Journal of Clinical Microbiology

Multilocus Sequence Typing for Studying Genetic Relationships among *Yersinia* Species

Mamuka Kotetishvili, Arnold Kreger, Georges Wauters, J. Glenn Morris Jr., Alexander Sulakvelidze and O. Colin Stine

J. Clin. Microbiol. **2005, 43(6):2674. DOI: 10.1128/JCM.43.6.2674-2684.2005.**

JOURNAL OF CLINICAL MICROBIOLOGY, June 2005, p. 2674–2684 Vol. 43, No. 6
0095-1137/05/$08.00+0 doi:10.1128/JCM.43.6.2674–2684.2005

Multilocus Sequence Typing for Studying Genetic Relationships among *Yersinia* Species

Mamuka Kotetishvili,[1] Arnold Kreger,[1] Georges Wauters,[2] J. Glenn Morris, Jr.,[1] Alexander Sulakvelidze,[1]* and O. Colin Stine[1]

Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201,[1] and University of Louvain, Faculty of Medicine, Microbiology Unit, B-1200 Brussels, Belgium[2]

Received 8 January 2005/Returned for modification 9 February 2005/Accepted 24 February 2005

The intra- and interspecies genetic relationships of 58 strains representing all currently known species of the genus *Yersinia* were examined by multilocus sequence typing (MLST), using sequence data from 16S RNA, *glnA*, *gyrB*, *recA*, and Y-HSP60 loci. *Yersinia aldovae*, *Y. bercovieri*, *Y. intermedia*, *Y. pestis*, *Y. pseudotuberculosis*, *Y. rohdei*, and *Y. ruckeri* were genetically more homogeneous than were *Y. enterocolitica*, *Y. frederiksenii*, *Y. kristensenii*, and *Y. mollaretii*. The MLST data concerning the genetic relatedness within and among various species of *Yersinia* support the idea that *Y. pestis* and *Y. pseudotuberculosis* are two lineages within the same species rather than two distinct species. *Y. ruckeri* is the genetically most distant species within the genus. There was evidence of O-antigen switching and genetic recombination within and among various species of *Yersinia*. The genetic relatedness data obtained by MLST of the four housekeeping genes and 16S RNA agreed in most, but not all, instances. MLST was better suited for determining genetic relatedness among yersiniae than was 16S RNA analysis. Some strains of *Y. frederiksenii* and *Y. kristensenii* are genetically less related to other strains within those species, compared to strains of all other species within the genus. The taxonomic standing of these strains should be further examined because they may represent currently unrecognized *Yersinia* species.

The genus *Yersinia* is a member of the gamma subdivision of *Proteobacteria* (60), and it is grouped in the family *Enterobacteriaceae*, based on biochemical tests and DNA-DNA similarity studies (10). Yersiniae have undergone extensive diversification during the course of their evolution, with one *Yersinia* species (*Yersinia pestis*) becoming the deadliest bacterium ever known in human history, and other species (e.g., *Y. aldovae*) diverging into completely nonpathogenic organisms (46, 55). Extensive research has been conducted to characterize 3 (*Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica*) of the 11 currently recognized species of *Yersinia*; however, the remaining 8 "*Y. enterocolitica*-like" species (*Y. frederiksenii*, *Y. intermedia*, *Y. kristensenii*, *Y. bercovieri*, *Y. mollaretii*, *Y. rohdei*, *Y. ruckeri*, and *Y. aldovae*) have been only moderately studied because they were not clearly recognized as human pathogens (55). This situation has resulted in a marked paucity of information pertaining to the phylogenetic interrelationships of all 11 species within the genus and to the mechanisms responsible for their serological and genetic divergences.

Initial studies examining the relatedness among *Yersinia* species used serotyping and other phenotypic characteristics, such as biochemical properties (biotyping), susceptibility to antibiotics (13), and phage typing (7, 44). Brenner et al. (11) subsequently introduced DNA hybridization techniques to classify *Y. enterocolitica*-like species and to determine the genetic relatedness among them. However, the methodology was limited because it did not provide information required to determine evolutionary relationships among yersiniae, and it was prone to yield potentially misleading results influenced by the level of gene acquisition and loss (39). Therefore, several molecular typing methodologies, such as plasmid profile analysis (22, 50), restriction fragment length polymorphism of chromosomal DNA (9), ribotyping (40, 42), sequence analysis of the 16S RNA gene (31), pulsed-field gel electrophoresis (PFGE) (43), and variable-number tandem repeat analysis (2) were subsequently applied for typing of yersiniae during epidemiological investigations and for determining genetic relatedness among *Yersinia* strains. Most of these approaches have good or excellent discriminatory power (e.g., PFGE was proposed as the "gold standard" for typing of *Yersinia* strains [34]) and are well suited for short-term epidemiological studies. However, they are less suited for long-term epidemiological studies and for determining evolutionary traits of, and phylogenetic relationships among, various strains or species (18). A relatively recently developed approach called multilocus enzyme electrophoresis (MLEE) addressed some of the shortcomings of the above-mentioned methodologies by combining excellent discriminatory power with relative clonal stability (49). Several studies (12, 17, 25, 26) have utilized MLEE to characterize yersiniae. However, problems with band resolution, and the ability of environmental conditions to alter phenotypic expression of the enzyme under study, can adversely affect the reproducibility of MLEE typing and complicate data analysis.

Nucleotide sequence-based approaches alleviate all of the above-mentioned drawbacks of MLEE. Indeed, sequencing of the entire genome of a bacterium, with subsequent in silico analysis of the data, constitutes the ultimate approach for delineating the genetic relatedness of various bacteria, and this approach is likely to become a common strategy during future studies of bacterial evolution and virulence traits (59). However, sequencing an entire bacterial genome is a major undertaking, and therefore it is usually limited to very few strains per

* Corresponding author. Mailing address: Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, MSTF Bldg., 10 South Pine Street, Baltimore, MD 21201. Phone: (410) 706-4587. Fax: (410) 706-4581. E-mail: asulakve@epi.umaryland.edu.

species. Sequencing a single gene—usually the 16S RNA gene, which has been used for the molecular analysis of many bacterial pathogens, including *Yersinia* (31, 32)—also is not optimal. For example, interspecies recombinations in 16S RNA genes (initially thought to be very rare) have recently been inferred (52) to occur in at least some bacterial species, which underscores the importance of not focusing on a single gene during studies to determine the phylogeny/taxonomy of a bacterial species (39). In this context, using a technique that allows sequence-based analysis of multiple loci simultaneously will provide an ideal balance between sequence-based resolution and technical feasibility. Such an approach, called multilocus sequence typing (MLST), was developed by Maiden et al. (41) to characterize meningococci, and it has rapidly been applied to the study of many other bacterial species as well, including selected species of *Yersinia* (1). However, MLST data pertaining to genetic relationship, phylogenetic traits, and clustering of all known species within the genus *Yersinia* are not currently available, and the present study is the first attempt to use MLST to examine the intra- and interspecies relationships among all of the currently described yersiniae.

MATERIALS AND METHODS

Bacterial strains, growth conditions, and DNA isolation. The bacterial collection contained 57 *Yersinia* strains representing all of the *Yersinia* species except *Y. pestis*. Sequence data for the appropriate loci from *Y. pestis* strain KIM (NC-004088) (62) were obtained from GenBank (http://ncbi.nlm.nih.gov) and were analyzed with the rest of our *Yersinia* isolates (Table 1). The isolates included *Y. aldovae* (two strains), *Y. bercovieri* (nine strains), *Y. enterocolitica* (five strains), *Y. frederiksenii* (eight strains), *Y. intermedia* (six strains), *Y. kristensenii* (eight strains), *Y mollaretii* (eight strains), *Y. rohdei* (five strains), *Y. ruckeri* (four strains), and *Y. pseudotuberculosis* (two strains). Five of these strains were ATCC isolates, and the remaining strains were from G. Wauters' and A. Sulakvelidze's bacterial strain collections. Of the five *Y. enterocolitica* strains, strains 8081 and YE37 belonged to biogroup 1B, strain W22708 was in biogroup 2, and strains A1074 and 1476 belonged to biogroup 1A. All of the strains were grown (30°C, overnight) in Luria-Bertani (LB) broth. Bacterial genomic DNA required for PCR amplification was extracted with the PrepMan TM Ultra minikit (Applied Biosystems, Inc., Foster City, Calif.).

MLST genes, primer design, and DNA sequencing. Five gene loci were selected for MLST analysis, including 16S RNA, Y-HSP60 (encoding a 60-kDa heat shock protein), *glnA* (glutamine synthetase), *gyrB* (DNA gyrase B subunit), and *recA* (DNA repair and recombination). The primers and PCR conditions for amplifying the 16S rRNA locus have been reported previously (29). Because of the limited sequence availability of some of the *Yersinia* species in GenBank, the primers for the remaining four loci were designed by using a two-step approach. The first set of primers was developed by using ClustalX (35) to align the corresponding *Yersinia* sequences contained in GenBank and then selecting conserved regions flanking potentially variable internal fragments of the targeted genes. The primers were examined for their ability to amplify the DNA in a limited number of isolates of various *Yersinia* species in our collection, and the amplified fragments were sequenced. A second set of primers (Table 2) was designed based on the sequences obtained, and efforts were made to select primers that would amplify the target loci in all strains of *Yersinia* species examined during this study. The PCR amplification conditions for the four housekeeping gene fragments (Y-HSP60, *glnA*, *gyrB*, and *recA*) were as follows: 94°C for 5 min, followed by 35 amplification cycles, each consisting of sequential incubation at 94°C (45 s), 51°C (45 s), and 72°C (1 min), followed by the final incubation at 72°C for 5 min. The amplified fragments were sequenced in both directions using the BigDye Terminator cycle sequencing kit (Applied Biosystems), and an ABI 3700 DNA analyzer was used to separate the labeled fragments by size.

