Large Direct Repeats Flank Genomic Rearrangements between a New Clinical Isolate of *Francisella tularensis* subsp. *tularensis* A1 and Schu S4

Ufuk Nalbantoglu  
*University of Nebraska-Lincoln*

Khalid Sayood  
*University of Nebraska-Lincoln*, ksayood1@unl.edu

Michael P. Dempsey  
*Armed Forces Institute of Pathology, Washington, D.C.*

Peter C. Iwen  
*University of Nebraska Medical Center*

Stephen C. Francesconi  
*Naval Medical Research Center, Silver Spring, Maryland*

See next page for additional authors

Follow this and additional works at: [https://digitalcommons.unl.edu/electricalengineeringfacpub](https://digitalcommons.unl.edu/electricalengineeringfacpub)

Part of the [Electrical and Computer Engineering Commons](https://digitalcommons.unl.edu/electricalengineeringfacpub)

Nalbantoglu, Ufuk; Sayood, Khalid; Dempsey, Michael P.; Iwen, Peter C.; Francesconi, Stephen C.; Barabote, Ravi D.; Xie, Gary; Brettin, Thomas S.; Hinrichs, Steven H.; and Fey, Paul D., "Large Direct Repeats Flank Genomic Rearrangements between a New Clinical Isolate of *Francisella tularensis* subsp. *tularensis* A1 and Schu S4" (2010). *Faculty Publications from the Department of Electrical and Computer Engineering*. 107.  
[https://digitalcommons.unl.edu/electricalengineeringfacpub/107](https://digitalcommons.unl.edu/electricalengineeringfacpub/107)

This Article is brought to you for free and open access by the Electrical & Computer Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Department of Electrical and Computer Engineering by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Authors
Ufuk Nalbantoglu, Khalid Sayood, Michael P. Dempsey, Peter C. Iwen, Stephen C. Francesconi, Ravi D. Barabote, Gary Xie, Thomas S. Brettin, Steven H. Hinrichs, and Paul D. Fey
Large Direct Repeats Flank Genomic Rearrangements between a New Clinical Isolate of *Francisella tularensis* subsp. *tularensis* A1 and Schu S4

Ufuk Nalbantoglu2, Khalid Sayood2, Michael P. Dempsey3, Peter C. Iwen1, Stephen C. Francesconi6, Ravi D. Barabote3,4, Gary Xie3,4, Thomas S. Brettin7, Steven H. Hinrichs1, Paul D. Fey1*

1 Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, United States of America, 2 Department of Electrical Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska, United States of America, 3 Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America, 4 DOE Joint Genome Institute, Walnut Creek, California, United States of America, 5 Division of Microbiology, Armed Forces Institute of Pathology, Washington, D.C., United States of America, 6 Naval Medical Research Center, Silver Spring, Maryland, United States of America, 7 Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States of America

Abstract

*Francisella tularensis* subspecies *tularensis* consists of two separate populations A1 and A2. This report describes the complete genome sequence of NE061598, an *F. tularensis* subspecies *tularensis* A1 isolated in 1998 from a human with clinical disease in Nebraska, United States of America. The genome sequence was compared to Schu S4, an *F. tularensis* subspecies *tularensis* A1a strain originally isolated in Ohio in 1941. It was determined that there were 25 nucleotide polymorphisms (22 SNPs and 3 indels) between Schu S4 and NE061598; two of these polymorphisms were in potential virulence loci. Pulsed-field gel electrophoresis analysis demonstrated that NE061598 was an A1a genotype. Other differences included repeat sequences (n = 11 separate loci), four of which were contained in coding sequences, and an inversion and rearrangement probably mediated by insertion sequences and the previously identified direct repeats I, II, and III. Five new variable-number tandem repeats were identified; three of these five were unique in NE061598 compared to Schu S4. Importantly, there was no gene loss or gain identified between NE061598 and Schu S4. Interpretation of these data suggests there is significant sequence conservation and chromosomal synteny within the A1 population. Further studies are needed to determine the biological properties driving the selective pressure that maintains the chromosomal structure of this monomorphic pathogen.

