Responses to Anthrax Outbreaks: A Review of Field Investigations, 1950–2001. Emerging Infectious Diseases 8:1163.

Bardell, D. 2002. Bacillus anthracis (Anthrax) has a unique place in the history of microbiology. Bios 73:133-136.

Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling 222:1810-1819.

Bellan, S. E., O. Gimenez, R. Choquet, and W. M. Getz. 2013a. A hierarchical distance sampling approach to estimating mortality rates from opportunistic carcass surveillance data. Methods in Ecology and Evolution.

Bellan, S. E., P. C. Turnbull, W. Beyer, and W. M. Getz. 2013b. Effects of experimental exclusion of scavengers from anthrax-infected herbivore carcasses on Bacillus anthracis sporulation, survival and distribution. Applied and environmental microbiology.

Berrang-Ford, L., O. Berke, L. Abdelrahman, D. Waltner-Toews, and J. McDermott. 2006. Spatial analysis of sleeping sickness, southeastern Uganda, 1970–2003. Emerg Infect Dis 12:813-820.

Beyer, W., S. Bellan, G. Eberle, H. H. Ganz, W. M. Getz, R. Haumacher, K. A. Hilss, W. Kilian, J. Lazak, and W. C. Turner. 2012. Distribution and molecular evolution of Bacillus anthracis genotypes in Namibia. PLoS Neglected Tropical Diseases 6:e1534.

Beyer, W. and P. Turnbull. 2009. Anthrax in animals. Molecular Aspects of Medicine 30:481-489.

Biek, R., J. C. Henderson, L. A. Waller, C. E. Rupprecht, and L. A. Real. 2007. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. Proceedings of the National Academy of Sciences 104:7993.

Blackburn, J. 2010. Integrating Geographic Information Systems and Ecological Niche Modeling into Disease Ecology: A Case Study of Bacillus anthracis in the United States and Mexico. Pages 59-88 *in* K. O'Connel, et. al. , editor. Emerging and Endemic Pathogens. Springer Science.

Blackburn, J. K. 2006. Evaluating the spatial ecology of anthrax in North America: Examining epidemiological components across multiple geographic scales using a GIS-based approach. Louisiana State University.

Blackburn, J. K., A. Curtis, T. L. Hadfield, B. O'Shea, M. A. Mitchell, and M. E. Hugh-Jones. 2010a. Confirmation of Bacillus anthracis from flesh-eating flies collected during a West Texas anthrax season. J Wildl Dis 46:918-922.

Blackburn, J. K., A. Curtis, T. L. Hadfield, B. O'Shea, M. A. Mitchell, and M. E. Hugh-Jones. 2010b. Confirmation of Bacillus anthracis from flesh-eating flies collected during a West Texas anthrax season. Journal of wildlife diseases 46:918-922.

Blackburn, J. K. and D. G. Goodin. 2013. Differentiation of Springtime Vegetation Indices Associated with Summer Anthrax Epizootics in West Texas, USA Deer. Journal of wildlife diseases 49.

Blackburn, J. K., K. M. McNyset, A. Curtis, and M. E. Hugh-Jones. 2007. Modeling the geographic distribution of Bacillus anthracis, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. American Journal of Tropical Medicine and Hygiene 77:1103-1110.

Blanchong, J. A., M. D. Samuel, K. T. Scribner, B. V. Weckworth, J. A. Langenberg, and K. B. Filcek. 2008. Landscape genetics and the spatial distribution of chronic wasting disease. Biology Letters 4:130-133.

Booth, M. G., J. Hood, T. J. Brooks, and A. Hart. 2010. Anthrax infection in drug users. The Lancet 375:1345-1346.

Bravata, D. M., G. S. Zaric, J.-E. C. Holty, M. L. Brandeau, E. R. Wilhelm, K. M. McDonald, and D. K. Owens. 2006. Reducing mortality from anthrax bioterrorism: strategies for stockpiling and dispensing medical and pharmaceutical supplies. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 4:244-262.

Broennimann, O. and A. Guisan. 2008. Predicting current and future biological invasions: both native and invaded ranges matter. Biology Letters 4:585-589.

Broennimann, O., U. A. Treier, H. Müller-Schärer, W. Thuiller, A. Peterson, and A. Guisan. 2007. Evidence of climatic niche shift during biological invasion. Ecology letters 10:701-709.

Carrel, M. A., M. Emch, R. T. Jobe, A. Moody, and X. F. Wan. 2010. Spatiotemporal Structure of Molecular Evolution of H5N1 Highly Pathogenic Avian Influenza Viruses in Vietnam. PLoS one 5:e8631.

Cherkasskiy, B. L. 1999. A national register of historic and contemporary anthrax foci. J Appl Microbiol 87:192-195.

Cheung, D. T., K. M. Kam, K. L. Hau, T. K. Au, C. K. Marston, J. E. Gee, T. Popovic, M. N. Van Ert, L. Kenefic, and P. Keim. 2005. Characterization of a Bacillus anthracis isolate causing a rare case of fatal anthrax in a 2-year-old boy from Hong Kong. Journal of clinical microbiology 43:1992-1994.

Chikerema, S., G. Matope, and D. Pfukenyi. 2013. Awareness and Attitude Toward Zoonoses with Particular Reference to Anthrax Among Cattle Owners in Selected Rural Communities of Zimbabwe. Vector-Borne and Zoonotic Diseases.

Chikerema, S. M., D. Pfukenyi, G. Matope, and E. Bhebhe. 2012. Temporal and spatial distribution of cattle anthrax outbreaks in Zimbabwe between 1967 and 2006. Tropical animal health and production 44:63-70.

Chirundu, D., S. Chihanga, A. Chimusoro, J. Chirenda, T. Apollo, and M. Tshimanga. 2009. Behavioural factors associated with cutaneous anthrax in Musadzi area of Gokwe North, Zimbabwe. The Central African journal of medicine 55:50.

