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*

Martien Broekhuijsen,[1] Pär Larsson,[2] Anders Johansson,[2,3,4] Mona Byström,[2] Ulla Eriksson,[2]
Eva Larsson,[2] Richard G. Prior,[5] Anders Sjöstedt,[3] Richard W. Titball,[5,6]
and Mats Forsman[2]*

TNO Prins Maurits Laboratory, Rijswijk, The Netherlands[1]; Swedish Defence Research Agency[2] and Department of Clinical Microbiology, Bacteriology,[3] and Department of Clinical Microbiology, Infectious Diseases,[4] Umeå University, Umeå, Sweden; and Defence Science and Technology Laboratory, CBS Porton Down, Salisbury,[5] and Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London,[6] United Kingdom

Francisella tularensis **is a potent pathogen and a possible bioterrorism agent. Little is known, however, to explain the molecular basis for its virulence and the distinct differences in virulence found between the four recognized subspecies,** *F. tularensis* **subsp.** *tularensis*, *F. tularensis* **subsp.** *mediasiatica*, *F. tularensis* **subsp.** *holarctica*, **and** *F. tularensis* **subsp.** *novicida*. **We developed a DNA microarray based on 1,832 clones from a shotgun library used for sequencing of the highly virulent strain** *F. tularensis* **subsp.** *tularensis* **Schu S4. This allowed a genome-wide analysis of 27 strains representing all four subspecies. Overall, the microarray analysis confirmed a limited genetic variation within the species** *F. tularensis*, **and when the strains were compared, at most 3.7% of the probes showed differential hybridization. Cluster analysis of the hybridization data revealed that the causative agents of type A and type B tularemia, i.e.,** *F. tularensis* **subsp.** *tularensis* **and** *F. tularensis* **subsp.** *holarctica*, **respectively, formed distinct clusters. Despite marked differences in their virulence and geographical origin, a high degree of genomic similarity between strains of** *F. tularensis* **subsp.** *tularensis* **and** *F. tularensis* **subsp.** *mediasiatica* **was apparent. Strains from Japan clustered separately, as did strains of** *F. tularensis* **subsp.** *novicida*. **Eight regions of difference (RD) 0.6 to 11.5 kb in size, altogether comprising 21 open reading frames, were identified that distinguished strains of the moderately virulent subspecies** *F. tularensis* **subsp.** *holarctica* **and the highly virulent subspecies** *F. tularensis* **subsp.** *tularensis*. **One of these regions, RD1, allowed for the first time the development of an** *F. tularensis***-specific PCR assay that discriminates each of the four subspecies.**

Tularemia is a zoonotic disease affecting a multitude of mammalian species, most notably, humans, rabbits, hares, and many rodents (15). Humans may contract the disease by direct contact with infected animals; by inhalation of infected material; through ingestion of contaminated water or food; or by bites from vectors such as biting flies, mosquitoes, or ticks (for a recent review, see the work by Dennis et al. [6]). The natural reservoir for the etiological agent, the facultative intracellular bacterium *Francisella tularensis*, is largely unknown, although studies have shown that the organisms may persist for more than a year in water or mud (32).

Inoculation or inhalation of as few as 10 *F. tularensis* organisms may cause disease in humans (37, 38), and due to its high level of virulence and infectivity, the bacterium is considered a potential biological warfare agent. Little is known to explain the high degree of virulence, the contagiousness, and the intracellular survival of the pathogen. Unlike many other facultative intracellular bacteria, *F. tularensis* has not been shown to produce toxins.

The genus *Francisella* contains two species, *F. tularensis* and *F. philomiragia*. Of these, *F. philomiragia* is an opportunistic pathogen, rarely causing disease in humans and often associated with water (14). At present, four subspecies of *F. tularensis* are recognized (41): *F. tularensis* subsp. *tularensis* (also referred to as type A), *F. tularensis* subsp. *holarctica* (type B), *F. tularensis* subsp. *mediasiatica*, and *F. tularensis* subsp. *novicida*. The four subspecies show marked differences in their virulence and originate from different regions of the world but still display a very close phylogenetic relationship. Moreover, the subspecies of *F. tularensis* are antigenically similar.

The most virulent *F. tularensis* isolates belong to *F. tularensis* subsp. *tularensis*, and the in humans rate of mortality of untreated infections caused by this subspecies may be as high as 30 to 60% (7, 43). Even today, with the availability of effective antibiotics, the mortality rate is a few percent (6). The genetic basis for the high degree of virulence of this subspecies is unknown. Although the subspecies is believed to be confined to North America, recent studies have reported the occasional isolation of strains belonging to this subspecies in Europe (13, 20). Virtually all European isolates belong to the subspecies *F. tularensis* subsp. *holarctica*, and although such isolates may cause severe disease and are highly infectious, the disease is

* Corresponding author. Mailing address: Swedish Defense Research Agency, SE 901 82, Umeå, Sweden. Phone: 46 90 106669. Fax: 46 90 106800. E-mail: mats.forsman@foi.se.

rarely fatal in humans. Strains belonging to the subspecies *F. tularensis* subsp. *holarctica* are also found in North America and Japan. Strains belonging to the subspecies *F. tularensis* subsp. *mediasiatica* are confined mainly to the Central Asian republics of the former USSR, and little is known about their properties or abilities to cause disease in humans (30). Strains of the subspecies *F. tularensis* subsp. *novicida* seem to have a close association with water and rarely cause human disease (3, 14).

The identification of *F. tularensis* and the differentiation of its subspecies have traditionally been based on growth characteristics, biochemical analyses, and virulence in rabbits (41). Only experimental infections of rabbits or monkeys appear to reliably reflect the difference in virulence between subspecies recorded in humans (29). Mice are highly susceptible and succumb from infection with all subspecies. Besides being time-consuming and tedious, cultivation of *F. tularensis* carries a high risk of laboratory-acquired infection. Diagnosis of human tularemia, therefore, traditionally relies on serology rather than detection of the infectious agent. Since antibodies are usually not detected before the second week of disease (22), there is a need for more rapid methods for diagnosis. To this end, DNA-based methods are of special interest; inactivated samples can be used, and extensive handling of live pathogens can be avoided. Previous studies have indicated a high degree of DNA sequence similarity among the subspecies of *F. tularensis* (5, 16, 20, 41). In a recent report, amplified fragment length polymorphism analysis demonstrated an 89% similarity between strains of *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica*, and different strains of a subspecies were typically >95% similar (12). In comparison, amplified fragment length polymorphism similarity values of 90 to 100% are considered to demonstrate identical strains in other bacterial species (39). To discriminate individual strains of *F. tularensis*, hypermutable parts of the genome with repeated sequences in tandem arrays have been successfully targeted (10, 19) In the study described here, we performed a genome-wide, microarray-based analysis of *F. tularensis* strains with the aim of identifying genomic regions that distinguish the four subspecies and regions unique to the highly virulent subspecies *F. tularensis* subsp. *tularensis*.

