2015-11-01

The Detection and Molecular Evolution of Francisella tularensis Subspecies

Mark K. Gunnell
Brigham Young University - Provo

The Detection and Molecular Evolution of

Francisella tularensis Subspecies

Mark K. Gunnell

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Richard A. Robison, Chair
Byron J. Adams
Joel S. Griffitts
William R. McCleary
Brent L. Nielsen

Department of Microbiology and Molecular Biology

Brigham Young University

November 2015

ABSTRACT

The Detection and Molecular Evolution of
Francisella tularensis Subspecies

Mark K. Gunnell
Department of Microbiology and Molecular Biology, BYU
Doctor of Philosophy

Francisella tularensis is the etiological agent of tularemia, a zoonotic disease with worldwide prevalence. *F. tularensis* is a highly pathogenic organism and has been designated as a potential biothreat agent. Currently there are four recognized subspecies of *F. tularensis*: *tularensis* (type A), *holarctica* (type B), *mediasiatica*, and *novicida*. In addition, genomic studies have further subdivided type A *tularensis* into two subclassifications, type A.I and type A.II. These two subclassifications differ in geographic distribution with type A.I appearing mainly in the Eastern United States and type A.II appearing mainly in the Western United States. Because of differences of virulence among the subspecies, it is important to be able to quickly identify each of the subspecies rapidly and accurately. This work describes the development of a multiplex real-time polymerase chain reaction (PCR) assay which was shown to be ~98% successful at identifying the known subspecies of *F. tularensis*. Furthermore, *F. tularensis* is thought be a genome in decay (losing genes) because of the relatively large number of pseudogenes present in its genome. We hypothesized that the observed frequency of gene loss/pseudogenes may be an artifact of evolution in response to a changing environment, and that genes involved in virulence should be under strong positive selection. Eleven arbitrarily chosen virulence genes were screened for positive selection along with 10 arbitrarily chosen housekeeping genes. Analyses of selection yielded one housekeeping gene and 7 virulence genes which showed significant evidence of positive selection. Our results suggest that while the loss of functional genes through disuse could be accelerated by negative selection, the genome decay in *Francisella* could also be the byproduct of adaptive evolution, as evidenced by several of its virulence genes which are undergoing strong, positive selection.

Keywords: *Francisella tularensis*, real-time PCR, detection, genome decay, genome sequencing natural selection, virulence, TreeSAAP

ACKNOWLEDGEMENTS

I would like to thank Dr. Richard A. Robison for his guidance, encouragement and mentoring throughout my academic career. I would also especially like to thank Dr. Byron J. Adams for serving on my advisory committee and for his guidance and thoughtful review of my manuscripts. Thanks also to the rest of my advisory committee: Drs. Joel Griffitts, William R. McCleary, and Brent Nielsen.

I would like to thank Dr. Bruce Harper for inspiring me and providing the avenue to pursue this degree as well as Dr. Angelo Madonna for his guidance and encouragement. Thanks also to Shawn Slater for helping to talk though concepts and helping to refine the ideas presented in this work. Most importantly, I would like to thank my family: my parents for instilling in me from a very young age, a desire for knowledge and the importance of education, my children, Spencer and Emily, for understanding when I had to be away working and at school, and my wife, Jennifer, for her encouragement, support, understanding and love.

For the literature review presented in Chapter 1, "The Genetic Diversity and Evolution of *Francisella tularensis* with Comments on Detection by PCR," (1) I thank my co-authors: Byron J. Adams and Richard A. Robison. Additionally, I thank Dr. Angelo Madonna of Dugway Proving Ground for his guidance and leadership in the production of this work.

For the research in Chapter 2, "A multiplex real-time PCR assay for the detection and differentiation of *Francisella tularensis*," (2) I thank my co-authors: Charity D. Lovelace, Benjamin A. Satterfield, Emily A. Moore, Kim L. O'Neill, and Richard A. Robison. Also, I thank the Utah Department of Health, the New Mexico Department of Health, Dr. Fran Nano at the University of Victoria, Rocky Mountain Laboratories, and Dugway Proving Ground for help in obtaining the *Francisella* isolates, without which this work would have been impossible. I

thank Dr. Paul Keim of Northern Arizona University for his MLVA database. Finally, I thank Scott Jonas of Dugway Proving Ground for expert guidance in performing the MLVA assays, as well as Dr. Angelo Madonna also of Dugway Proving Ground for guidance and technical help.

For the research in Chapter 3 "Natural selection in virulence genes of *Francisella tularensis*," I thank my coauthors: Richard A. Robison and Byron J. Adams. Also, I thank the Utah Department of Health and the New Mexico Department of Health for help in obtaining the isolates used in this study, and for the thoughtful and constructive criticisms of the anonymous reviewers, as well as Dr. Angelo Madonna of Dugway Proving Ground who provided expert guidance and assistance in completing this work.

Table of Contents

List of Tables

List of Figures

Chapter 1 The Genetic Diversity and Evolution of *Francisella tularensis* with Comments on Detection by PCR

1.1 Abstract

Francisella tularensis has been the focus of much research over the last two decades mainly because of its potential use as an agent of bioterrorism. *F. tularensis* is the causative agent of zoonotic tularemia and has a worldwide distribution. The different subspecies of *F. tularensis* vary in their biogeography and virulence, making early detection and diagnosis important in both the biodefense and public health sectors. Recent genome sequencing efforts reveal aspects of genetic diversity, evolution and phylogeography previously unknown for this relatively small organism, and highlight a role for detection by various PCR assays. This review explores the advances made in understanding the evolution and genetic diversity of *F. tularensis* and how these advances have led to better PCR assays for detection and identification of the subspecies.

1.2 Introduction

Francisella tularensis is a small, non-motile, Gram-negative coccobacillus and is the causative agent of the zoonotic disease tularemia. This facultative intracellular pathogen was first discovered in Tulare County California in 1911 where it caused a plague-like illness in local rodents (3). *F. tularensis* is able to cause disease in rabbits, squirrels, and other mammals, including humans (4). The transmission of *F. tularensis* to humans is mediated through arthropod vectors such as ticks and deer flies, by the ingestion of contaminated food or water, or by inhalation of aerosolized bacteria (5). *F. tularensis* subsp. *tularensis* is highly infectious. It is

estimated that an aerosol inoculation of as few as 10 organisms is sufficient to cause disease in humans (6). Because of its highly infectious nature, *F. tularensis* is considered a potential agent of bioterrorism and is categorized by the Centers for Disease Control and Prevention (CDC) as a Tier 1 select agent (7).

Through the years, the taxonomy of *Francisella* has gone through many changes. Upon its discovery, McCoy and Chapin named their new discovery *Bacterium tularense (3)*. Following the *Bacterium* genus, it was subsequently placed in *Pasteurella* and later *Brucella* (8). Finally in 1959, it was placed in a new genus, *Francisella*, in honor of Edward Francis, in which genus it resides today (9). There are currently 4 recognized subspecies of *Francisella tularensis*: *tularensis, holarctica and mediasiatica* and *novicida*. While the inclusion of *novicida* as a subspecies of *F. tularensis* is still contested (10, 11), much of the recent scientific literature, including Bergey's Manual of Systematic Bacteriology, recognizes this classification (12).

In 1950, the first *novicida* subspecies was isolated and characterized (13). This new isolate resembled *F.* tularensis morphologically, but differed in that it could ferment glucose, was not as virulent in humans, and did not cross-react with serum from rabbits inoculated with killed *F. tularensis.* Based on these differences, the authors proposed the name *Francisella novicida (13)*. However, in the 1950s, researchers did not have the genetic tools which became available in later decades. In the 1980s, DNA-DNA hybridization experiments between *F. tularensis* and *F. novicida* demonstrated up to 92% homology (14). Because of this high degree of genetic similarity, it was proposed that *F. novicida* be reclassified as a subspecies of *F. tularensis*. This reclassification was formally proposed in 2010 in the *International Journal of Systematic and Evolutionary Microbiology* (IJSEM) (15). This proposal received a formal objection in IJSEM, contending that genetic similarity was not enough to reclassify *F. novicida* as *F. tularensis*

subsp. *novicida*, but that the phenotypic differences were sufficient enough to justify separate species designation (16).

Finally, in a rebuttal to the objection of Johansson et al., Busse et al. (17), stood by their initial recommendation for reclassification, asserting that the genetic similarity meets the definition of a subspecies (18). Furthermore, Busse et al. acknowledge the phenotypic differences between *F. tularensis* and *F. novicida*, but contend that the 11 phenotypic differences noted are not sufficient enough for a new species (17). There are many other examples of bacteria with a greater percentage of phenotypic differences which are classified as the same species (e.g. the various biovars of *Pseudomonas fluorescens*) (17). Despite this evidence, a formal reclassification has yet to occur. Based on the high genetic similarity, and taking into account the relatively few phenotypic differences, we also propose the reclassification of *F. novicida* as a subspecies of *F. tularensis*, and will refer to it as such throughout this work.

Each subspecies is predominantly associated with a specific geographic distribution and severity of disease. The subspecies *tularensis* is typically found in North America (19) while the subspecies *holarctica* is found across much of the Northern Hemisphere (20). The subspecies *mediasiatica* has only been isolated from the central Asian republics of the former Soviet Union (21) and the subspecies *novicida* has been isolated from North America and Australia (14, 22). Phylogenetic relationships among these subspecies are inferred in Figure 1.

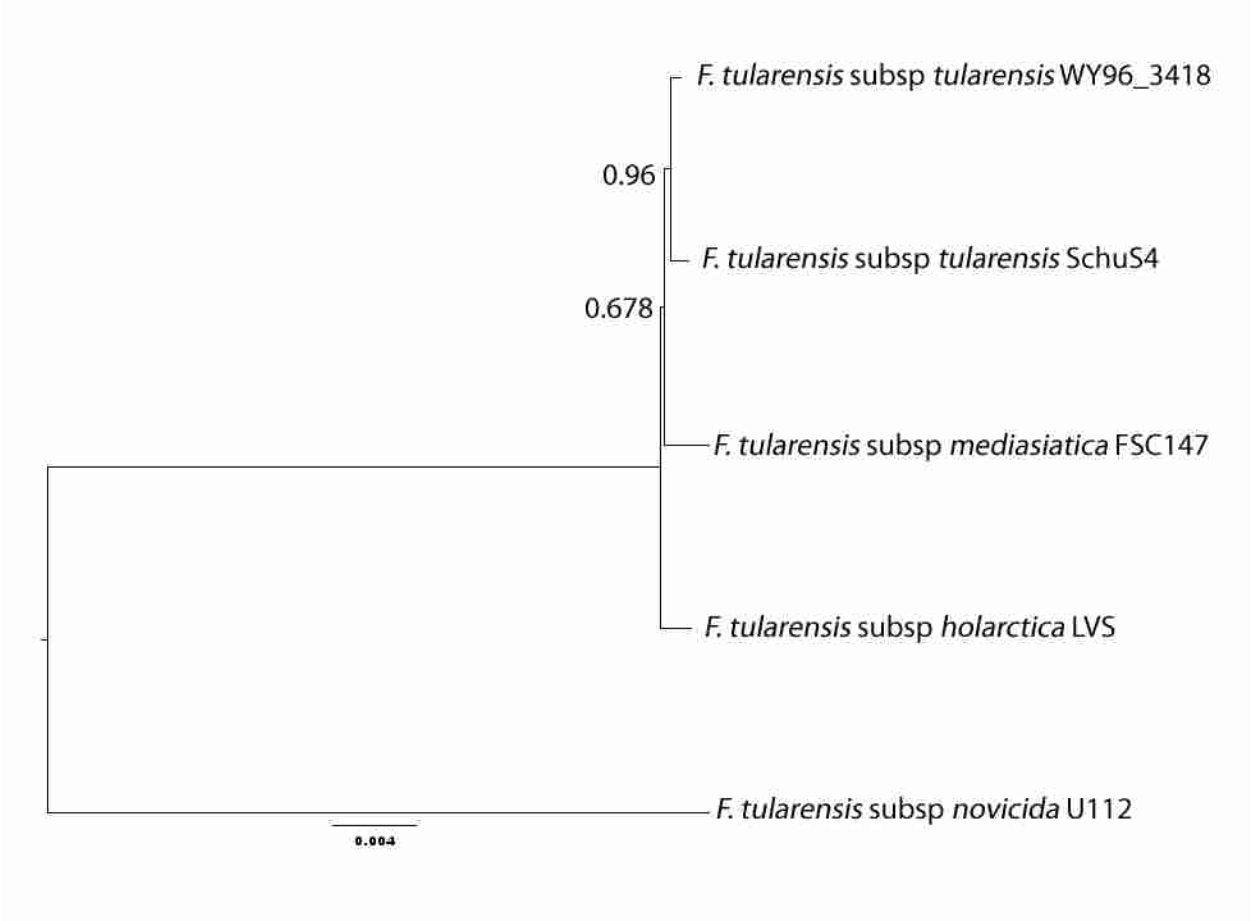


Figure 1 Maximum likelihood tree inferring the phylogenetic relationships of the *F. tularensis* subspecies. Tree was constructed by concatenating 10 housekeeping genes (*recA*, *gyrB*, *groEL*, *dnaK*, *rpoA1*, *rpoB*, *rpoD*, *rpoH*, *fopA*, and *sdhA*) followed by alignment with Clustal W and generation of the tree with MEGA 5.2. Bootstrap values are indicated at the nodes except where support was less than 0.65.

The two subspecies most associated with human disease are *tularensis* and *holarctica*. These are often abbreviated simply as Type A and Type B *tularensis*, respectively. Type A *tularensis* causes a more severe form of tularemia while the presentation of type B tularemia is somewhat milder (23, 24). The subspecies *mediasiatica* is fully virulent in mice, yet is believed to be of relatively mild virulence in humans (21, 25). Similar to the subspecies *mediasiatica*, the subspecies *novicida* is fully virulent in mice, yet rarely causes disease in humans (14).

Genetic analyses by multiple-locus variable-number tandem repeat analysis (MLVA) has identified further sub classifications and geographic structure of Type A and Type B *tularensis*. The major subdivisions of Type A *tularensis* include Type A.I and Type A.II, with the former generally isolated from the eastern United States and the latter generally isolated from the western United States (20). This biogeographic separation is correlated with the geographic distribution of specific vectors, hosts, and other abiotic factors such as elevation and rainfall (26, 27). The major divisions of Type B *tularensis* also display geographic structure, with Type B.I isolated from Eurasia, Type B.II isolated from North America and Scandinavia, Type B.III isolated from Eurasia and North America, Type B.IV isolated from North America and Sweden, and Type B.V isolated from Japan (20). Unlike type A *tularensis*, the distribution of Type B *tularensis* has not been shown to correlate with the distribution of any specific vectors (26).

The *F. tularensis* subsp. *holarctica* isolated from Japan was first differentiated from other *F. tularensis* subspecies based on its ability to ferment glucose (28). These isolates were further differentiated by demonstrating a reduced virulence from the subspecies *tularensis,* displaying a virulence similar to that of the subspecies *holarctica* (29). As genomic tools became more widely available, this division was confirmed by microarray analysis (21), restriction fragment length polymorphism (RFLP) analysis (30), and multiple-locus variable number tandem repeat analysis (MLVA) (20, 31). Genetic analyses have hinted that these isolates from Japan underwent a unique evolutionary process in a restricted area, separate from other *F. tularensis* subspecies (31). Because of the phenotypic differences, the genetic differences, and the apparent isolated evolution, it has been proposed that these strains from Japan be classified as another subspecies of *F. tularensis* called *F. tularensis* subsp. *japonica* (32). However, since relatively

few isolates from Japan have been analyzed, we recommend that this designation not be adopted at this time.

1.3 Genetic Diversity

The first complete genome of *Francisella tularensis* was sequenced in 2005 (33). This first sequence was the classical type strain of *Francisella tularensis* subsp. *tularensis* representing the Type A.I sub classification. Since then, numerous other whole and partial genomes of *F. tularensis* have been sequenced: *F. tularensis* subsp. *holarctica* strain OSU18 (Type B) (34), a European isolate of Type A *tularensis* (35), *F. tularensis* subsp. *novicida* strain U112 (36), a Type A.II *tularensis* (WY96-3418) (37), *F. tularensis* subsp. *mediasiatica* (10) and at least 10 more comprising the 4 subspecies of *F. tularensis* (38-41). With the advent of improved massively parallel sequencing technologies, more genomes continue to be sequenced at an ever-increasing rate (42). In all, there are currently 16 complete genomes of *Francisella tularensis* deposited in GenBank and even more partial genomes. This collection of genomic information allows for the comparative analysis of these genomes and provides insight into the evolution of *F. tularensis* genome architecture.

Even before the first *Francisella* genome was completed in 2005, studies analyzing the genomic diversity of *F. tularensis* were plentiful. Because of its potential use as a bioweapon and for public health reasons, rapid identification of *F. tularensis* became paramount (7). Early DNA based techniques focused on 16S rDNA typing. This proved difficult since among the 4 subspecies, the 16S rDNA genes exhibit between 98.5 – 99.9% similarity, the result of only 6 nucleotide differences among the most divergent strains (43). Other DNA based techniques for identification such as PCR, which is both rapid and accurate, helped spur further interest in the

genetic diversity of the *F. tularensis* subspecies (21, 44). A genome wide microarray that analyzed 27 strains of all four subspecies confirmed the limited genetic variation within the subspecies, but identified 8 variable regions that were used to develop a subspecies-specific PCR assay (21). Another microarray study analyzing the genetic diversity of 11 Type A isolates and 6 Type B isolates from various localities around the United States identified 13 regions of difference, including segments of several genes with implications for virulence (45). While microarray and other studies revealed valuable information about the regional distribution and differences in virulence, complete genome sequences reveal a more complete picture (20, 21, 45).

The first completed genome sequence of *F. tularensis* yielded insights to previously undiscovered features of its genetic makeup. Some of the genetic features discovered included previously uncharacterized virulence genes encoding type IV pili and iron acquisition systems (33). The complete sequence also revealed a duplication of an approximately 30 kb region previously identified as a pathogenicity island containing 17 open reading frames (ORFs), perhaps shedding light on the enhanced virulence of Type A *tularensis* (33, 46, 47). Finally, analysis of this genome indicated the loss of several biosynthetic pathways, which helps explain the fastidious nutritional requirements of *F. tularensis* and suggests the need to infect a host during its life cycle (33).

The first comparative genomic study of *F. tularensis* was of the Type A (Schu S4) and Type B (OSU18) strains. This study revealed an extensive genomic similarity of 97.63%, indicating that the differences in virulence between the two strains are likely not due to large differences in gene content (34). This degree of sequence identity was confirmed among the remaining subspecies as well (10, 25, 36). Perhaps the most striking difference between these

two strains is the vast amount of genomic rearrangement. These rearrangements can mostly be attributed to homologous recombination using insertion (IS) elements (34).

After the genome sequence of *F. tularensis* subsp. *novicida* was complete, a 3-way comparison between three of the subspecies (*tularensis*, *holarctica*, and *novicida*) was possible. Again, a high degree of sequence identity among the subspecies was confirmed, as was the large amount of genomic rearrangement (36). Even though the length and the gene content of the *novicida* subspecies (1.91 Mb and 1,731 protein coding genes) are both greater than that of the *tularensis* subspecies (1.89 Mb and 1,445 protein coding genes) and the *holarctica* subspecies (1.89 Mb and 1,380 protein coding genes), these human pathogenic strains contain 41 genes which the non-human pathogenic strains (*novicida*) do not (36). Initial comparisons of these genomes revealed that the human pathogenic strains carry 2 copies of the *Francisella* Pathogenicity Island (FPI) while the non human pathogenic strains carry only 1 copy, shedding further light on the differences in virulence among the subspecies (47).

Many studies have been completed comparing the various subsets of available *F. tularensis* genomes. A comparison of the genomes of two *holarcitca* subspecies, the live vaccine strain (LVS) and strain FSC200, sought to uncover the mode of attenuation for LVS (48), which was attenuated through the repeated passage of a *holarctica* strain between the 1930s and 1950s in the former Soviet Union (34, 49). The genomes of the LVS and FSC200 strains differ by only 0.08% but the LVS strain was able to confer immunity to infection with *F. tularensis* subsp. *tularensis* in BALB/c mice (48, 49). While the exact nature of genomic modifications leading to LVS attenuation were not found, comparison with other more virulent Type A strains revealed some candidate genes which could be targeted in the development of a future vaccine (48). When the sequence of *F. tularensis* subsp. *holarctica* FTNF002-00 was completed and compared

to both LVS and the OSU18 strains, it was found to have greater than 99.9% sequence similarity (38). Other studies have shown a stable genome architecture among Type B strains, but FTNF002-00 carries a 3.9 kb inversion compared to other Type B strains (34, 38, 50).

Other whole genome comparisons focused on comparing different strains of Type A *tularensis*. A comparison between *F. tularensis* subsp. *tularensis* Schu S4 (Type A.I) and WY96-3481 (Type A.II) revealed only one whole gene difference, a hypothetical protein with an unknown function (37). Despite the fact that these two strains are very closely related, there were still many other differences, including numerous single nucleotide polymorphisms (SNPs), small indels, differences in IS elements, and even 31 large chromosomal rearrangements (37). Many of the chromosomal rearrangements are frequently bordered by IS elements, providing a mechanism for the translocations (10, 37, 39). Another genome comparison of a Type A.I clinical isolate to the Schu S4 genome showed that except for some minor changes, the genomes were virtually identical, suggesting a high degree of sequence conservation within the Type A.I subgroup (39). The genome of another Type A.I strain (TI0902) isolated from a cat in Virginia, United States, is also highly similar to Schu S4 as it only differs by 103 SNPs (40). Other researchers compared a European isolate of Type A.I *tularensis* (which is typically restricted to North America) to Schu S4 and found that the two were virtually identical, with only 8 SNP and 3 variable number tandem repeat (VNTR) differences (35). The fact that these two strains are so alike suggests that the European isolates are descended from the Schu S4 strain and did not evolve independently in Europe (35).

The completion of a fourth subspecies genome of *F. tularensis*, the *mediasiatica* subspecies, enabled full genome comparisons of the four subspecies of *F. tularensis*. It was demonstrated that the subspecies *mediasiatica* and *tularensis* are highly similar, which raises

more questions about their differences in virulence (10, 21, 51). Phylogenetic analysis of the complete genomes of the subspecies *mediasiatica* also demonstrated that it is a monophyletic taxon of *F. tularensis*, contradicting previous evidence suggesting that the subspecies *mediasiatica* was not a member of the *F. tularensis* clade (52). However, since isolates of the *mediasiatica* subspecies are rare, it is difficult to know the true genetic diversity within the subspecies. Figure 2 shows the overall genome architecture of representative strains of *F. tularensis*, highlighting the large-scale genomic rearrangements between the subspecies.

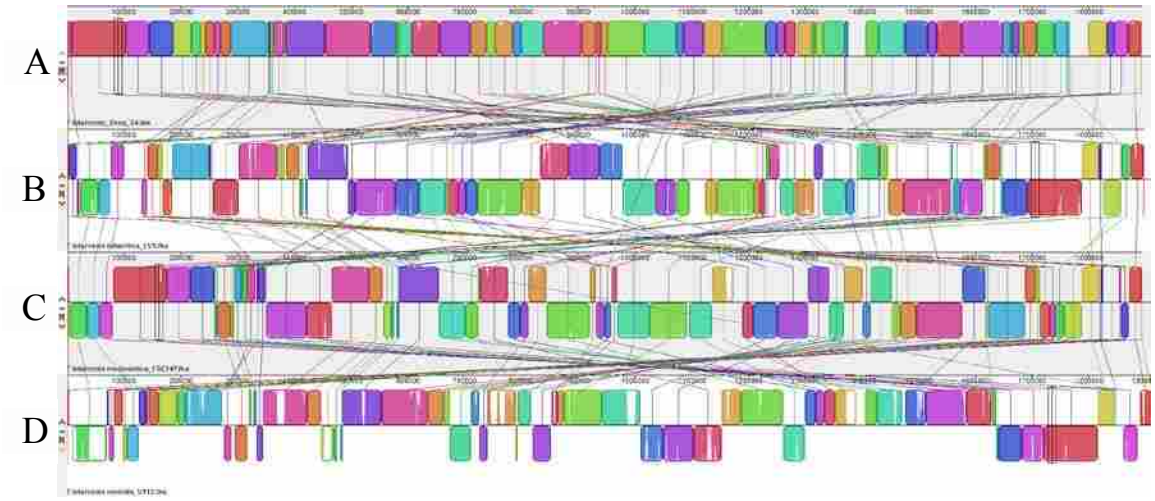


Figure 2 Whole genome alignment of *F. tularensis* subspecies
Whole genome alignment of representative strains from each of the four subspecies of *Francisella tularensis* using Mauve (53) highlighting differences in the macro genome architecture relative to the reference strain (A). Colored blocks represent homologous sections of each genome. A) *F. tularensis* subsp. *tularensis* Schu S4. B) *F. tularensis* subsp. *holarctica* LVS. C) *F. tularensis* subsp. *mediasiatica* FSC147. D) *F. tularensis* subsp. *novicida* U112.

The evolution of the *Francisellacaea* is complicated by the discovery of *Francisella*-like endosymbionts (FLEs) of ticks, which have an unknown pathogenicity in humans (54-57). While these endosymbionts lack sufficient evidence to be classified as *F. tularensis*, they are similar enough to cross react with many molecular-based methods of detection (58). Because of the potential to misidentify FLEs as *F. tularensis*, which could impact the diagnosis of tularemia

in public heath settings, many have cautioned about the use of PCR assays for the detection of *F. tularensis* (59, 60). Despite this caution, PCR remains the standard of practice for the detection and identification of *F. tularensis* subspecies (44).

1.3.1 Detection

The ability to accurately detect and diagnose *F. tularensis* infection carries significant implications in public health and bioterror (2, 7). Because of the different pathogenic profiles and biogeography of the various subspecies of *F. tularensis*, it is important to be able to accurately discriminate among them (2). Polymerase chain reaction (PCR) has become the method of choice for the identification of various pathogens because it is rapid, sensitive and highly specific (61-63). Detection and differentiation of the subspecies of *F. tularensis* by PCR is complicated by the lack of significant variability in their genomes (34, 43). Various methods for the detection of *F. tularensis* have been reviewed in the last decade, however much more work has since been completed on the detection of *F. tularensis* using PCR (44, 64).

1.3.1.1 Conventional PCR

Since 2008, research on the use of conventional PCR for the detection of *F. tularensis* has dropped off considerably, with only a handful of publications on the subject. In alignment with an earlier review (64), the gene *tul4* was a popular choice to detect all subspecies of *F. tularensis* (65, 66). Since *F. tularensis* is a potential agent of bioterrorism, some assays included the multiplex detection of other biothreat agents. One such study developed two multiplex assays to detect "Tier 1" select agents; one assay for DNA based organisms (Variola Major, *Bacillus anthracis*, *Yersinia pestis*, *Francisella tularensis*, and Varicella zoster virus) and another

assay with a reverse transcriptase for RNA based viruses (Ebola virus, Lassa fever virus, Rift Valley fever, Hantavirus Sin Nombre and the four serotypes of Dengue virus) (65). A major drawback to these multiplex assays however, is the use of a reporter dye and a colormetric detection system, because a positive result is unable to distinguish between the agents. The assay is intended only as a broad screening tool and further testing is required to differentiate between the organisms comprising the assay. Furthermore, since the genome of Variola Major (the causative agent of Smallpox) is so highly regulated, testing was completed with a plasmid control containing a small segment of the Variola Major genome (65).

Real-time PCR is known for being efficient and sensitive, but is not ideal for multiplexing beyond a 4- or 6-plex reaction because of the limited number of fluorescent channels available on most instrument platforms (67, 68). Researchers have overcome this limitation by using modified primers to bind the PCR products of a 15-plex reaction to fluorescent beads that can then be analyzed by a flow cytometer for the simultaneous detection of 11 pathogens with similar sensitivities to real-time reactions (69). While effective, flow cytometers can be large, difficult to use, and costly. The Luminex Corporation (Austin, TX) has developed a similar, yet easier to use technology in their MAGPIX® system. Rather than a flow cell, the MAGPIX® uses a magnet to capture fluorescently labeled magnetic beads and a CCD camera to capture images of up to 50 different analytes (70, 71). Because of its relatively low cost and ease of use, the MAGPIX® may be more ideally suited for integration in clinical labs for the simultaneous detection of multiple pathogens (70).

While it may be useful to detect broad categories of pathogens, because of the virulence status of various subspecies of *F. tularensis*, it is also important to be able to differentiate among them as well. Using the *tul4* gene and variations in the *pilA* gene, researchers were able to

differentiate the four subspecies of *F. tularensis* (66). Another study used suppression subtractive hybridization (SSH) to identify regions of difference between the genomes of Type A.I and Type A.II *tularensis*. This information was used to create a conventional PCR assay to differentiate between Type A.I, Type A.II, Type B, and *F. tularensis* subsp. *novicida* isolates (72). Later, this same assay was adapted to a real-time PCR platform (73).

1.3.1.2 Real-time PCR

Real-time PCR is a popular choice for the detection of *F. tularensis* because it is sensitive, reliable, cost-effective, and eliminates the need for time consuming gels, though this time commitment has been significantly reduced with the introduction of rapid dry gels (74). A popular method of real-time PCR incorporates the use of SYBR Green which will fluoresce upon binding double stranded DNA. Thus, the fluorescent signal will increase as PCR progresses and more amplicons are synthesized. SYBR green is a popular alternative to other real-time technologies because of its relatively low cost (75). However, it is not ideal for multiplex reactions since the dye will bind to all double stranded DNA in the reaction and produce a fluorescent signal. Sellek *et al*. (75) developed an assay to detect *F. tularensis* from soil using the *tul4* gene, previously used in conventional PCR assays (65, 66). However, the assay was only validated with *F. tularensis* subsp. *holarctica* and subsp. *novicida*. Lacking were representatives from the subsp. *tularensis* and *mediasiatica*. Furthermore, positive fluorescent signals were obtained from other non-related bacteria. These were later ruled out as true positives after analyzing the PCR products on a gel and finding only primer dimers (75).

Genome comparisons aided the development of SYBR green assays (76-78). Woubit *et al*. (78) compared several genomes from the *Escherichia, Francisella, Salmonella, Shigella,*

Vibrio, and Yersinia genera to develop a series of 27 assays to detect and differentiate these common food and biothreat pathogens. With respect to *Francisella*, the assays were so specific that assays intended to detect all subspecies of *Francisella* were only able to detect the *tularensis* and *novicida* subspecies (78).

The propensity of PCR assays to cross-react with environmental, non-pathogenic *Francisella* or other closely related organisms (59) requires the development of more specific assays to avoid false positives or incorrect diagnoses. To solve this problem, results from resequencing microarrays were compared to identify SNPs along the phylogeny of *F. tularensis* and build real-time PCR assays capable of differentiating Type A.I, A.II, A.Ia, A.Ib, Type B.I, and B.II *tularensis* (76). Similarly, another group analyzed publically available whole genome sequences to identify defining SNPs and small insertion/deletion elements (INDELs) to design a series of 35 assays capable of distinguishing the four subspecies of *F. tularensis* and the major subtypes of Type A and Type B *tularensis,* including Type A.I, A.II, and B.I, B.II, B.III, B.IV, and B.V (77). Both assays were able to accurately assign isolates to the correct subspecies and clade while avoiding any cross-reactivity to near neighbors (although the former includes only one *novicida* strain in the analysis).