Collinge, S. K. and C. Ray. 2006. Disease ecology: community structure and pathogen dynamics. Oxford University Press, USA.

Costa, J., A. T. Peterson, and C. B. Beard. 2002. Ecologic niche modeling and differentiation of populations of Triatoma brasiliensis neiva, 1911, the most important Chagas' disease vector in northeastern Brazil (hemiptera, reduviidae, triatominae). American Journal of Tropical Medicine and Hygiene 67:516-520.

Creel, S., N. M. Creel, J. Matovelo, M. Mtambo, E. Batamuzi, and J. Cooper. 1995. The effects of anthrax on endangered African wild dogs (Lycaon pictus). Journal of Zoology 236:199-209.

Cullingham, C. I., E. H. Merrill, M. J. Pybus, T. K. Bollinger, G. A. Wilson, and D. W. Coltman. 2011. Broad and fine-scale genetic analysis of white-tailed deer populations: estimating the relative risk of chronic wasting disease spread. Evolutionary Applications 4:116-131.

Davis, M., D. Hancock, T. Besser, D. Rice, C. Hovde, R. Digiacomo, M. Samadpour, and D. Call. 2003. Correlation between geographic distance and genetic similarity in an international collection of bovine faecal Escherichia coli O157 [ratio] H7 isolates. Epidemiology and Infection 131:923-930.

Dey, R., P. S. Hoffman, and I. J. Glomski. 2012. Germination and Amplification of Anthrax Spores by Soil Dwelling Amoeba. Applied and environmental microbiology.

Dietterich, T. 1995. Overfitting and undercomputing in machine learning. ACM computing surveys (CSUR) 27:326-327.

Dragon, D., B. Elkin, J. Nishi, and T. Ellsworth. 1999. A review of anthrax in Canada and implications for research on the disease in northern bison. Journal of Applied Microbiology 87:208-213.

Dragon, D. C., D. E. Bader, J. Mitchell, and N. Woollen. 2005. Natural dissemination of Bacillus anthracis spores in northern Canada. Applied and environmental microbiology 71:1610-1615.

Dragon, D. C. and R. P. Rennie. 1995. The ecology of anthrax spores: tough but not invincible. The Canadian Veterinary Journal 36:295.

Easterday, W. R., M. N. Van Ert, T. S. Simonson, D. M. Wagner, L. J. Kenefic, C. J. Allender, and P. Keim. 2005. Use of single nucleotide polymorphisms in the plcR gene for specific identification of Bacillus anthracis. Journal of clinical microbiology 43:1995-1997.

Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species. Methods in Ecology and Evolution 1:330-342.

Elliott, P. and D. Wartenberg. 2004. Spatial epidemiology: current approaches and future challenges. Environmental Health Perspectives 112:998.

Epp, T., C. Argue, C. Waldner, and O. Berke. 2010a. Spatial analysis of an anthrax outbreak in Saskatchewan, 2006. The Canadian Veterinary Journal 51:743.

Epp, T., C. Waldner, and C. K. Argue. 2010b. Case-control study investigating an anthrax outbreak in Saskatchewan, Canada--Summer 2006. Canadian Veterinary Journal-Revue Veterinaire Canadienne 51:973-978.

Epps, C. W., J. D. Wehausen, V. C. Bleich, S. G. Torres, and J. S. Brashares. 2007. Optimizing dispersal and corridor models using landscape genetics. Journal of Applied Ecology 44:714-724.

Farlow, J., K. L. Smith, J. Wong, M. Abrams, M. Lytle, and P. Keim. 2001. Francisella tularensis strain typing using multiple-locus, variable-number tandem repeat analysis. Journal of clinical microbiology 39:3186-3192.

Farlow, J., D. M. Wagner, M. Dukerich, M. Stanley, M. Chu, K. Kubota, J. Petersen, and P. Keim. 2005. Francisella tularensis in the United States. Emerg Infect Dis 11:1835-1841.

Fasanella, A., G. Garofolo, D. Galante, V. Quaranta, L. Palazzo, F. Lista, R. Adone, and M. H. Jones. 2010. Severe anthrax outbreaks in Italy in 2004: Considerations on factors involved in the spread of infection. The new microbiologica 33:83.

Fasanella, A., G. Garofolo, M. Hossain, M. Shamsuddin, J. Blackburn, and M. Hugh-Jones. 2012. Bangladesh anthrax outbreaks are probably caused by contaminated livestock feed. Epidemiology and Infection 1:1-8.

Fasanella, A., M. Van Ert, S. A. Altamura, G. Garofolo, C. Buonavoglia, G. Leori, L. Huynh, S. Zanecki, and P. Keim. 2005. Molecular diversity of Bacillus anthracis in Italy. Journal of clinical microbiology 43:3398.

Fisher, M. C., W. P. Hanage, S. De Hoog, E. Johnson, M. D. Smith, N. J. White, and N. Vanittanakom. 2005. Low effective dispersal of asexual genotypes in heterogeneous landscapes by the endemic pathogen Penicillium marneffei. PLoS pathogens 1:1986-1990.

Fitzpatrick, M. C. and W. W. Hargrove. 2009. The projection of species distribution models and the problem of non-analog climate. Biodiversity and Conservation 18:2255-2261.

Foster, J. T., S. M. Beckstrom-Sternberg, T. Pearson, J. S. Beckstrom-Sternberg, P. S. Chain, F. F. Roberto, J. Hnath, T. Brettin, and P. Keim. 2009. Whole-genome-based phylogeny and divergence of the genus Brucella. Journal of bacteriology 191:2864-2870.

Fouet, A., K. L. Smith, C. Keys, J. Vaissaire, C. Le Doujet, M. Levy, M. Mock, and P. Keim. 2002. Diversity among French Bacillus anthracis isolates. Journal of clinical microbiology 40:4732.