MLST data analysis. Reading of trace files and assembly of contigs were performed using the Phred (19, 20) and Phrap (available at http://www.washington.edu) programs, respectively. The contigs were viewed with Consed (24), the resulting DNA sequences were trimmed by removing low-quality nucleotide sequences from the ends, and the sequences were aligned by ClustalX (35). The START (sequence type analysis and recombinational tests) program (36) was used to determine the number of polymorphic sites, the ratio of synonymous and nonsynonymous nucleotide substitutions (d_N/d_S), and guanine-cytosine content. The method of split decomposition implemented in the Splitstree program (30) was applied to detect recombination between the strains and species. Splitsgraphs were computed from the sequences for each individual locus, and also from a "superlocus" consisting of 1,407 bp of concatenated *recA*, Y-HSP60, *gyrB*, and *glnA* sequences for each isolate (in order to obtain accurate alignments, the ends of some gene sequences were trimmed, which resulted in a total length of 1,407 bp for the "superlocus"). The genetic distances between the isolates were computed from the uncorrected distance matrices ("P") of pairwise differences (the total number of nucleotides that differed between the two sequences divided by the total number of nucleotides compared). The clustering of strains/species was determined with the PAUP program (Sinauer, version 3.1) using unrooted neighbor-joining trees based on maximum-likelihood criteria. Bootstrapping algorithms from Splitstree (30) were used to estimate the robustness of the phylogenetic analyses.

Nucleotide sequence accession numbers. The DNA sequences of the 16S RNA, *recA*, Y-HSP60, *gyrB*, and *glnA* gene loci have been deposited in GenBank under accession numbers AY332802 though AY333069.

RESULTS

16S RNA sequence variations in yersiniae. The 16S RNA gene fragments of all the isolates were conserved: 55 (ca. 95%) of the 58 *Yersinia* strains we analyzed had pairwise distances of <0.062 (Fig. 1). The three exceptions were *Y. frederiksenii* WE 83/02, *Y. intermedia* WA 11/94, and *Y. kristensenii* WA 948, which had larger genetic distances, up to 0.132. The 16S RNA sequences of four species (*Y. pestis*, *Y. pseudotuberculosis*, *Y. aldovae*, and *Y. ruckeri*), did not vary; i.e., a single 16S RNA allele type was identified in each species (Table 1). The isolates of two other species, *Y. intermedia* and *Y. rohdei*, were very closely related to the other isolates of the same species; i.e., the distances between conspecific isolates were <0.008. Similarly, eight of nine *Y. bercovieri* isolates had maximal distances of <0.005, and the ninth isolate, ATCC 43970, grouped in a distinct cluster together with three *Y. frederiksenii* strains (Fig. 1). The two *Y. pseudotuberculosis* strains we examined were also very closely related, and they clustered, together with *Y. pestis* KIM, in a distinct genetic cluster (Fig. 1).

The remaining four species (*Y. enterocolitica*, *Y. frederiksenii*, *Y. kristensenii*, and *Y. mollaretii*) appeared to be genetically more variable than the above-mentioned yersiniae. Two groups of *Y. enterocolitica* strains were identified, one containing three strains (WE 22708, A1074, and 1476) and one composed of two strains (8081 and YE 37). The two groups had within-group differences of <0.01 and between-group differences of ca. 0.025 (Fig. 1). Similarly, two groups were identified among the seven *Y. frederiksenii* strains we examined, one composed of four strains (ATCC 33641, WA 933, 7175, and WS 52/02) and one containing three strains (867, CNY 867, and WA 935). As noted above, the eighth *Y. frederiksenii* strain (WE 83/02) was a clear "outsider." The pairwise distances within the group containing four isolates were <0.009. Five isolates of *Y. kristensenii* (490, 105, WE 180/98, WS 45/98, and WE 414/93) were closely related and had pairwise distances of <0.009. However, *Y. kristensenii* WS 052 and ATCC 33638 were scattered across the 16S RNA maximum-likelihood tree (Fig. 1), and the distances to their nominal conspecifics ranged from 0.011 to 0.049. *Y. kristensenii* WA 948 was the most divergent strain among all of the *Yersinia* strains we analyzed, with the exception of the *Y. ruckeri* strains (Fig. 1). Six of the eight *Y. mollaretii* strains (including the ATCC 43969 type strain, also known as CNY 7263) were grouped together in a tight cluster (<0.003),

TABLE 1. *Yersinia* strains and the distribution of their 16S RNA, *glnA*, *gyrB*, *recA*, and Y-HSP60 gene alleles

Species	Strain	Serotype	Source	Country	Yr of isolation	Allele type (NT/AT)				
						16S RNA	*glnA*	*gyrB*	*recA*	Y-HSP60
Y. aldovae	35236	O:17				10	4/2	10/2		21/10
Y. aldovae	CNY7112	O:21				10	4/2	10/2	9/1 (4)[a]	21/10
Y. bercovieri	6519	O:58.73	Human	Georgia		12	25/6			7/6
Y. bercovieri	ATCC 43970	O:58.73	Human	France		12	25/6	2/3	23/2	7/7
Y. bercovieri	W601	O:8	Human	Belgium		13		1/3	23/2	
Y. bercovieri	WA 17/96	O:58.73	Human	New Zealand	1996	12	26/6	3/3	23/2	8/7
Y. bercovieri	WA 315	O:58.73	Vegetable	France		12		3/3	24/2	
Y. bercovieri	WE 82/87	O:70.36	Human	Belgium		13	25/6	3/3	23/2	8/7
Y. bercovieri	WS 19/96	O:58.73	Human	Belgium	1996	12	26/6	3/3	23/2	8/7
Y. bercovieri	WS 27/96	O:70.36	Human	Belgium	1996	12	25/6	3/3	23/2	8/7
Y. bercovieri	WS 39/98	O:69	Human	Belgium	1998	12	25/6	1/3	23/2	8/7
Y. enterocolitica	8081	O:8	Human	United States		7	11/3	16/8	20/2	22/2
Y. enterocolitica	A1074	O:6	Human	United States		8			21/2	24/12
Y. enterocolitica	W22708	O:9				6	10/9	26/7	19/2	
Y. enterocolitica	YE 37	O:8				7	11/3	27/8	20/2	22/2
Y. enterocolitica	1476	O:4.33				8	10/3	26/7		
Y. frederiksenii	7175	O:38	Pig	Belgium		24		21/13	15/3	
Y. frederiksenii	867	O:16a, 16b	Human	Belgium		3	24/1	19/12	10/2	17/1
Y. frederiksenii	ATCC 33641	O:38	Sewage	Denmark		22	18/6	21/13	17/3	10/4
Y. frederiksenii	CNY 867	O:16.29				3	24/1	19/12	10/2	17/1
Y. frederiksenii	WA 933	O:60	Unknown	Germany	1985	23	19/6	22/11	18/3	12/4
Y. frederiksenii	WA 935	O:58	Unknown	Germany	1985	2	35/1	35/4	31/2	31/8
Y. frederiksenii	WE 83/02	O:39	Human	Belgium	2002	23	17/6	20/13		10/5
Y. frederiksenii	WS 52/02	O:48	Human	Belgium	2002	24	18/6	20/13	16/3	11/4
Y. intermedia	955	O:17	Water	Norway		11	5/2	11/1	4/1	28/4
Y. intermedia	WA 11/94	O:55	Human	New Zealand	1994	26	6/2	12/1	8/1	29/4
Y. intermedia	WE 24/02	O:14	Human	Belgium	2002	11	8/2	12/1	6/1	29/4
Y. intermedia	WE 57/93	O:4.33	Human	Belgium	1993	14	9/2	14/1	5/1	31/8
Y. intermedia	WS 11/01	O:48	Human	Belgium	2001	14	7/2	13/1	4/1	29/4
Y. intermedia	WS 58/02	O:57	Human	Belgium	2002	10	9/2	15/1	7/1	30/8
Y. kristensenii	105	O:10.34	Human	Denmark		15	12/3	29/6	32/2	13/1
Y. kristensenii	490	O:12.25	Hare	France		19	16/3		35/1	
Y. kristensenii	ATCC 33638	O:11.23	Human			20	14/3	29/6	34/2	15/1
Y. kristensenii	WA 948	O:16	Human	Finland	1982	1	34/3	9/5	25/2	9/7
Y. kristensenii	WE 180/98	O:11.24	Human	Belgium	1998	15	15/3	30/6	33/2	14/1
Y. kristensenii	WE 414/93	O:12.25	Human	Belgium	1993	10	16/3	32/6	35/1	16/1
Y. kristensenii	WS 45/98	O:12.26	Human	Belgium	1998	16	13/3	31/6	32/2	13/1
Y. kristensenii	WS O52	O:52	Human	Belgium	1996	9	36/3	28/10	22/2	25/14
Y. mollaretii	ATCC 43969	O:59	Soil	United States		17	29/4	6/3	26/3	5/7
Y. mollaretii	WA 134/92	O:3	Food	Denmark	1992	19	33/4		28/3	1/9
Y. mollaretii	WA 309	O:59	Vegetable	France		10	27/4		30/3	
Y. mollaretii	WE 149/92	O:30.47	Human	Belgium	1992	17	30/4	5/3	26/3	4/8
Y. mollaretii	WE 302/92	O:3				17	31/4	8/3	27/3	
Y. mollaretii	WS 43/94	O:62	Meat	Belgium		18	32/4		29/3	3/9
Y. mollaretii	WS 44/98	O:59	Human	Belgium	1998	19	28/4	4/3	30/3	6/8
Y. mollaretii	WS 61/99	O:62	Human	Belgium	1990	17	32/4	7/3	27/3	2/9
Y. rohdei	ATCC 43380	O:38	Dog feces	Germany		4	21/5	23/14		19/13
Y. rohdei	WA 12/96	O:38	Pork	New Zealand	1996	4	22/5	23/14	13/2	18/13
Y. rohdei	WA 14/96	O:76	Pork	New Zealand	1996	5	23/5	23/14	11/2	19/13
Y. rohdei	WA 9/90	O:76	Unknown	Germany	1990	5	20/5	24/14	12/2	20/13
Y. rohdei	WS 59/02	O:76	Human	Belgium	2002	4	20/5	25/14	14/2	19/13
Y. ruckeri	WS 10/95		Fish	Belgium	1995	21	1/7	33/15	36/4	32/11
Y. ruckeri	WS 20/94		Fish	Belgium	1994	21	1/7	34/9	36/1 (4)[a]	32/11
Y. ruckeri	WS 25/94		Fish	Belgium	1994	21	1/7	34/9	37/4	32/11
Y. ruckeri	WS 34/93		Fish	Belgium	1993	21	1/7	34/9	36/4	32/11
Y. pseudotuberculosis	CDC542-84					25	2/8	16/8	1/2	27/2
Y. pseudotuberculosis	CDC801-84					25	3/8	17/8	2/2	27/2
Y. pestis KIM	KIM		Human	Kurdistan/Iran		25	3/8	18/8	3/2	26/2
Total ST/AT alleles						26	36/9	35/15	37/4	32/14

