JOURNAL OF BACTERIOLOGY, Feb. 2002, p. 1019–1027 Vol. 184, No. 4
0021-9193/02/$04.00+0 DOI: 10.1128/JB.184.4.1019–1027.2002

Genetic Variability of *Yersinia pestis* Isolates as Predicted by PCR-Based IS*100* Genotyping and Analysis of Structural Genes Encoding Glycerol-3-Phosphate Dehydrogenase (*glpD*)

Vladimir L. Motin,[1] Anca M. Georgescu,[1] Jeffrey M. Elliott,[1] Ping Hu,[1] Patricia L. Worsham,[2] Linda L. Ott,[1] Tomas R. Slezak,[1] Bahrad A. Sokhansanj,[1] Warren M. Regala,[1] Robert R. Brubaker,[3] and Emilio Garcia[1]*

Lawrence Livermore National Laboratory, University of California, Livermore, California 94550[1]; Bacteriology Division, U.S. Army Medical Research Institute of Infectious Diseases, Frederick, Maryland 21702[2]; and Department of Microbiology, Michigan State University, East Lansing, Michigan 48824[3]

Received 19 June 2001/Accepted 19 November 2001

A PCR-based genotyping system that detects divergence of IS*100* locations within the *Yersinia pestis* genome was used to characterize a large collection of isolates of different biovars and geographical origins. Using sequences derived from the glycerol-negative biovar orientalis strain CO92, a set of 27 locus-specific primers was designed to amplify fragments between the end of IS*100* and its neighboring gene. Geographically diverse members of the orientalis biovar formed a homogeneous group with identical genotype with the exception of strains isolated in Indochina. In contrast, strains belonging to the glycerol-positive biovar antiqua showed a variety of fingerprinting profiles. Moreover, strains of the biovar medievalis (also glycerol positive) clustered together with the antiqua isolates originated from Southeast Asia, suggesting their close phylogenetic relationships. Interestingly, a Manchurian biovar antiqua strain Nicholisk 51 displayed a genotyping pattern typical of biovar orientalis isolates. Analysis of the glycerol pathway in *Y. pestis* suggested that a 93-bp deletion within the *glpD* gene encoding aerobic glycerol-3-phosphate dehydrogenase might account for the glycerol-negative phenotype of the orientalis biovar. The *glpD* gene of strain Nicholisk 51 did not possess this deletion, although it contained two nucleotide substitutions characteristic of the *glpD* version found exclusively in biovar orientalis strains. To account for this close relationship between biovar orientalis strains and the antiqua Nicholisk 51 isolate, we postulate that the latter represents a variant of this biovar with restored ability to ferment glycerol. The fact that such a genetic lesion might be repaired as part of the natural evolutionary process suggests the existence of genetic exchange between different *Yersinia* strains in nature. The relevance of this observation on the emergence of epidemic *Y. pestis* strains is discussed.

Yersinia pestis, the causative agent of bubonic plague, is a recently evolved clone of *Yersinia pseudotuberculosis* serotype O:1b (1, 29). Strains of *Y. pestis* are divided into three biovars on the basis of their abilities to ferment glycerol and to reduce nitrate. These phenotypic differences have proven useful in distinguishing strains thought to be responsible for the three plague pandemics (7). Isolates of the biovar antiqua (able to ferment glycerol and reduce nitrate) are believed to remain as holdovers from the first pandemic that started with the Justinian plague of the 6th century. Strains of the biovar medievalis (glycerol positive and nitrate negative) evidently caused the second pandemic of Europe (Black Death), which was initiated during the 14th century, and strains of the biovar orientalis (glycerol negative and nitrate positive) are responsible for the third plague pandemic of modern times (15, 22).

Several molecular methods that generate fingerprinting patterns of *Y. pestis* DNA have been successfully used for genotyping strains of this microorganism, such as pulse-field gel electrophoresis and ribotyping (15, 16, 18, 25) as well as newly described variable-number tandem repeat analysis (2). Genotyping of *Y. pestis* was also accomplished by restriction fragment length polymorphism (RFLP) detected on Southern blots using probes originating from IS*100*, which is found at least singly on each of the three plasmids of the species and as multiple copies within the chromosome (24). This insertion sequence (IS) element was sequenced (11, 23) and then used by Filippov et al. (9) in conjunction with the novel IS*285* to obtain fingerprint patterns establishing phylogenetic relationships. Further IS*100*- and IS*285*-based RFLP analysis became the method of choice used by many researchers for extensive analysis of *Y. pestis* collections (1, 5, 14, 20). Finally, a third insertion sequence (IS*1541*) was discovered that disrupts *inv*, which encodes invasin (28). This element is now known to be present in many copies in the *Y. pestis* genome and has also been used to obtain RFLP fingerprinting profiles (21, 28).

Many insertions of IS*1541* in the genome of *Y. pestis* strain 6/69 M biovar orientalis were characterized and then tested to see if the insert flanked the same genes in other strains of *Y. pestis* (21). The authors employed five unrelated strains (four of biovar orientalis and one of medievalis) of different ribotypes and IS*1541* hybridization patterns that were isolated from different periods of time and geographical areas. All strains of the biovar orientalis possessed the IS*1541* insertions at the same place (nine flanking genes were determined); the biovar medievalis strain lacked only one such insertion. This

* Corresponding author. Mailing address: Biology and Biotechnology Research Program, L-452, 7000 East Ave., Livermore, CA 94550. Phone: (925) 422-8002. Fax: (925) 422-2282. E-mail: Garcia12@llnl.gov.