Another method for the real-time detection of *F. tularensis* is the 5' nuclease or TaqMan® assay. These assays incorporate fluorescently labeled DNA probes specific to the template DNA resulting in even more specific identification than the SYBR Green assays, eliminating the need to perform a melt curve analysis. Strategies for single-plex real-time assays for the detection of *F. tularensis* with TaqMan® assays are varied. Gene targets include a gene for an outer membrane protein, FopA, a single-copy gene for detection and quantification of all subspecies of *F. tularensis* (79), the 16S rRNA gene to detect all subspecies of *F. tularensis* (80,

81), the insertion element IS*Ftu2,* which is unique to *Francisella* species (82), intergenic regions of differentiation to distinguish Type A.I from Type A.II *tularensis* (73), and SNP-based assays to differentiate the species and subspecies of *Francisella* isolates (83). Some assays can be used in concert with others to detect a wide variety of agents. These include biothreat agents (80) or other organisms with similar disease presentations (81), while others were used solely for the differentiation of subspecies and subpopulations of *F. tularensis* (73, 83). The advantage of using a single-copy gene for detection is the ability to quantify the amount of the agent, which can be useful in clinical and diagnostic settings (79). Conversely, multicopy-genes such as the 16S rRNA gene and the IS*Ftu2* gene should achieve lower detection limits, which is ideal given the low infectious dose of *F. tularensis* (6, 80, 82). A significant drawback of using the 16S rRNA gene for detection is that since it is so conserved, there is some cross reactivity with near neighbors and other *Francisella*-like species, requiring further confirmatory analyses (43, 80).

Multiplex real-time TaqMan® assays incorporate the added convenience of running multiple reactions in a single tube using probes labeled with various fluorophores. However, as mentioned previously, multiplexing with TaqMan® assays is generally limited to a 4- or 6 plex reaction because of the limited number of fluorescent channels on the instruments (67, 68). One multiplex assay is a 2-plex assay designed from genome comparisons to detect the four subspecies of *F. tularensis* but does not differentiate among them. Another multiplex assay is capable of differentiating the four *F. tularensis* subspecies with only a 3-plex assay. This assay was developed using both unique and shared genome regions among the subspecies with the addition of a scoring matrix (2).

Since *F. tularensis* has the potential to be used as a bioweapon, a commercial market has arisen for field-ready detection of biothreat agents, including *Bacillus anthracis, Francisella*

tularensis, Yersinia pestsis, Brucella species, and others. A comparison of one such commercial instrument, the RAZOR®, (BioFire Defense; previously Idaho Technologies, Salt Lake City, UT) and another instrument designed for laboratory use, the Applied Biosystems 7300/7500 system (Thermo Fisher Scientific, Grand Island, NY) used assays developed for *B. anthracis, Brucella* species, *F. tularensis,* and *Y. pestis*, comparing sensitivities and specificities of the two platforms. Results showed that for all agents, the sensitivities were between 10-100 fg of target DNA per reaction, and no cross reactivity was observed with other closely related bacteria (84). Run time on the RAZOR® was notably shorter than that of the 7300/7500 instrument.

Another diagnostic tool, the FilmArray® system (BioFire Defense, Salt Lake City, UT), uses a lab-in-a-pouch approach to process raw samples and detect 17 biothreat pathogens with an array of single-plex real-time PCR assays in about an hour (85). An evaluation of the Biothreat Panel using DNA samples from *B. anthracis, F. tularensis,* and *Y. pestis* indicated sensitivities of 250 genome equivalents or lower and the authors conclude that the system is both sensitive and selective (85). However, since the FilmArray® system is designed to be a complete sample to answer system, sensitivities may vary when tested with whole organisms in different matrices like blood or serum rather than purified DNA.

Another evaluation compared the FilmArray® system with TaqMan® Array Cards developed for the detection of biothreat agents (86, 87). Here, researchers tested for *B. anthracis, F. tularensis,* and *Y. pestis* in the blood of murine infection models. Results showed that blood culture was the most sensitive means of detection followed by the FilmArray and Array Cards for *B. anthracis,* and *F. tularensis.* All three methods demonstrated similar detection levels for *Y. pestis* (87). While blood culture was the most sensitive means of detection for two of the three agents tested, it requires much more time for detection compared to the PCR

assays. Each of these methods for detection carries drawbacks and benefits and must be weighed appropriately to ensure the best possible outcome.

1.3.1.3 Other PCR assays

Recently, other PCR-based assays have been developed for the detection of *F. tularensis* and other bacteria. One such assay involves analyzing PCR products with electrospray ionization-mass spectrometry (ESI-MS). In this technique, the actual base composition of the PCR products are identified and compared to a library of sequences for identification rather than relying on the fluorescent signal obtained from real-time PCR (88). This PCR/ESI-MS technique has been applied to the wide-spread identification of biothreat agents, respiratory pathogens, and other pathogenic bacteria and viruses (88, 89). Others have used this technology specifically for identifying *F. tularensis* from natural sources (90) and even for typing the subspecies of *F. tularensis* (91).

Recombinase Polymerase Amplification (RPA) is a PCR-like assay in which amplification is carried out at one temperature (isothermal) instead of cycling temperatures as in PCR. Recently, RPA assays have been applied to the detection of *F. tularensis* and other biothreat agents (92-94). Two of these assays showed comparable sensitivities to real-time PCR assays with an instrument run time of about 10 minutes (92, 93). A third assay using electrochemical detection rather than fluorescent probes seemed less sensitive than other assays, with detection levels on the order of 10^4 copies/μL (94).

Finally, as the cost of sequencing continues to fall, more sequencing-based detection assays are being used to detect biological agents such as *F. tularensis*. One such assay used a pyrosequencing method to sequence the variable region of 16S rDNA to identify and group *F.*

tularensis isolates by subspecies (95). The results from analyzing the SNPs in 16S rDNA are more distinctive than SNP analysis from real-time PCR. Another sequencing assay was multiplexed for the detection and strain typing of *B. anthracis, F. tularensis,* and *Y. pestis* by interrogating 10 loci per pathogen (96). While sequencing assays provide some promise for the rapid detection and classification of *F. tularensis*, there is a noticeable lack of information on the sensitivity or detection limits of these assays. In the world of clinical diagnostics and biodefense, the ability to detect low quantities of *F. tularensis* and other agents is paramount.

1.4 Evolution

Numerous studies have been conducted on the evolution of the subspecies of *F. tularensis* to define specific clades and to reveal their evolutionary history. Before next generation whole genome sequencing was widely available, various techniques were used to recover the phylogenetic relationships among strains of *F. tularensis,* such as microarrays (21, 45), MLVA (20), and sequencing specific genes or other genetic loci (52, 97). One of the earliest of these studies produced a phylogenetic tree in which the subspecies *tularensis* and *mediasiatica* shared a major clade along with the Japanese isolates of the *holarctica* subspecies (21). A later analysis provided better resolution, differentiating the *tularensis* and *mediasiatica* subspecies, and grouping the Japanese isolates of the *holarctica* subspecies with the other *holarctia* subspecies (20). These authors also determined that *F. tularensis* subsp. *holarctia* appears to have recently spread globally from a single geographic origin, while *F. tularensis* subsp *tularensis* appears to have experienced most of its evolutionary history in North America, and may even have originated in the central United States (98). However, *F. tularensis* subsp. *tularensis* is now clearly distributed beyond North America into parts of Europe (35).

The finding that the subspecies *holarctica* recently spread from a single origin seems likely because of the small amount of genetic diversity within the subspecies, that has been identified by a variety of molecular methods (26, 48, 99-101). However, the precise area of origin of the subspecies *holarctica* is unknown. Based on phylogenetic analyses, there are two competing hypothesis as to its origin: 1) the subspecies *holarctica* originated in Asia or 2) the subspecies *holarctica* originated in North America before spreading around the Northern Hemisphere (102). There appears to be more evidence for the origination of the subspecies *holarctica* in North America, though this may be due to the lack of Asian isolates for analysis. Regardless, it appears that the *holarctica* subspecies is a highly fit clone that originated from a single source and spread throughout the Northern Hemisphere (100, 102). However, if *F. tularensis* subsp. *tularensis* originated in North America (20, 98) and the subspecies *holarctica* is descended from the *tularensis* subspecies (97), then it seems likely that the subspecies *holarctica* may have originated in North America as well. This hypothesis is supported by the fact that sequences of various housekeeping genes and some outer membrane proteins from the subspecies *tularensis* and *holarctica* align well, while those from the subspecies *novicida* and *mediasiatica* do not (52).

It is generally accepted that *F. tularensis* subsp. *novicida* is the oldest of the *F. tularensis* subspecies and evidence suggests that *F. tularensis* subsp. *novicida* and *Francisella philomiragia* share a common, aquatic ancestor (97, 103, 104). These two species are generally considered non-pathogenic to humans. However, their association with aquatic sources is further substantiated in that documented human infections by these two species have occurred in near-drowning victims (14, 105). Furthermore, *F. philomiragia* contains one copy of the FPI, similar

to *F. tularensis* subsp. *novicida* while the remaining subspecies of *F. tularensis* contain 2 copies (47, 104).

Molecular evidence suggests that the four subspecies of *F. tularensis* have evolved by vertical descent (97). A common method of acquiring genetic variation in bacteria is through horizontal gene transfer. This is well documented in many species of bacteria, and especially in the conference of antibiotic resistance (106-109). However, in the subspecies of *F. tularensis*, genetic variation, including antibiotic resistance seems to have arisen by mutation rather than the acquisition of new genes through horizontal gene transfer (110-112).

An *in silico* analysis has recently shown that the non human-pathogenic *F. tularensis* subsp. *novicida* possesses a CRISPER/Cas system to defend against invading genetic elements. This finding further supports the hypothesis that mutation is responsible for much of the evolution of *F. tularensis* (113, 114). Analyses of the other three virulent subspecies of *F. tularensis* (*tularensis*, *holarctica*, and *mediasiatica*), reveal that the genes responsible for the CRISPER/Cas system are non-functional (114). This is somewhat puzzling since deletion of the CRISPER/Cas system in other pathogens such as *Neisseria meningitidis*, *Camphylobacter jejuni*, *Legionella pneumophila*, and *Pseudomonas aeruginosa* result in decreased virulence. It is hypothesized that in the case of *F. tularensis*, other mutations in the genome have compensated for the degeneration of the CRISPER/Cas system in the virulent subspecies of *F. tularensis* (115).

1.5 Concluding Remarks

The genetic diversity of the subspecies of *F. tularensis* appears to be quite limited. Genome comparisons among the subspecies reveal similarities greater than 95% (10, 25). Many

of the differences in the genomes of *F. tularensis* are large-scale genomic rearrangements and a duplication of the pathogenicity island in the *tularensis*, *holarctica*, and *mediasiatica* subspecies (34, 47). However, because the *mediasiatica* subspecies is so rare, assessments of its true genetic diversity must be considered preliminary.

There are many pros and cons to the various PCR detection methods and the individual user's needs should dictate which method to use. Conventional PCR is easy and inexpensive but is known for being time consuming because of the need to run gels. However, since the introduction of rapid dry gels, the time commitment usually associated with gels has been shortened considerably. Utilizing fast PCR technology in combination with rapid dry gels, it is possible to get a result in approximately 50 minutes (74). In general, conventional PCR has fallen out of favor with many researchers. However, this approach allows for large multiplex reactions for the detection of many organisms at once, especially when coupled with another detection system such as the MAGPIX® (70, 71).

Real-time PCR is one of the most popular methods for detection because it is simple, cost effective, and sensitive. SYBR Green assays are inexpensive and accurate and can even be multiplexed with the incorporation of a melting curve analysis. TaqMan® assays are more expensive than SYBR Green assays, but carry an additional layer of specificity with the sequence of the probe. Multiplexing with TaqMan® assays is possible, but usually only up to a 4- or 6-plex because of the limited number of available fluorescent channels on most instruments (67, 68). The limited amount of multiplexing with TaqMan® assays can be overcome by setting up an array of single-plex reactions similar to the FilmArray® system (85).

Many current PCR assays lack the specificity to differentiate between environmental, non-pathogenic *Francisella* and other closely related organisms such as FLEs (58, 59). Perhaps

in these situations, it would be wise to use whole genome sequencing assays for the detection of *Francisella* subspecies (95, 96)

As whole genome sequencing has become more widely available, genome comparisons between the subspecies of *F. tularensis* are possible and shed further light on the genetic diversity and evolution of this pathogen. It is apparent that the more virulent subspecies of *F. tularensis* have evolved from *F. tularensis* subsp. *novicida* primarily by genomic decay, genomic rearrangements, and the duplication of the FPI (36). Many of the interrupted genes (pseudogenes) in the virulent subspecies of *F. tularensis* are metabolic genes, further supporting an intracellular life cycle, while other interrupted genes include secreted effector proteins that may have led to excessive virulence, furthering the patho-adaption of *F. tularensis* as an intracellular pathogen (10, 106, 116, 117).

Chapter 2 A multiplex real-time PCR assay for the detection and differentiation of *Francisella tularensis* subspecies

2.1 Summary

Francisella tularensis is the etiological agent of tularaemia, a zoonotic disease with world-wide prevalence. *F. tularensis* is a highly pathogenic organism and has been designated a category A biothreat agent by the Centers for Disease Control and Prevention (CDC). Tularaemia is endemic in much of the United States, Europe, and parts of Asia. It is transmitted by numerous vectors and vehicles such as deer flies, ticks, and rabbits. Currently, there are four recognized subspecies of *F. tularensis*: *tularensis* (Type A), *holarctica* (Type B), *mediasiatica,* and *novicida*. Within the Type A classification there are two subclassifications, Type A.I and A.II, each with a specific geographic distribution across the United States. Type B *tularensis* is found in both the United States and Europe. Because of virulence differences among subtypes, it is important that health departments, hospitals, and other government agencies be able to quickly identify each subtype. The purpose of this study was to develop a multiplex real-time PCR assay for the identification and discrimination of Type A.I, Type A.II, Type B, and *novicida* subspecies of *F. tularensis*. The assay was validated using 119 isolates of *F. tularensis*, 3 of its nearest neighbours, and 14 other bacterial pathogens. This assay proved to be ~98 % successful at identifying the known subspecies of *F. tularensis*, and could prove to be a useful tool in the characterization of this important pathogen.

2.2 Introduction

Francisella tularensis is a Gram-negative facultative intracellular bacterial pathogen and is the causative agent of the zoonotic disease, tularaemia. As few as 10 organisms can cause disease via the aerosol route (6). Due to its high infectivity, ease of dissemination and ability to cause illness and death, *F. tularensis* has long been considered a potential bioweapon by Japan, the former Soviet Union, and the United States (118). With the advent of the CDC Select Agent Program (a series of rules and regulations governing the possession and transfer of organisms that could be used as bioweapons), *F. tularensis* has been classified as a category A potential agent of bioterrorism. It was estimated by the World Health Organization (WHO) that 50 kg of *F. tularensis* dispersed as an aerosol over a highly populated area of 5 million people would result in 250 000 cases of tularaemia with 19 000 deaths (118).

Francisella tularensis is a member of the γ-subclass of the proteobacteria currently consisting of three accepted subspecies: *tularensis* (Type A), *holarctica* (Type B), and *mediasiatica.* The subspecies differ in their geographic distribution as well as virulence (20, 100). Much of the scientific literature, including this work, refers to *Francisella novicida* as a fourth subspecies of *F. tularensis* (97, 100, 102). Types A and B are most associated with human disease with Type A being the more virulent. The *mediasiatica* subspecies is more commonly found in the Central Asian republics of the former Soviet Union and little is known about its ability to cause disease in humans (21). The *novicida* subspecies is more associated with water and rarely causes human disease (119). The natural reservoir of *F. tularensis* remains largely unknown; though there is growing evidence that amoeba may play an important role in harboring the bacterium (120-122).

The genome of *F. tularensis* is highly conserved among the four subspecies. The 16S rRNA genes exhibit 98.5 % to 99.9 % similarity (43). Even with this high degree of sequence similarity, each of the subspecies demonstrates notable differences in virulence. Within the Type A *tularensis*, multi-locus variable-number tandem repeat analysis (MLVA) revealed a subdivision: Type A.I and Type A.II (20). An apparent geographical separation exists between these two subtypes. Type A.I isolates are primarily found in the Central and Eastern portions of the United States, while Type A.II isolates are generally found in the Western portion (19). Molins-Schneekloth *et al.*(72), using suppression subtractive hybridization (SSH), have successfully identified genetic markers used for the differentiation of Type A.I and A.II *tularensis* isolates.

Many molecular methods have been used for the identification of *F. tularensis* such as Pulsed Field Gel Electrophoresis (PFGE), amplified fragment length polymorphism fingerprinting (AFLP), 16S rRNA gene sequencing (123), RFLP (30), MLVA (20, 102), and PCR (59, 124-126). Many of these techniques can be labour intensive and cumbersome to perform, especially on a large number of samples. The previous PCR assays developed lack the convenience of real-time detection and are not performed in multiplex. Since tularaemia is endemic in many areas of the United States, and the potential exists for *F. tularensis* to be used as a bioweapon, rapid techniques are necessary to aid in the accurate identification and differentiation of *F. tularensis* subtypes. The goal of this study was to develop a multiplex real-time PCR assay for the rapid identification of *F. tularensis* isolates relevant to the subspecies commonly found in the United States and Europe.

2.3 Methods

2.3.1 Bacterial strains and culture conditions

The isolates used in this study are a part of a select agent archive housed at Brigham Young University and maintained by Dr. Richard Robison. The collection largely consists of isolates obtained from the State Health Departments of Utah and New Mexico over the past two decades. All *F. tularensis* isolates were grown on modified Mueller Hinton agar (MMHA)(Becton Dickinson and Company) for 3-4 days with 5 % CO_2 at 35 ºC. MMHA was prepared by autoclaving the Mueller Hinton base, which was chocolatized by adding 5 % sheep blood while the medium was approximately 80 °C. After the medium cooled to 50 °C, 10 mL of 10 % glucose and 20 mL of IsoVitaleX were added to 1 L. For near neighbours, genomic DNA was obtained from the Critical Reagents Program (CRP) (www.jpeocbd.osd.mil/packs/Default.aspx?pg=1205).

2.3.2 Preparation of DNA

Total genomic DNA was extracted from each isolate using the MagNA Pure System (Roche) and the MagNA Pure LC DNA Isolation Kit III (Roche) according to the manufactures directions. Briefly, cells grown on MMHA agar were suspended in 250 µL of Tris/EDTA buffer [10 mM Tris/HCl (pH 8.0), 1 mM EDTA] (TE buffer) containing 1.8 µg lysozyme μL^{-1} and incubated for 1 h at 37 °C. To this tube, 270 µL of bacterial lysis buffer and 100 µL of proteinase K were added and the tube was incubated for 10 min at 65 °C. Samples were then incubated in boiling water for 10 min to inactivate pathogens. DNA was eluted in a total volume of 100 µL. DNA concentration was measured using a PicoGreen assay (Invitrogen) and TBS-

380 fluorometer (Turner Biosystems). For optimization purposes, DNA stock solutions were diluted to a concentration of approximately 50 ng μL^{-1}.

2.3.3 Primer and probe design

Whole genome sequences of *F. tularensis* subspecies *holarctica* strains OSU18 (accession number CP000437), LVS (AM233362), and FTNF002-00 (CP000803), subspecies *novicida* strain U112 (CP000439), subspecies *tularensis* strains WY96-3418 (CP000608), FSC198 (AM286280), and Schu S4 (AJ749949), and subspecies *mediasiatica* strain FSC147 (CP000915) were obtained from GenBank (www.ncbi.nlm.nih.gov/genbank/) (Table 1). These genomes were aligned to each other using the genome alignment tool Mauve (53). With the *holarctica* genomes set as the reference sequences, the genomes were analyzed for regions of non-homology (Figure. S1). The process was repeated with each of the other genomes set as the reference and analyzed. Only in the *holarctia* (nucleotides 800268-800721 of the FTNF002-00 strain) and *novicida* (nucleotides 1579889-1580210 of the U112 strain) genomes were unique regions identified using this method.

Table 1 *F. tularensis* genome sequences analyzed

Subspecies	**Other designation**	**Type**	**Accession #**	**Reference**
holarctica	FTNF002-00	B	CP000803	Barabote *et al*., 2009
tularensis	WY96-3418	A.II	CP000608	Beckstrom-Sternberg *et al.*, 2007
tularensis	FSC198	A	AM286280	Chaudhuri *et al.*, 2007
novicida	U112		CP000439	Rohmer *et al.*, 2007
holarctica	OSU18	B	CP000437	Petrosino *et al*., 2006
holarctica	LVS	B	AM233362	Unpublished
tularensis	Schu S4	A.I	AF749949	Larsson *et al.*, 2005
mediasiatica	FSC147		CP000915	Larsson *et al*., 2009

For the A.I and A.II subtypes, the RD8 (A.I) and RD5 (A.II) regions described in Molins-Schneekloth *et al.* (72) were selected for Taqman probe design. Putative sequences were then

checked against known sequences in the NCBI database by using the BLAST engine (BLASTn) to confirm uniqueness. Once sequences were determined to be unique, primers and MGB Taqman probes (Table 2) were designed using Primer Express version 3.0 from Applied Biosystems (Foster City, CA). Primers and probes were also obtained from Applied Biosystems.

Table 2 Primer and probe sequences

Assay	Primer/Probe	Sequence (5' → 3')	Start Position 5'
tularensis A.I	Forward	AGCTTATGCATCGAGTTGAGGTATT	100,855 (Schu S4)
	Reverse	AAAGCTGGCGATCCAAGGT	100,921 (Schu S4)
	Probe	6FAM-ATGATAGATCCTTGGCTTGA-MGBNFQ	100,881 (Schu S4)
tularensis A.II	Forward	CGAGATTTTGTCCACGCTTCT	427,942 (WY96-3418)
	Reverse	TTTGCGCCAAGACCAGAGT	428,002 (WY96-3418)
	Probe	VIC-AAACTTAGTCAAAGGTCG-MGBNFQ	427,964 (WY96-3418)
holarctica	Forward	TTGCCTATCCAATACTCCGAGTTAG	800,609 (LVS)
	Reverse	CAAGCGCCTGGCTTTGATAA	800,670 (LVS)
	Probe	TET-CTCTGGCCAGTTATT-MGBNFQ	800,635 (LVS)
novicida	Forward	TCAATGTTGCTAAAGTCTCTGGAGTT	1,580,042 (U112)
	Reverse	AATGATGGTAATAAAAGAAGTGGAGCTT	1,580,141 (U112)
	Probe	NED-TGTAAAAGCCATATAAAGGC-MGBNFQ	1,580,092 (U112)

MGBNFQ - Minor Grove Binding, Non-Fluorescent Quencher

2.3.4 PCR cycling conditions

Real-time PCR assays were performed on an ABI 7900 using Taqman Universal Master Mix with UNG (Applied Biosystems, cat#4304437). Total reaction volume was 25 µL as recommended by the manufacturer. Individual assay conditions were are follows: 500 nM forward primer, 500 nM reverse primer, 250 nM probe, 20 ng target DNA, and PCR-grade H_20 (Quality Biological) to 25 µL. Thermal cycling conditions for individual reactions were 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 sec at 95 °C, and 60 sec at 60 °C. The multiplex assay consisted of the *tularensis* A.I, *tularensis* A.II, and *holarctica* primer probe sets. Conditions for the multiplex reaction were as follows: 250 nM of each forward primer, 250 nM of each reverse primer and 125 nM of each probe. Thermal cycling conditions for the multiplex

assay were 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 sec at 95 °C and 60 sec at 58 °C. A positive signal was determined by the crossing of a fluorescence threshold of 0.2 before cycle 40. Data analysis was performed using SDS v2.3 (Applied Biosystems). All tests were performed in at least triplicate to ensure reproducibility.

While a *novicida* specific primer probe set was initially designed and tested singularly, it was later determined that it could be omitted from the multiplex reaction. Rather than use all four assays in the multiplex reaction, shared genomic markers among the subspecies allowed a scoring matrix (Table 3) to be used to differentiate the subspecies using only three of the assays.

Table 3 Scoring matrix for triplex assay

Ft Type	A.I assay	A.II assay	*holarctica* assay
Ft A.I	Positive	Negative	Negative
Ft A.II	Negative	Positive	Negative
Ft *holarctica*	Positive	Positive	Positive
Ft *novicida*	Positive	Positive	Negative

Ft - *Francisella tularensis*

2.3.5 MLVA analysis

Multi-locus variable-number tandem repeat analysis was performed as described previously (20, 102). Briefly, the 11 marker MLVA was set up in five PCR mixes. Mix 1A contained three labeled primers, mix 1B contained one labeled primer, mix 2 contained 3 labeled primers, mix 3A contained three labeled primers, and mix 3B contained one labeled primer. After PCR amplification of the targets, mixes 1A and 1B were combined with water in a final ratio of 2:1:97 respectively, to produce mix 1. PCR products of mix 2 were diluted with water in a final ratio of 1:49. The PCR products of mix 3A and 3B were diluted with water in a final ratio of 2:1:97 to produce mix 3. The resulting three mixes were subjected to capillary electrophoresis on a 16 capillary 3130xl (Applied Biosystems). Each sample was run with 2.25 µL per well MapMarker 1000 (Bioventures) as a nucleic acid size standard. Resulting peaks (Figure. S2)

were compared to a MLVA database kindly provided by Paul Keim for subspecies identification. Positive identifications were called if a minimum of 8 of 11 peaks were present with at least one uniquely identifying peak to distinguish the isolate from others.

2.4 Results

The goal of this work was to address some of the shortfalls of previous PCR assays for *F. tularensis* and develop a multiplex real-time PCR assay useful for the rapid identification and characterization of *F. tularensis* isolates commonly found in the United States and Europe. Markers unique to each of the *F. tularensis* subspecies could not be identified by us or others (72). Molins-Schneekloth *et al.* (72) identified a genetic marker that was unique to type A.I *tularensis*, subspecies *holarctica* and subspecies *novicida*, and another marker unique to type A.II *tularensis*, subspecies *holarctica* and subspecies *novicida.* Based on these differences we found it more reasonable and economical to use a multiplex assay with a scoring matrix to type the four different types of *F. tularensis* (Table 3).

The development of this real-time PCR assay has broad application across the fields of medical surveillance and CDC select agent detection. The different types and subspecies of *F. tularensis* differ not only in their capacity to cause disease but in their geographic distribution as well (100).

2.4.1 Assay validation

For the validation of the multiplex PCR assay, results from 119 *Francisella* isolates were compared to other PCR assays (126) and the MLVA data from this study (Table S1). On 32 (~27 %) of the isolates, insufficient MLVA data was obtained for an identification. The

experiment was repeated 3 times with the same result. Even the U112 strain, which is widely known to be a *novicida* isolate, was inconclusive using the MLVA technique. This illustrates the need for additional typing methodologies.

Of the 87 isolates for which the MLVA was successful, 85 (~98 %) isolates had the same subspecies identification as the multiplexed PCR assay. The assay was also tested against 3 near neighbours, *Wolbachia persica, Francisella philomiragia* 25015, and *F. philomiragia* 25016; each with a negative result. The assay also showed no cross-reactivity with other laboratory species (Table 4).

Table 4 Other isolates tested

Brucella abortus
Burkholderia pseudomallei
Clostridium botulinum
Escherichia coli
Francisella philomiragia
Mycobacterium avium
Mycobacterium tuberculosis
Mycobacterium ulcerans
Pasteurella multocida
Salmonella choleraesuis
Shigella dysenteriae
Staphlococcus aureus
Streptococcus pneumoniae
Streptococcus pyogenes
Wolbachia persica
Yersinia pestis

Multiplex PCR results for isolates 70000907 and 6168 did not match up with the MLVA data. While the multiplexed PCR assay categorized isolate 70000907 as *novicida*, the MLVA analysis grouped it as type A.I. This is consistent with another PCR assay (126), which classified this isolate simply as Type A. The multiplex PCR data classified isolate 6168 as Type B while both the MLVA and the Kugeler *et al.*, (126) PCR assay classified it as *novicida*. The

reasons for these discrepancies are not immediately clear but may be due to genomic rearrangement. These discrepancies are a focus of ongoing research.

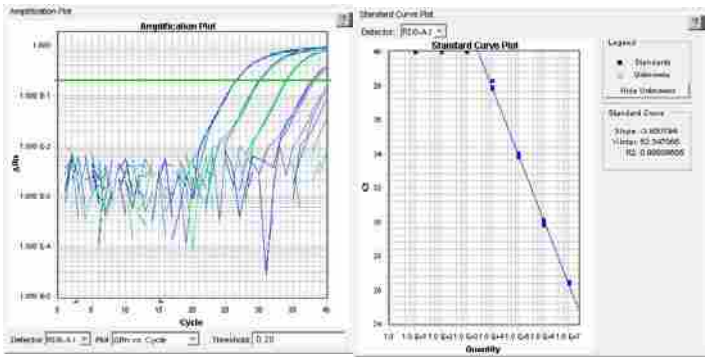
2.4.2 Assay sensitivity

Serial tenfold dilutions of genomic *F. tularensis* DNA were assayed to estimate the detection limits when used in singleplex and multiplex reactions. In singleplex reactions, the Type A.II *tularensis* was the most sensitive at 25 fg. Type A.I, Type B and *novicida* subspecies all had the same detection limit at 250 fg (Figure. S3). Assuming a genome weight of 2.2 fg, limits of detection were ~11 organisms for the Type A.II, and ~114 organisms for the Type A.I, Type B and *novicida* assays. In the multiplex reactions, limits of detection were reduced by one $\log_{10}$ step; the Types A.I, A.II and *novicida* all had limits of detection at 250 fg while Type B was at 2.5 pg (Figure. 3). Again, with a genome weight of 2.2 fg, this translates to ~114 organisms for the Type A.I, Type A.II and *novicida* assays, and ~1 136 organisms for the Type B assay. Each of the assays were run in triplicate with identical results.

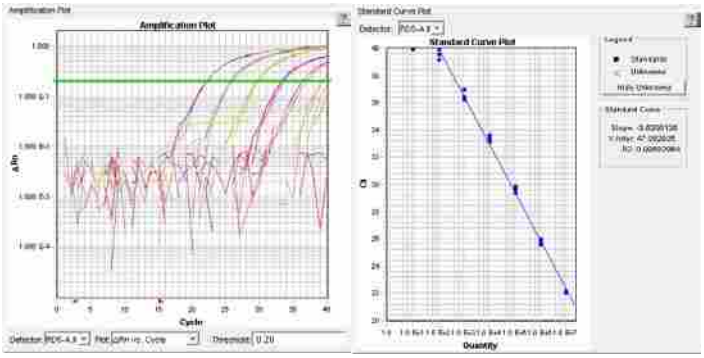
2.4.3 Characterization of isolates

We used this multiplex assay to characterize *F. tularensis* isolates, the majority of which were natural isolates from Utah and New Mexico. Of all the isolates tested (Table S1) 5.0 % were *F. tularensis* Type A.I, 76.5 % were *F. tularensis* Type A.II, 6.7 % were identified as *F. tularensis* Type B, and 11.8 % were identified as *F. tularensis* subspecies *novicida*.