Citation: Nalbantoglu U, Sayood K, Dempsey MP, Iwen PC, Francesconi SC, et al. (2010) Large Direct Repeats Flank Genomic Rearrangements between a New Clinical Isolate of *Francisella tularensis* subsp. *tularensis* A1 and Schu S4. PLoS ONE 5(2): e9007. doi:10.1371/journal.pone.0009007

Editor: Niyaz Ahmed, University of Hyderabad, India

Received: September 6, 2009; Accepted: January 7, 2010; Published February 3, 2010

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

Funding: This work was supported by a grant from the Department of Defense, Defense Advanced Research Program Agency (award W911NF0510275) to S.H.H. The opinions expressed in this manuscript are the private views of the authors and are not the official views of the Armed Forces Institute of Pathology, the Department of the Army, the United States Air Force, or the Department of Defense. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pfey@unmc.edu

Introduction

*Francisella tularensis* is a highly pathogenic gram-negative cocco-bacillus that is the causative agent of tularemia, commonly referred to as “rabbit fever.” The large majority of disease is ulceroglandular in nature and can be traced to contact with an infected host (e.g. rabbit or cat) or vector (e.g. tick or mosquito); however more serious forms of disease such as pneumonic tularemia can be life-threatening, and therefore *F. tularensis* is considered a potential biowarfare agent. There are three recognized subspecies of *F. tularensis* including *tularensis* (commonly referred to as type A), *holarctica* (commonly referred to as type B), and *mediasiatica* as well as a closely related species *F. novicida*. These subspecies are associated with important geographic differences in their distribution with *F. tularensis* *holarctica* found throughout the northern temperate regions of both hemispheres whereas subspecies *tularensis* is found primarily in North America. In addition, the population of *F. tularensis* subspecies *tularensis* consists of two major, geographically isolated clades, A1 and A2 [1,2]. The A2 population has been isolated in the western United States whereas the A1 population is found east of the Rocky Mountains, primarily in the Ozark mountain regions of Missouri, Oklahoma and Arkansas. The genomes of two *F. tularensis* subspecies *tularensis* A1 isolates (Schu S4 and FSC198) have recently been sequenced; FSC198 was isolated from Slovakia in 1986 whereas Schu S4, an often-utilized virulent laboratory strain, is a clinical isolate obtained from Ohio in 1941 [3,4]. In addition, a draft sequence of a separate *F. tularensis* subspp. *tularensis* A1 isolate, FSC033, was also recently published [5]. FSC033 was isolated from a squirrel in Georgia, USA. Genomic comparisons between FSC198 and Schu S4 revealed remarkable sequence conservation; only 8 SNP and three variable number tandem repeat (VNTR) differences were noted [3]. Chaudhri et al. [3] have suggested that the close similarity between FSC198 and Schu S4 indicated that the FSC198 strain may have derived from Schu S4. Preliminary analysis between a recent human clinical isolate of *F. tularensis* subspp. *tularensis* obtained in 1998 in Nebraska and Schu S4 revealed distinguishing characteristics [6]. This presented an
opportunity to further examine the genomic diversity within the A1 population, and therefore, the complete sequence of a *F. tularensis* subspecies *tularensis* A1 isolate NE061598 was determined. The genomes of the four A1 isolates that have been fully or partially sequenced (SchuS4, FSC198, NE061598 and FSC033) were compared in light of their temporal and spatial separation. This analysis demonstrated that the *F. tularensis* subspecies *tularensis* A1 population, as represented by these isolates, is highly clonal and displays a high degree of DNA sequence conservation and chromosomal synteny. The primary chromosomal differences between NE061598 and Schu S4/FSC198/FSC033 were due to rearrangements occurring between large direct repeats and insertion sequences.