Fowler, R. A., G. D. Sanders, D. M. Bravata, B. Nouri, J. M. Gastwirth, D. Peterson, A. G. Broker, A. M. Garber, and D. K. Owens. 2005. Cost-effectiveness of defending against bioterrorism: a comparison of vaccination and antibiotic prophylaxis against anthrax. Annals of internal medicine 142:601-610.

French, N., M. Barrigas, P. Brown, P. Ribiero, N. Williams, H. Leatherbarrow, R. Birtles, E. Bolton, P. Fearnhead, and A. Fox. 2005. Spatial epidemiology and natural population structure of Campylobacter jejuni colonizing a farmland ecosystem. Environmental microbiology 7:1116-1126.

Friedman, A. and A.-A. Yakubu. 2012. Anthrax Epizootic and Migration: Persistence or Extinction. Mathematical Biosciences.

Fulbright, T. E. and J. A. Ortega-s. 2006. White-tailed deer habitat: ecology and management on rangelands. Texas A & M University Press.

Furniss, P. and B. Hahn. 1981. A mathematical model of an anthrax epizoötic in the Kruger National Park. Applied Mathematical Modelling 5:130-136.

Gainer, R. S. and J. R. Saunders. 1989. Aspects of the epidemiology of anthrax in Wood Buffalo National Park and environs. The Canadian Veterinary Journal 30:953.

Garofolo, G., A. Ciammaruconi, A. Fasanella, S. Scasciamacchia, R. Adone, V. Pittiglio, and F. Lista. 2010. SNR analysis: molecular investigation of an anthrax epidemic. BMC Veterinary Research 6:11.

Garofolo, G., L. Serrecchia, M. Corrò, and A. Fasanella. 2011. Anthrax phylogenetic structure in Northern Italy. BMC Research Notes 4:273.

Gates, C. C., B. T. Elkin, and D. C. Dragon. 1995. Investigation, Control and Epizootiology of Anthrax in a Geographically Isolated, Free-Roaming Bison Population in Northern Canada. Canadian Journal of Veterinary Research-Revue Canadienne De Recherche Veterinaire 59:256-264.

Giles, J., A. T. Peterson, and A. Almeida. 2011. Ecology and Geography of Plague Transmission Areas in Northeastern Brazil. PLoS Neglected Tropical Diseases 5:e925.

Girard, J. M., D. M. Wagner, A. J. Vogler, C. Keys, C. J. Allender, L. C. Drickamer, and P. Keim. 2004. Differential plague-transmission dynamics determine Yersinia pestis population genetic structure on local, regional, and global scales. Proceedings of the National Academy of Sciences of the United States of America 101:8408-8413.

Gombe, N., B. Nkomo, A. Chadambuka, G. Shambira, and M. Tshimanga. 2010. Risk factors for contracting anthrax in Kuwirirana ward, Gokwe North, Zimbabwe. African health sciences 10:159-164.

Gonzalez, S., J. Soto-Centeno, and D. Reed. 2011. Population distribution models: Species distributions are better modeled using biologically relevant data partitions. BMC ecology 11:20.

Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84:2809-2815.

Gray, R. R., A. J. Tatem, S. Lamers, W. Hou, O. Laeyendecker, D. Serwadda, N. Sewankambo, R. H. Gray, M. Wawer, and T. C. Quinn. 2009. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. AIDS (London, England) 23:F9.

Grinnell, J. 1917. The niche-relationships of the California Thrasher. Auk 34:427-433.

Guh, A., M. L. Heyman, D. Barden, J. Fontana, and J. L. Hadler. 2010. Lessons learned from the investigation of a cluster of cutaneous anthrax cases in Connecticut. Journal of Public Health Management and Practice 16:201.

Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology letters 8:993-1009.

Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson. 2007. What matters for predicting the occurrences of trees: techniques, data, or species'characteristics? Ecological Monographs 77:615-630.

Hahn, B. D. and P. R. Furniss. 1983. A deterministic model of an anthrax epizootic: threshold results. Ecological Modelling 20:233-241.

Hampson, K., T. Lembo, P. Bessell, H. Auty, C. Packer, J. Halliday, C. A. Beesley, R. Fyumagwa, R. Hoare, and E. Ernest. 2011. Predictability of anthrax infection in the Serengeti, Tanzania. Journal of Applied Ecology 48:1333-1344.

Harrell, L. J., G. L. Andersen, and K. H. Wilson. 1995. Genetic variability of Bacillus anthracis and related species. Journal of clinical microbiology 33:1847-1850.

Hay, S. I., A. J. Tatem, A. J. Graham, S. J. Goetz, and D. J. Rogers. 2006. Global environmental data for mapping infectious disease distribution. Advances in Parasitology, Vol 62 62:37-77.

Henderson, I., C. Duggleby, and P. Turnbull. 1994. Differentiation of Bacillus anthracis from other Bacillus cereus group bacteria with the PCR. International journal of systematic bacteriology 44:99-105.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773-785.

Hickerson, M., B. Carstens, J. Cavender-Bares, K. Crandall, C. Graham, J. Johnson, L. Rissler, P. Victoriano, and A. Yoder. 2010. Phylogeography's past, present, and future: 10 years after. Molecular Phylogenetics and Evolution 54:291-301.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965-1978.

Himsworth, C. G. 2008. The danger of lime use in agricultural anthrax disinfection procedures: the potential role of calcium in the preservation of anthrax spores. Canadian Veterinary Journal-Revue Veterinaire Canadienne 49:1208-1210.

Hoffmaster, A. R., C. C. Fitzgerald, E. Ribot, L. W. Mayer, T. Popovic, A. Hoffmaster, C. Fitzgerald, and L. Mayer. 2002. Molecular subtyping of Bacillus anthracis and the 2001 bioterrorism-associated anthrax outbreak, United States. Emerging Infectious Diseases 8:1111.