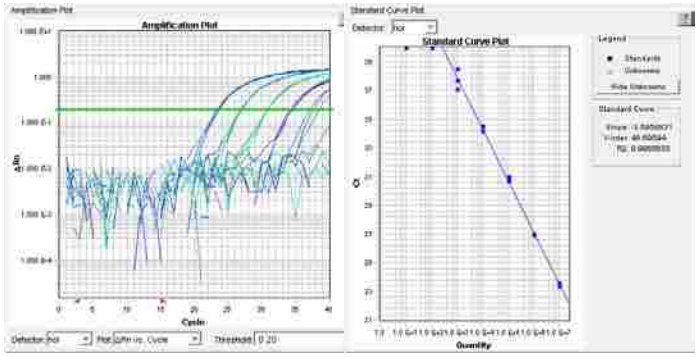
(a)


Amplification Plot
Amplification Plot
Cycle
ΔRn
Standard Curve Plot
Standard Curve Plot
Quantity
Ct

(b)


Amplification Plot
Amplification Plot
Cycle
ΔRn
Standard Curve Plot
Standard Curve Plot
Quantity
Ct

(c)


Amplification Plot
Amplification Plot
Cycle
ΔRn
Standard Curve Plot
Standard Curve Plot
Quantity
Ct

(d)

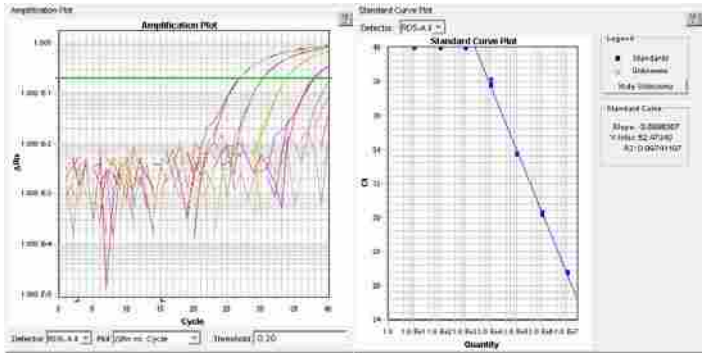

Amplification Plot
Amplification Plot
Cycle
ΔRn
Standard Curve Plot
Standard Curve Plot
Quantity
Ct

Figure 3 Sensitivities of multiplex assays
Tenfold serial dilutions of *F. tularensis* chromosomal DNA were tested to determine individual assay sensitivities. (a) Type A.I, (b) Type A.II, (c) Type B, (d) *novicida* (analysed using the A.II detector). No Template Control (NTC) samples were negative for each sample processed. Calculated PCR efficiencies (a) 81.6%, (b) 92.0%, (c) 89.7% (d) 86.7%.

2.5 Discussion

Many PCR assays for *F. tularensis* and its subspecies have been developed. An assay developed by Broekhuijsen *et al*., (21) is capable of discriminating the four subspecies of *F. tularensis*. However, it does not have the convenience of real-time detection and necessitates running the products on a gel to visualize amplicons up to 3 kilobases in length. Furthermore, the assay is not multiplexed nor can it discriminate between Type A.I and Type A.II *tularensis*. Kugeler *et al*. (126) developed a real-time PCR assay for *F. tularensis* that was able to distinguish Type A from Type B *tularensis*, but it could not differentiate Type A.I from Type A.II, nor could it identify the *novicida* subspecies. Finally, Molins-Schneekloth *et al*. (72) were able to develop a multiplex PCR assay that could differentiate Type A.I *tularensis* from Type A.II, but the assay was not real-time, nor could it differentiate among Type B *tularensis* or the *novicida* subspecies. This work describes the first multiplexed, real-time PCR assay for the characterization of the major types of *F. tularensis* found in the United States and around the world: Type A.I, Type A.II, Type B, and subspecies *novicida*. Since type A.I *tularensis* is usually found on the East Coast of the United States, an outbreak of type A.I on the West Coast could be an indication of an intentional release. This multiplex assay could help law enforcement agencies to identify possible bioterrorism events as well as guide the administration of therapeutics by health officials.

A major hurdle for creating *Francisella tularensis* PCR assays specific for subspecies is that their genomes are highly similar. The 16S rRNA genes which are typically used for

identification of microorganisms exhibit 98.5% to 99.9% similarity across all subspecies which equates to 2 to 23 nucleotides difference (43). A recent study further highlighted the similarities of *F. tularensis* subspecies. Their results show that a pairwise alignment of sequenced draft genomes of low and high virulence subspecies exhibited over 95 % similarity (25). Using various regions of difference identified by SSH (72) and comparative genomic data using Mauve software (53), we were able to identify genomic regions conducive to the creation of real-time PCR assays.

To validate the assays, we tested them against other laboratory strains (Table 4) in addition to two near-neighbours from the same genus: *F. philomiragia* as well as the nearest neighbour of *F. tularensis*, *Wolbachia persica* (43). The results showed no cross-reactivity among these other species, demonstrating that the identified regions were unique to the subspecies of *F. tularensis*. Furthermore, the subspecies identification based on the scoring matrix (Table 3) show a successful identification of ~98 % of the isolates, compared to the MLVA data. This discrepancy was not an artifact of the multiplex assay, as it was also seen in the singleplex reaction. Despite these slight differences, we believe that this multiplex assay remains a powerful, rapid, presumptive screening test for the subspecies of *F. tularensis*. In addition, this work further underscores the need for multiple identification assays to definitively identify these closely related subspecies.

In singleplex reactions, we demonstrated a sensitivity based on serial 10-fold dilutions of stock DNA of 11 organisms for the Type A.II, and 114 organisms for the Type A.I, Type B, and *novicida* assays. These sensitivities are within an order of magnitude of other published, real-time *F. tularensis* PCR assays (125, 126). In the triplex reactions, the Type A.I, Type A.II, and *novicida* assays yielded a sensitivity of 114 organisms and 1 136 organisms for the Type B

reaction. The sensitivities remained the same for the Type A.I and *novicida* assays with an order of magnitude decrease in sensitivity for the Type A.II and Type B assays. Another multiplex real-time PCR assay for *F. tularensis* demonstrated sensitivities as low as 0.5 genome equivalents (127). This increased sensitivity may be explained by two factors: 1) the assay targeted the 16S rDNA, of which multiple copies are present in the genome and 2) the multiplex assays only included two assays thus reducing the chance of competitive PCR problems. When Tomaso *et al.* (127) included three assays in their multiplex assay, no amplification curves were observed.

In larger multiplex reactions, it is not uncommon to see a decrease in sensitivity. In a quadruplex real-time PCR assay for *Yesinia pestis*, Stewart *et al*. (128) saw a 10 fold decrease in the sensitivity when their assay was moved from singlplex to quadruplex: from 150 pg to 1.5 ng. The results for our triplex assay are similar to these results showing a ten-fold decrease in sensitivity for two of the assays when used in a multiplex format: from 114 organisms to 1 136 organisms.

This multiplex assay was used to characterize a large collection of *F. tularensis* isolates. Many of the natural isolates were obtained from the Utah and New Mexico Departments of Health. Some standard laboratory strains of *F. tularensis*, such as Schu S4, LVS, U112 and other *novicida* variants were included for reference. The characterization including all isolates except the standard strains and *novicida* variants returned 104 isolates characterized. Of those, the vast majority (91 isolates or ~88 %) were identified as Type A.II. Four isolates (~3.8 %) were identified as Type A.I, seven isolates (~6.7 %) were identified as Type B, and one isolate (~0.9 %) was identified as *novicida*. This one isolate identified as *novicida* however is likely a

misidentification because another PCR assay (Kugler *et al*., 2006) as well as the MLVA data identified it as Type A.

It is not surprising that the vast majority of the tested isolates were Type A.II since most of the isolates used in this study were obtained from the Utah and New Mexico Departments of Health. This distribution of isolates is consistent with previous data (19, 100) suggesting that Type A.II *tularensis* is predominant in the western United Sates and that Type A.I *tularensis* is predominantly found in the eastern United States.

F. tularensis is classified as a Category A select agent by the CDC. Due to its potential use as a bioterrorism weapon, its endemic status in various locations around the world, and differences in virulence among the various types of *tularensis*, it is important that government agencies, health departments, and hospitals be able to rapidly identify each subtype. The assay presented in this work is a rapid, single-tube, multiplex, real-time PCR assay that can be used to quickly screen individual samples or adapted for high-throughput applications in either 96- or 384-well formats. We believe that this assay will be an invaluable tool in the presumptive identification and characterization of *F. tularensis* isolates during outbreaks of disease or possible bioterrorism events.

Chapter 3 Natural selection in virulence genes of *Francisella tularensis*

3.1 Abstract

A fundamental tenet of evolution is that alleles that are under negative selection are often deleterious and confer no evolutionary advantage. Negatively selected alleles are removed from the gene pool and are eventually extinguished from the population. Conversely, alleles under positive selection do confer an evolutionary advantage and lead to an increase in the overall fitness of the organism. These alleles increase in frequency until they eventually become fixed in the population. *Francisella tularensis* is a zoonotic pathogen and a potential biothreat agent. The most virulent type of *F. tularensis*, Type A, is distributed across North America with Type A.I occurring mainly in the east and Type A.II appearing mainly in the west. *F. tularensis* is thought to contain a genome in decay (losing genes) because of the relatively large number of pseudogenes present in its genome. We hypothesized that the observed frequency of gene loss/pseudogenes may be an artifact of evolution in response to a changing environment, and that genes involved in virulence should be under strong positive selection. To test this hypothesis we sequenced and compared whole genomes of Type A.I and A.II isolates. We analyzed a subset of virulence and housekeeping genes from several *F. tularensis* subspecies genomes to ascertain the presence and extent of positive selection. Eleven previously identified virulence genes were randomly chosen and screened for positive selection along with 10 randomly chosen housekeeping genes. Analyses of selection yielded one housekeeping gene and 7 virulence genes which showed significant evidence of positive selection at loci implicated in cell surface structures and membrane proteins, metabolism and biosynthesis, transcription, translation and cell separation, and substrate binding and transport. Our results suggest that while the loss of

functional genes through disuse could be accelerated by negative selection, the genome decay in *Francisella* could also be the byproduct of adaptive evolution, as evidenced by several of its virulence genes which are undergoing strong, positive selection.

3.2 Introduction

Francisella tularensis is a Gram-negative facultative intracellular pathogen and is the causative agent of the zoonotic disease, tularemia (rabbit fever). While many mammals are susceptible to the disease, they are not known to act as environmental reservoirs (129). Arthropods such as ticks and deer flies are likely the primary vectors in the spread of tularemia, with human infection usually resulting from the bite of an insect or inhalation of aerosolized bacteria (27, 130). While there are no confirmed reservoirs for *F. tularensis*, it has been demonstrated that some strains cause the rapid encystment of *Acanthamoeba castellanii* and are able to survive within these amoeba cysts for up to three weeks, thus increasing the environmental persistence of the bacteria (122, 129).

F. tularensis is highly virulent with an infectious dose of around 10 organisms (6). Because of its highly infectious nature, *F. tularensis* has, in the past, been characterized as a bioweapon by many nations (118). The Centers for Disease Control and Prevention (CDC) has classified *F. tularensis* as a Tier 1 potential agent of bioterrorism. While there is a vaccine for tularemia, it is still classified as an investigational new drug (IND) and is only available for high risk personnel such as the military and certain laboratory workers (131). In order to develop a licensed vaccine, it is imperative that the virulence mechanisms of *F. tularensis* be investigated and understood.

There are currently three recognized species within the *Francisella* genus: *F. tularensis, F. philomiragia,* and *F. novicida.* It has been suggested that *F. novicida* be reclassified as a subspecies of *F. tularensis*, and much of the scientific literature, including Bergey's manual of Systemic Bacteriology, already reflects this change (2, 97, 100, 130, 132). Excluding *F. novicida*, there are three recognized subspecies of *F. tularensis*: *tularensis, holarctica,* and *mediasiatica*. Two of the subspecies of *Francisella tularensis* are often abbreviated simply as Type A (*tularensis*) and Type B (*holarctica*). Furthermore, Type A tularensis has been further divided into A.I and A.II types based on geographic distribution and genome architecture (20). Type A tularensis is generally found in North America with Type A.I dominating the central and eastern portions of the continent while Type A.II is typically found in the western United States (19). Type B *tularensis* (or *holarctica*) exhibits a worldwide prevalence, while the *mediasiatica* subspecies appears to be confined to the central Asian republics of the former Soviet Union (20, 133) (Table 5). The phylogenetic relationships of these subspecies are shown in Figure 4. Type A *tularensis* is the most virulent subspecies of *F. tularensis* (23, 24).

Table 5 Subspecies of *F. tularesnsis* and their worldwide distribution

Subspecies	**Type**	**Distribution**
tularensis	A.I	Eastern United States
tularensis	A.II	Western United States
holarctica	B	North America, Eruope, Japan
mediasiatica	N/A	Central Asain Republics
novicida	N/A	Worldwide

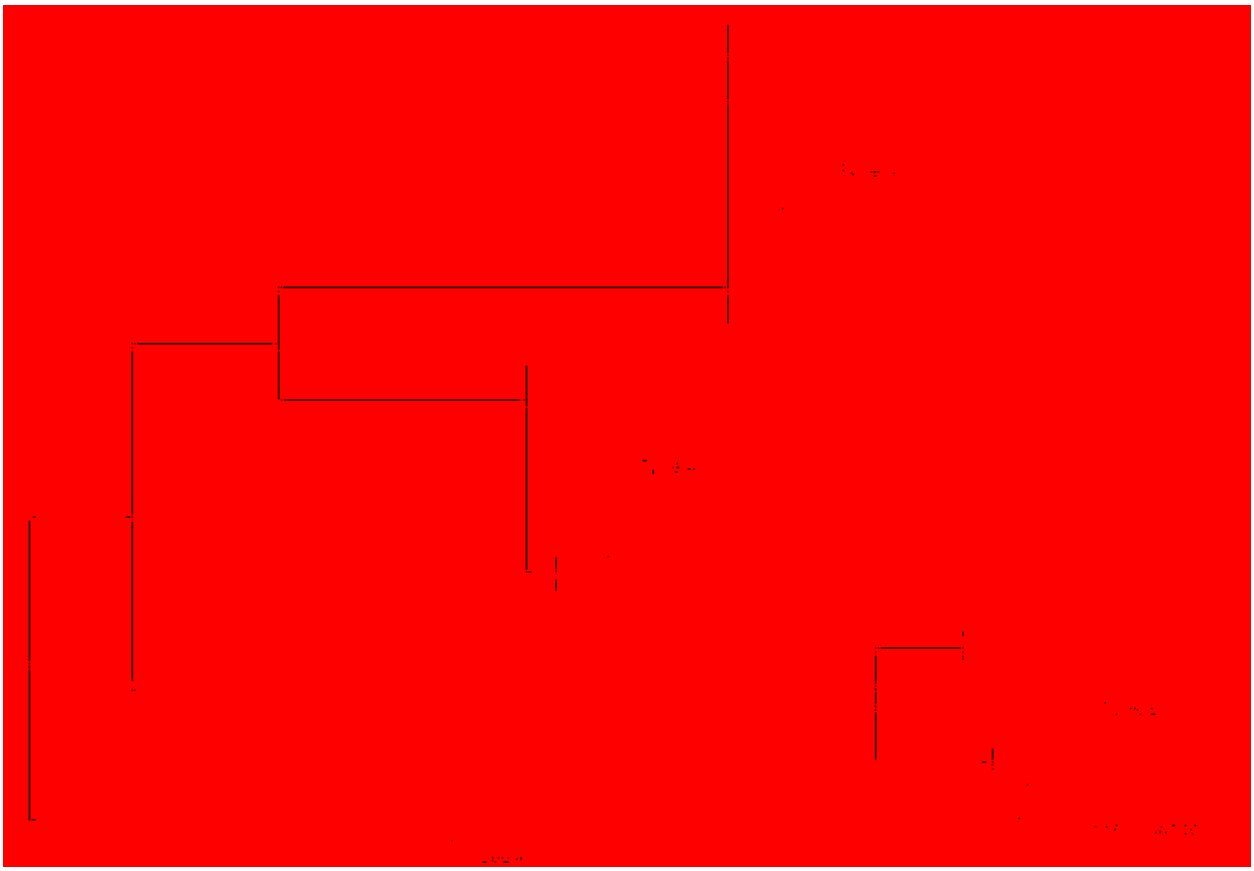


Figure 4 Maximum likelihood tree inferring the phylogenetic relationships of the F. tularensis subspecies Tree was constructed by concatenating 10 housekeeping genes (*recA*, *gyrB*, *groEL*, *dnaK*, *rpoA1*, *rpoB*, *rpoD*, *rpoH*, *fopA*, and *sdhA*) followed by alignment with Clustal W and generation of the tree with MEGA 5.2. Taxa names use the shorthand described in Table 2.

Virulence genes are those whose products contribute to an organism's ability to infect and colonize a host. Virulence genes are often located within mobile genetic elements such as plasmids, transposons, bacteriophages and pathogenicity islands, which enhance genetic diversity through the exchange with other bacteria (134). *F. tularensis* possesses a pathogenicity island of about 30 kb in length consisting of 17 open reading frames flanked by inverted repeat sequences (47). The *Francisella* Pathogenicity Island (FPI) has been shown to be required for intra-macrophage growth (46), phagosome escape and general virulence of F. *tularensis* (135). The subspecies *novicida* is most similar to the ancestral lineage of the *F. tularensis* subspecies

and contains only one copy of the FPI while the other more pathogenic subspecies contain two copies of the FPI (47). It should be noted that the inverted repeats flanking the FPI provide a direct means for this duplication to take place.

Despite the varied mechanisms for sharing genes among bacteria, molecular evidence suggests that the four subspecies of *F. tularensis* have evolved and acquired virulence genes by vertical descent rather than by horizontal gene transfer (97). A whole genome analysis of horizontal gene transfer (HGT) in *F. tularensis* identified 30 candidate genes as having been acquired through HGT (136), but none of them were putative virulence genes. Genetic variation in the subspecies of *F. tularensis* seems to have arisen by mutation rather than HGT (110-112). Furthermore, the subspecies *novicida* has been shown to possess a CRISPER/Cas system to defend against invading genetic elements, further supporting the notion that mutation and selection are the driving factors of evolution in *F. tularensis* virulence rather than the acquisition of mobile genetic elements by HGT (113, 114).

Whole genome comparisons between pathogenic and non-pathogenic subspecies of *F. tularensis* revealed that the pathogenic subspecies evolved from the non-pathogenic subspecies by mechanisms of genomic rearrangements, point mutations and small indels (36). The many pseudogenes present in the *F. tularensis* genome suggest that it is a genome in decay (137). While these pseudogenes may have arisen via negative selection through disuse, it is also possible that this evidence of decay is a byproduct of an adaptive response to a changing environment (138, 139) or even the evolution of avirulence (140, 141). Alternatively, we hypothesize that in addition to genome decay through the creation of pseudogenes, positive natural selection is also a driving force in the continued evolution of virulence of this important pathogen.

3.3 Materials and Methods

3.3.1 Bacterial strains and culture conditions

The isolates used in this study (Table 6) are a part of a select agent archive housed at the special pathogens laboratory at Brigham Young University (Provo, UT). The collection largely consists of isolates obtained from the State Health Departments of Utah and New Mexico over the past two decades. All *F. tularensis* isolates were grown on modified Mueller Hinton agar (MMHA) (Becton Dickinson and Company, Franklin Lakes, New Jersey, USA) for 3-4 days with 5 % CO_2 at 35 °C. MMHA was prepared by autoclaving the Mueller Hinton base, which was chocolatized by adding 5 % sheep blood while the medium was approximately 80 °C. After the medium cooled to 50 °C, 10 mL of 10 % glucose and 20 mL of IsoVitaleX (Becton Dickinson and Company) were added to 1 L.

3.3.2 Genome sequencing and annotation

Total genomic DNA was extracted from each isolate using the MagNA Pure System (Roche) and the MagNA Pure LC DNA Isolation Kit III (Roche) according to the manufacturer's directions. Briefly, cells grown on MMHA agar were suspended in 250 µL of TE buffer containing 1.8 µg/µL lysozyme and incubated for 1 h at 37 °C. To this tube, 270 µL of bacterial lysis buffer and 100 µL of proteinase K were added and the tube was incubated for 10 min at 65 °C. Samples were then incubated in boiling water for 10 min to inactivate pathogens. DNA was eluted in a total volume of 100 µL. DNA concentration was measured using a PicoGreen assay (Invitrogen) and TBS-380 fluorometer (Turner Biosystems). Prior to genome sequencing, isolates were typed using the multiplex PCR assay as described previously (2). Four Type A.I and six Type A.II strains were selected for sequencing in order to maximize the available

geographic and genetic diversity. Whole genome shotgun sequencing was accomplished using the Roche/454

Table 6 Genomes used for analysis of selection

Genome	Shorthand	Type	Reference	Isolated from
F. tularensis subsp. *tularensis* 70001275	tul_FT_01	A.II	This work	Utah, US
F. tularensis subsp. *tularensis* 70102010	tul_FT_02	A.II	This work	Utah, US
F. tularensis subsp. *tularensis* 80700103	tul_FT_03	A.II	This work	Utah, US
F. tularensis subsp. *tularensis* 831	tul_FT_05	A.II	This work	New Mexico, US
F. tularensis subsp. *tularensis* AS_713	tul_FT_06	A.II	This work	New Mexico, US
F. tularensis subsp. *tularensis* 3571	tul_FT_07	A.II	This work	New Mexico, US
F. tularensis subsp. *tularensis* 1378	tul_FT_08	A.I	This work	New Mexico, US
F. tularensis subsp. *tularensis* 80700075	tul_FT_09	A.I	This work	Utah, US
F. tularensis subsp. *tularensis* 79201237	tul_FT_10	A.I	This work	Utah, US
F. tularensis subsp. *tularensis* 80700069	tul_FT_11	A.I	This work	Utah, US
F. tularensis subsp. *tularensis* SCHU S4	tul_SchuS4	A.I	Larsson et al. 2005	Ohio, US
F. tularensis subsp. *tularensis* WY96-3418	tul_WY96_3418	A.II	Beckstrom-Sternberg et al. 2007	Wyoming, US
F. tularensis subsp. *tularensis* FSC198	tul_FSC198	A.I	Chaudhuri et al. 2007	Slovakia
F. tularensis subsp. *tularensis* NE061598	tul_NE061598	A.I	Nalbantoglu et al. 2010	Nebraska, US
F. tularensis subsp. *tularensis* TI0902	tul_TI0902	A.I	Modise et al. 2012	Virginia, US
F. tularensis subsp. *tularensis* TIGB03	tul_TIGB03	A.I	Modise et al. 2012	Virginia, US
F. tularensis subsp. *holarctica* F92	hol_F92	B	Antwerpen et al. 2013	Germany
F. tularensis subsp. *holarctica* FSC200	hol_FSC200	B	Svensson et al. 2013	Sweden
F. tularensis subsp. *holarctica* FTNF002_00	hol_FTNF002_00	B	Barabote et al. 2009	Spain
F. tularensis subsp. *holarctica* LVS	hol_LVS	B	Unpublished	Russia
F. tularensis subsp. *holarctica* OSU18	hol_OSU18	B	Petrosino et al. 2006	Oklahoma, US
F. tularensis subsp. *mediasiatica* FSC147	med_FSC147	N/A	Champion et al. 2009	Kazakhstan

GS FLX Pyrosequencer (Roche) and the Titanium® chemistry according to the manufacturer's recommendations. Raw sequencing reads were aligned to reference sequences using Newbler version 2.6. The reference sequence for A.I strains was SCHU S4 (GenBank accession number AJ749949.2), while WY96-3418 (NC_009257.1) was used for A.II strains. Aligned shotgun sequences were annotated using the Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (142).

3.3.3 Analysis of selection

A total of 22 taxa were included in the analysis of selection; the 10 new *F. tularensis* subspecies *tularensis* genomes described in this work, and 12 other *F. tularensis* genomes of the *tularensis, holarctica,* and *mediasiatica* subspecies retrieved from GenBank (Table 6).

In addition to 5 Type IV pilus genes which are known virulence factors of *F. tularensis*, 89 other previously identified virulence genes (143) were analyzed for positive selection. In total, 94 virulence genes were extracted from the genomes of the 22 taxa in Table 6 for analysis. While the taxa used in this work were isolated from various geographical locations (Table 6), the genes selected for analysis are common to all the isolates. Individual genes were aligned with ClustalW (144) using the default parameters. A hierarchical likelihood ratio test was performed on the aligned sequences to determine which model of evolution best fit the data (AIC) (145) using MEGA v. 5.2 (146). MEGA v. 5.2 was also used to generate phylogenetic trees via maximum likelihood assuming the previously identified best-fit substitution model (146). Maximum likelihood computations of dN and dS were conducted using HyPhy (147-151).

After analyses of selection with HyPhy, 11 genes were arbitrarily selected for further analysis using TreeSAAP version 3.2 (152), which measures selection based on 31 physiochemical properties of amino acids. TreeSAAP compares aligned sequences in the context of the specified phylogenetic topology, codon by codon, to infer amino acid replacement events. Results were divided into one of eight categories based on the magnitude of the change: categories 1-3 indicate a conservative change, categories 4-6 represent moderate change, while categories 7-8 represent more drastic changes and indicate positive selection. Only genes that

returned changes in categories 7-8 were examined further. A z-score was calculated by TreeSAAP for each of the 31 physiochemical properties. Drastic changes (categories 7-8) with a z-score of $p < 0.001$ were determined to indicate statistically significant positive selection. To summarize, we explored 94 virulence-associated genes. Of those, 64 were indicated by HyPhy analysis to be under positive selection, though the results were not statistically significant. From the 64 genes that showed some evidence of positive selection, we arbitrarily chose 11 for subsequent TreeSAAP analysis. To better visualize the evolutionary changes in the proteins, secondary structures of the proteins were predicted using the PSIPRED secondary structure prediction (http://bioinf.cs.ucl.ac.uk/psipred/) (153).

3.4 Results and Discussion

3.4.1 Genome sequencing and annotation

Of the 10 novel genomes sequenced, 4 were *F. tularensis* Type A.I and 6 were *F. tularensis* Type A.II. A summary of the sequencing statistics of these newly sequenced genomes can be found in Table 7.

The lengths of the reference sequences were 1,892,775 and 1,898,476 for the SCHU S4 and WY96-3418 strains, respectively. The average length of the sequenced Type A.I genomes was 1,814,938 and the average length of the Type A.II genomes was 1,814,544. Thus, the sequenced genomes are about 80 kb shorter than the reference sequences. This is not surprising since the shotgun method of sequencing cannot adequately sequence through large repetitive regions of DNA resulting in gaps in the final alignments.

Table 7 Summary of sequenced and aligned *F. tularensis* genomes

Strain Designation	70001275	70102010	80700103	831	AS_713	3571	1378	80700075	79201237	80700069
Shorthand Designation	FT-01	FT-02	FT-03	FT-05	FT-06	FT-07	FT-08	FT-09	FT-10	FT-11
Type	A.II	A.II	A.II	A.II	A.II	A.II	A.I	A.I	A.I	A.I
Average Sequencing Depth	212x	164.8x	76.6x	101.2x	107.7x	101.3x	93.2x	87.5x	92.2x	127.8x
Reference sequence	WY96-3418 (Accession number NC_009257.1)	WY96-3418 (Accession number NC_009257.1)	WY96-3418 (Accession number NC_009257.1)	WY96-3418 (Accession number NC_009257.1)	WY96-3418 (Accession number NC_009257.1)	WY96-3418 (Accession number NC_009257.1)	SCHU S4 (accession number AJ749949.2)	SCHU S4 (accession number AJ749949.2)	SCHU S4 (accession number AJ749949.2)	SCHU S4 (accession number AJ749949.2)
N50 contig size (bases)	137,296	124,943	43,424	61,138	73,318	103,637	360,449	348,066	239,117	348,045
Sequenced genome length	1815079	1816669	1810598	1814141	1814591	1816188	1815377	1814277	1814563	1815536

Table 8 Summary of annotation results for *F. tularensis* genomes

Shorthand Designation	FT-01	FT-02	FT-03	FT-05	FT-06	FT-07	FT-08	FT-09	FT-10	FT-11	SCHU S4	WY96-3418
Type	A.II	A.II	A.II	A.II	A.II	A.II	A.I	A.I	A.I	A.I	A.I	A.II
Number of protein sequences[a]	1784	1792	1808	1817	1809	1800	1785	1796	1781	1787	1804	1824
Number of rRNA sequences	6	6	5	7	6	6	7	8	6	6	10	10
Number of tRNA sequences	32	32	32	32	32	32	32	32	32	32	38	38
Total number of genes	1822	1830	1845	1856	1847	1838	1824	1836	1819	1825	1852	1872
GenBank accession number	AMPP00000000	AMPY00000000	AMPX00000000	AMPV00000000	AMPU00000000	AOUD00000000	APKY00000000	APKX00000000	APKW00000000	APKV00000000	AJ749949.2	NC_009257.1

[a] includes pseudogenes

The number of predicted protein-coding sequences for the references SCHU S4 and WY96-3418 are 1604 and 1634, respectively. Because of the shortened size of the newly sequenced genomes, it was expected that the total number of protein sequences would also be less. This, however, was not the case. The new genomes were processed through an automated annotation pipeline, and were not manually curated. This resulted in many sequences that the automated pipeline mistook for genes that the curators of the reference genomes had already removed. Further study and curation of these newly sequenced genomes will be necessary for detailed comparative analyses, but sufficiently resolved to facilitate our analysis of selection on their virulence-associated genes.

3.4.2 Analysis of selection

The analysis of d_N/d_S ratios is a well-known benchmark for identifying selection and was used as a first pass to identify genes that are most likely to be under positive selection. Results revealed that of the 94 previously identified virulence genes (Su et al. 2007), 64 gave some indication of positive selection (although with p-values ranging from 0.40 to 0.90, none of them were statistically significant). It has been suggested that some of the assumptions of the d_N/d_S ratio test, as well as the McDonald-Kreitman test of adaptive evolution (154) can be too conservative for accurately detecting positive selection and adaptation (152, 155-157). For example, d_N/d_S ratios alone were unable to detect positive selection in rapidly mutating genes of HIV in patients showing signs of drug resistance (157). Furthermore, Sharp (158) asserts that another problem with d_N/d_S ratio tests is that their significance falls as selection continues to "weed out" the effects of detrimental amino acid changes. Thus, more nuanced methods may be needed to accurately predict positive selection in biological systems (156). Finally, while d_N/d_S

ratios and the McDonald-Kreitman test may indicate the presence of selection on a gene, they do not give any indication of how selection affects the structure and function of the protein (159).

