Genomic Deletion Marking an Emerging Subclone of *Francisella tularensis* subsp. *holarctica* in France and the Iberian Peninsula

M. P. Dempsey  
*Armed Forces Institute of Pathology*

M. Dobson  
*Armed Forces Institute of Pathology*

C. Zhang  
*University of Nebraska-Lincoln*

M. Zhang  
*University of Nebraska-Lincoln*

C. Lion  
*Centre Hospitalier et Universitaire*

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

[http://digitalcommons.unl.edu/foodsciefacpub/174](http://digitalcommons.unl.edu/foodsciefacpub/174)
Genomic Deletion Marking an Emerging Subclone of Francisella tularensis subsp. holarctica in France and the Iberian Peninsula


Division of Microbiology, Armed Forces Institute of Pathology, Washington, D.C.; Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska; Department of Food Science Technology, University of Nebraska at Lincoln, Lincoln, Nebraska; Laboratoire de Bacteriologie, Centre Hospitalier et Universitaire, Nancy, France; Department of Animal Health, Faculty of Veterinary Medicine, University of León, León, Spain; Modernization Directorate, Office of the Air Force Surgeon General, Falls Church, Virginia; Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona; and Division of Infectious Diseases, Department of Clinical Microbiology, Umeå University, and Department of NBC Analysis, Swedish Defence Research Agency, Umeå, Sweden

Received 21 March 2007/Accepted 8 September 2007

Francisella tularensis subsp. holarctica is widely disseminated in North America and the boreal and temperate regions of the Eurasian continent. Comparative genomic analyses identified a 1.59-kb genomic deletion specific to F. tularensis subsp. holarctica isolates from Spain and France. Phylogenetic analysis of strains carrying this deletion by multiple-locus variable-number tandem repeat analysis showed that the strains comprise a highly related set of genotypes, implying that these strains were recently introduced or recently emerged by clonal expansion in France and the Iberian Peninsula.

The species Francisella tularensis is a gram-negative intracellular pathogen comprising four different subspecies: subsp. tularensis, subsp. holarctica, subsp. mediasiatica, and subsp. novicida. Of these four, the tularensis and holarctica subspecies are clinically significant for humans, with the tularensis subspecies causing a more serious disease in humans and with animal models (7, 25). In addition to virulence characteristics, F. tularensis subsp. tularensis and F. tularensis subsp. holarctica also display unique geographic distributions. F. tularensis subsp. holarctica is found in the Eastern and Western hemispheres, whereas F. tularensis subsp. tularensis appears to be limited to North America (7).

The basis for these apparent differences in ecological and virulence characteristics is not well understood. Comparative genomics studies show that genomic organization among the subspecies is distinct, with multiple rearrangements, translocations, inversions, and recombination events that have occurred during divergence of the subspecies (1, 6, 15, 20, 23). Nonetheless, the composition of the genomes is remarkably similar among the subspecies.

Though F. tularensis subsp. tularensis populations display a more limited geographic range, at least two different sublineages (AI and AII) have been detected in the Eastern and Western portions of the United States by at least two different genotyping methods (8, 13, 18, 24). In contrast, F. tularensis subsp. holarctica populations found throughout North America and the Palearctic ecozone of Eurasia show surprisingly little genetic diversity, leading many to speculate that these populations are derived from a relatively recent bottleneck or clonal expansion event (6, 18, 20–22, 25).

The dearth of genetic diversity in F. tularensis poses substantial epidemiological problems, particularly with F. tularensis subsp. holarctica, constraining molecular epidemiological analyses to the most rapidly evolving genetic markers, such as tandem repeats. Because these markers can be susceptible to homoplasy, additional genomic markers or events can greatly enhance both epidemiological and phylogenetic inference. With this problem in mind, we used our previously described shotgun DNA microarray (23) to search for F. tularensis subsp. holarctica strains for genetic diversity by comparative genome hybridization (CGH). Our strain collection included a set of 42 F. tularensis subsp. holarctica strains derived from an epidemic of tularemia in animals and humans in Spain and Southern France in 1997 to 1999 (5, 10), as well as three strains from the Czech Republic (Tu-28, Tu-29, and Tu-35) and one from the Russian Federation (Tu-42) (see Table S1 in the supplemental material). A set of 16 F. tularensis subsp. tularensis strains, an F. tularensis subsp. novicida strain, and a Francisella philomiragia strain that had previously been tested with this same array were also included as controls (23). DNA isolation, labeling, hybridization, washing, and data analyses were performed as we have previously described, with binary conversion of the ratios based on >2 standard deviations from the mean of each array (23, 26).

