DISSERTATION

DETERMINATION OF *YERSINIA PESTIS* POPULATION STRUCTURE AS A MODEL FOR LOW PREVALENCE/EPIZOOTIC DISEASE DYNAMICS

Submitted by

Jennifer L. Lowell

Department of Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2007

COLORADO STATE UNIVERSITY

December 13, 2006

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY JENNIFER L. LOWELL ENTITLED "DETERMINATION OF *YERSINIA PESTIS* POPULATION STRUCTURE AS A MODEL FOR LOW PREVALENCE/EPIZOOTIC DISEASE DYNAMICS" BE ACCEPTED AS FULLFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Advisor

Department Head

ABSTRACT OF DISSERTATION

DETERMINATION OF *YERSINIA PESTIS* POPULATION STRUCTURE AS A MODEL FOR LOW PREVALENCE/EPIZOOTIC DISEASE DYNAMICS

Plague is a highly virulent zoonotic disease caused by the gram-negative bacterium *Yersinia pestis*. It is estimated that *Y. pestis* evolved only 1,500-20,000 years ago from the mild enteric pathogen *Yersinia pseudotuberculosis*. Despite this recent emergence, *Y. pestis* has spread around the globe killing millions of people in three major pandemics and currently infects more than 200 mammal species world-wide. Plague was introduced in the U. S. from Asia in the early 1900s through shipping, and has established primarily in the west and southwest. Currently in the U.S., *Y. pestis* is most active in rodents in the Four-Corners region where frequent fluctuations in transmission rates result in the amplification of disease to epidemic (epizootic) proportions, or a decline to endemic (enzootic) states. During rodent epizootic events, *Y. pestis* occasionally spills over into humans causing severe and often fatal illness. It is hypothesized that these disease fluctuations are the result of changing rodent population densities and contact rates between hosts and/or disease vectors whose populations are affected by changes in temperature and moisture triggering cascades in terrestrial net primary productivity. While increasing evidence shows that plague can be influenced by factors like large-scale climatic patterns, the mechanisms of these disease cycles are unknown. Determining the spatial arrangement of disease foci using molecular genetic analyses will provide tools for increasing our understanding of disease dynamics. Genotyping *Y. pestis* isolates collected from geographically distinct areas that experience epizootic plague cycles may

help us better understand the spatial dynamics of plague and the mechanisms behind epizootics. The objectives of this study focused on utilizing *Y. pestis* molecular diversity to identify epizootic sources of human plague infections, and to determine whether molecular typing of *Y. pestis* strains corresponds to purported geographic foci or hypotheses of epizootic spread in the Southwestern U. S. and in Kazakhstan, the hypothesized origin of *Y. pestis*. Isolates of *Y. pestis* were chosen from the collection at the Centers for Disease Control and Prevention in Fort Collins, CO and were collected during human plague case investigations, animal-based surveillance activities, and during an ecological study of *Cynomys ludivicianus* (black-tailed prairie dog) epizootics at the Pawnee National Grasslands during 2004, 2005, and 2006. Samples were chosen in clusters representing three geographic scales, and were analyzed to identify isolates that arose from the same epizootic source. The three geographic scales represented a close scale, an intermediate scale, and a distant scale. The molecular markers and the associated analyses used were variable number of tandem repeats (VNTRs) and multi-locus variable number of tandem repeat analysis (MLVA) respectively. The close geographic scale analysis consisted of *Y. pestis* isolates obtained from infected humans, and from environmental sources such as fleas and rodents collected during epidemiological investigations associated with these human cases. The objective of initial analyses was to determine whether MLVA could identify the epizootic event responsible for each human infection. Bacterial isolates collected from distinct clusters of human plague cases in New Mexico during a 1980s epidemic, and isolates collected from predefined foci in the central-Asian country of Kazakhstan were used to make inferences about *Y. pestis* population structure on intermediate scales. The distant

geographic scale analyses were carried out using isolates from the three *Y. pestis* biovars, antiqua, medievalis, and orientalis, and the samples compared were from the U. S. and Kazkakhstan. The last part of this study consisted of identifying single nucleotide polymorphisms (SNPs) in the *Y. pestis* genome. Molecular markers like SNPs are more stable than repeat elements and will provide a more adequate genotyping system for phylogenetic studies of *Y. pestis* than the current MLVA marker system.

We determined that MLVA was useful for inferring *Y. pestis* isolate relationships on small and large geographic scales but was not useful for determining the population structure of epizootic foci. Molecular diversity detected using MLVA was extremely useful for the identification of epizootic sources of human plague infections, with isolates from the human and infective epizootic source inferring genetic relationships with greater than 70 % jackknife (JK) support in the phylogenetic trees. Biovars antiqua, medievalis, and orientalis also formed monophyletic clades when rooted with ancestral *Yersinia pestoides* isolates.

Our intermediate scale analyses using MLVA were inconclusive. The stochastic variability generated by the mutation rates of the VNTR markers at intermediate scales made it impossible to determine the population structure and the arrangement of epizootic plague foci in the U. S. and in Kazakhstan. To address this problem, DNA microarrays were used to discover SNPs. These markers are commonly used in eukaryotic population studies because they have slow mutation rates, are distributed evenly throughout the genomes of most organisms, and are good genetic markers for testing hypotheses about hierarchy in populations. Our *Y. pestis* set chosen for SNP discovery represented isolates from two port cities in California and fifteen Colorado isolates from pre-defined plague

foci. Colorado isolates were chosen from the foothills, the foothills/prairie interface, and the PNG to determine population structure among the three areas. Sixty-nine SNPs identified separate *Y. pestis* populations on the eastern and the western PNG. Isolates from the mountains and one isolate from the eastern plains were similar but weakly supported in phylogenetic analyses.

Jennifer Lynne Lowell
Biology Department
Colorado State University
Fort Collins, CO 80523
Spring 2007

TABLE OF CONTENTS

Chapter 1

Review of the Literature

1.1 A history of the plague

Yersinia pestis is the causative agent of the plague, a disease that has wreaked havoc on many societies throughout the last two millennia. Three major pandemics have killed millions of people around the world. Although relatively infrequent in their reccurence, human plague cases are still reported world-wide (80, 123). These human cases are caused by each of the three major biovars, Antiqua, Medievalis, and Orientalis, which were defined on the bases of their ability to ferment glycerol and to reduce nitrate (32). Antiqua ferments glycerol and reduces nitrate, Medievalis does not reduce nitrate but ferments glycerol, and Orientalis reduces nitrate, but does not ferment glycerol. It has been hypothesized that each biovar was responsible for one of three major pandemics (32). This original hypothesis stated that Justinian's plague arose around 500 A.D. in Constantinople and was caused by the biovar Antiqua. The Black

Death killed millions of people in Europe during the 14th century and was caused by biovar Medievalis. The last and ongoing pandemic, which began in the Yunan Province, was caused by the biovar Orientalis (2, 32). This biovar has spread around the world, including to North America via steam-shipping routes in the late 1800s. Recently a fourth biovar, Microtus, was defined, but this biovar is avirulent in humans (113). Estimated human death tolls from the three major plague pandemics exceeded 30% of the human population in affected regions, but the most important demographic effects of the plague were the multiple and insistent epidemic cycles that followed. Between the years 544 and 700 A. D., and 1361 and 1480 A. D., plague epidemics occurred in 2-5 year cycles and killed 10-15% of the population in each epidemic area. These cycles led to chronic depopulation into the 17th century (94).

Today we see relatively little human plague, but the same fluctuating disease cycles are evident in wild rodent populations world-wide where *Y. pestis* has established in nature. Antiqua, Medievalis, and Orientalis circulate in central-Asian rodents where *Y. pestis* evolved (2). Orientalis is the only biovar present in the U. S. where it has circulated in rodent populations for approximately 100 years. Regardless of the biovar, with the exception of Microtus, *Y. pestis* remains extremely virulent, exhibits widespread epizootic behavior, and occasionally infects humans. Modern day hygiene and medical technology will most likely prevent human epidemics of historical proportions, but the movement of *Y. pestis* through rodent populations will continue to present a route for human infection.

1.2 Plague in the U.S.

Currently *Y. pestis* circulates in U.S. rodent populations and their fleas (Fig. 1.1). Human risks are greatest in certain regions in the southwestern states that are thought to have exceptionally high epizootic activity. The heaviest activity is seen in the Four-Corners region extending southeast to Santa Fe and Albuquerque, NM where the frequency of human cases increases following wet winters and during cool summers like those seen during certain El Nino Southern Ocillation (ENSO) years (38). In Colorado, the PNG and the Rocky Mountain front range maintain fairly consistent enzootic/epizootic cycles but these also seem to be influenced by ENSO events (115). Black-tailed prairie dogs that reside on the PNG in northeastern Colorado, and chipmunks (*Tamias* spp.) and ground squirrels (*Citellus* spp., *Spermophilus* spp.) that reside in the mountainous areas in north-central Colorado experience epizootic plague cycles followed by high mortality rates. Rodent mortality caused by *Y. pestis* has been estimated to be as high as 99% in prairie dogs (28). Ground squirrels and woodrats are also highly susceptible to *Y. pestis* infection (78, 99). Chipmunks experience massive die-offs (CDC unpublished data) and were recently identified as important species for high peridomestic plague risk to humans in Colorado (Lowell dissertation chapter 3, unpublished data). In spite of its persistence in mammal populations throughout the world, including the U. S., surprisingly little is understood about how plague is maintained in nature and how *Y. pestis* spreads through host populations (46).

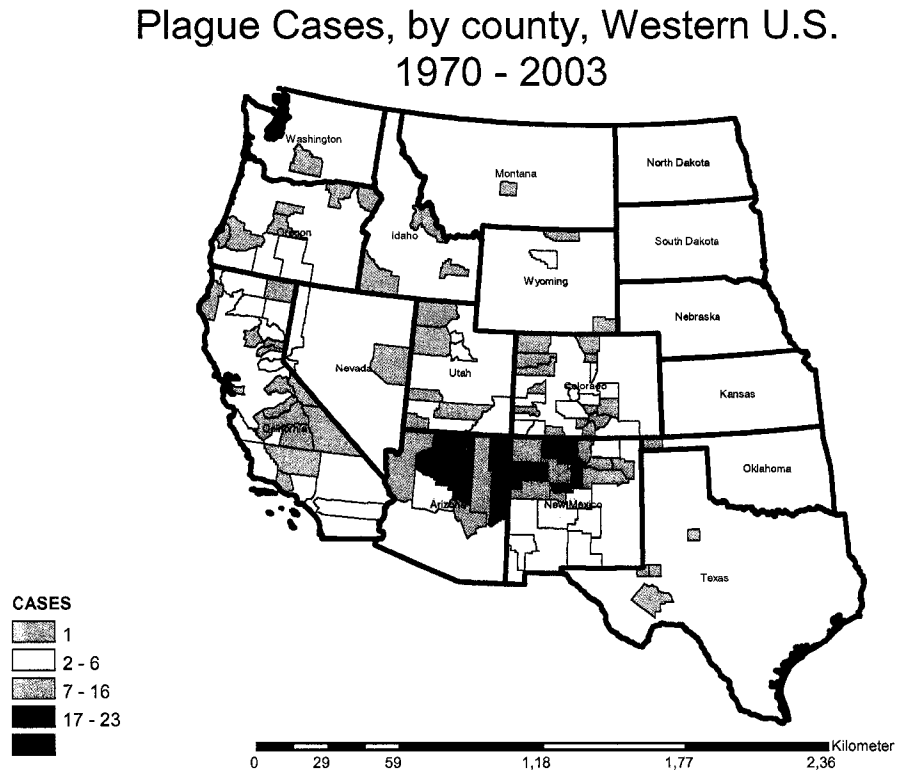

Figure 1.1. Western U. S. map of plague positive counties.

1.3 *Yersinia pestis* in nature

1.3a Definition of disease foci. Zoonotic diseases generally inhabit geographical regions that are favorable to the survival of the pathogen. Parts of the world where plague occurs have historically been divided into geographically distinct regions referred to as foci. Unfortunately, the criteria for defining the extent of foci have varied greatly in the past. Plague foci were defined by their differing ecological characteristics, such as plant communities, elevation, and the rodent and flea species residing in areas that experience epizootic plague cycles on a regular basis. The distribution of human cases and even political boundaries have also been used to define foci. Plague foci have not been defined by bacterial population structure, which might be a more objective way to define focus boundaries. In the former Soviet Union (FSU), 43 natural plague foci have been identified (5). Foci were first defined based on geographical distribution and then further subdivided by the primary rodent species that experienced plague epizootics in each area. Zhou (2004) defines Chinese natural plague foci as the geographic regions where the pathogen has persisted, coexisted, and caused plague epizootics in the range of a particular animal host. Soviet and Chinese definitions suggested that a "primary" rodent species acts as the major reservoir in each focus but in the U. S. plague foci were defined by epizootic hosts with no implication for maintenance (46). In the FSU and China *Y. pestis* from diverse rodent hosts have been categorized as different phenotypes. Isolates differed in nitrate reduction and sugar fermentation capabilities, amino acid utilization, and water-soluble protein profiles. These properties correlated somewhat with the geographic areas from which

the isolates were collected and were used to help define plague foci (5, 77, 126). Although plague foci have historically been well described in this manner, data are lacking that identify the geographic scale on which *Y. pestis* bacteria circulate in nature. It is not known how far clonal lineages travel across landscapes, or if geographic or biological barriers play a role in population segregation and adaptation of *Y. pestis*. While there is some evidence based on biochemical tests that *Y. pestis* exhibits geographic variation, which might have resulted partially from adaptation to certain rodent hosts used to define FSU and Chinese plague foci (5, 126), sample sizes are small and stable genetic markers are lacking. It is difficult to determine how far a bacterial clone can travel during an epizootic and there is no evidence that the same bacterial clone is responsible for epizootics spanning a predefined focus. Currently, plague foci are based on the ecology of each area. The genetic population structure of *Y. pestis* may define foci on a different scale.

1.3b Establishment and epizootic spread. Our understanding of *Y. pestis* maintenance in nature is limited. The spread of *Y. pestis* across expansive land areas is largely based on speculation gained through observation of animal and human behaviour. Plague epizootics and the spread of *Y. pestis* over vast geographic areas probably depends on many interrelated factors including *Y. pestis* virulence, rodent host/flea vector interactions and densities, host susceptibility, and abiotic factors such as temperature and moisture that presumably play into whether or not an epizootic can occur (46, 116).

First, *Y. pestis* must be introduced into an area capable of supporting an epizootic, and then it must spread. This may only occur if the area encompasses

susceptible rodent host species, susceptible rodent densities high enough to perpetuate an epizootic, migration of a carrier or infectious species into the population, proper ratios of susceptible to infectious species, (29, 116), ideal soil moisture and temperature in the rodent burrows (Savage 2006, unpublished data) for long term off-host flea-vector survival (for possible maintenance of the pathogen), and virulence of the plague bacterium (38, 45, 46). If these factors are favourable for the establishment and epizootic cycling of *Y. pestis*, how does it spread?

Plague is transmitted cyclically between rodent hosts and their fleas, or by direct contact with infected tissues (46, 95). Within a relatively small geographic area like a prairie dog colony, *Y. pestis* transmission occurs through an exchange of fleas within nesting areas, or through social or foraging contact of the rodents. As animals within a colony die, the flea densities on the remaining susceptible hosts increase. Higher loads of presumably infectious fleas may speed transmission throughout the remaining animal population (36). Following a rodent die-off, off-host fleas may maintain a transmissible infection while residing in empty rodent nests. Subsequently, *Y. pestis* is transmitted to susceptible rodents that recolonize the empty burrows (6). If the new susceptible rodents reach sufficiently high numbers before *Y. pestis* transmission begins, this population will begin a new epizootic cycle. Alternatively, if no new susceptible hosts are introduced into the population and conditions are not favorable for off-host infectious flea survival, the epizootic will die out and plague will disappear from the area.

On larger geographic scales, *Y. pestis* may spread to adjacent rodent populations or colonies by normal migration patterns, dispersal of young, or during food acquisition.

Mass migration or dispersal covering greater distances has also been observed in the wake of natural disasters, such as fire, floods, food shortages, or overpopulation (95). Scavengers may also contribute to the spread of *Y. pestis* over large geographic distances by acquiring infective fleas while feeding on carcasses (47). Carnivores are resistant to plague and can transfer infected fleas to new areas without falling ill (95).

Several cases of humans traveling during the *Y. pestis* incubation period have been documented (33, 74, 81). One patient was infected in south-central Colorado, and traveled to Tucson, AZ where he died. Two humans infected in Santa Fe, NM in traveled to New York City where one fell seriously ill. Although there was no subsequent transmission, both situations presented scenarios where humans carried disease to distant geographic areas.

Domestic animals that roam can acquire *Y. pestis* from eating infected carcasses or through fleas that are seeking alternative hosts (47). They can also relocate infected fleas to a human's home site providing an infectious source where one was not previously present. The patient infected in south-central Colorado was presumably infected through aerosol droplets from an infected domestic cat that he attempted to remove from under a house (33).

Large geographic spread of *Y. pestis*, such as over or between continents has undoubtedly been facilitated by humans. Rodent to rodent contact, rodent to carnivore contact, or other mammalian contact apparently caused a chain of transmission events across large land masses such as Kazakhstan, China (5, 77, 82, 126), and across oceans to other countries, including the U.S. Humans were responsible for carrying plague infected rats across oceans and for the rapid spread seen between highly populated cities

during large pandemics. A lack of understanding of plague etiology (95), fear, and the attempt to escape mass human die-offs surely contributed to pandemic spread. Humans probably carried infected rats and their fleas with them as they fled epidemics, only to establish disease at their destination points. Pandemic spread by humans can be exemplified by the spread of biovar Orientalis via shipping routes in the late 19th century.

1.3c Plague maintenance theories. Although *Y. pestis* has successfully established over large geographic areas, there are no definitive explanations as to how this pathogen is maintained in nature. Fluctuating epizootic/enzootic cycles and high mortality rates in rodents suggest that survival of the pathogen is dependent on a mammalian maintenance host (28) or infective off-host fleas that survive in rodent burrows (11, 67). There is also speculation that *Y. pestis* survives in soil, infecting burrowing rodents as they forage (35). Whatever the mechanism, it is one that allows epizootics to occur over large geographic areas and to persist over many years.

1.4 The evolution of *Y. pestis*

Although *Y. pestis* is a recently evolved species (1,500-20,000 years ago), and has >90 % genome identity with its most recent common ancestor *Y. pseudotuberculosis* (2); *Y. pestis* ecology, epidemiology, and plague etiology are markedly different from that of *Y. pseudotuberculosis* (2). *Yersinia pseudotuberculosis* causes mild enteric disease and mesenteric lymphadenopathy and *Y. pestis* causes extreme lymphadenopathy, followed by septicemia, pneumonic disease, disseminated

intravascular coagulation (120), and frequent death. The route of transmission has also evolved from a water and foodborne illness to a vector-borne illness. As a consequence, *Y. pestis* has rapidly adapted from an enteropathogenic pathogen that can survive in water and soil, to a blood-borne pathogen that can survive in an insect host vector (flea). There is evidence that *Y. pestis* cannot readily survive as a free-living organism in the environment. Survival in the flea vector is probably one of the most striking phenotypic differences seen in *Y. pestis* when compared to its ancestor *Y. pseudotuberculosis*. Although *Y. pestis* can be transmitted by direct contact with infected tissues or through aerosols, it is mechanistically competent for transmission by a flea vector, and currently the flea vector is hypothesized as the major route of infection between mammals (36, 45, 56, 96). This drastic change in the *Y. pestis* phenotype compared to that of *Y. pseudotuberculosis* has primarily been explained by the acquisition of plasmid DNA and several chromosomal virulence factors through horizontal gene transfer (HGT) (91).

Yersinia pestis contains one plasmid (pCD1) that is common to the Yersiniae, and two unique plasmids encoding for proteins that partially explain the virulence and vectorborne lifestyle of *Y. pestis*. Briefly, pCD1 encodes for the *Yersinia* outer proteins (YOPS) and the type III secretion system. These factors are involved in antiphagocytosis and other mechanisms responsible for evasion of the mammalian immune system. Two plasmids are unique to *Y. pestis*. The first is pMT1 encoding for murine toxin (phospholipase D) and F1 capsular antigen. The second is pPCP1 encoding for plasminogen activator and coagulase activity. Murine toxin and plasminogen activator are critical components for survival in the flea and for

dissemination of the bacteria from the flea bite-site respectively (56, 112). The F1 capsular antigen is thought to be important in antiphagocytosis (121). Chromosomal virulence factors include the *pgm* locus, containing the high pathogenicity island (HPI) and the pigmentation system containing the heme storage locus (*hms*) (15). The HPI encodes for several iron scavenging systems and *hms* encodes for proteins causing biofilm formation. Biofilms of *Y. pestis* "block" the flea vector and facilitate transmission to the mammalian host by certain flea species (58). The HPI and *hms* loci are not unique to *Y. pestis* and this region was probably acquired by *Y. pseudotuberculosis* through HGT prior to *Y. pestis* speciation. Many other enteric pathogens contain the HPI (15).

1.5 Prokaryotic genetic diversity

Generally prokaryotes have been considered ubiquitous in nature and unaffected by limited dispersal because of their small size, abundance, and metabolic plasticity (43). Until recently, prokaryotes have been viewed as organisms that undergo little genetic exchange and contain low genetic variation (43, 108). These views have been challenged in the last ten years with 16S rDNA analyses, evidence for horizontal gene exchange in several medically important bacteria (88, 110), and with genetic diversity detected by complete genome sequencing (http://www.ncbi.nlm.nih.gov/). Carl Woese (1980) discovered secondary structure in 16S rRNA and Archaea, redefining the tree of life, and analyses of 16S rDNA has allowed researchers to discern species diversity in uncultivatible microorganisms from many different environments including soil, water, sediments, and animals (30, 90).

The discovery of diversity in prokaryotes has allowed large numbers of biogeographical and population studies on pathogenic bacteria commensal to man, such as *Escherichia coli*, *Haemophilus influenza*, *Niesseria meningiditis*, and *Staphylococcus aureus* (89, 108). The results from these studies suggested that bacterial populations were in linkage disequilibrium. Populations showed a lack of recombination, and therefore were comprised of a limited number of clones with worldwide distribution. Undoubtedly these pathogens' lifestyles and necessity to remain with their human hosts meant that their global distributions, clonal epidemic structures, and resulting linkage disequilibrium were largely affected by human activity.

Bacterial biogeography studies that focused on local populations of free living bacteria showed that these populations were also in linkage disequilibrium on worldwide scales, but approached linkage equilibrium on local scales (63, 114). A study by (21) used several genotyping techniques and spatial statistics to determine the population structure of fluorescent *Pseudomonas* strains collected from soil in six regions on four continents. Their study indicated that the genetic diversity of those microbes increased with geographic distance. Another study on the hyperthermophilic Archaea *Sulfolobus sulfataricus* showed that genes encoding for a variety of putative cellular functions resulted in large, significant F_{st} values (122). These studies indicated that the bacterial lifestyle is important in determining population structure and distribution. Commensals tended toward cosmopolitan structure, while free-living bacteria showed endemic structure. Free living pathogenic bacteria and those infecting non-human hosts also contain genetic diversity allowing population differentiation. Pathogenic bacterial housekeeping genes *recA* and *glnA* of *Vibrio vulnificus* (52) and

variable number tandem repeats throughout the genomes of *Bacillus anthracis* and *Y. pestis* (3, 69) have been used for genetic studies that grouped bacteria geographically on several scales. But none of these environmental pathogens have been analyzed for linkage disequilibrium or clonal structure. In spite of these recent findings, ecological population genetics and phylogeographic studies for zoonotic pathogens are scarce, and theories stating that epizootic bacterial dispersal is not limited by geographic barriers must be challenged.

1.6 *Yersinia pestis* genetic diversity

Because *Y. pestis* is a recently evolved bacterium, researchers traditionally believed that the genome did not contain sufficient within species variation to conduct population genetics or phylogeographic studies. Despite the relatively recent divergence of *Y. pestis*, genotyping efforts have identified genetic and phenotypic variation among *Y. pestis* isolates. Several historic studies revealed phenotypic differences in isolates associated with certain rodent hosts, and differential protein variants were identified in ancient foci world-wide. These included isolates with differing sugar fermentation capabilities and auxotrophic amino acid properties (77). Phenotypic differences compelled Russian investigators to group *Y. pestis* into subspecies that corresponded to the geographic regions and the specific hosts from which the isolates originated (5). Subspecies loosely corresponded to these broadly defined geographic foci (7, 82).

Achtman (2004) tested hypotheses of *Y. pestis* biovar hierarchy by comparing world-wide isolate collections. Consensus trees were generated using *Y. pestis* sugar

fermentation capabilities, IS100 elements, and SNPs. The results were informative for biovar relationships but these differences were not informative for studies on smaller geographic scales that compared isolates within biovars. Zhou (2004) identified 14 unique *Y. pestis* types termed "genomovars" based on difference regions (DFRs) identified with microarray analyses. The DFRs consisted of insertions and deletions in defined dynamic regions of the *Y. pestis* genome in natural populations. Based on these 14 genomovars, the authors deduced the expansion of *Y. pestis* across China in a hierarchal manner and identified unique foci specific to each genomovar.

Ribotyping (restriction fragment length polymorphism analysis followed by IS*100* and IS*285* element hybridization), pulse-field gel electrophoresis (PFGE) profiles (61), IS*100* element and glycerol-3-phosphate dehydrogenase (*glpD*) polymorphisms (84), and variable number tandem repeat (VNTRs) polymorphisms (69) are all evident between *Y. pestis* isolates. Huang (2002) used ribotyping and PFGE to differentiate between geographically isolated North American isolates from the southwestern U.S. All methods showed variation within the isolates. The most variability was demonstrated by PFGE (five types contained 26 subtypes). IS*100* typing showed 16 genotypes, IS*285* typing showed four types, and ribotyping only identified one genotype. Although the five major types identified by PFGE analysis did not group the isolates by geographic area, the authors concluded that PFGE generally correlated genetic variability to the geographic distance between the isolates. The individual genotypic subtypes within these groups often consisted of isolates collected within close proximity to each other (61).

Motin (2002) identified shifts in IS*100* elements throughout the genomes of the three biovars, *Y. pestoides*, and *Y. pseudotuberculosis* from a worldwide isolate collection. Sixteen distinct IS*100* insertions discriminated isolates between biovars and in some cases within biovars. This method had limited discriminatory power within Orientalis and could not distinguish Orientalis isolates from different regions. One exception was an Orientalis isolate from Indochina, which differed from other worldwide Orientalis isolates. Medievalis and Antiqua isolates differed by geographical origin, with the exception of one Antiqua isolate that was identical to an Orientalis isolate and one Antiqua isolate that was identical to a Medievalis isolate. Overall, this method was effective for discriminating biovars, and broad geographic scale differentiation within biovars.

The completion of several *Y. pestis* genome sequences has provided great insight into the genetic diversity within and between the three biovars (20, 31, 91, 113). Most recently, (20) compared the genomes of CO92 (Orientalis), KIM (Medievalis), 91001 (microtus) and the two recently sequenced Antiqua isolates, Antiqua (Congo) and Nepal516 (20). These authors came to the conclusion that since its divergence from *Y. pseudotuberculosis*, *Y. pestis* has undergone reductive evolution through loss of gene function. Many genes that supported an environmental lifestyle, such as those that encoded for flagellar and chemotactic properties, have been inactivated through a variety of mechanisms. These consist of insertions and deletions, frameshifts, non-synonymous SNPs (nsSNPs), and interruptions by insertional elements. Functional losses probably occurred as *Y. pestis* made a transition from an orally and fecally transmitted organism, to that of a flea-borne pathogen.

Insertional elements were of great interest in this comparison because they appear to remain active. Additionally, IS*100* elements have successfully been used for typing and grouping biovars in the past (1, 84). Chain (2004) examined the locations of the IS*100*, IS*1541*, IS*285*, and IS*1661* elements, and determined how many were shared by the five isolates that are now fully sequenced, and which were unique. While 12 IS elements shared by the five genomes, and *Y. pseudotuberculosis* were previously identified (19), (20) identified 45 IS*1541*, 15 IS*100*, 11 IS*285*, and six IS*1661* that were shared by the five sequenced *Y. pestis* isolates but not by *Y. pseudotuberculosis*. These elements may be indicative of those present in the most recent common *Y. pestis* ancestor of the sequenced strains.

Subsets of shared and unique IS elements and synonymous SNPs (sSNPs) yielding congruent trees have led to phylogenies that contradict previous hypotheses stating that Antiqua is ancestral to Medievalis and Orientalis (2, 32). The recent hypothesis elucidated phylogenetic relationships placing the two Antiqua isolates clearly on two separate branches rather than on one branch ancestral to Medievalis and Orientalis. Nepal516 (Antiqua) and KIM (Medievalis) share one branch, while Antiqua (Congo) and CO92 share a different branch (1, 20). This phylogeny is only based on five samples. Additional sampling of Antiqua isolates from central-Asia, where *Y. pestis* is thought to have existed the longest, might reveal a paraphyletic group with the other biovars forming a nested monolphyletic group.

Diversity in the *Y. pestis* genome has also been examined using highly variable repeat regions. Adair (2000), Klevytska (2001), and Girard (2004) identified 43 VNTRs in *Y. pestis*, and demonstrated that these markers were capable of

discrimination between the three biovars. It was later demonstrated that these repeat regions were also extremely useful in epidemiological investigations linking environmental flea and rodent isolates to human plague cases (74). Unique clusters of *Y. pestis* genotypes among prairie dog colonies during plague epizootics were also identified using VNTRs (49). Many of these markers have relatively rapid mutation rates (49) that make them useful in close geographic scale isolate comparisons for identifying isolates that potentially arose from the same source during an epizootic (74, 75).