To perform a more sensitive analysis we used TreeSAAP (Selection on Amino Acid Properties using phylogenetic Trees) which tests for selection of expected random distributions of possible amino acid changes based on 31 physiochemical properties of amino acids and associated phylogenetic trees (152). The methodological approach that TreeSAAP uses to detect selection has been shown to be better suited at detecting selection where more traditional methods such as d_N/d_S ratios and McDonald-Kreitman analyses cannot (152, 158). We note that TreeSAAP doesn't have a built-in correction for multiple-hypothesis testing when calculating p-values. Although a Bonferroni correction of TreeSAAP significance tests is indicated for comparative purposes, previous work has cast doubt on whether this will yield an appropriate level of confidence (160). Thus, we employed TreeSAAP as the best-suited approach currently available for detecting selection.

Since the HyPhy results were not statistically significant, 11 virulence genes were arbitrarily selected for further analyses of selection using TreeSAAP (Table 9). In genes FTT_1611, FTT_1156c, FTT_0935c, and FTT_1525c, no significant positive selection was detected. However, significant positive selection (categories 7-8) was detected in one or more physiochemical properties of the remaining seven genes: FTL_1134, FTT_0683, FTT_0881c, FTT_0504c, FTT_0936c, FTT_0766, and FTT_1125. The physiochemical properties for which positive selection was detected include the following: isoelectric point, power to be at the N-terminal, composition, power to be at the middle of alpha helix, solvent accessible reduction ratio, equilibrium constant, power to be at the C-terminal, coil tendencies, compressibility, bulkiness, turn tendencies, and average number of surrounding residues.

Table 9 Virulence genes randomly chosen for TreeSAAP analysis

Locus tag[a]	Gene Name	Positive selection detected by TreeSAAP
FTT_0881c	rocE	Yes
FTL_1134		Yes
FTT_1156c	pilQ	No
FTT_1611		No
FTT_1125	metQ	Yes
FTT_0504c	sucC	Yes
FTT_0935c	bioC	No
FTT_0683	pilD	Yes
FTT_1525c		No
FTT_0766	deoD	Yes
FTT_0936c	bioF	Yes

[a] Locus tag designations for FTT are adopted from the annotation of the SCHU S4 genome (accession no. AJ749949.2) and locus tag designations for FTL are adopted from the annotation of the LVS genome (accession no. AM233362)

For analytical purposes, these seven genes were placed into four functional categories (143): FTL_1134 and FTT_0683 fall into the cell surface structures and membrane proteins category, FTT_0881c, FTT_0504c and FTT_0936c are metabolism and biosynthesis proteins, FTT_0766 is categorized as a protein involved in transcription, translation and cell separation, and FTT_1125 is classified in the substrate binding and transport functional group.

3.4.2.1 Cell surface structures and membrane proteins

In the functional group of cell surface structures and membrane proteins, the gene FTL_1134 demonstrated selection for three of the 31 physiochemical properties tested: isoelectric point in codons 183-203 and 383-394, power to be at the N-terminus in codons 271-285 and 323-337, and composition in codons 183-197 (Figure. 5). Codons 183-203 begins in an α-helix and ends in a loop, codons 271-285 includes portions of two α-helices and a loop in

between, codons 323-337 begins in an α-helix and ends in a loop and codons 383-394 occurs near the C-terminal of an α-helix (Figure S4). The product of FTL_1134 is a hypothetical membrane protein of unknown function. This gene (as were all other genes cited in this work) was identified as a virulence gene by a whole genome transposon screen for virulence genes of *F. tularensis* (143). Thus, in this case, a protein of unknown function has been identified as a virulence gene.

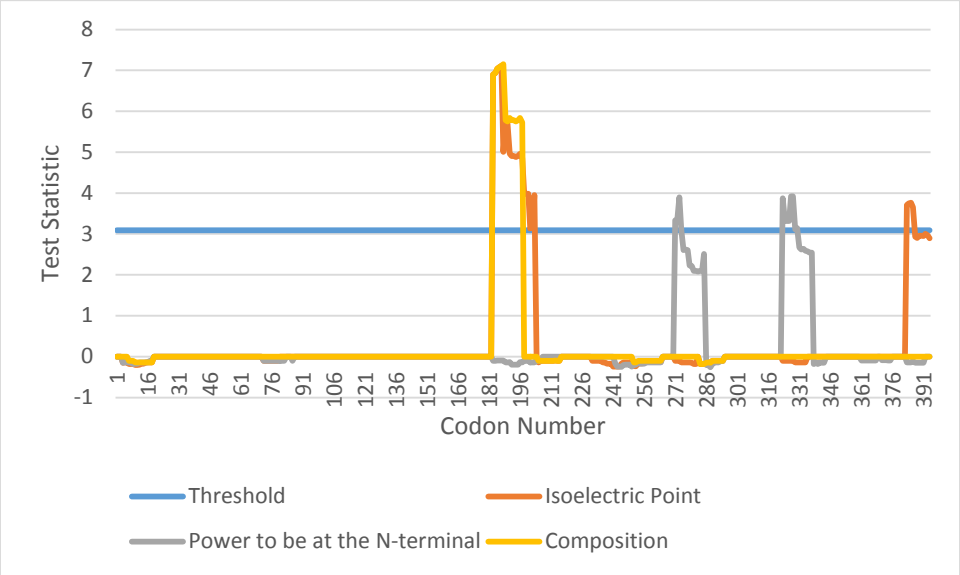


Figure 5 Selection on FTL_1134.
Areas of the gene under selection are indicated where the test statistic is greater than 3.09 (horizontal line) ($p < 0.001$). Selection is present in the gene for the following amino acid properties where the lines cross the threshold: Isoelectric Point (orange line), Power to be at the N-terminal (gray line) and Composition (yellow line).

F. tularensis is an intracellular pathogen which is routinely phagocytized by host cells and taken into a phagosome where the pH is lowered in an effort to rid the cell of the bacterium. Selection of FTL_1134 for the isoelectric point, the pH at which a molecule carries no net electrical charge, hints that this membrane protein may be exposed to changes in pH and is adapting to withstand these changes.

FTL_1134 also demonstrated selection for the amino acid property of power to be at the N-terminal. This amino acid property is defined as the intrinsic property of an amino acid residue to be located at the N-terminus of an α-helix (161). The codons under selection occur in varied domains of the secondary structure of this protein including the N-terminus of an α-helix, the C-terminus of an α-helix and the loop between two α-helices (Figure S4). While this property and its relation to positive selection is not well understood, we speculate that the regions under selection are important to the ability of this virulence gene to function intracellularly since in an infection model with this gene knocked out, the host is able to clear the infection (143).

The third and final amino acid property under selection for FTL_1134 is composition. Composition refers to the individual amino acid make-up of a protein. Since FTL_1134 is a hypothetical membrane protein of an unknown function, it is difficult to know how positive selection for the composition property may affect it. Further study is needed to determine the function of this protein and its role in the virulence mechanisms of *F. tularensis* before we can gain an understanding of how selection is influencing this protein.

The other protein to demonstrate positive selection in the cell surface structures and membrane proteins functional group is FTT_0683, which codes for the type IV pilus protein PilD. Type IV pili are common in Gram-negative pathogens and mediate the attachment of the pathogen to various host receptors (162). Specifically, PilD is a bifunctional, cytoplasmic membrane protein responsible for the cleavage of the N-teminal leader sequences of prepilin and also catalyzes the N-methylation of the N-terminal phenylalanine of mature pilins (163). Without a functional PilD, the type-IV pilus apparatus cannot be assembled, thus restricting the pathogenicity of *F. tularensis*. TreeSAAP detected selection in PilD for the amino acid property of power to be at the middle of α-helix, which is described as the intrinsic property of a specific

amino acid to be located in the middle of an α-helix (161). The catalytic domains of PilD are predicted to be specific cytoplasmic cysteine motifs found toward the N-terminus of the protein (164), well distanced from the selection occurring toward the C-terminus of the protein in amino acids 190-204 (Figure 6). Selection in these membrane bound α-helices do not affect the catalytic activity of PilD, but may alter membrane positioning of the protein (Figure S5).

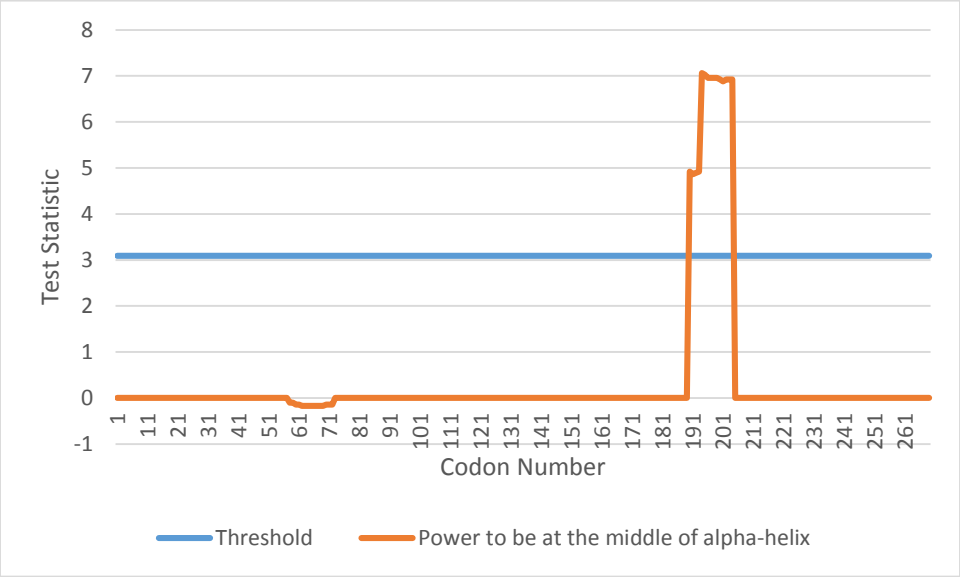


Figure 6 Selection on FTT_0683 (*pilD*).
Areas of the gene under selection are indicated where the test statistic is greater than 3.09 (horizontal line) ($p < 0.001$). Selection is present in the gene for the amino acid property of Power to be at the middle of alpha-helix where the orange line crosses the threshold.

3.4.2.2 Metabolism and biosynthesis

Three genes in the metabolism and biosynthesis functional group demonstrated selection, FTT_0881c (*rocE*), FTT_0504c (*sucC*), and FTT_0936c (*bioF*). The study which identified these genes as virulence factors used a signature-tagged mutagenesis approach (143). Briefly, this approach introduces random transposons to disrupt genes in the genome before infection using a mouse model. If an organism with a particular disrupted gene is not recovered from the

infection model, that gene is then classified as necessary for virulence. This approach has been used to identify virulence factors in many other organisms (165, 166). In the strictest sense, genes categorized as metabolism and biosynthesis should not be considered virulence factors, since impairing them would naturally restrict growth regardless of the status of infection (143, 167). However, it is important not to discount the role these genes may play in infection altogether, since they may be valuable targets for the future development of therapeutics or vaccines.

The first gene in the metabolism and biosynthesis category for which selection was identified was FTT_0881c. This locus codes for the protein RocE which is an amino acid permease, specifically, a transmembrane arginine transporter protein. *Francisella tularensis* lacks the ability to synthesize the amino acid arginine. Furthermore, recent studies have shown that the cytosol of the host cell doesn't contain sufficient free nutrients to support the rapid growth of the pathogen (168). Without access to arginine, the pathogen would not be able to reproduce and infection with *F. tularensis* would fail. To compensate, *F. tularensis* induces the host cell into an autophagy state in which proteins are broken down into their constituent amino acids and made available to the pathogen (168). *F. tularensis* and its transmembrane amino acid permease, RocE, is then ready to scavenge for the newly available amino acids it is not able to produce on its own, among which is arginine. Furthermore, macrophages use arginine to produce nitric oxide (NO) as an antimicrobial defense (169). In addition to scavenging arginine for growth, the ability of *F. tularensis* to remove arginine would also deprive the cell of this potent antimicrobial agent, thus allowing the infection to progress.

The *rocE* gene demonstrated positive selection for the solvent accessible reduction ratio property of amino acids in codons 109-116, 438, and 442 (Figure 7). In the secondary structure

of RocE, codons 109-116 encompass a β-sheet and part of a loop while codons 438 and 442 occur within the same α-helix (Figure S6). The solvent accessible reduction ratio is defined as the decrease in solvent accessibility for an amino acid residue when the protein molecule moves from a hypothetically extended state to the native folded state (170). Consequently, as the solvent accessibility decreases, the hydrophobicity of the protein increases. RocE is a transmembrane protein and hydrophobic domains are necessary for its placement and function in the membrane. Host cells have evolved a mechanism of nutrient restriction as a protection from intracellular pathogens such as *F. tularensis* (168). These results suggest that *F. tularensis* is also evolving mechanisms to scavenge for, and acquire, the necessary nutrients for intracellular survival.

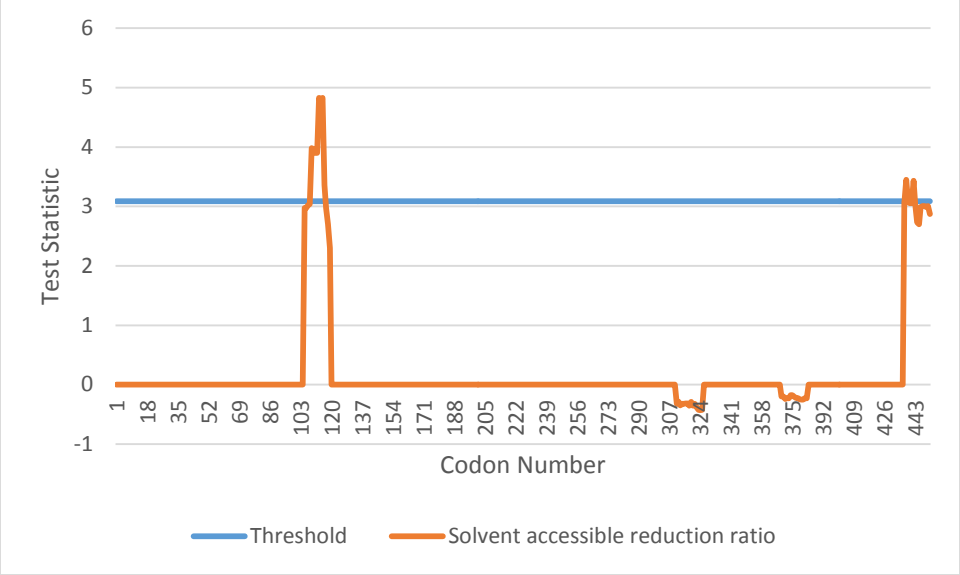


Figure 7 Selection on FTT_0881c (*rocE*).
Areas of the gene under selection are indicated where the test statistic is greater than 3.09 (horizontal line) ($p < 0.001$). Selection is present in the gene for the amino acid property of Solvent accessible reduction ratio where the orange line crosses the threshold.

The second protein in the metabolism and biosynthesis group to demonstrate positive selection is SucC, coded for by the FTT_0504c locus. SucC is the β-chain of the succinyl-CoA

synthetase which when completed with another β-chain and 2 α-chains (SucD), catalyzes the reaction of succinyl-CoA to succinate in the citric acid cycle. Positive selection was identified for the amino acid property of equilibrium constant (ionization of COOH) in codons 31-45 and 84-89 (Figure 8). These codons encompass all three major secondary structures: an α-helix, a β-sheet and a loop, all near the N-terminus of the protein (Figure S7). The areas identified under selection don't appear to be involved in any of the catalytic regions of the protein (171, 172), thus amino acid substitutions in these non-catalytic regions should still allow the completed enzyme to function normally. It has also been demonstrated that the activity of the succinyl-CoA synthetase increases when *Escherichia coli* is cultured in an acidic environment. If succinyl-CoA synthetase behaves silmilarly in *F. tularensis*, this may be part of a stress response initiated when the bacterium is inside a phagolysosome of a host cell (173, 174).

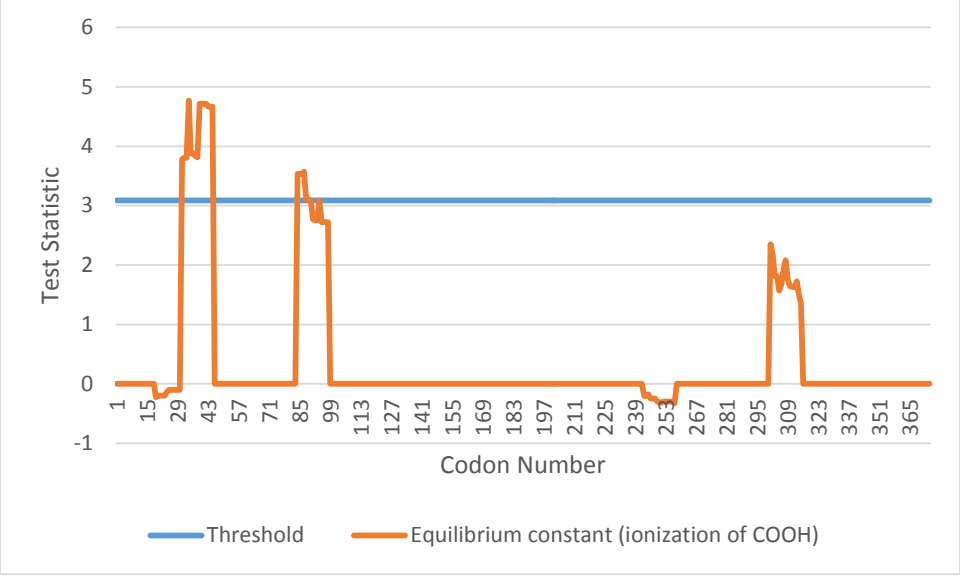


Figure 8 Selection on FTT_0504 (*sucC*).
Areas of the gene under selection are indicated where the test statistic is greater than 3.09 (horizontal line) ($p < 0.001$). Selection is present in the gene for the amino acid property of Equilibrium constant (ionization of COOH) where the orange line crosses the threshold.

The third and final gene under selection in the metabolism and biosynthesis group is FTT_0936c which codes for the protein BioF. BioF is a 8-Amino-7-oxononanoate synthase which is responsible for the first committed step in the biosynthesis of biotin (175). TreeSAAP identified selection for 4 different amino acid properties: power to be at the N-terminal in codons 8-22 and 198-210, power to be at the C-terminal in codons 8-22 and 198-202, coil tendencies in codons 8-22 and 200-203, and compressibility at codons 8-22, and 198-209 (Figure 9). The secondary structures involved in the areas of selection include α-helices, part of a β-sheet and loops (Figure S8). The active site for BioF is at residues 47 and 361 (176), well away from the areas of selection in this gene. It's worth reiterating that genes involved in metabolism and biosynthesis aren't considered virulence genes in the strictest sense, but they shouldn't be discounted either as they may prove to be effective drug targets. Such is the case with plumbagin (5-hydroxy-2-methyl-1,4-naphthoquinone), a natural compound found to be an effective herbicide through the inhibition of 8-Amino-7-oxononanoate synthase (BioF) (177). Because plants, microbes, and some fungi synthesize their own biotin, while animals require trace amounts of biotin in their diets, plumbagin may be a safe and effective herbicide by disrupting the biotin biosynthesis pathway (176-178).

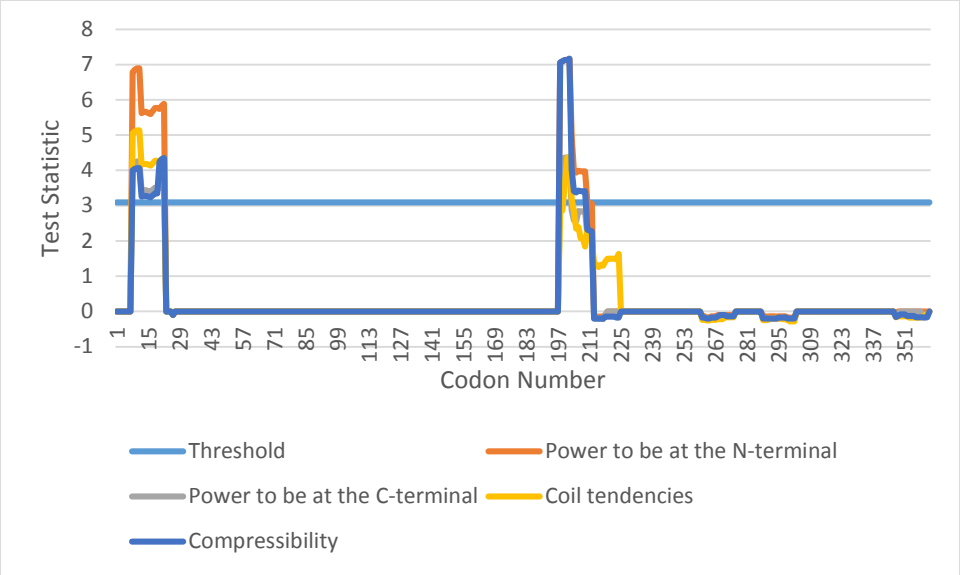


Figure 9 Selection on FTT_0936c (*bioF*).
Areas of the gene under selection are indicated where the test statistic is greater than 3.09 (horizontal line) ($p < 0.001$). Selection is present in the gene for the following amino acid properties where the lines cross the threshold: Power to be at the N-terminal (orange line), Power to be at the C-terminal (gray line), Coil tendencies (yellow line), and Compressibility (dark blue line).

3.4.2.3 Transcription, translation and cell separation

One gene in the transcription, translation and cell separation functional group demonstrated selection, DeoD, coded for by the locus FTT_0766. Of the 31 physiochemical properties of amino acids tested, selection was identified for bulkiness in codons 169-183 and turn tendencies in codons 127 and 176-183 (Figure 10). Bulkiness refers to the relative size of the side chain of a particular amino acid. For example, leucine is considered to be "bulkier" than alanine (179, 180). Substituting amino acids with different bulkiness properties can affect the overall hydrophobicity of a protein or even protein folding and substrate binding where specific steric interactions are important (179). Turn tendencies refers to the propensity of a particular amino acid to be in a β-turn (181) . In terms of secondary structure, the areas of the gene under selection include an α-helix, a β-sheet and loops (Figure S9).

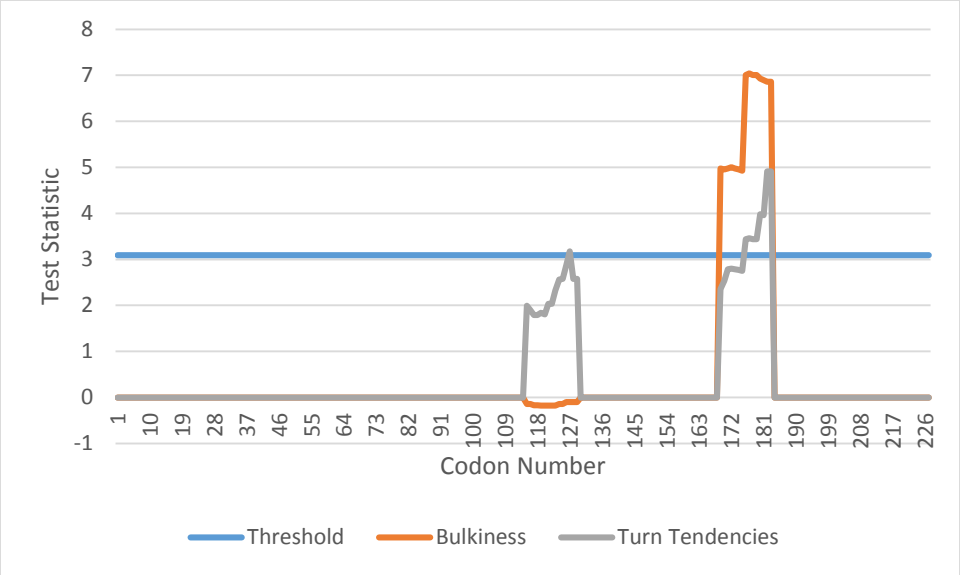


Figure 10 Selection on FTT_0766 (*deoD*).
Areas of the gene under selection are indicated where the test statistic is greater than 3.09 (horizontal line) ($p < 0.001$). Selection is present in the gene for the following amino acid properties where the lines cross the threshold: Bulkiness (orange line), and turn tendencies (gray line).

The product of the *deoD* gene is a purine nucleoside phophoylase (PNP) which catalyzes the phosphorylation of purine ribonucleosides and 2'-deoxyribonucleosides as part of the purine salvage pathway (182). The active site for DeoD is near the N-terminus of the protein while selection was detected further downstream toward the C-terminus, indicating that selection isn't changing the specific function of this enzyme. However, enzymes that are part of the nucleotide biosynthesis pathway, such as PNP, are a favorite target for antimicrobial therapy since they are significantly different from the eukaryotic enzymes that perform the same function (183). Understanding how selection operates on this gene, could lead to better, more effective antimicrobial therapies.

3.4.2.4 Substrate binding and transport

The final functional group to have a gene implicated in selection is substrate binding and transport. Selection was detected for two amino acid properties in the gene *metQ* (FTT_1125):

composition and average number of surrounding residues, both in codons 105-119 (Figure 11). These regions occur within α-helices (Figure S10). Composition refers to the individual amino acid make up of a protein. Most amino acid residues in a protein cannot be replaced without a change in function (184). The average number of surrounding residues refers to the number of residues surrounding a particular amino acid within the effective distance of influence, and is an important factor in changing the hydrophobicity index of the region (170).

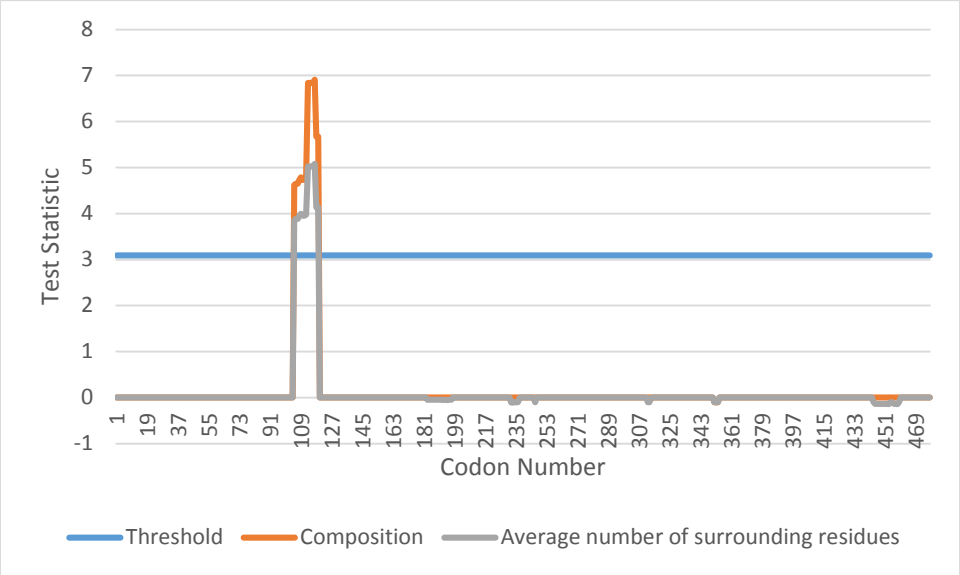


Figure 11 Selection on FTT_1125 (*metQ*).
Areas of the gene under selection are indicated where the test statistic is greater than 3.09 (horizontal line) ($p < 0.001$). Selection is present in the gene for the following amino acid properties where the lines cross the threshold: Composition (orange line), and Average number of surrounding residues (gray line).

MetQ is a D-Methionine-binding protein, which together with Met N (an ATPase) and MetI (a D-methionine permease) make up the D-methionine ABC transporter (185). Similar to RocE discussed earlier in the metabolism and biosynthesis section, which is an arginine scavenger, this D-methionine transporter scavenges for D- methionine from its environment, since *F. tularensis* cannot synthesize methionine. Once transported inside the cell, D-methionine

can be converted to L-methionine and used in protein synthesis (186). This is an important strategy used by *F. tularensis* to survive intracellularly. The fact that MetQ is experiencing selection, implies that *F. tularensis* is continuing to evolve strategies to acquire nutrients from its environment to survive within a host cell.

3.4.2.5 Housekeeping genes

As a control to determine if the virulence genes of *F. tularensis* are experiencing more selection than other genes in the genome, 10 housekeeping genes (*recA*, *gyrB*, *groEL*, *dnaK*, *rpoA1*, *rpoB*, *rpoD*, *rpoH*, *fopA*, and *sdhA*) were subjected to the same tests of selection (HyPhy and TreeSAAP) as the virulence genes described earlier. Since housekeeping genes are responsible for the fundamental functions of the cell, these are typically highly conserved and are not expected to be evolving at the same rate, or under similar selective pressures as the genes responsible for virulence. Of the 10 housekeeping genes surveyed, only one gene, the molecular chaperone GroEL, was shown to experience significant positive selection (category 7-8), confirming the hypothesis that the virulence genes of *F. tularensis* are more likely to be undergoing concerted selective pressures. The other nine housekeeping genes analyzed showed no significant positive selection by either the HyPhy or TreeSAAP methods.

We initially expected that all of the housekeeping genes would not show significant selection. However, while GroEL is classified as a housekeeping gene, its function as a molecular chaperone includes responding to stress. In this way, GroEL may be functioning like a virulence gene, experiencing positive selection in much the same way as the other virulence genes.

3.5 Conclusions

Francisella tularensis virulence genes failed to recover statistically significant evidence of positive selection using d_N/d_S ratios. However, subsequent analyses of 11 arbitrarily-chosen virulence genes via TreeSAAP (152) identified 7 genes undergoing statistically significant positive selection. The biological functions of the identified mutations are cautiously inferred. However, since these genes appear to be under positive selection, they likely confer some evolutionary advantage leading to an increase in the overall fitness of the organism. The genes undergoing selection participate in a variety of functions, such as membrane transport, host defense evasion, stress response, intracellular survival, and even certain metabolic and biosynthetic pathways. Since the genes we identified as being under selection are virulence genes, it can be inferred that the altered genes confer an adaptive benefit and increase the ability of the organism to infect the host and/or evade host defenses. Although the number of pseudogenes present in the genome *F. tularensis* indicate that it is a genome in decay (137), our findings suggest that it is also undergoing an adaptive response to changes in its intracellular environment by way of positive selection on its virulence genes.

Chapter 4 Conclusions and Future Work

Francisella tularensis is a zoonotic pathogen with a worldwide distribution and is the causative agent of tularemia. The four subspecies of *F. tularensis* differ in their geographic distributions and in their virulence. The subspecies *tularensis* (Type A) is the most virulent and is found only in North America(20). The subspecies *holarctica* (Type B) has a milder symptoms than Type A, but is found typically in the northern hemisphere (20, 133). The subspecies *mediasiatica* is even less virulent than either Type A or Type B and is only found in the Central Asain republics of the former Soviet Union (133). Finally, the subspecies *novicida* is the least virulent and has been isolated around the world (14). Since *F. tularensis* is highly virulent and has the potential to be used as an agent of bioterror, it is important for both public health and defense agencies to rapidly diagnose and identify the various subspecies of this important pathogen.

This work describes the development of a real-time PCR assay for the characterization of the subspecies of *F. tularensis* commonly found in the United States. The biggest challenge in designing an assay to differentiate the subspecies of *F. tularensis* is that the genomes are so similar (25, 43). To overcome this, we used various regions of difference identified by SSH (72) and the comparative genomics tool, Mauve, (53) to identify genomic regions conducive to the creation of real time PCR assays. The resulting multiplex assay was designed using an economical scoring matrix (Table 3) and some shared genomic regions to identify four subtypes of *F. tularensis* with only three assays.