MATERIALS AND METHODS

Bacterial strains and genomic DNA. The *F. tularensis* strains were obtained from the *Francisella* Strain Collection (FSC) of more than 300 strains at the Swedish Defense Research Agency. All strains have been characterized by means of specific agglutination, by PCR specific for a gene encoding an *F. tularensis*-specific 17-kDa lipoprotein (40), and by biochemical characterization. Twenty-seven *F. tularensis* strains, representing all four recognized subspecies, were selected for this study to ensure maximum genetic diversity (Table 1). The *F. tularensis* strains were grown for 2 days on modified Thayer-Martin agar plates at 37°C in 5% CO_2, harvested by scraping, and suspended in saline at a concentration of 10^9 CFU/ml. The cells were lysed with guanidine isothiocyanate, and the DNA was captured on silica particles as described previously (2, 17). Avirulent strain *Yersinia pestis* KIM5 (obtained from R. R. Brubaker, Michigan State University, East Lansing) was included as a control strain. Heat-killed whole-cell bacterial lysates were used as templates to confirm the microarray results by PCR.

Genomic *F. tularensis* microarray construction. A total of 1,832 inserts cloned into pUC18 from a shotgun DNA library of *F. tularensis* Schu S4 (34) were selected by using the following criteria: good-quality readings, minimal overlap with other clones, and maximal coverage of the genome. The mean ± standard deviation insert sizes were 1,405 ± 205 bp. The DNA probes used for microarray construction were obtained by amplifying the inserts by PCR with vector-specific primers 5′-TGT AAA ACG ACG GCC AGT-3′ and 5′-ATG TTG TGT GGA

TABLE 1. *F. tularensis* strains hybridized to the whole-genome DNA microarray

Strain source information	FSC no.	Alternative strain designation	No. (%) of differentially hybridizing probes
F. tularensis subsp. *tularensis*[a]			
Tick, 1935, British Columbia, Canada	041	Vavenby	20 (1.2)
Canada	042	Utter	12 (0.7)
Human ulcer, 1941, Ohio	043	Schu	2 (0.1)
Human pleural fluid, 1940, Ohio	046	Fox	17 (1.0)
Human ulcer, laboratory infection	053	F. tul AC	15 (0.9)
Hare, 1953, Nevada	054	Nevada	20 (1.2)
Mite, 1988, Slovakia	198	SE-219/38	1 (0.1)
Mite, 1988, Slovakia	199	SE-221/38	13 (0.8)
Human lymph node, 1920, Utah	230	ATCC 6223	43 (2.6)
Human ulcer, 1941, Ohio	237	Schu S4	0 (0)
F. tularensis subsp. *holarctica*			
Beaver, 1976, Montana	035	B423A	49 (3.0)
Tick, 1953, Norway	032	Norway	36 (2.2)
Hare, 1981, Sweden	076	SVA T9	34 (2.1)
Hare, 1984, Finland	080	SVA T13	49 (3.0)
Human blood, 1989, Norway	089	45F2	42 (2.6)
Water, 1990, Ukraine	124	14588	23 (1.4)
Norway rat, 1988, Russia	150	250	31 (1.9)
Live vaccine strain, Russia	155	LVS, ATCC 29684	28 (1.7)
Human blood, 1994, Sweden	157	CCUG 33270	37 (2.3)
Human ulcer, 1998, Sweden	200		37 (2.3)
Human, 1958, Japan	021	Tsuchiya	17 (1.0)
Human, 1950, Japan	022	Ebina	26 (1.6)
Tick, 1957, Japan	075	Jama	32 (2.0)
F. tularensis subsp. *mediasiatica*			
Miday gerbil, 1965, Central Asia	147	543	12 (0.7)
Tick, 1982, Central Asia	148	240	19 (1.2)
Hare, 1965, Central Asia	149	120	13 (0.8)
F. tularensis subsp. *novicida*, water, 1950, Utah	040	ATCC 15482	60 (3.7)

[a] Subspecies nomenclature according to Sjöstedt (41).

ATT GTG-3′, and the sizes were visually verified on agarose gels. The PCR products were purified by using 96-well MultiScreen-PCR filter plates (Millipore), suspended in 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), transferred to 384-well plates, and stored at −20°C. Two replicates of each PCR amplicon were printed onto glass slides (25 by 75 mm; CMT-GAPS Amino Silane-coated slides; Corning) with an SDDC-2 robot (Engineering Services, Inc., Toronto, Canada) equipped with Stealth micropins (ArrayIt; TeleChem). The microarrays, which featured eight grids of 480 probes (20 by 24) with a spot-to-spot distance of 160 μm and a total size of 1.4 cm², were postprocessed with UV cross-linking, blocked by a previously described method (http://cmgm.stanford.edu/pbrown/), and stored at room temperature with protection from light.

Microarray hybridization. Three micrograms of bacterial genomic DNA was digested with *Nla*III and labeled with Cy-3–dCTP (Amersham Pharmacia Biotech) by using Klenow DNA polymerase and random octamer primers. The labeled DNA was precipitated with isopropanol, dissolved in 20 μl of DIG Easy Hyb hybridization solution (Roche) supplemented with 2 μl of herring sperm DNA (10 mg/ml), and then hybridized immediately. Microarray slides were prehybridized beneath glass coverslips with 20 μl of DIG Easy Hyb solution (which lowers the hybridization temperature as though it contained 50% formamide) and herring sperm DNA for 1 h at 37°C under humid conditions. Hybridization with labeled DNA was done overnight at 37°C. The slides were washed three times in prewarmed 0.1× SSC–0.1% sodium dodecyl sulfate for 10 min at 50°C with occasional agitation, washed in 0.1× SSC, and dried by centrifugation at 46 × g for 5 min. The slides were shielded to avoid exposure to excessive light during washing and subsequent handling.