Most of the genetic diversity used to type *Y. pestis* is chromosomal, but different plasmid profiles have been noted in isolates from central Asia, Brazil, and the U. S. (23, 34, 42, 71). Only those from central Asia corresponded to the region from which they were collected (34, 42). Isolates collected from the Tien Shan mountain focus in Kyrgystan bordering Kazakhstan, were lacking the pPCP1 plasmid. Plasmid loss has been documented in the laboratory (71), but the central-Asian isolates were naturally missing pPCP1 suggesting adaptation of a phenotype capable of a transmission route different than the flea vector (55). Plasmid profiles are not typically used for phylogenetic studies, but are included in isolate characterizations for strain collections (75).

These observations and molecular techniques have provided invaluable insight into *Y. pestis* diversity and evolution, however studies that correlate *Y. pestis* genotypes to enzootic/epizootic fluctuations are lacking. High-resolution bacterial strain typing methods are needed to better define disease dynamics and spatial arrangement of natural disease foci.

1.7 Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are single base changes that have been identified throughout the genomes of many eukaryotic organisms. In the human genome SNPs occur in approximately 1 out of 1,000 bp (103, 120) and in most other sequenced eukaryotic genomes in about 1 out of 300-1000 bp (13, 103). It has been hypothesized that SNPs are better than repeat motifs for estimating genetic distance (Fst values) because they represent a more complete sample of the entire genome and reduced interlocus sampling variance (4). It is also believed that SNPs are more stable than other molecular markers such as repeat motifs eliminating much of the potential for homoplasy in phylogenetic inference (4).

While MLVA has revealed genetic variability useful for epidemiologic studies on small and large geographic scales (49, 69, 74, 75) phylogeographic inference on intermediate scales is not easily resolved using current VNTR markers (75). SNPs are expected to be useful for studying the population dynamics of *Y. pestis* to delineate populations and answer questions about the movement of *Y. pestis* in the environment.

1.8 Conclusion

The last six years have brought a deluge of information concerning *Y. pestis* genetics, diversity, and evolution. However, the molecular markers identified in *Y. pestis* don't provide the tools necessary for bacterial population analyses that identify epizootic borders. The high mutation rates associated with MLVA markers introduce stochastic variation into the population making it impossible to identify clonal structure on geographic scales that currently represent epizootic foci. Markers such as IS

elements do not provide enough diversity to define epizootic structure. The following projects describe the usefulness of MLVA and repeat markers on three spatial scales, and the discovery and utility of SNPs in North American *Y. pestis* isolates.

1.9 Objectives/Hypotheses

Objective 1: To determine if MLVA can link human *Y. pestis* isolates to the associated environmental source *Y. pestis* isolates collected during human plague case investigations (close geographic scale).

Hypothesis 1: MLVA and the associated 17 VNTR markers that mutate the fastest will infer positive genetic relationships between human plague case *Y. pestis* isolates and clones of the infective *Y. pestis* source collected from the environment.

Objective 2: To identify clusters of human plague cases from the 1980s in New Mexico as genetically distinct disease foci using MLVA (intermediate geographic scale).

Hypothesis 2: MLVA and the associated 17 VNTR markers will identify separate *Y. pestis* isolate populations corresponding to plague case clusters occurring in the 1980s in New Mexico.

Objective 3: To determine if MLVA can discriminate *Y. pestis* biovars Medievalis and Antiqua collected from Kazakhstan from biovar Orientalis collected from North America (distant and intermediate geographic scales).

Hypothesis 3: MLVA and the associated 42 VNTR markers will infer relationships that discriminate between North American and Kazakh groups of *Y. pestis* isolates, and between isolates collected from predefined Kazakh plague foci.

Objective 4: To identify SNPs in the *Y. pestis* genome for use in population genetics and epidemiologic studies of plague.

Hypothesis 4: SNPs will be abundant in the *Y. pestis* genome and will provide population level genetic diversity suitable for discerning *Y. pestis* population dynamics on multiple geographic and temporal scales.

Objective 5: To determine whether plague epizootics on the PNG are a result of infection being carried from the front-range foothills.

Hypothesis 5: If plague outbreaks occur in waves of infection from the foothills, isolates collected from the PNG and the foothills during the same year will show synapomorphic relationships.

Chapter 2

Identifying sources of human exposure to plague

This chapter was published in *The Journal of Clinical Microbiology*. The text of this manuscript is presented as it appeared in this journal. All figures that appeared in this manuscript are presented exactly as they appeared in this journal. The reference for this chapter is shown below.

Lowell, J. L., Wagner, D. M., Atshabar, B. Antolin, M. F., Vogler, A. J., Keim, K., Chu, M. C., and K. L. Gage. 2005. Identifying sources of human exposure to plague. J. Clin. Microbiol. Feb; 43: 650-656.

2.1 Abstract

Yersinia pestis, the etiologic agent of plague, has shaped the course of human history, killing millions of people in three major pandemics. This bacterium is still endemic in parts of Asia, Africa, and the Americas, where it poses a natural disease threat to human populations, and *Y. pestis* has also recently received attention as a possible bioterrorism agent. Thus, rapid methods to distinguish between bioterrorism and naturally occurring plague infections are of major importance. Our study is the first to demonstrate that variable number tandem repeats (VNTRs) in the *Y. pestis* genome can link human case isolates to those obtained from suspected environmental sources of infection. We demonstrate the valuable utility of VNTR markers in epidemiological investigations of naturally occurring plague and the forensic analysis of possible bioterrorism events.

2.2 Introduction

Plague, which is caused by the bacterium *Yersinia pestis,* has wreaked devastation around the globe, killing millions of people in three major disease pandemics. Natural transmission of plague to humans remains a possibility in many regions of the world where foci exist in sylvatic rodent populations (9). Approximately 3000 human cases occur world-wide annually, with 12-15 cases reported each year in the western U.S (123). *Y. pestis* has also been identified also as a potential bioterrorism agent (62), and the threat for bioterrorism or biocrimes, combined with the continuing occurrence of natural outbreaks, emphasizes the need for methods for differentiating victims of deliberate exposures from those who become infected from natural sources (68). One of the primary objectives of routine epidemiological plague investigations is to identify the source of human exposure and to assess the exposure site for potential continuing risk. This objective is sometimes difficult to meet when more than one epizootic source exists or when a patient's history is ambiguous. Despite its epidemic potential, outbreak investigations and prevention efforts are often hampered by our limited knowledge about how *Y. pestis* spreads through host populations, and by a lack of methods for unambiguously identifying individual exposure sites, local sources of infection, and local populations of bacteria. The use of molecular epidemiological techniques in these investigations has been particularly difficult for *Y. pestis* because of its apparent lack of genetic variation (2). *Y. pestis* is currently grouped into three biovars (32), and while previous genotyping techniques are efficient for biovar identification, detection of genetic

variability within biovars has not been consistent (51, 64, 84, 97). Furthermore, a lack of high-resolution bacterial strain-typing methods has made molecular epidemiology and surveillance of *Y. pestis* difficult.

The completion of the first *Y. pestis* genome sequence (91) revealed DNA repeats that have the potential to identify variability among plague isolates on small geographic scales, and mutation rates of these DNA repeats has provided additional information on the feasibility of using these markers to identify genetically similar *Y. pestis* isolates on a local scale (49). This information has led to development of a highly effective typing system for use in molecular epidemiology and forensic analyses (69). We show the applicability of 17 multiple locus variable number tandem repeat (VNTR) markers (MLVA) (3, 69) to the molecular epidemiology and identification of environmental infection sources for human plague cases. When combined with epidemiological information, the analysis of these highly mutable VNTR markers (69, 91) enabled us to identify exposure sites and likely environmental sources of infection for past human plague cases, including a highly publicized case that occurred in New York City (NYC) in November 2002 (50).

2.3 Materials and Methods

2.3a Isolate selection. We examined 13 sets of *Y. pestis* isolates collected during epidemic investigations conducted in New Mexico in the early 1980s, and in New Mexico, Arizona, and Colorado in 1992, 1996, 1999, 2001, and 2002. Three sets, used as positive location controls, consisted of paired isolates collected from different fleas or

hosts but at the same time and location (Table 2.1a). Positive control pair A was collected from an antelope ground squirrel and a flea removed from this animal, control pair B was collected from fleas found in neighbouring burrows in the same prairie dog colony, and control pair C was collected from fleas in the same rodent burrow. A fourth isolate set served as a negative location control and consisted of two isolates collected in the same year but at separate sites located approximately 300 km apart (Table 2.1a). The nine remaining sets of isolates were collected during plague case investigations in which isolates were obtained from both human patients and associated environmental samples from fleas and other mammalian hosts (Table 2.1b). Isolates obtained from other mammals or fleas during each plague case investigation were identified and were compared genetically to the corresponding human isolate. A *mediavalis* biovar isolate from Kazakhstan was included in the phylogenetic analyses as an outgroup.

2.3b Selection of VNTR markers. A subset of 17 VNTR markers was selected from the 43 VNTR markers described for *Y. pestis* (49, 15). The most polymorphic markers were selected because they are considered more effective for forensic analysis and for identifying genetic similarity on small geographic scales (69). In *Y. pestis*, those markers with the highest number of the repeated motif copies show the highest degree of polymorphism across isolates tested (69) and some of the highest mutation rates *in vitro* (49).

Table 2.1a. Known environmental isolate pairs used as location controls.

	Location	Source of Paired Isolates/Accession number		
		First of pair	Second of pair	Genetic Match y=yes n=no
Control A	Sandoval Co, NM	*Ammospermophilus leucurus* (ground squirrel) NM000293	*Thrassis bacchi* (ground squirrel flea) NM000293-42	y
Control B	La Plata Co, CO	*Oropsylla hirsuta* (prairie dog flea) CO021867-142	*Oropsylla hirsuta* (prairie dog flea) CO021868-143	y
Control C	San Miguel Co, NM	*Thrassis bacchi* (ground squirrel flea) NM8301675-1885	*Oropsylla hirsuta* (prairie dog flea) NM830674-1879	y
Control D	Santa Fe Co, NM (first of pair) Bernalillo Co, NM (second of pair)	Human NM012147	*Oropsylla hirsuta* (prairie dog flea) NM0113239-539	n

Table 2.1b. Human cases and associated environmental isolates collected during case investigations.

Case Letter/ Date	Human Isolate CDC Accession Number/ Case Location	Distance of Environmental Isolate(s) from Potential Exposure Site	Environmental Isolate(s) Source	Environmental Isolate(s) CDC Accession Number	Genetic Match y=yes n=no
Case A 11/2002	NM024452 Santa Fe Co, NM	Residence yard	*Orchopeas sexdentatus* (wood rat flea)	NM021852-138	y
			O. neotomae (wood rat flea)	NM021856-140	y
			Peromyscopsylla hesperomys (deer mouse flea)	NM024476-306 NM024477-309 NM024479-310	y y y
			Anomiopsyllus nudatus (wood rat flea)	NM024484-315	y
Case B 07/1992	AZ921389 Apache Co, AZ	200 m from residence (site 1)	*Oropsylla hirsuta* (prairie dog flea) (site 1)	AZ921367-360	y
		27 km from residence (site 2)	*Spermophilus variegates* (rock squirrel) (site 2)	AZ921377	n
Case C 1999	NM990061 Santa Fe Co, NM	Patient skinned rabbit	*Sylvilagus auduboni* (rabbit)	NM990030	y
Case D 1992	AZ962456 Coconino Co, AZ	Patient visited prairie dog town (site 1)	*Oropsylla hirsuta* (prairie dog flea) (site 1)	AZ962544-528	y
		Prairie dog town 22. 4 km from site 1 (site 2)	No isolate (site 2)		

Case Letter/ Date	Human Isolate CDC Accession Number/ Case Location	Distance of Environmental Isolate(s) from Potential Exposure Site	Environmental Isolate(s) Source	Environmental Isolate(s) CDC Accession Number	Genetic Match y=yes n=no
Case E 10/1992	CO92 Chaffee Co, CO	1 km from residence	*Tamius quadrivittatus* (Colorado chipmunk)	CO921715	n
Case F 08/1983	NM830692 San Miguel Co, NM	Gravel pit (site 1)	*Oropsylla hirsuta* (prairie dog flea) (site 1)	NM830651-885	n
			Thrassis bacchi (ground squirrel flea) (site 1)	NM8306741-879	n
		4 km from site 1 (site 2)	No isolate (site 2)		
		4 – 8 km from site 1 (site 3)	No isolate (site 3)		
		Roaming dogs contacted patient	No isolates from roaming dogs (seropositive)		
Case G 06/1983	NM830483 Santa Fe Co, NM	300 m (site 1)	*Oropsylla montana* (ground squirrel flea) (site 1)	NM8304881-284	n
		Roaming cats slept with patient	No isolates from roaming cats		
Case H 08/1983	NM830694 McKinley Co, NM	Residence (site 1)	*Oropsylla hirsuta* (prairie dog flea) (site 1)	IJ831816-1920	n
		354 km (site 2)	No isolate (site 2)		
Case I 04/1983	NM830202 McKinley Co, NM	400 m from residence	*Aetheca wagnerii* (deer mouse flea)	NM83-IJ823	n

2.3c DNA extraction and PCR amplification. DNA was prepared from *Y. pestis* isolates using a heat soak method (3). Each 20 µl PCR reaction contained 1 X PCR buffer with 1.5 mM $MgCl_2$, 200µM dNTPs, and 0.5 U *Taq* polymerase (Promega Madison, WI), 1.0 µl of DNA template (approximately 0.5 ng DNA) and one of the following six multiplex phosporamidite linkage dye labelled primer sets: mix 1, 0.1 µM primer M09 and 0.2 µM each of primers M21 and M18, mix 2, 0.1 µM of primer M06, and 0.2 µM M58, mix 3, 0.1 µM of primer M34, and 0.2 µM each of primers M23 and M28, mix 4, 0.1 µM of primer M31 and 0.2 µM of primer M12, mix 5, 0.1 µM of primer M27 and 0.2 µM each of primers M29 and M33, mix 6, 0.1 µM of primer M22 and 0.2 µM each of primers M25 and M59, and mix 7, 0.2 µM M19. Reactions were placed on a PTC-100 thermal cycler (MJ Research, Waltham, MA) at 94 °C for 5 min, 40 cycles of 94 °C for 20 s, 57 °C for 20 s, and 72 °C for 45 s, with a final extension step of 72 °C for 5 min. Following thermal cycling, samples were diluted 1:5 with sterile, DNAase free water for fragment analysis.

2.3d Fragment analysis. PCR fragments were analyzed on a CEQ 8000 DNA capillary sequencer (Beckman Coulter, Fullerton, CA) by adding 1.25 µl of the amplified samples to 39.5 µl of sample loading solution (SLS) (Beckman Coulter) and 0.5 µl of 600 bp D1 dye labelled size standard (Beckman Coulter). Method parameters consisted of a 35 °C capillary temperature, 120 s 90 °C denaturation, 30 s 2.0 kV injection, and 35.0 min 7.5 kV separation. PCR fragment sizes were determined from the raw data using the CEQ 8000 fragment analysis software version 5.0 (Beckman Coulter, Fullerton, CA). After fragment sizes were determined, the number of tandem repeats per allele was calculated

in reference to the published CO92 repeat sizes (69, 91). Repeat numbers were scored as characters for each taxon, and these data were entered into a data matrix to infer relationships among isolates.

2.3e Statistical Analysis. The data matrix containing repeat numbers was entered into PAUP version 4.0b10 (118). A strict consensus tree was generated using maximum parsimony analysis, and jackknife support was determined based on 37% deletion and 500 replications. Isolates that were supported in at least 70% of jackknifed parsimony trees fit our first criterion for inferring a match between isolate pairs. Greater than 70% jackknife support represents a greater than 95% probability of obtaining the correct clade (79).

2.3f Database Query. To place the genetic relationships within our set of isolates in a global context, each of the nine human isolates and one isolate from each of the four control pairs were queried against a large *Y. pestis* DNA collection at the Keim Genetics Laboratory at Northern Arizona University (NAU). Each of the 13 isolates was compared to the 30 other isolates listed in Tables 2.1a and 2.1b, and 632 additional isolates from NAU for a total of 662 pair-wise comparisons for each sample. Each isolate was queried against the database in a non-nested hierarchal design on worldwide, continent, and local scales. The worldwide scale consisted of 346 *Orientalis* isolates collected outside North America, whereas the local scale consisted of 169 isolates, including our 31, from New Mexico, Arizona, Colorado, and Utah (Four Corners region). The continent scale consisted of 147 isolates from various states in the USA, excluding

the Four Corners region (Table 2.2). Pair-wise genetic distances between the each of the 13 isolates of interest (one from each location control pair, and each of the nine human isolates) and the other 662 isolates were generated in PAUP 4.0b10 (118) based on VNTR fragment sizes. These pair-wise distances were converted to the number of marker differences, the average marker difference between samples was 9.6 (99% CI = 8.64, 10.56). Isolates that matched each other (i.e. had very few, or no marker differences when compared) were identified as extreme outliers compared to the lower tail of the dataset, thereby fitting our second criterion for inferring a match between isolate pairs.

2.3g Epidemiologic data collection. Health officials from federal (CDC), state, and local agencies routinely conduct investigations of human plague cases in an effort to identify likely sources of infection and persons who also might be at risk. As part of these investigations, efforts are made to collect samples from potential mammalian hosts and their fleas. These samples are analyzed for *Y. pestis* and bacterial isolates are deposited in the plague strain reference collection at CDC's Division of Vector-Borne Infectious Diseases. Investigators also determine the patients' travel histories and potential exposure sites, the proximity of patients' residences to rodent populations known to be common sources of infection (e.g. prairie dogs), patient recollection of flea bites, direct contact with wild mammals or domestic pets that may have been exposed to a plague epizootic, and other pertinent epidemiologic information. These data were used in addition to statistical analyses as a third criterion for inferring a match between isolate pairs.

Table 2.2. Origins of the 632 isolates queried from the NAU *Y. pestis* MLVA type database, and the 31 isolates typed in this study.

World: Country	**Number of Isolates**	**USA: State**	**Number of Isolates**	**Four Corners: State/County**	**Number of Isolates**
Belgian Congo	1	CA	127	**AZ**	
Bolivia	2	KS	9	Apache	10
Brazil	3	MT	2	Coconino	38
Burma	2	NV	2	Navajo	4
China	4	OR	1	Yavapai	1
Former USSR	1	TX	5	**CO**	
Germany	3	WY	1	Chafee	3
India	3			Denver	1
Indonesia	2			Larimer	7
Madagascar	308			La Plata	3
Namibia	5			Park	1
Senegal	2			**NM**	
South Africa	1			Bernalillo	11
Turkey	2			Cibola	1
Vietnam	7			Harding	1
				Los Alamos	1
				McKinley	3
				Rio Arriba	3
				San Juan	1
				San Miguel	5
				Sandoval	5
				Santa Fe	31
				Unspecified	39
				UT	
				Jaub	1
Total	346		147		170

2.3h Calculation of VNTR mutation rate. An overall mutation rate for the 17 VNTR markers used in this study was calculated based on data from an *in vitro* parallel, serial passage experiment recently described by Girard et al. (49), where mutations observed across ~21,000 *Y. pestis* generations yielded an overall mutation rate of 1.3×10^{-3} mutations/generation for the 43 VNTR markers. Because the markers used in this study are a subset of the 43 used by (49), it was possible to calculate an overall mutation rate of 1.0×10^{-3} mutations/generation for our 17 VNTR loci. Based on the Poisson distribution, the probability of observing *n* mutation(s) is maximized at the number of generations that is equal to the inverse of the rate X *n*. For example, the probability of observing one mutation is maximized at 1,000 generations (95% CI = 26, 5370 generations).

2.3i Transmission cycle estimates. The number of transmission cycles that occurred between some of the human and paired environmental isolates was estimated using a recently described transmission model for *Y. pestis* (49). This model predicts that ~52 *Y. pestis* generations (doublings) occur in a single transmission cycle, which involves a single infected flea passing on a *Y. pestis* infection to a single mammalian host (49). When coupled with a mutation rate estimate for VNTR markers, this transmission model provides predictions of the number of transmission cycles that have occurred between two isolates. For example, the probability of observing two mutations in the 17 VNTR markers is maximized at 2,000 generations (95% CI = 254, 6840 generations), which corresponds to ~38 transmission cycles (95% CI = 5, 132 transmission cycles).

2.4 Results

All positive location controls met our three criteria for inferring a positive match between isolates. First, the most parsimonious trees generated from these samples and 17 markers had three branches that included both isolates from each positive location control (A, B, and C). Jackknifing analysis showed that support for unique pairing of isolates from each area was 98, 93, and 84%, respectively (Figure 2.1). Second, each isolate pair in the positive location controls differed at just one marker compared to its match, did not match any of the other isolates in the NAU database query, and fell well outside the lower limit of the 99% CI for the average number of marker differences. Third, paired isolates were collected from different fleas or hosts but at the same time and location, as would be the case if matching human and environmental isolate pairs were obtained during an epidemiologic investigation. The geographically distant negative location control pair D was not supported in parsimony analyses and fell outside the upper end of the 99% CI for the average number of marker differences for all pair-wise comparisons (11 marker differences). Epidemiologic information was combined with MLVA data from human and environmental samples to verify specific plague exposure sites for each patient. The 2002 NYC plague case (case A) was a high profile case and, because it was diagnosed in an area where plague does not cycle in wild rodent populations, urgent identification of the infective source was needed to rule out bioterrorism. The human isolate matched multiple flea isolates collected near the patient's home in Santa Fe, NM before and after the date when the patient was first exposed (jackknife support 88%).

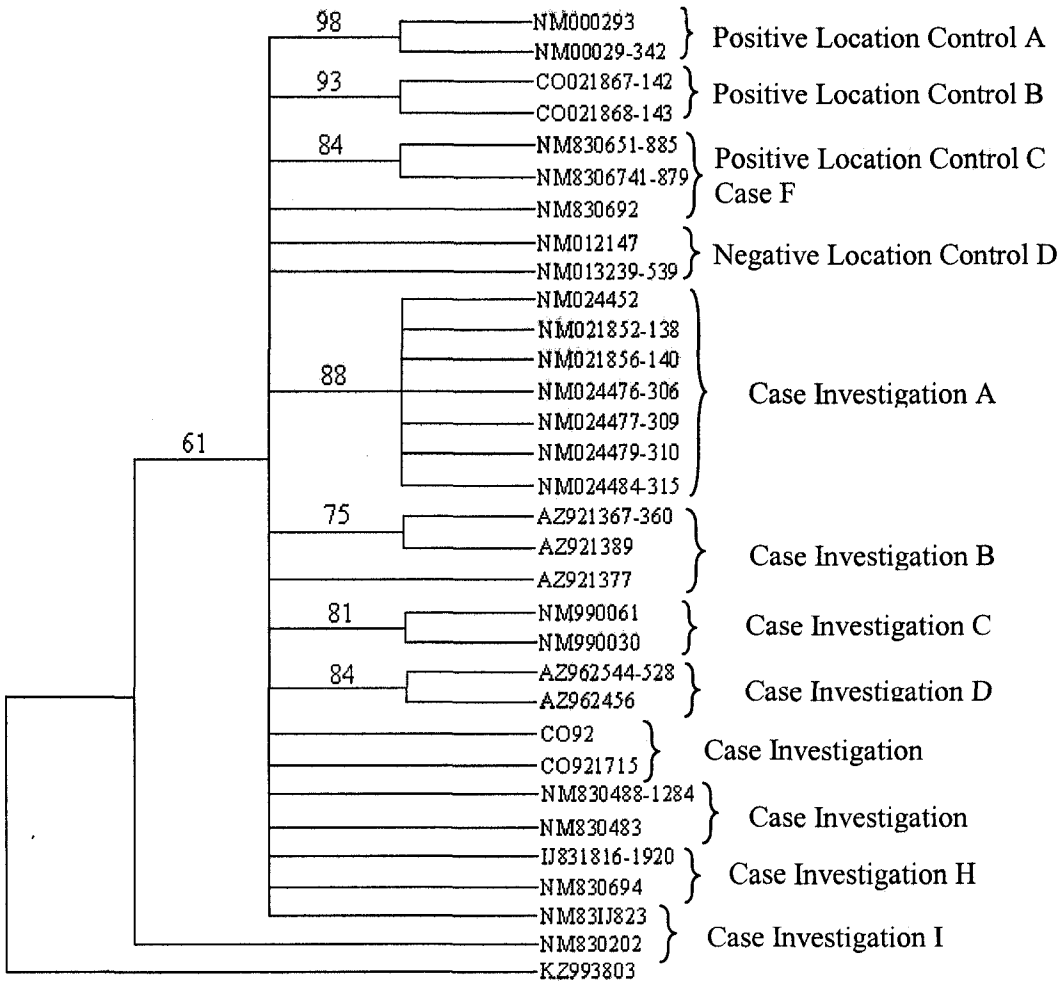

Figure 2.1. Strict consensus tree constructed using maximum parsimony analysis identified genetically similar isolates. Number labels on the tree are accession numbers and may be referenced in Tables 1a and 1b. Jackknife support values are based on 500 simulations. Branches with no numbers had values of less than 50. *NOTE: Isolates NM830651-885 and NM8306741-879 were used in both Positive Location Control C and as environmental isolates for Human Case F.

Interestingly, matching environmental samples included not only *Y. pestis*-infected fleas collected during the follow-up case investigation, but also samples obtained during routine surveillance on the patient's New Mexico property four months prior to the onset of illness (Table 2.1b). MLVA results for this group of samples yielded high identity between the human isolate and the flea isolates collected before and after the case occurred (Figure 2.1). Five of the environmental samples differed at only two markers, while one of the environmental samples differed at three markers. The 2002 NYC human case isolate was also highly dissimilar to isolates collected in surrounding regions during other case investigations and highly dissimilar to isolates queried from the Four Corners, the U. S., and from the world (Figure 2.2). Exceptions were two isolates collected in 1998 approximately 61 km from the 2002 human case and one collected in 1991 in the same county. One of the 1998 isolates also differed from the human isolate by 2 markers, and the other 1998 isolate and the 1991 isolate differed at 3 markers compared to the human case A isolate.

Our MLVA also linked certain human case isolates each with an environmental isolate from a single suspected exposure site even when isolates from more than one exposure site existed. Case B had two known potential exposure sites, one approximately 27 kilometers from the patient's residence where he was collecting wood, and the other 200 meters north of his residence, where a plague epizootic had occurred in prairie dogs and other nearby rodent populations. During the case investigation, several field mice and a rock squirrel (*Spermophilus variegatus*) were collected in the immediate vicinity of the wood collecting site, and a *Y. pestis* isolate was obtained from the carcass of a rock squirrel that died in a live trap. Several fleas were also collected from different types of

rodent burrows near the patient's residence, and *Y. pestis* was isolated from an *Oropsylla montana* flea pool from one of these burrows. When the human isolate from case B was tested against the environmental isolates from the distant wood collecting site and the rodent burrows near the patient's residence, it showed high similarity to the nearby site, with a jackknife value of 75% and only two marker differences. The isolate from the wood collecting site had very little similarity to either the human or the above *Oropsylla montana* isolate (Figure 2.1), differing at 10 markers. In contrast to the above case studies, patient cases C, D, E, F, G, H, and I, had multiple potential exposure sites, but with environmental isolates from only one of the sites. Case C involved a hunter who shot and skinned a rabbit from a plague enzootic area. A tissue sample from the dead rabbit, which was stored in the patient's freezer, yielded a *Y. pestis* isolate that was highly similar to the patient isolate with 81% jackknife support and only one marker difference, well outside of the 99% CI of mean marker differences for the NAU database query. This match verified that the likely infection source was the rabbit, and that it was unlikely that the patient was exposed to *Y. pestis* in other areas where he might have been hunting. The patient in Case D reportedly visited two potential exposure sites approximately 22.4 km apart during a three day period. Prairie dog die-offs, suggestive of plague, were observed at both of these sites and rodents and fleas were sampled from both areas. The epidemiological investigation only yielded positive *Y. pestis* fleas from one of the sites, perhaps because the other site had been affected much earlier by epizootic activity and the burrows no longer harbored infected fleas. Maximum parsimony analysis generated a highly supported clade (84% jackknife) between the positive flea pool isolate and the

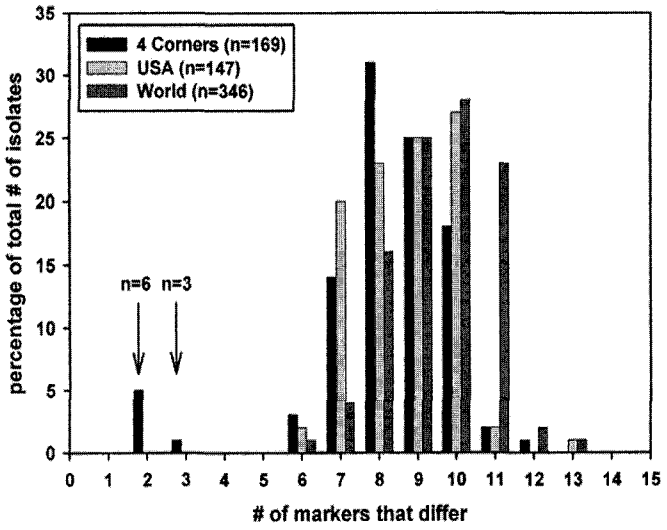

Figure 2.2. This figure shows the case A marker difference distribution when all of the isolates from the NAU database (n=632), as well as the isolates typed in our study (n=30), were compared to the human case A isolate. Isolates showing 2 marker differences to the human case A sample, (n=7) consist of the 6 environmental isolates associated with the case and one 1998 isolate. Isolates showing 3 marker differences to the human case A isolate (n=2) consist of one 1998 and one 1991 isolate. Matching isolates fall in the extreme lower tail of the distribution, while the non-matching isolates fall in the 99% CI, or in the extreme upper tail of the distribution.

patient isolate, with three marker differences. This result provided strong evidence that the infection source was from the area in which the positive fleas were recovered.