Holt, R., M. Barfield, and R. Gomulkiewicz. 2005. Theories of niche conservatism and evolution: could exotic species be potential tests. Species invasions: insights into ecology, evolution, and biogeography:259-290.

Hugh-Jones, M. 1999. 1996-97 Global Anthrax Report. J Appl Microbiol 87:189-191.

Hugh-Jones, M. and J. Blackburn. 2009. The ecology of Bacillus anthracis. Molecular Aspects of Medicine 30:356-367.

Hugh-Jones, M. E. and V. de Vos. 2002. Anthrax and wildlife. Rev Sci Tech 21:359-383.

Hutchinson, G. E. 1957. Population Studies - Animal Ecology and Demography - Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415-427.

Huynh, L. Y., M. N. V. Ert, T. Hadfield, W. S. Probert, B. H. Bellaire, M. Dobson, R. J. Burgess, R. S. Weyant, T. Popovic, and S. Zanecki. 2008. Multiple locus variable number tandem repeat (VNTR) analysis (MLVA) of Brucella spp. identifies species-specific markers and insights into phylogenetic relationships. National Institute of Allergy and Infectious Diseases, NIH:47-54.

Jackson, P. J., E. A. Walthers, A. S. Kalif, K. L. Richmond, D. M. Adair, K. K. Hill, C. R. Kuske, G. L. Andersen, K. H. Wilson, and M. Hugh-Jones. 1997. Characterization of the variable-number tandem repeats in vrrA from different Bacillus anthracis isolates. Applied and environmental microbiology 63:1400-1405.

Jacquez, G. M. 1996a. A k nearest neighbour test for space–time interaction. Statistics in medicine 15:1935-1949.

Jacquez, G. M. 1996b. Statistical software for the clustering of health events. Statistics in medicine 15:951-952.

Jiménez-Valverde, A., Y. Nakazawa, A. Lira-Noriega, and A. T. Peterson. 2009. Environmental correlation structure and ecological niche model projections. Biodiversity informatics 6.

Johansson, A., J. Farlow, P. Larsson, M. Dukerich, E. Chambers, M. Byström, J. Fox, M. Chu, M. Forsman, and A. Sjöstedt. 2004. Worldwide genetic relationships among Francisella tularensis isolates determined by multiple-locus variable-number tandem repeat analysis. Journal of bacteriology 186:5808-5818.

Joyner, T. A. 2010. Ecological niche modeling of a zoonosis: A case study using anthrax outbreaks and climate change in Kazakhstan. University of Florida.

Joyner, T. A., L. Lukhnova, Y. Pazilov, G. Temiralyeva, M. E. Hugh-Jones, A. Aikimbayev, and J. K. Blackburn. 2010. Modeling the potential distribution of Bacillus anthracis under multiple climate change scenarios for Kazakhstan. PLoS one 5:e9596.

Karahocagil, M., N. Akdeniz, H. Akdeniz, Ö. Çalka, H. Karsen, A. Bilici, S. Bilgili, and Ö. Evirgen. 2008. Cutaneous anthrax in Eastern Turkey: a review of 85 cases. Clinical and experimental dermatology 33:406-411.

Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, C. R. Kuske, and P. Jackson. 1997. Molecular evolution and diversity in Bacillus anthracis as detected by amplified fragment length polymorphism markers. Journal of bacteriology 179:818.

Keim, P., L. Price, A. Klevytska, K. Smith, J. Schupp, R. Okinaka, P. Jackson, and M. Hugh-Jones. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis. Journal of bacteriology 182:2928.

Keim, P., K. L. Smith, C. Keys, H. Takahashi, T. Kurata, and A. Kaufmann. 2001. Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan. Journal of clinical microbiology 39:4566.

Keim, P., M. N. Van Ert, T. Pearson, A. J. Vogler, L. Y. Huynh, and D. M. Wagner. 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. Infection, Genetics and Evolution 4:205-213.

Kenefic, L., J. Beaudry, C. Trim, R. Daly, R. Parmar, S. Zanecki, L. Huynh, M. Van Ert, D. Wagner, and T. Graham. 2008a. High resolution genotyping of Bacillus anthracis outbreak strains using four highly mutable single nucleotide repeat markers. Letters in applied microbiology 46:600-603.

Kenefic, L., J. Beaudry, C. Trim, L. Huynh, S. Zanecki, M. Matthews, J. Schupp, M. Van Ert, and P. Keim. 2008b. A high resolution four-locus multiplex single nucleotide repeat (SNR) genotyping system in Bacillus anthracis. Journal of microbiological methods 73:269-272.

Kenefic, L. J., T. Pearson, R. T. Okinaka, W. K. Chung, T. Max, C. P. Trim, J. A. Beaudry, J. M. Schupp, M. N. Van Ert, C. K. Marston, K. Gutierrez, A. K. Swinford, A. R. Hoffmaster, and P. Keim. 2008c. Texas isolates closely related to Bacillus anthracis Ames. Emerg Infect Dis 14:1494-1496.

Kenefic, L. J., T. Pearson, R. T. Okinaka, J. M. Schupp, D. M. Wagner, J. Ravel, A. R. Hoffmaster, C. P. Trim, W. K. Chung, and J. A. Beaudry. 2009. Pre-columbian origins for north american anthrax. PLoS one 4:e4813.

Kidd, D. M. and M. G. Ritchie. 2006. Phylogeographic information systems: putting the geography into phylogeography. Journal of Biogeography 33:1851-1865.

Knouft, J. H., J. B. Losos, R. E. Glor, and J. J. Kolbe. 2006. Phylogenetic analysis of the evolution of the niche in lizards of the Anolis sagrei group. Ecology 87:29-38.

Knowles, L. L. 2009. Statistical phylogeography. Annual review of ecology, evolution, and systematics 40:593-612.