[a] For these strains, NTs were different; however, their ATs were identical to AT1. Because only short fragments were available for amino acid-based analysis, the AT4 designation has been conditionally assigned to the strains, to indicate that their ATs may be different from AT1.

and the remaining two strains, *Y. mollaretii* WS 43/94 and WA 309, appeared to be more closely related to *Y. enterocolitica* and *Y. frederiksenii*, respectively, than to the other strains of *Y. mollaretii.*

Some of the *Yersinia* strains had identical 16S RNA sequences, e.g., *Y. aldovae* CNY 7112 and 35236 and *Y. kristensenii* WE 414/93, *Y. intermedia* WS 58/02 and *Y. kristensenii* 490, and *Y. pestis* KIM and *Y. pseudotuberculosis* CDC801-84 (Fig. 1). Eight strains did not cluster with their conspecifics: *Y. frederiksenii* WE 83/02, *Y. intermedia* WA 11/94, *Y. kristensenii* WA 948, *Y. mollaretii* WA 309 and WS 43/94, *Y. bercovieri* ATCC 43970, and *Y. kristensenii* WS 052 and ATCC 33638.

TABLE 2. Primers used for MLST of the *Yersinia* species, average G+C content, polymorphic sites, and d_N/d_S ratios determined by START for the target gene loci among 35 strains

Gene	Primers (5′→3′)	Amplicon size (bp)	Fragment (bp) analyzed by START	Avg G+C content (%)	No. of polymorphic sites	d_N/d_S
16S RNA	AGTTTGATCATGGCTCAG TTACCGCGGCTGCTGGCA	500	261	57.84	12	
glnA	CGATTGGTGGCTGGAAAGGC TTGGTCATRGTRTTGAAGCG	530	387	51.15	109	0.0289
gyrB	CGGCGGTTTGCAYGGYGTRGG CAGSGTRCGRGTCATYGCCG	545	387	46.81	105	0.0378
recA	GGGCCAAATTGAAAARCARTTCGG CGCCRATYTTCATRCGRATYTGGT	560	351	47.90	92	0.0052
Y-HSP60	GACGTNGTAGAAGGTATGYAG CGCCGCCAGCCAGTTTAGC	565	372	47.93	74	0.0186

Y-HSP60, *glnA*, *recA*, and *gyrB* sequence variations in *Yersinia*. All of the four loci were polymorphic: *glnA*, *gyrB*, *recA*, and Y-HSP60 had 109, 105, 92, and 74 polymorphic sites, respectively (Table 2). The calculated d_N/d_S ratios for those housekeeping genes ranged from 0.0052 for the *recA* locus to 0.0378 for the *gyrB* locus. The G+C contents of the loci ranged from 46.81% for the *gyrB* locus to 57.84% for the 16S RNA locus (Table 2).

Two different allelic designations were assigned to each sequence, one based on the nucleotide sequence (NT) and the other on the translated amino acid sequence (AT). Thirty-six, 35, 37, and 32 NTs were identified for the *glnA*, *gyrB*, *recA*, and Y-HSP60 genes, respectively (Table 1). With the exception of *Y. pestis* and *Y. pseudotuberculosis*, identical NTs were observed only in strains within the same species, and strains possessing identical NTs for all of the loci were very rare (e.g., *Y. bercovieri* WA 17/96 and WS 19/96 had identical alleles for all four loci). One exception, where the strains of two different species had the same NTs, was *Y. pestis* KIM and the two *Y. pseudotuberculosis* isolates: these three strains shared identical ATs for each housekeeping gene locus examined. Translation of the nucleotide sequences into amino acid sequences resulted in fewer alleles (ATs) than NTs, which suggests that most of the nucleotide substitutions were silent (Table 1). For example, the 37 NTs for the *recA* locus translated into only 4 ATs. Overall, there were 9, 15, 4, and 14 ATs for *glnA*, *gyrB*, *recA*, and Y-HSP60, respectively.

Some isolates of several *Yersinia* species shared identical ATs at one or more of the four housekeeping loci. As noted above, isolates of *Y. pestis* and *Y. pseudotuberculosis* shared ATs at all four loci. For *glnA*, *Y. aldovae* and *Y. intermedia* shared AT2, *Y. enterocolitica* and *Y. kristensenii* shared AT3, and *Y. frederiksenii* and *Y. bercovieri* shared AT3. For *gyrB*, *Y. bercovieri* and *Y. mollaretii* shared AT3, and *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis* shared AT8. For Y-HSP60, *Y. pseudotuberculosis*, *Y. pestis*, and *Y. enterocolitica* shared AT2, *Y. intermedia* and *Y. frederiksenii* shared AT4, *Y. kristensenii* and *Y. frederiksenii* shared AT1, and one strain each of *Y. kristensenii* and *Y. mollaretii* shared AT7, which was commonly found in the *Y. bercovieri* strains. The smallest number of ATs was observed for *recA*. The number of *recA*-based ATs identified was less than the number of currently known species of *Yersinia*; therefore, not surprisingly, the *recA*-based ATs were commonly shared across the species. For example, 28 isolates representing six *Yersinia* species possessed *recA* AT2, and 12 strains of *Y. frederiksenii* and *Y. mollaretii* had *recA* AT3. In contrast, with the exception of one possible match with *Y. aldovae* CNY 7112, *Y. ruckeri*'s *recA* AT4 was not found in any other species. Moreover, the data obtained during our analysis of the other three housekeeping genes of *Y. ruckeri* indicated that their ATs also were unique to *Y. ruckeri*, i.e., they were not shared with any strains of the other *Yersinia* species.

Genetic relationships among yersiniae. Inter- and intraspecies relationships among various yersiniae were examined by analyzing a 16S RNA-based maximum-likelihood tree (Fig. 1) and an unrooted phylogram (Fig. 2) constructed from the 1,407-bp concatenated sequences of all four housekeeping MLST loci. The 16S RNA sequences were not included in the concatenated locus, because multiple copies of the gene exist in various *Yersinia* strains (e.g., 6 and 7 copies in *Y. pestis* strains CO92 and KIM, respectively). The availability of the two trees enabled us to compare the genetic relatedness data obtained by 16S RNA sequencing with those obtained by MLST analysis of a concatenated tree containing four housekeeping gene loci.