TABLE 1. *Yersinia* spp. strains and IS*100* genotype classification

Strain(s)	Geographical origin[a]	Biovar[b]	IS*100* type
CO92,[c] Shasta, NM77-538, Yreka, South Park, 763546NM(752), Sci Alexander, A1122, 1171, A12 and A12(D9), Dodson	United States	orientalis	O1
JAVA (D88), Java2, TS (D5), MP6 (D77), M23(D73), 769, 770, Java9	Indonesia	orientalis	O1
Salazar (D14), La Paz and replicate	Bolivia	orientalis	O1
P Exu, Exu #2 (581), Exu #3 (582), Exu #9 (586), Exu #13 (590)	Brazil	orientalis	O1
EV76H (D59), EVC1962, M111 (74), H3	Madagascar	orientalis	O1
195P, I-254 (70), I-18 (71), YP1-RR5 (Yeo186)	India	orientalis	O1
Stavropol, SEV, 471, Russian Vaccine	FSU	orientalis	O1
390	Israel	orientalis	O1
31, 33, P1178VN (219)	Vietnam	orientalis	O2a
PMB9, RFPM19 (350), PBM23 (Yeo194)	Burma	orientalis	O2a
703	Cambodia	orientalis	O2a
316	Unknown	Atypical[d]	O2a
27, 32	Vietnam	orientalis	O2b
25 and five replicates	Vietnam	orientalis	O2c
KIM,[e] PRK159, PKH-10, PRK108	Iran/Kurdistan	medievalis	M1a
366	Yemen	medievalis	M1b
Harbin 35, Nicholisk 41	Manchuria	medievalis	M2[f]
Nicholisk 51 and five replicates	Manchuria	antiqua	O1
Yokohama	Japan	antiqua	A1a[f]
Kuma (D94)	Manchuria	antiqua	A1b
Nairobi and five replicates	Kenia	antiqua	A2
A16 and Antiqua	DROC	antiqua	A3
Angola	Angola	antiqua	A4
Pestoides A, B, C, D, Aa, Ba	FSU	medievalis	P1
Pestoides J	FSU	Atypical[d]	M1a
Pestoides E and F	FSU	antiqua	P2
Y. pseudotuberculosis PB1 (B16)	USA	I	S1
Y. pseudotuberculosis IP33953	France	I	S2
Y. pseudotuberculosis YPIII	Unknown	III	NA[g]
Y. enterocolitica WA (C22)	United States	O:8	NA[g]

[a] FSU, former Soviet Union; DROC, Democratic Republic of the Congo.
[b] See text for *Y. pestis* biovar definition; serovar is shown for other *Yersinia* spp.
[c] Sixteen variants of CO92 were analyzed with identical results.
[d] Both glycerol and nitrate negative.
[e] Nine variants of Kim were analyzed with identical results.
[f] IS*100* types A1a and M2 are identical.
[g] NA, no IS*100*-based amplification.

observation indicated that insertion of IS*1541* at given loci constitutes a stable event and suggested that the intracellular mobility of the IS elements in the *Y. pestis* genome is limited (21). These assumptions, if correct, legitimize genotyping by determining divergence of the exact positions of IS elements. In fact, IS element positional fingerprinting reveals itself as a technique with just the right discriminatory power, since in contrast with techniques such as RFLP, pulsed-field gel electrophoresis, or ribotyping (which can detect more frequent random variations of single nucleotides that lead to changes in restriction endonuclease recognition sites), it detects exclusively genetic rearrangements. Furthermore, the loss or change of an IS fingerprinting pattern automatically maps where on the genome the rearrangement event has taken place.

In this study, we describe the results of PCR-based genotyping of a large *Y. pestis* collection, comparing the variance of the location of the IS*100* element in the genome of these strains with that of the reference isolate, strain CO92. In addition, we have identified and employed for genotyping purposes the defective genes of the glycerol pathway that are likely to account for the glycerol-negative phenotype of the biovar orientalis strains. Furthermore, sequence analyses of these genes suggested that the glycerol-positive phenotype of an orientalis-like *Y. pestis* strain may have been the result of genetic exchange between glycerol-negative and -positive yersiniae strains. If true, this fact might indicate that naturally occurring recombinational events across strains may modulate a mechanism whereby an entire pathway can be inactivated and reactivated in response to environmental pressure.

MATERIALS AND METHODS

Bacterial strains. *Yersinia* spp. strains used for these studies were obtained from the culture collections of the U.S. Army Medical Research Institute of Infectious Diseases and Michigan State University and are listed in Table 1. Bacterial cultures were started from frozen stocks and DNA preparations were done as described earlier (2). The set of derivative variants of strain CO92 that underwent a number of laboratory manipulations, such as multiple passages in vitro or in vivo, had different plasmid composition, deletion of pigmentation locus, etc., also has been described previously (2).

Cloning of IS*100* flanking regions and primer design. A genomic library of *Y. pestis* CO92 (generous gift from G. Andersen) constructed by cloning DNA fragments of 3 to 6 kb into the pGEM vector (Promega, Madison, Wis.) was used to retrieve IS*100* elements and associated regions. The library was screened using oligonucleotide probes representing the IS*100* sequence bp 197 to 216 and bp 1754 to 1773 (accession number U59875). The nucleotide sequence of positive clones was determined and the location of putative neighboring open reading frames (ORFs) was ascertained by BLAST analysis (www.ncbi.nlm.nih.gov/BLAST/). Two primers corresponding to sequences located at the right and left ends of IS*100* were selected: ISfor1754 (5′-GGTGATGCAGCACTGACCTC) and ISrev216 (5′-GCTCAGATTTTGCCTGCAAA). The design of locus-specific primers corresponding to the flanking regions was based on an identified nucleotide sequence that could pair with one of the IS*100*-specific primers. A total of 27 individual primers were employed for the final analysis of DNA of

TABLE 2. Locus-specific primers for PCR-IS*100*-based fingerprinting[a]