To validate performance of this microarray data set, the
binary-converted data were sorted to identify probes hybridizing specifically to *F. tularensis* subsp. *tularensis* strains (Fig. 1A) but not the *F. tularensis* subsp. *holarctica*, *F. tularensis* subsp. *novicida*, or *F. philomiragia* strains. This yielded 36 probes, which DNA sequence and contig analysis subsequently showed to all be derived from the previously identified RD* tularensis*1 to RD* tularensis*13, validating that the array performance was similar to that in our previous study (23). To screen the data for diversity within *F. tularensis* subsp. *holarctica*, the data were then sorted among the *F. tularensis* subsp. *holarctica* strains according to geography. As shown in Fig. 1B, four of the microarray probes hybridized to *F. tularensis* subsp. *holarctica* strains isolated from the United States, the Czech Republic, and Russia but not to the large set of *F. tularensis* subsp. *holarctica* strains derived from Spain and France (marked by blue squares at the leaves) but binary 1 for all other *F. tularensis* subsp. *holarctica* stains (red squares at leaves) and binary 1 for all *F. tularensis* subsp. *tularensis* strains and *F. philomiragia* strains (white rectangles at leaves).

To precisely map the endpoints of the RD* 23* deletion, PCR primers were designed within the FTT1005c and FTT1009 coding regions abutting the 1.8-kb RD* 23* contig (RD* 23*Forward, 5'-GTCTTTGTAGCAAATGCC-3'; RD* 23*Reverse, 5'-CGGACAGGCCTAAAATAGTG-3') and PCR was performed...
on *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica* strains. As shown in Fig. 3, *F. tularensis* subsp. *holarctica* strains showing the RD23 CGH pattern (Tu-2, Tu-4, Tu-7, and Tu-10) gave rise to PCR products of the same size, 1.38 kb, whereas all other strains produced the expected 2.97-kb amplicon, in common with the SCHU S4 strain.

DNA sequence analysis of the shorter, 1.38-kb fragment from strain Tu-19 showed that the RD23 deletion extends from the FTT1006 coding region to that of FTT1008, as predicted from the microarray data. The deletion begins in the central region of the FTT1006 (403 amino acids; hypothetical membrane protein) coding region at the Leu160 codon (SCHU S4 coordinate 1019329) and extends into the FTT1008c (54 amino acids; hypothetical membrane protein) coding region at the Leu23 (SCHU S4 coordinate 1020924). The protein encoded by FTT1006 shares similarity with several different proteins, specifically amidases. FTT1008 shares similarity with several different proteins, including transporter proteins and *flgJ* product-like amidases. FTT1007 and FTT1008 encode hypothetical proteins.

**RD23 is limited to strains from Spain and Southern France.**

To test whether the RD23 deletion is a clonal marker for an epidemic population, a real-time PCR assay was developed to specifically identify the RD23 deletion and subsequently used to test a large set of *F. tularensis* strains. The design of the assay was based on use of TaqMan fluorophore probes, which can be detected on the ABI 7900HT Fast real-time PCR system (Applied Biosystems, Foster City, CA) and Dyad (MJ Research, Reno, Nevada) thermocyclers according to the following cycling parameters: initial hold at 95°C for 2 min, 30 s; 30 cycles of 95°C for 30 s, 64°C for 1 min, and 72°C for 1 min; final extension at 72°C; and a final indefinite hold at 4°C. PCR products were electrophoresed on an agarose gel and stained with ethidium bromide. Strain designations are indicated above the respective lanes. The marker lane on the left contains the 1-kb ladder with sizes of the 1- to 5-kb fragments indicated to the left. (For more information on strain designations, see Table S1 in the supplemental material.)

The real-time (RT)-PCR assay was developed as a duplex assay, consisting of the leftward forward primer as the common forward primer, the leftward probe just inside the deletion and flanked by its reverse primer, and the rightward probe outside the right junction and flanked by its reverse primer. This scheme allows detection of the RD23 Left Probe in instances of an intact region and detection of the RD23 Right Probe in strains carrying the RD23 deletion. The RT-PCRs were performed in 25-μl reaction volumes containing 1× Universal Master Mix (Applied Biosystems), 400 nM of each primer, 600 nM of each probe, and 0.55 U of Platinum Taq (Invitrogen). The thermocycling conditions included 50°C for 2 min; a hold for 10 min at 95°C; 45 cycles of 95°C for 15 s, 56°C for 15 s, and 61°C for 1 min; and a final indefinite hold at 4°C. The RT-PCR was validated by first comparing a set of 90 strains, using the conventional PCR and the RT-PCR, where perfect correlation was observed (data not shown). We subsequently used the RT-PCR to screen a large strain collection comprising isolates of *F. tularensis* subsp. *tularensis* (n = 88), *F. tularensis* subsp. *holarctica* (n = 300), *F. tularensis* subsp. *novicida* (n = 7), *F. tularensis* subsp. *mediasiatica* (n = 4), and *F. philomiragia* (n = 6) strains.