Scanning and data analysis. Fluorescent scanning was performed with a ScanArray 4000 instrument (GSI Lumonics, Billerica, Mass.) at a resolution of 5 μm and a 16-bit image depth. Laser strength and photomultiplier gain were adjusted

in order to give saturated signals for <1% of the probes in the hybridizations. Probe signal intensities were quantified with ImaGene (version 3.04 and 4.0; BioDiscovery Inc., Marina del Rey, Calif.). Anomalous spots were identified by visual inspection and were excluded from analysis. To reduce the impact of inconsistent printing, only the higher-intensity value for each duplicate probe was used for further analysis. The data were globally normalized by dividing the intensity signal by the mean signal intensity for each hybridization experiment. The probe noise for 18 of the probe signal variables that had close to equal hybridization signal levels with reference strain Schu S4 was examined with *F. tularensis* subsp. *tularensis* strains (excluding ATCC 6223). Probes hybridizing with weak signal intensities for reference strain Schu S4 were found to have low signal-to-noise ratios and were excluded from further analysis, resulting in a data matrix consisting of 1,635 signal intensity values for each hybridization experiment. Differentially hybridizing probes were identified by using a criterion that required <30% of the signal intensity compared with that for reference strain Schu S4. A binary data matrix was derived for analysis of strain relatedness. Relationships between strains were determined by using hierarchical cluster analysis with complete linkage and the uncentered correlation metric (CLUSTER) (9). The cladogram was visualized by conversion of the output file into parenthetical notation (Newick Standard) and by use of TreeView software (31).

Analysis of RDs and development of a subspecies-specific PCR. The genetic differences between *F. tularensis* subsp. *tularensis* strains and non-Japanese *F. tularensis* subsp. *holarctica* strains detected by microarray analysis were designated genomic regions of difference (RDs). Each RD was verified by PCR analysis across each region in all strains. PCR amplification was performed with the following flanking primer pairs: 5′-TTT ATA TAG GTA AAT GTT TTA CCT GTA CCA-3′ and 5′-GCC GAG TTT GAT GCT GAA AA-3′ (RD1); 5′-CAT GTC TAT GTT ATT TAT CAC TCA TGA TTT-3′ and 5′-GCT TTA GCA TTC ATT ATC ATG AGA AGT AT-3′ (RD2); 5′-TCA ACT ACA GCA GAG TAT AAG CAA-3′ and 5′-ACA TTT TAA GAG CTC CTT GGT AGA T-3′ (RD3); 5′-TTT TAT AAA TCA AGA ATC AAT ATG CCT AA-3′ and 5′-TGA TGA AGG TCA TTG TCT TAG AGA T-3′ (RD4); 5′-TCA GTT ATC ATT TCA CTA AGT AAA GCA TT-3′ and 5′-GTA GAG TTT AAG GGC GAT GAG TT-3′ (RD5); 5′-TGA TTG TAC GCT TTG TTC TAT GAA TA-3′ and 5′-TTT GTC AAA GCA GGA ATC TTA-3′ (RD6); and 5′-CGA TAT GTT TTG CTA TAG ATA ACC TTG AT-3′ and 5′-CAC TTA TCG TAG CCG CTA AAC AT-3′ (RD7). The PCRs were performed in volumes of 25 μl with an initial denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 60 s. The PCR products were sized on agarose gels and visualized by ethidium bromide staining. The amplified DNA fragments from *F. tularensis* subsp. *tularensis* strain Schu S4 and the *F. tularensis* subsp. *holarctica* live vaccine strain (LVS) were purified with Micro-Spin S400HR columns (Amersham Pharmacia Biotech, Uppsala, Sweden), and both strands were sequenced by using the same primers as those used in the initial PCR amplifications and the Big-Dye terminator cycle-sequencing ready reaction kit on an ABI 377XL DNA sequencer (PE Applied Biosystems). An ~11.5-kb region, RD8, could not be amplified from any strain of *F. tularensis* subsp. *holarctica*; hence, this region was not mapped in detail.

In addition, the highly variable RD1 was analyzed by sequencing the genomic DNA from strains representing each of the four *F. tularensis* subspecies. The large PCR product corresponding to RD1 of strain FSC040 was sequenced by primer walking.

The open reading frames identified in the sequences of all RDs were compared with the sequences in the GenBank database by using the BLASTX program with standard parameters and low-complexity filtering.

Nucleotide sequence accession numbers. The RD1 sequence data have been submitted to GenBank and assigned accession numbers AF469614 (Schu S4, FSC237), AF469615 (ATCC 6223, FSC230), AF469616 (FSC147), AF469617 (LVS FSC155), AF469618 (FSC075), and AF469619 (FSC040).

RESULTS

***F. tularensis* whole-genome microarray.** The *F. tularensis* microarray was constructed by using 1,832 clones from a shotgun library of highly virulent strain Schu S4. DNA from the homologous organism hybridized with 98% of the probes. These genome sequence data are available online at http://artedi.ebc .uu.se/Projects/Francisella/. Since the genome size is not yet precisely determined, the genome coverage was estimated by comparing the size of the present gene assembly with that

estimated by separation of the *F. tularensis* genome by restriction enzyme cleavage and pulsed-field gel electrophoresis (unpublished data). On this basis, it is estimated that the probes of the DNA microarray represent more than 95% of the *F. tularensis* genome. As a control, 16 probes containing various mouse DNA sequences were included in the microarray. In all experiments, they did not hybridize to *F. tularensis* DNA. Chromosomal DNA from *Y. pestis* hybridized with 10 of 1,832 probes.