Primer name	Primer sequence	IS*100* pair	PCR fragment (bp)[b]	Primer location within ORF of CO92 genome	ORF status due to IS*100* insertion
vlm01Arev	TGGAAGCCAACTGTACTTTCACT	ISfor1754	491	*artJ*, arginine-binding periplasmic protein 2	Disrupted
vlm02Bfor	**ATAATGACCAACTGCGACTCAAC**	**ISrev216**	**504**	**IS*1541*-IS*100* intergenic region**	**N/A[c]**
vlm03rev	**ACCCGTTAAATATGCCGTTG**	**ISfor1754**	**486**	**Homology to *ycfS, E. coli* hypothetical 34.6-kDa protein in *ndh-mfd* intergenic region**	**Disrupted**
vlm04for	CTGCACCCGTTTCAGTCTTT	ISrev216	503	*gst*, glutathione *S*-transferase (EC 2.5.1.18)	Intact
vlm05rev	CCAAGTTAACCCGCGTAGAG	ISfor1754	385	Homology to *yjbD, E. coli* hypothetical protein 10.5 kDa in *pepE-lysC* intergenic region	Intact
vlm06for	**CACGCCCAATACGTTATTCA**	**ISrev216**	**307**	**Homology to 5′ end of *yurG*, putative aspartate aminotransferase, *Bacillus subtilis* (Z99120)**	**Disrupted**
vlm07Arev	**GAACGGCTGAATCATCATACCG**	**ISfor1754**	**317**	**Homology to 3′ end of *yurG*, putative aspartate aminotransferase, *B. subtilis* (Z99120)**	**Disrupted**
vlm08for	AAAGCCTTGCATGGTGATTC	ISrev216	930	Homology to *yajF, E. coli* hypothetical protein 32.5 kDa in *aroM-araJ* intergenic region	Intact
vlm09rev	ATGAAGGCTTCGGTACGTTG	ISfor1754	368	Homology to *sbcC, E. coli* ATP-dependent dsDNA exonuclease	Disrupted
vlm12for	**TATCGTCCGCTATTGCCTGT**	**ISrev216**	**393**	**Putative OFR with no significant homology in GenBank**	**Disrupted**
vlm14for	CCACCATAATGACCACCAATAGT	ISrev216	671	IS*100-'yp48* intergenic region, the 102-kb *pgm* locus of *Y. pestis* (AL031866)	N/A
vlm20for	**GCAAACCCATTGATGTATCTGAG**	**ISrev216**	**300**	***pheA*, prephenate dehydratase (EC 4.2.1.51)**	**Disrupted**
vlm21rev	**TCTTGGTAGAGAGCCGTTCATTA**	**ISfor1754**	**587**	***tyrA*, chorismate mutase-T (EC 5.4.99.5)**	**Intact**
vlm22for	TTTTCATCATTGGTTGCTTTACC	ISrev216	356	Homology to *ybaN, E. coli* hypothetical protein 14.8 kDa in *priC-apt* intergenic region	Disrupted
vlm23rev	TAAAAACCTTGATAGCCGAATCA	ISfor1754	436	*aprT*, adenine phosphoribosyl transferase (EC 2.4.2.7)	Intact
vlm25rev	**CGCCGTGGTACCTACTTTATTTT**	**ISfor1754**	**384**	***mutT*, mutator MutT protein**	**Intact**
vlm26for	TCAATGGAAAATACCACATAGCC	ISrev216	583	Low homology to *ytaP*, hypothetical protein *B. subtilis* (Z99119); dienelactone hydrolase family motif	Disrupted
vlm27rev	CTACCGTCTTGGATATTTGAACG	ISfor1754	380	Homology to *yrdA, E. coli* hypothetical protein, 20.4 kDa, in *rrnD-aroE* intergenic region	Disrupted
vlm28for	**GCTCTGCAGTTCATTGAAGTTTT**	**ISrev216**	**256**	**Homology to *ycjZ, E. coli* hypothetical transcriptional regulator in *tpx-fnr* intergenic region, LysR family**	**Disrupted (promoter)**
vlm29rev	**GAACAGTTTCAGGGAACGCTAT**	**ISfor1754**	**1,325**	**Homology to *ynbD, E. coli* hypothetical protein 49.6 kDa in *maoC-acpD* intergenic region**	**Intact**
vlm30for	TATCCCATGGCAGCTAAAAATAA	ISrev216	334	Homology to methylation-accepting chemotaxis proteins	Disrupted
vlm31rev	GAGGTGCGCACTATTGAGATG	ISfor1754	1,107	Homology to *ybiN, E. coli* hypothetical protein 34.2 kDa in *dinG-gluQ* intergenic region	Intact
vlm32for	**ACATCATAACCCGCACACTTAAT**	**ISrev216**	**383**	**Homology to probable polysaccharide deacetylase *Schizosaccharomyces pombe* (Z97209)**	**Disrupted**
vlm33rev	**GAGCTTTTCAGCTTCTTCATCAG**	**ISfor1754**	**399**	**Putative ORF containing prophage integrase domain**	**Disrupted[d]**
vlm35rev	TTTTCAAACAACCTTGCCAAC	ISfor1754	978	Low homology to *irp2*, high molecular weight protein 2, *Yersinia* spp.	Could be disrupted[e]
vlm36for	**ATTGATTGCTGACGAGTTAATCC**	**ISrev216**	**1,169**	**Low homology to *yidJ, E. coli* hypothetical protein 57.3 kDa in *emrD-glvG* intergenic region**	**Intact[f]**
vlm37rev	**GCAACTGGTGTTTACCAGCTACT**	**ISfor1754**	**543**	**Homology to chondroitin sulfate lyase family**	**Intact[f]**

[a] Alternating IS*100* loci are displayed in boldface. In some cases only one primer exists for a given locus.
[b] Size of PCR fragment amplified from *Y. pestis* CO92 genome.
[c] N/A, not applicable due to selection of primer within intergenic region.
[d] This ORF is not functional due to another IS element inserted downstream of the vlm033rev primer.
[e] ORF is overlapped with IS*100* sequence.
[f] IS*100* is inserted within the IS*1541* transposase.

yersiniae strains (Table 2). These primers represented 16 unique chromosomal loci. In most cases, primers originated from both sides of the adjacent regions of independent IS*100* copies. However, primers derived from one side were utilized in a few cases. We purposely tried to place a locus-specific primer within the structural part of the putative gene rather than within an intergenic area. Therefore, the online unfinished genome of *Y. pestis* CO92 available at the Sanger Centre site (www.sanger.ac.uk/Projects/Y_pestis/) was widely used for the ORF border determination and for the locus selection itself.

PCR amplification. PCR mixtures contained 50 pmol of one of the IS*100*-specific primers (ISfor1754 or ISrev216), 5 pmol of the locus-specific primer (vlm series), and 1 to 3 ng of template per 50 μl of total reaction mixture volume. The reactions were run in duplicate using two different polymerases, such as AmpliTaq and AmpliTaqGold (PE Applied Biosystems, Foster City, Calif.).

The primer pairs designed to detect defective genes of the glycerol pathway included glpK-F2 (5′-ACTCATGGCCTGTTGACCACCATT), glpX-R4 (5′-GGGCATGTAAGAATGTGCCTTGGT), glpD-F1 (5′-GGCTAGCCGCCTCAACAAAAACAT), and glpD-R1 (5′-GGTCATACAAGAACAAGCCGGTGC).

Numerical analysis. PCR results were visualized by using agarose gel, and each strain was given a score of 1 (present) or 0 (absent) for each possible band position. Similarities in band patterns were analyzed by the unweighted pair group method using average linkages (UPGMA) clustering software package PAUP 4.0 (30). To validate the robustness of this analysis, the neighbor-joining

method of Saitou and Nei (26) (which uses a different set of underlying assumptions of unequal rate of character evolution) was also used. The topologies of the trees obtained by both methods were nearly identical (aside from small shifts in internal nodes).

Nucleotide sequence determination. Sequencing of cloned and PCR fragments was undertaken as described previously (17).