Kolbe, J. J., P. S. VanMiddlesworth, N. Losin, N. Dappen, and J. B. Losos. 2012. Climatic niche shift predicts thermal trait response in one but not both introductions of the Puerto Rican lizard Anolis cristatellus to Miami, Florida, USA. Ecology and Evolution.

Kracalik, I., L. Lukhnova, A. Aikimbayev, Y. Pazilov, G. Temiralyeva, and J. K. Blackburn. 2011. Incorporating retrospective clustering into a prospective cusum methodology for anthrax: Evaluating the effects of disease expectation. Spatial and Spatio-temporal Epidemiology 2:11-21.

Kracalik, I. T., J. K. Blackburn, L. Lukhnova, Y. Pazilov, M. E. Hugh-Jones, and A. Aikimbayev. 2012. Analysing the spatial patterns of livestock anthrax in Kazakhstan in relation to environmental factors: a comparison of local (Gi*) and morphology cluster statistics. Geospatial health 7:111-126.

Kugeler, K. J., P. S. Mead, A. M. Janusz, J. E. Staples, K. A. Kubota, L. G. Chalcraft, and J. M. Petersen. 2009. Molecular epidemiology of Francisella tularensis in the United States. Clinical Infectious Diseases 48:863-870.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in bioinformatics 5:150-163.

Lamothe, W. D. 2011. Gastrointestinal Anthrax in New Hampshire: A 2009 Case Report. Lab Medicine 42:363-368.

Last, J. M. 2000. A dictionary of epidemiology. Oxford University Press, USA.

Lembo, T., K. Hampson, H. Auty, C. A. Beesley, P. Bessell, C. Packer, J. Halliday, R. Fyumagwa, R. Hoare, and E. Ernest. 2011. Serologic Surveillance of Anthrax in the Serengeti Ecosystem, Tanzania, 1996–2009.

Lindeque, P. M. and P. C. B. Turnbull. 1994. Ecology and Epidemiology of Anthrax in the Etosha-National-Park, Namibia. Onderstepoort Journal of Veterinary Research 61:71-83.

Lista, F., G. Faggioni, S. Valjevac, A. Ciammaruconi, J. Vaissaire, C. Le Doujet, O. Gorgé, R. De Santis, A. Carattoli, and A. Ciervo. 2006. Genotyping of Bacillus anthracis strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. BMC microbiology 6:33.

Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17:145-151.

MacMillan, K., A. J. Monaghan, T. Apangu, K. S. Griffith, P. S. Mead, S. Acayo, R. Acidri, S. M. Moore, J. T. Mpanga, and R. E. Enscore. 2012. Climate predictors of the spatial distribution of human plague cases in the West Nile region of Uganda. The American journal of tropical medicine and hygiene 86:514-523.

Maher, S. P., C. Ellis, K. L. Gage, R. E. Enscore, and A. T. Peterson. 2010. Range-wide Determinants of Plague Distribution in North America. The American journal of tropical medicine and hygiene 83:736-742.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189-197.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer research 27:209-220.

Marston, C. K., C. A. Allen, J. Beaudry, E. P. Price, S. R. Wolken, T. Pearson, P. Keim, and A. R. Hoffmaster. 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. PLoS one 6:e28274.

McNyset, K. M. 2005. Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. Ecology of Freshwater Fish 14:243-255.

Meade, M. and M. Emch. 2010. Medical geography. Guilford Press.

Moore, D. A. 1999. Spatial diffusion of raccoon rabies in Pennsylvania, USA. Preventive veterinary medicine 40:19-32.

Morelli, G., Y. Song, C. J. Mazzoni, M. Eppinger, P. Roumagnac, D. M. Wagner, M. Feldkamp, B. Kusecek, A. J. Vogler, and Y. Li. 2010. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nature genetics 42:1140-1143.

Morens, D. M. 2003. Characterizing a “new” disease: epizootic and epidemic anthrax, 1769–1780. Journal Information 93.

Mukherjee, A., R. Diaz, M. Thom, W. Overholt, and J. Cuda. 2012. Niche-based prediction of establishment of biocontrol agents: an example with Gratiana boliviana and tropical soda apple. Biocontrol Science and Technology 22:447-461.

Mukherjee, A., D. Williams, G. Wheeler, J. Cuda, S. Pal, and W. Overholt. 2011. Brazilian peppertree (Schinus terebinthifolius) in Florida and South America: evidence of a possible niche shift driven by hybridization. Biological Invasions:1-16.

Mullins, J., G. Garofolo, M. Van Ert, A. Fasanella, L. Lukhnova, M. Hugh-Jones, and J. Blackburn. 2013. Ecological Niche Modeling of *Bacillus anthracis* on Three Continents: Evidence for Genetic-Ecological Divergence. Ecosphere In review.

Mullins, J., L. Lukhnova, A. Aikimbayev, Y. Pazilov, M. Van Ert, and J. K. Blackburn. 2011. Ecological Niche Modelling of the Bacillus anthracis A1. a sub-lineage in Kazakhstan. BMC ecology 11:32.

Nakazawa, Y., R. Williams, A. T. Peterson, P. Mead, E. Staples, and K. L. Gage. 2007. Climate change effects on plague and tularemia in the United States. Vector-Borne and Zoonotic Diseases 7:529-540.

Nakazawa, Y., R. A. Williams, A. T. Peterson, P. S. Mead, K. J. Kugeler, and J. M. Petersen. 2010. Ecological niche modeling of Francisella tularensis subspecies and clades in the United States. Am J Trop Med Hyg 82:912-918.

Ndiva Mongoh, M., R. Hearne, and M. Khaitsa. 2008. Private and public economic incentives for the control of animal diseases: the case of anthrax in livestock. Transboundary and emerging diseases 55:319-328.

O'Brien, D. J., J. B. Kaneene, A. Getis, J. W. Lloyd, M. R. Rip, and R. W. Leader. 1999. Spatial and temporal distribution of selected canine cancers in Michigan, USA, 1964–1994. Preventive veterinary medicine 42:1-15.