Another challenge in designing real-time PCR assays for *F. tularensis* is the discovery of *Francisella*-like endosymbionts (FLEs) of ticks. FLEs lack sufficient similarity to be classified as *F. tularensis*, but are similar enough to cross-react with many PCR assays (58). Future work

could involve testing existing PCR assays, including the one described in Chapter 2 of this work against FLEs and/or designing new assays or detection methods able to distinguish *F. tulrarensis* from FLEs.

The completed multiplex assay was about 98% effective at identifying the subtypes of *F. tularensis* compared to the MLVA. The assay did not cross react with other common laboratory strains or near neighbors tested. The multiplex assay demonstrated a limit of detection of approximately 1,136 organisms. While there are many varied assays and methods available to identify the subtypes of *F. tularensis*, we believe that the multiplex assay presented here will be another valuable tool for the identification and characterization of *F. tularensis* during outbreaks of disease or events of bioterroism.

Whole genome comparisons among the subspecies of *F. tularensis* have revealed that the pathogenic subspecies have evolved from the non-pathogenic subspecies (36). In addition, the many pseudogenes present in *F. tularensis*, which are largely metabolic genes, suggest a genome in decay furthering the pathoadaptation of *F. tularensis* to an intracellular lifestyle (10, 106, 111, 116, 137). We hypothesized that in addition to genome decay by the creation of pseudogenes, positive natural selection in virulence genes is also a driving force in the continued evolution of virulence in *F. tularensis*.

Analyses of 11 arbitrarily chosen virulent genes for positive selection via TreeSAAP identified 7 genes experiencing statistically significant positive selection. In contrast, only one of ten housekeeping genes showed positive selection. It can be inferred that those genes under positive selection confer an adaptive benefit to the organism. These results suggest that *F. tularensis* is undergoing an adaptive response, to its environment furthering pathoadaptation through positive selection on its virulence genes.

Currently there is no automated pipeline for the analysis of selection with TreeSAAP on a large or whole genome scale. For this reason, only 11 arbitrarily chosen virulence genes and 10 housekeeping genes were chosen for analysis. Future work could include the development of a large scale pipeline for analyzing large sets of genes for positive selection. After such a pipeline is developed, we recommend the reanalysis of the entire set of virulence genes, or even the whole genome of *F. tularensis*. We suspect that such an analysis would yield further insights into the evolution and pathoadaptation of this important pathogen.

Supplementary Tables

Table S 1 Assay results

Isolate #	Source	Kugler *et al.*, 2006 PCR	Multiplexed PCR	MLVA
70000907	Tick Bite, Human, Utah	A	N	A.I
1378	Human, New Mexico	A	A.I	A.I
80700075	Wound, Human, Utah	A	A.I	A.I
SCHU S4	Jean Celli, Rocky Mountain Laboratories, NAID, NH	A	A.I	A.I
79201237	Utah	A	A.I	A.I
2338	Human, New Mexico	A	A.II	A.II
1365	Human, New Mexico	A	A.II	A.II
3570	Human, New Mexico	A	A.II	A.II
831	Animal, New Mexico	A	A.II	A.II
3571	Human, New Mexico	A	A.II	A.II
311	Animal, New Mexico	A	A.II	A.II
2434	Human, New Mexico	A	A.II	A.II
597	Rabbit, New Mexico	A	A.II	A.II
19a	Animal, New Mexico	A	A.II	A.II
19b	Animal, New Mexico	A	A.II	A.II
1773a	Human, New Mexico	A	A.II	A.II
1773b	Human, New Mexico	A	A.II	A.II
1773c	Human, New Mexico	A	A.II	A.II
1801	Human, New Mexico	A	A.II	A.II
159	Rabbit, New Mexico	A	A.II	A.II
210	Rabbit, New Mexico	A	A.II	A.II
74	Rabbit, New Mexico	A	A.II	A.II
1419	Human, New Mexico	A	A.II	A.II
1643	Rabbit, New Mexico	A	A.II	A.II
9433	Cat, New Mexico	A	A.II	A.II
1024	Animal, New Mexico	A	A.II	A.II
1741	Human, New Mexico	A	A.II	A.II
70001275	Tissue, Human, Utah	A	A.II	A.II
1109	Animal, New Mexico	A	A.II	A.II
1277	Animal, New Mexico	A	A.II	A.II
6817	Animal, New Mexico	A	A.II	A.II
2250	Animal, New Mexico	A	A.II	A.II
2506	Animal, New Mexico	A	A.II	A.II
2109	N. mexicana, New Mexico	A	A.II	A.II
1965	Animal, New Mexico	A	A.II	A.II
2014	Animal, New Mexico	A	A.II	A.II
1465	Animal, New Mexico	A	A.II	A.II
70101546	Lesion, Human, Utah	A	A.II	A.II
1205	Animal, New Mexico	A	A.II	A.II

1385	Rabbit, New Mexico	A	A.II	A.II
AS 1284	Rodent, New Mexico	A	A.II	A.II
70102010	Axillary Lesion, Human, Utah	A	A.II	A.II
70102009	Back Wound, Human, Utah	A	A.II	A.II
804	Rabbit, New Mexico	A	A.II	A.II
1103a	Human, New Mexico	A	A.II	A.II
1275	Animal, New Mexico	A	A.II	A.II
6069	Animal, New Mexico	A	A.II	A.II
14149	New Mexico	A	A.II	A.II
1103b	Human, New Mexico	A	A.II	A.II
BA 4414	New Mexico	A	A.II	A.II
80402860	right Arm Abcess, Human, Utah	A	A.II	A.II
80402637	Skin, Human, Utah	A	A.II	A.II
80606987	Blood, Feline, Utah	A	A.II	A.II
80606984	Lymph node, Feline, Utah	A	A.II	A.II
80606985	Spleen, Feline, Utah	A	A.II	A.II
80700103	Hand drainage, Human, Utah	A	A.II	A.II
80700051	Abscess, Human, Utah	A	A.II	A.II
80800086	Neck lesion aspirate, Human, Utah	A	A.II	A.II
80800071	Lymph node, Feline, Utah	A	A.II	A.II
5906	Animal, New Mexico	A	A.II	A.II
1513	Dog, New Mexico	A	A.II	A.II
6733	New Mexico	A	A.II	A.II
6734	New Mexico	A	A.II	A.II
0018 7/88	Utah	A	A.II	A.II
08902000 (11/16)	Utah	A	A.II	A.II
79000875	Utah	A	A.II	A.II
79101487	Utah	A	A.II	A.II
79101488	Utah	A	A.II	A.II
79101615	Utah	A	A.II	A.II
79102173	Utah	A	A.II	A.II
79301108	Utah	A	A.II	A.II
79400960	Utah	A	A.II	A.II
880178-1	Utah	A	A.II	A.II
880178-2	Utah	A	A.II	A.II
880321	Utah	A	A.II	A.II
93703	Utah	A	A.II	A.II
LVS	Sheep, Australia, Dugway Proving Ground	B	B	B
1518	Human, New Mexico	B	B	B
79901959	3rd finger wound, Human, Utah	B	B	B
3459	Animal, New Mexico	A	B	B
70001144	Eye Drainage, Human, Utah	B	B	B
80501131	Pleural fluid, Human, Utah	B	B	B

U112: CG69	Sheep, Australia, University of Victoria, Fran Nano	N	N	N
U112: CG57	University of Victoria, Fran Nano	N	N	N
U112: GB2/pGB40 #7	University of Victoria, Fran Nano	N	N	N
U112: SC119	University of Victoria, Fran Nano	N	N	N
79201152	Utah	A	A.II	NM
1815	Human, New Mexico	A	A.II	NM
305	Animal, New Mexico	A	A.II	NM
75	Rabbit, New Mexico	A	A.II	NM
736	Rabbit, New Mexico	A	A.II	NM
70102163	Blood, Human, Utah	A	A.II	NM
AS 2058	Lapine, New Mexico	A	A.II	NM
1276	Animal, New Mexico	A	A.II	NM
AS 816	Feline, New Mexico	A	A.II	NM
AS 713	Lapine, New Mexico	A	A.II	NM
80502541	Groin, Human, Utah	A	A.II	NM
80606986	Liver, Feline, Utah	A	A.II	NM
80700132	Bite wound, Human, Utah	B	B	NM
80700107	Face, Human, Utah	A	A.II	NM
80700069	Lesion, Human, Utah	A	A.I	NM
80800087	Wound, Human, Utah	A	A.II	NM
80800082	Spleen, Feline, Utah	A	A.II	NM
80800111	Finger wound, Human, Utah	A	A.II	NM
AS 200801417	New Mexico	A	A.II	NM
AS 200801507	New Mexico	A	A.II	NM
U112: 1L	University of Victoria, Fran Nano	N	N	NM
U112: 68-11	University of Victoria, Fran Nano	N	N	NM
U112: CG116	University of Victoria, Fran Nano	N	N	NM
U112: CG62	University of Victoria, Fran Nano	N	N	NM
U112: GB2	University of Victoria, Fran Nano	N	N	NM
U112: GB5	University of Victoria, Fran Nano	N	N	NM
U112: KM14	University of Victoria, Fran Nano	N	N	NM
U112: KM14S	University of Victoria, Fran Nano	N	N	NM
1574	Utah	A	A.II	NM
7/27	Utah	A	A.II	NM
79001541 4/45	Utah	A	A.I	NM
U112	Utah, Dugway	N	N	NM
6168	Human, University of Victoria, Fran Nano	N	B	N

A- *tularensis* Type A, A.I - *tularensis* Type A.I, A.II - *tularensis* Type A.II, B - *holarctica*, N - *Novicida*, NM - no match.

Supplementary Figures

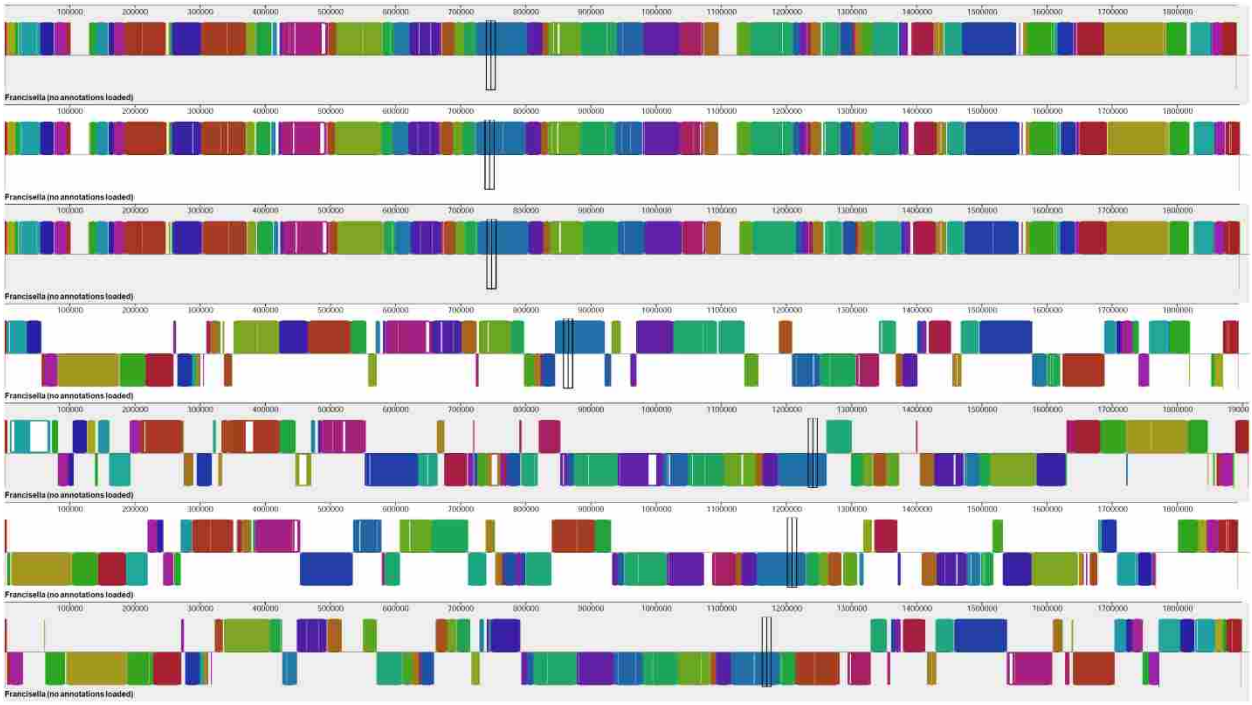


Figure S 1 Mauve alignment of *Francisella* genomes
The complete genomes of 7 *Francisella* isolates were obtained from GenBank and aligned using Mauve. The coloured blocks represent homologous sequences among the different strains. Blocks above the center line represent a 'forward' reading frame relative to the reference strain (lane 1) while blocks below the center line represent a 'reverse' reading frame. White spaces within the colored blocks represent areas of non-homology between strains. The search for unique genomic markers was focused within these regions of non-homology. (1) *holarctica* FTNF002-00, (2) *holarctica* OSU18, (3) *holarctica* LVS, (4) *mediasiatica* FSC147, (5) *novicida* U112, (6) *tularensis* Schu S4, (7) *tularensis* WY96-3418

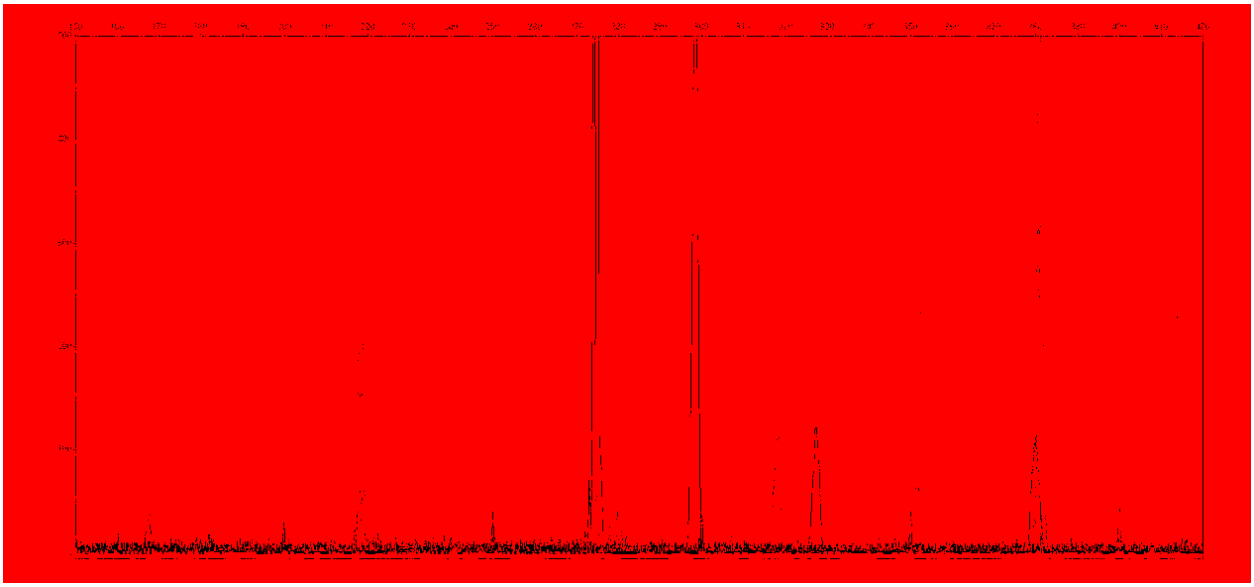
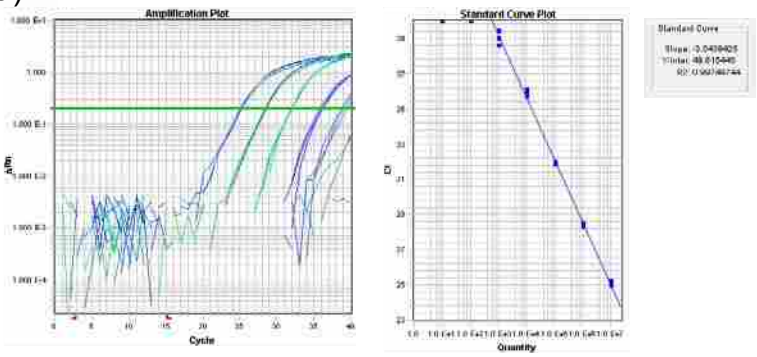
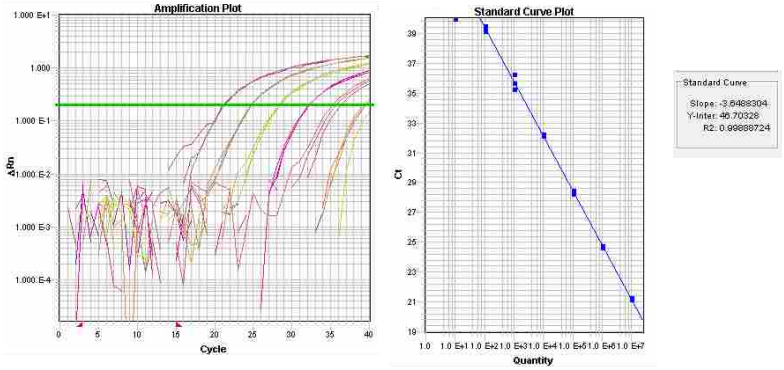


Figure S 2 Representative MLVA analysis
The peaks represent the detection of amplified PCR products from all 3 mixes using labeled primers. These peaks are compared to a database of known subspecies for identification.

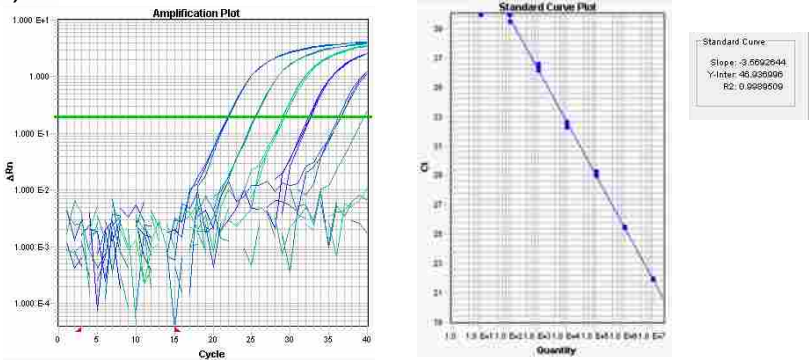
(a)


Amplification Plot
ΔRn
Cycle
Standard Curve Plot
Ct
Quantity
Standard Curve

(b)


Amplification Plot
ΔRn
Cycle
Standard Curve Plot
Ct
Quantity
Standard Curve
Slope: -3.6488304
Y-Inter: 46.70328
R2: 0.9988724

(c)


Amplification Plot
ΔRn
Cycle
Standard Curve Plot
Ct
Quantity
Standard Curve
Slope: -3.5692644
Y-Inter: 46.936996
R2: 0.9989509

(d)

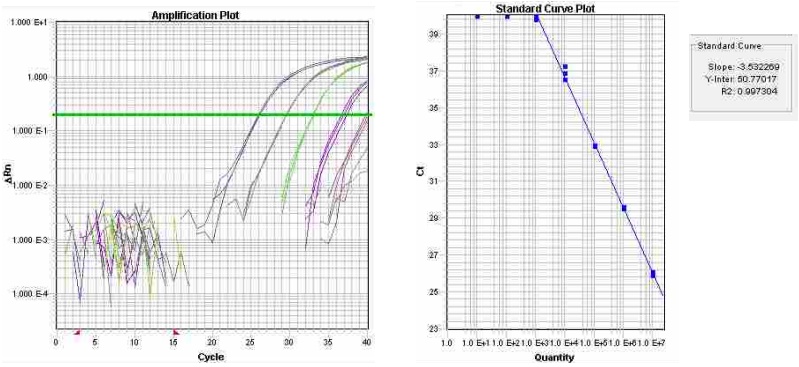
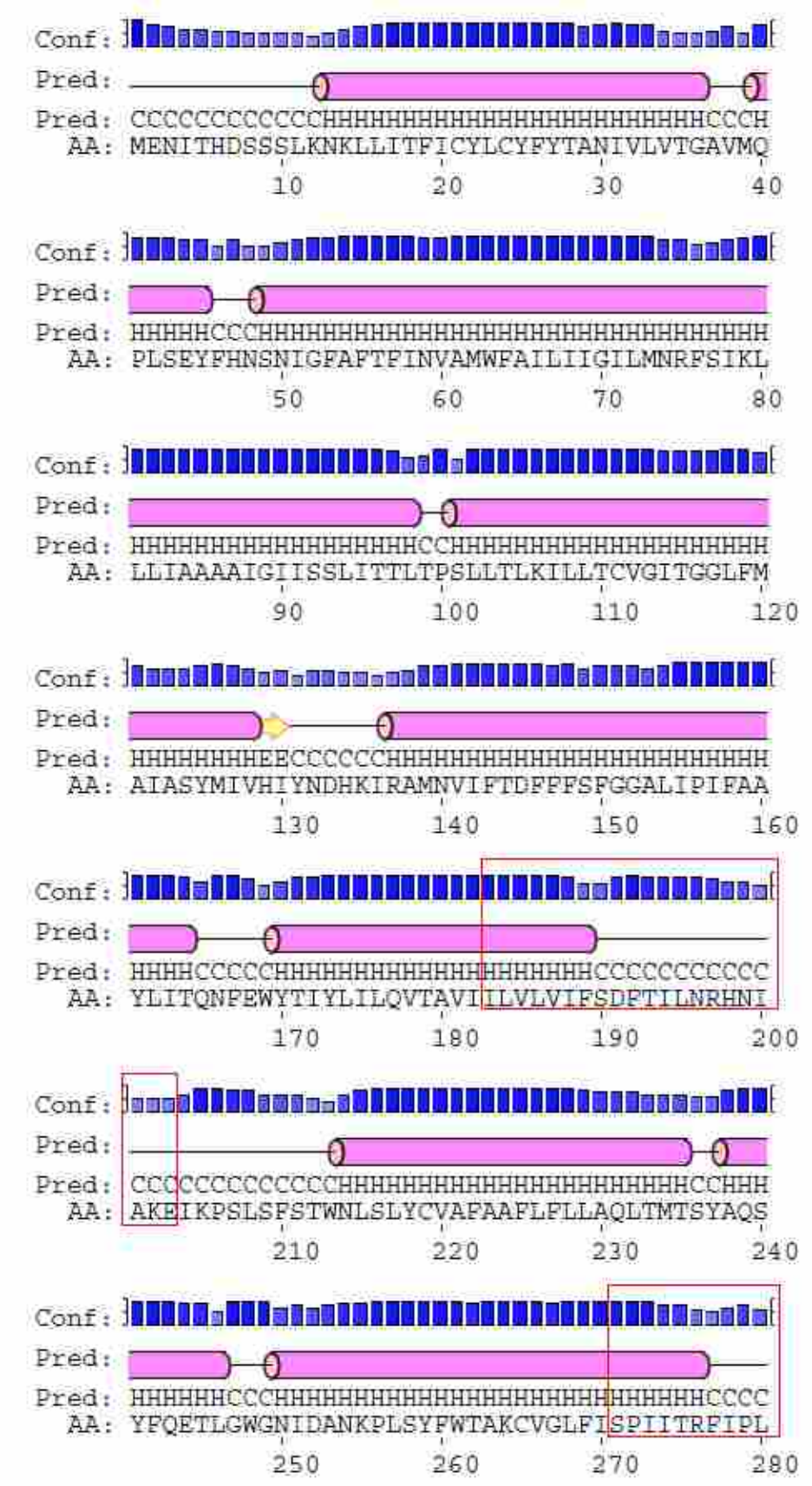

Amplification Plot
ΔRn
Cycle
Standard Curve Plot
Ct
Quantity
Standard Curve
Slope: -3.532269
Y-Inter: 50.77017
R2: 0.997304

Figure S 3 Sensitivities of singleplex assays
Tenfold serial dilutions of *F. tularensis* chromosomal DNA were tested to determine individual assay sensitivities. (a) Type A.I, (b) Type A.II, (c) Type B (d) *novicida*. No Template Control (NTC) samples were negative for each sample processed. Calculated PCR efficiencies: (a) 91.5 %, (b) 88.0 %, (c) 90.6%, (d) 91. 9%.


Conf:
Pred:
Pred: CCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCH
AA: MENITHDSSSLKNKLLITFICYLCYFYTANIVLVTGAVMQ
10 20 30 40
Conf:
Pred:
Pred: HHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: PLSEYFHNSNIGFAFTFINVAMWFAILIIGILMNRFSIKL
50 60 70 80
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHHHHHHHH
AA: LLIAAAAIGIISSLITTLTPSLLTLKILLTCVGITGGLFM
90 100 110 120
Conf:
Pred:
Pred: HHHHHHHHHEECCCCCCHHHHHHHHHHHHHHHHHHHHHH
AA: AIASYMIVHIYNDHKIRAMNVIFTDFFFSFGGALIPIFAA
130 140 150 160
Conf:
Pred:
Pred: HHHHCCCCCHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCC
AA: YLITQNPEWYTIYLILQVTAVIILVLVIFSDPTILNRHNI
170 180 190 200
Conf:
Pred:
Pred: CCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHCCHHH
AA: AKEIKPSLSFSTWNLSLYCVAFAAFLFLLAQLTMTSYAQS
210 220 230 240
Conf:
Pred:
Pred: HHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHCCCC
AA: YFQETLGWGNIDANKPLSYFWTAKCVGLFISPIITRFIPL
250 260 270 280

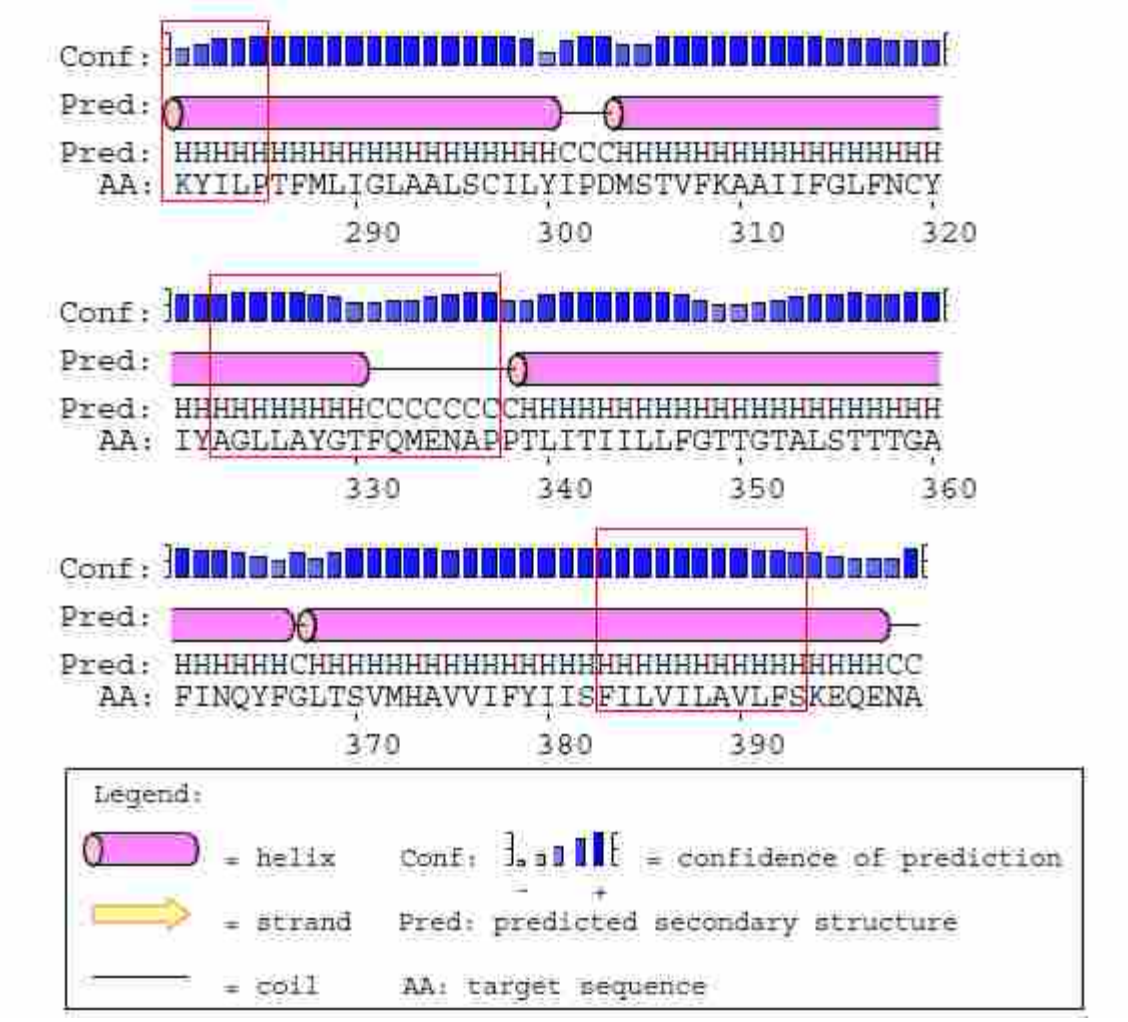
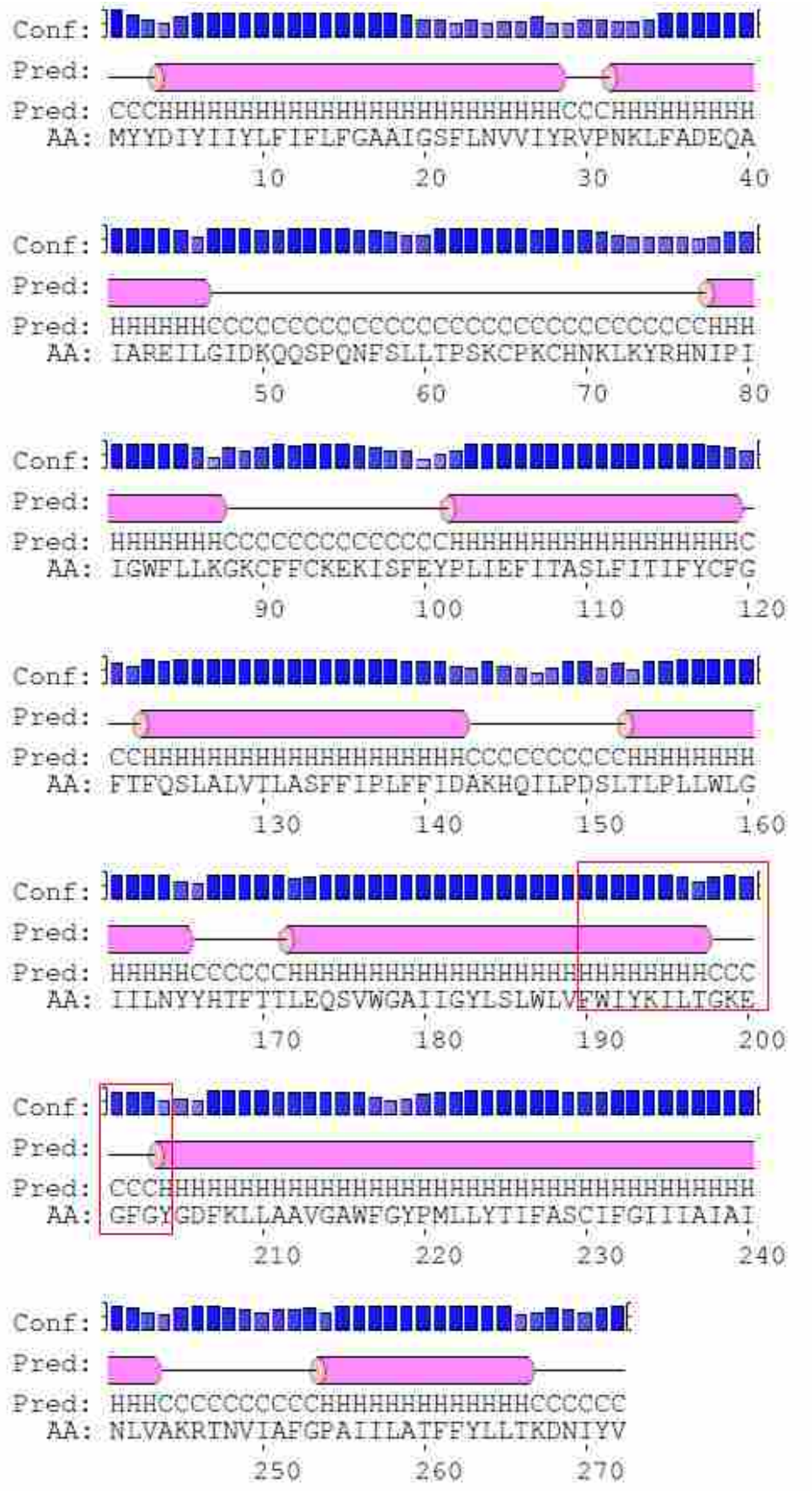


Figure S 4 Predicted secondary structure of FTL_1134.
Predicted secondary structure of FTL_1134 using the PSIPRED server at http://bioinf.cs.ucl.ac.uk/psipred/. The red boxes at codons 183-203, 271-285, 232-337 and 383-394 indicate the areas under selection as determined by TreeSAAP.


Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHH
AA: MYYDIYIIYLFIFLFGAAIGSFLNVVIYRVPNKLFADEQA
10 20 30 40
Conf:
Pred:
Pred: HHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHH
AA: IAREILGIDKQQSPQNFSLLTPSKCPKCHNKLKYRHNIPI
50 60 70 80
Conf:
Pred:
Pred: HHHHHHCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHC
AA: IGWFLLKGKCFFCKEKISFEYPLIEFITASLFITIFYCFG
90 100 110 120
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHH
AA: FTFQSLALVTLASFFIPLFFIDAKHQILPDSLTLPLLWLG
130 140 150 160
Conf:
Pred:
Pred: HHHHHCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHCCC
AA: IILNYYHTFTTLEQSVWGAIIGYLSLWLVFWIYKILIGKE
170 180 190 200
Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: GFGYGDFKLLAAVGAWFGYPMLLYTIFASCIFGIIIAIAI
210 220 230 240
Conf:
Pred:
Pred: HHHCCCCCCCCCCHHHHHHHHHHHHHCCCCCC
AA: NLVAKRTNVIAFGPAIILATFFYLLIKDNIYV
250 260 270

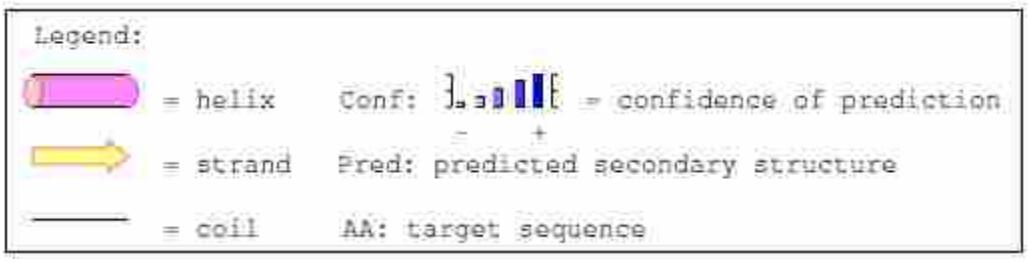
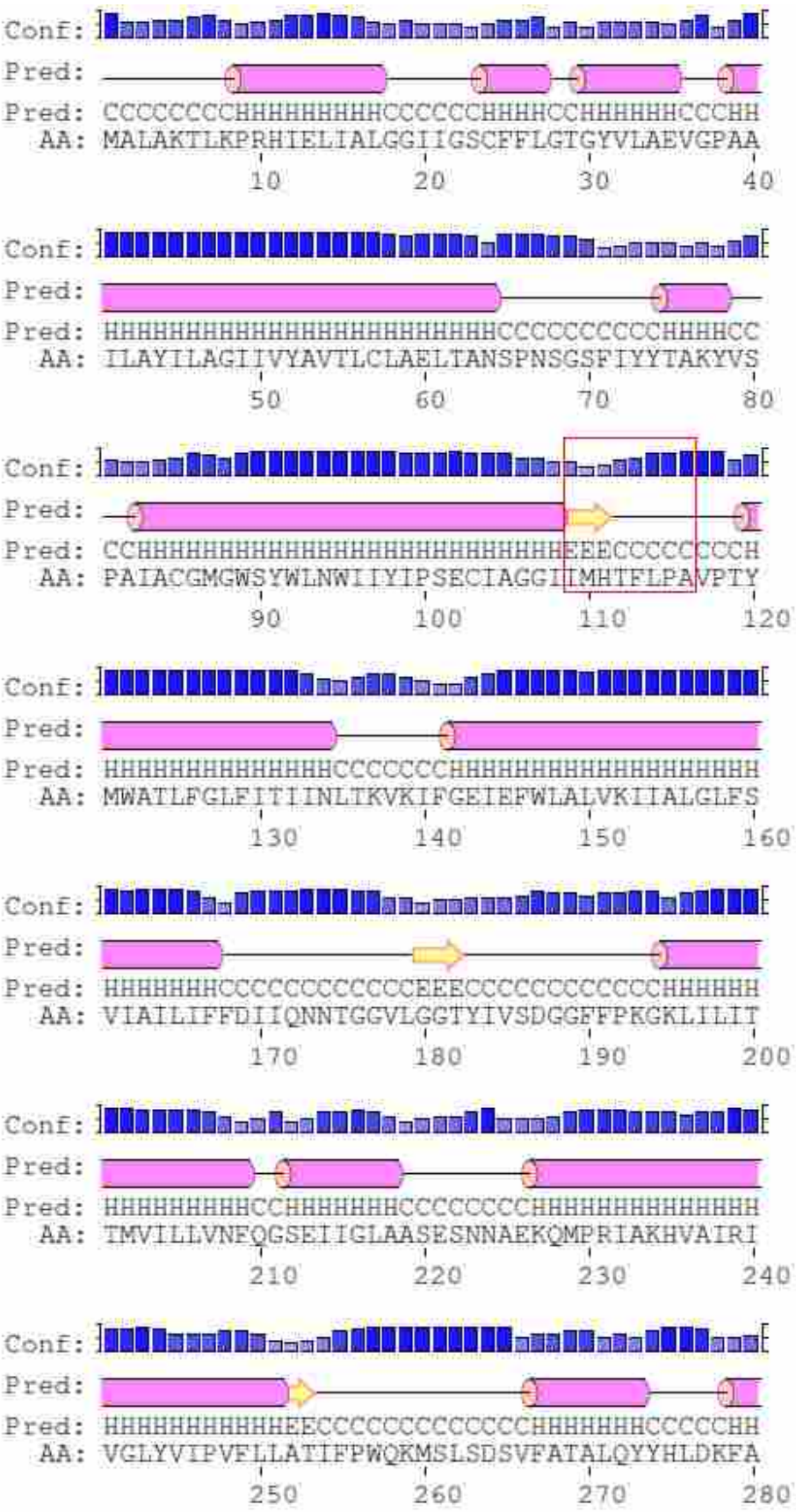


Figure S 5 Predicted secondary structure of FTT_0683 (*pilD*).
Predicted secondary structure of FTT_0683 (*pilD*) using PSIPRED server at http://bioninf.cs.ucl.ac.uk/psipred/. The red boxes at codons 190-204 indicate the areas under selection as determined by TreeSAAP.


Conf:
Pred:
Pred: CCCCCCCCCHHHHHHHHHCCCCCCHHHHCCHHHHHHCCCHH
AA: MALAKTLKPRHIELIALGGIIGSCFFLGTGYVLAEVGPAA
10 20 30 40
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHCC
AA: ILAYILAGIIVYAVTLCLAELIANSPNSGSFIYYTAKYVS
50 60 70 80
Conf:
Pred:
Pred: CCHHHHHHHHHHHHHHHHHHHHHHHHHHHEEECCCCCCCCH
AA: PAIACGMGWSYWLNWIIYIPSECIAGGIIMHTFLPAVPTY
90 100 110 120
Conf:
Pred:
Pred: HHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHHHHHHHHHH
AA: MWATLFGLFITIINLIKVKIFGEIEFWLALVKIIALGLFS
130 140 150 160
Conf:
Pred:
Pred: HHHHHHHCCCCCCCCCCCCCEEECCCCCCCCCCCCCHHHHHH
AA: VIAILIFFDIIQNNTGGVLGGTYIVSDGGFFPKGKLILIT
170 180 190 200
Conf:
Pred:
Pred: HHHHHHHHHCCHHHHHHHCCCCCCCCHHHHHHHHHHHHHH
AA: TMVILLVNFQGSEIIGLAASESNNAEKQMPRIAKHVAIRI
210 220 230 240
Conf:
Pred:
Pred: HHHHHHHHHHEECCCCCCCCCCCCCCHHHHHHHCCCCCHH
AA: VGLYVIPVFLLATIFPWQKMSLSDSVFATALQYYHLDKFA
250 260 270 280

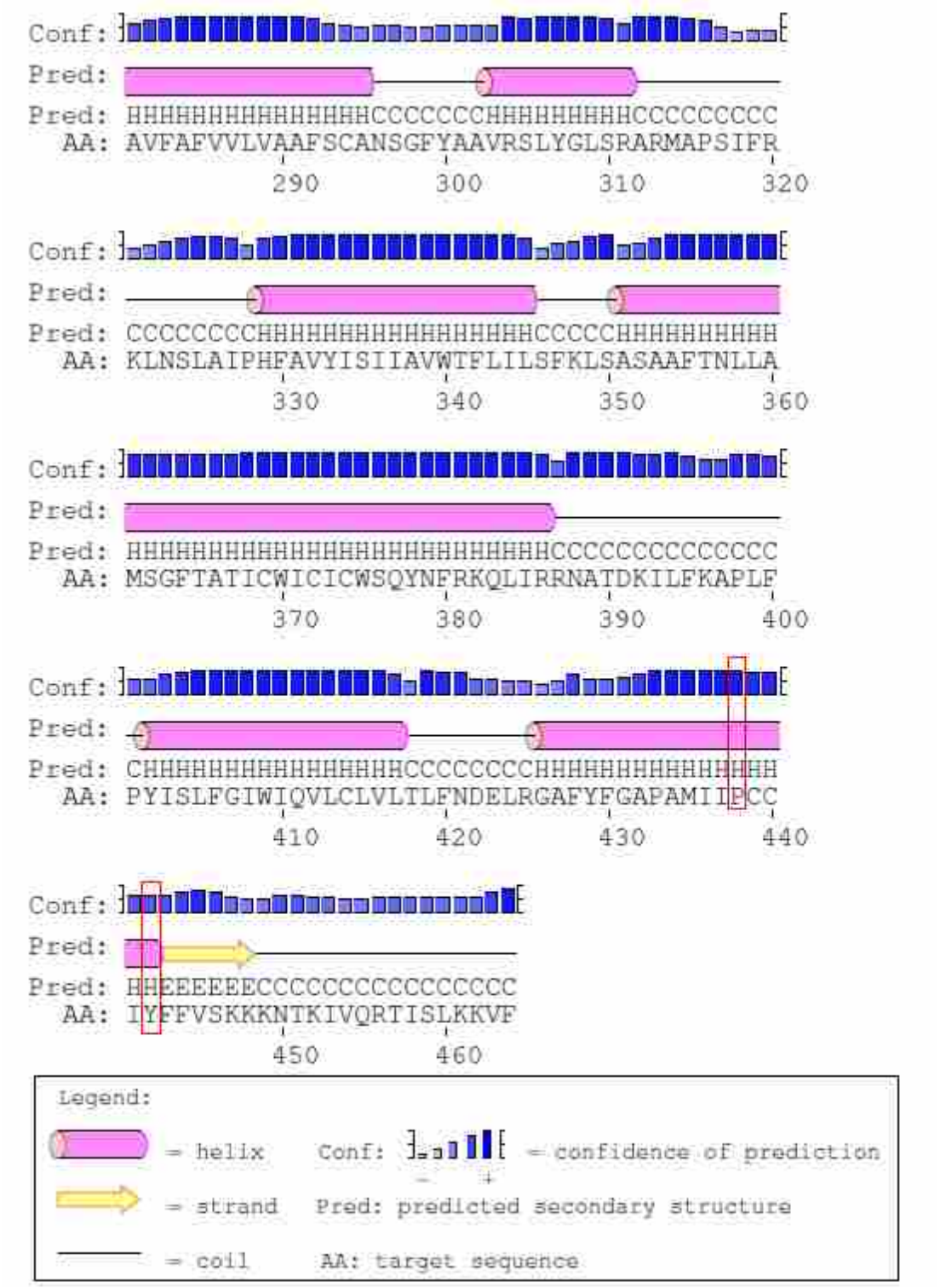
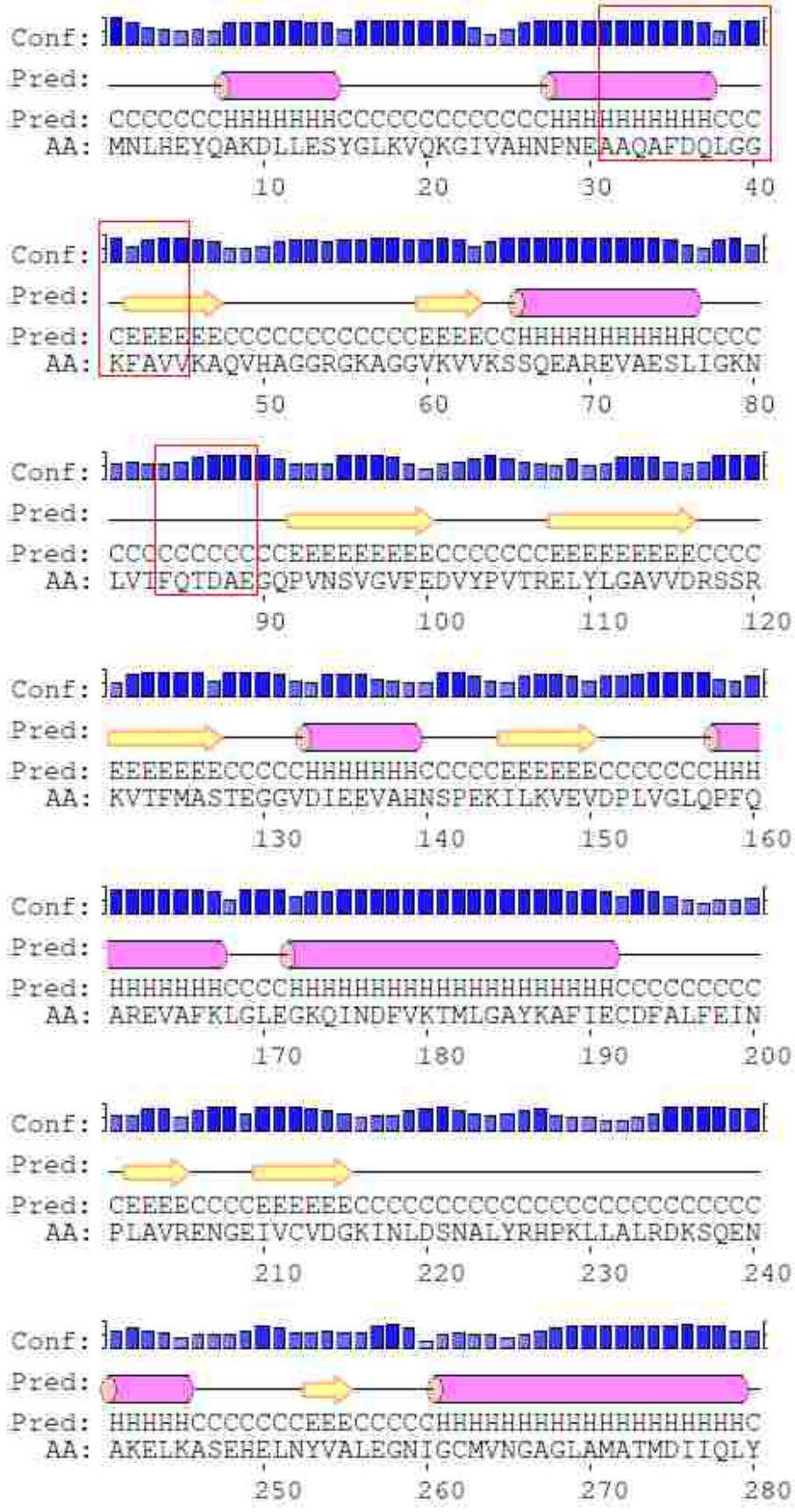


Figure S 6 Predicted secondary structure of FTT_0881 (*rocE*).
Predicted secondary structure of FTT_0881 (*rocE*) using the PSIPRED server at http://bioninf.cs.ucl.ac.uk/psipred/. The red boxes at codons 109-116, 438 and 442 indicate the areas under selection as determined by TreeSAAP.


Conf:
Pred:
Pred: CCCCCCCHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHCCC
AA: MNLHEYQAKDLLESYGLKVQKGIVAHNPNEAAQAFDQLGG
10 20 30 40
Conf:
Pred:
Pred: CEEEEEECCCCCCCCCCCCEEEECCHHHHHHHHHHHCCCC
AA: KFAVVKAQVHAGGRGKAGGVKVVKSSQEAREVAESLIGKN
50 60 70 80
Conf:
Pred:
Pred: CCCCCCCCCCCEEEEEEEEECCCCCCCCEEEEEEEECCCC
AA: LVTFQTDAEGQPVNSVGVFEDVYPVTRELYLGAVVDRSSR
90 100 110 120
Conf:
Pred:
Pred: EEEEEEECCCCCHHHHHHHCCCCCEEEEEECCCCCCCHHH
AA: KVTFMASTEGGVDIEEVAHNSPEKILKVEVDPLVGLQPFQ
130 140 150 160
Conf:
Pred:
Pred: HHHHHHHCCCCHHHHHHHHHHHHHHHHHHHCCCCCCCCC
AA: AREVAFKLGLEGKQINDFVKTMLGAYKAFIECDFALFEIN
170 180 190 200
Conf:
Pred:
Pred: CEEEECCCCEEEEEECCCCCCCCCCCCCCCCCCCCCCCC
AA: PLAVRENGEIVCVDGKINLDSNALYRHPKLLALRDKSQEN
210 220 230 240
Conf:
Pred:
Pred: HHHHHCCCCCCCEEECCCCCHHHHHHHHHHHHHHHHHHC
AA: AKELKASEHELNYVALEGNIGCMVNGAGLAMATMDIIQLY
250 260 270 280

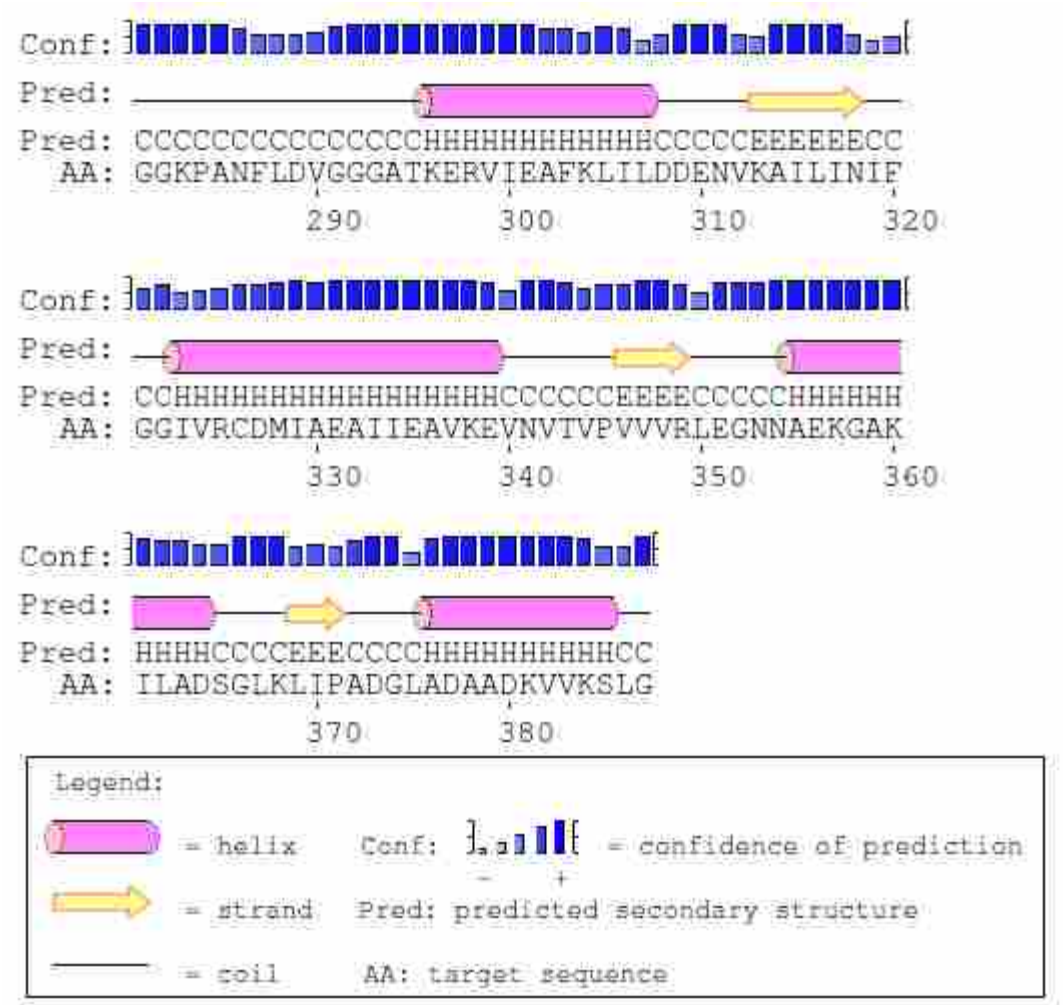


Figure S 7 Predicted secondary structure of FTT_0504 (*sucC*).
Predicted secondary structure of FTT_0504 (*sucC*) using the PSIPRED server at http://bioninf.cs.ucl.ac.uk/psipred/. The red boxes at codons 31-45 and 84-89 indicate the areas under selection as determined by TreeSAAP.

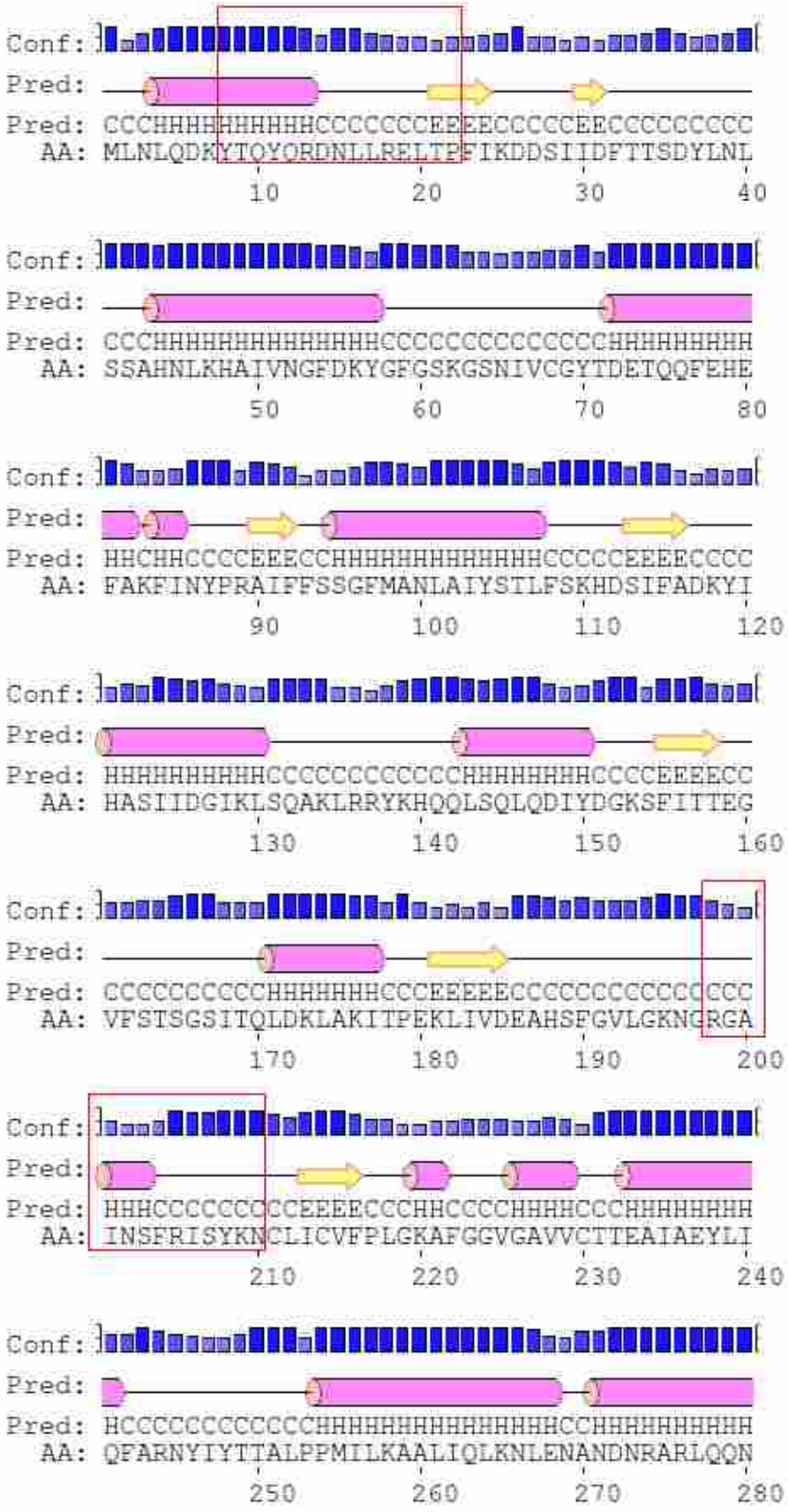

Conf:
Pred:
Pred: CCCHHHHHHHHHHCCCCCCCCEEEECCCCCCEECCCCCCCCC
AA: MLNLQDKYTQYQRDNLLRELTPFIKDDSIIDFTTSDYLNL
10
20
30
40
Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHHCCCCCCCCCCCCCCCHHHHHHHHH
AA: SSAHNLKHAIVNGFDKYGFGSKGSNIVCGYTDETQQFEHE
50
60
70
80
Conf:
Pred:
Pred: HHCHHCCCCEEECCHHHHHHHHHHHHHHCCCCCEEEECCCC
AA: FAKFINYPRAIFFSSGFMANLAIYSTLFSKHDSIFADKYI
90
100
110
120
Conf:
Pred:
Pred: HHHHHHHHHHCCCCCCCCCCCCCHHHHHHHHCCCCEEEECC
AA: HASIIDGIKLSQAKLRRYKHQQLSQLQDIYDGKSFITTEG
130
140
150
160
Conf:
Pred:
Pred: CCCCCCCCCCCHHHHHHHCCCEEEEECCCCCCCCCCCCCCCC
AA: VFSTSGSITQLDKLAKITPEKLIVDEAHSFGVLGKNGRGA
170
180
190
200
Conf:
Pred:
Pred: HHHCCCCCCCCCEEEEECCCHHCCCCHHHHCCCHHHHHHHH
AA: INSFRISYKNCLICVFPLGKAFGGVGAVVCTTEAIAEYLI
210
220
230
240
Conf:
Pred:
Pred: HCCCCCCCCCCCCHHHHHHHHHHHHHHHCCHHHHHHHHHH
AA: QFARNYIYTTALPPMILKAALIQLKNLENANDNRARLQQN
250
260
270
280

Figure S 8 Predicted secondary structure of FTT_0936c (*bioF*).
Predicted secondary structure of FTT_0936c (*bioF*) using the PSIPRED server at http://bioinf.cs.ucl.uk/psipred/. The red boxes at codons 8-22 and 198-210 indicate the areas under selection as determined by TreeSAAP.