Only 51 of 405 strains tested carried the RD23 deletion (see Table S1 in the supplemental material). All 51 of the RD23 deletion strains were *F. tularensis* subsp. *holarctica* strains, and 50 of the 51 RD23-positive strains were isolated from Spain or France. The six strains isolated from France were from Chateauneu (AFIP3 and F0295), St. Germaine (AFIP4), Lorraine (FR-LauR), Langres (FR-Syl-S), and Vosges (F0020), implying that RD23-positive strains may be widespread in France. The FR-LauR and FR-Syl-S strains were isolated in 1993 and 2000, respectively, with the FR-Syl-S strain being isolated from one of the first cases of human *F. tularensis* bacteremia reported in Southern Europe (12). However, the F0020 strain was isolated in 1952, implying that the RD23-positive strains may have been in France for some time.

The larger set of RD23-positive strains from Spain comprised isolates from the Valladolid, Palencia, Zamora, and Leon regions, representing a wide geographic distribution in this country. These strains were all isolated during a narrow time frame of 1997 to 1999, coinciding with an epidemic of tularemia found in humans and animals during that time (11). The only exception of a strain carrying RD23 that was isolated outside of Spain or France was F0228, which was isolated in Uppsala, Sweden. Thus, with the exception of the F0228 isolate, the RD23 allele appears to be concentrated among strains isolated from the Iberian Peninsula and France.

**Strains carrying RD23 comprise a diverging clonal complex.**

Many of the RD23-positive isolates in our collection from Spain have previously been shown to share a close genetic
relationship by amplified fragment length polymorphism and other genotyping methods (5, 9). If the RD23 deletion is indeed a clonal marker of an epidemic population, then we would expect that the strains from France carrying RD23 should also be genetically related. The RD23-positive strains were therefore genotyped by multiple-locus variable-number tandem repeat analysis (MLVA) and normalized to MLVA data generated previously (8, 13). MLVA was performed as described previously (13) using primers amplifying regions Ft-M1 through Ft-M25. All reverse primers were labeled with IRD800 (Li-Cor, Lincoln, NE), and the reactions were resolved on 41-cm gels using a Li-Cor 4000L automated DNA sequencer. Control reactions were run using the SCHU S4 and LVS strains. Sizes of MLVA products were predicted from a combination of known SCHU S4 and LVS sizes, the 50- to 350-bp and 50- to 700-bp IRD800 size standards (Li-Cor), and a 1-bp sequencing ladder. The data from each strain were tabulated with the previously published MLVA data (13) run on an ABI 377 system (Applied Biosystems). Data from the two data sets were converted to numeric base pair values and normalized for allele size by the NAU Laboratory. An unweighted, midpoint-rooted, neighbor-joining tree was created using MEGA 3.1 software (14).

As shown in Fig. 4, all strains carrying the RD23 deletion comprised a single cluster based on MLVA genotypes. All strains in this cluster shared a unique 466-bp allele at M24 that was not observed in any other strain, supporting the conclusion that they are clonally related. Divergence within the cluster was due to multiple alleles at the M3, M6, M8, and M10 loci. Polymorphism at the M3 and M6 loci alone accounts for six of the nine different MLVA genotypes in the RD23 cluster. These loci are among the most highly polymorphic of the MLVA loci examined to date and would be expected to show some differentiation even within a clonal complex (13). Polymorphic alleles at the M8 and M10 loci within the RD23 cluster were detected only in strain F0228 (unique allele at M8) and strain Tu-31 (unique allele at M10). Thus, strains carrying the RD23 deletion appear to form a clone complex with divergence largely marked by events at the most rapidly evolving loci. The most divergent member of the RD23 clade was strain F0228, which carries unique alleles at M3, M6, and M8. The F0228 strain is an interesting case because it is the only isolate carrying RD23 which was not originally isolated from the Iberian Peninsula or France. It was isolated from a skin lesion of a human with tularemia in Uppsala, Sweden, in the year 2000 during a large epidemic in that country but was the only isolate among 28 tested from that epidemic that carries RD23. Further surveillance using a combination of the RD23 deletion and the MLVA genotypes identified in our study will now provide a
basis on which to further track the distribution and spread of this apparent clone complex. Within the *F. tularensis* holarctica subsp., genetic variation is scarce, even among strains from unlinked geographies. This lack of diversity is perplexing given that the *F. tularensis* subsp. holarctica genome has undergone substantial rearrangements with respect to the *F. tularensis* subsp. *tularensis* genome (6, 20, 21). MLVA studies also bear this out, showing little allelic diversity in *F. tularensis* subsp. *holarctica* populations compared to *F. tularensis* subsp. *tularensis* populations (13). These findings have been taken to indicate that the *F. tularensis* subsp. *holarctica* population may have recently been through a global evolutionary bottleneck or, more likely, that a clone has emerged and spread globally within a relatively recent time frame.