The concatenated tree was constructed based on the 39 isolates for which we had complete sequence data for all four loci (Fig. 2). The intraspecies genetic distances for seven species (*Y. bercovieri*, *Y. enterocolitica*, *Y. intermedia*, *Y. mollaretii*, *Y. pseudotuberculosis*, *Y. rohdei*, and *Y. ruckeri*) were <0.02. However, in two species (*Y. kristensenii* and *Y. frederiksenii*), all of the isolates did not cluster together. For example, four isolates of *Y. kristensenii* (WE 180/98, WS 45/98, 105, and ATCC 33638) diverged from each other by <0.01, while the other two isolates, WE 414/93 and WS 052, differed from the others by about 0.04 and 0.09, respectively (Fig. 2). The pairwise distances for four strains of *Y. frederiksenii* were <0.04, and they were >0.06 for the remaining isolates (Fig. 2). Only one *Y. pestis* isolate was analyzed; therefore, the genetic distances among strains of that species could not be determined. However, data obtained from the 16S RNA- and MLST-based trees (Fig. 1 and 2, respectively) indicated that the single *Y. pestis* strain we examined clustered tightly with the *Y. pseudotuberculosis* strains. Furthermore, as noted above, both species shared the same 16S RNA allele (NT25) and all four ATs (AT8, AT8, AT2, and AT2 for *glnA*, *gyrB*, *recA*, and Y-HSP60, respectively [Table 1]).

The genetic distances between species ranged from 0.062 to 0.134, with the exception of *Y. pestis* and *Y. pseudotuberculosis*, for which it was <0.01. Other species that were closely related

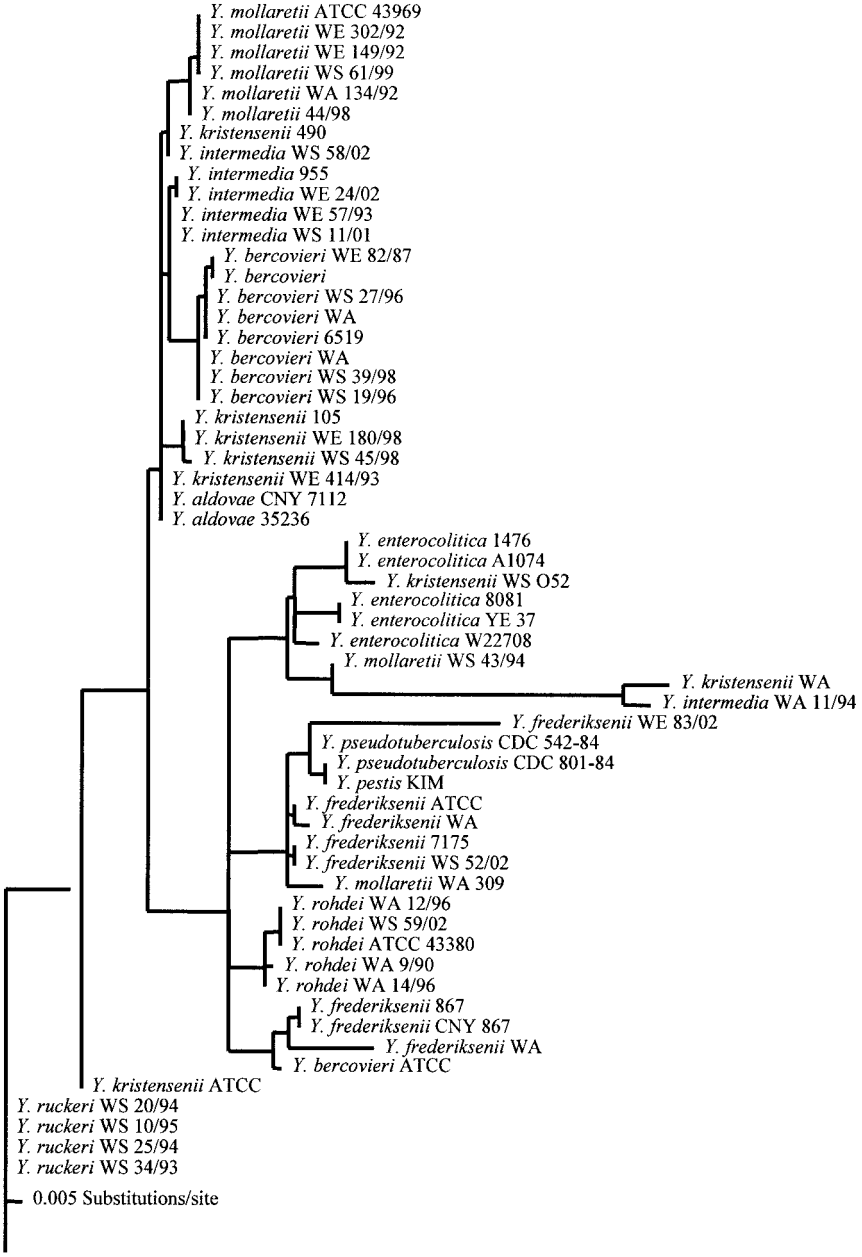

FIG. 1. Genetic relatedness among *Yersinia* species, based on 16S RNA sequences (maximum-likelihood tree).

were *Y. aldovae* and *Y. intermedia* (pairwise genetic distances between isolates ranged from 0.066 to 0.068) and *Y. bercovieri* and *Y. mollaretii* (pairwise genetic distances ranged from 0.062 to 0.066). The *Y. ruckeri* isolates exhibited the largest genetic distances (0.116 to 0.134) from the other *Yersinia* species.

Comparison of the 16S RNA and MLST data. The data obtained by 16S RNA analysis (Fig. 1) and MLST (Fig. 2) of 39 *Yersinia* strains placed 35 of the isolates in their nominal species-specific group. The four discordant strains were *Y. kristensenii* WS 052, *Y. kristensenii* WA 948, *Y. frederiksenii* WA 935, and *Y. bercovieri* ATCC 43970. *Y. kristensenii* WS 052 appeared to be more closely related to the *Y. enterocolitica* group (based on the 16S RNA analysis [Fig. 1]) and to the *Y. enterocolitica*, *Y. frederiksenii*, *and Y. rohdei* groups (based on the MLST analysis [Fig. 2]) than to other *Y. kristensenii* strains. In addition, according to the 16S RNA analysis (Fig. 1), *Y. kristensenii* WA 948 discordantly clustered with *Y. intermedia* WA 11/94—another "outsider" strain. The former strain also was observed, by MLST, to be discordantly but not tightly clustered with *Y. bercovieri* (Fig. 2). Based on the 16S RNA analysis (Fig. 1), *Y. frederiksenii* WA 935 was genetically distinct and not closely related to other *Y. frederiksenii* strains; however, it clustered with two other *Y. frederiksenii* strains that formed one of the two *Y. frederiksenii* clusters on the concatenated tree (Fig. 2). Similarly, based on the 16S RNA analysis (Fig. 1), *Y. bercovieri* ATCC 43970 did not cluster with the other *Y. bercovieri* isolates; however, it clustered tightly with four other *Y. bercovieri* strains on the MLST tree (Fig. 2).

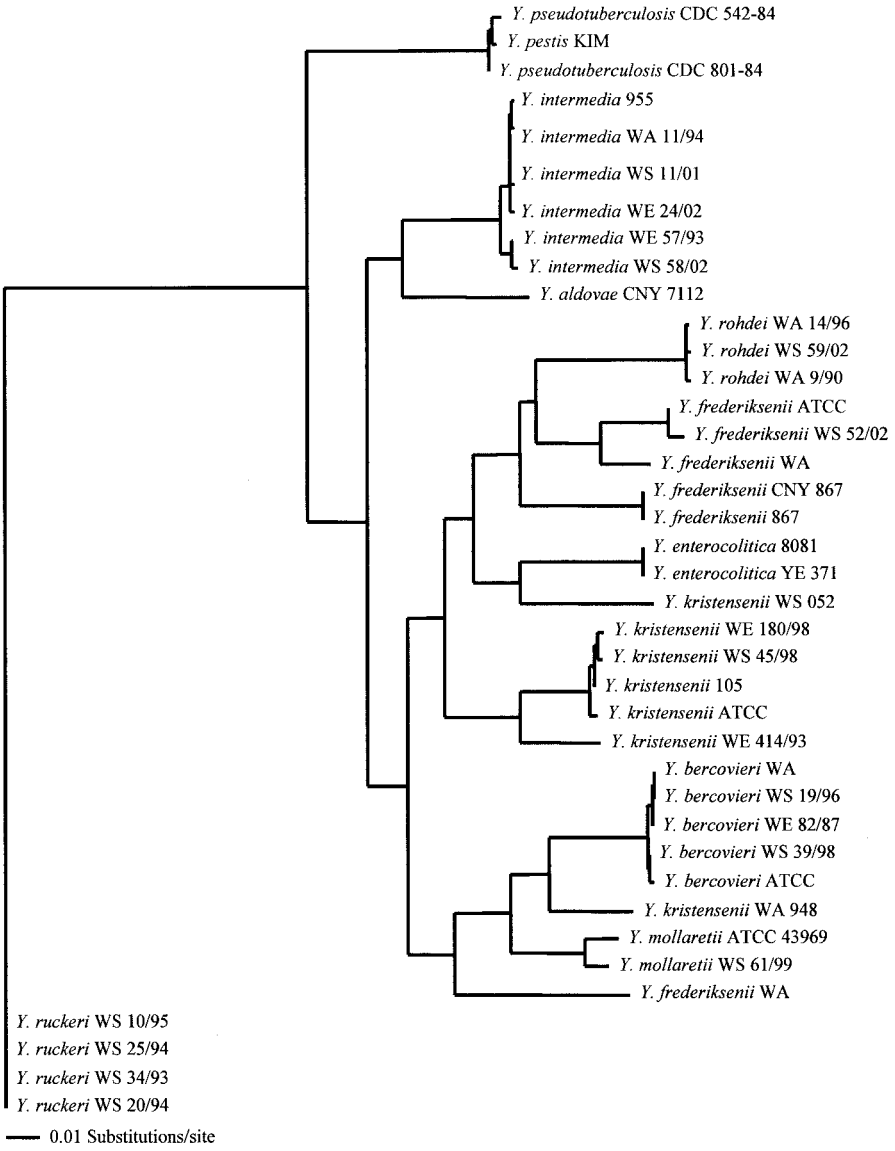

FIG. 2. Genetic relatedness among *Yersinia* species, based on concatenated sequences of the *glnA*, *gyrB*, *recA*, and Y-HSP60 loci (maximum-likelihood tree).