Comparative genomics. The comparison of the two unfinished genomes of *Y. pestis* strains CO92 and KIM was done in two steps. First, potential coding regions of the CO92 isolate were determined using the ORF-finding program CRITICA (4); identified ORFs were then subjected to homology search against the KIM database (www.genome.wisc.edu/html/pestis.html) using BLAST analysis. The ORFs that displayed identities of less than 90% were selected and further investigated.

Genes of *Y. pestis* participating in the fermentation of glycerol were identified using the collection of metabolic pathways from WIT2 (wit.mcs.anl.gov/WIT2/) and KEGG (www.genome.ad.jp/kegg/metabolism.html) databases. To find potential defects in the glycerol pathway of CO92, each of the ORFs identified from the KIM (glycerol-positive) and CO92 (glycerol-negative) strains were compared to each other and analyzed for sequence differences.

Nucleotide sequence accession numbers. Newly determined *glpD* allele sequences were deposited in GenBank under accession numbers AF377928 (strain Nicholisk 51), AF377936 (strain Nicholisk 41), AF377929 (strain Pestoides C), AF377930 (strain Pestoides F), AF377937 (strain P EXU), AF377931 (strain Yokohama), AF377938 (strain 32), AF377939 (strain 390), AF377940 (strain A12), AF377932 (strain A16), AF377941 [strain Salazar (D14)], AF377942 [strain EV76H (D59)], AF377943 [strain TS (D5)], AF377933 [strain Kuma (D94)], AF377935 (strain Harbin 35), AF377934 (strain Nairobi).

RESULTS

Molecular characterization of IS*100* insertions. A total of 185 IS*100*-positive clones were selected from the screening of the genomic library of *Y. pestis* strain CO92. Sequencing of these clones revealed the presence of 60 independent contigs representing copies of the IS*100* element and their unique flanking regions. Variants were obtained that contained entire IS*100* copies together with their up- or downstream regions and partial copies possessing unique sequence on one side only. The putative ORFs located within the adjacent area of each IS*100* copy were identified, and internal primers corresponding to the sequence of such ORFs were made. PCRs on *Y. pestis* CO92 templates were performed using the set of constructed primers and either the forward or reverse primer corresponding to the IS*100* sequence (see Materials and Methods). Those 27 primers that showed a strong specific signal of amplification were chosen for further analysis (Table 2). These primers represented 16 distinct IS*100* insertions, with 11 of them having both sides amplified. Analysis of the sequences surrounding the chosen IS*100* elements showed that most of the insertions led to the disruption of putative ORFs. We found no indication of specificity for IS*100* integration into genomic sites.

IS*100* genotyping. A total 116 bacterial DNAs were employed for the analyses: 77 represented unique *Yersinia* isolates, and the rest were variants of one of these 77 strains (Table 1). The results of the fingerprinting are shown in Fig. 1. Out of 27 locus-specific primers tested, 4 amplified the same PCR fragment in all *Y. pestis* strains studied (vlm12for, vlm29rev, vlm33rev, and vlm37rev). In these cases (except vlm12for), both sides of the locus were employed for the primer selection (however, the second portion of the locus failed to amplify in at least one strain). For example, the vlm28for primer (same locus as vlm29rev) did not amplify strains of the O2b group, the vlm32for primer (same locus as vlm33rev) failed on the P2 group, and the vlm36for primer (same locus as vlm37rev) was negative on the O2b and A5 groups. Such variations, where only one and not a second side of the same IS*100* insertion locus could be amplified, were further seen on vlm26for/vlm27rev and vlm30for/vlm31rev loci, suggesting possible rearrangements of the regions adjacent to the IS*100* element.

Geographically distinct members of the orientalis biovar showed an identical fingerprint pattern (whether derived from South or North America, India, the former USSR, or Indonesia), with the exception of strains isolated from Indochina (Vietnam, Burma, Cambodia), which lacked at least one characteristic IS*100* copy at the vlm04for/vlm05rev locus. Indeed, PCR on DNA of these strains that directly employed the flanking vlm04for/vlm05rev pair of primers generated a fragment that was about 2 kb shorter (the size of IS*100*) than that seen for other orientalis strains (data not shown). We designated these strains as IS*100* type O2 and introduced additional subtyping based on the random variability of other fragments (O2a, O2b, and O2c). Thus, our criteria for assigning a new genotype name within a biovar was a difference between strains in at least one IS*100* copy position that could be detected by the absence of the amplification on both sides of IS*100* when using the locus-specific primers. Different subtypes between strains were assigned a, b, or c notation (e.g., M1a, M1b) if they shared amplification of a locus on at least one side of IS*100* and if the second side of that insert displayed either a shifted fragment or failed to amplify.

Fingerprints of biovar medievalis strains differed from those of the reference orientalis biovar strain CO92 by the absence of at least 10 bands, as well as by a shifting in the size of the vlm03rev- and vlm21rev-derived fragments. All strains originating from Iran and Kurdistan showed an identical IS*100* type, M1a, which differed by two band positions from the strain of the M1b type isolated in Yemen. Interestingly, both medievalis strains of the M2 type from Manchuria resembled antiqua strains of the A1 type (identical to A1a and differing from the A1b type only in one position), which was also isolated from the same geographical areas (Japan and Manchuria, respectively). The latter strains differed significantly from other antiqua isolates such as strains found in Africa and former USSR. We analyzed a few representatives of a heterogeneous group of *Y. pestis* strains isolated in the former USSR, known as Pestoides variants. These strains (except one) formed a separate cluster on the dendrogram that included also an antiqua Angola strain (Fig. 1). The exceptional strain, Pestoides J, possessed a medievalis-specific IS*100* type M1a.

Both *Y. pseudotuberculosis* strains of serovar I displayed at least two positive amplification bands corresponding to locus-specific primers vlm26for and vlm14for, suggesting that these IS*100* copies are present at the same locus as in the *Y. pestis* strains. Not surprisingly, one such IS*100* copy was positioned near the high-pathogenicity island of the 102-kb unstable pigmentation region. Probably, *Y. pseudotuberculosis* strain PB1 has an additional *Y. pestis*-specific insertion within the vlm20for/vlm21rev-associated locus.