Case E represents a situation where epidemiologic evidence clearly indicated a domestic cat as the infective source. The patient presumably became infected while removing this domestic cat from the crawlspace of a friend's home. The sick cat displayed symptoms strongly suggestive of pneumonic plague, and the patient was diagnosed post-mortem with primary pneumonic plague (33). Unfortunately, the cat died prior to examination and was incinerated at a local veterinary practice before investigators arrived, precluding *Y. pestis* isolation attempts. One environmental sample was isolated from the carcass of a Colorado Chipmunk (*Tamius quadrivittatus*) collected approximately 1 km from the patient's residence. This isolate was paired with the human isolate from case E to see if it may have been related to the infective source. MLVA detected some similarity with 13 markers in common, however this isolate pair was not considered a match because it was not collected at the actual exposure site.

Successful environmental sample collection for cases F, G, H, and I ranged from 300 m to 0.4 km from the patients' residences or potential exposure sites, however, in each instance, epidemiological data indicated that patients' had traveled as far as 354 km to other potentially plague affected areas in New Mexico (case H, Table 1b). No isolates were obtained from environmental investigations done at these alternative exposure sites. In addition, cases F and G had roaming household pets that potentially covered several kilometers surrounding patients' residences before returning home with dead rodents and live fleas. When tested by MLVA, the environmental and human isolate pairs were not supported in jackknifed parsimony trees, and marker differences ranged from 7 to 15 loci,

which is typical of the number of marker differences observed for non-matching isolates seen in the NAU database. We therefore concluded that the correct exposure sites were not successfully sampled.

2.5 Discussion

Three inferences may be used in combination to support our conclusions that particular isolate pairs do indeed represent a match. First, in jackknifed parsimony trees, samples that were considered a match were highly supported with jackknife values ranging from 75 to 98%, providing > 99% confidence of a correct match. Second, isolates that were called matches were extreme outliers from the lower limit of the 99% CI of mean marker differences (n=663). Third, patient history and data collected during epidemiological investigations supported the match on a temporal and geographical scale. Based on mutation rate data and transmission modelling we would expect to see some slight genetic variation between matching human and environmental samples in those markers that mutate the fastest, and it is not surprising that those isolate pairs with the strongest statistical support differ at 1-3 markers (average = 2). Based on the transmission model, 2000 (95% CI = 254, 6840) generations or approximately 38 transmission cycles are required to see 2 mutations (49), and this number of transmission cycles probably would occur during an epizootic period or during one or two seasons of ongoing enzootic transmission in a plague focus such as that identified in case A. In contrast, the genetic similarity seen between the older 1991 and 1998 isolates, and the human case A isolate may be the result of these samples arising from a common origin but undergoing limited enzootic transmission and few mutations during those years in

which epizootic activity was not very evident. It is also possible that these isolates share alleles that are not identical by descent, but are similar because of parallel or convergent evolution. The markers chosen for this comparison are rapidly evolving and therefore increased likelihood that the same allelic state could arise through separate mutations (homoplasy) exists, however, we do not feel that this was a common phenomenon in our data set because of the lack of additional randomly matching isolates in the NAU database query. Whether these isolates illustrate an example of homoplasy, or they arose from the same epizootic source, the epidemiological data does not support the possibility of a match as the sample isolation precedes the human case by 11 and 4 years respectively. If the NYC 2002 plague case had been a case of bioterrorism, the human isolate still would have been traced to the correct region and even pinpointed to Santa Fe County, even when compared to plague isolates from around the world.

The epidemiological information collected during investigations is helpful in deciding where sampling should occur, but our data indicate that a definitive decision as to where the infective source arose should not be made based on these data alone. For example, in the original 1992 investigation of case B, the identification of an abscess on the patient's abdomen led investigators to believe that he was exposed to an infectious flea bite while carrying wood to his vehicle. Our MLVA results however, strongly suggest that the patient was exposed near his home, as indicated by the close match between his isolate and the one obtained from the flea pool collected from a prairie dog town near his residence. This example demonstrates the power and importance of using genomic diversity to ascertain likely exposure scenarios when epidemiologic data are inconclusive or contradictory.

Case D visited multiple potential exposure sites but samples could not be obtained from all of them. It was important in this case to identify the correct exposure site as the case was fatal and various members of the patient's family resided near two of the rural areas that the patient visited, with an additional site near the patient's residence and a high school (119). Although plague warnings are posted and appropriate precautions are taken in all suspect areas in cases such as these, a definitive answer as to where the infective source arose can greatly assist public health officials in allocating limited personnel and other resources.

These cases provide examples of how MLVA verified infective plague sources when it was not clear in the original investigation. By combining epidemiological information with matching isolate MLVA data, the likely exposure sites and often the infective sources can be identified. A non-matching environmental isolate can help investigators appropriately decide whether environmental sampling should be continued at a particular site, if additional potentially infective sites should be investigated further, or if a simple warning should be issued in those areas not successfully sampled. The human isolates in cases F, G, H, and I, which all occurred during 1983 in northern New Mexico, did not match the corresponding environmental isolates, or any of the isolates in the NAU database. This result could be expected during periods of widespread intense epizootic activity that occurred in 1983, when more cases were reported in the U. S. than had been seen since 1920 (26, 72). Isolates that did not match were collected over a widespread area in the Southwest during the 1980s plague epidemic, suggesting that this outbreak did not arise form a single source but rather from activity in many small plague foci scattered throughout the Southwest. We believe that such results are to be expected

when a very widespread outbreak occurs and multiple *Y. pestis* clones arising from many sources spread quickly across a region, eventually overlapping in distribution with each other. The plague outbreak of the early to mid-1980s represented such an event. Because of the high number of cases, only those sites likely to pose threats to other humans were thoroughly investigated by intensive trapping of rodent hosts and collection of flea vectors. Given the genetic dissimilarity between isolates obtained from cases F, G, H, and I, it seems that these cases were infected at alternate sites, or by additional widely circulating genotypes that might have spread from neighboring plague affected areas. While the epidemiological investigations and *Y. pestis* sampling efforts in these cases yielded helpful information, definitive statements about the actual infective source could not be made.

Our study presents an analytic strategy involving both epidemiologic data and MLVA, and demonstrates its use on multiple case studies, including one where the diagnosis was made a half continent away from the infective source. When combined with epidemiologic information, judicious use of genetic data from non-human organisms is highly attractive because of the power of DNA-based analyses to identify exposure sources (68, 69). However, this has proven contentious as experts disagree upon valid criteria for determining a match among samples (79). Our MLVAs of the above-described human and environmental *Y. pestis* isolates clearly demonstrate the value of this technique for the identification of likely sources of infection, and sites of exposure for human plague cases. When coupled with case histories and other epidemiological information, MLVA should also be useful for differentiating naturally occurring cases from those occurring from an intentional *Y. pestis* release.

2.6 Acknowledgements

This work was supported by the ISTC Biotechnology Engagement Program award K-584 p.

We thank Dr. Mark P. Simmons from the Colorado State University Department of Biology for technical assistance with phylogenetic analyses, Sandra K. Urich from the Centers for Disease Control and Prevention, Division of Vector Borne Infectious Diseases for laboratory support, and the Arizona Department of Health Services, New Mexico Department of Health, Colorado Department of Health and Environment, New York State Department of Health, and Navajo Area Indian Health Services for assistance with field investigations.

Chapter 3

Colorado animal-based plague surveillance systems: relationships between targeted animal species and areas at risk for humans

This chapter will be submitted to the *Journal of Zoonotic and Vector Borne Diseases* with the authors listed below. The submission date is dependent upon comments from participating authors.

Jennifer L. Lowell, Rebecca J. Eisen, Liang Xiaocheng, John A. Montenieri, Dale Tanda, John Pape, Michael F. Antolin, and Kenneth L. Gage

3.1 Abstract

Plague is a rare but highly virulent flea-borne zoonotic disease characterized by quiescent and epizootic periods. Static spatial models were created previously to identify areas posing significant risk to humans living in the southwestern United States. Animal-based surveillance complements these models by identifying when epizootics occur within known risk areas. The primary objective of our study was to determine the accuracy of predicting human risk of exposure to the causative agent of plague in Colorado based on passive plague surveillance data collected from 1991 to 2005 by the Centers for Disease Control and Prevention (CDC) and their state and local collaborators. By comparing locations of plague-positive animal samples with a previously constructed GIS-based plague risk model, we determined that the majority of plague-positive Gunnison's prairie dogs and non-prairie dog sciurids, and moderately high percentages of sigmodontine rodents, domestic cats, coyotes, and domestic dogs were recovered within 1 km of the nearest area posing high peridomestic risk to humans. In contrast, the majority of white-tailed prairie dog, Leporidae, and black-tailed prairie dog samples originated more than 5 km from the nearest human risk habitat. The spatial accuracy of identifying epizootic activity prior to a human case was fairly low, suggesting that other mammalian species or their fleas are likely to be more important sources of human infection in the areas posing the highest human plague risk. Epidemiological observations and multilocus variable number tandem repeat analyses (MLVA) identified chipmunks as a potentially important species for human plague risk.

3.2 Introduction

Plague is a rare but highly virulent flea-borne zoonotic disease caused by the Gram-negative bacterium *Yersinia pestis*. Identification of areas with high peridomestic risk for human exposure to the etiological agent of plague provides both a useful tool for targeting limited public health resources and reduces the likelihood of misdiagnosis by raising awareness of the disease. In a previous study, we created a temporally static spatial risk model for human exposure to the plague bacterium in the four corners region (37). Although the overall accuracy of the static model was high (~82%), plague is characterized by quiescent and epizootic periods, each posing different risks of human exposure (9, 26, 38). Consequently, there is a need for surveillance strategies, such as animal-based plague surveillance systems, that are capable of providing early warnings of plague epizootics.

In Colorado, the CDC and the Colorado Department of Public Health and Environment and their local collaborators have implemented a plague surveillance system that screens submitted animal samples for evidence of exposure to *Y. pestis*. This strategy represents modifications of earlier World Health Organization recommendations (9). The primary objective of our study was to determine if existing Colorado animal-based plague surveillance practices focus on species that are indicative of increased peridomestic human plague risk. Peridomestic areas are defined by the CDC as those areas within 2 km of the human home site. We identified animal groups associated with high human peridomestic risk areas and determined that some species important for human plague may be under-sampled during routine surveillance.

3.3 Materials and Methods

As part of an ongoing passive national plague surveillance program, the CDC Division of Vector-Borne Infectious Diseases tests submitted diagnostic samples for evidence of exposure to *Y. pestis*. Animal carcasses are necropsied and livers and spleens are tested for infection by direct fluorescent antibody staining and culturing. Cultures are confirmed as *Y. pestis* by specific bacteriophage lysis. Serum and blood samples collected on Nobuto strips are tested by passive hemagluttination and inhibition for antibodies to *Y. pestis*; titers $\geq$1:16 and $\geq$ 1:32 are considered positive for serum samples and serum recovered from Nobuto strips, respectively (22, 124). Fleas collected from submitted animals, or in association with their nests or burrows are identified, pooled by species and collection site and tested for presence of *Y. pestis* by inoculating triturated flea pools into susceptible laboratory mice; plague bacteria are then recovered from liver and spleen, cultured and confirmed by bacteriophage lysis (22).

All *Y. pestis*-positive non-human samples collected in Colorado from 1990-2005 were geo-referenced, added to a Geographic Information System (GIS) and projected to NAD 1983 Albers (Fig. 3.1). To avoid biasing results in favor of small mammals and fleas collected during investigation of human plague cases, any samples collected during epidemiological investigations of these cases were excluded from analyses. To evaluate accuracy of the animal-based surveillance groups in predicting human risk of exposure to the etiologic agent of plague, we first assigned each sample to one of nine groups (Table 3.1): coyotes, domestic cats, domestic dogs, one of three species of prairie dogs, non-prairie dogs sciurids, sigmodontine rodents, and lagamorphs. For each point, we

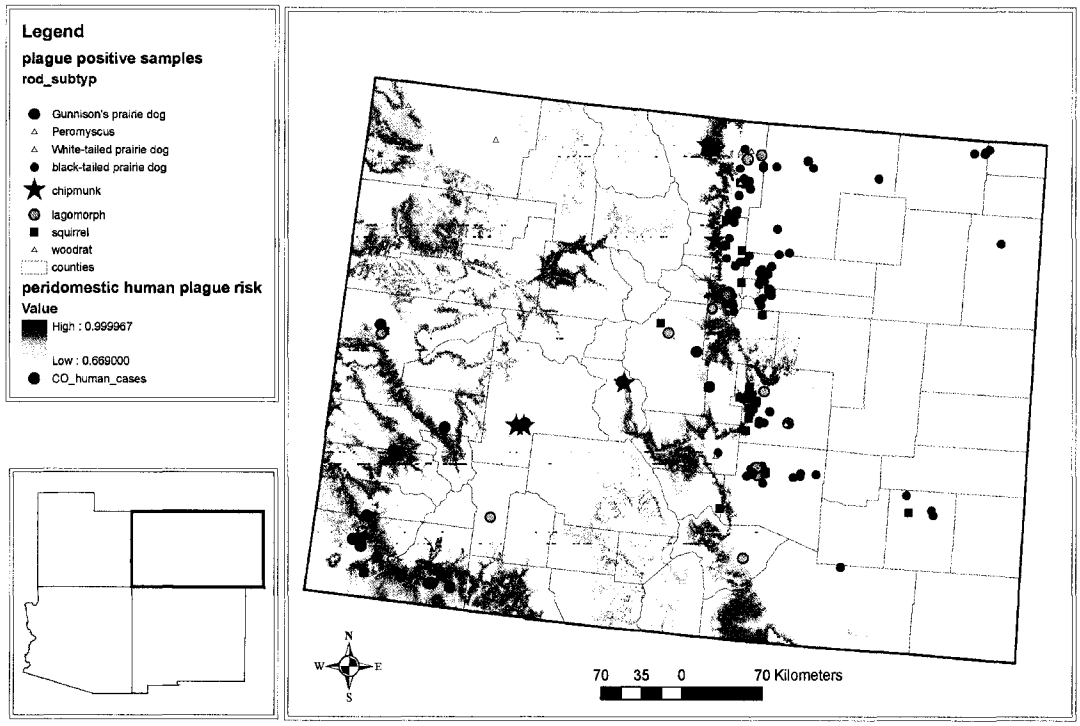

Figure 3.1. Origins of *Y. pestis* positive rodents and fleas that were used to predict high peridomestic risk habitat. Origins of human and environmental samples compared by MLVA.

Table 3.1. Plague positive samples associated with human case investigation data by animal-based surveillance group. Letter indicates the tier into which animal group fell when the distance from positive samples to high-risk habitat was determined. [a]tier one; [b]tier two, [c]tier three, [d]tier four, [e]tier five

Animal-based group and specimen type	**No. of positive samples**
Sciuridae, Cynomys gunnisoni and associated fleas[a]	**40**
Cynomys gunnisoni	19
Oropsylla hirsuta	13
Oropsylla tuberculata cynomuris	7
Oropsylla tuberculata tuberculata	1
Sciuridae non-Cynomys and associated fleas[b]	**133**
Ammospermophilus leucurus	1
Sciurus aberti	4
Sciurus niger	90
Spermophilus richardsonii	3
Spermophilus variegatus	14
Tamias quadrivittatus	2
Tamias spp.	1
Eumolpianus eumolpi	1
Hoplopsyllus anomalus	1
Orchopeas howardi	6

Oropsylla idahoensis	3
Oropsylla labis	2
Oropsylla montana	3
Thrassis bacchi	1
Thrassis fotus	1
Sigmodonitinae[c]	**7**
Neotoma mexicana	2
Peromyscus spp.	2
Malaraeus spp.	1
Orchopeas neotomae	2
Coyotes[c]	**132**
Canis latrans	132
Domestic cats[c]	**140**
Felis catus	140
Domestic dogs[c]	**8**
Canis familiaris	8

Table 3.1 cont.

Lagomorphidae and associated fleas[d]	**21**
Lepus californicus	1
Sylvilagus audubonii	19
Euhoplopsyllus glacialus	1
Sciuridae, *Cynomys leucuris* and associated fleas[d]	**6**
Cynomys leucuris	2
Oropsylla hirsuta	4
Sciuridae, *Cynomys ludovicianus* and associated fleas[e]	**229**
Cynomys ludovicianus	99
Oropsylla hirsuta	111

determined the shortest linear distance to high peridomestic risk plague areas described elsewhere (37). This model (37) was constructed using the known locations of human case exposure sites and habitat data available from GAP analyses and a digital elevation model.

Distance values were assigned to one of four categories. A value of zero km indicates that the plague-positive animal was collected in a location classified by the model as high peridomestic risk (37). Distance to risk habitat categories increased at the following increments: 0, 0.1-1, 1.1-5, >5 km. Distance to risk habitat among all groups was compared using goodness of fit tests. Fisher's exact tests were used for pair-wise comparisons and results were considered significant when $P \leq 0.05$.

To determine which of the surveillance strategies was most effective at identifying plague epizootics prior to a human case, we created buffers of 2, 5, 10, 25, 50 and 100 km around each of the 19 human cases occurring in Colorado from 1991 to 2005. We then identified the number of positive samples collected the year prior to the human case, or during the year of the case but prior to the date of onset of human illness. Each positive animal sample was assigned to a surveillance strategy group and linear regression was used to determine the relationship between the percentage of human cases associated with plague-positive animal specimens and distance of the specimen collection site from the human exposure site. Statistical tests were carried out using the JMP statistical package (SAS Institute, Cary, North Carolina).

Multilocus variable number of tandem repeat analysis (MLVA) was used to identify environmental sources of human plague exposure for four Colorado human plague cases that were investigated jointly by the CDC and the Colorado Department of

Public Health and Environment (CO-DPHE). Three cases were acquired in high peridomestic risk areas in Colorado, and one case was acquired in a low peridomestic risk area (Fig. 3.1) previously identified by (37). In each case the associated environmental investigations resulted in the isolation of *Y. pestis* from mammals or fleas collected at the sites where these cases presumably were exposed to plague. The environmental isolates collected during the corresponding human plague case investigations were categorized as high peridomestic risk if they were from plague positive animals given values of 0-1 km in the above analyses. Environmental isolates collected during the corresponding human case investigations were categorized as low risk if the plague positive animal was given a value > 1 km in the above analyses. The isolates from human case investigations were used for MLVA comparisons with the corresponding human isolates according to (74) (Table 3.2). Additional environmental isolates from 2004 surveillance sampling in surrounding Colorado counties and isolates from human cases in Arizona and New Mexico were included in the analyses to confirm that the Colorado human isolates did not genetically match isolates from different geographic areas or from different years.

DNA was extracted from 48 hour bacterial cultures using the heat soak method (3) for each of the above human cases, their corresponding environmental samples, and the 2004 isolates collected from surrounding counties. DNA from the 1983 to 2002 isolates from surrounding states were isolated and analysed previously for the identification of infective sources for human plague cases (74). All samples were subjected to MLVA and phylogenetic analysis using PAUP according to (74, 118).

Table 3.2. Origins of human and environmental *Y. pestis* isolates used in MLVA comparisons. A genetic match indicated that JK support was greater than 70%. *The case from Buena Vista was supported in 52% of JK trees.

Human Plague Case Locations and Years	**Associated Environmental Isolate Source**	**Did Case Occur in High Peridomestic Risk Habitat?**	**Was Environmental Source Species Associated with High Risk Habitat?**	**Genetic Match?**
Red Feather, CO 2004	*Eumolpianus eumolpi* (chipmunk flea)	Yes	Yes	Yes
Red Feather, CO 1999	*Cediopsylla inaequalis* (rabbit flea)	Yes	No	No
Buena Vista, CO 1992	*Tamias quadrivittatus* (Colorado chipmunk)	Yes	Yes	Yes (weak)*
Park County, CO 2004	*Sylvilagus audubonii* (rabbit)	No	No	Yes

inferred by constructing a strict consensus maximum parsimony phylogenetic tree, and 1000 jackknife replicates were used to assess support around the matching isolate groups. A *Y. pestis* isolate from Kazakhstan was used as the outgroup.

3.4 Results

3.4a Evaluation of surveillance groups in relation to peridomestic risk areas.

Five tiers of animal groups and their relationships to high human peridomestic plague risk areas were identified. Gunnison's prairie dogs compared with all other groups combined showed the highest proportion of samples collected within 1 km of areas considered high peridomestic risk for human exposure to plague (i.e., distance to nearest risk pixel ranges from 0 to 1 km) (100%; Fisher's exact 2-tailed $P<0.0001$). Significantly fewer non-prairie dog sciurid samples were collected within 1 km of risk areas (Fisher's exact $P = 0.0092$), but this second animal group also showed a high proportion of positive samples associated with high-risk areas (87.31%). The third group consisted of sigmodontine rodents (71.43%), domestic cats (71.43), coyotes (63.64%), and domestic dogs (62.50%) collected within 1 km of risk habitat ($\chi^2 = 2.03$, d.f. = 3, $P=0.5669$); as a group these animals were less frequently captured within 1 km of risk area than both Gunnison's prairie dogs and non-prairie dog sciurids (Fisher's exact $P<0.0001$). The proportion of plague-positive lagomorphs (28.57%) and white-tailed prairie dogs (33.33%) were not commonly collected within 1 km of risk area and did not indicate a significant association with this area (Fisher's exact, $P = 1.00$). Few positive black-tailed

prairie dog (9.61%) were collected within 1 km of risk area, and yielded fewer samples within risk habitat than lagomorphs and white-tailed prairie dogs combined (Fisher's exact P = 0.0060) (Fig. 3.2).

3.4b Occurrence of positive animal samples in relation to human plague exposure sites. From 1991 to 2005, 19 human cases of plague were reported in Colorado. The exposure sites for 12 of the 19 cases were determined to be within the high peridomestic risk environment. Within fixed distances (0, 0-1, 1.1-5, >5 km) of the human exposure site for each of these cases, we identified positive animal samples collected the calendar year prior to the occurrence of the case, or during a period within the same year but prior to the date of onset for the case. Among these cases, only a single positive domestic cat carcass was collected within 2 km of the human exposure site within the year prior to onset of symptoms in the human. At a distance of 100 km from the human exposure site, at least a single positive specimen was collected during the specified time period for all human cases.

To determine the relationship between the percentage of human cases associated with a positive animal sample and the distance from the human case where the sample was collected, we constructed linear regression models for detection based on all animal groups combined, or each of the groups separately; each of the models yielded a statistically significant linear fit ($P < 0.02$) with r^2 values ranging from (0.78-0.99). From these linear regression models, we estimated the buffer size needed to identify a positive animal specimen for 75% of the human cases that occurred from 1991-2005. Using all animal-based surveillance groups and a buffer size of 71 km, 75% of human cases would

be associated with a positive animal sample (percent of cases associated with a positive animal specimen = 0.96 km^{-1} (buffer size) + 7.1; F = 387.38, d.f. = 1, 5, r^2 = 0.99, P<0.0001). That is, for 75% of human cases, epizootic or enzootic plague activity was detected within 71 km of the human exposure site the year prior to or the year of, but prior to onset of human illness. In contrast, any individual animal group alone would require a buffer size of 114 to 1,515 km, larger than the width or length of Colorado.

3.4c Identification of sources of human plague exposure by MLVA. In an attempt to identify likely sources of infection for human cases within high peridomestic risk areas, we compared human isolates two high risk areas with environmental isolates collected during investigations of these same human cases. Human plague case isolates collected in high peridomestic risk areas in Red Feather, CO in 2004 and in Buena Vista, CO in 1992 were genetically similar to chipmunk associated environmental isolates from E. eumolpi fleas and T. quadrivittatus (Colorado chipmunk) and were supported in 84% and 52% of jackknifed parsimony trees respectively (Fig. 3.3), suggesting these fleas and their hosts might have been a source of exposure for these cases. The human plague case isolate collected in a high peridomestic risk area from Red Feather, CO in 1999 was not genetically similar to an isolate from a *C. inaequalis* (rabbit flea), which was the only environmental isolate recovered during this case investigation. An additional human plague case isolate collected from a patient exposed in a low peridomestic risk area in 2004 was genetically similar with high jackknife support (84%) to the environmental

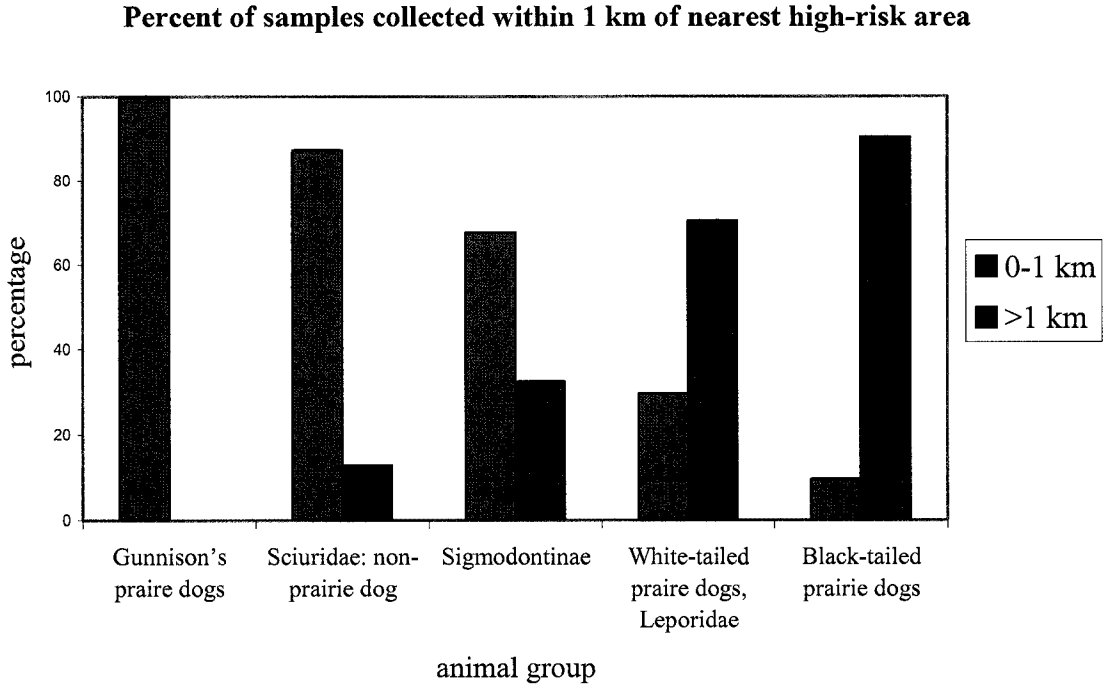

Figure 3.2. Percentages of each animal species or group associated with high peridomestic risk areas for human plague.

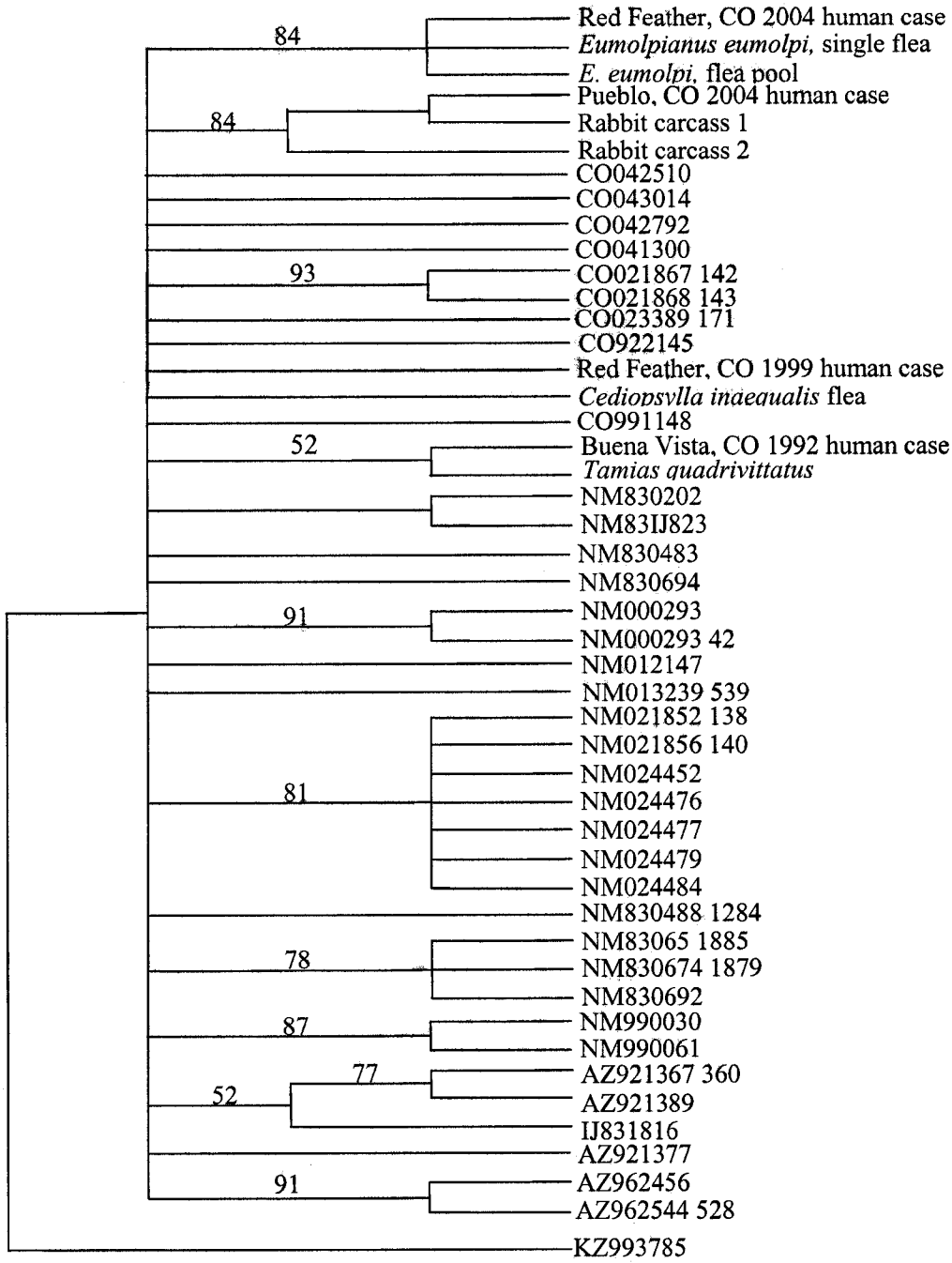

Figure 3.3. Strict consensus tree constructed using parsimony analysis identified genetically similar isolates. Each CO 2004 human and environmental isolate is labeled as described in the text. Additional isolates are labeled with CDC accession numbers according to their origins: NM = New Mexico, IJ = Indian Health Services, AZ = Arizona, CO = Colorado, KZ = Kazakhstan. The two numbers following the state abbreviation indicate the year collected. The remaining numbers are unique identifiers. Jackknife support values are based on 1000 replicates. Branches with no numbers received less than 50% jackknife support.