Evidence for O-antigen switching and genetic recombination in *Yersinia*. Several strains with identical or closely related genetic backgrounds had distinct O antigens; e.g., although *Y. intermedia* WS 11/01 and WE 24/02 were very closely related (Fig. 1 and 2), they had distinct O antigens (O:48 and O:14, respectively) (Table 1). Also, in a few cases, O antigens were shared by strains of species that were not closely related; e.g., *Y. bercovieri* W601 shared the O:8 antigen with the genetically distant *Y. enterocolitica* 8081 and YE 37 (Table 1; Fig. 1 and 2).

The Splitstree analysis revealed parallelograms, for each individual locus and for the concatenated sequences, among strains within and among various *Yersinia* species. Parallelograms in the splitsgraph were accepted as evidence for parallel changes only if (i) the fit values were above 95% (indicating that most of the variation was accounted for in the analysis) and (ii) the bootstrap values on adjacent sides were greater than 90%. The criteria were never met for data sets containing the sequences for all the strains, presumably because the data from the genetically diverse species were too complex. However, the criteria were routinely met when small subsets of the sequences (i.e., simpler data sets) were analyzed. An example of such an analysis for a subset of *Y. enterocolitica*, *Y. frederiksenii*, *Y. intermedia*, *Y. kristensenii*, and *Y. mollaretii* strains is presented in Fig. 3. As shown in Fig. 3, the parallelograms often had adjacent sides with very high bootstrap values (e.g., all bootstrap values for the *gyrB* and *recA* loci had a maximum possible value of 100), and the fit values for all loci were at their maximum possible value of 100.

DISCUSSION

Genetic relatedness among *Yersinia* species. MLST is a relatively new approach that has been used to determine the genetic relatedness among strains of various bacterial species

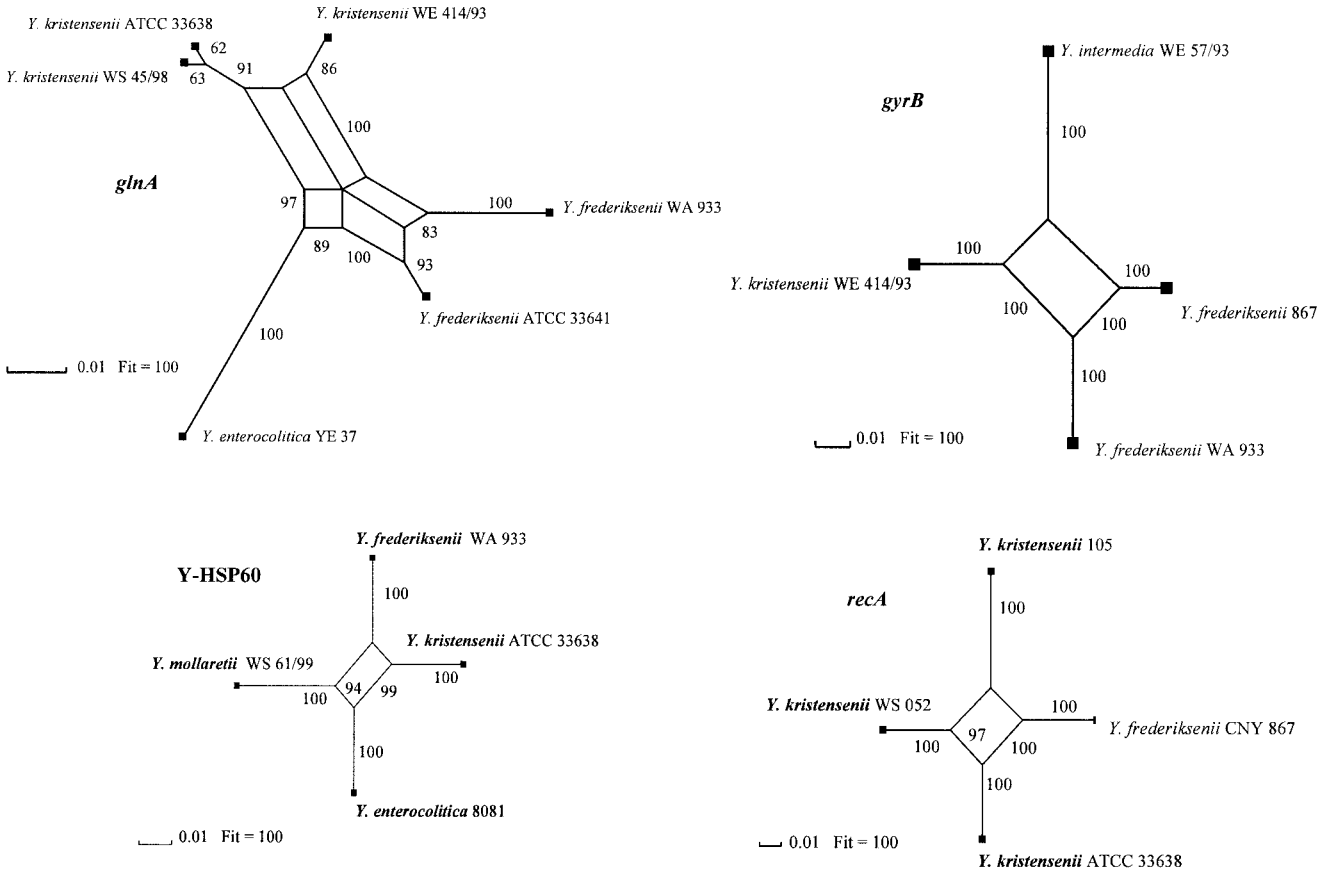

FIG. 3. Split decomposition analysis of randomly selected strains of various *Yersinia* species.

and, less commonly, among various species. The three most intensively studied, pathogenic *Yersinia* species (*Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica*) have been characterized by MLST (1). In that study, no sequence diversity was found in any of the six genes among the *Y. pestis* isolates examined, and the alleles identified were identical or nearly identical to alleles in *Y. pseudotuberculosis*. Based on these observations, the authors concluded that *Y. pestis* is an extremely homogeneous species that evolved from *Y. pseudotuberculosis* shortly before the first known pandemic of human plague. Other investigators have also suggested close genetic relatedness between these two species (1, 27, 45). In agreement with the above-cited reports, *Y. pestis* and *Y. pseudotuberculosis* were found to be very closely related in our study too, based on all criteria examined, including 16S RNA- and MLST-based clustering (Fig. 1 and 2, respectively) and their allelic profiles (both species shared the same 16S RNA allele and all four ATs [Table 1]).

In addition to *Y. pestis* and *Y. pseudotuberculosis*, five *Yersinia* species (*Y. aldovae*, *Y. bercovieri*, *Y. intermedia*, *Y. rohdei*, *and Y. ruckeri*) were genetically fairly homogeneous. For example, based on the sequence analysis of the 16S RNA gene loci, all four *Y. ruckeri* strains and both of the *Y. aldovae* strains we examined were identical for each species, even though some of the strains were isolated in different years (e.g., *Y. ruckeri* WS 34/93 and WS 10/95 were isolated in Belgium in 1993 and 1995, respectively) or had different serotypes (e.g., the serotypes of *Y. aldovae* 35236 and CNY 7112 were O:17 and O:21, respectively). The *Y. intermedia* and *Y. rohdei* strains were also genetically homogeneous. The distances between conspecific isolates were <0.008, and some strains isolated ca. 6 years apart on different continents (e.g., *Y. rohdei* WA 9/90 and WA 14/96, which had the same serotype, O:76, were isolated in Germany in 1990 and in New Zealand in 1996, respectively) were identical based on their 16S RNA sequences. The two identical strains of *Y. bercovieri*, WA 17/96 and WS 19/96, were isolated from humans in 1996. They had identical alleles for all five loci (Table 1), and they shared the O:58,73 antigen; however, they were isolated from geographically distinct locations (New Zealand and Belgium, respectively).

The genetic homogeneity of *Y. aldovae*, *Y. intermedia*, *Y. rohdei*, and *Y. ruckeri* has been suggested by other investigators (22, 48), and our observations are in agreement with those earlier reports. For example, our 16S RNA and MLST data regarding the genetic homogeneity of *Y. ruckeri* are in agreement with a previous report (22) in which *Y. ruckeri* strains isolated from a wide variety of sources were shown to be clonal, based on their plasmid profile and ribotyping pattern analysis, and one plasmid type and ribotype predominated. In addition, MLEE analysis (48) of *Y. ruckeri* strains indicated that they were primarily of one electrophoretic type. In contrast, *Y. enterocolitica*, *Y. frederiksenii*, *Y. kristensenii*, and *Y.*

mollaretii were genetically more heterogeneous, which is also in agreement with previous reports (12, 17, 57).