Genomic differences within the species *F. tularensis*. Chromosomal DNA from 27 *F. tularensis* strains (Table 1) showed differential hybridization to certain probes on the microarray. There were fewer differences between strains belonging to the same subspecies than between strains belonging to different subspecies (Fig. 1). A total of 0.1 to 3.7% differentially hybridizing probes (DHPs) were detected for a given strain (Table 1). A hierarchical cluster analysis was performed with the binary-coded signal-intensity microarray data, and a cladogram illustrating the similarity between strains is shown in Fig. 1. Generally, strains from the same subspecies formed clusters. Moreover, the cluster analysis indicated that (i) strains belonging to *F. tularensis* subsp. *mediasiatica* show close genetic similarity to strains of *F. tularensis* subsp. *tularensis*; (ii) the type strain of *F. tularensis*, ATCC 6223, is distinct from other strains belonging to *F. tularensis* subsp. *tularensis*; (iii) the Japanese strains cluster separately from the European and American *F. tularensis* subsp. *holarctica* strains; and (iv) the single representative of *F. tularensis* subsp. *novicida* showed a unique localization. Overall, only a few probes discriminated the strains. All DHPs present in LVS FSC155 were present in other European or North American strains of *F. tularensis* subsp. *holarctica*. The two most divergent strains were FSC040 (*F. tularensis* subsp. *novicida*) and ATCC 6223 (*F. tularensis* subsp. *tularensis*, avirulent). The strains showed 3.7 and 2.6% DHPs, respectively, many of which were unique for each strain (Table 1 and Fig. 1). The Japanese strains lacked a number of DHPs present among other strains of *F. tularensis* subsp. *holarctica*, which explains their location on the cluster plot (Fig. 1).

***F. tularensis* RDs.** Many of the DHPs clustered in contiguous genomic regions and were designated *F. tularensis* RDs. It was apparent that the presence or absence of *F. tularensis* RDs reflected the different subspecies of the strains. We focused further studies on eight such regions of 0.6 to 11.5 kb that differentiated strains of the highly virulent subspecies *F. tularensis* subsp. *tularensis* (type A) from strains of the moderately virulent subspecies *F. tularensis* subsp. *holarctica* (type B, non-Japanese) (Table 2). PCR and sequence analysis of these RDs revealed that in each case a 9- to 38-bp direct repeat motif was present in the corresponding Schu S4 sequence (Table 2 and Fig. 2). The only exception was RD8, in which the presence of such motifs could not be verified. Analysis of *F. tularensis* RD1 to RD8 identified a total of 21 open reading frames in the sequences absent from *F. tularensis* subsp. *holarctica* strains, with several open reading frames sharing high degrees of similarity to known bacterial genes (Table 2).

F. tularensis RD1 was analyzed in detail, as it appeared to be highly variable between strains from all four subspecies. Sequence analysis of RD1 showed the following characteristics compared to the sequence of strain Schu S4: strain FSC147 (*F. tularensis* subsp. *mediasiatica*) showed a 68-bp deletion be-
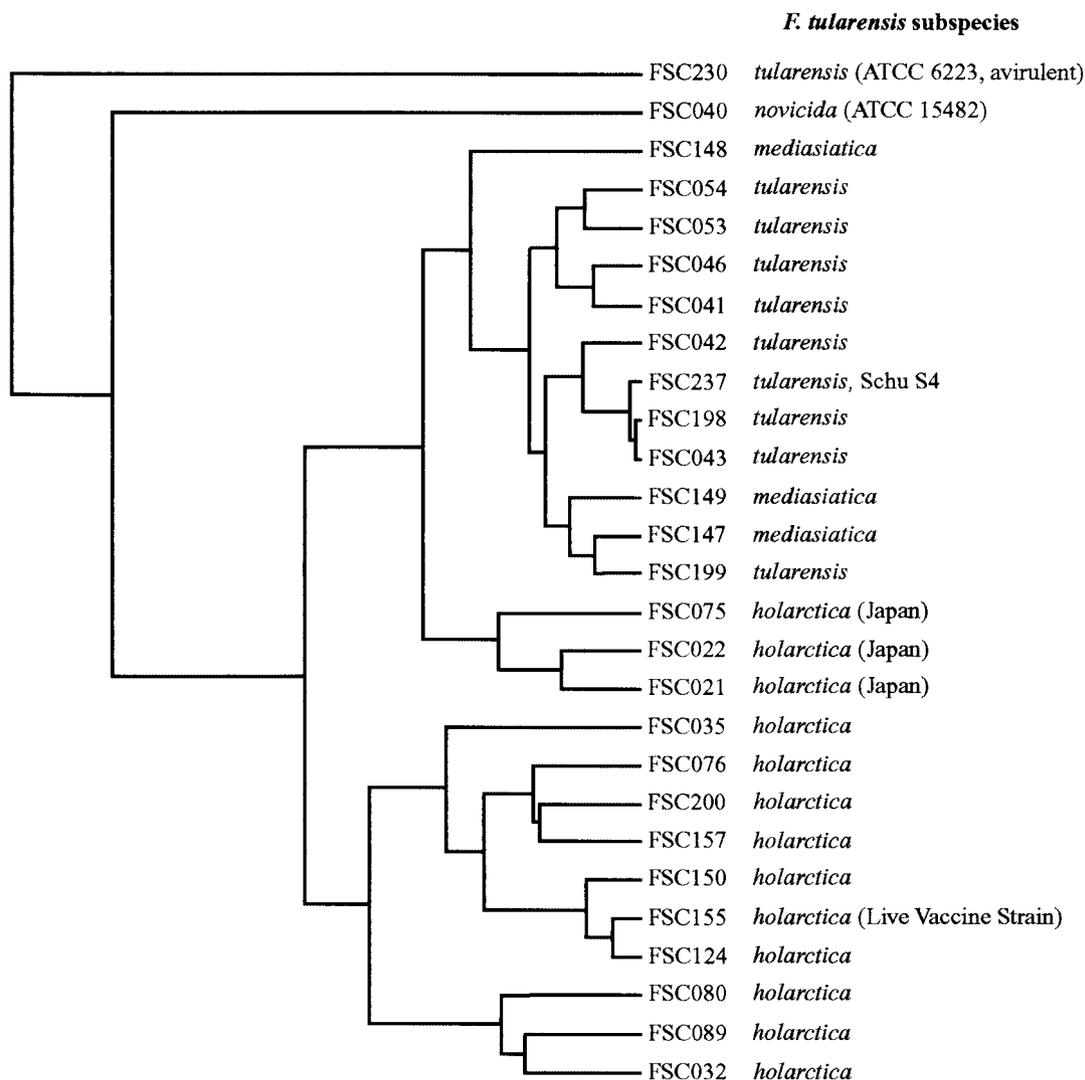
F. tularensis subspecies

FIG. 1. Hierarchical cluster analysis of *F. tularensis* DNA microarray hybridizations. Strain identity is indicated by FSC number (Table 1) and subspecies.

ginning after position 262; the LVS, strain FSC155 (*F. tularensis* subsp. *holarctica*, Europe), exhibited a 598-bp deletion beginning after position 418; and strain FSC075 (*F. tularensis* subsp. *holarctica*, Japan) had a 389-bp deletion beginning after position 1037. Strain FSC040 (*F. tularensis* subsp. *novicida*) showed a more complex and longer RD1 sequence (Fig. 2 and 3).