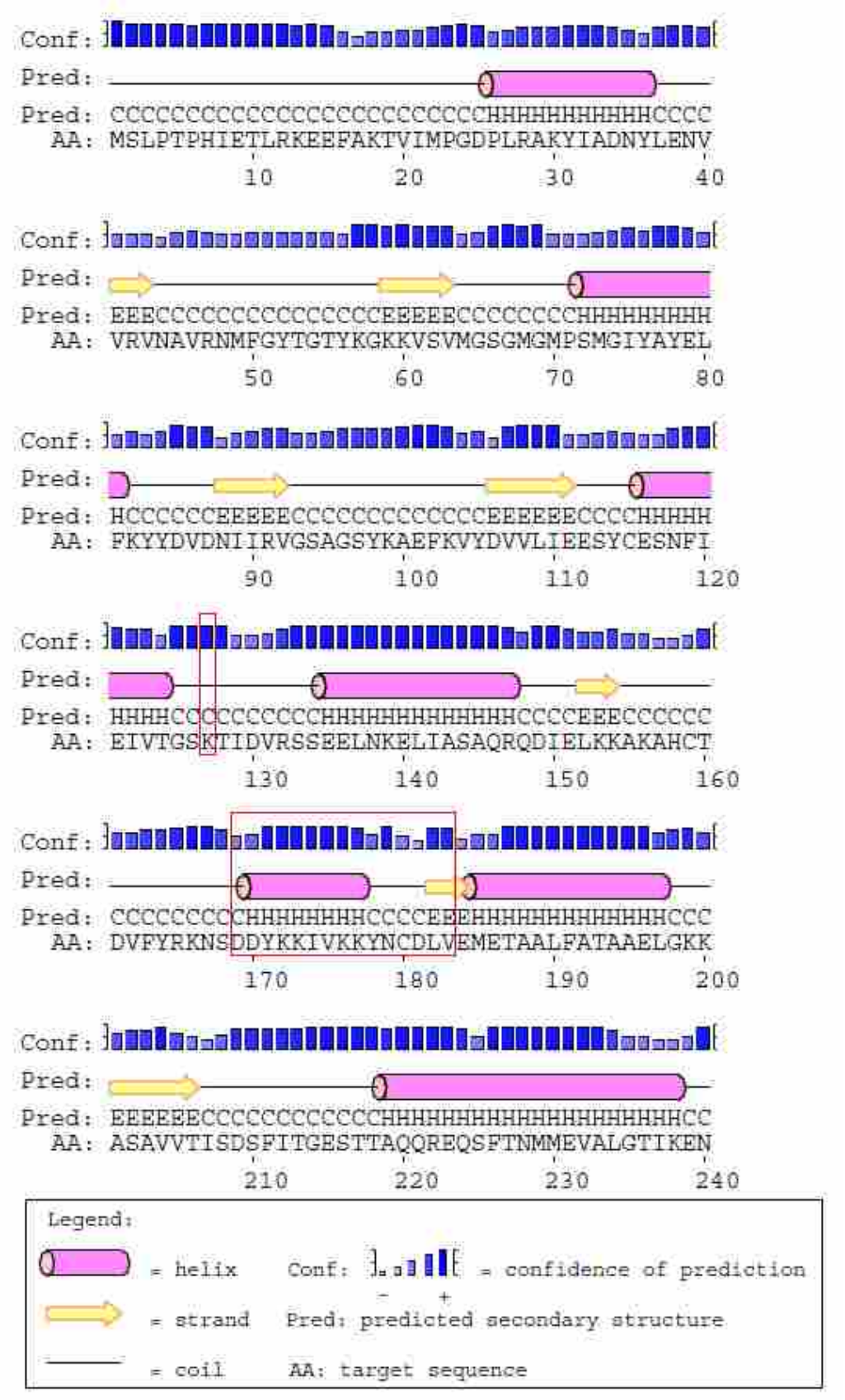
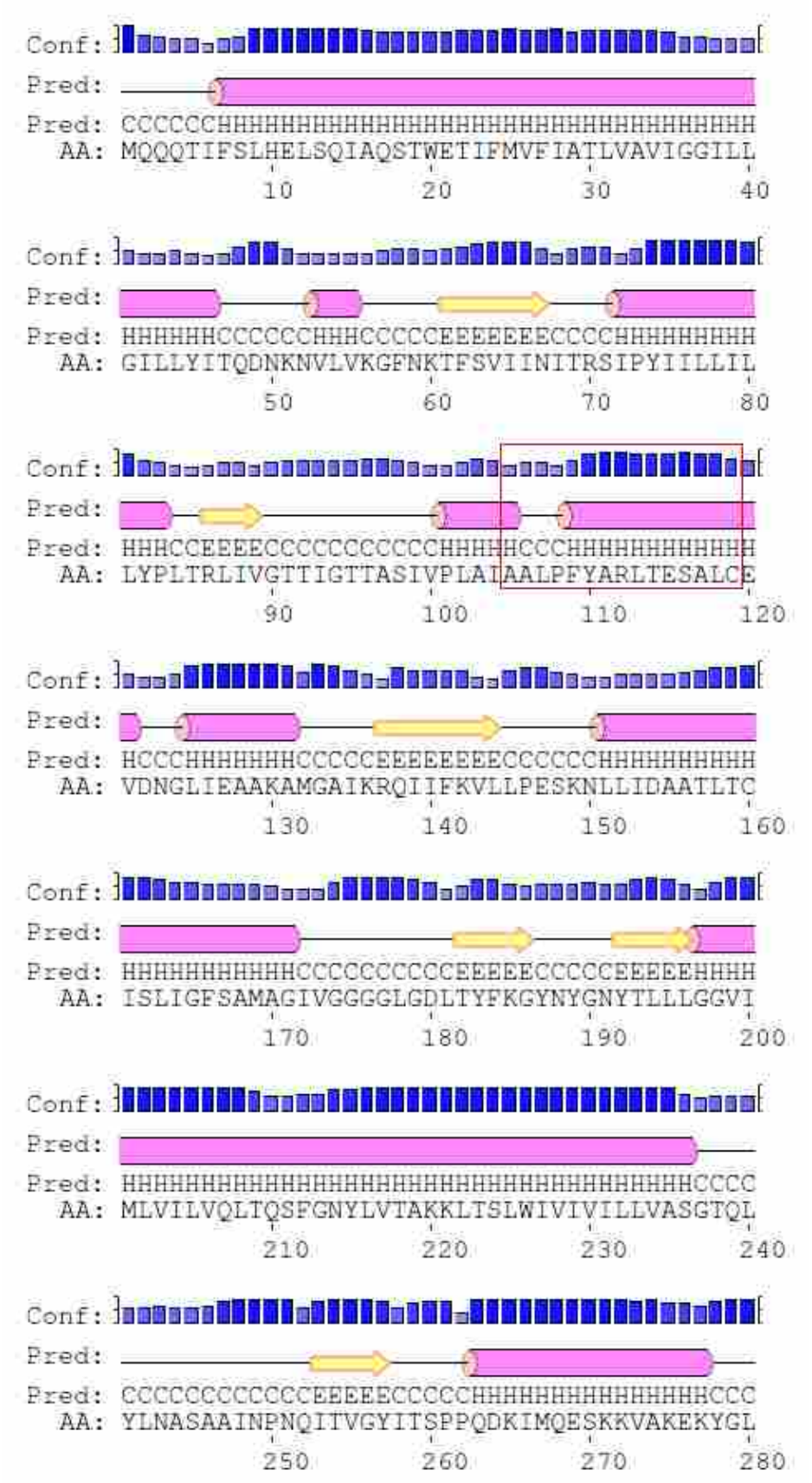

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCC
AA: MSLPTPHIETLRKEEFAKTVIMPGDPLRAKYIADNYLENV
10 20 30 40
Conf:
Pred:
Pred: EEECCCCCCCCCCCCCCCCEEEEECCCCCCCCHHHHHHHHH
AA: VRVNAVRNMFGYTGTYKGKKVSVMGSGMGMPSMGIYAYEL
50 60 70 80
Conf:
Pred:
Pred: HCCCCCCEEEEECCCCCCCCCCCCCCEEEEECCCCHHHHH
AA: FKYYDVDNIIRVGSAGSYKAEFKVYDVVLIEESYCESNFI
90 100 110 120
Conf:
Pred:
Pred: HHHHCCCCCCCCCCHHHHHHHHHHHHHHCCCCEEECCCCCC
AA: EIVTGSKTIDVRSSEELNKELIASAQRQDIELKKAKAHCT
130 140 150 160
Conf:
Pred:
Pred: CCCCCCCCCHHHHHHHHCCCCEEEHHHHHHHHHHHHHHCCC
AA: DVFYRKNSDDYKKIVKKYNCDLVEMETAALFATAAELGKK
170 180 190 200
Conf:
Pred:
Pred: EEEEEECCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHCC
AA: ASAVVTISDSFITGESTTAQQREQSFTNMMEVALGTIKEN
210 220 230 240
Legend:
= helix
= strand
= coil
Conf: - + = confidence of prediction
Pred: predicted secondary structure
AA: target sequence

Figure S 9 Predicted secondary strucutre of FTT_0766 (*deoD*).
Predicted secondary strucutre of FTT_0766 (*deoD*) using the PSIPRED server at http://bioinf.cs.ucl.ac.uk/psipred/. The red boxes at codons 127 and 169-183 indicate the areas under selection as determined by TreeSAAP.


Conf:
Pred:
Pred: CCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: MQQQTIFSLHELSQIAQSTWETIFMVFIATLVAVIGGILL
10 20 30 40
Conf:
Pred:
Pred: HHHHHHCCCCCCHHHCCCCCEEEEEEECCCCHHHHHHHHH
AA: GILLYITQDNKNVLVKGFNKTFSVIINITRSIPYIILLIL
50 60 70 80
Conf:
Pred:
Pred: HHHCCEEEECCCCCCCCCCCHHHHHCCCHHHHHHHHHHHH
AA: LYPLTRLIVGTTIGTTASIVPLAIAALPFYARLTESALCE
90 100 110 120
Conf:
Pred:
Pred: HCCCHHHHHHHCCCCCEEEEEEEECCCCCCHHHHHHHHHH
AA: VDNGLIEAAKAMGAIKRQIIFKVLLPESKNLLIDAATLTC
130 140 150 160
Conf:
Pred:
Pred: HHHHHHHHHHHCCCCCCCCCCEEEEEECCCCCCEEEEEHHHH
AA: ISLIGFSAMAGIVGGGGLGDLTYFKGYNYGNYTLLLGGVI
170 180 190 200
Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCC
AA: MLVILVQLTQSFGNYLVTAKKLTSLWIVIVILLVASGTQL
210 220 230 240
Conf:
Pred:
Pred: CCCCCCCCCCCCCEEEEECCCCCHHHHHHHHHHHHHHHCCC
AA: YLNASAAINPNQITVGYITSPPQDKIMQESKKVAKEKYGL
250 260 270 280

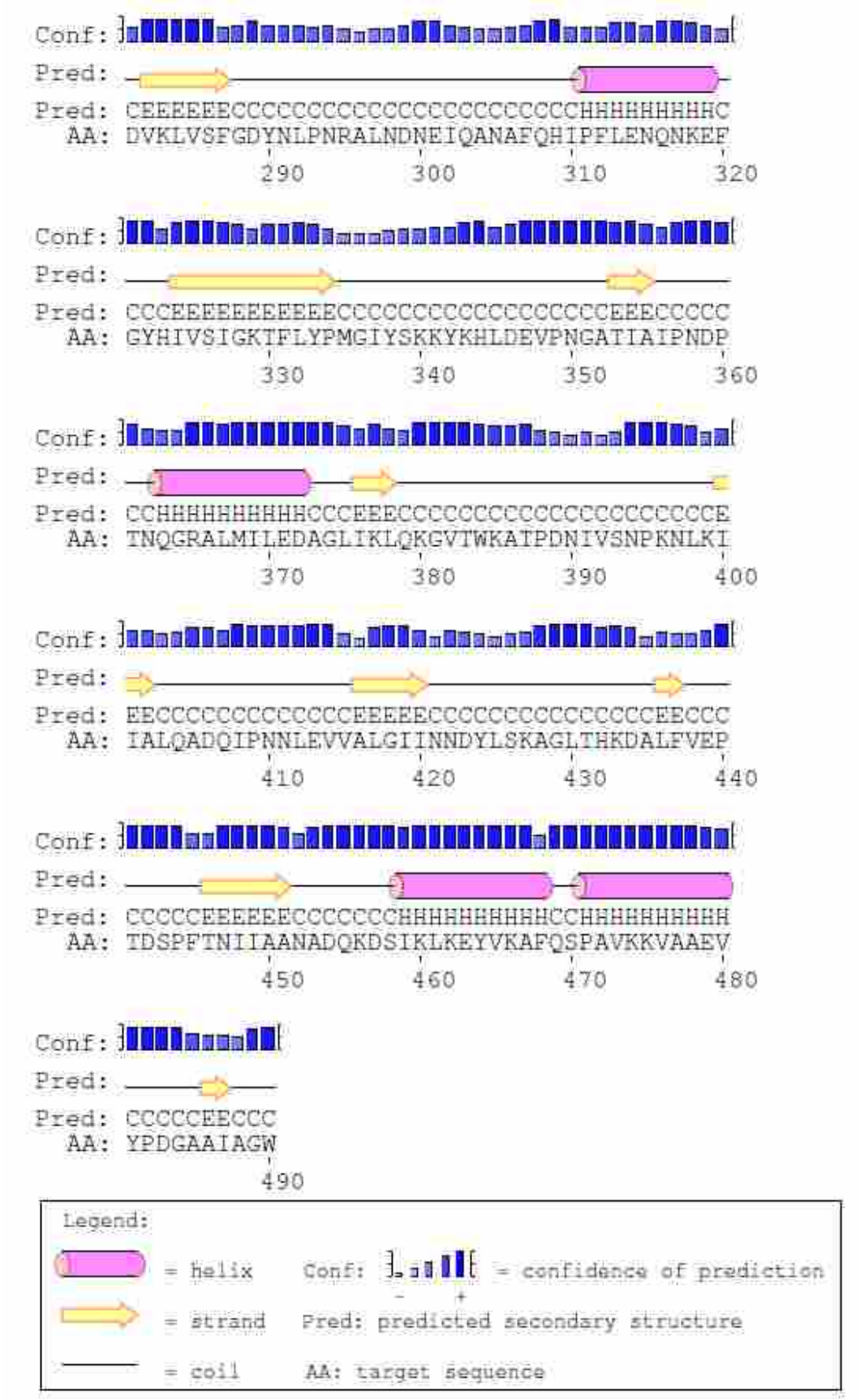

Conf:
Pred:
Pred: CEEEEECCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHC
AA: DVKLVSFGDYNLPNRALNDNEIQANAFQHIPFLENQNKEF
290
300
310
320
Conf:
Pred:
Pred: CCCEEEEEEEEEECCCCCCCCCCCCCCCCCCCCEEECCCCC
AA: GYHIVSIGKTFLYPMGIYSKKYKHLDEVPNGATIAIPNDP
330
340
350
360
Conf:
Pred:
Pred: CCHHHHHHHHHHCCCEEECCCCCCCCCCCCCCCCCCCCCE
AA: TNQGRALMILEDAGLIKLQKGVTWKATPDNIVSNPKNLKI
370
380
390
400
Conf:
Pred:
Pred: EECCCCCCCCCCCCCEEEEECCCCCCCCCCCCCCCEECCC
AA: IALQADQIPNNLEVVALGIINNDYLSKAGLTHKDALFVEP
410
420
430
440
Conf:
Pred:
Pred: CCCCCEEEEEECCCCCCCHHHHHHHHHHCCHHHHHHHHHH
AA: TDSPFTNIIAANADQKDSIKLKEYVKAFQSPAVKKVAAEV
450
460
470
480
Conf:
Pred:
Pred: CCCCCEECCC
AA: YPDGAAIAGW
490
Legend:
= helix
= strand
= coil
Conf:
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

Figure S 10 Predicted secondary structure of FTT_1125 (*metQ*).
Predicted secondary structure of FTT_1125 (*metQ*) using the PSIPRED server at http://bioninf.cs.ucl.uk/psipred/. The red boxes at codons 105-119 indicate the areas under selection as determined by TreeSAAP.

REFERENCES

1. **Gunnell MK, Adams BJ, Robison RA.** 2016. The genetic diversity and evolution of *Francisella tularensis* with comments on detection by PCR. Current Issues in Molecular Biology **18:**79-92.
2. **Gunnell MK, Lovelace CD, Satterfield BA, Moore EA, O'Neill KL, Robison RA.** 2012. A multiplex real-time PCR assay for the detection and differentiation of *Francisella tularensis* subspecies. Journal of Medical Microbiology **61:**1525-1531.
3. **McCoy G, Chapin C.** 1912. Further observations on a plaguelike disease of rodents with a preliminary note on the causative agent *Bacterium tularense*. J Infect Dis **10:**61-72.
4. **Wherry WB, Lamb BH.** 1914. Infection of man with *Bacterium tularense*. J Infect Dis **15:**331-340.
5. **Akimana C, Abu Kwaik Y.** 2011. *Francisella*-arthropod vector interaction and its role in patho-adaptation to infect mammals. Frontiers in Microbiology **2:** DOI: 10.3389/fmicb.2011.00034.
6. **McCrumb FR.** 1961. Aerosol infection of man with *Pasteurella tularensis*. Bacteriological Reviews **25:**262-267.
7. **Dennis DT, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, Fine AD, Friedlander AM, Hauer J, Layton M, Lillibridge SR, McDade JE, Osterholm MT, O'Toole T, Parker G, Perl TM, Russell PK, Tonat K, Working Grp Civilian B.** 2001. Tularemia as a biological weapon - Medical and public health management. Jama-Journal of the American Medical Association **285:**2763-2773.
8. **Salomonsson E.** 2008. The role of the Type IV pili system in the virulence of *Francisella tularensis*. Doctoral. Umea University, Umea.
9. **Olsufjev NG, Emelyanova OS, Dunaeva TN.** 1959. Comparative study of strains of *B. tularense* in the Old and New World and their taxonomy. J Hyg Epdimiol Microbiol Immunol **3:**139-149.
10. **Larsson P, Elfsmark D, Svensson K, Wikstrom P, Forsman M, Brettin T, Keim P, Johansson A.** 2009. Molecular evolutionary consequences of niche restriction in *Francisella tularensis*, a facultative intracellular pathogen. Plos Pathogens **5**.
11. **Kingry LC, Petersen JM.** 2014. Comparative review of *Francisella tularensis* and *Francisella novicida*. Frontiers in Cellular and Infection Microbiology **4:** DOI 10.3389/fcimb.2014.00035.
12. **Garrity G.** 2005. Bergey's Manual of Systematic Bacteriology, 2 ed, vol 2. Williams & Wilkins.
13. **Larson CL, Wicht W, Jellison WL.** 1955. A new organism resembling *P. tularensis* isolated from water. Public Health Reports **70:**253-258.
14. **Hollis DG, Weaver RE, Steigerwalt AG, Wenger JD, Moss CW, Brenner DJ.** 1989. *Francisella philomiragia* comb. nov. (formerly *Yersinia philomiragia*) and *Francisella tularensis* biogroup novicida (formerly *Francisella novicida*) associated with human disease. Journal of Clinical Microbiology **27:**1601-1608.
15. **Huber B, Escudero R, Busse HJ, Seibold E, Scholz HC, Anda P, Kampfer P, Splettstoesser WD.** 2010. Description of *Francisella hispaniensis* sp *nov*., isolated from human blood, reclassification of *Francisella novicida* (Larson et al. 1955) Olsufiev et al. 1959 as *Francisella tularensis* subsp *novicida* comb. *nov* and emended description of the

genus *Francisella*. International Journal of Systematic and Evolutionary Microbiology **60:**1887-1896.

16. **Johansson A, Celli J, Conlan W, Elkins KL, Forsman M, Keim PS, Larsson P, Manoil C, Nano FE, Petersen JM, Sjostedt A.** 2010. Objections to the transfer of *Francisella novicida* to the subspecies rank of *Francisella tularensis*. International Journal of Systematic and Evolutionary Microbiology **60:**1717-1718.
17. **Busse HJ, Huber B, Anda P, Escudero R, Scholz HC, Seibold E, Splettstoesser WD, Kampfer P.** 2010. Objections to the transfer of *Francisella novicida* to the subspecies rank of *Francisella tularensis* - response to Johansson et al. International Journal of Systematic and Evolutionary Microbiology **60:**1718-1720.
18. **Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E, Starr MP, Truper HG.** 1987. Report of the ad-hoc-committee on reconciliation of approaches to bacterial systematics. International Journal of Systematic Bacteriology **37:**463-464.
19. **Staples JE, Kubota KA, Chalcraft LG, Mead PS, Petersen JM.** 2006. Epidemiologic and molecular analysis of human tularemia, United States, 1964-2004. Emerging Infectious Diseases **12:**1113-1118.
20. **Johansson A, Farlow J, Larsson P, Dukerich M, Chambers E, Bystrom M, Fox J, Chu M, Forsman M, Sjostedt A, Keim P.** 2004. Worldwide genetic relationships among *Francisella tularensis* isolates determined by multiple-locus variable-number tandem repeat analysis. Journal of Bacteriology **186:**5808-5818.
21. **Broekhuijsen M, Larsson N, Johansson A, Bystrom M, Eriksson U, Larsson E, Prior RG, Sjostedt A, Titball RW, Forsman M.** 2003. Genome-wide DNA microarray analysis of *Francisella tularensis* strains demonstrates extensive genetic conservation within the species but identifies regions that are unique to the highly virulent *F. tularensis* subsp *tularensis*. Journal of Clinical Microbiology **41:**2924-2931.
22. **Whipp MJ, Davis JM, Lum G, de Boer J, Zhou Y, Bearden SW, Petersen JM, Chu MC, Hoggi G.** 2003. Characterization of a *novicida*-like subspecies of *Francisella tularensis* isolated in Australia. Journal of Medical Microbiology **52:**839-842.
23. **Owen CR, Lackman DB, Jellison WL, Buker EO, Bell JF.** 1964. Comparative studies of *Francisella tularensis* + *Francisella novicida*. Journal of Bacteriology **87:**676-&.
24. **Weiss DS, Brotcke A, Henry T, Margolis JJ, Chan K, Monack DM.** 2007. In vivo negative selection screen identifies genes required for *Francisella* virulence. Proceedings of the National Academy of Sciences of the United States of America **104:**6037-6042.
25. **Champion MD, Zeng QD, Nix EB, Nano FE, Keim P, Kodira CD, Borowsky M, Young S, Koehrsen M, Engels R, Pearson M, Howarth C, Larson L, White J, Alvarado L, Forsman M, Bearden SW, Sjostedt A, Titball R, Michell SL, Birren B, Galagan J.** 2009. Comparative genomic characterization of *Francisella tularensis* strains belonging to low and high virulence subspecies. Plos Pathogens **5**.
26. **Farlow J, Wagner DM, Dukerich M, Stanley M, Chu M, Kubota K, Petersen J, Keim P.** 2005. *Francisella tularensis* in the United States. Emerging Infectious Diseases **11:**1835-1841.
27. **Oyston PCF.** 2008. *Francisella tularensis*: unravelling the secrets of an intracellular pathogen. Journal of Medical Microbiology **57:**921-930.

28. **Olsufjev NG, Meshcheryakova IS.** 1983. Subspecific taxonomy of *Francisella tularensis* McCoy and Chapin 1912. International Journal of Systematic Bacteriology **33:**872-874.
29. **Sandstrom G, Sjostedt A, Forsman M, Pavlovich NV, Mishankin BN.** 1992. Characterization and classification of strains of *Francisella tularensis* isolated in the central Asain focus of the Soviet Union and in Japan. Journal of Clinical Microbiology **30:**172-175.
30. **Thomas R, Johansson A, Neeson B, Isherwood K, Sjostedt A, Ellis J, Titball RW.** 2003. Discrimination of human pathogenic subspecies of *Francisella tularensis* by using restriction fragment length polymorphism. Journal of Clinical Microbiology **41:**50-57.
31. **Fujita O, Uda A, Hotta A, Okutani A, Inoue S, Tanabayashi K, Yamada A.** 2008. Genetic diversity of *Francisella tularensis* subspecies *holarctica* strains isolated in Japan. Microbiology and Immunology **52:**270-276.
32. **Vivekananda J, Kiel JL.** 2006. Anti-*Francisella tularensis* DNA aptamers detect tularemia antigen from different subspecies by aptamer-linked immobilized sorbent assay. Laboratory Investigation **86:**610-618.
33. **Larsson P, Oyston PCF, Chain P, Chu MC, Duffield M, Fuxelius HH, Garcia E, Halltorp G, Johansson D, Isherwood KE, Karp PD, Larsson E, Liu Y, Michell S, Prior J, Prior R, Malfatti S, Sjostedt A, Svensson K, Thompson N, Vergez L, Wagg JK, Wren BW, Lindler LE, Andersson SGE, Forsman M, Titball RW.** 2005. The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. Nature Genetics **37:**153-159.
34. **Petrosino JF, Xiang Q, Karpathy SE, Jiang HY, Yerrapragada S, Liu YM, Gioia J, Hemphill L, Gonzalez A, Raghavan TM, Uzman A, Fox GE, Highlander S, Reichard M, Morton RJ, Clinkenbeard KD, Weinstock GM.** 2006. Chromosome rearrangement and diversification of *Francisella tularensis* revealed by the type B (OSU18) genome sequence. Journal of Bacteriology **188:**6977-6985.
35. **Chaudhuri RR, Ren CP, Desmond L, Vincent GA, Silman NJ, Brehm JK, Elmore MJ, Hudson MJ, Forsman M, Isherwood KE, Gurycova D, Minton NP, Titball RW, Pallen MJ, Vipond R.** 2007. Genome sequencing shows that European isolates of *Francisella tularensis* subspecies *tularensis* are almost identical to US laboratory strain Schu S4. Plos One **2**.
36. **Rohmer L, Fong C, Abmayr S, Wasnick M, Freeman TJL, Radey M, Guina T, Svensson K, Hayden HS, Jacobs M, Gallagher LA, Manoil C, Ernst RK, Drees B, Buckley D, Haugen E, Bovee D, Zhou Y, Chang J, Levy R, Lim R, Gillett W, Guenthener D, Kang A, Shaffer SA, Taylor G, Chen JZ, Gallis B, D'Argenio DA, Forsman M, Olson MV, Goodlett DR, Kaul R, Miller SI, Brittnacher MJ.** 2007. Comparison of *Francisella tularensis* genomes reveals evolutionary events associated with the emergence of human pathogenic strains. Genome Biology **8**.
37. **Beckstrom-Sternberg SM, Auerbach RK, Godbole S, Pearson JV, Beckstrom-Sternberg JS, Deng ZM, Munk C, Kubota K, Zhou Y, Bruce D, Noronha J, Scheuermann RH, Wang AH, Wei XY, Wang JJ, Hao J, Wagner DM, Brettin TS, Brown N, Gilna P, Keim PS.** 2007. Complete genomic characterization of a pathogenic A.II strain of *Francisella tularensis* subspecies *tularensis*. Plos One **2**.
38. **Barabote RD, Xie G, Brettin TS, Hinrichs SH, Fey PD, Jay JJ, Engle JL, Godbole SD, Noronha JM, Scheuermann RH, Zhou LW, Lion C, Dempsey MP.** 2009.

Complete genome sequence of *Francisella tularensis* subspecies *holarctica* FTNF002-00. Plos One **4**.

39. **Nalbantoglu U, Sayood K, Dempsey MP, Iwen PC, Francesconi SC, Barabote RD, Xie G, Brettin TS, Hinrichs SH, Fey PD.** 2010. Large direct repeats flank genomic rearrangements between a new clinical isolate of *Francisella tularensis* subsp *tularensis* A1 and Schu S4. Plos One **5**.
40. **Modise T, Ryder C, Mane SP, Bandara AB, Jensen RV, Inzana TJ.** 2012. Genomic comparison between a virulent type A1 strain of *Francisella tularensis* and its attenuated O-antigen mutant. Journal of Bacteriology **194:**2775-2776.
41. **Svensson K, Sjodin A, Bystrom M, Granberg M, Brittnacher MJ, Rohmer L, Jacobs MA, Sims-Day EH, Levy R, Zhou Y, Hayden HS, Lim R, Chang J, Guenthener D, Kang A, Haugen E, Gillett W, Kaul R, Forsman M, Larsson P, Johansson A.** 2012. Genome sequence of *Francisella tularensis* subspecies *holarctica* strain FSC200, isolated from a child with tularemia. Journal of Bacteriology **194:**6965-6966.
42. **La Scola B, Elkarkouri K, Li WJ, Wahab T, Fournous G, Rolain JM, Biswas S, Drancourt M, Robert C, Audic S, Lofdahl S, Raoult D.** 2008. Rapid comparative genomic analysis for clinical microbiology: The *Francisella tularensis* paradigm. Genome Research **18:**742-750.
43. **Forsman M, Sandstrom G, Sjostedt A.** 1994. Analysis of 16S ribosomal DNA sequences of *Francisella* strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR. International Journal of Systematic Bacteriology **44:**38-46.
44. **Pohanka M, Hubalek M, Neubauerova V, Macela A, Faldyna M, Bandouchova H, Pikula J.** 2008. Current and emerging assays for *Francisella tularensis* detection: a review. Veterinarni Medicina **53:**585-594.
45. **Samrakandi MM, Zhang C, Zhang M, Nietfeldt J, Kim J, Iwen PC, Olson ME, Fey PD, Duhamel GE, Hinrichs SH, Cirillo JD, Benson AK.** 2004. Genome diversity among regional populations of *Francisella tularensis* subspecies *tularensis* and *Francisella tularensis* subspecies *holarctica* isolated from the US. Fems Microbiology Letters **237:**9-17.
46. **Nano FE, Zhang N, Cowley SC, Klose KE, Cheung KKM, Roberts MJ, Ludu JS, Letendre GW, Meierovics AI, Stephens G, Elkins KL.** 2004. A *Francisella tularensis* pathogenicity island required for intramacrophage growth. Journal of Bacteriology **186:**6430-6436.
47. **Nano FE, Schmerk C.** 2007. The *Francisella* pathogenicity island. Francisella Tularensis: Biology, Pathogenicity, Epidemiology, and Biodefense **1105:**122-137.
48. **Rohmer L, Brittnacher M, Svensson K, Buckley D, Haugen E, Zhou Y, Chang J, Levy R, Hayden H, Forsman M, Olson M, Johansson A, Kaul R, Miller SI.** 2006. Potential source of *Francisella tularensis* live vaccine strain attenuation determined by genome comparison. Infection and Immunity **74:**6895-6906.
49. **Green M, Choules G, Rogers D, Titball RW.** 2005. Efficacy of the live attenuated *Francisella tularensis* vaccine (LVS) in a murine model of disease. Vaccine **23:**2680-2686.
50. **Dempsey MP, Dobson M, Zhang C, Zhang M, Lion C, Gutierrez-Martin CB, Iwen PC, Fey PD, Olson ME, Niemeyer D, Francesconi S, Crawford R, Stanley M, Rhodes J, Wagner DM, Vogler AJ, Birdsell D, Keim P, Johansson A, Hinrichs SH,**

Benson AK. 2007. Genomic deletion marking an emerging subclone of *Francisella tularensis* subsp *holarctica* in France and the Iberian Peninsula. Applied and Environmental Microbiology **73:**7465-7470.

51. **Olsufjev NG, Meshcheryakova IS.** 1982. Infraspecific taxonomy of tularemia agent *Francisella tularensis* McCoy et Chapin. Journal of Hygiene Epidemiology Microbiology and Immunology **26:**291-299.
52. **Nübel U, Reissbrodt R, Weller A, Grunow R, Porsch-Ozcurumez M, Tomaso H, Hofer E, Splettstoesser W, Finke EJ, Tschape H, Witte W.** 2006. Population structure of *Francisella tularensis*. Journal of Bacteriology **188:**5319-5324.
53. **Darling ACE, Mau B, Blattner FR, Perna NT.** 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research **14:**1394-1403.
54. **Niebylski ML, Peacock MG, Fischer ER, Porcella SF, Schwan TG.** 1997. Characterization of an endosymbiont infecting wood ticks, *Dermacentor andersoni*, as a member of the genus *Francisella*. Applied and Environmental Microbiology **63:**3933-3940.
55. **Scoles GA.** 2004. Phylogenetic analysis of the *Francisella*-like endosymbionts of *Dermacentor* ticks. Journal of Medical Entomology **41:**277-286.
56. **Goethert HK, Telford SR.** 2005. A new *Francisella* (Beggiatiales : *Francisellaceae*) inquiline within *Dermacentor variabilis* say (Acari : *Ixodidae*). Journal of Medical Entomology **42:**502-505.
57. **Dergousoff SJ, Chilton NB.** 2012. Association of different genetic types of *Francisella*-like organisms with the Rocky Mountain Wood Tick (*Dermacentor andersoni*) and the American Dog Tick (*Dermacentor variabilis*) in localities near their northern distributional limits. Applied and Environmental Microbiology **78:**965-971.
58. **Szigeti A, Kreizinger Z, Hornok S, Abichu G, Gyuranecz M.** 2014. Detection of *Francisella*-like endosymbiont in *Hyalomma rufipes* from Ethiopia. Ticks and Tick-Borne Diseases **5:**818-820.
59. **Kugeler KJ, Gurfield N, Creek JG, Mahoney KS, Versage JL, Petersen JM.** 2005. Discrimination between *Francisella tularensis* and *Francisella*-like endosymbionts when screening ticks by PCR. Applied and Environmental Microbiology **71:**7594-7597.
60. **Sreter-Lancz Z, Szell Z, Sreter T, Marialigeti K.** 2009. Detection of a novel *Francisella* in *Dermacentor reticulatus*: A need for careful evaluation of PCR-based identification of *Francisella tularensis* in Eurasian ticks. Vector-Borne and Zoonotic Diseases **9:**123-126.
61. **Foroshani NS, Karami A, Pourali F.** 2013. Simultaneous and rapid detection of *Salmonella typhi*, *Bacillus anthracis*, and *Yersinia pestis* by using multiplex polymerase chain reaction (PCR). Iranian Red Crescent Medical Journal **15**.
62. **Sting R, Runge M, Eisenberg T, Braune S, Muller W, Otto P.** 2013. Comparison of bacterial culture and polymerase chain reaction (PCR) for the detection of *F. tularensis* subsp *holarctica* in wild animals. Berliner Und Munchener Tierarztliche Wochenschrift **126:**285-290.
63. **Celebi B, Kilic S, Yesilyurt M, Acar B.** 2014. Evaluation of a newly-developed ready-to-use commercial PCR kit for the molecular diagnosis of *Francisella tularensis*. Mikrobiyoloji Bulteni **48:**135-142.
64. **Splettstoesser WD, Tomaso H, Al Dahouk S, Neubauer H, Schuff-Werner P.** 2005. Diagnostic procedures in tularaemia with special focus on molecular and immunological

techniques. Journal of Veterinary Medicine Series B-Infectious Diseases and Veterinary Public Health **52:**249-261.