Comparison of 16S RNA analysis and MLST data regarding genetic relatedness among *Yersinia* species. In most, but not all, instances, the MLST and 16S RNA data correlated well with each other, and they agreed with the species' designations determined by biochemical tests. The MLST and 16S RNA data also correlated well in a few cases (e.g., *Y. kristensenii* WS 052 and WA 948) where the results of both analyses did not agree with the biochemical designation of the species; i.e., according to the 16S RNA and MLST analyses, the two strains mentioned above were clear outsiders in relation to other strains of the same species. Therefore, as discussed below, their taxonomic designations should be reexamined.

In instances where the 16S RNA and MLST data did not correlate with each other, MLST was more consistent with the species' biochemical designations than was 16S RNA analysis. For example, with the exception of one isolate (*Y. bercovieri* strain ATCC 43970), all of the *Y. bercovieri* strains were either identical or very closely related, based on their 16S RNA sequence analysis (Fig. 1). The ATCC 43970 strain (also known as CNY 7506) was a clear outsider, and it appeared to be more closely related to other *Yersinia* species (e.g., *Y. frederiksenii*) than to the other *Y. bercovieri* strains we analyzed. Similarly, according to the 16S RNA data (Fig. 1), *Y. kristensenii* ATCC 33638 (also known as CIP 80-30) did not cluster with the other *Y. kristensenii* strains. Both ATCC strains (*Y. bercovieri* ATCC 43970 and *Y. kristensenii* ATCC 33638) are the type isolates of their respective species; therefore, our observation that they were genetically distinct (according to the 16S RNA analysis) from other strains of their respective species was puzzling. However, when five of the nine *Y. bercovieri* strains (including ATCC 43970) and seven of the eight *Y. kristensenii* strains (including ATCC 33638) were analyzed by MLST using concatenated trees based on Y-HSP60, *glnA*, *gyrB*, and *recA* loci, both type strains correctly clustered with isolates of their respective species (Fig. 2). Another example of MLST being better than 16S RNA sequencing for determining phylogenetic identity is *Y. frederiksenii* WA 935. According to 16S RNA analysis, that strain was genetically distinct from other *Y. frederiksenii* strains (Fig. 1), but it correctly clustered with other *Y. frederiksenii* strains on the concatenated MLST tree (Fig. 2).

16S RNA sequencing has been and still is commonly used to identify various species, and its ability to yield reproducible and accurate information is commonly accepted. However, an increasing body of data suggests that 16S RNA does not always ideally identify/resolve strains on a species level. For example, a recent study (15) of *Y. frederiksenii* reported that *gyrB* was a better phylogenetic marker than that species' 16S RNA sequence. Also, as noted above, interspecies recombinations in 16S RNA genes, which initially were thought to be rare, have recently been suggested (52) to occur in at least some bacterial species, and such genetic exchanges may affect the determination of the phylogeny/taxonomy of bacterial species in which the changes have occurred (39). Our data support these observations, and they suggest that (i) although 16S RNA sequences are reasonably good phylogenetic markers for differentiating *Yersinia* species, genetic relatedness data obtained at least by the partial sequencing of this locus may not clearly distinguish between some members of this genus, and (ii) MLST is better suited for studying genetic relatedness among yersiniae than is 16S RNA-based analysis. On a more general level, our data support the idea that MLST can be an invaluable tool during continued refinement of the phylogenetic relationships among enteric bacteria (3).

16S RNA analysis, MLST, and *Yersinia* taxonomy. In the genus *Yersinia* (as in other bacteria), species have been identified mainly according to certain biochemical tests. The introduction of DNA-based approaches has revolutionized bacterial taxonomy and has helped assign species designations to numerous bacteria, including yersiniae (11). 16S RNA analysis has been particularly useful in that regard, and it has become a molecular standard for identification of many bacterial species, including uncultivable bacteria (29). However, the question about the minimal percent difference in 16S RNA sequences needed to differentiate bacterial species has not yet been resolved, and the issue of whether a 2% or larger difference is required to assign strains to different species is still a matter of debate (6, 14). The results of our present analysis of various *Yersinia* species suggest that MLST can be useful in that regard and that MLST is better than 16S RNA analysis for determining the taxonomic designations of the yersiniae.

The MLST data obtained during our study were based on four housekeeping gene loci that are generally considered to be conserved; therefore, a selection of conserved loci would be expected to represent the minimum differences between homologous loci in two genomes. Relatively few strains of the various *Yersinia* species were analyzed during our study, and therefore it is difficult to state with confidence whether there are "rules of thumb" for identifying the genetic distances between isolates of different *Yersinia* species. However, our data provide some preliminary insight into this issue. For example, based on our MLST data for seven *Yersinia* species (*Y. bercovieri*, *Y. enterocolitica*, *Y. intermedia*, *Y. mollaretii*, *Y. pseudotuberculosis*, *Y. rohdei*, and *Y. ruckeri*), the intraspecies genetic distances were <0.02, while the interspecies distances were >0.06. From these data, we observed a bimodal distribution in which intraspecies distances were in the first mode and interspecies distances were in the second mode. Thus, there is the potential to separate species using a bimodal distribution and a suitable statistic such as Giacomelli's dip statistic (23).

Several strains could not be distinguished on a species level by the partial sequencing of the 16S RNA gene but were identified by MLST, which suggests that MLST, but not 16S RNA analysis, has sufficient resolving ability to unambiguously differentiate the isolates of the closely related *Yersinia* species. Additional studies are required to fine-tune and validate the cutoff values for interspecies and intraspecies genetic distances, and their results may have a profound impact on clarifying the taxonomic classification of various bacterial species (53). Indeed, MLST already has been proposed (3) as a useful tool for refining the taxonomy and phylogenetic relationships among enteric bacteria.

We observed four exceptions to the above-mentioned bimodal distribution of the interspecies and intraspecies genetic distances between *Yersinia* isolates. The first two exceptions are *Y. pestis* and *Y. pseudotuberculosis* strains, which had genetic distances similar to those of intraspecies genetic distances. In addition, the same strains also (i) had the same 16S RNA allele (NT25), (ii) had identical ATs for all four housekeeping genes

we analyzed (Table 1), and (iii) clustered tightly, as determined by 16S RNA sequencing (Fig. 1) and MLST (Fig. 2). The genomes of *Y. pestis* and *Y. pseudotuberculosis* are known to be very similar, based on the results of (i) DNA hybridization studies (8), (ii) 16S RNA analyses (56), and (iii) MLST studies (1), which has led to the hypothesis that the two species are actually two subgroups of the same species. Our results are in agreement with those previous reports, and they support the idea that *Y. pestis* and *Y. pseudotuberculosis* are two lineages within the same species rather than two distinct species. The third and fourth exceptions are *Y. kristensenii* and *Y. frederiksenii*, respectively. Some strains of those species (e.g., *Y. kristensenii* WA 948 and WS 052, and *Y. frederiksenii* WA 935) were separated from other strains of their respective species by genetic distances larger than those observed for other strains within the same species (i.e., the genetic distances were >0.06). Furthermore, on the maximum-likelihood trees, the same strains were not closely related to other strains of their respective species, based on the 16S RNA (Fig. 1) and concatenated MLST (Fig. 2) analyses. *Y. frederiksenii* has been shown (15–17, 33), by the results of comparative 16S RNA and *gyrB* sequence analyses, DNA hybridization studies, and MLEE analysis, to include strains that may represent a new subline within the genus *Yersinia* which is most closely related to *Y. frederiksenii* hybridization group 2 (unnamed genomospecies 2) and is biochemically indistinguishable from typical *Y. frederiksenii* (*Y. frederiksenii* genomospecies 1). Based on our 16S RNA sequencing and MLST data (Fig. 1 and 2), the *Y. frederiksenii* strains in our collection grouped in distinct clusters, which provides additional support for the idea that the species contains multiple subgroups, some of which may represent unknown species in *Yersinia*. Also, our MLST analyses (Fig. 2) identified at least one genetically atypical *Y. frederiksenii* isolate (WA 935) that was a clear outsider with regard to the other *Y. frederiksenii* isolates we examined.

Dolina and Peduzzi (17) found that some of the *Y. kristensenii* strains in their collection were genetically more closely related to *Y. mollaretii* than to other strains of *Y. kristensenii*, as determined by MLEE analysis. Our MLST data obtained during our study of the atypical *Y. kristensenii* strains (e.g., *Y. kristensenii* WA 948 and WS 052) in our collection are consistent with their report. The number of *Y. kristensenii* and *Y. frederiksenii* strains analyzed during our study was relatively small, and we detected only a few atypical strains; thus, it is not clear how common those strains are in nature. However, our data support the idea that such strains do exist, and they emphasize the need for additional MLST studies to elucidate the prevalence of such strains and to delineate their taxonomic classification, since they may represent a new species within the genus *Yersinia*.