By using the primer pair flanking RD1 for PCR amplification, the four subspecies and, in addition, Japanese strains could be discriminated by their distinct amplicon sizes (Fig. 3). DNA was amplified from all strains tested, and each subspecies showed a unique amplicon size. Moreover, also strains from Japan showed one amplicon of the same size distinct from those of all other strains. In total, 13 strains of *F. tularensis* subsp. *tularensis*, 22 strains of *F. tularensis* subsp. *holarctica* (non-Japanese origin), 3 strains of *F. tularensis* subsp. *mediasiatica*, 1 strain of *F. tularensis* subsp. *novicida*, and 4 strains of Japanese origin were tested (data not shown). All 27 strains listed in Table 1 were included in the analysis. The specificity

of the PCR assay for *F. tularensis* was verified by analysis of nine other bacterial pathogens: *F. philomiragia*, *Staphylococcus aureus*, *Salmonella enterica* serovar Typhimurium, *Yersinia enterocolitica*, *Bacillus subtilis*, *Escherichia coli*, *Coxiella burnetii*, *Legionella pneumophila*, and *Listeria monocytogenes*. No visible PCR amplicons were obtained (data not shown).

DISCUSSION

Twenty-seven strains of *F. tularensis* that represented each of the four subspecies and that were isolated in different parts of the world over a period of more than 70 years were selected for the study. When DNA from the strains was analyzed with the microarray, DHPs that showed reduced hybridization signal intensities compared to that for reference strain Schu S4 were detected. The differences in the hybridization patterns detected among strains of the same subspecies were smaller than those detected among strains of different subspecies. Five distinct hybridization patterns were apparent. Overall, the strains

TABLE 2. Compilation of RDs between *F. tularensis* subsp. *tularensis* and non-Japanese *F. tularensis* subsp. *holarctica* strains

RD[a]	Size (bp)	Open reading frame	Homolog (GenBank accession no.)	E value[b]	Direct repeat motif[e]
RD1	598	Hypothetical protein[c]			AgAATGGATTCTACGGTCAAGCCGTAGAATGA AaAATGGATTCTACGGTCAAGCCGTAGAATGA
RD2	975	Oligopeptide ABC transporter, ATP-binding protein, *oppD*	NP_230739.1	2 E-96	CATCCGTATACAAAAGtcCTtCTAAATgCtATtCCTA CATCCGTATACAAAAGgtCTgCTAAATaCaATgCCTA
		Oligopeptide ABC transporter, ATP-binding protein, *oppF*	NP_439277.1	1 E-88	
RD3	1,731	Putative modulator of drug activity B, *mdaB*	NP_289602.1	2 E-72	TATCTAAtGAATGGGAGCTA-TC TA-CTAAa-AATGG-AG-TAATC
		Conserved hypothetical protein	NP_436326.1	6 E-29	
		Hypothetical arginine-ornithine operon transport protein	NP_249582.1	1 E-09	
RD4	1,314	Piperideine-6-carboxylate dehydrogenase (strongly homologous to aldehyde dehydrogenases)	BAB19801.1	1 E-119	TCgGCTTTTAAC TCaGCTTTTAAC
		Conserved hypothetical protein	NP_232243.1	6 E-55	
RD5	1,117	Adenine-specific DNA methyltransferase	NP_478077.1	6 E-17	AGtAgAAGAGCAAgAGCAAtATaAAActG AGcAcAAGAGCAAaAGCAAgATaAAAaaG
		Type I restriction-modification system, subunit	NP_345026.1	1 E-11	
RD6	4,260	Conserved hypothetical protein	NP_488560.1	2 E-41	GtTATGTGCTT GcTATGTGCTT
		Hypothetical protein			
RD7	1,458	Aspartokinase I, homoserine dehydrogenase I	NP_414543.1	1 E-77	GcTAgCAAAATTAG GaTAtCAAAA-TAG
		Homoserine kinase (possibly truncated)	NP_299503.1	3 E-18	
RD8	~11,500	Snf2 family ATP-dependent helicase, HepA	NP_603318.1	0.0	Insertion sequence elements[d]
		Type III restriction-modification system, subunit (possibly truncated)	NP_603321.1	1 E-133	
		Antirestriction protein	NP_478454.1	4 E-25	
		Hypothetical protein	ZP-00044100.1	1 E-10	
		Hypothetical protein	NP_282940.1	7 E-09	
		MobC-like protein	CAB62410.1	6 E-08	
		Hypothetical protein	ZP_00086265.1	4 E-04	

[a] All RDs except RD8 were confirmed by PCR analysis and sequencing of the *F. tularensis* subsp. *holarctica* LVS FSC 155.
[b] E values were obtained from NCBI Blast2 searches for sequences in the GenBank database by using the low-complexity filter option.
[c] Open reading frame A in Fig. 2.
[d] GenBank accession no. AF357005.
[e] Lowercase letters indicate deviation from the consensus sequence.

clustered according to their subspecies, and the clustering supported the present subspecies division. However, the whole-genome approach uncovered some additional information. Interestingly, the *F. tularensis* subsp. *mediasiatica* strains, which comprise strains from the Central Asian republics of the former USSR, and the *F. tularensis* subsp. *tularensis* strains, which represent strains from North America, clustered together in the analysis. The *F. tularensis* subsp. *mediasiatica* strains exhibited a moderate degree of virulence for mammals, but most of their biochemical characteristics are common with those of strains of the highly virulent subspecies *F. tularensis* subsp. *tularensis* (30). Thus, virulence appears to be the most important criterion for discrimination of these two subspecies. A whole-genome sequence analysis of an *F. tularensis* subsp. *mediasiatica* strain could possibly pinpoint the cause of this difference in virulence.