Phylogenetic relationships between isolates were isolates from two S. audubonii (rabbits) shot and handled by the case-patient. The Colorado human case samples from these four cases did not genetically match any of the other isolates from surrounding areas or from different years.

3.5 Discussion

Our methods provided a framework approach to evaluate animal-based surveillance techniques for zoonotic diseases that pose a risk to humans. In Colorado, this study indicated that despite the close association between site of collection of plague-positive surveillance samples and high peridomestic risk habitat, the spatial accuracy of identifying epizootic activity prior to a human case was fairly low. To identify epizootic or enzootic plague activity for 75% of human cases the year prior to or the year of, but prior to onset of human illness, a 71 km buffer around the human exposure site was needed. For any individual animal group alone, the percentage of cases associated with positive animals within the same geographic distance would be less than 42%. This lack of precision is most likely the result of collecting samples opportunistically or by convenience rather than through continuous surveillance involving an experimental design.

Although some animal groups were more closely associated with high peridomestic risk habitat, combining multiple animal-based surveillance groups provided better landscape coverage of plague activity than any animal group alone with the exception of Gunnison's prairie dogs. When this approach was used to assess plague surveillance sampling efforts, we identified five tiers of animal groups that varied in their

relationships with high peridomestic risk areas. Gunnison's prairie dogs showed the highest association with 100% of plague-positive animals collected < 1 km of high peridomestic risk areas. Non-prairie dog sciurids were the second most closely associated group with 87.31% of plague-positive animals collected < 1 km of high peridomestic risk areas. Relative to the first two groups, moderately high percentages of sigmodontine rodents (71.43%), domestic cats (71.43%), coyotes (63.64%), and domestic dogs (62.50%) were recovered < 1 km of the nearest area posing a peridomestic risk to humans. In contrast, the fourth and fifth tiers consisted of white-tailed prairie dogs and the Leporidae, and black-tailed prairie dog samples with only 29.63% and 9.61% of positive samples respectively collected < 1 km of the nearest human risk area (Fig. 3.2).

Prairie dogs are generally implicated as a high peridomestic risk species for human plague transmission because their epizootics are easy to identify (60) and they are highly susceptible to plague infection with colonies suffering nearly 100% mortality (28). Our study indicated that Gunnison's prairie dogs are the only *Cynomys* species associated with the high peridomestic risk areas identified using GIS (37). The majority of plague-positive white-tailed and black-tailed prairie dogs or associated fleas in our study were collected more than 5 km from the nearest high peridomestic risk areas. This finding is consistent with the observation that white-tailed prairie dogs have not been implicated in human plague in Colorado, and black-tailed prairie dogs have been implicated as source of infection for only 10% (two of 19 cases) of human infections in Colorado from 1991-2004. The remaining 6 prairie dog associated human cases in Colorado occurred within the distribution range of Gunnison's prairie dogs (CDC, unpublished data). The need for larger expanses of land by black-tailed prairie dogs may reduce risk to humans by this

species. Large colonies are restricted to areas where the density of human dwellings is low. Within urban areas where high densities of humans and prairie dogs co-occur, habitat fragmentation (e.g. roads) may be protective against epizootics (24). Alternatively, the primary flea infesting prairie dogs, *Oropsylla hirsuta,* does not frequently bite humans (9), thus reducing the number of human infections associated with prairie dogs. In contrast, Gunnison's prairie dog habitat tends to be interspersed with rural areas inhabited by humans where *Y. pestis* transmission is likely to occur during an epizootic. Plague in Gunnison's prairie dogs has been well studied, and past human plague investigations in Colorado, New Mexico, and Arizona have linked this species with human plague (27, 28, 72). It is not surprising that 100% of these animals were within 1 km of high peridomestic risk areas.

Moderately high percentages of sigmodontine rodents, domestic cats, coyotes, and domestic dogs were recovered within 1 km of the nearest area posing high peridomestic risk to humans. Numerous studies have recognized the utility of screening serum samples from coyotes for antibodies to *Y. pestis* (33, 109) because the home range of coyotes in Colorado is large, ranging from 2.8 to 185 km^2 (48). Within that area they consume high numbers of potentially *Y. pestis* infected rodents or are exposed to infectious fleas (33, 104, 109). Experimental challenges and the high seroprevalence rates among canids suggest that most hosts exposed to infection survive and seroconvert (9, 33, 102) and high antibody titers persist for up to four to eight months (9). For these reasons, it has been proposed that sampling a single coyote is equivalent to sampling

hundreds of rodents (33). However, because a coyote's home range is so large, the exact geographic location of epizootic activity is less certain than for any of the other surveillance groups.

Due to their close association with humans, domestic cats and dogs are invaluable sentinels for plague. They typically roam closer to the home than coyotes, making them extremely useful for identifying plague activity in the peridomestic environment where most humans are exposed (37). Indeed an infected domestic cat carcass was the only positive animal sample collected within 2 km of a human exposure site within a year and a half prior to onset of symptoms.

The non-prairie dog sciurid group was second only to Gunnison's prairie dogs in their association with high peridomestic risk plague areas. Within this group rock squirrels and their fleas (*Oropsylla montana*) have previously been identified as important species for transmission of *Y. pestis* to humans (9, 26), but human several human plague cases in Colorado have occurred outside of rock squirrel and Gunnison's prairie dog habitat. We used molecular techniques and human case investigations from high and low peridomestic risk areas to identify alternative species within the non-prairie dog sciurid group that may be important indicators for human plague in these risk areas. By comparing human *Y. pestis* isolates with case-associated environmental isolates collected in high peridomestic risk areas outside of rock squirrel habitat, MLVA inferred genetic matches between humans and chipmunks in two cases. The third human case that occurred in a high peridomestic risk area was compared to a rabbit flea (Leporidae group), the only isolate found during the case investigation, and these isolates were not genetically similar. The human case from low risk area was genetically similar to *Y.*

pestis infected rabbit carcasses that had been handled by the patient. These results lend more support to the hypothesis that although rabbits often are sources of infection for hunters and can infect humans in areas that probably have low risk of fleabite exposure, they do not appear important in peridomestic human plague.

The 2004 Red Feather case was especially interesting and lends the highest support to a relationship between chipmunks and human plague risk. Rather than an obvious animal die-off as indicated by the presence of dead rodents, the flea-rodent host associations occurring in the area provided evidence of a plague epizootic. Although one of the *Y. pestis*-positive *E. eumolpi* collected during this investigation was found on a *T. minimus* (least chipmunk), which is one of this flea's preferred hosts, a second *Y. pestis* isolate was obtained from *E. eumolpi* fleas recovered from a *S. lateralis* (golden-mantled ground squirrel), which is not a preferred host for this flea. This observation suggested that the Red Feather area was experiencing a plague epizootic involving chipmunks. "Host-switching" by fleas is commonly observed during plague epizootics and the presence of a particular flea species on atypical hosts is often considered evidence that the flea's normal hosts were killed by plague (9). Based on this argument, the *E. eumolpi* found on the golden-mantled ground squirrel in 2004 could indicate that the chipmunk fleas were seeking alternative hosts on which to feed, thereby making it more likely that an *E. eumolpi* would feed other animals including humans. Barnes (1982) demonstrated that *E. eumolpi* fleas readily feed on humans when they are lacking their natural host and suggested that they could be sources of human infection. Plague positive *E. eumolpi* isolated from two different rodent species, the epidemiologic information acquired for this case, and the MLVA genotyping provided investigators with the

evidence needed to justify their decision to focus their prevention and control efforts in the Red Feather area rather than other potential exposure sites visited by this case during the incubation period of this disease. The human case from 1992 also showed genetic similarity with a chipmunk isolate. Although this genetic match was not as high as those samples from the 2004 case, the positive chipmunk carcass was found approximately 1 km from the patient's residence, and when combined with the epidemiological investigation of the case (33), the genetic match was high enough to infer that this sample was from the same epizootic source as the human isolate (74).

The majority of mountain-dwelling non-prairie dog sciurids, including 100% of the chipmunks used in this study, were recovered within 1 km of the nearest area posing high peridomestic risk to humans. Although combining multiple animal-based surveillance groups provides better landscape coverage of plague activity than any animal group alone, the 2004 Red Feather case molecular data suggested that chipmunks may be important species for human plague outside of rock squirrel and Gunnison's prairie dog range providing a local indicator of human plague risk. In addition, chipmunk distribution corresponds to high peridomestic risk habitat, they are susceptible to *Y. pestis*, *Eumolpianus spp.* (chipmunk fleas) willingly bite humans, and their non-colonial lifestyle enables them to survive in fragmented habitat that often results from urban development or other human activities common in the Colorado mountains. Chipmunks have also been implicated in human plague in coniferous forest communities in California where they are considered excellent sentinel animals, and have been used in integrated surveillance programs (86). Chipmunks may also be important species for

human plague risk in Colorado and should be targeted in routine surveillance activities, especially outside of rock squirrel and Gunnison's prairie dog habitat.

Current animal based surveillance techniques in Colorado are not effective for predicting when and where human cases will occur. This study provided a framework approach to evaluate animal-based surveillance techniques for zoonotic diseases that pose a risk to humans and provided information about which animal species and groups are associated with high human peridomestic plague risk areas in Colorado. Gunnison's prairie dogs and rock squirrels were previously implicated in human plague cases (9, 26, 27) but molecular data suggested that outside of these species' range, chipmunks may also be important and surveillance strategies should include more intensive sampling of these rodents.

Chapter 4

Multi-locus variable number of tandem repeat analysis of two human plague case cluster in New Mexico in the early 1980s

This chapter will be submitted for publication in the summer of 2007. The submission date is dependent on the completion of the *Y. pestis* spanning tree analyses. The following authors will be included on this manuscript.

Lowell, J. L., M. F. Antolin, and K. L. Gage.

4.1 Abstract

Although plague is often viewed in an historical context, this flea-borne zoonosis is still prevalent in rodent populations of the southwestern U. S. and the etiological agent *Yersinia pestis* is occasionally transmitted to humans. Widespread epizootic activity is especially evident in New Mexico (NM) where it can affect many rodent species. Rodent die-offs caused by the plague appear to span large continuous regions, but geographic and biological barriers may provide mechanisms that spatially separate plague epizootics. Bacterial population genetics studies can provide evidence of pathogen population structure and in turn reveal whether epizootic events are small and isolated or large with the same bacterial clone occupying multiple ecological niches. We analyzed a set of *Y. pestis* isolates from a human plague epidemic that occurred in NM during the early 1980s using multi-locus variable number of tandem repeat analysis (MLVA) to determine if *Y. pestis* populations structure corresponded to human case clusters. Variable number of tandem repeat (VNTR) loci were PCR-amplified and analyzed using MLVA and maximum parsimony analysis. Four groups of isolates collected from the same counties formed distinct clades and received over 70 % parsimony jackknife support, but isolates from the larger human plague case clusters were not supported and were unresolved in a polytomy.

4.2 Introduction

Yersinia pestis was introduced to the U. S. from Asia around 1900. Since then it has spread eastward to the hundredth meridian and circulates in rodent populations of the western U. S. where the disease occasionally infects humans. Epizootic activity and subsequent human infections are especially evident in California, NM, Arizona, and Colorado. Between 1944 and 1991, 296 human plague cases were reported in those states, including 166 (56%) cases in NM (26). Although more recent years have seen fewer human cases, percentages by state have remained the same and NM still experiences the highest rates of human plague in the western U. S. (80). The plague was first documented in Caltron County, NM in 1938 when *Y. pestis* was isolated from a prairie dog carcass and its associated fleas (73). Within the following ten years *Y. pestis* was found in several other rodent species including ground squirrels (*Citellus spp.*), wood rats (*Neotoma albigula*), grasshopper mice (*Onychomys spp.*), and others spanning 11 counties (73). Subsequent rapid urban growth in plague endemic areas of NM provided a scenario in which humans were living in close proximity to rodents that experienced plague die-offs and human cases began increasing in the 1970s (26). Between 1980 and 1985, widespread epizootic activity in the southwest led to a human plague epidemic where approximately 63 human cases spanning 13 counties were documented in NM (10, 25). The majority of these cases formed two geographically distinct clusters with a few outlying cases. One cluster was in north-central NM, and one was in west-central NM. Most of the outlying cases radiated outward from the north-central cluster and two cases occurred in the southwestern part of the state (Fig. 4.1).

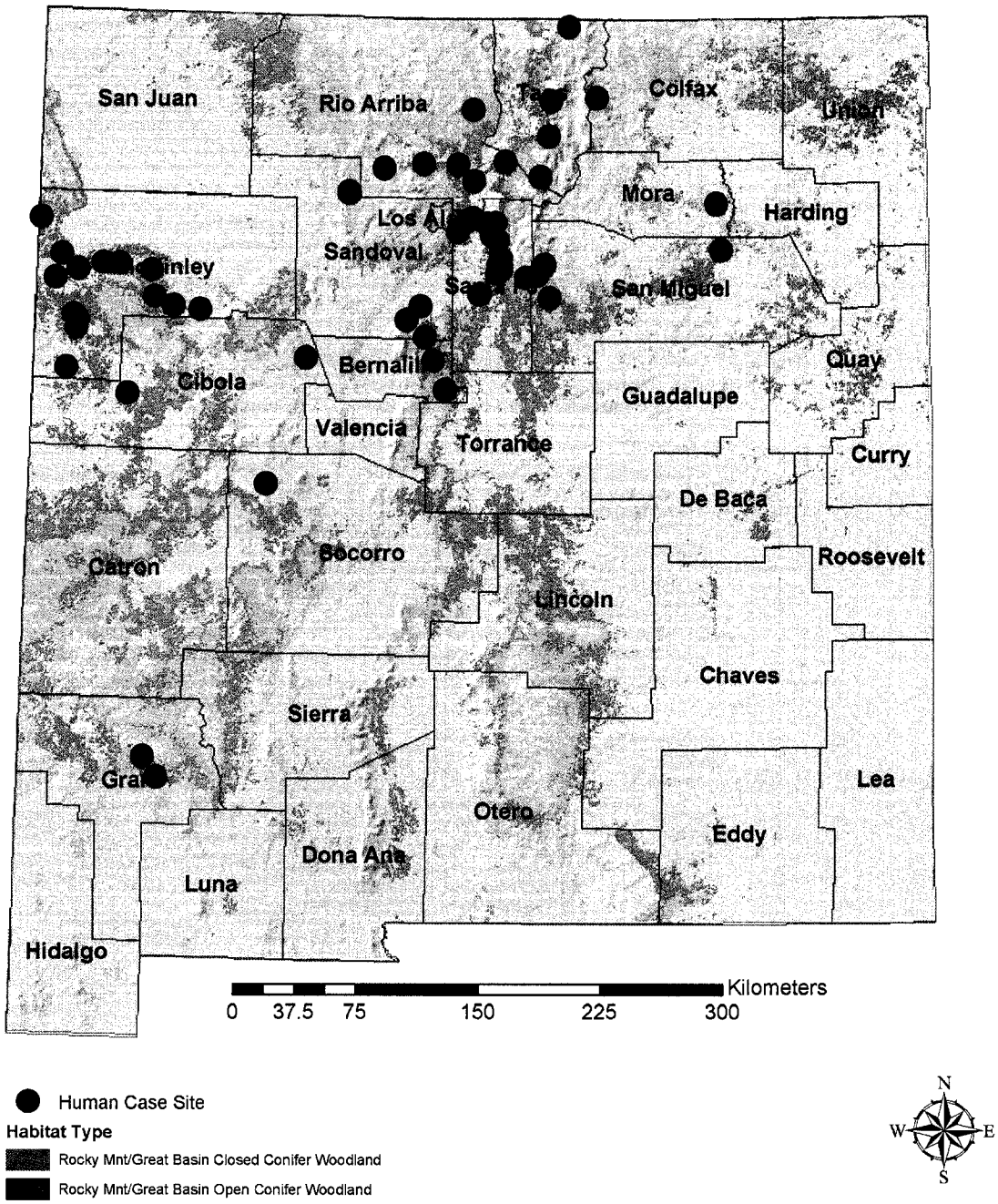

Figure 4.1. Map of New Mexico indicating human case isolate origins. Name labels identify counties.

Plague epizootics were extremely widespread in the southwestern U. S. in the early 1980s and many different rodent species were affected. Species and habitat variation may have provided biological or ecological barriers to geographically separate plague epizootics. Rock squirrels have been implicated most often for human plague transmission in NM (25) but prairie dogs and other squirrels that occupy different habitat types also experience plague epizootics (73) and probably played a role in transmitting infection to humans during the 1980s epidemic (74). Transmission of *Y. pestis* by different rodent species that occupy distinct ecological niches or are separated by geographic barriers may have created clusters of human disease that were caused by geographically separate bacterial clones. Defining the bacterial population structure of seemingly separate disease foci may answer questions about how far individual epizootic events can spread. If rodent populations are geographically or biologically isolated, then the bacteria circulating in the rodent populations may consist of genetically distinct populations. The bacterial population structure among human case clusters may provide clues to how *Y. pestis* moves across landscapes and whether widespread epizootic activity such as that seen in NM during the 1980s was caused by a single epizootic clone, or by individual bacterial clones that arose separately when environmental conditions were favorable for epizootic activity. The goal of this project was to determine the spatial arrangement of disease foci seen during the 1980s NM plague outbreak using MLVA to analyze *Y. pestis* population structure.

4.3 Materials and Methods

4.3a Isolate selection.

Sixty-two isolates were selected from 13 counties containing each main human case cluster from northwestern and from north-central NM. Isolates from counties surrounding the two clusters and two isolates from Grant County in southwestern NM were also included in the analysis (Fig. 4.1, Table 4.1). A set of four isolates from Santa Fe county collected in 2002 were included as a positive control because these isolates were known to infer genetic matches by MLVA (74).

4.3b MLVA and statistical analyses.

All isolates were plated on sheep's blood agar (SBA) and incubated at 25 °C for 48 hours. DNA was extracted using a heat soak method (3). Polymerase chain reaction was performed using 17 VNTR markers according to (74), and a subset of these isolates was also tested using all 42 VNTR markers according to (75). Fragment analysis and the associated phylogenetic analyses were performed according to (74, 75). A *Y. pestis* isolate from Kazakhstan was used as the outgroup.

4.4 Results and Discussion

MLVA and parsimony analysis of the NM samples did not identify distinct populations of *Y. pestis* by county or by human case cluster (Fig. 4.2). A subset of these

Table 4.1. Isolates used in NM analysis. *The CDC accession number consists of the state abbreviation followed by the last two digits of the collection year with the exception of several isolates from 1980 in which the state abbreviation and year are reversed. SC is the abbreviation for the New Mexico State Health Department. The last four digits are unique identifiers.

CDC Accession number*	County	Isolate origin
NM830695	Bernalillo	Bubo
NM840731	Bernalillo	Cervical Bubo
NM855387	Bernalillo	Inguinal Bubo
80NM517	Cibola	Axillary Bubo
NM830854	Cibola	Inguinal Bubo
NM830692	Colfax	Inguinal Bubo
NM830385	Grant	Inguinal Bubo
NM857152	Grant	Multiple Bubo
NM 830203	McKinley	Axillary Bubo
NM 830538	McKinley	Bubo
NM 830586	McKinley	Inguinal Bubo
NM 830694	McKinley	Inguinal Bubo
NM 830889	McKinley	Axillary Bubo
NM 840082	McKinley	Axillary Bubo
NM 858134	McKinley	Blood
NM 858135	McKinley	Blood
NM820180	McKinley	Axillary Bubo
NM820394	McKinley	Inguinal Bubo
NM 853945	Mora	Axillary Bubo
80 NMWE909	Rio Arriba	Blood
NM 830426	Rio Arriba	Bubo
NM 830444	Rio Arriba	Inguinal Bubo
NM 840532	Rio Arriba	Bubo
NM 840691	Rio Arriba	Axillary Bubo
NM840470	Rio Arriba	Inguinal Bubo
80 NM595	San Miguel	Inguinal Bubo
NM 830480	San Miguel	Bubo
NM 840692	San Miguel	Inguinal Bubo
NM 841002	San Miguel	Inguinal Bubo
NM 855938	San Miguel	Axillary Bubo
80 NM769B	Sandoval	Axillary Bubo
NM 810240	Sandoval	Blood
NM 810381	Sandoval	Blood
NM 810635	Sandoval	Inguinal Bubo
NM 856807	Sandoval	Axillary Bubo
80 NM826-2	Santa Fe	Inguinal Bubo
80 NM877	Santa Fe	Cervical Bubo
80 NM913	Santa Fe	Inguinal Bubo
80 NM980	Santa Fe	Inguinal Bubo
80 WE0375B	Santa Fe	Inguinal Bubo

NM 820444	Santa Fe	Axillary Bubo
NM 830445	Santa Fe	Inguinal Bubo
NM 830482	Santa Fe	Inguinal Bubo
NM 830483	Santa Fe	Blood
NM 830539	Santa Fe	Inguinal Bubo
NM 830587	Santa Fe	Blood
NM 830693	Santa Fe	Inguinal Bubo
NM 840518	Santa Fe	Blood
NM 840610	Santa Fe	Blood
NM 840923	Santa Fe	Axillary Bubo
SC 830003	Santa Fe	Multiple Bubo
80 NM 725B	Socorro	Inguinal Bubo
NM 856390	Taos	Inguinal Bubo
NM830522	Taos	Bubo
NM858324	Taos	Bubo

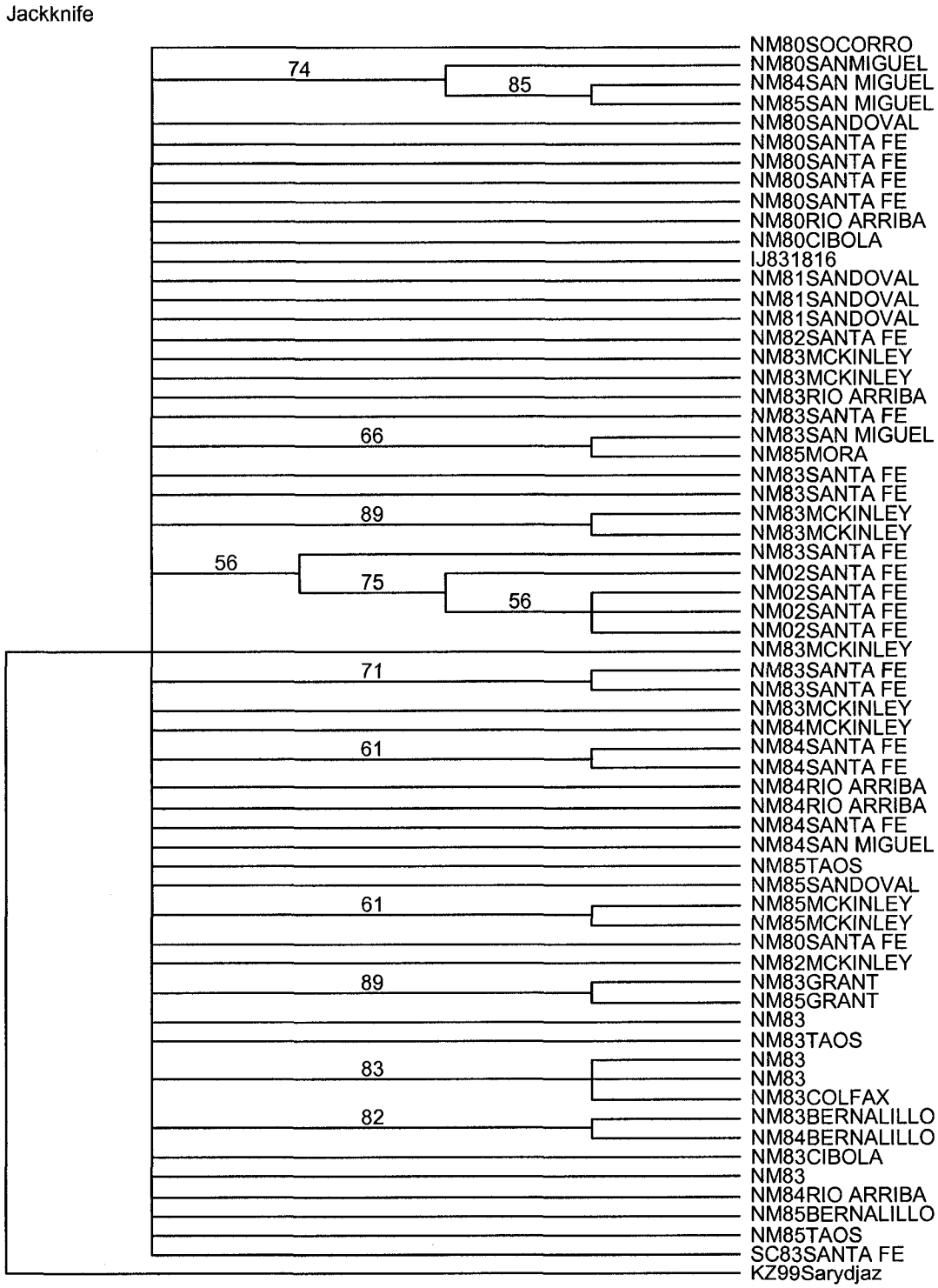

Figure 4.2. Jackknife consensus tree of samples collected from two NM human case clusters during the early 1980s. Terminals are labeled with the CDC accession number for each sample with the state abbreviation followed by the collection year and the county of origin. NM, New Mexico; IJ, Indian Health Services of New Mexico; SC. Numbers above the branches indicate jackknife support.

isolates was analyzed using the entire 42 marker set to determine if adding less variable markers would help identify groups of isolates that were genetically related, but this analysis did not provide any additional resolution (75).

Several isolate pairs or groups were supported in greater than 70% JK support. These consisted of three isolates from San Miguel county collected in three different years (74 %), two isolates from McKinley county collected in the same year (89% JK), two isolates from Santa Fe county collected in the same year (71%), two isolates from Bernalillo county collected in different years (82%) and three isolates from Colfax counties and the same year (82%). The isolates from Grant county were highly supported (89%) and the four isolates used as a control and collected in 2002 were supported in 75% JK support.

Prior MLVA demonstrated that samples collected within approximately 3 km of each other are generally well supported in monophyletic groups by VNTRs (74). The samples in this analysis that were collected in the same years and counties, and exhibited high JK support were also collected in close proximity to each other. Samples that matched from the same county but collected in different years were from areas with relatively few human cases. One group of supported isolates (San Miguel county, (1980, (1984, 1985)) were separated in time by four years. Epizootic activity in these counties may not have been as widespread or as intense as in the counties with higher human case incidences and enzootic periods may have occurred in those particular foci. If *Y. pestis* transmission and subsequent bacterial multiplication were reduced in these counties, marker mutation events may also have been reduced allowing the informative markers to remain the same between these isolate over a longer time period. It is also possible that

these marker matches were homoplastic because of the high mutation rates associated with VNTRs (49). The overall lack of resolution in the tree is most likely due to stochastic variation caused by the more rapidly evolving VNTR markers (49, 75), and the lack of parsimony informative characters provided by the slowly mutating markers. It is also possible that when widespread epizootics occur *Y. pestis* moves across large geographic distances while undergoing many replications, causing high variability in the markers that mutate the fastest. Slower evolving genetic markers are needed to definitively answer the question of *Y. pestis* clonality across a landscape scale represented by an area such as NM.

Chapter 5

Phenotypic and molecular characterizations of *Yersinia pestis* isolates from Kazakhstan and adjacent regions

This chapter was published in the journal *Microbiology*. The text of this manuscript and all figures are presented as they appeared in the journal. The reference for this chapter is shown below.

Jennifer L. Lowell, Aigul Zhansarina, Brook Yockey, Tatyana Meka-Mechenko, Gulnaz Stybayeva, Bakyt Atshabar, Larissa Nekrassova, Rinat Tashmetov, Kuralai Kenghebaeva, May C. Chu, Michael Kosoy, Michael F. Antolin, and Kenneth L. Gage. 2007. Phenotypic and molecular characterizations of *Yersinia pestis* isolates from Kazakhstan and adjacent regions. Microbiology. 2007 Jan;153 (Pt 1):169-77

5.1 Summary

Recent interest in characterizing infectious agents associated with bioterrorism has resulted in the development of effective pathogen genotyping systems, but this information is rarely combined with phenotypic data. *Yersinia pestis*, the etiologic agent of plague, has been well defined genotypically on local and world wide scales using multi-locus variable number tandem repeat analysis (MLVA), with emphasis on evolutionary patterns using old isolate collections from countries where *Y. pestis* has existed the longest. World wide MLVA studies are largely based on isolates that have been in long term laboratory culture and storage, or on field material from parts of the world where *Y. pestis* has potentially circulated in nature for thousands of years. Diversity in these isolates suggests that they may no longer represent the wild type organism phenotypically, including the possibility of altered pathogenicity. Our study focused on the phenotypic and genotypic properties of 48 *Y. pestis* isolates collected from ten plague foci in Kazakhstan. Phenotypic characterization was based on diagnostic tests typically performed in reference laboratories working with *Y. pestis*. MLVA was used to define the genotypic relationships between the central-Asian isolates and a group of North American (NA) isolates, and to examine Kazakh *Y. pestis* diversity according to predefined plague foci and on an intermediate geographic scale. Phenotypic properties revealed that a large portion of this collection lacks one or more plasmids necessary to complete the blocked flea/mammal transmission cycle, has lost Congo Red binding capabilities, or both. MLVA analysis classified isolates into previously identified biovars, and in some cases groups of isolates collected within the same plague focus

formed a clade. Overall MLVA did not distinguish unique phylogeographic groups of *Y. pestis* isolates as defined by plague foci and indicated higher genetic diversity among older biovars.