A most intriguing finding of our IS*100*-based fingerprinting study was the fact that one of the antiqua strains (Nicholisk 51 from Manchuria) displayed an identical IS*100* pattern to that found in strains corresponding to the orientalis IS*100* type O1 (Table 1). This observation was validated by thorough analyses

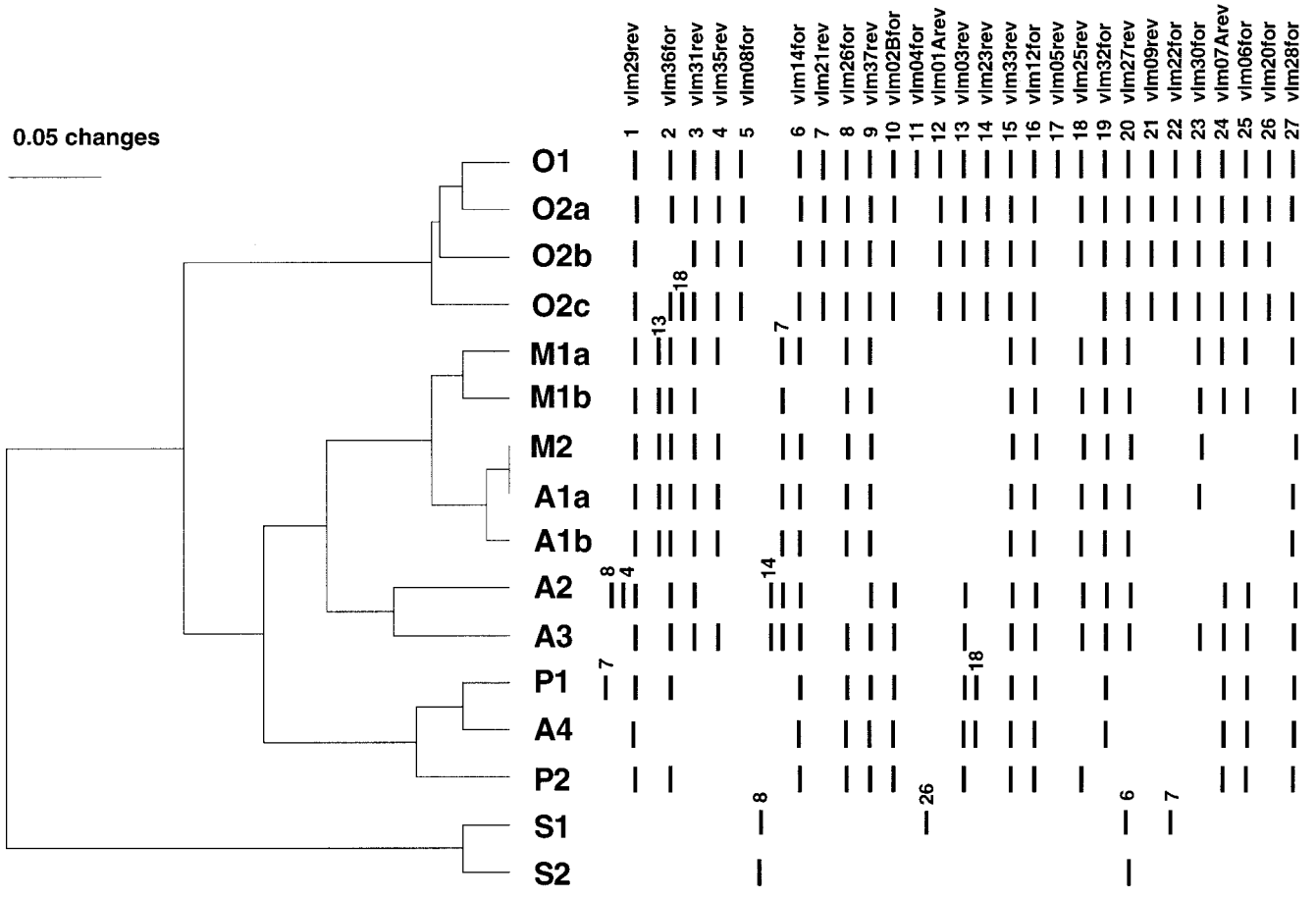

FIG. 1. Digitized IS*100*-based fingerprints and dendrogram constructed by the UPGMA clustering method. Similarity analysis was performed using the Dice coefficient. The scale bar indicates dissimilarity proportions. The numbers 1 to 27 were assigned to each locus-specific primer vlm according to the size of the PCR fragment formed, using *Y. pestis* strain CO92 DNA as a template. The positions of the shifted fragments were followed in other strains. The IS*100* genotypes of different groups shown are as follows: O1, O2a, O2b, and O2c are biovar orientalis; M1a, M1b, and M2 are biovar medievalis; A1a, A1b, A2, A3, and A4 are biovar antiqua; P1, and P2 are strains of the Pestoides group; and S1 and S2 are strains of *Y. pseudotuberculosis*.

of individual colonies of this strain with respect to biochemical properties and by reconfirmation of their IS*100* genotype.

Glycerol pathway defects in *Y. pestis*: the *glpK-glpX* locus. We needed to first resolve the nature of the metabolic lesion in the glycerol pathway of biovar orientalis strains to facilitate identifying the mutational event accounting for the glycerol-positive phenotype of the Nicholisk 51 strain. To do so, we compared genes involved in glycerol utilization in both the unfinished *Y. pestis* strains CO92 and KIM genome projects (see Materials and Methods) and identified two regions within the strain CO92 chromosome that could account for its glycerol-negative phenotype. The first region was a cluster of the three genes, two of which were affected by a 956-bp deletion extending within both *glpK* and *glpX* (Fig. 2A). The deletion resulted in the omission of 142 amino acids from the C-terminal portion of the glycerol kinase as well as 88 amino acids from the N-terminal portion of the GlpX protein. Interestingly, this 956-bp deletion could be a result of reciprocal recombination between imperfect direct repeats of 12 bp present in both the *glpK* and *glpX* genes of the KIM strain. We prepared the PCR primer pair glpK-F2/glpX-R4 flanking the 956-bp deletion and analyzed our entire culture collection to detect the distribution of this defect in glycerol fermentation. All glycerol-positive strains (including Nicholisk 51) lacked the deletion as evidenced by production of a fragment of 1,258 bp corresponding to the size of that in strain KIM. As expected, glycerol-negative strain CO92 gave rise to a significantly shorter fragment of 302 bp. Surprisingly, most of the glycerol-negative strains of biovar orientalis showed an intact *glpK-glpX* region, and only strains originating in the United States (with two exceptions) possessed a deletion identical to that found in the CO92 strain (PCR results were confirmed with sequencing data). The two U.S.-originated strains that did not have the described 956-bp deletion were A1122 and its variant strain A12. Thus, the 956-bp deletion found within the *glpK-glpX* region in the CO92 strain most likely is not responsible for the glycerol-negative phenotype of biovar orientalis because most of the strains of this biovar do not possess this deletion.