Our study has now shown that a distinct but highly related clone complex of *F. tularensis* subsp. *holarctica*, carrying the distinct RD23 genomic signature, has emerged within France and the Iberian Peninsula. Because tularemia has not commonly been reported in Spain and Southern France, we believe the isolation of highly related strains carrying RD23 is a consequence of recent epidemic spread of a single founding population. Further differentiation of the clone during its spread across the peninsula likely resulted in the eight distinct genotypes that are observed within the RD23 clade of the MLVA tree (Fig. 4). The RD23 strain F0228 appears unusual among the other RD23 strains. It was isolated outside Spain or France, and it bears a relatively distantly related MLVA genotype. It is unclear whether F0228 share ancestry with isolates from France and Spain or if RD23 occurred independently in a separate *F. tularensis* subsp. *holarctica* subclone.

Among isolates from France and Spain, phylogenetic inference from the clustering shown in Fig. 4 implies that two primary subpopulations carrying the RD23 signature are circulating. The F0020 strain, isolated from France in 1993, and the F0295 strain, isolated from France in 1952, are within one of these two large subpopulations. If a founder effect can explain the relative abundance of this subpopulation, then the F0295 strain implies that the RD23 clone or its immediate ancestor has been present in the region for >50 years. Although the exact origin of the clone is unknown, its emergence in Northwestern Spain is now well documented. The RD23 marker implies that the RD23 clone or its immediate ancestor is ancestral to all other RD23 strains. It was isolated outside Spain or France, and its genome is evolving over time. The RD23 deletion itself does not include any obvious genes associated with virulence functions, but the adjacent genomic region contains several pseudogenes and genes that have been inactivated by multiple stop codons. One of these degraded coding regions encodes a member of the Mip family (macrophage infectivity potentiator), suggesting some type of association of the region with virulence. Clearly more study will be necessary to determine if the RD23 deletion itself influences any of the apparent virulence characteristics of the clone complex.

We thank the Centers for Disease Control and Prevention in Fort Collins, CO, for access to DNAs. This work was supported in part by grant 1R21AI057755-01 to A.K.B. from the National Institutes of Health and a grant from the Homeland Security Advanced Research Projects Agency Bioinformatics and Assays Development Program to P.K. and D.M.W.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the U.S. Air Force, the Department of Defense, or the U.S. Government.

REFERENCES


15. Larsson, P., P. O. Ostrop, P. Chain, M. C. Chu, M. Duffield, H. H. Fuxelius, E. Garcia, G. Halltorp, D. Johansson, K. E. Isherwood, P. D. Karp, E. Larsson, Y. Liu, S. Michell, J. Prior, R. Prior, S. Malfatti, A. Sjostedt, K. Orduna, and A. Rodriguez-Torres. 2003. Genotyping of *Francisella tularensis* subsp. *holarctica* strains (7). The outbreak of tularemia in northwest Spain, which occurred between 1997 and 1998, is the first recorded outbreak in that region, resulting in over 500 human cases (11). It is not yet clear whether the RD23 clone complex is a hypervirulent clone or whether other factors may explain its epidemic spread. Although a gain of virulence characteristics is usually associated with allelic diversity or gene acquisition, there are well-documented instances of specific deletions enhancing virulence of other species (2, 4, 16, 17) and of genome decay resulting in emergence of highly virulent species (3). The RD23 deletion itself does not include any obvious genes associated with virulence functions, but the adjacent genomic region contains several pseudogenes and