5.2 Introduction

Interest in microorganisms with potential for use as bioweapons has steadily increased in the last five years (5, 19, 46, 49, 74), providing insight into the epidemiology, ecology, evolution, and the molecular diversity of pathogenic bacteria. *Yersinia pestis*, the etiological agent of plague has caused millions of deaths world wide and holds greater historical significance for human health than perhaps any other microbe (72, 94), but little is understood about its maintenance in nature, and whether phenotypic diversity is important. Recent collaborative efforts have provided a synthesis of old and new information describing the phenotypic and molecular diversity of world-wide *Y. pestis* isolate collections (5) and additional molecular characterizations have summarized the evolutionary and epidemiologic relationships of *Y. pestis* (49, 61, 74, 84, 91). This knowledge combined with ecological properties and life-history traits may help answer questions about epizootic plague cycles and *Y. pestis* transmission in nature (5, 46).

Defining plague foci on the basis of phenotypic, biochemical, and ecological properties of bacterial isolates has proven useful. Anisimov (2004) summarized the characteristics of *Y. pestis* isolates from central Asia, one of the hypothesized regions from which *Y. pestis* emerged as a flea-borne pathogen of rodents and other mammals (1). This summary described plague focus designations, epizootic activity, main rodent hosts, main flea vectors, and geographic extent for each of the forty-three purportedly distinct plague foci found in the southern and southeastern regions of the former Soviet Union (FSU). Subgroups of *Y. pestis* were also characterized using biochemical properties, rodent host, and geographic region of isolation.

While these phenotypic properties have historically represented *Y. pestis* subtypes and helped define the foci from which they originated, relatively new molecular techniques have provided bacteriologists with tools capable of revealing genetic diversity in *Y. pestis*. Molecular typing methods like MLVA have helped to define *Y. pestis* population structure on several spatial scales (1, 49, 74). On local scales where isolates were collected within a few kilometers of each other, MLVA genotyping has demonstrated matches between environmental isolates and associated human plague infections in the southwestern United States (74), and genotype relationships of *Y. pestis* isolates among 19 prairie dog colonies in Arizona (49). At broader scales, MLVA has inferred relationships between biovars based on clustering algorithms, but has not typically been used to show hierarchal relationships between the biovars (1, 69). Recent single nucleotide polymorphism (SNP) analyses have indicated that biovar Antiqua is polyphyletic (1, 20), and that the atypical Antiqua isolate "Angola", and the *Y. pestis* "Pestoides," isolates are ancestral to all three biovars (1).

While phenotypic and molecular characterizations have provided valuable insight into worldwide *Y. pestis* collections and their relationships, studies examining *Y. pestis* isolates from an intermediate scale in central Asia are lacking. Combining ecological, phenotypic, and molecular information into single data sets could be especially useful for central-Asian isolates, as they potentially contain diversity not seen in parts of the world where *Y. pestis* is relatively new. The goals of this study were two fold. The first was to phenotypically characterize these isolates according to the tests commonly used by plague diagnostic and reference laboratories around the world (22). The second was to

examine intermediate geographic scale MLVA on *Y. pestis* isolates that have potentially circulated in nature for thousands of years generating different MLVA relationships than those in NA isolates.

5.3 Materials and Methods

5.3a Isolate collection. The FSU was previously divided into five natural plague regions based on the rodent host species, the ecology of the area, and the *Y. pestis* phenotypes generally found within these regions (5). These natural plague regions were further subdivided by FSU scientists and designated "autonomous plague foci," creating independent areas according to the predominant rodents and vectors, and also by administrative units (105). Sixteen autonomous plague foci reside within and adjacent to Kazakhstan. For this study, 48 isolates were selected from the *Y. pestis* collection at the M. Aikimbayev's Kazkah Scientific Center for Quarantine and Zoonotic Diseases (KSCQZD), Almaty, Kazakhstan. These isolates were originally collected from 10 of the 16 autonomous plague foci within and adjacent to Kazakhstan (Table 5.1, Fig. 5.1). The isolates selected are representative of the mammal host, vector host, and geographic diversity found in Kazakhstan. The isolate source, geographical origin, and collection year were recorded in Table 5.1. Although some of the isolates were collected adjacent to Kazakh borders, all were kept in the Kazakh collection at KSCQZD, and here referred to as Kazakh isolates. Isolates were previously stored at KSCQZD at 4 °C on Hottinger's

Table 5.1. Phenotypic characteristics of the Kazakh isolates.
*Origin is the autonomous plague focus contained within designated natural plague regions.
†Natural region containing autonomous plague foci
‡Original biological source yielding *Y. pestis* isolate
§M, medievalis; A, antiqua; O, orientalis
‖Plasmid sizes 110 = pMT1, 70 = pCD1, and 9.5 = pPCP1
¶Pgm, pigmentation result on Congo Red Agar, white = congo red binding negative, red = congo red binding positive
#Number of times isolate has been passed during storage
**T, Typical ferments Glucose, Rhamnose, Maltose, and Melibiose; Leu, Leucine dependent; Try, Tryptophan dependent; phage res., phage resistant..

Isolate ID #	Origin*	Source‡	Date Isolated	Passage number#	Biovar§	Plasmid Profile‖	Pgm¶	Biochemical properties**
North Pre-Caspian region†								
3785	Volga-Ural sandy	*Meriones* fleas	05/08/1971	13	M	110, 70, 9.5	white	T
3786	Volga-Ural sandy	*Meriones tamariscinus*/spleen	05/19/1998	6	M	110, 70, 9.5	white	T
3787	Volga-Ural sandy	*M. tamariscinus/Nosopsyllus laeviceps*	11/05/1998	6	M	110, 70, 9.5	red	T
3788	Volga-Ural sandy	human/blood	06/19/1997	7	M	110, 70, 45-50	red	T
3823	Volga-Ural steppe	*Citellus pygmaeus/Nosopsyllus laeviceps*	04/25/1966	38	M	110, 70, 9.5	white	T, Try
3824	Volga-Ural steppe	*M. tamariscinus*/blood	04/21/1966	38	M	110, 70, 9.5	white	T
3825	Volga-Ural steppe	*M. tamariscinus/Haemaphysalis sp* (tick)	04/25/1966	38	M	110, 70, 9.5	white	T
3826	Volga-Ural steppe	*M. tamariscinus*/fleas	06/21/1982	22	M	110, 70	white	T
3827	Volga-Ural steppe	*Citellus pygmaeus*	06/26/1997	7	M	110, 70, 9.5	white	T
Central-Asian Desert region†								
3789	Ustyurt	*Rhombomys opimus*/liver	05/10/1963	40	M	110, 70, 9.5	white	T, Leu
3790	Ustyurt	*M. meridianus*/blood	11/03/1964	40	O	110, 70, 9.5	white	T
3791	Ustyurt	*Spermophilopsis leptodactylus*/spleen	10/08/1965	39	M	110, 70, 9.5	white	T
3792	Ustyurt	*Mus musculus*/blood	11/21/1971	33	M	110, 70	white	T, Leu
3793	Ustyurt	*R. opimus*/lice	04/25/1957	41	M	110, 70	white	T, Leu

3794	Ustyurt	camel/blood	08/06/1968	42	M	110, 70, 9.5	white	T, Leu
3795	Ustyurt	*R. opimus/Haemaphysalis sp* (tick)	10/09/1973	35	M	110, 70, 9.5	white	T, Leu
3796	Ustyurt	*R. opimus/Echidnophaga oschanini*	10/24/1973	35	M	110, 70	white	Leu
3797	Pre-Balkhash	*R. opimus*/liver	05/31/1938	66	A	70	white	T
3798	Pre-Balkhash	*R. opimus*/blood	08/02/1949	55	M	110, 70, 9.5	white	T
3799	Pre-Balkhash	*R. opimus*/spleen	10/05/1950	54	M	110, 70, 9.5	white	T
3800	Pre-Balkhash	*R. opimus*/blood	05/08/1961	43	M	110, 70, 9.5	white	T
3801	Pre-Balkhash	*M. meridianus*/blood	09/07/1965	39	M	110, 70, 9.5	white	T
3802	Pre-Balkhash	*R. opimus*/fleas	06/01/1998	6	M	70, 9.5	white	T
3803	Pre-Balkhash	*R. opimus*/fleas	05/31/1993	11	M	9.5	red	T, phage res.
3804	Pre-Balkhash	*R. opimus*/blood	07/15/1988	15	M	110, 70, 9.5	red	T
3805	Pre-Balkhash	*R. opimus*/blood	06/17/1989	55	M	70, 9.5	white	T
3806	N. Pre-Aral	human/blood	08/22/1955	59	M	110, 70	white	T
3807	N. Pre-Aral	human/pharynx	07/03/1999	5	M	110, 70	white	T
3808	N. Pre-Aral	human/bubo	08/14/1999	5	M	110, 70, 9.5	white	T
3809	N. Pre-Aral	human/blood	08/07/1999	5	M	110, 70, 9.5	red	T
3810	Muyun-Kum	*R. opimus* flea/*Coptopsylla lamellifer*	10/22/1962	44	M	110, 70	white	T
3811	Muyun-Kum	*R. opimus*	05/30/1969	37	M	110, 70, 9.5	white	T, Try
3812	Muyun-Kum	*R. opimus*/spleen	06/11/1972	40	M	110, 70, 9.5	red	T, Leu
3813	Muyun-Kum	*R. opimus*	10/17/1998	7	M	110, 70, 9.5	red	T
3814	Pre-Aral -Kara-Kum	*R. opimus*/liver	05/08/1998	6	M	110, 70, 9.5	red	T
3815	Pre-Aral -Kara-Kum	human/bubo	06/07/1999	5	M	110, 70, 9.5	red	T
3816	Betpak-Dala	unidentified fleas/burrow	10/21/1962	32	M	110, 70, 9.5	red	T
3817	Betpak-Dala	*Meriones libycus*	11/22/1983	21	M	110, 70, 9.5	red	T
3818	Tau-Kum	unidentified fleas/burrow	05/15/1966	38	M	110, 70, 9.5	white	T
3819	Tau-Kum	*R. opimus*/liver	10/09/1968	36	M	110, 70, 9.5	white	T, Try
3820	Tau-Kum	*R. opimus/Echidnophaga oschanini*	04/24/1971	33	M	110, 70, 9.5	red	T
3821	Tau-Kum	*M. meridianus*/liver	04/18/1971	33	M	110, 70, 9.5	red	T
3822	Tau-Kum	unidentified fleas/burrow	05/29/1988	16	M	110, 70, 9.5	red	T, Try
Tien Shan region[†]								
3828	Sarydzhaz	*Ixodes* Ticks	08/24/1949	55	A	110, 70, 9.5	white	T, Leu
3829	Sarydzhaz	*Rhadinopsylla li ventricosa*	05/12/1962	44	A	110, 70, 9.5	white	T
3830	Sarydzhaz	*Marmota baibacina*	06/28/1962	44	A	110, 70, 9.5	red	T
3831	Sarydzhaz	*Ixodes* Ticks	06/14/1966	40	O	70	white	T
3832	Sarydzhaz	*Marmota baibacina*	06/27/1971	33	M	110, 70	red	T

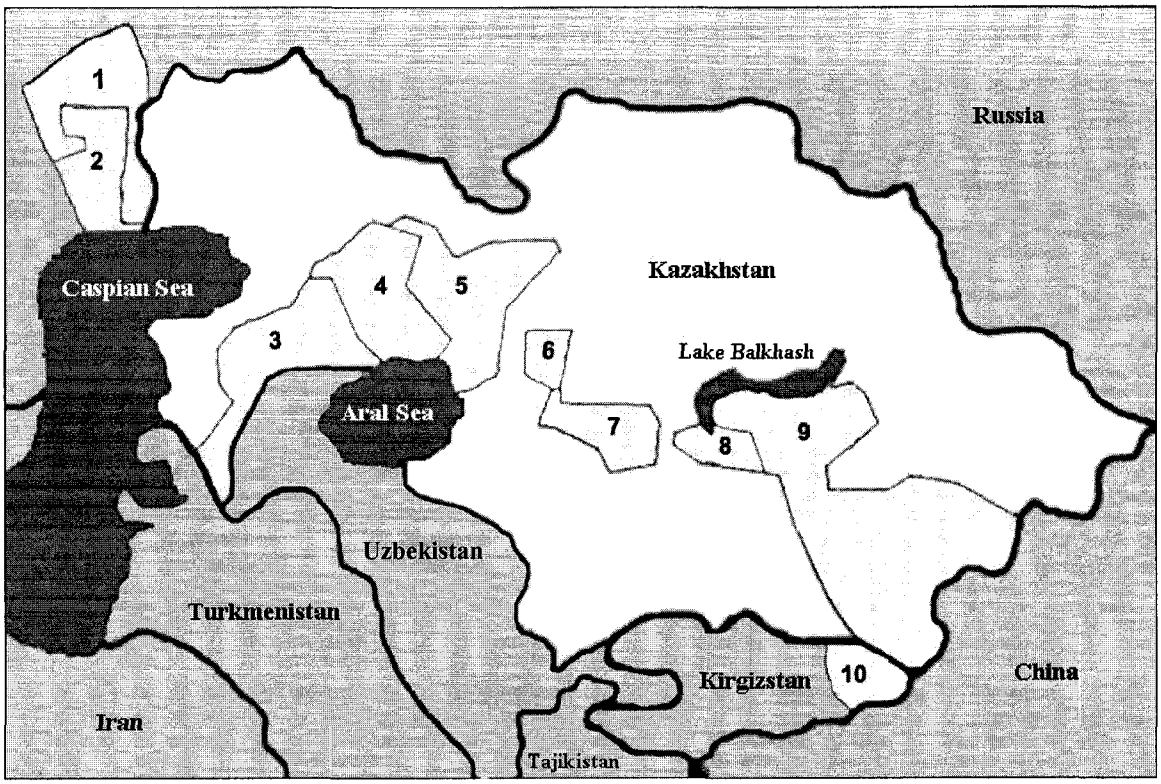

Figure 5.1. Kazakh and adjacent autonomous plague foci from which the 48 Kazakh isolates were collected. 1, Volga-Ural steppe, 2, Volga-Ural sandy, 3, Ustyurt, 4, North Pre-Aral, 5, Pre-Aral Kara-Kum, 6, Betpak-Dala, 7, Muyun-Kum, 8, Tau-Kum, 9, Pre-Balkhash, 10, Sarydzhaz.

media, until 1999, when they arrived at the Centers for Disease Control and Prevention (CDC) in Fort Collins, CO. There they were stored in Heart Infusion Broth (Beckton Dickinson, Sparks, MD) with 10% glycerol at -80 °C.

5.3b Phenotypic analyses. Kazakh isolate biovars were determined using nitrate reduction and glycerol fermentation tests (32). Plasmid profiles were performed according to (66), and the *caf-1, pla,* and *repA1* genes were analyzed by PCR according to (12). Isolates lacking the 110 kb pMT1 plasmid were also confirmed for F1 negativity by Western Blot (101). Amino acid auxotroph screening was performed by plating isolates on minimal medium containing or lacking the corresponding auxotroph amino acid, according to (77). Amino acids tested included Leucine, Tryptophan, Cystein, Phenylalanine, Threonine, and Arginine. Plates were incubated at 25 °C and 37 °C for 72 hours and auxotrophs were defined by negative growth on the minimal medium lacking the corresponding amino acid, and positive growth on minimal medium containing the specifically required amino acid. As controls, five typical *Y. pestis* isolates were plated on the minimal medium and all isolates were plated on sheep blood agar. Bacteriophage lysis tests were also performed according to (18).

Hemin binding capabilities were determined by plating each isolate on Congo Red Agar and incubating for 48 hours at 25 °C (16, 117). The presence or absence of the *pgm* locus was verified for the using *irp2* directed PCR. Each 50 µl PCR reaction contained 1X PCR buffer with 1.5 mM $MgCl_2$, 200µM dNTPs, 0.5 U *Taq* polymerase (Promega, Madison, WI), 0.5 ng of *Y. pestis* DNA, and 10uM each of primers IRP2F (5′-TTTCCGGGGGCGTCTCAAC-3′) and IRP2R

(5′-CTCGACCTTTTGCAGTTTGATGTG-3′). Primers were designed using Primer Select (Dnastar v 7.0, Madison, WI) and the *irp2* sequences of *Y. pestis* CO92 and KIM (GenBank accession numbers NC_003143 and NC_004088 respectively). A negative control containing master mix only and a *Y. pestis* CO92 positive control were included in the set of reactions. All reactions were amplified in a PTC-200 thermal cycler (MJ Research, Watertown, MA) with an initial denaturation step of 95 °C for 5 minutes, followed by 35 cycles of 95 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min, and a final extension step of 72 °C for 10 minutes. Products were 234 bp long and visualized by separating 5μl of the reaction mixture on 2% agarose gels containing 0.1% ethidium bromide.

Sugar requirements and carbohydrate fermentation capabilities of all isolates were determined at KSCQZD according to (7).

5.3c Phylogenetic analyses. MLVA was used to compare isolates from Kazakhstan and from the southwestern U.S. (NA isolates) on three geographic scales. A geographically distant scale was represented by comparing Orientalis, Medievalis, and Antiqua isolates from Kazakhstan to Orientalis isolates from NA. An intermediate geographic scale was represented by comparing the Kazakh isolates from 10 autonomous plague foci, while the NA *Y. pestis* set represented a close geographic scale analysis with isolates from human cases in Arizona, New Mexico, and Colorado (Four Corners region), compared to the animal or flea isolates obtained during the associated epidemiological investigations. These isolates were previously described in (74), but for this analysis an additional 25

variable number of tandem repeat (VNTR) markers were added to the dataset to complete the set described in (69), plus marker M19 from (49), for a total of 42 markers.

DNA was extracted and VNTR marker genotypes were determined according to (74). Each 20 μl PCR multiplex reaction for the additional 25 markers contained 1 X PCR buffer with 1.5 mM $MgCl_2$, 200μM dNTPs, and 0.5 U *Taq* polymerase (Promega Madison, WI), 1.0 μl of DNA template (approximately 0.5 ng DNA) and one of the following multiplex phosphoramidite linkage dye labeled primer sets: 0.1 μM of primer M26 and 0.25 μM of M36, 0.1 μM of M15 and 0.25 μM of M37, 0.1 μM of M42 and 0.2 μM of M49, 0.1 μM of M51 and 0.2 μM of M52, 0.1 μM of M43 and 0.25 μM of M54, 0.1 μM M66 and 0.2 μM of M55, 0.2 μM of M68 and 0.25 μM of M65, 0.1 μM of M69 and 0.2 μM of M56, 0.2 μM of M61 and 0.25 μM of M70, 0.1 μM each of M73 and M74, 0.1 μM each of M72 and M76, 0.1 μM each of M75and M79, and 0.1 μM of M71. Primer sets were run in pairs, as more consistent amplifications were achieved with fewer primers per multiplex. Pairs were chosen based on the primer dye label and on the size of the PCR fragment. All 42 markers were also analyzed for the Kazakh isolates according to (74) and the above primer combinations.

Of the 42 markers, three were variable and 34 were parsimony-informative. The phylogeny was inferred using equally weighted parsimony and 1,000 tree-bisection-reconnection (TBR) searches with a maximum of 20 trees held per search. TBR branch swapping was then performed on all of the most parsimonious trees found with a maximum of 100,000 trees held, from which a strict consensus tree was calculated (76). Jackknife (JK) support (41) was inferred using 1,000 replicates, each consisting of 10

TBR searches and a maximum of 20 trees held (Fig. 5.2). The atypical Antiqua Angola (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=16067) and *Y. pestis* strain 15-70 (Pestoides F) isolates (https://maple.lsd.ornl.gov/microbial/ypes_1570/) were chosen as outgroups (39). Negative PCR results for Angola and Pestoides F were scored as missing data. All phylogenetic analyses were performed using PAUP 4.0b10 (118).

5.4 Results and Discussion

5.4a Phenotypic characteristics. Plasmid profiles revealed that many of the Kazakh isolates were missing one or more of the three unique *Y. pestis* plasmids. Two isolates lacked only pMT1, eight lacked only pPCP1, and two were missing both pPCP1 and pMT1 (Table 5.1). All isolate plasmid profiles corresponded to directed PCR for plasmidial genes, with the exception of isolate 3803. Although this isolate appeared to be missing pMT1 and pCD1, it was PCR positive for *caf-1* and *repA1*, and positive for F1 antigen by Western Blot, indicating chromosomal F1 integration (98). It was possible that the pCD1 plasmid was atypical in size, a characteristic noted in some central-Asian isolates (42), and was not detected in the plasmid profile.

It has been established that the *Y. pestis* plasmids may spontaneously delete during laboratory storage (94), but recent studies have shown that isolates undergoing several passages under laboratory conditions were highly variable in their plasmid

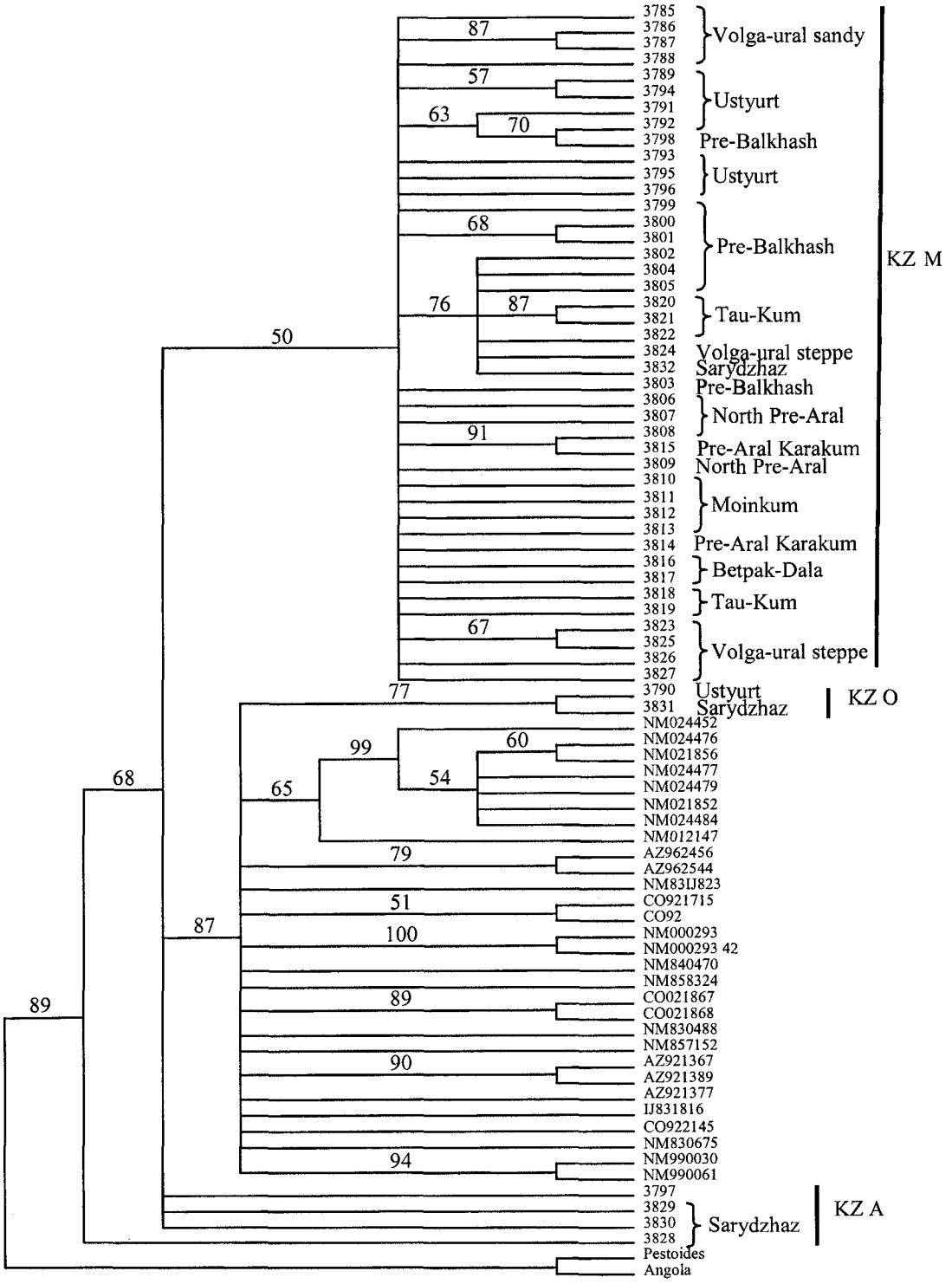

Figure 5.2. Maximum parsimony tree of Kazakh and NA isolates. Numbers above branches refer to jackknife support. Plague focus corresponding to each Kazakh isolate is also listed in Table 1. KZ M, Kazakh Medievalis isolates, KZ O, Kazakh Orientalis isolates, KZ A, Kazakh Antiqua isolates, NA O, NA Orientalis isolates. NA isolates are labeled with CDC accession numbers, NM, New Mexico, CO, Colorado, AZ, Arizona. Numbers following state abbreviation refer to the collection year and the CDC unique identifier.

stability and that one subculture lost both the pPMT1 and pPCP1 plasmids after only four passages, while others lost none after 32 passages (71). Atypical isolates lacking plasmids have also been recovered in nature from around the world (17), including FSU foci (42). Because plasmid profiles were not performed on our Kazakh isolates at the time of collection, it is not possible to determine if plasmid loss occurred naturally or during storage. These isolates may have circulated in nature for thousands of years, undergoing high passage during epizootic activity and isolates lacking pPCP1 or pMT1 may persisted in mammals if the route of transmission was oral or pulmonary (57, 59).

Thirty-two of the 48 Kazakh isolates (66%) did not bind Congo Red (Table 5.1), and were also negative for *irp2* PCR. Absence of the *pgm* locus including the yersiniabactin siderophore iron acquisition system implies avirulence in these isolates (58, 93). This result is most likely a laboratory artifact caused by spontaneous deletion because the isolates were passed many times during storage, or biased selection of pgm^- colonies during original bacterial isolation. This locus is known to spontaneously delete *en bloc* at a high rate (94) and mixed populations of pgm^+ and pgm^- isolates, with high frequency of pgm^- isolates, have been observed from the field (KSCQZD unpublished data, (17). If the majority of the bacterial colonies from the original post epizootic isolations were pgm^-, selection of this phenotype for laboratory storage would be a likely scenario.

Nine Leucine and four Tryptophan dependent auxotrophic phenotypes were found in the isolate collection (Table 5.1). Seven of the nine Leucine dependent auxotrophs were from the Pre-Ustyurt autonomous focus and one was from the Muyun-Kum autonomous focus, both located in the central-Asian natural region, where Leucine

dependence is typically seen. One isolate (3828) was also Leucine dependent, but it was collected from the Sarydzhaz autonomous focus, located in the Tien Shan natural focus, an area that does not typically yield this phenotype. Trypotphan dependent auxotrophs are not usually identified in Kazakh isolate subtypes or autonomous plague foci, however, considering the variety of amino acid auxotrophs noted in the subtypes (77), finding these phenotypes in this collection was not a surprise. No additional amino acid auxotrophs were present.

Sugar fermentation characteristics for these 48 Kazakh isolates were typical for *Y. pestis* found in central-Asian plague foci. Isolate 3803 was phage resistant (Table 5.1).

5.4b Kazakh and NA MLVA comparison. NA (Orientalis) and Kazakh (Medievalis) isolates formed distinct clades. Kazakh isolates 3831 and 3790 fell in the NA clade because they are biovar Orientalis. The NA clade was supported by 87% of JK replicates while the remaining Kazakh isolates formed a monophyletic group with 50% JK support (Fig. 5.2). Markers M58, M59, and M52 were unique to biovar Orientalis.

5.4c Kazakh predefined foci versus Kazakh MLVA. The Kazakh phylogeny generally did not group isolates according to focus, with most isolates unresolved in a polytomy, although there were some exceptions (Fig. 5.2). Two isolates from Volga-Ural sandy (3786, 3787) clustered with 87% JK support, two from Pre-Ustyurt (3789, 3794) with 57% support, two from Pre-Balkhash (3800, 3801) with 68% support, two from the Volga-Ural steppe (3823, 3825) with 67% support, and several isolates from three different foci ((3802, 3804, 3805, (3820, 3821), 3822, 3824, 3832)) formed a clade with

76% support. Within this clade, two isolates from the Tau-Kum focus (3820, 3821) were supported in 87% of JK replicates. An additional group of isolates from the North Pre-Aral (3808) and Pre-Aral Kara-Kum (3815) were supported with 91% JK (Fig. 5.2). These foci border one another, so it is possible that the isolates supported in this clade were collected in close geographic proximity.

Isolates that formed clades supported by at least 70% JK, differed by as many as four VNTR markers and clades that were supported between 50 and 70% differed by as many as seven markers. Clades without support differed by at least eight markers.