Glycerol pathway defects in *Y. pestis*: the *glpD* gene. The second defect in the glycerol pathway found in strain CO92 was the loss of 93 bp that resulted in a 31-amino acid in-frame deletion within *glpD*, encoding aerobic glycerol-3-phosphate dehydrogenase (Fig. 2B). In a manner analogous to that observed for the *glpK-glpD* deletion described above, we located

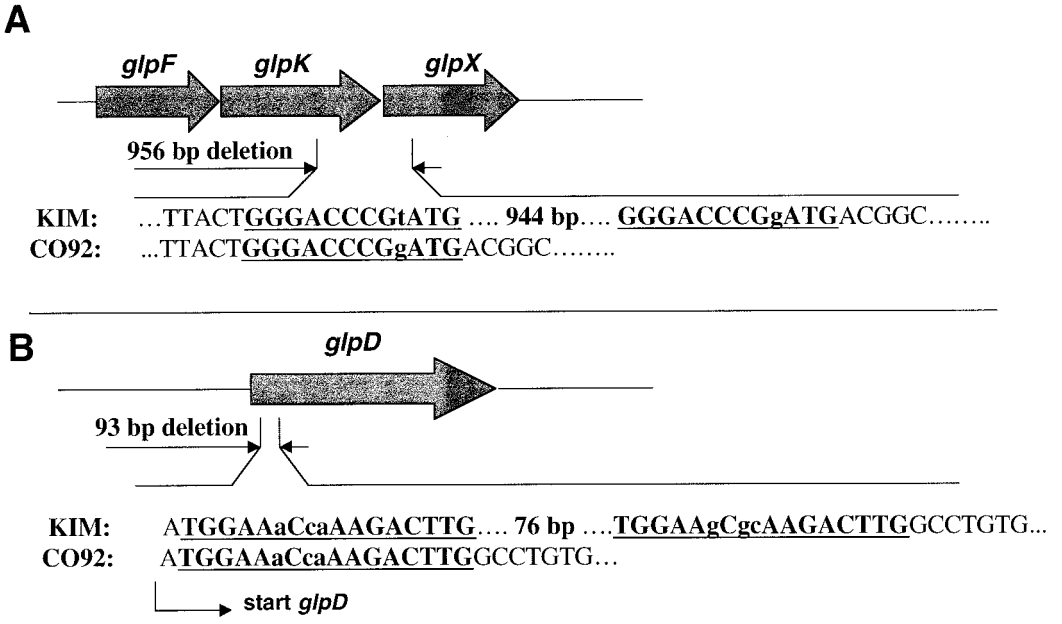

FIG. 2. Genes of glycerol metabolism that are defective in *Y. pestis* strain CO92 in comparison with strain KIM. The imperfect repeats that could account for a deletion are bold and underlined; lowercase letters represent mismatches within a repeat. (A) Cluster of genes, including the following: *glpF*, glycerol uptake facilitator; *glpK*, glycerol kinase; *glpX*, unknown function in glycerol metabolism. (B) *glpD* gene for glycerol-3-phosphate dehydrogenase.

within the *glpD* gene two 17-bp imperfect direct repeats flanking the region that was lost in strain CO92. We then tested our culture collection for the presence of the defect within *glpD* using a deletion-flanking pair of primers, glpD-F1 and glpD-R1. The results of PCR analysis showed that all glycerol-positive strains (including Nicholisk 51) had an intact *glpD* gene (fragment size of 508 bp, identical to that of strain KIM), and all glycerol-negative strains had a defective *glpD* gene (fragment size of 415 bp, identical to strain CO92). This 100% correlation between the glycerol-negative phenotype and the presence of a defective *glpD* indicates that most likely the inability of biovar orientalis strains to ferment glycerol is due to inactivation of this gene. Comparison of nucleotide sequences of the entire *glpD* gene of strains CO92 and KIM revealed the presence of two nucleotide differences (causing amino acid substitutions) (Fig. 3) in addition to the 93-bp deletion described above. Sequence analysis of the *glpD* gene of the glycerol-positive Nicholisk 51 strain showed that it represented a hybrid variant possessing both the CO92-specific point substitutions and the intact KIM-specific region that is deleted in orientalis strains.

DISCUSSION

To characterize a large collection of *Y. pestis* strains of different biovars and geographical origins, we have used a technique that enables the determination of the position of the IS element IS*100* present in the *Y. pestis* genome (IS-positional genotyping). The method employed a fragment polymorphism originated from the amplification of the region located between one or the other end of the IS*100* element and the neighboring ORF. To design primers for these experiments, the IS*100* sequences together with the surrounding regions were cloned from the genome of *Y. pestis* strain CO92 of biovar orientalis. Therefore, this isolate became a reference strain and subsequent genotyping of other isolates was based on comparison with the CO92 fingerprinting pattern. A total of 16 distinct IS*100* insertions were employed for the analysis, with 11 of them having both sides amplified and the remainder using neighboring ORFs positioned only up- or downstream. In contrast to the IS*1541* that preferably integrates within a T/A-rich short segment (21), analysis of the sequences located in the vicinity of the IS*100* insertion site did not reveal any integration specificity. In addition, most of the IS*100* insertions led to the disruption of putative ORFs (Table 2). The low specificity of the IS*100* integration observed by us is in good correlation with previously published data (10). The IS*100* typing was done on a total of 116 yersiniae DNA samples using 27 primer pairs. Comparison of the PCR fragments formed using DNA of the different strains with those originating from CO92 DNA showed that the particular fragments are either present, shifted, or missing. Curiously, we noted in many cases that some strains amplified only one side of the same IS*100* insertion, suggesting possible rearrangements of the regions adjacent to the IS*100* element. However, multiple laboratory passages in vitro or in vivo did not change the IS*100* fingerprint profile in any analyzed strain derivatives. In general, we observed a correlation between the IS*100* genotype and the biovar of the isolates except with Manchurian strains and Pestoides isolates (see below).