**Results**

**General Features**

The genomic sequence of *Francisella tularensis* subspecies *tularensis* NE061598 (GenBank accession number CP001633 or at http://bioinf.unl.edu/NE061598genome) consists of a single circular chromosome of size 1,892,681 base pairs (bp). General characteristics of the NE061598 genome are shown in Table 1. Using pulsed-field gel electrophoresis, Kugeler et al have demonstrated the population of *F. tularensis* subspecies *tularensis* A1 can be divided into at least two separate groups, A1a and A1b [2]. Previous PFGE analysis of NE061598 using both *Pvu*I and BamHI suggested that it was a subtype A1a (data not shown) and [6]).

**Comparison to the Other Type A1 Strains**

The NE061598 genome sequence contains 65 bp more than the FSC190 sequence [3] and 94 bp less than the Schu S4 sequence [4]. Previous bioinformatic analysis of the FSC190 and Schu S4 genomes demonstrated that there were only eight single nucleotide polymorphisms (SNPs) and three VNTR differences between these two isolates [3]. Therefore, based on the known genomic similarity between Schu S4 and FSC190, NE061598 was compared with Schu S4 (Genbank accession number AJ749949 and the Refseq accession no. NC_006570). The regions of difference between Schu S4 and NE061598 were divided into 2 types: small tandem repeats (Table 2) and rearrangements (Table 3). The VNTR’s listed in Table 2 accounted for the difference in size between the two isolates. Table 2 consists of known VNTR markers used previously for MLVA analysis [6,7] in addition to five newly identified tandem repeat differences (VNTR 1–5) discovered between NE061598 and Schu S4. Only one of the five new VNTRs was found within an open reading frame.

Compared to the published Schu S4 genome sequence, NE061598 had 25 polymorphisms (22 SNPs and 3 indels; Table 4). All SNP and indel differences were confirmed by repeat sequence analysis. Of the 22 confirmed SNPs, 6 were synonymous SNPs, 5 were intergenic SNPs, and 11 were nonsynonomous. There were no SNPs in rRNA or tRNA genes. Petrozino et al. [8] have identified 268 virulence genes associated with *F. tularensis*. Comparing NE061598 to Schu S4, only two of the proposed virulence genes identified by Petrozino et al. [8] were determined to have SNPs. These include a ferrous iron transport protein (FTT0249) and 2-isopropylmalate synthase (FTT0252). Both contain non-synonymous polymorphisms that result in a non-conservative amino acid substitution; it is unknown whether these mutations have any effect on protein function.

Apart from the rearrangements and polymorphisms, the main reason for the remaining genomic differences in composition and length between NE061598 and Schu S4 were found to be due to differences in the VNTR’s. VNTR analysis has been very useful in epidemiological and population analyses of *Francisella* [6,7]. Of the twelve tandem repeats that have a unique number of repeats in NE061598 in comparison to Schu S4, 7 (FmM5, FmM9, FmM10, FmM21, VNTR-1, VNTR-2, and VNTR-4) occur in intergenic regions, and the remaining 4 (FmM2, FmM3, FmM6, and VNTR-3) are in coding regions (Table 3). Of these four, one repeat in the gene for a hypothetical protein (FmM2; FTT1800c [Schu S4] and NE6158_10490 [NE061598]) inserted two amino acids into the translated sequence. Another repeat in a gene for a hypothetical protein (VNTR3; FTT0877c [Schu S4]) resulted in a premature stop codon in NE061598. An insertion of 7 amino acids was observed in an ATP-dependent DNA helicase protein in NE061598 compared to Schu S4 (FTT1395c [Schu S4] and NE6158_07740 [NE061598]). Lastly, one tandem repeat difference (FmM3) appeared to eliminate a premature stop codon in a pseudogene in Schu S4 (TPR repeat region protein; FTT0294 [Schu S4] and NE6158_0160 [NE061598]). This difference resulted in a deletion of the repeat NKDNKD. Importantly, NE061598 does not encode any unique genes that are not found in Schu S4.