This result is expected, given the local variation generated by relatively fast mutation rates in VNTR markers (49) and by sampling on an intermediate geographic scale. Our sample set from Kazakhstan consisted of only a few isolates from each focus but the overall collection spanned most of the country. Isolates from Kazakh plague foci that have potentially existed for thousands of years may have undergone adaptation to local hosts, vectors, or micro-environments, creating a high level of genetic diversity throughout the larger geographic area. Even the relatively young NA isolates that clustered in geographic space (~ collected less than 3 km apart) were strongly supported statistically (49, 74), providing evidence that individual *Y. pestis* clones causing plague epizootics tend to remain fairly isolated. While epizootics may eventually spread to adjacent rodent colonies after rapid animal die-offs like those observed within some rodent species, outbreaks typically remain isolated within a burrow system or rodent community for some time before the epizootic spreads to another location via dispersing juveniles or other mammals carrying infected fleas (9, 49). This pattern of isolated epizootics with little dispersal would maintain local bacterial populations, and supports

the hypotheses that *Y. pestis* populations are maintained in relatively small areas rather than on spatial scales defined by plague foci spanning several thousand square kilometers (Fig. 5.1). The low resolution of the Kazakh phylogeny using MLVA likely resulted from a combination of the geographic scale at which the Kazakh samples were collected, the relative age of Medievalis and Antiqua isolates, and the molecular marker mutation rates.

In contrast, MLVA distinguished between biovars and also provided information about their relative diversity. Biotypes Orientalis and Medievalis formed distinct clades (Fig. 5.2). Antiqua did not form a monophyletic clade, indicating they were more diverse in their MLVA profiles. Orientalis showed the least amount of marker variability with 87% JK support for the NA clade, Medievalis only showed 50% support, and Antiqua was not supported. This result showed signs of genetic diversity that reflect previous analyses that described Antiqua as ancestral to Orientalis and Medievalis (2, 32). Recent single nucleotide polymorphism (SNP) analyses have demonstrated that one Antiqua isolate from Asia (Nepal 516) and several from Africa (Kenya) are polyphyletic and not ancestral to Asian Medievalis and world wide Orientalis isolates respectively (1, 20). However, Asian Antiqua isolates from Japan and the USSR are still shown ancestral to Asian Medievalis, and the atypical Antiqua isolate Angola is ancestral to all other biovars (1). When our MLVA tree was rooted with Angola and *Y. pestis* 15-70, the overall *Y. pestis* phylogeny was highly supported and one Kazakh Antiqua isolate (3828) formed a branch leading to the other *Y. pestis* biovars (Fig. 5.2).

These isolates provided insight into the phenotypic and molecular diversity that exists in Kazakh *Y. pestis*. If these isolates naturally lacked plasmids important for

blocked flea transmission, it may suggest that direct mammal to mammal transmission or mechanical transmission by fleas may be important in epizootic cycles, but evidence for this is sparse. If plasmids were lost because of storage practices, perhaps researchers should consider other possible mutations that may occur in vitro. It was possible that laboratory storage also affected VNTR marker diversity in the faster mutating markers (49). Although MLVA did not group these isolates by focus, we are confident that additional collections of fresh isolates on local scales would yield similar results to those seen in the U.S. While current molecular typing and whole-genome sequencing of *Y. pestis* have provided invaluable insight into the evolution and diversity of *Y. pestis*, much of this work has been performed on old isolates that have probably been stored in a similar manner to this set. Future studies should include phenotypic and genotypic analyses conducted on fresh field isolates and stored isolates that have undergone minimal passages to differentiate between naturally occurring diversity and laboratory artifacts.

5.5 Acknowledgements

This work was supported by the ISTC Biotechnology Engagement Program award K-584 p. We thank Dr. Mark P. Simmons from the Colorado State University Department of Biology for technical assistance with phylogenetic analyses, Sandra Urich, Leon Carter, and Aimee Janusz from the Centers for Disease Control and Prevention, Division of Vector Borne Infectious Diseases for laboratory support, and the Arizona Department of Health Services, New Mexico Department of Health, Colorado Department of Health and Environment, New York State Department of Health, and Navajo Area Indian Health Services for assistance with field investigations. We also thank our colleagues at M. Aikimbayev's Kazakh Scientific Center for Quarantine and Zoonotic Diseases in Almaty, Kazakhstan.

Chapter 6

Single nucleotide polymorphism discovery to determine *Yersinia pestis* population structure in Colorado

This chapter will be submitted for publication in the spring of 2007. The submission date is dependent on the completion of the *Y. pestis* microarray genotyping chip that was developed in part using the isolates presented here. The finished array will be used to validate all SNP data and more isolates will be added to the set prior to manuscript submission. The following authors will be included on this publication.

Lowell, J. L., G. L. Andersen, R. P. Stowkowski, M. F. Antolin, and K. L. Gage.

6.1 Summary

Yersinia pestis, the etiological agent of plague, originated on the central-Asian plateau between 1,500 and 20,000 years ago. It has since spread around the world and occupies every continent except Australia and Antarctica. In the central-Asian countries where *Y. pestis* has existed the longest, ecologically unique areas are defined as plague foci. Studies have indicated that certain bacterial phenotypes and genotypes are confined to these geographically and biologically unique plague foci. *Yersinia pestis* has been in the United States for just over 100 years and in Colorado for approximately 40 years. Rodent epizootic activity is regularly observed statewide in CO where plague foci were previously defined using the same criteria as those in central-Asia, but in the U.S. there is no evidence that unique *Y. pestis* genotypes exist in predefined ecological niches. Studying the genetic diversity of *Y. pestis* in CO is expected to provide information on bacterial population structure. We used DNA microarray technology to discover single nucleotide polymorphisms (SNPs) in the genomes of CO *Y. pestis* isolates that were chosen to represent different predefined plague foci statewide. Isolates from the mountainous front range of Northern CO were compared to isolates from the eastern plains to determine if clonal spread occurs between these two regions or if plague epizootics are independent events. We also examined the presence of non-synonymous SNPs (nsSNPs) compared with synonymous SNPs (sSNPs) for indications of purifying selection or niche adaptation within distinct foci. Microarray analyses yielded higher than expected numbers of SNPs between CO isolates with 62 % nsSNPs overall. Phylogenetic inference indicated that plague foci on the eastern and western plains of CO are separate from mountainous foci. Eastern and western plains clades were supported by

nsSNPs indicating that *Y. pestis* niche adaptation may have occurred in CO. Discovery of SNPs in this set of *Y. pestis* isolates provided molecular markers that were informative for *Y. pestis* population studies. These markers will provide a valuable tool for future studies of *Y. pestis* evolution and population genetics in the U.S.

6.2 Introduction

Yersinia pestis is a pathogenic facultative intracellular bacterium and the causative agent of the plague, a zoonotic disease that affects more than 200 mammal species worldwide. Historically this highly virulent pathogen caused millions of human deaths in three major pandemics. Currently, *Y. pestis* is established in hundreds of natural plague foci around the world where it causes epizootic disease cycling in many rodent species(5, 9, 72, 126). Despite its global distribution and wide host range, there is little information on the population structure of *Y. pestis* in nature, and few studies have identified genetically unique populations of *Y. pestis* that correlate with natural focus boundaries. While *Y. pestis* contains little genetic diversity compared with many other pathogenic bacterial species (83, 87, 89, 108, 111), historic and recent analyses have revealed interstrain variation in *Y. pestis* that may be biologically significant (5, 77, 126).

The completion of several *Y. pestis* complete genome sequences revealed genetic diversity (20) and indicated that *Y. pestis* experienced horizontal gene transfer (HGT) events during speciation. Genome reduction and adaptation followed in response to a change from an enteric lifestyle to a highly virulent insect-borne pathogen (19, 20, 31, 91, 113). Examples included 149 pseudogenes in the *Y. pestis* CO92 genome that were inactivated through a variety of mechanisms such as insertions and deletions, frameshifts, nsSNPs, and interruptions by insertional (IS) elements. Genes that were no longer beneficial to survival were lost during *Y. pestis* adaptation to its vector-borne lifestyle. It follows that the variety of ecological niches, hosts, and vectors that *Y. pestis* occupies may provide additional selective pressures to create genetically distinct populations of *Y. pestis.*

Evidence that *Y. pestis* has undergone niche adaptation and that specific genotypes corresponded to natural disease foci was demonstrated in China and central-Asia (5, 42, 77, 126). Specific genotypes were identified by genomic difference regions (DFRs) (100, 126) phenotypic properties, protein profiles (5, 77), and differential plasmid profiles (42). These analyses provided important clues about the biological importance of *Y. pestis* genetic diversity and possible selective mechanisms, but these studies were limited to isolates from China and central-Asia where *Y. pestis* evolved 1,500 to 20,000 years ago (2). Further, these analyses were based only on a small set of phenotypic differences and on the presence or absence of 4,000 putative coding regions (5, 53, 77) rather than on whole genome differences. Additionally, the geographic regions that defined the unique genotypes were extremely large. The selected molecular markers may not have provided the resolution needed to differentiate bacterial populations between small ecological niches that potentially exist across large landscapes.

Studies examining population dynamics and niche adaptation have not been performed on isolates from the U.S., where *Y. pestis* arrived around the year 1900. While natural plague foci have been identified in the U.S., using the same criteria as in China and central Asia (i.e. by rodent species and ecotypes), there is no evidence that U.S. *Y. pestis* genotypes are defined by species or geographic boundaries. American isolates have been compared for genotypic diversity using a variety of methods such as pulse-field gel electrophoresis (PFGE) (61), IS*100* element and glycerol-3-phosphate dehydrogenase (*glpD*) polymorphisms (84), and multi-locus variable number of tandem repeat analysis (MLVA) (49, 69, 74, 75). None of these methods have identified populations of bacteria that correspond to the natural foci from which they were

collected. Although *Y. pestis* has probably not diverged as much in the U. S. as it has in central Asia, it is more likely that small sample sizes collected opportunistically during human case investigations rather than by systematic and continual surveillance have not provided adequate spatial and temporal sampling of isolates to address questions about population structure. Currently available molecular markers also may not provide the correct information to answer questions concerning hierarchal structure (75).

Higher resolution typing methods that examine the whole genome at the single nucleotide level are expected to provide additional phylogenetic characters and better population resolution than current techniques. By comparing samples of interest to reference sequences, high density oligonucleotide array analysis has been used as an alternative to whole genome sequencing to assess DNA variation (Andersen proposal). High-density arrays have been used for SNP discovery in human chromosomes (54) and can be used to perform comparative sequence analysis between samples of any species for which a reference genome sequence is available. Arrays also provide a method to rapidly compare whole genomes between many isolates of the same species, or closely related species for nucleotide variation (54).

Perelgen Sciences (Mountain View, CA) and Lawerence Berkeley National Laboratory (Berkeley, CA) designed a high density oligonucleotide array using six sequenced *Y. pestis* genomes and one sequenced *Y. pseudotuberculosis* genome. We compared a collection of U.S. *Y. pestis* isolates to the reference array to determine whether *Y. pestis* genotypes are unique to the natural plague foci found in Colorado. Our specific our aim was to determine if plague epizootics spread in waves from the north western CO mountains and foothills to the Pawnee National Grasslands (PNG). The

mountains and the PNG are separate foci consisting of distinct habitats differing in rodent species, flea species, vegetation, elevation, annual rainfall, and seasonal temperatures. Biological and abiotic factors may provide mechanisms for natural selection of *Y. pestis* phenotypes and niche adaptation. By comparing this set of *Y. pestis* genomes we developed a SNP genotyping array capable of capturing genetic diversity of isolates collected throughout Colorado, and of distinguishing *Y. pestis* populations from the mountains and the PNG.

6.3 Methods

6.3a Isolates for SNP discovery using microarray technology. For this study, 15 isolates from CO and two from California were compared for SNP discovery. The CO isolates were chosen from the western slope, central CO, Red Feather, Livermore, the eastern and the western PNG, and from the CO/Nebraska border in Red Lion Wildlife Refuge (Table 6.1, Fig. 6.1). Each of these areas represents a distinct ecological niche with different vegetation, rainfall, and seasonal temperature variation. Rodent species differ from the mountains to the plains but overlap in mountain foci. These isolates were also chosen to increase the likelihood of identifying as much SNP diversity as possible in Colorado *Y. pestis*. Two isolates from CA were included to assess the potential of westward spread of *Y. pestis* from CA to CO.

6.3b Isolate preparation and DNA extractions. All isolates were grown on Congo red agar for 48 hours at 25 °C to screen for the presence of the *pgm* locus (16, 117). One

inoculation loop full of red colonies was selected for each isolate, transferred to 5 ml of Heart Infusion Broth (8) (Beckton Dickinson, Franklin Lakes, NJ) and shaken overnight (ON) at 25 °C and 160 rpm. One ml of each 5 ml ON culture was transferred to 25 ml of HIB, and ON culturing was repeated. Cells were pelleted in oakridge tubes by centrifugation in a Beckman Coulter J series centrifuge (Fullerton, CA) at 8000 rpm for 10 min. Cell pellets were stored at -80 °C until DNA extractions were performed. DNA was extracted using the PureGene DNA Purification Kit™ (Minneapolis, MN) protocol with the following modifications. Cell pellets from 25 ml ON cultures were used rather than 50 ml cultures. Pellets were resuspended in 1.0 ml of 50 mM Tris/50 mM EDTA and transferred to a 50 ml conical tube prior to the cell lysis step. For RNA digestion, a working stock of 100 mg/ml RNase A was prepared and 30 µl (200 µg/ml final concentration) was added to each sample, mixed by inverting the tubes 25 times, and incubated at 37 °C for 30 minutes. Proteinase K (Invitrogen, city, ST) was then added to a final concentration of 100 µg/ml, and the samples were incubated at 55 °C for 1 hour. Following DNA precipitation, the pellets were transferred to 1.5 ml microcentrifuge tubes containing 1.0 ml of 70% ethanol, washed, and centrifuged for 3 minutes at 6000 rpm. The ethanol was removed, the pellets were air dried, and resuspended in 1.0 ml of

Table 6.1. Origins of *Y. pestis* isolates used for SNP discovery. *The CDC accession number consists of the state abbreviation followed by the last two digits of the collection year, except for isolate 80CA2178 in which these two identifiers are reversed. The last four digits are unique identifiers. The extension abbreviations on each accession number are as follows: Chaffee, Chaffee county; RF, Red Feather; Liv, Livermore; PNGE, Pawnee National Grasslands east; PNGW, Pawnee National Grasslands west; Delta, Delta county; Mesa, Mesa county; RL, Red Lion National Wildlife Refuge; Shasta, Shasta county; LA, Los Angeles county.

CDC Accession number*	County	Isolate origin
CO92Chaffee	Chaffee	human blood
CO991133RF	Larimer	human blood
CO042349RF	Larimer	*Cediopsylla inaequalis* (rabbit flea)
CO042792Liv	Larimer	cat blood
CO042794Liv	Larimer	cat blood
CO052871PNGE	Weld	*Cynomys ludovicianus* (black-tailed prairie dog)
CO052847PNGE	Weld	*Cynomys ludovicianus*
CO053907PNGW	Weld	*Cynomys ludovicianus*
CO054186APNGW	Weld	*Pulex sp.* (flea)
CO054186BPNGW	Weld	*Cynomys ludovicianus*
CO041300PNGW	Weld	*Cynomys ludovicianus*
CO041305PNGW	Weld	*Cynomys ludovicianus*
CO052715Delta	Delta	unidentified flea species
CO052626Mesa	Mesa	*Spermophilus variegatus* (rock squirrel)
CO021867RL	Selgwick	*Oropsylla hirsuta* (prairie dog flea)
80CA2178Shasta	Shasta	human blood
CA812298LA	Los Angeles	human blood

Yersinia pestis Isolate Origins

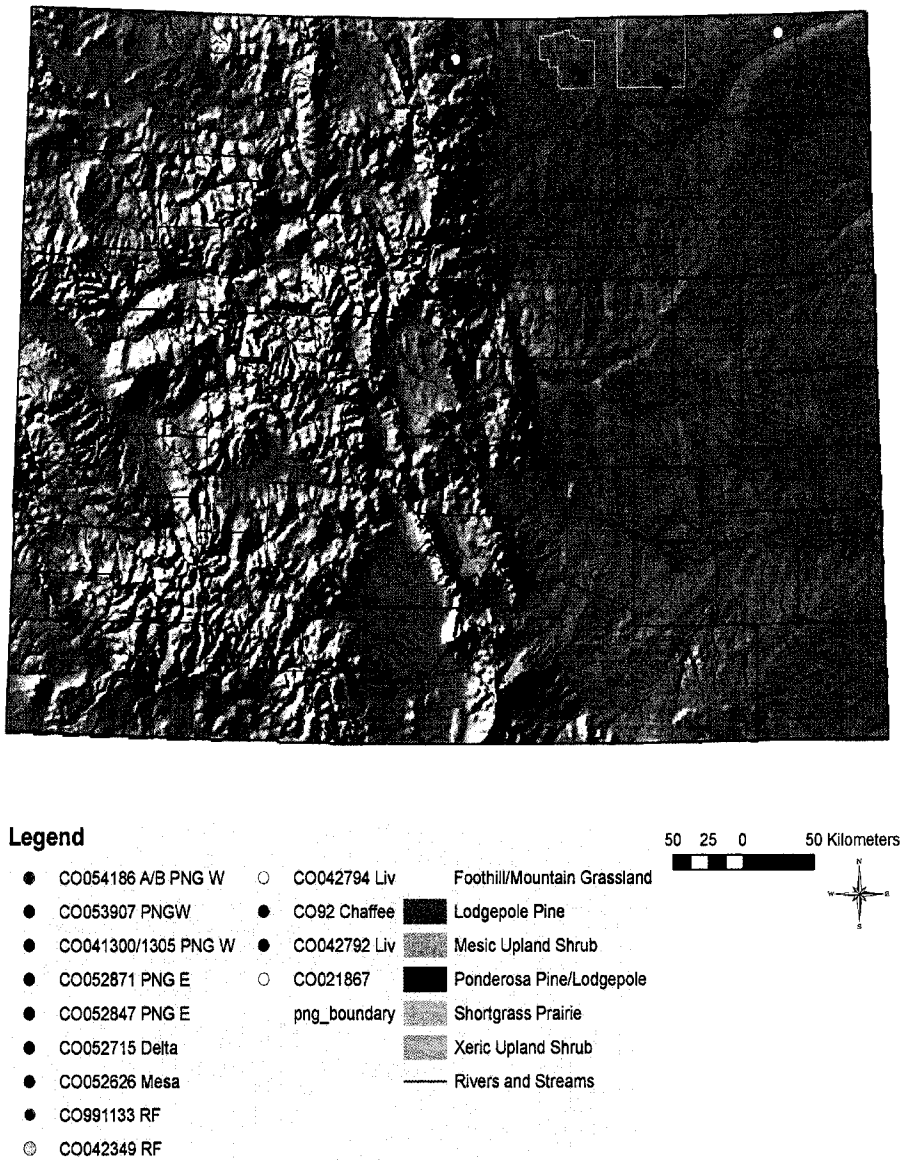

Figure 6.1. Map of *Y. pestis* isolate origins. The letters on the isolate names refer to the state abbreviation followed by the year the isolate was collected. The remaining numbers are CDC unique identifiers. The letter extensions are as follows: PNGW, PNG west; PNGE, PNG east; Delta, Delta county; Mesa, Mesa county; RF, Red Feather; CO, Liv, Livermore, CO; Chafee, Chaffee county; RL, Red Lion Wildlife Refuge.

kit supplied DNA Hydration Solution. DNA was shipped overnight to Lawerence Berkeley Laboratory, (Berkeley, CA) where it was transferred to Perlegen Sciences, (Mountain View, CA) for microarray analysis.

6.3c SNP Discovery Array. Genotyping discovery arrays were developed using 25-bp oligonucleotides designed in four sets (eight sets per SNP) corresponding to forward and reverse strands (Fig. 6.2). Oligonucleotides were designed using the full genome sequences of *Y. pestis* isolates CO92, KIM, Antiqua, Microtus 91001, Nepal 516, and *Y. pseudotuberculosis* IP32953, (websites), and were tiled onto the array according to (44). These reference sequences provided 7.85 Mb of potential SNP positions after redundant regions such as IS and elements repeat regions were removed. Comparisons were also made between each of the reference sequences in order to determine base calling accuracy. For the discovery array, base calling was determined to be 99.8 % accurate.

Our *Y. pestis* isolate DNAs were compared to the reference isolates by hybridization of the samples to the high-density oligonucleotide array containing the reference oligos. A total of 200 μg of DNA per sample was fragmented using DNase I and biotinylated according to Hinds (2004). The resulting labeled oligos were hybridized to the arrays according to Hinds (2004). Hybridization of the labeled sample to the microarray was detected using a confocal laser scanner (92).

6.3d SNP genotyping. To find SNP locations, the thirteenth position of each oligonucleotide was analyzed for a match or mismatch between the reference and queried sequences. SNPs were determined by measuring the ratios of mean intensity of perfect match features to mismatch features on the microarray according to (54) (Fig. 6.2). Each

25 bp oligonucleotide that contained a SNPs at the thirteenth position was queried against the published CO92 *Y. pestis* genome (GenBank Accession NC_003143), using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). SNPs were categorized as intergenic or within an open reading frame (ORF), and as nsSNP or sSNP.

6.3e Phylogenetic analysis. Each SNP was entered into a data matrix and nucleotides were used as character states for phylogenetic analyses. Ambiguous SNPs were identified as Ns and were treated as missing data. The data matrix containing the SNP states was entered into PAUP* version 4.0b10 (118). The phylogeny was inferred using equally weighted parsimony and 1,000 tree-bisection-reconnection (TBR) searches with a maximum of 20 trees held per search. TBR branch swapping was then performed on all of the most parsimonious trees found with a maximum of 100,000 trees held, from which a strict consensus tree was calculated (106) Jackknife (JK) support (41) was inferred using 37% deletion and 1,000 replicates, each consisting of 10 TBR searches and a maximum of 20 trees held. An Antiqua isolate used in previous SNP analyses (1, 20) was added to the analysis and a *Yersinia pestis* pestoides F isolate was used as an outgroup (1, 75)

6.3f Sequence analysis. The ORF sequences containing SNPs were copied into Seqbuilder (Lasergene v.7.0) and the codons containing SNPs were located. The position of the SNP within the codon was used to determine whether mutations were synonymous

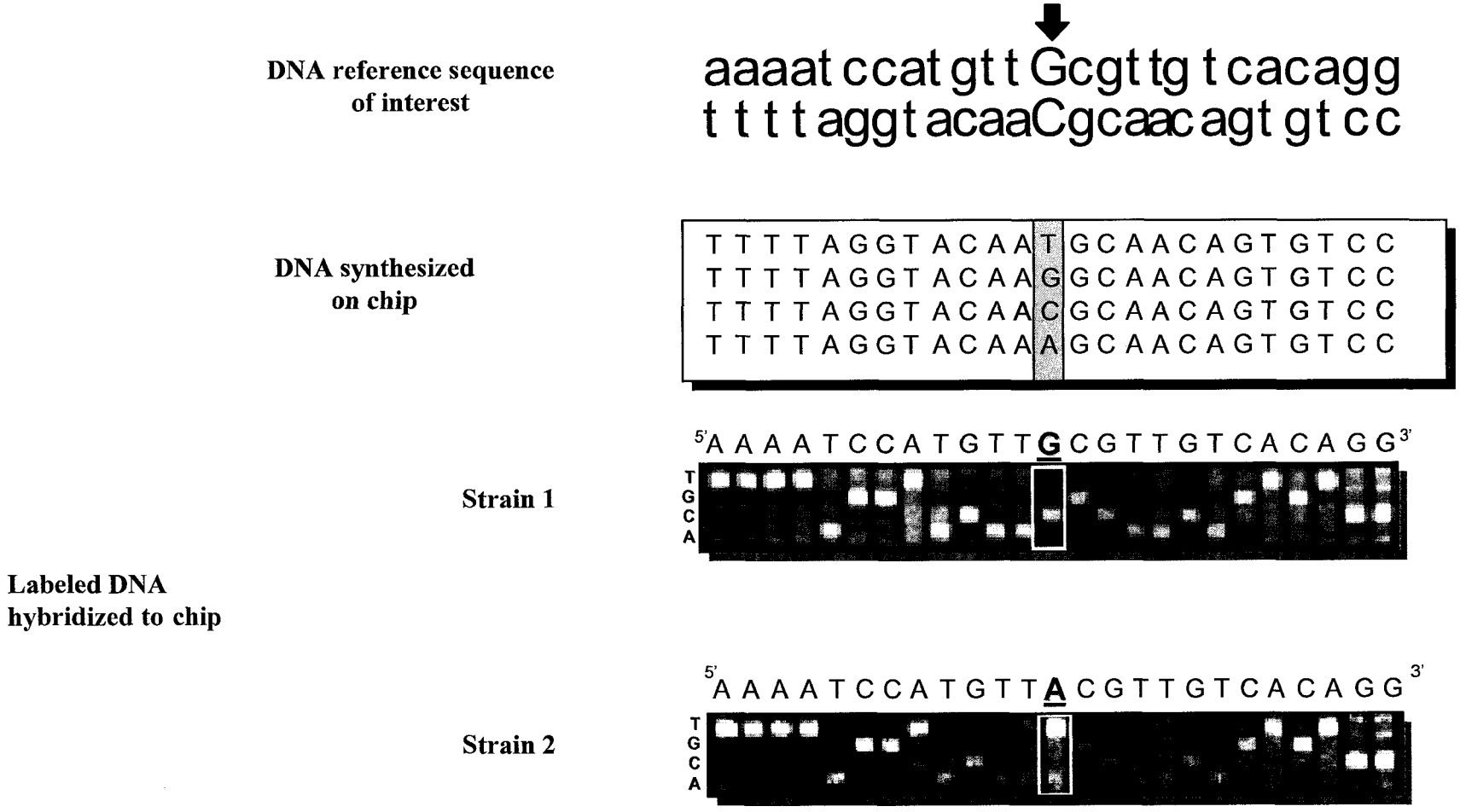

Figure 6.2. Detection of new DNA variation using a high density nucleotide array

or nonsynonymous and whether they were transitions or transversions. Although we expected recombination to be absent from this population based on *Y. pestis* physiology, we tested for linkage disequilibrium to verify this expectation. Tests of linkage disequilibrium were performed between populations using the Index of Association (I_A) in the program START (65) to determine whether population samples were undergoing frequent recombination. Disequilibrium is an indicator of low recombination (14, 110). Each SNP was subcategorized into a functional gene class using Clusters of Orthologous Groups (COGs) according to (20, 125). Once the SNPs were categorized according to functional class, isolates containing unique SNPs and isolates supported in clades by parsimony informative SNPs, were compared to determine if they contained functionally unique mutations. The ratio of mean nsSNPs to sSNPs per site (d_N/d_S) was calculated across each COG functional group (20, 125).

6.4 Results

6.4a SNP discovery. When the 15 CO and two CA isolates were compared to the ancestral *Y. pestis* pestoides F isolate, 69 SNPs (43 nsSNPS) were identified. Sixty SNPs (86 %) were in chromosomal ORFs (35 nsSNPs), nine SNPs (13 %) were in intergenic regions (seven nsSNPs), and three (4 %) were in virulence plasmids (one nsSNP). Forty-six SNPs (66 %) (31 nsSNPS) were isolate specific and 23 SNPs (33 %) (15 nsSNPs) were present in more than one isolate (Table 6.2). Forty-three SNP (62 %) mutations caused transitions and 21 (31 %) caused transversions. Five SNPs (7 %) could not be

characterized because they were ambiguously identified and referred to as Ns. All SNPs were classified into15 COG functional groups.

6.4b Phylogenetic analysis. Parsimony analysis inferred two well supported and two weakly supported clades (Fig. 6.3). Twenty-two SNPs were parsimony informative, 45 were variable, and two were constant. The ensemble consistency index (CI) (70) was 0.88 when uninformative characters were excluded and the ensemble retention index (RI) was 0.95 (40). Isolates from the eastern and western PNG formed monophyletic groups that each received 100% jackknife support with eight character state changes supporting the PNGE group and six character state changes supporting the PNGW group. These synapomorphies were uncontradicted by character reversals or convergences. All SNPs supporting the PNGE and PNGW clades were nsSNPs with the exception of one sSNP supporting the PNGE clade and one sSNP supporting the PNGW. The isolates collected from the western slope (Delta and Mesa counties), Red Feather, Livermore, and the Red Lion Wildlife Refuge, formed a weakly supported clade with 69% jackknife support. The two isolates from CA were unresolved at the base of the tree. The isolate from Chaffee county (CO92) formed a weakly supported clade (63 %) with the Antiqua isolate (Fig 6.3). This result is similar to those of (1, 20).

6.4c Test for recombination and selection. The START program (65) was used to test for horizontal gene transfer (HGT) and recombinational versus random mutational events for the diversity seen in this collection of isolates. The calculated I_A was significantly different from zero (3.6265, P=0.000). The expected variance (V_e) between alleles was

6.859 and the observed variance (V_o) was 31.733. The V_o was significantly greater than the V_e obtained in 1000 trials indicating that this population is in significant linkage disequilibrium. Linkage disequilibrium is the non-random association of alleles from between generations and indicates a clonal population structure in bacteria (111).

6.4d Distributions of nonsynonymous and synonymous SNPs. We investigated whether strain specific mutations and the SNPs belonging to various branches were biased towards COG functional gene classes. Functional class G contained the highest number of isolate specific nsSNPs. Classes C, K, L and R also contained more than one isolate specific nsSNPs (Fig. 6.4a). The remaining isolate specific nsSNPs were unique to individual functional classes. Six isolate specific nsSNPs could not be classified using COG functional groups and three were intergenic (Fig 6.4a). Functional classes E and P each contained two isolate specific sSNPs while the remaining isolate specific sSNPs were unique to individual COG functional classes (Fig. 6.4b). Five isolate specific sSNPs could not be classified using COGs, and none were intergenic (Fig. 6.4b).

Branches leading to the PNGE and PNGW clades consisted almost entirely of nsSNP mutations. Functional groups classified using COGs were unique to each branch (Fig 6.4b). The PNGE and PNGW clades were also each supported by two SNPs that could not be classified using functional groups. The branch leading to the weakly supported mountain clade consisted of one intergenic SNP and one SNP that could not be classified using COGs. The ratio of mean nsSNPs to sSNPs per site (d_N/d_S) was calculated across each COG functional group (20, 125). No nsSNP versus sSNP bias was

Table 6.2. SNP characterization for 15 Colorado isolates and two California isolates.