Y. ruckeri (the etiologic agent of red mouth disease in fish) was a clear outsider with regard to all other yersiniae, based on (i) 16S RNA analysis (Fig. 1) and (ii) analysis of the concatenated sequences (Fig. 2). Moreover, the ATs of the four *Y. ruckeri* strains in our collection were unique to that species; i.e., AT7 (for *glnA*), AT15 and AT9 (for *gyrB*), AT4 (for *recA*), and AT11 (for Y-HSP60) were not found in any strains of any other *Yersinia* species we examined. The taxonomic status of *Y. ruckeri* has been controversial since its initial classification in 1978 (21), and data reported by several investigators (5, 17, 38, 48) suggest that *Y. ruckeri* should not be classified in the genus *Yersinia*. Our 16S RNA and MLST data suggest that *Y. ruckeri* is the genetically most distant species within the genus and that the taxonomic standing of *Y. ruckeri* may need to be reevaluated.

Correlation between the genetic backgrounds of various *Yersinia* species and their O antigens. Many O antigens are shared among various species of *Yersinia*; e.g., *Y. frederiksenii* shares several O antigens with *Y. enterocolitica* (57), and many *Y. kristensenii* strains are agglutinable in the Wauters scheme (58). However, the genetic relatedness of *Yersinia* strains with common O antigens has never been rigorously examined, and the mechanisms responsible for this phenomenon are currently not well understood. During the present study we identified two *Y. mollaretii* strains (WA 134/92 and WE 302/92) and one *Y. bercovieri* strain (W601) that shared the O:3 and O:8 antigens (51, 61), respectively, with *Y. enterocolitica* (Table 1). *Y. mollaretii* strains with *Y. enterocolitica* O antigens have been described previously (4), and our data provide additional evidence for the existence of such strains. Genetically, the two *Y. mollaretii* strains with the O:3 antigen were closely related to one another, according to their 16S RNA sequences (Fig. 1) and their shared NTs (Table 1). However, both strains were genetically distinct from all of the *Y. enterocolitica* strains we examined (Table 1 and Fig. 1). Similarly, *Y. bercovieri* W601 was not closely related to the *Y. enterocolitica* strains, including the two strains (8081 and YE 371) with which it shared the O:8 antigen (Fig. 1 and 2). These observations suggest that O-antigen switching, which has been proposed (54) to have occurred among several other bacteria, may have also occurred between at least some *Yersinia* species.

Evidence of intraspecies and interspecies genetic recombination in *Yersinia*. The START analysis of our MLST data revealed a noticeable difference in the G+C contents of the genes we analyzed (Table 2). The G+C content differed by as much as 11% among the various loci; e.g., the mean G+C contents for the *gyrB* and 16S RNA loci were 47% and 58%, respectively. The backbone G+C content of the *Yersinia* genome is estimated to range from 46 to 48% (5), and with the exception of the 16S RNA loci, which had a higher G+C content, the remaining loci we analyzed were within this typical G+C range for *Yersinia*. Furthermore, species-specific differences in the G+C content were not observed for any of the loci we examined; i.e., there was no apparent evidence that the loci were acquired from distinct ancestors.

As an additional approach to detect recombinational changes, we used the Splitstree analysis, utilizing small subsets of randomly selected strains of various *Yersinia* species (simultaneously analyzing many strains overwhelmed the program, in agreement with what has been reported previously for other bacteria [37, 47]). The observation of parallelograms during Splitstree analysis indicates that recombination played a role in the evolution of the gene under examination, and we observed such parallelograms during our analyses (one example of such analyses for each of the four housekeeping genes is presented in Fig. 3). The bootstrapping values were greater than 90% (Fig. 3)—in many cases, the bootstrapping values were at their maximum of 100% and the fit values were the maximum possible value of 100—indicating robust results based on multiple genetic changes. We consider the possibility that these multiple

changes resulted from multiple, independent, recurrent mutations to be very small, particularly since nearly all of the nucleotide changes failed to result in amino acid changes (Table 1). Thus, the parallel changes are likely to be the result of recombination rather than independent recurrent mutations. Also, the d_N/d_S ratios for all loci were <1 (Table 2), suggesting that there is no selective pressure on nucleotide changes in these genes.

Parallelograms were also formed during our Splitstree analysis of the 16S RNA loci (data not shown) in strains within the same species and in strains of various *Yersinia* species. Interspecies recombinations in 16S RNA genes have recently been proposed (52) to occur in some non-*Yersinia* species, and our data suggest that the same phenomenon also may occur among the yersiniae. However, caution is required in interpreting these data, because a relatively small subset of strains was analyzed, and the number of nucleotide differences between the sequences is very small. Nevertheless, our data suggest that recombination in the 16S RNA genes of various *Yersinia* species is a possibility, and they indicate that additional studies in that direction are warranted. Such interspecies recombinations among the yersiniae could have important practical implications; e.g., they may complicate the development of 16S RNA-targeted tools to detect and identify various *Yersinia* species. These observations also support the idea (28) that gene trees based on a single gene (including conserved genes, such as those for 16S RNA) are not likely to permit accurate determination of the evolutionary history of the species. Therefore, analysis of several loci of the bacterial chromosome (e.g., by using MLST) is required for a rigorous understanding of the evolutionary history of yersiniae and for the development of advanced tools capable of accurately identifying and differentiating *Yersinia* species.

ACKNOWLEDGMENTS

This study was supported by a BTEP 10/ISTC G597 grant from the Department of Health and Human Services Biotechnology Engagement Program (to A.S.) and a 1R21 AI055660-01A1 grant from the National Institutes of Health (to A.S.).

REFERENCES

1. **Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule, and E. Carniel.** 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. USA **96:**14043–14048.
2. **Adair, D. M., P. L. Worsham, K. K. Hill, A. M. Klevytska, P. J. Jackson, A. M. Friedlander, and P. Keim.** 2000. Diversity in a variable-number tandem repeat from *Yersinia pestis*. J. Clin. Microbiol. **38:**1516–1519.
3. **Ahmad, S., W. G. Weisburg, and R. A. Jensen.** 1990. Evolution of aromatic amino acid biosynthesis and application to the fine-tuned phylogenetic positioning of enteric bacteria. J. Bacteriol. **172:**1051–1061.
4. **Aleksic, S.** 1995. Occurrence of *Y. enterocolitica* antigens O:3, O:9 and O:8 in different *Yersinia* species, their corresponding H antigens and origin. Contrib. Microbiol. Immunol. **13:**89–92.
5. **Aleksic, S., and J. Bockemühl.** 1999. *Yersinia* and other *Enterobacteriaceae*, p. 483–496. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of clinical microbiology, 7th ed. American Society for Microbiology, Washington, D.C.
6. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59:**143–169.
7. **Baker, P. M., and J. J. Farmer III.** 1982. New bacteriophage typing system for *Yersinia enterocolitica*, *Yersinia kristensenii*, *Yersinia frederiksenii*, and *Yersinia intermedia*: correlation with serotyping, biotyping, and antibiotic susceptibility. J. Clin. Microbiol. **15:**491–502.
8. **Bercovier, H., H. H. Mollaret, J. M. Alonso, J. Brault, G. R. Fanning, A. G. Steigerwalt, and D. J. Brenner.** 1980. Intra- and interspecies relatedness of *Yersinia pestis* by DNA hybridization and its relationship to *Yersinia pseudotuberculosis*. Curr. Microbiol. **4:**225–229.
9. **Blumberg, H. M., J. A. Kiehlbauch, and I. K. Wachsmuth.** 1991. Molecular epidemiology of *Yersinia enterocolitica* O:3 infections: use of chromosomal DNA restriction fragment length polymorphisms of rRNA genes. J. Clin. Microbiol. **29:**2368–2374.
10. **Bottone, E. J.** 1997. *Yersinia enterocolitica*: the charisma continues. Clin. Microbiol. Rev. **10:**257–276.
11. **Brenner, D. J., A. G. Steigerwalt, D. P. Falxo, R. E. Weaver, and G. R. Fanning.** 1976. Characterization of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* by deoxyribonucleic acid hybridization and by biochemical reactions. Int. J. Syst. Bacteriol. **26:**180–194.
12. **Caugant, D. A., S. Aleksic, H. H. Mollaret, R. K. Selander, and G. Kapperud.** 1989. Clonal diversity and relationships among strains of *Yersinia enterocolitica*. J. Clin. Microbiol. **27:**2678–2683.
13. **Cavalcanti, Y. V., N. C. Leal, and A. M. De Almeida.** 2002. Typing of *Yersinia pestis* isolates from the state of Ceara, Brazil. Lett. Appl. Microbiol. **35:** 543–547.
14. **Dang, H., and C. R. Lovell.** 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. Appl. Environ. Microbiol. **66:**467–475.
15. **Demarta, A., S. De Respinis, M. Dolina, and R. Peduzzi.** 2004. Molecular typing of *Yersinia frederiksenii* strains by means of 16S rDNA and *gyrB* gene sequence analyses. FEMS Microbiol. Lett. **238:**423–428.
16. **Dolina, M., V. Gaia, and R. Peduzzi.** 1995. Molecular typing of *Yersinia frederiksenii* strains by means of ribotyping and DNA-DNA hybridization. Contrib. Microbiol. Immunol. **13:**140–144.
17. **Dolina, M., and R. Peduzzi.** 1993. Population genetics of human, animal, and environmental *Yersinia* strains. Appl. Environ. Microbiol. **59:**442–450.
18. **Enright, M. C., and B. G. Spratt.** 1999. Multilocus sequence typing. Trends Microbiol. **7:**482–487.
19. **Ewing, B., and P. Greene.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185.
20. **Ewing, B., and P. Greene.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.
21. **Ewing, W. H., A. J. Ross, D. J. Brenner, and G. R. Fanning.** 1978. *Yersinia ruckeri* sp. nov., the redmouth (RM) bacterium. Int. J. Syst. Bacteriol. **28:** 37–44.
22. **Garcia, J. A., L. Dominguez, J. L. Larsen, and K. Pedersen.** 1998. Ribotyping and plasmid profiling of *Yersinia ruckeri*. J. Appl. Microbiol. **85:**949–955.
23. **Giacomelli, F., J. Wiener, J. B. Kruskal, J. V. Pomeranz, and A. V. Loud.** 1971. Subpopulations of blood lymphocytes demonstrated by quantitative cytochemistry. J. Histochem. Cytochem. **19:**426–433.
24. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
25. **Goullet, P., and B. Picard.** 1988. Characterization of *Yersinia enterocolitica*, *Y. intermedia*, *Y. aldovae*, *Y. frederiksenii*, *Y. kristensenii* and *Y. pseudotuberculosis* by electrophoretic polymorphism of acid phosphatase, esterases, and glutamate and malate dehydrogenases. J. Gen. Microbiol. **134:**317–325.
26. **Goullet, P., and B. Picard.** 1984. Distinctive electrophoretic and isoelectric focusing patterns of esterases from *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. J. Gen. Microbiol. **130:**1471–1480.
27. **Hinchliffe, S. J., K. E. Isherwood, R. A. Stabler, M. B. Prentice, A. Rakin, R. A. Nichols, P. C. Oyston, J. Hinds, R. W. Titball, and B. W. Wren.** 2003. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. Genome Res. **13:**2018–2029.
28. **Holmes, E. C., R. Urwin, and M. C. Maiden.** 1999. The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. Mol. Biol. Evol. **16:**741–749.
29. **Hugenholtz, P., C. Pitulle, K. L. Hershberger, and N. R. Pace.** 1998. Novel division level bacterial diversity in a Yellowstone hot spring. J. Bacteriol. **180:** 366–376.
30. **Huson, D. H.** 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics **14:**68–73.
31. **Ibrahim, A.** 1995. Genetic diversity among *Yersinia enterocolitica* strains as revealed by sequence analysis of the 16S rRNA gene. Contrib. Microbiol. Immunol. **13:**277–280.
32. **Ibrahim, A., B. M. Goebel, W. Liesack, M. Griffiths, and E. Stackebrandt.** 1993. The phylogeny of the genus *Yersinia* based on 16S rDNA sequences. FEMS Microbiol. Lett. **114:**173–177.
33. **Ibrahim, A., W. Liesack, A. G. Steigerwalt, D. J. Brenner, E. Stackebrandt, and R. M. Robins-Browne.** 1997. A cluster of atypical *Yersinia* strains with a distinctive 16S rRNA signature. FEMS Microbiol. Lett. **146:**73–78.
34. **Iteman, I., A. Guiyoule, and E. Carniel.** 1996. Comparison of three molecular methods for typing and subtyping pathogenic *Yersinia enterocolitica* strains. J. Med. Microbiol. **45:**48–56.
35. **Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson.** 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23:** 403–405.
36. **Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden.** 2001. Sequence type analysis and recombinational tests (START). Bioinformatics **17:**1230–1231.
37. **Kotetishvili, M., O. C. Stine, A. Kreger, J. G. Morris, Jr., and A. Sulak-**