**Chromosomal Rearrangements**

In order to describe the chromosomal rearrangements between NE061598 and Schu S4, the genomes were divided into six local collinear blocks (LCBs) as shown in Table 3 and Figure 1. The initial division was performed using the genome rearrangement analysis tool SPRING (Sorting Permutation by Reversals and block-interchange) [9]. These analyses demonstrated that the first, third and sixth LCBs are conserved whereas the second LCB is inverted in NE061598 with respect to Schu S4. The fourth and fifth LCBs are rearranged (Table 3 and Figure 1). These data are consistent with a previous comparison of two type A strains of *Francisella tularensis* subspecies *tularensis*, WY96 (A2) and Schu S4 (A1), which demonstrated the presence of various genome rearrangements due to inversions and block rearrangements mediated by insertion sequences [10]. The remaining LCBs have flanking duplicated regions. Several insertion elements were also observed juxtaposed to the flanking regions of the LCBs (Table 3) that might promote further chromosomal rearrangements during strain evolution.
divergence. For example, the second LCB is inverted between NE061598 and Schu S4. This inversion is hypothesized to be due to 2969 bp long flanking regions on each side of the inverted region that are reverse complements of each other. These flanking regions are comprised of one ISFtu2 and two additional ISFtu1 insertion sequence elements.

The rearrangements in LCBs four and five are most probably mediated by two large duplicated regions (DR1 and DR2) previously discussed in the genome report comparing WY96 and Schu S4 [10]. These duplicated regions include the Francisella Pathogenicity Island (FPI) containing the iglABCD operon [11] required for intramacrophage growth. This operon is regulated by the transcription factor MglA that has been shown to regulate a number of virulence factors [12]. These two regions (33,910 bp) occur at locations 1,374,336–1,408,246 (DRI) and 1,767,671–1,801,581 (DRII) in Schu S4. In addition, a 5358 bp segment of

**Table 2. VNTR markers and their differences between Schu S4 and NE061598.**

| VNTR Marker | Repeat motif | Repeat size (nt) | Genomic Location | Repeat copy no., strain SCHU S4 | Repeat copy no., strain NE061598 |
|-------------|--------------|-----------------|------------------|-------------------------------|-------------------------------|
| Ft-M1       | AAT          | 3               | I (– 76)         | 3                             | 3                             |
| Ft-M2       | TAGCTTTTTTTT | 10              | I (– 113)        | 2                             | 2                             |
| Ft-M3       | CTCAGGACCAA  | 12              | G (+1174)        | 2                             | 2                             |
| Ft-M4       | CTACCCATATA  | 6               | G (+67)          | 3                             | 3                             |
| Ft-M5       | TTTCTCAAGATTC | 16              | I (– 120)        | 2                             | 2                             |
| Ft-M6       | TTTCTACAATCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M7       | TTTCTACAATCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M8       | TTTCTACAATCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M9       | TTTCTACAATCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M10      | TTTCTACAATCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M11      | AATTATAAAT   | 10              | I (– 113)        | 5                             | 5                             |
| Ft-M12      | TAGCTTTTTTTT | 10              | I (– 113)        | 2                             | 2                             |
| Ft-M13      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M14      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M15      | AATTATAAAT   | 10              | I (– 113)        | 5                             | 5                             |
| Ft-M16      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M17      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M18      | AATTATAAAT   | 10              | I (– 113)        | 5                             | 5                             |
| Ft-M19      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M20      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M21      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M22      | AATTATAAAT   | 10              | I (– 113)        | 5                             | 5                             |
| Ft-M23      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M24      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| Ft-M25      | TTTCTCAAGATTCTTTT | 16          | I (– 120)        | 2                             | 2                             |
| VNTR-1      | CAAAGACA     | 8               | I (+222)         | 2                             | 2                             |
| VNTR-2      | TTTATATAAG   | 11              | I (– 42)         | 3                             | 2                             |
| VNTR-3      | GAAAATAA     | 8               | G (+282)         | 1                             | 2                             |
| VNTR-4      | TCTCTCAATACTTTT | 16          | I (– 32)         | 1                             | 1                             |
| VNTR-5      | AAAATGCATATATAGCCAGATTTTAG | 29     | I (– 32)         | 1                             | 1                             |