‡Location in ORF refers to SNP position in the open reading frame or intergenic region. If SNP is intergenic, the gene name is denoted using YPO###_##.

*T, transition; V, transversion

‖AA indicates amino acid change in nsSNPs. If a change is indicated the SNP is non-synonymous

gene SNP location in ORF‡	gene orientation	Isolates containing SNP	SNP location in CO92 COG class	Features related to subject sequence	base change T or V*	Codons containing SNPs‖ AA change
YP0006	+	CO052871	6629	potassium transport protein	C/T	cct
808		CO052847	P		T	P to S
YPCD1.32c_33c	+	CO991133c	23228	pCD1	T/C	tat
25		CO053907	-		T	Y to H
		CO052847				
YPMT1.40c	+	CO052871	40212	pMT1	G/A	tcg
1773		CO052847	-	hypothetical protein	T	
YPMT1.65	+	CO052715_32	66046	pMT1	C/T	cac
306			-	hypothetical protein	T	
YPO0242	-	CO052871	241300	polypeptide deformylase	T/C	V to A
155		CO052847	-		T	
YPO0325_26	+	All except CO92	335311	14 bp at 5' side: single-strand DNA-binding	C/T	tct
26				protein	T	S to F
				209 bp at 3' side: hypothetical protein		
YPO0415	+	CO052871	432784	putative carbohydrate kinase	C/T	gcc
113		CO052847	G		T	A to V
YPO0517	-	CO052626	557456	ATP-dependent helicase HepA	G/A	tgc
2396			KL		T	C to Y

YPO0581 1037	+	CO052626	626912 -	altronate hydrolase	C/T T	ccg P to L
YPO0599 2512	+	CO052871 CO052847	651829 U	putative adhesin	C/T T	cag Q to Stop
YPO0620 23	-	CO052626	683180 -	putative exported protein	G/T V	cgg R to L
YPO0711 496	-	all except CO92	772652 -	putative flagellar motor switch protein	A/G T	agc S to G
YPO0785 1273	-	CO021867	862712 T	phosphoenolpyruvate-protein phosphotransferase	G/T V	?
YPO0798 1072	-	All except CO92	877870 -	putative sugar transport protein	C/T T	ctg
YPO1036_37 219	+	CO92 CO052715-32 CO052626 CO991133c CO042349 CO042792 CO042794 CO021867	1178177	207 bp at 5' side: transposase for the IS1541 insertion element 23 bp at 3' side: conserved hypothetical protein	T/C T	gct
YPO1106 605	-	CO053907 CO054186 CO05 4186-297 CO041300 CO041305	1249983 G	conserved hypothetical protein	C/A T	cca P to Q
YPO1200 1184	+		1351273 E	putative amino acid permease	A/N ?	gta ?
YPO1373 91	-	CO05 4186-297	1543493 CO	transport ATP-binding protein	C/A V	cgc R to S

Table 1 cont.

YPO1527_28 188	+	CO92	1738117	176 bp at 5' side: putative membrane protein 242 bp at 3' side: putative ferric iron reductase	T/C T	ctg L to P
YPO1601 1056	+	CO053907 CO054186 CO05 4186-297 CO041300 CO041305	1825505 IQ	3-oxoacyl-[acyl-carrier-protein] synthase II	C/A V	atc
YPO1709 2525	+	CO052715_32	1952442 NU	putative outer membrane usher protein	G/T V	agt S to I
YPO1791 732	+	80CA2178	2036302 NU	flagellar biosynthesis protein	G/A T	ttg
YPO1815 434	+	CO052626	2063009 G	putative sugar ABC transporter permease	T/G V	ctc L to R
YPO1820 348	-	All except CO92	2067177 NU	flagellar biosynthesis protein	C/G V	ccc
YPO1867 165	+	CO042794	2114501 I	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltr...	T/G V	cgt
YPO1882_83 2311	+	All except CO92	2125486	2298 bp at 5' side: hypothetical protein 484 bp at 3' side: hypothetical protein	T/C T	tat Y to H
YPO1910 8701	-	CO021867	2147260 Q	yersiniabactin biosynthetic protein	G/A T	gag E to K
YPO1910 2676	-	CO021867	2153285 Q	yersiniabactin biosynthetic protein	C/T T	gag
YPO1910 1464	-	CO052871 CO052847	2154497 Q	yersiniabactin biosynthetic protein	C/T T	gcc
YPO1926 760	-	CO042792	2184032 C	putative 4-hydroxybutyrate coenzyme A transferase	G/A T	gga G to R
YPO1930 307	-	CO021867	2187463 -	putative lipoprotein	G/A T	ggc G to S

YPO2002 915	+	CA812298	2275640 R	hypothetical protein	A/C V	gca
YPO2005 119	-	CO92 CO052715_32 CO052626 CO991133c CO042349 CO042792 CO042794 CO021867	2278316 -	putative exported protein	T/C T	gtt V to A
YPO2071 643	+	CO021867	2351074 KL	putative DEAD box family helicase	C/T T	ctt L to F
YPO2153 1014	-	CO021867	2423680 T	hypothetical protein	G/A T	gag
YPO2158 159	+	CO021867	2428649 -	methionine sulfoxide reductase B	C/T T	ctc
YPO2176_77 266	+	CO05 4186-297	2450160	254 bp at 5' side: thymidine kinase 91 bp at 3' side: transposase/IS protein	A/N ?	cag
YPO2182_83 22	+	CO052847	2459227	10 bp at 5' side: periplasmic oligopeptide-binding protein precursor 53 bp at 3' side: oligopeptide transport system permease protein	A/C V	agt S to R
YPO2232 173	+	CO053907 CO054186 CO05 4186-297 CO041300 CO041305	2507280 -	putative transaldolase	C/T T	gcc A to V

Table 1 cont.

YPO2274 606	+	CO92 CO052715_32 CO052626 CO991133c CO042349 CO042792 CO042794 CO052871 CO052847 CO021867 80CA2178 CA812298	2556384 -	putative phage protein	G/A T	tgg W to stop
YPO2310 1431	+	CO021867	2599025 E	putative carboxypeptidase	C/T T	atc
YPO2333 132	-	CO991133c	2623166 P	quaternary ammonium compound-resistance protein	T/C T	att
YPO2348 202	-	CO05 4186-297	2642991 -	hypothetical protein	C/N ?	cgg
YPO2387 814	+	CO052715_32	2681847 K	purine nucleotide synthesis repressor	A/T V	atc I to F
YPO2425 639	-	80CA2178	2726118 H	vtamin B12-transporter permease	C/T T	tgc
YPO2450 62	-	CO042794	2750694 R	hypothetical protein	T/C T	gtg V to A
YPO2660 979	+	CO021867	2986382 E	putative solute-binding protein	G/T V	gat D to Y

YPO2681	+	CO053907	3007059	cel operon repressor	T/C	tac
736		CO054186	-		T	Y to H
		CO05 4186-297				
		CO041300				
		CO041305				
YPO2701	+	CO052626	3029229	hypothetical protein	G/T	gcg
213			-		V	
YPO2816	-	CA812298	3147188	putative phosphatidate cytidylyltransferase	C/A	tac
924			R		V	Y to Stop
YPO2943_44		CO052871	3286072	280 bp at 5' side: transposase for the IS285	C/G	cgt
292		CO052847		insertion element	V	R to G
				1396 bp at 3' side: putative pili chaperone protein		
YPO3032	+	80CA2178	3385893	coproporphyrinogen III oxidase	G/A	gtg
822			-		T	
YPO3097	-	CA812298	3455366	phosphomannomutase	C/T	gct
1016			G		T	A to V
YPO3280_81	+	80CA2178	3662399	910 bp at 5' side: transposase for the IS1541	A/C	aaa
922				insertion element	V	K to Q
				369 bp at 3' side: transposase for insertion sequence IS100		
YPO3300	-	CO042349	3679051	S-ribosylhomocysteinase	T/G	atc
158			-		V	I to S
YPO3322	-	CO052871	3705698	conserved hypothetical protein	C/T	cct
356		CO052847	R		T	P to L
YPO3471	-	CO052847	3876889	sugar transport system permease protein	G/C	atg
744			G		V	M to I

Table 1 cont.

YPO3588 649	+	CO053907 CO054186 CO05 4186-297 CO041300 CO041305	3997221 F	aspartate carbamoyltransferase catalytic subunit	G/A T	gaa E to K
YPO3635 184	+	CO052847	4056094 G	putative branched-chain amino acid transport system perm...	C/T T	ctg
YPO3678 1337	-	CO054186	4108907 -	insecticidal toxin complex	G/N ?	agc
YPO3728 1390	+	80CA2178	4182506 -	bifunctional phosphoribosylaminoimidazolecarboxamide form...	G/A T	gaa E to K
YPO3766 1004	+	CO052871	4229825 I	fatty acid oxidation complex alpha subunit	A/G T	aaa K to R
YPO3768 434	-	CO052871	4232614 HC	NAD(P)H-flavin reductase	A/C V	cac H to R
YPO3841 1072	-	CO052871	4311139 L	DNA helicase II	G/C V	gcc A to P
YPO3905 6	-	CO052871	4379789 G	putative sugar transport system permease	A/T V	tta L to F
YPO3937 939	+	80CA2178	4422197 C	aerobic glycerol-3-phosphate dehydrogenase (partial)	G/A T	cag
YPO3967 581	+	CO021867	4467345 P	phosphate transport protein	C/T T	cct P to L
YPO3980 1004	+	CA812298	4483148 -	putative glycerate kinase	A/G T	caa Q to R
YPO4054 1287	-	CO052847	4571608 E	selenocysteine synthase	A/C V	aca

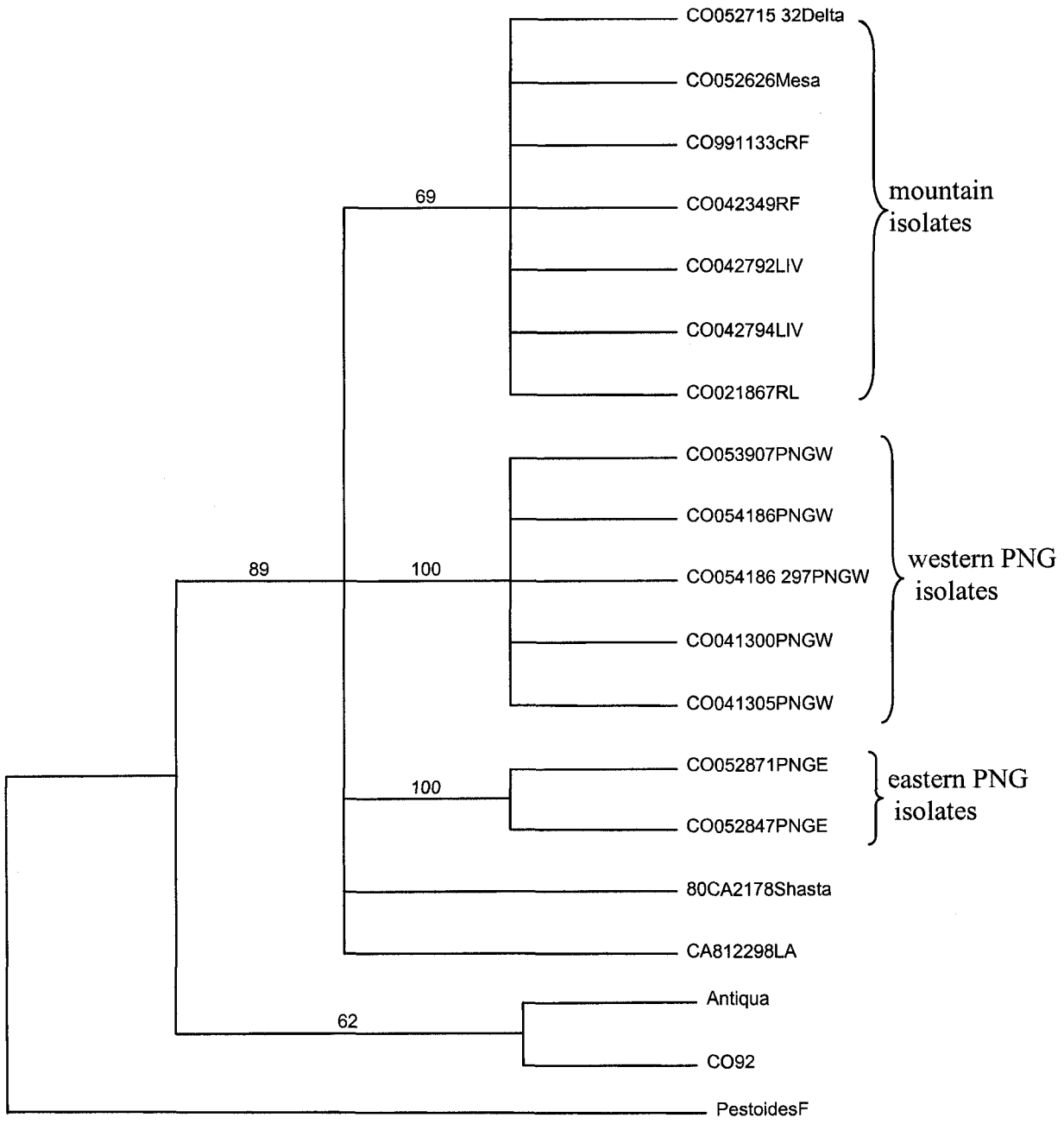

Figure 6.3. Maximum parsimony tree of isolates used for SNP discovery. Isolate abbreviations are listed in Table 1 and Fig. 2. Jackknife support is indicated above branches. All mountain clade isolates are from the CO western slope, Red Feather, or Livermore with the exception of CO021867RL, which is from the eastern plains (Fig. 6.1).

apparent for the isolate specific SNPs with a mean d_N/d_S of 1.5 (20), except for functional class G, which had a large proportion of nsSNPs relative to the mean ratio with a d_N/d_S of 4.0. Although several isolate specific SNPs could be classified into unique functional groups, the numbers of samples and of SNPs were too few to determine if these mutations were due to selection, or just random events. The parsimony informative SNPs that supported each of the PNGE and PNGW clades showed a bias toward nsSNPs with a d_N/d_S of 7.0 for the PNGE and 5.0 for the PNGW.

6.5 Discussion

This was the first comparison of SNPs in a collection of U. S. *Y. pestis* isolates. Even though *Y. pestis* has only resided in the U. S. for approximately 100 years, SNP discovery microarrays revealed large numbers of mutations among only a few isolates from CO. A high overall proportion of SNPs were nonsynonymous (62%) and 81 % of the nsSNPs were within ORFs. Parsimony analysis indicated that populations of *Y. pestis* on the eastern PNG and the western PNG consisted of separate genotypes, and that those populations differed from isolates collected in the mountain foci.

Most of the synapomorphies for the PNGE and PNGW clades were nsSNPs. When classified by COG functional genes, each branch showed a bias toward separate sets of functional classes (Fig 6.4c). The PNGE group contained SNPs that fell into classes G (carbohydrate metabolism), P (inorganic ion metabolism), R (general function prediction only), and U (intracellular trafficking and secretion), and the PNGW group contained SNPs that fell into classes F (nucleotide metabolism), I (lipid metabolism), and Q

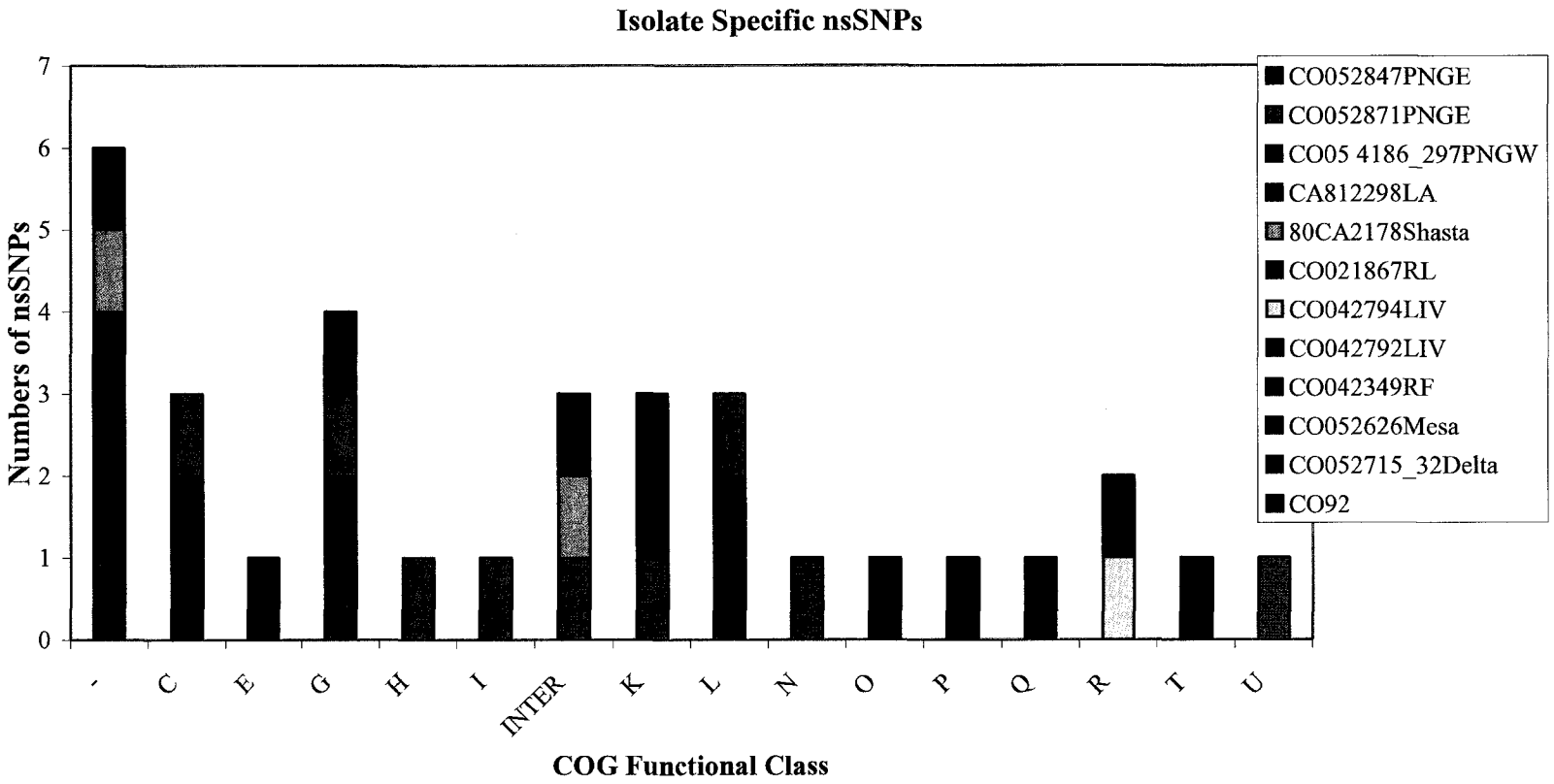

Figure 6.4a. Functional distribution of loci containing isolate specific nsSNPs. The number of isolate specific nsSNPs were grouped by COG functional classes. -, SNP could not be classified by COGs; C, energy production and conversion; E, amino acid transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme metabolism; I, lipid metabolism; INTER, intergenic region; K, transcription; L, DNA replication, recombination and repair; N, cell motility and secretion; O, posttranslational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function (poorly characterized); T, signal transduction; U, intracellular trafficking and secretion.

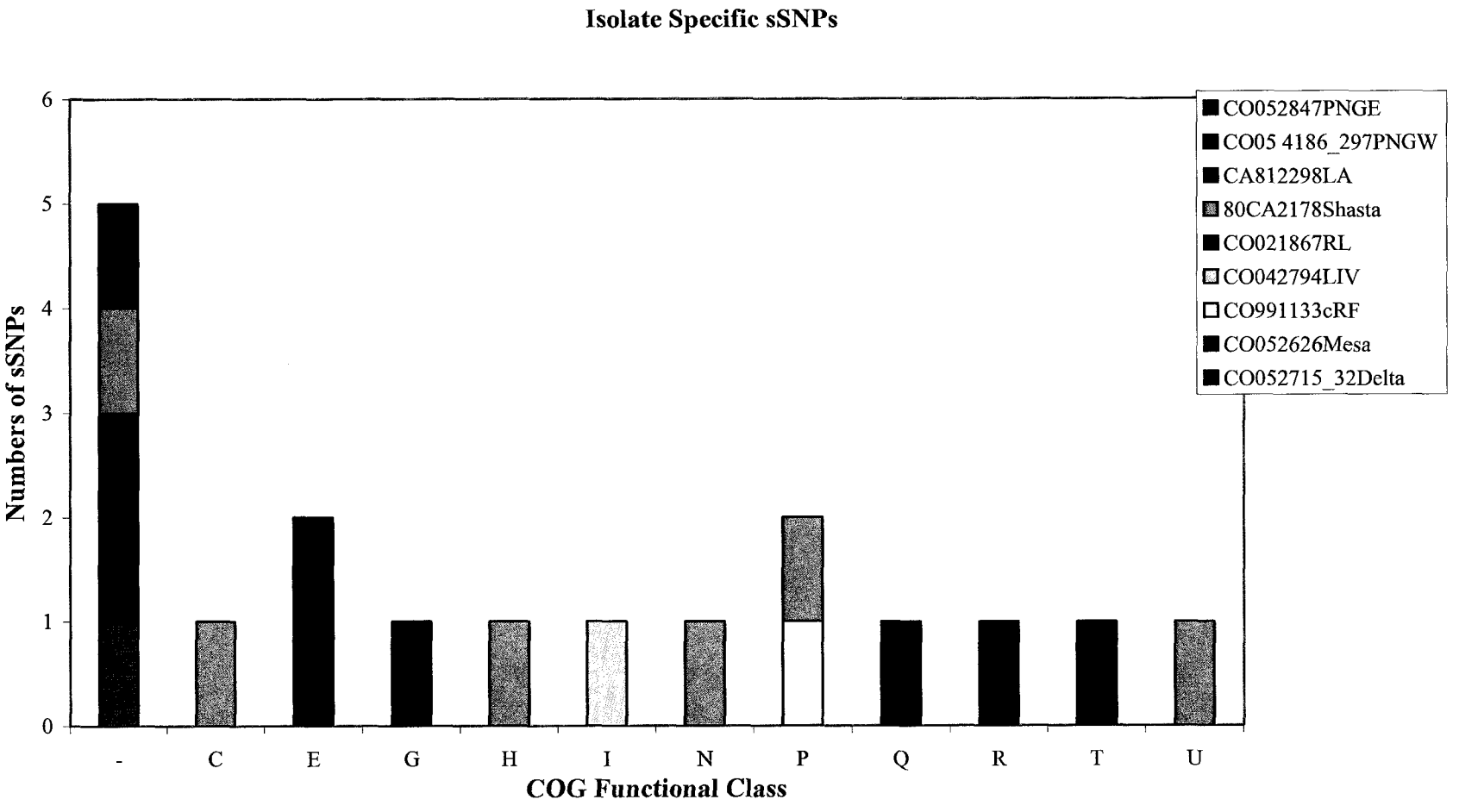

Figure 6.4b. Functional distribution of loci containing isolate specific sSNPs. The number of isolate specific sSNPs were grouped by COG functional classes. COG functional class abbreviations are described in the Fig. 3A legend.

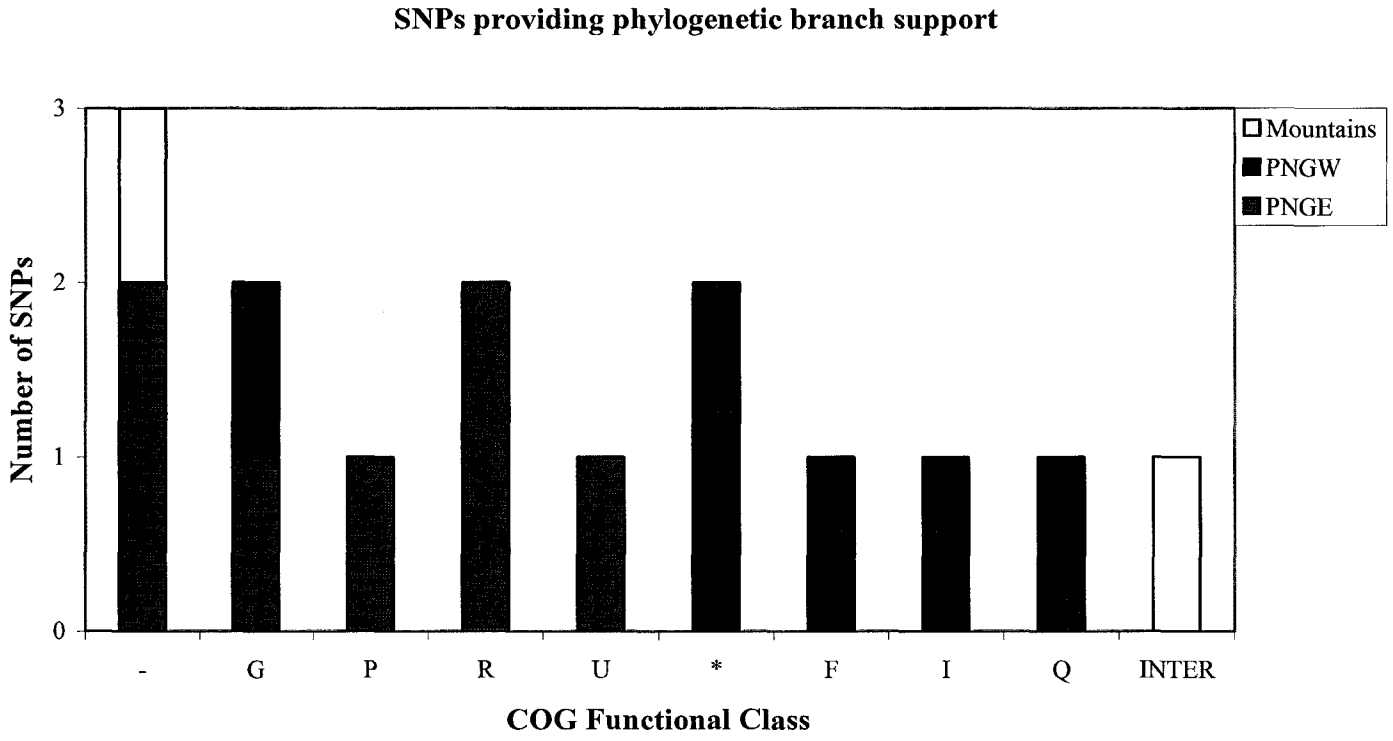

Figure 6.4c. Functional distribution of parsimony informative characters supporting each of the three clades inferred in the maximum parsimony analysis. -, loci supporting the PNGE clade that cannot be classified using COGs; *, loci supporting the PNGW clade that cannot be classified using COGs; F, nucleotide metabolism. The remaining COG functional class abbreviations are described in the Fig. 3A legend.

(secondary metabolite biosynthesis). Each of the two clades were also supported by unique nsSNPs that could not be classified by functional groups but were in ORFs. Only one SNP residing in a non-classified ORF and one SNP in an intergenic region were shared by the isolates from the mountains, western slope, and from Red Lion Wildlife Refuge. These results indicated that the PNGE and PNGW are represented by distinct populations of *Y. pestis* and are separated from populations of *Y. pestis* collected in the mountains in the same years. It was not surprising that bacterial populations differed on the PNG when compared with the mountains as these areas are well separated by diverse habitats and physical barriers. The PNG however, appear non-fragmented, do not contain large geographic or species barriers. It was therefore surprising to see local bacterial populations within the PNG.

The mean d_N/d_S ratio across COG functional classes was 7.0 for the PNGE and 5.0 for the PNGW, indicating weak or no purifying selection, but it is difficult to say whether the functional group bias between the PNGE and PNGW SNPs actually affected bacterial phenotypes. Local population structure is likely a result of natural selection and niche adaptation on small scales, genetic drift, or sampling of epizootic clones instead of the true carrier populations (85, 111).

Isolate specific SNPs did not show any particular bias toward functional classes. The majority of isolate specific SNPs were found in *Y. pestis* collected in the mountains and in the isolate from the Red Lion Wildlife Refuge (Figs. 6.1, 6.4a and 6.4b). These isolates formed a weakly supported clade but the overall group was quite diverse in their SNP profiles and low jackknife support really indicated a polytomy. Although the

isolates from Red Feather and Livermore were collected only three to ten kilometres apart, the topology of the area is mountainous and large geographic barriers create a very fragmented landscape.

The Red Lion isolate contains two SNPs that are also in the mountain isolate genomes indicating they arose from the most recent ancestor. Although the Red Lion Wildlife Refuge is several hundred kilometres from the mountains and ecologically different it is possible that clonal spread occurred between these two areas historically. The Poudre and Platte River drainages extend from the front range foothills to the eastern plains (Fig. 6.1). This drainage may have provided a route for animal dispersal and spread of an ancestral *Y. pestis* clone between the two areas, or human activity may have facilitated bacterial transposition. When more samples from the same areas are added to this analysis, or when ambiguous SNPs currently scored as Ns (missing data) are identified, singleton mutations may become parsimony informative characters that will help provide further resolution.

According to this set of SNPs CO92 is more closely related to the Antiqua isolate than it is to the rest of the CO isolates. This result is congruent with other inferred phylogenies based on sSNPs (1) and on SNPs found through comparison of the five published *Y. pestis* genomes (20). In our study, CO92 and Antiqua shared five of the Pestoides F ancestral SNP states and one derived character independent of the other CO isolates. Perhaps if more sSNPs were available for this study the characters would have been more informative about hierarchy of the CO isolates (1, 20, 125). Multiple

introductions into the U. S. of *Y. pestis* isolates with different evolutionary histories may have also occurred. We will not know until additional isolates from the same area and nationwide from earlier years are analyzed.