Odum, E. 1959. Fundamentals of ecology. Sauders. Philadelphia and London.

Oksanen, J., G. Blanchet, R. Kindt, P. Legendre, P. Minchin, G. Simpson, P. Solymos, M. H. M. Stevens, and H. Wagner. 2013. vegan:Community Ecology Package. R package v. 2.0-7.

Ostfeld, R. S., G. E. Glass, and F. Keesing. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. Trends in Ecology & Evolution 20:328-336.

Özden, K., Z. ÖZKURT, S. EROL, M. H. UYANIK, and M. PARLAK. 2012. Cutaneous anthrax patients in Eastern Anatolia, Turkey: a review of 44 adults cases.

Palazzo, L., E. De Carlo, G. Santagada, L. Serrecchia, A. Aceti, A. Guarino, R. Adone, and A. Fasanella. 2012. Recent Epidemic-Like Anthrax Outbreaks in Italy: What Are the Probable Causes? Open Journal of Veterinary Medicine 2:74-76.

Parkinson, R., A. Rajic, and C. Jenson. 2003. Investigation of an anthrax outbreak in Alberta in 1999 using a geographic information system. Canadian Veterinary Journal-Revue Veterinaire Canadienne 44:315-318.

Parmenter, R. R., E. P. Yadav, C. A. Parmenter, P. Ettestad, and K. L. Gage. 1999. Incidence of plague associated with increased winter-spring precipitation in New Mexico. The American journal of tropical medicine and hygiene 61:814-821.

Pearman, P. B., A. Guisan, O. Broennimann, and C. F. Randin. 2008. Niche dynamics in space and time. Trends in Ecology & Evolution 23:149-158.

Pearson, T., J. D. Busch, J. Ravel, T. D. Read, S. D. Rhoton, J. M. U'ren, T. S. Simonson, S. M. Kachur, R. R. Leadem, and M. L. Cardon. 2004. Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. Proceedings of the National Academy of Sciences of the United States of America 101:13536-13541.

Peck, R. N. and D. W. Fitzgerald. 2007. Cutaneous anthrax in the Artibonite Valley of Haiti: 1992–2002. The American journal of tropical medicine and hygiene 77:806-811.

Pepperell, C. S., J. M. Granka, D. C. Alexander, M. A. Behr, L. Chui, J. Gordon, J. L. Guthrie, F. B. Jamieson, D. Langlois-Klassen, and R. Long. 2011. Dispersal of Mycobacterium tuberculosis via the Canadian fur trade. Proceedings of the National Academy of Sciences 108:6526-6531.

Petersen, J. M., J. K. Carlson, G. Dietrich, R. J. Eisen, J. Coombs, A. M. Janusz, J. Summers, C. B. Beard, and P. S. Mead. 2008. Multiple Francisella tularensis subspecies and clades, tularemia outbreak, Utah. Emerging Infectious Diseases 14:1928.

Peterson, A. 2006. Ecologic niche modeling and spatial patterns of disease transmission. Emerging Infectious Diseases 12:1822.

Peterson, A. and Y. Nakazawa. 2008. Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. Global Ecology and Biogeography 17:135-144.

Peterson, A. T. 2001. Predicting species'geographic distributions based on ecological niche modeling. The Condor 103:599-605.

Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. Quarterly Review of Biology 78:419-433.

Peterson, A. T., L. G. Ball, and K. P. Cohoon. 2002a. Predicting distributions of Mexican birds using ecological niche modelling methods. Ibis-Journal of the British Ornithologists Union 144:27.

Peterson, A. T., J. T. Bauer, and J. N. Mills. 2004. Ecologic and geographic distribution of filovirus disease. Emerging Infectious Diseases 10:40-47.

Peterson, A. T. and R. D. Holt. 2003. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. Ecology letters 6:774-782.

Peterson, A. T., M. Pape , and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30:550-560.

Peterson, A. T., M. Papes, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling 213:63-72.

Peterson, A. T., V. Sánchez-Cordero, C. B. Beard, and J. M. Ramsey. 2002b. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico.

Peterson, A. T., V. Sanchez-Cordero, C. Ben Beard, and J. M. Ramsey. 2002c. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. Emerging Infectious Diseases 8:662-667.

Peterson, A. T. and J. Shaw. 2003. Lutzomyia vectors for cutaneous leishmaniasis in Southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. International Journal for Parasitology 33:919-931.

Peterson, A. T., J. Soberon, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, and M. B. Araujo. 2011. Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press.

Pfukenyi, D. M., G. Matope, S. Chikerema, and E. Bhebhe. 2011. Temporal and spatial distribution of cattle anthrax outbreaks in Zimbabwe between 1967 and 2006.

Price, E. P., M. L. Seymour, D. S. Sarovich, J. Latham, S. R. Wolken, J. Mason, G. Vincent, K. P. Drees, S. M. Beckstrom-Sternberg, and A. M. Phillippy. 2012. Molecular Epidemiologic Investigation of an Anthrax Outbreak among Heroin Users, Europe. Emerging Infectious Diseases 18:1307.

Prins, H. and F. Weyerhaeuser. 1987. Epidemics in populations of wild ruminants: anthrax and impala, rinderpest and buffalo in Lake Manyara National Park, Tanzania. Oikos:28-38.

Proffitt, K. M., J. L. Grigg, R. A. Garrott, K. L. Hamlin, J. Cunningham, J. A. Gude, and C. Jourdonnais. 2010. Changes in Elk Resource Selection and Distributions Associated With a Late-Season Elk Hunt. The Journal of Wildlife Management 74:210-218.

Randin, C. F., T. Dirnböck, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography 33:1689-1703.

Rice, N. H., E. Martinez-Meyer, and A. T. Peterson. 2003. Ecological niche differentiation in the Aphelocoma jays: a phylogenetic perspective. Biological Journal of the Linnean Society 80:369-383.