*FtM1-FtM25 VNTR markers as previously reported by Johansson et al. [7]. New VNTR polymorphisms identified in this study are listed as VNTR1 through VNTR-5.

Indicates repeat size in nucleotides.

"G" indicates that the repeat is located within an open reading frame (genic) whereas "I" indicates that the repeat is located within an intergenic region. Distance to predicted translation start site is indicated in nucleotides. “r” or “—” indicates that the translation start site is downstream or upstream of repeat motif, respectively (as reported by Johansson et al. [7]).

doi:10.1371/journal.pone.0009007.t002

**Table 3. Description of six local collinear blocks (LCBs) between NE061598 and Schu S4.**

| LCB Type | NE061598 Position | Schu S4 position |
|----------|-------------------|------------------|
| 1        | Conserved         | 1-352156         | 1-352087         |
| 2        | Inversion         | 352157-381876    | 381807-352088    |
| 3        | Conserved         | 381877-1312701   | 381808-1312781   |
| 4        | Rearrangement     | 1312702-1700690  | 1379901-1767877  |
| 5        | Rearrangement     | 1700691-1767602  | 1307424-1374335  |
| 6        | Conserved         | 1767603-1892681  | 1767671-1892775  |

doi:10.1371/journal.pone.0009007.t003
the duplicated regions between the 208th and 5565th bases of the duplicated regions, was also duplicated at positions 1,307,425 bp–
1,312,781 bp in Schu S4. No structural alterations in the
duplicated regions, was also duplicated at positions 1,307,425 bp–
1,312,781 bp in Schu S4. No structural alterations in the

duplicated regions, was also duplicated at positions 1,307,425 bp–
1,312,781 bp in Schu S4. No structural alterations in the

duplicated regions, was also duplicated at positions 1,307,425 bp–
1,312,781 bp in Schu S4. No structural alterations in the

duplicated regions, was also duplicated at positions 1,307,425 bp–
1,312,781 bp in Schu S4. No structural alterations in the

The location of DRI and DRII in both Schu S4 and NE061598 are shown in figures 2A and 2B. In addition, DRIII (III, red) is shown which contains the aforementioned 5358 bp long segment of the duplicated regions [10]. Relating these regions to the LCBs noted in Figure 2, DRII is contained in LCB 6 while the other components are contained in LCBs four and five. The rearrangement can be explained as an edit operation in which one block with a partially duplicated flanking region is replaced by another block having DRI as the flanking region (Figure 3). Consequently, DRIII (III, red) is contained in LCB 6 while the other components are contained in LCBs four and five. The rearrangement was determined to be the most compositionally different segments of the genome using the Alien Hunter program [13].

While it is known that IS elements are significantly involved in intrachromosomal rearrangement, only one rearrangement associated with insertion sequences was observed when comparing NE061598 to Schu S4. The most parsimonious transformation using the rearrangements and inversions of the collinear blocks involved an inversion of LCB2 and the edit process discussed in Figure 2.