velidze. 2002. Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. J. Clin. Microbiol. **40:**1626–1635.
38. **Kwaga, J., J. O. Iversen, and V. Misra.** 1992. Detection of pathogenic *Yersinia enterocolitica* by polymerase chain reaction and digoxigenin-labeled polynucleotide probes. J. Clin. Microbiol. **30:**2668–2673.
39. **Lan, R., and P. R. Reeves.** 2001. When does a clone deserve a name? A perspective on bacterial species based on population genetics. Trends Microbiol. **9:**419–424.
40. **Lobato, M. J., E. Landeras, M. A. Gonzalez-Hevia, and M. C. Mendoza.** 1998. Genetic heterogeneity of clinical strains of *Yersinia enterocolitica* traced by ribotyping and relationships between ribotypes, serotypes, and biotypes. J. Clin. Microbiol. **36:**3297–3302.
41. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:**3140–3145.
42. **Mendoza, M. C., R. Alzugaray, E. Landeras, and M. A. Gonzalez-Hevia.** 1996. Discriminatory power and application of ribotyping of *Yersinia enterocolitica* O:3 in an epidemiological study. Eur. J. Clin. Microbiol. Infect. Dis. **15:**220–226.
43. **Najdenski, H., I. Iteman, and E. Carniel.** 1994. Efficient subtyping of pathogenic *Yersinia enterocolitica* strains by pulsed-field gel electrophoresis. J. Clin. Microbiol. **32:**2913–2920.
44. **Nilehn, B.** 1971. Some aspects on phage typing of *Yersinia enterocolitica*. Acta Pathol. Microbiol. Scand. B **79:**446–447.
45. **Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature **413:**523–527.
46. **Perry, R. D., and J. D. Fetherston.** 1997. *Yersinia pestis*—etiologic agent of plague. Clin. Microbiol. Rev. **10:**35–66.
47. **Revazishvili, T., M. Kotetishvili, O. C. Stine, A. S. Kreger, J. G. Morris, Jr., and A. Sulakvelidze.** 2004. Comparative analysis of multilocus sequence typing and pulsed-field gel electrophoresis for characterizing *Listeria monocytogenes* strains isolated from environmental and clinical sources. J. Clin. Microbiol. **42:**276–285.
48. **Schill, W. B., S. R. Phelps, and S. W. Pyle.** 1984. Multilocus electrophoretic assessment of the genetic structure and diversity of *Yersinia ruckeri*. Appl. Environ. Microbiol. **48:**975–979.
49. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873–884.
50. **Shubin, F. N., A. L. Gintsburg, V. M. Kitaev, N. V. Ianishevskii, and Z. G. Zenkova.** 1989. Analysis of the plasmid composition of *Yersinia pseudotuberculosis* strains and its use for typing pseudotuberculosis pathogens. Mol. Gen. Mikrobiol. Virusol. **6:**20–25.
51. **Skurnik, M., and P. Toivanen.** 1993. *Yersinia enterocolitica* lipopolysaccharide: genetics and virulence. Trends Microbiol. **1:**148–152.
52. **Smith, J. M., E. J. Feil, and N. H. Smith.** 2000. Population structure and evolutionary dynamics of pathogenic bacteria. Bioessays **22:**1115–1122.
53. **Stackebrandt, E., W. Frederiksen, G. M. Garrity, P. A. Grimont, P. Kampfer, M. C. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H. G. Truper, L. Vauterin, A. C. Ward, and W. B. Whitman.** 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol. **52:**1043–1047.
54. **Stroeher, U. H., and P. A. Manning.** 1997. *Vibrio cholerae* serotype O139: swapping genes for surface polysaccharide biosynthesis. Trends Microbiol. **5:**178–180.
55. **Sulakvelidze, A.** 2000. Yersiniae other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: the ignored species. Microbes Infect. **2:**497–513.
56. **Trebesius, K., D. Harmsen, A. Rakin, J. Schmelz, and J. Heesemann.** 1998. Development of rRNA-targeted PCR and in situ hybridization with fluorescently labelled oligonucleotides for detection of *Yersinia* species. J. Clin. Microbiol. **36:**2557–2564.
57. **Ursing, J., and S. Aleksic.** 1995. *Yersinia frederiksenii*, a genotypically heterogeneous species with few differential characteristics. Contrib. Microbiol. Immunol. **13:**112–116.
58. **Wauters, G., S. Aleksic, J. Charlier, and G. Schulze.** 1991. Somatic and flagellar antigens of *Yersinia enterocolitica* and related species. Contrib. Microbiol. Immunol. **12:**239–243.
59. **Weinstock, G. M.** 2000. Genomics and bacterial pathogenesis. Emerg. Infect. Dis. **6:**496–504.
60. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74:**5088–5090.
61. **Zhang, L., J. Radziejewska-Lebrecht, D. Krajewska-Pietrasik, P. Toivanen, and M. Skurnik.** 1997. Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O:8. Mol. Microbiol. **23:**63–76.
62. **Zhou, S., W. Deng, T. S. Anantharaman, A. Lim, E. T. Dimalanta, J. Wang, T. Wu, T. Chunhong, R. Creighton, A. Kile, E. Kvikstad, M. Bechner, G. Yen, A. Garic-Stankovic, J. Severin, D. Forrest, R. Runnheim, C. Churas, C. Lamers, N. T. Perna, V. Burland, F. R. Blattner, B. Mishra, and D. C. Schwartz.** 2002. A whole-genome shotgun optical map of *Yersinia pestis* strain KIM. Appl. Environ. Microbiol. **68:**6321–6331.