Ripley, B. D. 1977. Modelling spatial patterns. Journal of the Royal Statistical Society. Series B (methodological):172-212.

Rissler, L. J. and J. J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (Aneides flavipunctatus). Systematic Biology 56:924-942.

Rödder, D. and S. Lötters. 2010. Explanative power of variables used in species distribution modelling: an issue of general model transferability or niche shift in the invasive Greenhouse frog (Eleutherodactylus planirostris). Naturwissenschaften 97:781-796.

Rödder, D., S. Schmidtlein, M. Veith, and S. Lötters. 2009. Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? PLoS one 4:e7843.

Ron, S. R. 2005. Predicting the Distribution of the Amphibian Pathogen Batrachochytrium dendrobatidis in the New World1. Biotropica 37:209-221.

Saile, E. and T. M. Koehler. 2006. Bacillus anthracis multiplication, persistence, and genetic exchange in the rhizosphere of grass plants. Applied and environmental microbiology 72:3168-3174.

Saupe, E., V. Barve, C. Myers, J. Soberón, N. Barve, C. Hensz, A. Peterson, H. Owens, and A. Lira-Noriega. 2012. Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. Ecological Modelling 237:11-22.

Schuch, R. and V. A. Fischetti. 2009. The secret life of the anthrax agent Bacillus anthracis: bacteriophage-mediated ecological adaptations. PLoS one 4:e6532.

Schuch, R., A. Pelzek, S. Kan, and V. Fischetti. 2010. Prevalence of Bacillus anthracis-like organisms and bacteriophages in the intestinal tract of the earthworm Eisenia fetida. Applied and environmental microbiology 76:2286-2294.

Siamudaala, V. M., J. M. Bwalya, H. M. Munang'andu, P. G. Sinyangwe, F. Banda, A. S. Mweene, A. Takada, and H. Kida. 2006. Ecology and epidemiology of anthrax in cattle and humans in Zambia. Japanese Journal of Veterinary Research 54:15-23.

Simonson, T. S., R. T. Okinaka, B. Wang, W. R. Easterday, L. Huynh, J. M. U'Ren, M. Dukerich, S. R. Zanecki, L. J. Kenefic, and J. Beaudry. 2009. Bacillus anthracis in China and its relationship to worldwide lineages. BMC microbiology 9:71.

Smith, K. L. 1999. Epidemiology of anthrax in the Kruger National Park, South Africa: Genetic diversity and environment. Dissertation. LSU, Baton Rouge.

Smith, K. L., V. De Vos, H. B. Bryden, M. E. Hugh-Jones, A. Klevytska, L. B. Price, P. Keim, and D. T. Scholl. 1999. Meso-scale ecology of anthrax in southern Africa: a pilot study of diversity and clustering. J Appl Microbiol 87:204-207.

Smith, K. L., V. DeVos, H. Bryden, L. B. Price, M. E. Hugh-Jones, and P. Keim. 2000. Bacillus anthracis diversity in Kruger National Park. J Clin Microbiol 38:3780-3784.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology letters 10:1115-1123.

Soberón, J. and M. Nakamura. 2009. Niches and distributional areas: concepts, methods, and assumptions. Proceedings of the National Academy of Sciences 106:19644-19650.

Soberon, J. and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity informatics 2.

Staples, J. E., K. A. Kubota, L. G. Chalcraft, P. S. Mead, and J. M. Petersen. 2006. Epidemiologic and molecular analysis of human tularemia, United States, 1964–2004. Emerging Infectious Diseases 12:1113.

Steele, J. H. and R. J. Helvig. 1953. Anthrax in the United States. Public Health Reports 68:616.

Stein, C. 1945. The history and distribution of anthrax in livestock in the United States. Vet Med 40:340-349.

Stein, C. and G. Van Ness. 1955. A ten year survey of anthrax in livestock with special reference to outbreaks in 1954. Vet Med 50:579-588.

Sternbach, G. 2003. The history of anthrax. The Journal of emergency medicine 24:463-467.

Stockwell, D. and D. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science 13:143-158.

Stockwell, D. R. B. and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148:1-13.

Storfer, A., M. Murphy, J. Evans, C. Goldberg, S. Robinson, S. Spear, R. Dezzani, E. Delmelle, L. Vierling, and L. Waits. 2006. Putting the 'landscape'in landscape genetics. Heredity 98:128-142.

Stratilo, C. W. and D. E. Bader. 2012. Genetic Diversity among Bacillus anthracis Soil Isolates at Fine Geographic Scales. Applied and environmental microbiology 78:6433-6437.

Stratilo, C. W., C. T. Lewis, L. Bryden, M. R. Mulvey, and D. Bader. 2006. Single-nucleotide repeat analysis for subtyping Bacillus anthracis isolates. Journal of clinical microbiology 44:777.

Svensson, K., E. Bäck, H. Eliasson, L. Berglund, M. Granberg, L. Karlsson, P. Larsson, M. Forsman, and A. Johansson. 2009. Landscape epidemiology of tularemia outbreaks in Sweden. Emerging Infectious Diseases 15:1937.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. Science; Science.

Sylvestre, P., E. Couture-Tosi, and M. Mock. 2003. Polymorphism in the collagen-like region of the Bacillus anthracis BclA protein leads to variation in exosporium filament length. Journal of bacteriology 185:1555-1563.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution 24:1596-1599.

Turnbull, P., R. Bell, K. Saigawa, F. Munyenyembe, C. Mulenga, and L. Makala. 1991. Anthrax in wildlife in the Luangwa Valley, Zambia. Veterinary Record 128:399-403.

Turnbull, P., M. Doganay, P. Lindeque, B. Aygen, and J. McLaughlin. 1992. Serology and anthrax in humans, livestock and Etosha National Park wildlife. Epidemiology and Infection 108:299-313.