Comparison of NE061598 and Schu S4 with the Draft Sequence of F. tularensis Subsp. Tularensis FSC033

Kugeler et al have demonstrated the population of F. tularensis subsp. tularensis A1b is associated with higher mortality rates [2]. A prototype A1b isolate, FSC033, has recently been partially sequenced [2, 5]. In order to perform preliminary genomic comparisons between FSC033, NE061598 and Schu S4, the genomes were divided into 10 LCBs as described above (Figure 3). This analysis found that the only major difference between FSC033 and NE061598/Schu S4 was the rearrangement of LCB2.
(Figure 3). The genomic organization of FSC033 surrounding DRI and DRII as shown in Figures 1 and 2 was similar to the Schu S4 genomic arrangement. Although few significant differences were observed regarding the genomic synteny between FSC033 (subtype A1b) and NE061598/Schu S4 (subtype A1a), SNP analysis indicated that 123 SNPs and 8 indels were detected between NE061598 and FSC033.

Transposable Elements

Seven different types (n = 75) of IS elements were found within NE061598 (Table 5). In addition to 50 ISFtu1 elements, NE061598 contains 16 ISFtu2 elements (of which one flanks the inverted LCB 2), 3 ISFtu3 and ISFtu6 elements, and one copy each of ISFtu4, ISFtu5 and ISSod13. All of the insertion sequences found in NE061598 are also present in Schu S4.

Discussion

Due to the remarkable sequence conservation between Schu S4 and FSC198 [3], speculation was made that these two isolates may have the same origin. Therefore, we proposed to sequence a separate virulent isolate of F. tularensis subsp. tularensis A1 and compare it with Schu S4 to evaluate the issue of sequence divergence over time. NE061598 was isolated in Nebraska in 1998 from the blood of a patient with ulceroglandular tularemia, Schu S4 was derived in 1941 and FSC198 was isolated in 1986. The availability of a recent clinically virulent isolate of F. tularensis subsp. tularensis A1 isolate obtained in the mid-western portion of the United States provided the opportunity for an in-depth sequence comparison with other A1 isolates. Because of the significant temporal separation (45 years) between Schu S4 and NE061598, the sequence conservation between these two isolates was unexpected. Even though VNTR analysis yielded 11 distinct polymorphisms (see Table 2), analysis of the entire genome only yielded 25 additional SNPs/indels. The most significant difference detected was an inversion associated with LCB 2 and rearrangements associated with LCBs 4 and 5 (see Figures 1 and 2); both events were predictably mediated through IS element recombination (LCB 2) or rearrangement mediated by large duplicated regions (LCBs 4 and 5). Significantly, there was no net gain (or loss) of genes within the NE061598 genome in relationship to Schu S4. These data may suggest that the minimal differences observed...
in pulsed-field RFLP patterns of the *F. tularensis* subsp. *tularensis* A1 population may be due to IS- or direct repeat-mediated rearrangements and is not due to the acquisition of new genes [1,2,6]. Furthermore, these data support the notion that this highly monomorphic pathogen [14] may have undergone a recent population bottleneck which may be related to its specific host preference (e.g. lagomorphs, humans) and vectors (e.g. ticks). The further elucidation of the natural reservoir, hosts, and vectors of *F. tularensis* may lead to novel hypotheses of the selective pressure of this A1 population.

Due to the lack of genetic diversity noted within the *F. tularensis* subsp. *tularensis* A1 population, phylogenetic and population structure analyses are problematic and biased especially due to the rapid evolution of VNTR loci and lack of sensitivity of other methodologies [14,15]. However, whole genome SNP analysis has been successful at probing the population structure of highly monomorphic pathogens such as *B. anthracis* and other highly virulent pathogens [14,16]. A recent report using a variety of SNP analyses identified 11 subclades within *F. tularensis* subsp. *holarctica* [15]. Phylogenetic analysis suggested that *F. tularensis* subsp. *holarctica* originated from North America and was introduced multiple times into Eurasia. Further studies need to be performed to delineate the complicated population structure of *F. tularensis* subsp. *tularensis* A1 (both A1a and A1b) and its relationship to the *F. tularensis* subsp. *tularensis* A2 population. Data provided in our study may yield canonical SNPs that provide lineage- or strain-specific phylogeny within this subspecies. The utility of these unique SNPs will be evaluated using large repositories of *F. tularensis* subspecies. Lastly, our study suggests that the genomic organization between the A1a and A1b populations may not significantly differ; however, preliminary SNP/indel analysis provides evidence that the increased virulence observed with A1b strains may reside in specific nucleotide alterations and not gene acquisition or loss.