The microarray analysis revealed that strains of *F. tularensis* subsp. *holarctica* showed a greater degree of heterogeneity than strains belonging to the subspecies *F. tularensis* subsp. *mediasiatica* and *F. tularensis* subsp. *tularensis*. Although subdivisions of the subspecies *F. tularensis* subsp. *holarctica* have been proposed (28), strains of the subspecies display very similar phenotypic characteristics and there is little evidence supporting this proposal. The present investigation and previous studies based on genetic analyses have demonstrated that Japanese strains show genetic patterns distinct from those of European or American strains of *F. tularensis* subsp. *holarctica* (16, 20). Compared to strain Schu S4, Japanese strains lacked some DHPs found in other strains of *F. tularensis* subsp. *holarctica*. One speculative reason for this may be that Japanese strains represent an intermediate between the subspecies *F. tularensis* subsp. *tularensis* or *F. tularensis* subsp. *mediasiatica* and *F. tularensis* subsp. *holarctica*. Since the Japanese strains appear to be less virulent than their European or American counterparts, further studies are warranted to determine if these strains constitute a separate subspecies.
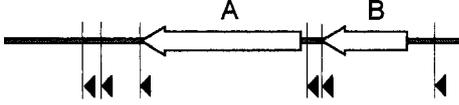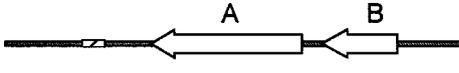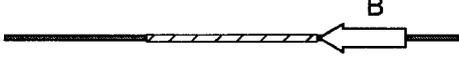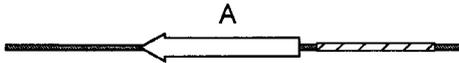
FIG. 2. Schematic depiction of the variable chromosomal region, RD1, identified among the different *F. tularensis* subspecies with insertions and deletions. Two direct repeats were found to define the boundaries of each deleted genomic segment. Within RD1, two complete open reading frames, designated A and B, were present. Both open reading frames A and B were found in strains Schu S4, ATCC 6223 (*F. tularensis* subsp. *tularensis*), and FSC147 (*F. tularensis* subsp. *mediasiatica*), although in strain FSC147 open reading frame A was truncated by 10 amino acids from the C terminus and open reading frame B was truncated by 12 amino acids from the N terminus. Only open reading frame B was found to be present in the LVS (*F. tularensis* subsp. *holarctica*), and only open reading frame A was found to be present in strain FSC075 (*F. tularensis* subsp. *holarctica*, Japan). Four open reading frames, designated C, D, E, and F, were found in strain FSC040 (*F. tularensis* subsp. *novicida*). Searches of the sequences in the GenBank database with the BLAST program could not identify any known gene function of the open reading frames of RD1.

ATCC 6223, the type strain of *F. tularensis* subsp. *tularensis*, appeared to be genetically distinct from all other strains investigated and displayed many unique DHPs. This strain became avirulent during multiple passages on artificial media (18) and has more fastidious growth requirements than most other *F. tularensis* strains (41). Our findings suggest that during the repeated passages the isolate lost regions of its genome. In contrast, FSC155 (the LVS), which also became attenuated after repeated passages in vitro (45), did not show any unique change in hybridization signals compared to those of the other European or North American strains of *F. tularensis* subsp. *holarctica*. This suggests that its attenuation be due to a more subtle genetic change.

The microarray analysis of the different subspecies of *F. tularensis* indicates a high degree of genetic conservation within the species. Among the 27 *F. tularensis* strains analyzed, only 0.1 to 3.7% of the probes showed differential hybridization. A previous survey of *Pseudomonas aeruginosa* strains with a random-fragment phage library of strain X24509 and comparative hybridization to *P. aeruginosa* strain PAO1 suggested genomic variation between these two strains in the range of 3% (24). This is in contrast to the findings from microarray studies performed with certain other bacterial species. Among 15 *Helicobacter pylori* strains tested, 22% of the open reading frames on the array were absent from at least one strain (36), and among 36 *S. aureus* isolates tested, up to 12% of the open
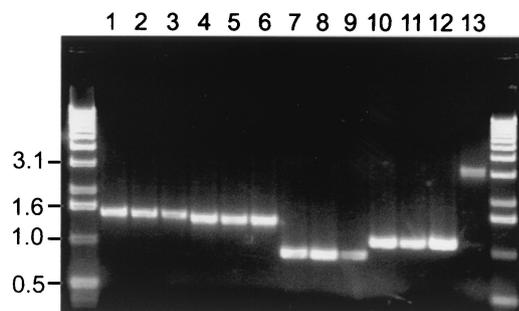
FIG. 3. PCR with primers flanking RD1 resulted in amplicons of different sizes for the different subspecies of *F. tularensis*. Lanes 1 to 3, strains ATCC 6223, Schu S4, and FSC199 (*F. tularensis* subsp. *tularensis*), respectively; lanes 4 to 6, strains FSC147, FSC148, and FSC149 (*F. tularensis* subsp. *mediasiatica*), respectively; lanes 7 to 9, strains FSC035, FSC155, and FSC200 (*F. tularensis* subsp. *holarctica*, non-Japanese), respectively; lanes 9 to 12, strains FSC021, FSC022, and FSC075 (*F. tularensis* subsp. *holarctica*, from Japan), respectively; lane 13, strain FSC040 (*F. tularensis* subsp. *novicida*). Molecular sizes (in kilobases) are indicated on the left.

reading frames were absent (11). Since the *F. tularensis* microarray also includes intergenic regions, a higher degree of genetic variability could be expected compared to that found with an array with genes only. In this context, it should be noted that sequence duplications as well as insertions relative to the sequence of the reference strain might be undetected in any comparative DNA microarray analysis. Moreover, very similar sequences such as homologous genes may cross-hybridize in microarray experiments. However, our results indicate that there are indeed few deletion events and few major sequence differences within the species *F. tularensis*. This is similar to recent findings in a microarray study with *Mycobacterium tuberculosis*, a species considered to be clonal and of recent global dissemination (21, 42). A possible cause for the genetic homogeneity of *F. tularensis* could be related to the intracellular lifestyle of the bacterium. According to a recently proposed model, bacterial lifestyle correlates with genomic stability (27). According to this model, obligate intracellular symbionts show the highest degree of genomic stability, and on the other end of the scale, free-living species undergo frequent rearrangement due to continuous fluctuations in their ambient environment.

Analyses of the RDs (Table 2) indicated that homologous recombination might be the cause for the missing genomic segments (i.e., deletions). The boundaries of the missing genomic segments are defined by two functional direct repeats at which excision of the segments has taken place. One region, RD8, could not be amplified by PCR from the LVS. This may be due to a rearrangement, since insertion sequence elements flank this region in Schu S4. The size of RD8 is therefore estimated according to the positions of the insertion sequence elements in the Schu S4 sequence (Table 2).