The method described here did not have enough discriminatory power to differentiate orientalis strains of different geographical origins, except for those originating in Indochina (Vietnam, Cambodia, and Burma). The latter did not possess an IS*100* copy located between the primers vlm04for and vlm05rev, and these flanking primers, in contrast to other ori-

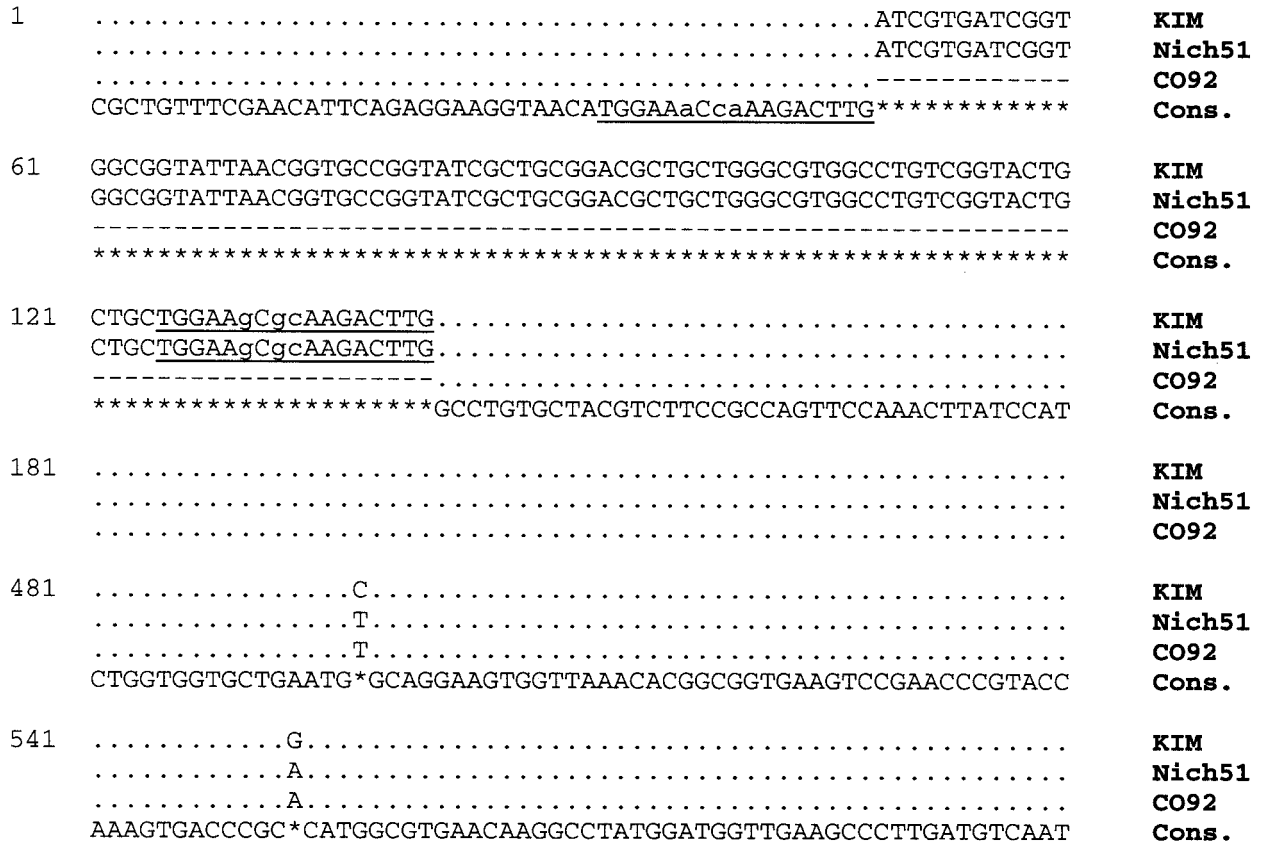

FIG. 3. Alignment of the partial sequences of the *glpD* gene of *Y. pestis* strains KIM, Nicholisk 51, and CO92. The consensus sequence is shown, and the asterisks represent differences. The imperfect repeats are designated in the same way as for Fig. 2.

entalis strains, generated a fragment identical to the one amplified from DNA of *Y. pestis* strains of biovars antiqua and medievalis as well as tested isolates of *Y. pseudotuberculosis*. This fact indicates that the strains from Indochina either lost this particular IS*100* copy due to some recombination event (i.e., precise excision of the IS*100* element, crossing over, etc.) or never had it. The latter possibility would imply that invasion of the Indochina region by plague bacilli during the last pandemic was a separate event unrelated to the distribution of a distinct clone of the orientalis biovar over the world. To improve the discriminatory power of this technique so as to enable further differentiation among orientalis strains, it may be necessary to include additional copies of the IS*100* (or other IS elements) into the PCR-based genotyping scheme. Release of the *Y. pestis* CO92 genome reveals the potential use of more than a hundred total copies of IS*100*, IS*285*, and IS*1541*.

Fingerprinting patterns of the biovar medievalis strains differed from the orientalis isolates by more than 10 missing or shifted bands. We had fewer representatives of the former in our collection; nevertheless, it was possible to group the medievalis strains according to their geographical origin. Interestingly, clusters formed on the dendrogram by the biovars medievalis and antiqua overlapped within the Manchurian and Japanese isolates (M2 and A1 genotypes), and the profile of the Yokohama strain (A1a) was identical to that of the Harbin 35 and Nicholisk 41 isolates (M2). This fact indicates a close association between the strains of the different biovars isolated in this region. The use of a larger quantity of strains of both biovars combined with analysis of a greater number of genetic loci should clarify our observation of the relationship of medievalis strains with one of the lineage of the antiqua strains found in Southeast Asia. The remaining strains of the antiqua biovar in our collection originated in Africa and the former USSR, and their IS*100* genotypes were significantly different from the antiqua isolates described above.

Atypical plague isolates can be found in the ancient reservoirs in the territory of the former USSR as well as typical epidemic forms of *Y. pestis*. This heterogeneous atypical group received the common name Pestoides (19). Strains isolated in the Transcaucasian Alpine, Daghestan Mountain, Mountain-Altai, Hissar, and Talas plague foci as well as in some regions of Mongolia display unusual abilities to ferment rhamnose and melibiose and have a reduced virulence for guinea pigs. The Caucasian isolates fail to produce pesticin and plasminogen activator and the Altai strains are pesticin-sensitive irrespective of the presence of the pPCP plasmid. There are some unusual auxotrophic features of these atypical strains, which parasitize unique species of rodents (3). Finally, in many cases the Pestoides strains reported from these foci possess unique plasmid profiles (12). Although we do not know the exact geographical origins of the Pestoides strains present in our collection, IS*100*-based fingerprinting divided our isolates into three genotypes. The strain Pestoides J (both glycerol and nitrate negative) showed a fingerprint pattern identical to that

of biovar medievalis strains such as KIM. In contrast, other Pestoides isolates formed a separate cluster on the dendrogram. The Pestoides strains (P1 genotype) of biovar medievalis resembled the antigua isolate Angola. Such connection of Pestoides with the African strain did not surprise us, due to the fact that a possible relationship of strains from Central Asia and Africa has been suggested (7, 15, 34). The strains Pestoides E and F belonged to the IS*100*-genotype P2, and they most probably represented the Caucasian isolates (Transcaucasian Alpine and Daghestan Mountain foci). The characteristic feature of these isolates, as confirmed in our study, is the absence of the pPCP plasmid encoding plasminogen activator, pesticin, and its immunity protein (12, 13). These unusual strains differ from the typical epidemic forms of plague in that pPCP is not required for tissue invasion (27, 31, 33). Recently, based on RFLP derived from the hybridization of Southern blots with the probes for IS*100* and IS*285* elements, Bobrov and Filippov (5) concluded that the Caucasian isolates are the most phylogenetically ancient of *Y. pestis* strains. Our data are in agreement with this assumption.