**Materials and Methods**

**Genome Sequencing of NE061598**

The genome coverage determined at the end of the draft-sequencing phase was 11x and resulted in 19 contigs mapped into 12 scaffolds. The draft phase involved two clone libraries, one small insert library (2200 bp average insert size) and one medium insert library (6289 bp average insert size). Paired end shotgun reads from each of these libraries produced 12218 and 13156 reads respectively. During the finishing phase, seven transposon libraries were created and sequenced to assist with repeat resolution. Four PCR shatter libraries were created and sequenced to assist with hard stops. An additional 528-primer walk reads were created as needed to address low quality regions of the draft assembly. The final genome was a complete genome with no gaps consisting of 1892901 base pairs. The overall average error rate of the finished genome was less than one error in 100,000 bp. The total number of reads used in the final assembly was 25,531.

**Annotation**

The open reading frames of Schu S4 strains were extracted and each ORF was searched for in the NE061598 chromosome using the standard Smith-Waterman algorithm [17]. The hits having accuracy higher than 98% identity were detected as initial annotations. Next, the NCBI annotation pipeline (http://www.ncbi.nlm.nih.gov/genome/guide/build.html) was employed and any missed ORFs were extracted from the output of this pipeline. Eliminating the ORFs and overlapping genes that had already been recognized, protein BLAST searches were performed on filtered predictions of the pipeline.

**Insertion Sequence Element Mapping**

Annotated insertion sequence elements that are specific to *F. tularensis* were detected in the NE061598 genome using Smith-Waterman alignment [17].

---

**Table 5. IS element found in NE061598 compared to Schu S4.**

| IS Elements      | Number in NE061598 | Number in Schu S4 |
|------------------|--------------------|-------------------|
| ISFtu1 (IS630 family) | 50                 | 50                |
| ISFtu2           | 16                 | 16                |
| ISFtu3 (ISNCY family, ISHpal-IS1016) | 3                  | 3                 |
| ISFtu4 (IS982 family) | 1                  | 1                 |
| ISFtu5 (ISf family) | 1                  | 1                 |
| ISFtu6 (IS1595 family) | 3                  | 3                 |
| ISSod13          | 1                  | 1                 |
| **TOTAL**        | **75**             | **75**            |

doi:10.1371/journal.pone.0009007.t005
SNP Discovery

SNP polymorphisms between Schu S4 and NE061598 were discovered using the SNPFinder program of Los Alamos Laboratories (http://snpsfinder.lanl.gov/UsersManual/index.html). SNP predictions were then curated manually using BLAST (with parameters match: 1 mismatch: −4 existence and extension gaps: −1).

Genome Rearrangement Discovery

In order to determine the local collinear blocks (LCB), the SPRING tool [7] was utilized. The SPRING parameters for LCB discovery included the following. Block search mode: reversals (inversions) plus block interchange mode; minimum multi-MUM length: 21 bp (closest integer to log2 [1892 Kbp], where 1892 is the average genome length); minimum LCB length: 63 bp (3 x minimum multi-MUM); chromosome type: linear. The boundaries of the rearrangements were further optimized using BLAST (expect threshold: 10; word size: 64; match score: 1; mismatch score: −4; existence and extension gaps: −1) around the 10 Kb flanking regions of LCB ends.

Pulsed-Field Gel Electrophoresis

Agarose embedded DNA was prepared and digested with Pmel and BsmHI as previously described [18]. RFLP analysis was performed using BioNumerics software (Applied Maths).