Almost nothing is known about the virulence mechanisms of *F. tularensis*. However, this study has not addressed the function of genes identified in the RDs, and their possible role in virulence will remain speculative, as genetic methods to study the mutagenesis of individual genes in *F. tularensis* are lacking. Nevertheless, some of the genes present in the highly virulent subspecies *F. tularensis* subsp. *tularensis* but absent from the

less virulent *F. tularensis* subsp. *holarctica* might be of interest for future studies. For example, oligopeptide permease (Opp), an ATP-binding cassette transporter, has been shown to play a key role in nutrient uptake, growth regulation, transport of intracellularly signaling peptides, and resistance to host defensins and toxic peptides (23, 33). An Opp operon, *oppABCDF*, is present in *F. tularensis* subsp. *tularensis* strain Schu S4, whereas the *oppD* and *oppF* genes are missing from strains of *F. tularensis* subsp. *holarctica* (Table 2). Moreover, RD5 corresponds to the deletion of a type I restriction-modification system in strains of *F. tularensis* subsp. *holarctica*. The presence of this type of gene in *H. pylori* strains colonizing mice has recently been found to correlate with the magnitude of the host response (1). It was also suggested that these genes might function by modulating the expression of genes affecting bacterial virulence (1). The *rapA* homolog of strain Schu S4 is missing from all *F. tularensis* subsp. *holarctica* strains. RapA (also known as HepA) of *Escherichia coli* is an ATP-dependent RNA helicase, and it was recently suggested that it influences DNA supercoiling or topology (44). The level of DNA supercoiling can influence gene expression (25, 35). At a local level, differences in supercoiling created during transcription can affect the expression of neighboring genes (35), and at a global level, environmental stimuli may at least transiently affect overall DNA supercoiling and expression (8). In *Vibrio cholerae*, the RapA gene has been linked to acid tolerance and virulence (26). However, perhaps the most intriguing finding is the deletion (RD7) in *F. tularensis* subsp. *holarctica* strains of a gene encoding the enzyme aspartokinase-homoserine dehydrogenase, which catalyzes two proximal steps required for homocysteine biosynthesis. Homocysteine has a central role as an endogenous nitric oxide antagonist and may counteract the antimicrobial activity of the host cell during intracellular infection (4). Thus, it will be important to determine whether the two *F. tularensis* subspecies show differential susceptibilities to nitric oxide.

One of the genomic regions discriminating *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica*, RD1, became the focus of a detailed sequence analysis, as it appeared to be highly variable. Amplification of this region allowed for the first time the discrimination of all four subspecies of *F. tularensis* by use of a single PCR, as well as the discrimination of Japanese strains. This assay holds promise as a rapid tool for taxonomy and clinically relevant typing.

ACKNOWLEDGMENTS

REFERENCES

1. **Björkholm, B. M., J. L. Guruge, J. D. Oh, A. J. Syder, N. Salama, K. Guillemin, S. Falkow, C. Nilsson, P. G. Falk, L. Engstrand, and J. I. Gordon.** 2002. Colonization of germ-free transgenic mice with genotyped *Helicobacter pylori* strains from a case-control study of gastric cancer reveals a correlation between host responses and HsdS components of type I restriction-modification systems. J. Biol. Chem. **277:**34191–34197.
2. **Boom, R., C. J. Sol, M. M. Salimans, C. L. Jansen, P. M. Wertheim-van Dillen, and J. van der Noordaa.** 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. **28:**495–503.
3. **Clarridge, J. E., III, T. J. Raich, A. Sjöstedt, G. Sandström, R. O. Darouiche, R. M. Shawar, P. R. Georghiou, C. Osting, and L. Vo.** 1996. Characterization

of two unusual clinically significant *Francisella* strains. J. Clin. Microbiol. **34:**1995–2000.

4. **De Groote, M. A., T. Testerman, Y. Xu, G. Stauffer, and F. C. Fang.** 1996. Homocysteine antagonism of nitric oxide-related cytostasis in *Salmonella typhimurium*. Science **272:**414–417.

5. **de la Puente-Redondo, V. A., N. G. del Blanco, C. B. Gutierrez-Martin, F. J. Garcia-Pena, and E. F. Rodriguez Ferri.** 2000. Comparison of different PCR approaches for typing of *Francisella tularensis* strains. J. Clin. Microbiol. **38:**1016–1022.

6. **Dennis, D. T., T. V. Inglesby, D. A. Henderson, J. G. Bartlett, M. S. Ascher, E. Eitzen, A. D. Fine, A. M. Friedlander, J. Hauer, M. Layton, S. R. Lillibridge, J. E. McDade, M. T. Osterholm, T. O'Toole, G. Parker, T. M. Perl, P. K. Russell, and K. Tonat.** 2001. Tularemia as a biological weapon: medical and public health management. JAMA **285:**2763–2773.

7. **Dienst, F. T.** 1963. Tularemia: a perusal of three hundred thirty-nine cases. J. La. State Med. Soc. **115:**114–127.

8. **Dorman, C. J.** 1991. DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. Infect. Immun. **59:**745–749.

9. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA **95:**14863–14868.

10. **Farlow, J., K. L. Smith, J. Wong, M. Abrams, M. Lytle, and P. Keim.** 2001. *Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis. J. Clin. Microbiol. **39:**3186–3192.

11. **Fitzgerald, J. R., D. E. Sturdevant, S. M. Mackie, S. R. Gill, and J. M. Musser.** 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. USA **98:**8821–8826.

12. **Garcia Del Blanco, N., M. E. Dobson, A. I. Vela, V. A. De La Puente, C. B. Gutierrez, T. L. Hadfield, P. Kuhnert, J. Frey, L. Dominguez, and E. F. Rodriguez Ferri.** 2002. Genotyping of *Francisella tularensis* strains by pulsed-field gel electrophoresis, amplified fragment length polymorphism fingerprinting, and 16S rRNA gene sequencing. J. Clin. Microbiol. **40:**2964–2972.

13. **Gurycova, D.** 1998. First isolation of *Francisella tularensis* subsp. *tularensis* in Europe. Eur. J. Epidemiol. **14:**797–802.

14. **Hollis, D. G., R. E. Weaver, A. G. Steigerwalt, J. D. Wenger, C. W. Moss, and D. J. Brenner.** 1989. *Francisella philomiragia* comb. nov. (formerly *Yersinia philomiragia*) and *Francisella tularensis* biogroup novicida (formerly *Francisella novicida*) associated with human disease. J. Clin. Microbiol. **27:**1601–1608.