Turnbull, P. C. B. and S. V. Shadomy. 2010. Anthrax from 5000 BC to AD 2010. Bacillus Anthracis and Anthrax:1.

Turner, W. C., P. Imologhome, Z. Havarua, G. P. Kaaya, J. K. Mfune, I. D. Mpofu, and W. M. Getz. 2013. Soil ingestion, nutrition and the seasonality of anthrax in herbivores of Etosha National Park. Ecosphere 4:art13.

Unwin, D. J. 1975. An introduction to trend surface analysis. Geo Abstracts Limited.

Van Ert, M. N., W. R. Easterday, L. Y. Huynh, R. T. Okinaka, M. E. Hugh-Jones, J. Ravel, S. R. Zanecki, T. Pearson, T. S. Simonson, J. M. U'Ren, S. M. Kachur, R. R. Leadem-Dougherty, S. D. Rhoton, G. Zinser, J. Farlow, P. R. Coker, K. L. Smith, B. X. Wang, L. J. Kenefic, C. M. Fraser-Liggett, D. M. Wagner, and P. Keim. 2007a. Global Genetic Population Structure of Bacillus anthracis. PLoS one 2:-.

Van Ert, M. N., W. R. Easterday, T. S. Simonson, J. M. U'Ren, T. Pearson, L. J. Kenefic, J. D. Busch, L. Y. Huynh, M. Dukerich, C. B. Trim, J. Beaudry, A. Welty-Bernard, T. Read, C. M. Fraser, J. Ravel, and P. Keim. 2007b. Strain-specific single-nucleotide polymorphism assays for the Bacillus anthracis Ames strain. Journal of clinical microbiology 45:47-53.

Van Ness, G. 1967. Geologic Implications of Anthrax. Geological Society of America 90:61-64.

Van Ness, G. 1971. Ecology of Anthrax. Science 172:1303-1307.

Van Ness, G., Stein CD. 1956. Soils of the United States favorable for Anthrax. J Am Vet Med Assoc 128:7-9.

Vogler, A., D. Birdsell, J. Lee, J. Vaissaire, C. Doujet, M. Lapalus, D. Wagner, and P. Keim. 2011a. Phylogeography of Francisella tularensis ssp. holarctica in France. Letters in applied microbiology 52:177-180.

Vogler, A. J., D. Birdsell, L. B. Price, J. R. Bowers, S. M. Beckstrom-Sternberg, R. K. Auerbach, J. S. Beckstrom-Sternberg, A. Johansson, A. Clare, and J. L. Buchhagen. 2009. Phylogeography of Francisella tularensis: global expansion of a highly fit clone. Journal of bacteriology 191:2474-2484.

Vogler, A. J., F. Chan, R. Nottingham, G. Andersen, K. Drees, S. M. Beckstrom-Sternberg, D. M. Wagner, S. Chanteau, and P. Keim. 2013. A Decade of Plague in Mahajanga, Madagascar: Insights into the Global Maritime Spread of Pandemic Plague. mBio 4.

Vogler, A. J., F. Chan, D. M. Wagner, P. Roumagnac, J. Lee, R. Nera, M. Eppinger, J. Ravel, L. Rahalison, and B. W. Rasoamanana. 2011b. Phylogeography and molecular epidemiology of Yersinia pestis in Madagascar. PLoS Neglected Tropical Diseases 5:e1319.

Ward, M. P. and T. E. Carpenter. 2000. Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. Preventive veterinary medicine 45:257-284.

Wiens, J. J. and C. H. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Annual review of ecology, evolution, and systematics:519-539.

Wiley, E., K. M. McNyset, A. T. Peterson, C. R. Robins, and A. M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography 16:120–127.

Williams, G., E. Linley, R. Nicholas, and L. Baillie. 2013. The role of the exosporium in the environmental distribution of anthrax. Journal of Applied Microbiology 114:396-403.

Woods, C. W., K. Ospanov, A. Myrzabekov, M. Favorov, B. Plikaytis, and D. A. Ashford. 2004. Risk factors for human anthrax among contacts of anthrax-infected livestock in Kazakhstan. The American journal of tropical medicine and hygiene 71:48.

Zaric, G. S., D. M. Bravata, J.-E. C. Holty, K. M. McDonald, D. K. Owens, and M. L. Brandeau. 2008. Modeling the logistics of response to anthrax bioterrorism. Medical Decision Making 28:332-350.

BIOGRAPHICAL SKETCH

Jocelyn Mullins was born in New York and shortly thereafter moved to Gainesville, Florida, living there for 2 years while her stepfather finished his doctoral degree. At the age of four she moved to Miami where she lived until graduation from Coral Gables High School. As a child Jocelyn was famous for her addiction to “bug and leaf shows” and harbored a dream of someday narrating nature specials in the footsteps of David Attenborough and George Page. After a brief stint at a college in Connecticut, Jocelyn spent some time exploring the country before landing at the University of Oregon where she earned a Bachelor of Science in biology with a concentration in ecology and evolution.

While in Eugene, Jocelyn worked for a small animal and exotic veterinarian and for a group that rehabilitated birds of prey. It was here that she discovered medicine and epidemiology and decided to pursue veterinary school, eventually obtaining her DVM at the University of Florida. Following veterinary school Jocelyn worked in small animal practice in St. Augustine, FL and as a pharmacovigilance manager for a veterinary pharmaceutical company in Duluth, GA. Life then brought her back to Gainesville for a third time where she decided to finally go back to school for a Master in Public Health and then a PhD in geography. A veterinarian in geography seems a bit surprising to many. However, the work combines many of Jocelyn’s interests and tackles real world animal and human problems. Now her children await her graduation.

After graduation Jocelyn will enter a two year applied epidemiology fellowship with the Centers for Disease Control and Prevention and will be stationed in Hartford, CT.