A concern when determining bacterial evolution and hierarchical population structure is HGT and recombination, which create reticulated evolutionary patterns. We tested this population for linkage disequilibrium, an indicator of clonal population structure in bacteria (111). The I_A for this isolate set was 3.6 which was significantly different from zero (P=0.000). A zero value is expected from a recombining population (65, 111). These SNPs indicated that our *Y. pestis* population is in linkage disequilibrium and recombinational events are extremely rare. In addition, the C.I. for the maximum parsimony tree was 0.88, reflecting the few inferred convergences and reversals.

6.6 Conclusion

High density microarrays are a fast and effective alternative for interspecies SNP detection independent of whole genome sequencing. Our isolates were useful for the discovery of a large and diverse set of SNPs in CO *Y. pestis* isolates. Although questions remain about population structure and directional spread of *Y. pestis* populations, we now have a tool to answer these questions with additional isolate analyses. As plague surveillance techniques and experimental designs improve, isolate collections will grow and consist of samples that can better address population structure.

Our study was the first analysis of U.S. *Y. pestis* SNPs and the results were extremely interesting. *Yersina pestis* does not appear to have spread from the northern mountains of CO to the PNG during the 2004 or 2005 epizootic events. Rather, these

SNPs indicated that *Y. pestis* is maintained in very local populations, even in areas where plague affects rodents with a metapopulation structure (6). Translocation of *Y. pestis* across landscapes may not occur as frequently as once thought (9). If *Y. pestis* is maintained locally, soil organisms (35, 107), surviving infected fleas, or mammalian hosts with small habitat range may act as reservoirs.

We have not analyzed the same set of isolates using VNTRs or PFGE, however VNTRs have previously revealed similarly local population structure in *Y. pestis* (74). The only major inconsistency noted between VNTRs and our SNP analyses was the placement of CO92 outside of the U.S. clade. Although CO92 was sister to the same Antiqua isolate in other SNP experiments (1, 20), our VNTR analysis (75) placed CO92 within the North American clade and separate from Antiqua.

This data set has provided a starting point for elucidating the evolutionary history, maintenance, and persistence in nature, of a pathogen recently introduced into the U.S. As larger sets of *Y. pestis* isolates are analyzed, questions about population structure and adaptation of this zoonotic pathogen can be answered more definitively.

References

1. **Achtman, M., G. Morelli, P. Zhu, T. Wirth, I. Diehl, B. Kusecek, A. J. Vogler, D. M. Wagner, C. J. Allender, W. R. Easterday, V. Chenal-Francisque, P. Worsham, N. R. Thomson, J. Parkhill, L. E. Lindler, E. Carniel, and P. Keim.** 2004. Microevolution and history of the plague bacillus, Yersinia pestis. Proc Natl Acad Sci U S A **101:**17837-42.
2. **Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule, and E. Carniel.** 1999. Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proc Natl Acad Sci U S A **96:**14043-8.
3. **Adair, D. M., P. L. Worsham, K. K. Hill, A. M. Klevytska, P. J. Jackson, A. M. Friedlander, and P. Keim.** 2000. Diversity in a variable-number tandem repeat from Yersinia pestis. J Clin Microbiol **38:**1516-9.
4. **Aitken, N., S. Smith, C. Schwarz, and P. A. Morin.** 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol Ecol **13:**1423-31.
5. **Anisimov, A. P., L. E. Lindler, and G. B. Pier.** 2004. Intraspecific diversity of Yersinia pestis. Clin Microbiol Rev **17:**434-64.
6. **Antolin, M. F., S. L. T., and E. R. J.** 2006. Landscape features influence genetic structure of black-tailed prairie dogs (Cynomys ludovicianus). Landscape Ecology **21:**867-875.
7. **Aparin, G. P., and E. P. Golubinskii.** 1989. Plague microbiology manual. Irkutsk State University.
8. **Archibald, W. S., and S. J. Kunitz.** 1971. Detection of plague by testing serums of dogs on the Navajo Reservation. HSMHA Health Rep **86:**377-80.
9. **Barnes, A. M.** 1982. Surveillance and control of bubonic plague in the United States. Symp Zool Soc Lond:237-270.
10. **Barnes, A. M., and J. D. Poland.** 1983. Plague in the United States, 1982. MMWR CDC Surveill Summ **32:**19SS-24SS.
11. **Bazanova, L. P., and M. P. Maevskii.** 1996. The duration of the persistence of plague microbe in the body of flea *Citellophilus tesquorum altaicus* in Tuva plague focus. Sci. J. Cent. Control Res. Nat. Infect. Dis. **8:**67-74.
12. **Begier, E. M., G. Asiki, Z. Anywaine, B. Yockey, M. E. Schriefer, P. Aleti, A. Ogden-Odoi, J. E. Staples, C. Sexton, S. W. Bearden, and J. L. Kool.** 2006. Pneumonic plague cluster, Uganda, 2004. Emerg Infect Dis **12:**460-7.
13. **Brouillette, J. A., J. R. Andrew, and P. J. Venta.** 2000. Estimate of nucleotide diversity in dogs with a pool-and-sequence method. Mamm Genome **11:**1079-86.
14. **Brown, A. H. D., M. W. Feldman, and E. Nevo.** 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics **96:**523-536.
15. **Buchrieser, C., C. Rusniok, L. Frangeul, E. Couve, A. Billault, F. Kunst, E. Carniel, and P. Glaser.** 1999. The 102-kilobase pgm locus of Yersinia pestis: sequence analysis and comparison of selected regions among different Yersinia pestis and Yersinia pseudotuberculosis strains. Infect Immun **67:**4851-61.

16. **Burrows, T. W., and S. Jackson.** 1956. The pigmentation of Pasteurella pestis on a defined medium containing haemin. Br J Exp Pathol **37:**570-6.
17. **Cavalcanti, Y. V., N. C. Leal, and A. M. De Almeida.** 2002. Typing of Yersinia pestis isolates from the state of Ceara, Brazil. Lett Appl Microbiol **35:**543-7.
18. **Cavanaugh, D. C., and S. F. Quan.** 1953. Rapid identification of Pasteurella pestis using specific bacteriophage lyophilized on strips of filter paper; a preliminary report. Am J Clin Pathol **23:**619-20.
19. **Chain, P. S., E. Carniel, F. W. Larimer, J. Lamerdin, P. O. Stoutland, W. M. Regala, A. M. Georgescu, L. M. Vergez, M. L. Land, V. L. Motin, R. R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V. Chenal-Francisque, B. Souza, D. Dacheux, J. M. Elliott, A. Derbise, L. J. Hauser, and E. Garcia.** 2004. Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. Proc Natl Acad Sci U S A **101:**13826-31.
20. **Chain, P. S., P. Hu, S. A. Malfatti, L. Radnedge, F. Larimer, L. M. Vergez, P. Worsham, M. C. Chu, and G. L. Andersen.** 2006. Complete genome sequence of Yersinia pestis strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. J Bacteriol **188:**4453-63.
21. **Cho, J. C., and J. M. Tiedje.** 2000. Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. Appl Environ Microbiol **66:**5448-56.
22. **Chu, M.** 2000. Laboratory manual of plague diagnostic tests. *In* C. f. D. C. a. Prevention (ed.), 1 ed. Department of Health and Human Services.
23. **Chu, M. C., X. Q. Dong, X. Zhou, and C. F. Garon.** 1998. A cryptic 19-kilobase plasmid associated with U.S. isolates of Yersinia pestis: a dimer of the 9.5-kilobase plasmid. Am J Trop Med Hyg **59:**679-86.
24. **Collinge, S. K., W. C. Johnson, C. Ray, R. Matchett, J. Grensten, J. F. Cully, K. L. Gage, M. Y. Kosoy, J. E. Loye, and A. P. Martin.** 2005. Landscape structure and plague occurrence in black-tailed prairie dogs on grasslands of the western USA. Landscape Ecology **20:**941-955.
25. **Craven, R. B., and A. M. Barnes.** 1991. Plauge and Tularemia. Infectious Disease Clinics of North America **5:**165-175.
26. **Craven, R. B., G. O. Maupin, M. L. Beard, T. J. Quan, and A. M. Barnes.** 1993. Reported cases of human plague infections in the United States, 1970-1991. Journal of Medical Entomology **30:**758-61.
27. **Cully, J. F., Jr.** 1997. Dynamics of plague from a Gunnison's prairie dog complex in New Mexico. Journal of Wildlife Diseases **33:**706-719.
28. **Cully, J. F., and E. S. Williams.** 2001. Interspecific comparisons of sylvatic plague in prairie dogs. Journal of Mammalogy **82:**894-905.
29. **Davis, S., M. Begon, L. De Bruyn, V. S. Ageyev, N. L. Klassovskiy, S. B. Pole, H. Viljugrein, N. C. Stenseth, and H. Leirs.** 2004. Predictive thresholds for plague in Kazakhstan. Science **304:**736-8.
30. **DeLong, E. F., and N. R. Pace.** 2001. Environmental diversity of bacteria and archaea. Syst Biol **50:**470-8.
31. **Deng, W., V. Burland, G. Plunkett, 3rd, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna, D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L.**

Nilles, J. S. Matson, F. R. Blattner, and R. D. Perry. 2002. Genome sequence of Yersinia pestis KIM. J Bacteriol **184:**4601-11.

32. **Devignat, R.** 1951. [Varieties of Pasteurella pestis; new hypothesis.]. Bull World Health Organ **4:**247-63.
33. **Doll, J. M., P. S. Zeitz, P. Ettestad, A. L. Bucholtz, T. Davis, and K. Gage.** 1994. Cat-transmitted fatal pneumonic plague in a person who traveled from Colorado to Arizona. Am J Trop Med Hyg **51:**109-14.
34. **Dong, X. Q., L. E. Lindler, and M. C. Chu.** 2000. Complete DNA sequence and analysis of an emerging cryptic plasmid isolated from Yersinia pestis. Plasmid **43:**144-8.
35. **Drancourt, M., L. Houhamdi, and D. Raoult.** 2006. Yersinia pestis as a telluric, human ectoparasite-borne organism. Lancet Infect Dis **6:**234-41.
36. **Eisen, R. J., S. W. Bearden, A. P. Wilder, J. A. Montenieri, M. F. Antolin, and K. L. Gage.** 2006. Early-phase transmission of Yersinia pestis by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics. Proc Natl Acad Sci U S A **103:**15380-5.
37. **Eisen, R. J., R. E. Enscore, B. J. Biggerstaff, P. J. Reynolds, P. Ettestad, T. Brown, J. Pape, D. Tanda, C. E. Levy, D. M. Engelthaler, J. Cheek, J. Bueno, R. , J. Targhetta, J. A. Montenieri, and K. L. Gage.** 2007. Human plague in the southwestern United States, 1957-2004: Spatial models of elevated risk of human exposure to *Yersinia pestis*. Journal of Medical Entomology **In Press.**
38. **Enscore, R. E., B. J. Biggerstaff, T. L. Brown, R. E. Fulgham, P. J. Reynolds, D. M. Engelthaler, C. E. Levy, R. R. Parmenter, J. A. Montenieri, J. E. Cheek, R. K. Grinnell, P. J. Ettestad, and K. L. Gage.** 2002. Modeling relationships between climate and the frequency of human plague cases in the southwestern United States, 1960-1997. Am J Trop Med Hyg **66:**186-96.
39. **Farris, J. S.** 1972. Estimating phylogenetic trees from distance matrices. American Naturalist **106:**645-668.
40. **Farris, J. S.** 1989. The retention index and the rescaled consistency index. Cladistics **5:**417-419.
41. **Farris, J. S., V. A. Albert, M. Kallersjo, D. Lipscomb, and A. Kluge.** 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics **12:**99-181.
42. **Filippov, A. A., N. S. Solodovnikov, L. M. Kookleva, and O. A. Protsenko.** 1990. Plasmid content in Yersinia pestis strains of different origin. FEMS Microbiol Lett **55:**45-8.
43. **Finlay, B. J.** 2002. Global dispersal of free-living microbial eukaryotic species. Science **296:**1061-1063.
44. **Fodor, S. P., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas.** 1991. Light-directed, spatially addressable parallel chemical synthesis. Science **251:**767-73.
45. **Gage, K. L.** 1998. Plague, p. 885. *In* L. Colier, A. Balows, and M. Sussman (ed.), Topley and Wilson's Microbiology and Microbial Infections, 9 ed, vol. 3. Oxford University Press, Oxford.
46. **Gage, K. L., and M. Y. Kosoy.** 2005. Natural history of plague: perspectives from more than a century of research. Annu Rev Entomol **50:**505-28.

47. **Gage, K. L., J. A. Montenieri, and R. E. Thomas.** 1994. The role of predators in the ecology, epidemiology, and surveillance of plague in the United States. Proceedings of the Sixteenth Vertebrate Pest Conference **16:**200-206.
48. **Geese, E. M., O. J. Rongstad, and W. R. WMytton.** 1988. Home range and habitat use of coyoes in souteastern Colorado. Journal of Wildlife Management **52:**640-646.
49. **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. Proc Natl Acad Sci U S A **101:**8408-13.
50. **Guarner, J., W. J. Shieh, M. Chu, D. C. Perlman, J. Kool, K. L. Gage, P. Ettestad, and S. R. Zaki.** 2005. Persistent Yersinia pestis antigens in ischemic tissues of a patient with septicemic plague. Hum Pathol **36:**850-3.
51. **Guiyoule, A., B. Rasoamanana, C. Buchrieser, P. Michel, S. Chanteau, and E. Carniel.** 1997. Recent emergence of new variants of Yersinia pestis in Madagascar. J Clin Microbiol **35:**2826-33.
52. **Gutacker, M., N. Conza, C. Benagli, A. Pedroli, M. V. Bernasconi, L. Permin, R. Aznar, and J. C. Piffaretti.** 2003. Population genetics of Vibrio vulnificus: identification of two divisions and a distinct eel-pathogenic clone. Appl Environ Microbiol **69:**3203-12.
53. **Hinchliffe, S. J., K. E. Isherwood, R. A. Stabler, M. B. Prentice, A. Rakin, R. A. Nichols, P. C. Oyston, J. Hinds, R. W. Titball, and B. W. Wren.** 2003. Application of DNA microarrays to study the evolutionary genomics of Yersinia pestis and Yersinia pseudotuberculosis. Genome Res **13:**2018-29.
54. **Hinds, D. A., R. P. Stokowski, N. Patil, K. Konvicka, D. Kershenobich, D. R. Cox, and D. G. Ballinger.** 2004. Matching strategies for genetic association studies in structured populations. Am J Hum Genet **74:**317-25.
55. **Hinnebusch, B. J.** 1997. Bubonic plague: a molecular genetic case history of the emergence of an infectious disease. J Mol Med **75:**645-52.
56. **Hinnebusch, B. J.** 2003. Transmission factors: Yersinia pestis genes required to infect the flea vector of plague. Adv Exp Med Biol **529:**55-62.
57. **Hinnebusch, B. J., E. R. Fischer, and T. G. Schwan.** 1998. Evaluation of the role of the Yersinia pestis plasminogen activator and other plasmid-encoded factors in temperature-dependent blockage of the flea. J Infect Dis **178:**1406-15.
58. **Hinnebusch, B. J., R. D. Perry, and T. G. Schwan.** 1996. Role of the Yersinia pestis hemin storage (hms) locus in the transmission of plague by fleas. Science **273:**367-70.
59. **Hinnebusch, J., P. Cherepanov, Y. Du, A. Rudolph, J. D. Dixon, T. Schwan, and A. Forsberg.** 2000. Murine toxin of Yersinia pestis shows phospholipase D activity but is not required for virulence in mice. Int J Med Microbiol **290:**483-7.
60. **Hoogland, J. L.** 1995. The black-tailed prairie dogs: social life of a burrowing mammal. The University of Chicago Press, Chicago.
61. **Huang, X. Z., M. C. Chu, D. M. Engelthaler, and L. E. Lindler.** 2002. Genotyping of a homogeneous group of Yersinia pestis strains isolated in the United States. J Clin Microbiol **40:**1164-73.

62. **Inglesby, T. V., D. T. Dennis, D. A. Henderson, J. G. Bartlett, M. S. Ascher, E. Eitzen, A. D. Fine, A. M. Friedlander, J. Hauer, J. F. Koerner, M. Layton, J. McDade, M. T. Osterholm, T. O'Toole, G. Parker, T. M. Perl, P. K. Russell, M. Schoch-Spana, and K. Tonat.** 2000. Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. Jama **283:**2281-90.
63. **Istock, C. A., K. E. Duncan, N. Ferguson, and X. Zhou.** 1992. Sexuality in a natural population of bacteria--Bacillus subtilis challenges the clonal paradigm. Mol Ecol **1:**95-103.
64. **Iteman, I., C. Baril, I. Saint Girons, and E. Carniel.** 1991. Pulse field electrophoresis of the chromosome of the pathogenic yersiniae. Contrib Microbiol Immunol **12:**198-202.
65. **Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden.** 2001. Sequence type analysis and recombinational tests (START). Bioinformatics **17:**1230-1.
66. **Kado, C. I., and S. T. Liu.** 1981. Rapid procedure for detection and isolation of large and small plasmids. J Bacteriol **145:**1365-73.
67. **Kartman, L., S. F. Quan, and R. R. Lechleitner.** 1962. Die-off of a Gunnison's prairie dog colony in central Colorado: Retrospective determination of plague infection in flea vectors, rodents, and man. Zoonoses Research **1:**201-224.
68. **Keim, M. E., N. Pesik, and N. A. Twum-Danso.** 2003. Lack of hospital preparedness for chemical terrorism in a major US city: 1996-2000. Prehospital Disaster Med **18:**193-9.
69. **Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim.** 2001. Identification and characterization of variable-number tandem repeats in the Yersinia pestis genome. J Clin Microbiol **39:**3179-85.
70. **Kluge, A., and J. S. Farris.** 1969. Quantitative phyletics and the evolution of Anurans. Systematic Zoology **18:**1-32.
71. **Leal-Balbino, T. C., N. C. Leal, C. V. Lopes, and A. M. Almeida.** 2004. Differences in the stability of the plasmids of Yersinia pestis cultures in vitro: impact on virulence. Mem Inst Oswaldo Cruz **99:**727-32.
72. **Link, V. B.** 1955. A history of plague in the United States of America, p. 1-2. *In* E. United States Department of Health, and Welfare (ed.). Washington, D. C.
73. **Link, V. B.** 1949. Plague among wild rodents in Rio Arriba County, New Mexico. American Journal of Tropical Medicine & Hygiene **29:**493-500.
74. **Lowell, J. L., D. M. Wagner, B. Atshabar, M. F. Antolin, A. J. Vogler, P. Keim, M. C. Chu, and K. L. Gage.** 2005. Identifying sources of human exposure to plague. J Clin Microbiol **43:**650-6.
75. **Lowell, J. L., A. Zhansarina, B. Yockey, T. Meka-Mechenko, G. Stybayeva, B. Atshabar, L. Nekrassova, R. Tashmetov, K. Kenghebaeva, M. C. Chu, M. Kosoy, M. F. Antolin, and K. L. Gage.** 2007. Phenotypic and molecular characterizations of Yersinia pestis isolates from Kazakhstan and adjacent regions. Microbiology **153:**169-77.
76. **Maddison, D. R.** 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. **40:**315-328.

77. **Martinevsky, I. L.** 1973. Materials for typing natural foci of plague according to the genetic properties of plague bacillus strains. J Hyg Epidemiol Microbiol Immunol **17:**272-8.
78. **McCoy, G. W.** 1910. The susceptibility to plague of the weasel, the chipmunk, and the pockey gopher. Journal of Infectious Diseases.
79. **Metzer, M. L., D. P. Mindell, X. M. Liu, R. G. Ptak, R. A. Gibbs, and D. M. Hillis.** 2002. Molecular evidence of HIV-1 transmission in a criminal case. Proc Natl Acad Sci U S A **99:**14292-14297.
80. **MMWR.** 2006. Human plague-four states, 2006. Morbidity and Mortality Weekly Report **55:**1-3.
81. **MMWR.** 2002. Imported plague: New York City, 2002. MMWR **52:**725-728.
82. **Mollaret, H. H., and C. Mollaret.** 1965. Melibiose fermentation in the genus *Yersinia* and its importance in the diagnosis of the varieties of *Yersinia pestis*. Bull. Soc. Pathol. Exot. Fil. **58:**154-156.
83. **Moorhead, S. M., G. A. Dykes, and R. T. Cursons.** 2003. An SNP-based PCR assay to differentiate between Listeria monocytogenes lineages derived from phylogenetic analysis of the sigB gene. J Microbiol Methods **55:**425-32.
84. **Motin, V. L., A. M. Georgescu, J. M. Elliott, P. Hu, P. L. Worsham, L. L. Ott, T. R. Slezak, B. A. Sokhansanj, W. M. Regala, R. R. Brubaker, and E. Garcia.** 2002. Genetic variability of Yersinia pestis isolates as predicted by PCR-based IS100 genotyping and analysis of structural genes encoding glycerol-3-phosphate dehydrogenase (glpD). J Bacteriol **184:**1019-27.
85. **Musser, J. M.** 1996. Molecular population genetic analysis of emerged bacterial pathogens: selected insights. Emerg Infect Dis **2:**1-17.
86. **Nelson, B. C.** 1980. Plague studies in California-the roles of various species of sylvatic rodents in plague ecology in California. Proceedings of the Ninth Vertebrate Pest Conference:89-96.
87. **Nubel, U., R. Reissbrodt, A. Weller, R. Grunow, M. Porsch-Ozcurumez, H. Tomaso, E. Hofer, W. Splettstoesser, E. J. Finke, H. Tschape, and W. Witte.** 2006. Population structure of Francisella tularensis. J Bacteriol **188:**5319-24.
88. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.
89. **Ochman, H., T. S. Whittam, D. A. Caugant, and R. K. Selander.** 1983. Enzyme polymorphism and genetic population structure in Escherichia coli and Shigella. J Gen Microbiol **129:**2715-26.
90. **Pace, N. R.** 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734-40.
91. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature **413:**848-52.

92. **Patil, N., A. J. Berno, D. A. Hinds, W. A. Barrett, J. M. Doshi, C. R. Hacker, C. R. Kautzer, D. H. Lee, C. Marjoribanks, D. P. McDonough, B. T. Nguyen, M. C. Norris, J. B. Sheehan, N. Shen, D. Stern, R. P. Stokowski, D. J. Thomas, M. O. Trulson, K. R. Vyas, K. A. Frazer, S. P. Fodor, and D. R. Cox.** 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science **294:**1719-23.

93. **Perry, R. D., S. W. Bearden, and J. D. Fetherston.** 2001. Iron and heme acquisition and storage systems of Yersinia pestis. Rec Res Dev Microbiol**:**13-27.

94. **Perry, R. D., and J. D. Fetherston.** 1997. Yersinia pestis--etiologic agent of plague. Clin Microbiol Rev **10:**35-66.

95. **Poland, J. D., and A. M. Barnes.** 1979. Plague, p. 515-559. *In* J. H. Steele (ed.), CRC Handbook Series in Zoonoses. Section A: Bacterial, Rickettsial and Mycotic Diseases. Volume I. CRC Press Inc., Boca Raton, FL.

96. **Pollitzer, R., and K. F. Meyer.** 1961. Studies in Disease Ecology. Hafner, New York.

97. **Pourcel, C., F. Andre-Mazeaud, H. Neubauer, F. Ramisse, and G. Vergnaud.** 2004. Tandem repeats analysis for the high resolution phylogenetic analysis of Yersinia pestis. BMC Microbiol **4:**22.

98. **Protsenko, O. A., A. A. Filippov, and V. V. Kutyrev.** 1991. Integration of the plasmid encoding the synthesis of capsular antigen and murine toxin into Yersinia pestis chromosome. Microb Pathog **11:**123-8.

99. **Quan, T. J., A. M. Barnes, L. G. Carter, and K. R. Tsuchiya.** 1985. Experimental plague in rock squirrels, *Spermophilus variegatus* (Erxleben). Journal of Wildlife Diseases **21:**205-210.

100. **Radnedge, L., P. G. Agron, P. L. Worsham, and G. L. Andersen.** 2002. Genome plasticity in Yersinia pestis. Microbiology **148:**1687-98.

101. **Rahalison, L., E. Vololonirina, M. Ratsitorahina, and S. Chanteau.** 2000. Diagnosis of bubonic plague by PCR in Madagascar under field conditions. J Clin Microbiol **38:**260-3.

102. **Rust, J. H., D. C. Cavanaugh, R. O'Shita, and J. D. Marshall.** 1971. The role of domestic animals in the epidemiology of plague. I. Experimental infection of dogs and cats. Journal of infectious diseases **124:**522-526.

103. **Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, and D. Altshuler.** 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409:**928-33.

104. **Salkeld, D. J., and P. Stapp.** 2006. Seroprevalence rates and transmission of plague (*Yersinia pestis*) in mammalian carnivores. Vector borne and zoonotic diseases **in press**.

105. **Saratov.** 1979. Instruction on epidemiological surveillance in the natural plague foci of Soviet Union, p. 60.

106. **Schuh, R. T., and J. T. Polhemus.** 1980. Analysis of taxonomic congruence among morphological, ecological, and biogeographic data sets for the Leptopodomorpha (Hemiptera). Systematic Zoology **29:**1-26.
107. **Sebbane, F., C. O. Jarrett, D. Gardner, D. Long, and B. J. Hinnebusch.** 2006. Role of the Yersinia pestis plasminogen activator in the incidence of distinct septicemic and bubonic forms of flea-borne plague. Proc Natl Acad Sci U S A **103:**5526-30.
108. **Selander, R. K., and B. R. Levin.** 1980. Genetic diversity and structure in Escherichia coli populations. Science **210:**545-7.
109. **Smith, C. R., B. C. Nelson, and A. M. Barnes.** 1984. The use of wild carnivore serology in determining patterns of plague activity in rodents in California. Proceedings of the Eleventh Vertebrate Pest Conference **11:**71-76.
110. **Smith, J. M., and N. H. Smith.** 1998. Detecting recombination from gene trees. Mol. Biol. Evol. **15:**590-599.
111. **Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc Natl Acad Sci U S A **90:**4384-4388.
112. **Sodeinde, O. A., and J. D. Goguen.** 1989. Nucleotide sequence of the plasminogen activator gene of *Yersinia pestis*: Relationship to *ompT* of *Escherichia coli* and gene E of *Salmonella typhimurium*. Infect Immun **57:**1517-1523.
113. **Song, Y., Z. Tong, J. Wang, L. Wang, Z. Guo, Y. Han, J. Zhang, D. Pei, D. Zhou, H. Qin, X. Pang, Y. Han, J. Zhai, M. Li, B. Cui, Z. Qi, L. Jin, R. Dai, F. Chen, S. Li, C. Ye, Z. Du, W. Lin, J. Wang, J. Yu, H. Yang, J. Wang, P. Huang, and R. Yang.** 2004. Complete genome sequence of Yersinia pestis strain 91001, an isolate avirulent to humans. DNA Res **11:**179-97.
114. **Souza, V., T. T. Nguyen, R. R. Hudson, D. Pinero, and R. E. Lenski.** 1992. Hierarchical analysis of linkage disequilibrium in Rhizobium populations: evidence for sex? Proc Natl Acad Sci U S A **89:**8389-93.
115. **Stapp, P., M. F. Antolin, and M. Ball.** 2004. Patterns of extinction in prairie dog metapopulations: plague outbreaks follow El Nino events. Front Ecol Environ **2:**235-240.
116. **Stenseth, N. C., N. I. Samia, H. Viljugrein, K. L. Kausrud, M. Begon, S. Davis, H. Leirs, V. M. Dubyanskiy, J. Esper, V. S. Ageyev, N. L. Klassovskiy, S. B. Pole, and K. S. Chan.** 2006. Plague dynamics are driven by climate variation. Proc Natl Acad Sci U S A **103:**13110-5.
117. **Surgalla, M. J., and E. D. Beesley.** 1969. Congo red-agar plating medium for detecting pigmentation in Pasteurella pestis. Appl Microbiol **18:**834-7.
118. **Swofford, D. L.** 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), 4.0b10 ed ed. Sinauer Associates, Sunderland, MA.
119. **Tenborg, M., B. Davis, D. Smith, C. E. Levy, B. Koehler, D. Tanda, J. Pape, R. Hoffman, R. E. Fulgham, B. Joe, and J. Cheek.** 1996. Fatal human plague-Arizona and Colorado. Morbidity and Mortality Weekly Report **46:**617-620.
120. **Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas,**

J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al. 2001. The sequence of the human genome. Science **291:**1304-51.

121. **Welkos, S. L., K. M. Davis, L. M. Pitt, P. L. Worsham, and A. M. Freidlander.** 1995. Studies on the contribution of the F1 capsule-associated plasmid pFra to the virulence of Yersinia pestis. Contrib Microbiol Immunol **13:**299-305.
122. **Whitaker, R. J., D. W. Grogan, and J. W. Taylor.** 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science **301:**976-8.
123. **WHO.** 2004. Weekly epidemiological record. WHO Weekly Epidemiological Record **33:**301-308.
124. **Wolff, K. L., and B. W. Hudson.** 1974. Paper-strip blood-sampling technique for the detection of antibody to the plague organism *Yersinia pestis*. Appl Microbiol **28:**323-325.
125. **Zhang, W., W. Qi, T. J. Albert, A. S. Motiwala, D. Alland, E. K. Hyytia-Trees, E. M. Ribot, P. I. Fields, T. S. Whittam, and B. Swaminathan.** 2006. Probing genomic diversity and evolution of Escherichia coli O157 by single nucleotide polymorphisms. Genome Res **16:**757-67.
126. **Zhou, D., Y. Han, Y. Song, P. Huang, and R. Yang.** 2004. Comparative and evolutionary genomics of Yersinia pestis. Microbes Infect **6:**1226-34.