15. **Hopla, C. E.** 1974. The ecology of tularemia. Adv. Vet. Sci. Comp. Med. **18:**25–53.

16. **Ibrahim, A., P. Gerner-Smidt, and A. Sjöstedt.** 1996. Amplification and restriction endonuclease digestion of a large fragment of genes coding for rRNA as a rapid method for discrimination of closely related pathogenic bacteria. J. Clin. Microbiol. **34:**2894–2896.

17. **Ibrahim, A., L. Norlander, A. Macellaro, and A. Sjöstedt.** 1997. Specific detection of *Coxiella burnetii* through partial amplification of 23S rDNA. Eur. J. Epidemiol. **13:**329–334.

18. **Jellison, W. L.** 1972. Tularemia: Edward Francis and his first 23 isolates of *Francisella tularensis*. Bull. Hist. Med. **46:**477–485.

19. **Johansson, A., I. Göransson, P. Larsson, and A. Sjöstedt.** 2001. Extensive allelic variation among *Francisella tularensis* strains in a short-sequence tandem repeat region. J. Clin. Microbiol. **39:**3140–3146.

20. **Johansson, A., A. Ibrahim, I. Göransson, U. Eriksson, D. Gurycova, J. E. Clarridge III, and A. Sjöstedt.** 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*. J. Clin. Microbiol. **38:**4180–4185.

21. **Kato-Maeda, M., J. T. Rhee, T. R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P. M. Small.** 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res **11:**547–554.

22. **Koskela, P., and A. Salminen.** 1985. Humoral immunity against *Francisella tularensis* after natural infection. J. Clin. Microbiol. **22:**973–979.

23. **Lazazzera, B. A.** 2001. The intracellular function of extracellular signaling peptides. Peptides **22:**1519–1527.

24. **Liang, X., X. Q. Pham, M. V. Olson, and S. Lory.** 2001. Identification of a genomic island present in the majority of pathogenic isolates of *Pseudomonas aeruginosa*. J. Bacteriol. **183:**843–853.

25. **Liu, L. F., and J. C. Wang.** 1987. Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA **84:**7024–7027.

26. **Merrell, D. S., D. L. Hava, and A. Camilli.** 2002. Identification of novel factors involved in colonization and acid tolerance of *Vibrio cholerae*. Mol. Microbiol. **43:**1471–1491.

27. **Mira, A., L. Klasson, and S. Andersson.** 2002. Microbial genome evolution: sources of variability. Curr. Opin. Microbiol. **5:**506.

28. **Olsufjev, N. G.** 1970. Taxonomy and characteristic of the genus *Francisella* Dorofeev, 1947. J. Hyg. Epidemiol. Microbiol. Immunol. **14:**67–74.

29. **Olsufjev, N. G., and I. S. Meshcheryakova.** 1982. Infraspecific taxonomy of tularemia agent *Francisella tularensis* McCoy et Chapin. J. Hyg. Epidemiol. Microbiol. Immunol. **26:**291–299.

30. **Olsufjev, N. G., and I. S. Meshcheryakova.** 1983. Subspecific taxonomy of *Francisella tularensis*. Int. J. Syst. Bacteriol. **33:**872–874.

31. **Page, R. D.** 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:**357–358.

32. **Parker, R. R., E. A. Steinhaus, G. M. Kohls, and W. L. Jellison.** 1951. Contamination of natural waters and mud with *Pasteurella tularensis*. Natl. Inst. Health Bull. **193:**7–33.

33. **Parra-Lopez, C., M. T. Baer, and E. A. Groisman.** 1993. Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*. EMBO J. **12:**4053–4062.

34. **Prior, R. G., L. Klasson, P. Larsson, K. Williams, L. Lindler, A. Sjöstedt, T. Svensson, I. Tamas, B. W. Wren, P. C. Oyston, S. G. Andersson, and R. W. Titball.** 2001. Preliminary analysis and annotation of the partial genome sequence of *Francisella tularensis* strain Schu 4. J. Appl. Microbiol. **91:**614–620.

35. **Rhee, K. Y., M. Opel, E. Ito, S. Hung, S. M. Arfin, and G. W. Hatfield.** 1999. Transcriptional coupling between the divergent promoters of a prototypic LysR-type regulatory system, the ilvYC operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **96:**14294–14299.

36. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. USA **97:**14668–14673.

37. **Saslaw, S., H. T. Eigelsbach, J. A. Prior, H. E. Wilson, and S. Carhart.** 1961. Tularemia vaccine study. I. Intracutaneous challenge. Arch. Intern. Med. **107:**689–701.

38. **Saslaw, S., H. T. Eigelsbach, J. A. Prior, H. E. Wilson, and S. Carhart.** 1961. Tularemia vaccine study. II. Respiratory challenge. Arch. Intern. Med. **107:**702–714.

39. **Savelkoul, P. H., H. J. Aarts, J. de Haas, L. Dijkshoorn, B. Duim, M. Otsen, J. L. Rademaker, L. Schouls, and J. A. Lenstra.** 1999. Amplified-fragment length polymorphism analysis: the state of an art. J. Clin. Microbiol. **37:**3083–3091.

40. **Sjöstedt, A., K. Kuoppa, T. Johansson, and G. Sandström.** 1992. The 17 kDa lipoprotein and encoding gene of *Francisella tularensis* LVS are conserved in strains of *Francisella tularensis*. Microb. Pathog. **13:**243–249.

41. **Sjöstedt, A.** Family XVII. Francisellaceae. Genus I. *Francisella. In* D. J. Brenner (ed.), Bergey's manual of systematic bacteriology, in press. Springer-Verlag, New York, N.Y.

42. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **94:**9869–9874.

43. **Stuart, B. M., and R. I. Pullen.** 1945. Tularemic pneumonia: review of American literature and report of 15 additional cases. Am. J. Med. Sci. **210:**223–236.

44. **Sukhodolets, M. V., J. E. Cabrera, H. Zhi, and D. J. Jin.** 2001. RapA, a bacterial homolog of SWI2/SNF2, stimulates RNA polymerase recycling in transcription. Genes Dev. **15:**3330–3341.

45. **Tigertt, W. D.** 1962. Soviet viable *Pasteurella tularensis* vaccines. Bacteriol. Rev. **26:**354–373.