Author Contributions

Conceived and designed the experiments: UN KS SHH PDF. Performed the experiments: UN KS RDB GX TSB PDF. Analyzed the data: UN KS RDB GX TSB SHH PDF. Contributed reagents/materials/analysis tools: UN KS SHH PDF. Wrote the paper: UN KS SHH PDF.

References

1. Farlow J, Wagner DM, Dukerich M, Stanley M, Chu M, et al. (2005) Francisella tularensis in the United States. Emerg Infect Dis 11: 1035–1841.
2. Kugeler KJ, Mead PS, Janusz AM, Staples JE, Kubota KA, et al. (2009) Molecular Epidemiology of Francisella tularensis in the United States. Clin Infect Dis 48: 863–870.
3. Chaudhuri RR, Ren CP, Desmond L, Vincent GA, Silman NJ, et al. (2007) Genome sequencing shows that European isolates of Francisella tularensis subspecies tularensis are almost identical to US laboratory strain Schu S4. PLoS ONE 2:e352.
4. Larsson P, Oyston PC, Chaim P, Chu MC, Duffield M, et al. (2005) The complete genome sequence of Francisella tularensis, the causative agent of tularemia. Nat Genet 37: 153–159.
5. Champion MD, Zeng Q, Nix EB, Nano FE, Keim P, et al. (2009) Comparative genomic characterization of pathogenic A11 strain of Francisella tularensis subspecies tularensis. PLoS ONE 2:e947.
6. Nano FE, Zhang N, Cooley SC, Klose KE, Chengh KK, et al. (2004) A Francisella tularensis pathogenicity island required for intramacrophage growth. J Bacteriol 186: 6430–6436.
7. Brotcke A, Weiss DS, Kim CC, Chain P, Malfatti S, et al. (2006) Identification of MgI2-regulated genes reveals novel virulence factors in Francisella tularensis. Infect Immun 74: 6642–6655.
8. Vogler AJ, Birdsell D, Price LB, Bowers JR, Beckstrom-Sternberg SM, et al. (2009) Phylogography of Francisella tularensis: Global Expansion of a Highly Fit Clone. J Bacteriol.
9. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, et al. (2007) Global genetic population structure of Bacillus anthracis. Annu Rev Microbiol 61: 623–657.
10. Beckstrom-Sternberg SM, Auerbach RK, Godbole S, Pearson JV, Beckstrom-Sternberg JS, et al. (2007) Complete genomic characterization of a pathogenic A11 strain of Francisella tularensis subspecies tularensis. PLoS ONE 2:e947.
11. Chaudhuri RR, Ren CP, Desmond L, Vincent GA, Silman NJ, et al. (2007) Genome sequencing shows that European isolates of Francisella tularensis subspecies tularensis are almost identical to US laboratory strain Schu S4. PLoS ONE 2:e352.
12. Petrosino JF, Xiang Q, Karpathy S, Jiang H, Verrapragada S, et al. (2006) Chromosome rearrangement and diversification of Francisella tularensis revealed by the type B (OSU18) genome sequence. J Bacteriol 188: 6977–6985.
13. Lin YC, Lu CL, Lin YC, Tang CY (2006) SPRING: a tool for the analysis of genome rearrangement using reversals and block-interchanges. Nucleic Acids Res 34: W696–699.
14. Achtman M (2008) Evolution, population structure, and phylogeography of Francisella tularensis species. Microbiol Mol Biol Rev 72: 721–745.
15. Vogler AJ, Birdsell D, Price LB, Bowers JR, Beckstrom-Sternberg SM, et al. (2009) Phylogography of Francisella tularensis: Global Expansion of a Highly Fit Clone. J Bacteriol.
16. Vernikos GS, Parkhill J (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. Bioinformatics 22: 2296–2303.
17. Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62: 53–70.
18. Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147: 193–197.
19. Staples JE, Kabota KA, Chalcraft LG, Mead PS, Petersen JM (2006) Epidemiologic and molecular analysis of human tularemia, United States, 1964–2004. Emerg Infect Dis 12: 1113–1118.