65. **He J, Kraft AJ, Fan JA, Van Dyke M, Wang LH, Bose ME, Khanna M, Metallo JA, Henrickson KJ.** 2009. Simultaneous detection of CDC category "A" DNA and RNA bioterrorism agents by use of multiplex PCR & RT-PCR enzyme hybridization assays. Viruses-Basel **1:**441-459.
66. **Kormilitsyna MI, Meshcheryakova IS, Mikhailova TV.** 2013. Molecular and genetic characterization of *Francisella tularensis* strains of differing taxonomic status and virulence. Molecular Genetics Microbiology and Virology **28:**110-114.
67. **Varma-Basil M, El-Hajj H, Marras SAE, Hazbon MH, Mann JM, Connell ND, Kramer FR, Alland D.** 2004. Molecular beacons for multiplex detection of four bacterial bioterrorism agents. Clinical Chemistry **50:**1060-1063.
68. **Skottman T, Piiparinen H, Hyytiainen H, Myllys V, Skurnik M, Nikkari S.** 2007. Simultaneous real-time PCR detection of *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis*. European Journal of Clinical Microbiology & Infectious Diseases **26:**207-211.
69. **Hsu HL, Huang HH, Liang CC, Lin HC, Liu WT, Lin FP, Kau JH, Sun KH.** 2013. Suspension bead array of the single-stranded multiplex polymerase chain reaction amplicons for enhanced identification and quantification of multiple pathogens. Analytical Chemistry **85:**5562-5568.
70. **Bergval I, Sengstake S, Brankova N, Levterova V, Abadia E, Tadumaze N, Bablishvili N, Akhalaia M, Tuin K, Schuitema A, Panaiotov S, Bachiyska E, Kantardjiev T, de Zwaan R, Schurch A, van Soolingen D, van 't Hoog A, Cobelens F, Aspindzelashvili R, Sola C, Klatser P, Anthony R.** 2012. Combined species identification, genotyping, and drug resistance detection of *Mycobacterium tuberculosis* cultures by MLPA on a bead-based array. Plos One **7**.
71. **Munro SB, Kuypers J, Jerome KR.** 2013. Comparison of a multiplex real-time PCR assay with a multiplex Luminex assay for influenza virus detection. Journal of Clinical Microbiology **51:**1124-1129.
72. **Molins-Schneekloth CR, Belisle JT, Petersen JM.** 2008. Genomic markers for differentiation of *Francisella tularensis* subsp *tularensis* A.I and A.II strains. Applied and Environmental Microbiology **74:**336-341.
73. **Molins CR, Carlson JK, Coombs J, Petersen JA.** 2009. Identification of *Francisella tularensis* subsp *tularensis* A1 and A2 infections by real-time polymerase chain reaction. Diagnostic Microbiology and Infectious Disease **64:**6-12.
74. **Zasada AA, Forminska K, Zacharczuk K.** 2013. Fast identification of *Yersinia pestis*, *Bacillus anthracis*, and *Francisella tularensis* based on conventional PCR. Polish Journal of Microbiology **62:**453-455.
75. **Sellek R, Jimenez O, Aizpurua C, Fernandez-Frutos B, De Leon P, Camacho M, Fernandez-Moreira D, Ybarra C, Cabria JC.** 2008. Recovery of *Francisella tularensis* from soil samples by filtration and detection by real-time PCR and cELISA. Journal of Environmental Monitoring **10:**362-369.
76. **Pandya GA, Holmes MH, Petersen JM, Pradhan S, Karamycheva SA, Wolcott MJ, Molins C, Jones M, Schriefer ME, Fleischmann RD, Peterson SN.** 2009. Whole genome single nucleotide polymorphism based phylogeny of *Francisella tularensis* and its application to the development of a strain typing assay. Bmc Microbiology **9**.

77. **Svensson K, Granberg M, Karlsson L, Neubauerova V, Forsman M, Johansson A.** 2009. A real-time PCR array for hierarchical identification of *Francisella* isolates. Plos One **4**.
78. **Woubit A, Yehualaeshet T, Habtemariam T, Samuel T.** 2012. Novel genomic tools for specific and real-time detection of biothreat and frequently encountered foodborne pathogens. Journal of Food Protection **75:**660-670.
79. **Abril C, Nimmervoll H, Pilo P, Brodard I, Korczak B, Seiler M, Miserez R, Frey J.** 2008. Rapid diagnosis and quantification of Francisella tularensis in organs of naturally infected common squirrel monkeys (Saimiri sciureus). Veterinary Microbiology **127:**203-208.
80. **Yang S, Rothman RE, Hardick J, Kuroki M, Hardick A, Doshi V, Ramachandran P, Gaydos CA.** 2008. Rapid polymerase chain reaction-based screening assay for bacterial biothreat agents. Academic Emergency Medicine **15:**388-392.
81. **Angelakis E, Roux V, Raoult D, Rolain JM.** 2009. Real-time PCR strategy and detection of bacterial agents of lymphadenitis. European Journal of Clinical Microbiology & Infectious Diseases **28:**1363-1368.
82. **Simsek H, Taner M, Karadenizli A, Ertek M, Vahaboglu H.** 2012. Identification of *Francisella tularensis* by both culture and real-time TaqMan PCR methods from environmental water specimens in outbreak areas where tularemia cases were not previously reported. European Journal of Clinical Microbiology & Infectious Diseases **31:**2353-2357.
83. **Birdsell DN, Vogler AJ, Buchhagen J, Clare A, Kaufman E, Naumann A, Driebe E, Wagner DM, Keim PS.** 2014. TaqMan real-time PCR assays for single-nucleotide polymorphisms which identify *Francisella tularensis* and its subspecies and subpopulations. Plos One **9**.
84. **Matero P, Hemmila H, Tomaso H, Piiparinen H, Rantakokko-Jalava K, Nuotio L, Nikkari S.** 2011. Rapid field detection assays for *Bacillus anthracis*, *Brucella* spp., *Francisella tularensis* and *Yersinia pestis*. Clinical Microbiology and Infection **17:**34-43.
85. **Seiner DR, Colburn HA, Baird C, Bartholomew RA, Straub T, Victry K, Hutchison JR, Valentine N, Bruckner-Lea CJ.** 2013. Evaluation of the FilmArray (R) system for detection of *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis*. Journal of Applied Microbiology **114:**992-1000.
86. **Rachwal PA, Rose HL, Cox V, Lukaszewski RA, Murch AL, Weller SA.** 2012. The potential of TaqMan array cards for detection of multiple biological agents by real-time PCR. Plos One **7**.
87. **Weller SA, Cox V, Essex-Lopresti A, Hartley MG, Parsons TM, Rachwal PA, Stapleton HL, Lukaszewski RA.** 2012. Evaluation of two multiplex real-time PCR screening capabilities for the detection of *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis* in blood samples generated from murine infection models. Journal of Medical Microbiology **61:**1546-1555.
88. **Jacob D, Sauer U, Housley R, Washington C, Sannes-Lowery K, Ecker DJ, Sampath R, Grunow R.** 2012. Rapid and high-throughput detection of highly pathogenic bacteria by Ibis PLEX-ID technology. Plos One **7**.
89. **Jeng K, Hardick J, Rothman R, Yang S, Won H, Peterson S, Hsieh YH, Masek BJ, Carroll KC, Gaydos CA.** 2013. Reverse transcription-PCR-electrospray ionization mass

spectrometry for rapid detection of biothreat and common respiratory pathogens. Journal of Clinical Microbiology **51:**3300-3307.

90. **Whitehouse CA, Kesterson KE, Duncan DD, Eshoo MW, Wolcott M.** 2012. Identification and characterization of *Francisella* species from natural warm springs in Utah, USA. Letters in Applied Microbiology **54:**313-324.
91. **Duncan DD, Vogler AJ, Wolcott MJ, Li F, Sarovich DS, Birdsell DN, Watson LM, Hall TA, Sampath R, Housley R, Blyn LB, Hofstadler SA, Ecker DJ, Keim P, Wagner DM, Eshoo MW.** 2013. Identification and typing of *Francisella tularensis* with a highly automated genotyping assay. Letters in Applied Microbiology **56:**128-134.
92. **Euler M, Wang YJ, Otto P, Tomaso H, Escudero R, Anda P, Hufert FT, Weidmann M.** 2012. Recombinase polymerase amplification assay for rapid detection of *Francisella tularensis*. Journal of Clinical Microbiology **50:**2234-2238.
93. **Euler M, Wang YJ, Heidenreich D, Patel P, Strohmeier O, Hakenberg S, Niedrig M, Hufert FT, Weidmann M.** 2013. Development of a panel of recombinase polymerase amplification assays for detection of biothreat agents. Journal of Clinical Microbiology **51:**1110-1117.
94. **del Rio JS, Adly NY, Acero-Sanchez JL, Henry OYF, O'Sullivan CK.** 2014. Electrochemical detection of *Francisella tularensis* genomic DNA using solid-phase recombinase polymerase amplification. Biosensors & Bioelectronics **54:**674-678.
95. **Jacob D, Wahab T, Edvinsson B, Peterzon A, Boskani T, Farhadi L, Barduhn A, Grunow R, Sandstrom G.** 2011. Identification and subtyping of *Francisella* by pyrosequencing and signature matching of 16S rDNA fragments. Letters in Applied Microbiology **53:**592-595.
96. **Turingan RS, Thomann HU, Zolotova A, Tan E, Selden RF.** 2013. Rapid focused sequencing: A multiplexed assay for simultaneous detection and strain typing of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. Plos One **8**.
97. **Svensson K, Larsson P, Johansson D, Bystrom M, Forsman M, Johansson A.** 2005. Evolution of subspecies of *Francisella tularensis*. Journal of Bacteriology **187:**3903-3908.
98. **Birdsell DN, Johansson A, Ohrman C, Kaufman E, Molins C, Pearson T, Gyuranecz M, Naumann A, Vogler AJ, Myrtennas K, Larsson P, Forsman M, Sjodin A, Gillece JD, Schupp J, Petersen JM, Keim P, Wagner DM.** 2014. *Francisella tularensis* subsp *tularensis* group A.I, United States. Emerging Infectious Diseases **20:**861-865.
99. **Dempsey MP, Nietfeldt J, Ravel J, Hinrichs S, Crawford R, Benson AK.** 2006. Paired-end sequence mapping detects extensive genomic rearrangement and translocation during divergence of *Francisella tularensis* subsp *tularensis* and *Francisella tularensis* subsp *holarctica* populations. Journal of Bacteriology **188:**5904-5914.
100. **Keim P, Johansson A, Wagner DM.** 2007. Molecular epidemiology, evolution, and ecology of *Francisella*. Francisella Tularensis: Biology, Pathogenicity, Epidemiology, and Biodefense **1105:**30-66.
101. **Larsson P, Svensson K, Karlsson L, Guala D, Granberg M, Forsman M, Johansson A.** 2007. Canonical insertion-deletion markers for rapid DNA typing of *Francisella tularensis*. Emerging Infectious Diseases **13:**1725-1732.
102. **Vogler AJ, Birdsell D, Price LB, Bowers JR, Beckstrom-Sternberg SM, Auerbach RK, Beckstrom-Sternberg JS, Johansson A, Clare A, Buchhagen JL, Petersen JM,**

Pearson T, Vaissaire J, Dempsey MP, Foxall P, Engelthaler DM, Wagner DM, Keim P. 2009. Phylogeography of *Francisella tularensis*: Global expansion of a highly fit clone. Journal of Bacteriology **191:**2474-2484.

103. **Sjödin A, Svensson K, Ohrman C, Ahlinder J, Lindgren P, Duodu S, Johansson A, Colquhoun DJ, Larsson P, Forsman M.** 2012. Genome characterisation of the genus *Francisella* reveals insight into similar evolutionary paths in pathogens of mammals and fish. Bmc Genomics **13**.
104. **Zeytun A, Malfatti SA, Vergez LM, Shin M, Garcia E, Chain PSG.** 2012. Complete genome sequence of *Francisella philomiragia* ATCC 25017. Journal of Bacteriology **194:**3266-3266.
105. **Wenger JD, Hollis DG, Weaver RE, Baker CN, Brown GR, Brenner DJ, Broome CV.** 1989. Infection caused by *Francisella philomiragia* (formerly *Yersinia philomiragia* - a newly recognized human pathogen. Annals of Internal Medicine **110:**888-892.
106. **Bliven KA, Maurelli AT.** 2012. Antivirulence genes: insights into pathogen evolution through gene loss. Infection and Immunity **80:**4061-4070.
107. **Turner PE, Williams E, Okeke C, Cooper VS, Duffy S, Wertz JE.** 2014. Antibiotic resistance correlates with transmission in plasmid evolution. Evolution **68:**3368-3380.
108. **Dunlop PSM, Ciavola M, Rizzo L, McDowell DA, Byrne JA.** 2015. Effect of photocatalysis on the transfer of antibiotic resistance genes in urban wastewater. Catalysis Today **240:**55-60.
109. **Ying JC, Wang HF, Bao BK, Zhang Y, Zhang JF, Zhang C, Li AF, Lu JW, Li PZ, Ying J, Liu Q, Xu T, Yi HG, Li JS, Zhou L, Zhou TL, Xu ZY, Ni LY, Bao QY.** 2015. Molecular variation and horizontal gene transfer of the homocysteine methyltransferase gene mmuM and its distribution in clinical pathogens. International Journal of Biological Sciences **11:**11-21.
110. **Gestin B, Valade E, Thibault F, Schneider D, Maurin M.** 2010. Phenotypic and genetic characterization of macrolide resistance in *Francisella tularensis* subsp. *holarctica* biovar I. Journal of Antimicrobial Chemotherapy **65:**2359-2367.
111. **Siddaramappa S, Challacombe JF, Petersen JM, Pillai S, Kuske CR.** 2012. Genetic diversity within the genus *Francisella* as revealed by comparative analyses of the genomes of two North American isolates from environmental sources. Bmc Genomics **13**.
112. **Sutera V, Levert M, Burmeister WP, Schneider D, Maurin M.** 2014. Evolution toward high-level fluoroquinolone resistance in *Francisella species*. Journal of Antimicrobial Chemotherapy **69:**101-110.
113. **Gallagher LA, McKevitt M, Ramage ER, Manoil C.** 2008. Genetic dissection of the *Francisella novicida* restriction barrier. Journal of Bacteriology **190:**7830-7837.
114. **Schunder E, Rydzewski K, Grunow R, Heuner K.** 2013. First indication for a functional CRISPR/Cas system in *Francisella tularensis*. International Journal of Medical Microbiology **303:**51-60.
115. **Sampson TR, Weiss DS.** 2013. Degeneration of a CRISPR/Cas system and its regulatory target during the evolution of a pathogen. Rna Biology **10:**1618-1622.
116. **Hager AJ, Bolton DL, Pelletier MR, Brittnacher MJ, Gallagher LA, Kaul R, Skerrett SJ, Miller SI, Guina T.** 2006. Type IV pili-mediated secretion modulates *Francisella* virulence. Molecular Microbiology **62:**227-237.

117. **Siddaramappa S, Challacombe JF, Petersen JM, Pillai S, Hogg G, Kuske CR.** 2011. Common ancestry and novel genetic traits of *Francisella novicida*-like isolates from North America and Australia as revealed by comparative genomic analyses. Applied and Environmental Microbiology **77:**5110-5122.
118. **Dennis DT, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, Fine AD, Friedlander AM, Hauer J, Layton M, Lillibridge SR, McDade JE, Osterholm MT, O'Toole T, Parker G, Perl TM, Russell PK, Tonat K.** 2001. Tularemia as a biological weapon: medical and public health management. JAMA (Journal of the American Medical Association) **285:**2763-2773.
119. **Hollis DG, Weaver RE, Steigerwalt AG, Wenger JD, Moss CW, Brenner DJ.** 1989. *Francisella philomiragia* comb *nov* (formerly *Yersinia philomiragia*) and *Francisella tularensis* biogroup *novicia* (formerly *Francisella novicida*) associated with human disease. Journal of Clinical Microbiology **27:**1601-1608.
120. **Gustafsson K.** 1989. Growth and survival of 4 strains of *Francisella tularensis* in a rich medium preconditioned with *Acanthamobea palestiensis*. Canadian Journal of Microbiology **35:**1100-1104.
121. **Abd H, Johansson T, Golovliov I, Sandstrom G, Forsman M.** 2003. Survival and growth of *Francisella tularensis* in *Acanthamoeba castellanii*. Applied and Environmental Microbiology **69:**600-606.
122. **El-Etr SH, Margolis JJ, Monack D, Robison RA, Cohen M, Moore E, Rasley A.** 2009. *Francisella tularensis* Type A strains cause the rapid encystment of *Acanthamoeba castellanii* and survive in amoebal cysts for three weeks postinfection. Applied and Environmental Microbiology **75:**7488-7500.
123. **Del Blanco NG, Dobson ME, Vela AI, De La Puente VA, Gutierrez CB, Hadfield TL, Kuhnert P, Frey J, Dominguez L, Ferri EFR.** 2002. Genotyping of *Francisella tularensis* strains by pulsed-field gel electrophoresis, amplified fragment length polymorphism fingerprinting, and 16S rRNA gene sequencing. Journal of Clinical Microbiology **40:**2964-2972.
124. **Johansson A, Ibrahim A, Goransson I, Eriksson U, Gurycova D, Clarridge JE, Sjostedt A.** 2000. Evaluation of PCR-based methods for discrimination of *Francisella* species and subspecies and development of a specific PCR that distinguishes the two major subspecies of *Francisella tularensis*. Journal of Clinical Microbiology **38:**4180-4185.
125. **Versage JL, Severin DDM, Chu MC, Petersen JM.** 2003. Development of a multitarget real-time TaqMan PCR assay for enhanced detection of *Francisella tularensis* in complex specimens. Journal of Clinical Microbiology **41:**5492-5499.
126. **Kugeler KJ, Pappert R, Zhou Y, Petersen JM.** 2006. Real-time PCR for *Francisella tularensis* types A and B. Emerging Infectious Diseases **12:**1799-1801.
127. **Tomaso H, Reisinger EC, Al Dahouk S, Frangoulidis D, Rakin A, Landt O, Neubauer H.** 2003. Rapid detection of *Yersinia pestis* with multiplex real-time PCR assays using fluorescent hybridisation probes. Fems Immunology and Medical Microbiology **38:**117-126.
128. **Stewart A, Satterfield B, Cohen M, O'Neill K, Robison R.** 2008. A quadruplex real-time PCR assay for the detection of *Yersinia pestis* and its plasmids. Journal of Medical Microbiology **57:**324-331.

129. **Jones CL, Napier BA, Sampson TR, Llewellyn AC, Schroeder MR, Weiss DS.** 2012. Subversion of host recognition and defense systems by *Francisella* spp. Microbiology and Molecular Biology Reviews **76:**383-404.
130. **Sjostedt A.** 2007. Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. *Francisella Tularensis*: Biology, Pathogenicity, Epidemiology, and Biodefense **1105:**1-29.
131. **Oyston PCF, Quarry JE.** 2005. Tularemia vaccine: past, present and future. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology **87:**277-281.
132. **Anonymous.** 2005. Bergey's Manual of Systematic Bacteriology, 2 ed, vol 2. Williams & Wilkins.
133. **Broekhuijsen M, Larsson P, Johansson A, Bystrom M, Eriksson U, Larsson E, Prior R, Sjostedt A, Titball RW, Forsman M.** 2003. Genome-wide DNA microarray analysis of *Francisella tularensis* strains demonstrates extensive genetic conservation within the species but identifies regions that are unique to the highly virulent *F. tularensis* subsp. *tularensis*. Journal of Clinical Microbiology **41:**2924-2931.
134. **Kaper JB.** 2005. Pathogenic Escherichia coli. International Journal of Medical Microbiology **295:**355-356.
135. **Barker JR, Chong A, Wehrly TD, Yu JJ, Rodriguez SA, Liu JR, Celli J, Arulanandam BP, Klose KE.** 2009. The pathogenicity island encodes a secretion system that is required for phagosome escape and virulence. Molecular Microbiology **74:**1459-1470.
136. **Sridhar S, Sharma A, Kongshaug H, Nilsen F, Jonassen I.** 2012. Whole genome sequencing of the fish pathogen *Francisella noatunensis* subsp *orientalis* Toba04 gives novel insights into *Francisella* evolution and pathogenecity. Bmc Genomics **13**.
137. **Santic M, Al-Khodor S, Abu Kwaik Y.** 2010. Cell biology and molecular ecology of Francisella tularensis. Cellular Microbiology **12:**129-139.
138. **Ward PN, Holden MTG, Leigh JA, Lennard N, Bignell A, Barron A, Clark L, Quail MA, Woodward J, Barrell BG, Egan SA, Field TR, Maskell D, Kehoe M, Dowson CG, Chanter N, Whatmore AM, Bentley SD, Parkhill J.** 2009. Evidence for niche adaptation in the genome of the bovine pathogen *Streptococcus uberis*. Bmc Genomics **10**.
139. **Feng Y, Chen Z, Liu SL.** 2011. Gene decay in *Shigella* as an incipient stage of host-adaptation. Plos One **6**.
140. **Hain T, Chatterjee SS, Ghaia R, Kuenne CT, Billion A, Steinweg C, Domann E, Karst U, Jansch L, Wehland J, Eisenreich W, Bacherc A, Joseph B, Schar J, Kreft J, Klumpp J, Loessner MJ, Dorscht J, Neuhaus K, Fuchs TM, Scherer S, Doumith M, Jacquet C, Martin P, Cossart P, Rusniock C, Glaser P, Buchrieser C, Goebel W, Chakraborty T.** 2007. Pathogenomics of *Listeria* spp. International Journal of Medical Microbiology **297:**541-557.
141. **Harris SR, Robinson C, Steward KF, Webb KS, Paillot R, Parkhill J, Holden MTG, Waller AS.** 2015. Genome specialization and decay of the strangles pathogen, *Streptococcus equi*, is driven by persistent infection. Genome Research **25:**1360-1371.
142. **Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O.** 2008.

Toward an online repository of standard operating procedures (SOPs) for (meta) genomic annotation. Omics-a Journal of Integrative Biology **12:**137-141.
143. **Su J, Yang J, Zhao D, Kawula TH, Banas JA, Zhang J-R.** 2007. Genome-wide identification of *Francisella tularensis* virulence determinants. Infection and Immunity **75:**3089-3101.
144. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and clustal X version 2.0. Bioinformatics **23:**2947-2948.
145. **Nei M, Kumar S.** 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.
146. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution **28:**2731-2739.
147. **Muse SV, Gaut BS.** 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Molecular Biology and Evolution **11:**715-724.
148. **Hasegawa M, Kishino H, Yano TA.** 1985. Dating of the human ape splitting by a molecular clock of mintochondrial-DNA. Journal of Molecular Evolution **22:**160-174.
149. **Pond SLK, Frost SDW.** 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution **22:**1208-1222.
150. **Suzuki Y, Gojobori T.** 1999. A method for detecting positive selection at single amino acid sites. Molecular Biology and Evolution **16:**1315-1328.
151. **Pond SLK, Frost SDW, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676-679.
152. **Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA.** 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics **19:**671-672.
153. **Jones DT.** 1999. Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology **292:**195-202.
154. **McDonald JH, Kreitman M.** 1991. Adaptive protein evolution at the adh locus in *Drosophila*. Nature **351:**652-654.
155. **Fay JC.** 2011. Weighing the evidence for adaptation at the molecular level. Trends in Genetics **27:**343-349.
156. **Messer PW, Petrov DA.** 2013. Frequent adaptation and the McDonald-Kreitman test. Proceedings of the National Academy of Sciences of the United States of America **110:**8615-8620.
157. **Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP.** 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Molecular Biology and Evolution **16:**372-382.
158. **Sharp PM.** 1997. In search of molecular darwinism. Nature **385:**111-112.
159. **Taylor SD, de la Cruz KD, Porter ML, Whiting MF.** 2005. Characterization of the long-wavelength opsin from Mecoptera and Siphonaptera: Does a flea see? Molecular Biology and Evolution **22:**1165-1174.

160. **McClellan D, Ellison D.** 2010. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. International Journal of Bioinformatics Research and Applications **6:**120-133.
161. **Prahhakaran M, Ponnuswamy PK.** 1979. The spatial distribution of physical, chemical, energetic and conformational properties of amino acid residues in globular proteins. Journal of Theoretical Biology **80:**485-504.
162. **Strom MS, Lory S.** 1993. Structure-function and biogenesis of the type-IV pili. Annual Review of Microbiology **47:**565-596.
163. **Strom MS, Nunn DN, Lory S.** 1993. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type-IV pilin family. Proceedings of the National Academy of Sciences of the United States of America **90:**2404-2408.
164. **Lory S, Strom MS.** 1997. Structure-function relationship of type-IV prepilin peptidase of Pseudomonas aeruginosa - A review. Gene **192:**117-121.
165. **Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW.** 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science **269:**400-403.
166. **Chiang SL, Mekalanos JJ, Holden DW.** 1999. In vivo genetic analysis of bacterial virulence. Annual Review of Microbiology **53:**129-154.
167. **Casadevall A, Pirofski LA.** 1999. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. Infection and Immunity **67:**3703-3713.
168. **Steele S, Brunton J, Ziehr B, Taft-Benz S, Moorman N, Kawula T.** 2013. *Francisella tularensis* harvests nutrients derived via ATG5-independent autophagy to support intracellular growth. Plos Pathogens **9**.
169. **Kone BC, Kuncewicz T, Zhang WZ, Yu ZY.** 2003. Protein interactions with nitric oxide synthases: controlling the right time, the right place, and the right amount of nitric oxide. American Journal of Physiology-Renal Physiology **285:**F178-F190.
170. **Ponnuswamy PK, Prabhakaran M, Manavalan P.** 1980. Hydrophobic packing and spatial arrangement of amino-acid-residues in globular-proteins. Biochimica Et Biophysica Acta **623:**301-316.
171. **Fraser ME, James MNG, Bridger WA, Wolodko WT.** 1999. A detailed structural description of Escherichia coli succinyl-CoA synthetase. Journal of Molecular Biology **285:**1633-1653.
172. **Wolodko WT, Fraser ME, James MNG, Bridger WA.** 1994. The crystal-structure of cussinyl-CoA synthetase from *Escherichia coli* at 2.5 angstrom resolution. Journal of Biological Chemistry **269:**10883-10890.
173. **Stancik LM, Stancik DM, Schmidt B, Barnhart DM, Yoncheva YN, Slonczewski JL.** 2002. pH-dependent expression of periplasmic proteins and amino acid catabolism in Escherichia coli. Journal of Bacteriology **184:**4246-4258.
174. **Walshaw DL, Wilkinson A, Mundy M, Smith M, Poole PS.** 1997. Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in *Rhizobium leguminosarum*. Microbiology-Uk **143:**2209-2221.
175. **Alexeev D, Alexeeva M, Baxter RL, Campopiano DJ, Webster SP, Sawyer L.** 1998. The crystal structure of 8-amino-7-oxononanoate synthase: A bacterial PLP-dependent, acyl-CoA-condensing enzyme. Journal of Molecular Biology **284:**401-419.

176. **Kerbarh O, Campopiano DJ, Baxter RL.** 2006. Mechanism of alpha-oxoamine synthases: identification of the intermediate Claisen product in the 8-amino-7-oxononanoate synthase reaction. Chemical Communications doi:10.1039/b511837a**:**60-62.
177. **Choi JS, Park NJ, Lim HK, Ko YK, Kim YS, Ryu SY, Hwang IT.** 2012. Plumbagin as a new natural herbicide candidate for Sicyon angulatus control agent with the target 8-amino-7-oxononanoate synthase. Pesticide Biochemistry and Physiology **103:**166-172.
178. **Nudelman A, Marcovici-Mizrahi D, Nudelman A, Flint D, Wittenbach V.** 2004. Inhibitors of biotin biosynthesis as potential herbicides. Tetrahedron **60:**1731-1748.
179. **Dhanasekaran M, Negi S, Imanishi M, Suzuki M, Sugiura Y.** 2008. Effects of bulkiness and hydrophobicity of an aliphatic amino acid in the recognition helix of the GAGA zinc finger on the stability of the hydrophobic core and DNA binding affinity. Biochemistry **47:**11717-11724.
180. **Sengupta D, Kundu S.** 2012. Do topological parameters of amino acids within protein contact networks depend on their physico-chemical properties? Physica a-Statistical Mechanics and Its Applications **391:**4266-4278.
181. **Chou PY, Fasman GD.** 1978. Empirical predictions of protein conformation. Annual Review of Biochemistry **47:**251-276.
182. **Luo W, Liu YP, Zhu XC, Zhao WJ, Huang L, Cai J, Xu ZN, Cen PL.** 2011. Cloning and characterization of purine nucleoside phosphorylase in *Escherichia coli* and subsequent ribavirin biosynthesis using immobilized recombinant cells. Enzyme and Microbial Technology **48:**438-444.
183. **Grenha R, Levdikov VM, Fogg MJ, Blagova EV, Brannigan JA, Wilkinson AJ, Wilson KS.** 2005. Structure of purine nucleoside phosphorylase (DeoD) from Bacillus anthracis. Acta Crystallographica Section F-Structural Biology and Crystallization Communications **61:**459-462.
184. **Hormoz S.** 2013. Amino acid composition of proteins reduces deleterious impact of mutations. Scientific Reports **3**.
185. **Merlin C, Gardiner G, Durand S, Masters M.** 2002. The *Escherichia coli metD* locus encodes an ABC transporter which includes Abc (MetN), YaeE (MetI), and YaeC (MetQ). Journal of Bacteriology **184:**5513-5517.
186. **Santic M, Abu Kwaik Y.** 2013. Nutritional virulence of *Francisella tularensis*. Frontiers in Cellular and Infection Microbiology **3**.