Our observation that one of the glycerol-positive antiqua strains (Nicholisk 51) had an IS*100* genotype identical to that of glycerol-negative biovar orientalis isolates prompted a closer investigation of this relationship. Using a comparison of unfinished genomes of two *Y. pestis* strains followed by screening the entire strain collection, we found an IS*100*-independent marker region (archival phage remains) that was present exclusively in orientalis strains as well as in the Nicholisk 51 strain (data not shown). The positive glycerol fermentation phenotype of Nicholisk 51 places it squarely within the antiqua biovar, yet its IS*100* genotype and the presence of specific phage remains categorizes this isolate as a member of the orientalis biovar. Since the orientalis biovar may have arisen from a biovar antiqua clone (7, 15), strain Nicholisk 51 could represent either an ancestor of biovar orientalis or a variant of this biovar that had undergone phenotypic reversion back to the glycerol positive phenotype. Our analysis of the glycerol pathway in *Y. pestis* showed that mutations in two loci could potentially account for the glycerol-negative phenotype of biovar orientalis isolates. The first defect was a 956-bp deletion affecting both *glpK* and *glpX*. However, not all glycerol-negative strains possessed this mutation. In fact, most of the strains of biovar orientalis tested did not have this deletion. Only strains isolated in the United States (the CO92 reference isolate) possessed this deletion. The exceptions were strains A1122 obtained in California in 1943 (32) and A12, a derivative of the former strain. Both of the two other California strains in our collection (Yreka and Shasta) had the deletion. Thus, the 956-bp deletion within the *glpK-glpX* region could not, by itself, account for the glycerol-negative phenotype because it does not exist in most biovar orientalis strains. Obviously, the glycerol-positive Nicholisk 51 strain did not possess this deletion.

The second defect found in the glycerol pathway of biovar orientalis strains was a 93-bp in-frame deletion in the *glpD* gene encoding aerobic glycerol-3-phosphate dehydrogenase. Analysis of our culture collection showed that all glycerol-negative strains contained this deletion, whereas all glycerol-positive strains possessed intact *glpD*. This observation suggested that inactivation of GlpD accounts for the inability of biovar orientalis strains to ferment glycerol. This finding is in agreement with previously published data (8). In addition to this 93-bp deletion, the *glpD* genes of strains CO92 and KIM exhibited two nucleotide differences. We performed additional sequencing of *glpD* from *Y. pestis* strains of different geographical origins, including seven isolates of biovar orientalis, six of antiqua, and three of medievalis. The sequences of all orientalis strains were identical to that of strain CO92 and the remainder were identical to that of strain KIM. Accordingly, we envision that the most likely scenario for the generation of the *glpD* sequence found in the glycerol-negative orientalis biovar was an initial deletion event followed by the accumulation of two point mutations within the already inactivated gene. Of course, there is a possibility for the reverse sequence of events in which two point mutations occurred prior to the deletion. However, the latter possibility would put *glpD* somewhat at odds with the observed conservatism of sequence found during the extensive analyses conducted on housekeeping genes in *Y. pestis* (1). Interestingly, the *glpD* sequence of Nicholisk 51 strain represents a hybrid variant possessing both strain CO92-specific base substitutions and an intact KIM-specific region that was deleted in the orientalis strains. The signature of orientalis-specific point substitutions within *glpD* of the Nicholisk 51 strain suggests that this strain originated from a typical orientalis isolate as result of repairing the originally defective GlpD. The exact mechanism accounting for this phenotypic reversion is not clear. A reciprocal recombination between flanking observed imperfect direct repeats of 12 and 17 bp may account for the deletions within *glpK-glpX* and *glpD*, respectively. We were not able to identify by BLAST searches within either strain CO92 or KIM the sequence that resembled a second copy of *glpD* or that was deleted in biovar orientalis strains. This fact suggests that repair of the 93-bp deletion in *glpD* of strain Nicholisk 51 reflects genetic exchange between the Nicholisk 51 precursor and a strain capable of fermenting glycerol. Obviously, the range of the area that underwent recombination was short rather than large, because the orientalis-specific *glpD* point mutations present in Nicholisk 51 are located close to the deletion site. The overall conclusion of the glycerol fermentation scenario of strain Nicholisk 51 is that this isolate provides solid, but indirect, evidence of genetic exchange between *Y. pestis* strains facilitating repair of a biochemical pathway. The existence of the potential for such phenotypic reversion among *Y. pestis* isolates that have lost discrete physiological functions (6) may point to a possible mechanism by which a more virulent epidemic version of the plague bacillus could arise from time to time.

ACKNOWLEDGMENTS

We thank Andrew A. Filippov for helpful discussions and Gary Andersen for providing a genomic library of *Y. pestis* CO92.

This work was performed under the auspices of the U.S. Department of Energy by the University of California Lawrence Livermore National Laboratory under Contract no. W-7405-Eng-48. Work conducted at the U.S. Army Medical Research Institute of Infectious Diseases was supported through funding provided by the U.S. Army Medical Research and Materiel Command.

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. **Aparin, G. P., and E. P. Golubinsky.** 1989. Microbiology of plague. University of Irkutsk, Irkutsk, Russia.
4. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. E vol. **16:**512–524.
5. **Bobrov, A. G., and A. A. Filippov.** 1997. Prevalence of IS*285* and IS*100* in *Yersinia pestis* and *Yersinia pseudotuberculosis* genomes. Mol. Gen. Mikrobiol. Virusol. **2:**36–40.
6. **Brubaker, R. R.** 1991. Factors promoting acute and chronic diseases caused by yersiniae. Clin. Microbiol. Rev. **4:**309–324.
7. **Devignat, R.** 1951. Varietes de l'espece *Pasteurella pestis*. Nouvelle hypothese. Bull. O. M. S. **4:**247–263.
8. **Domaradskii, I. V., L. V. Linnikova, and E. P. Golubinskii.** 1968. Investigation of glycerol metabolism in the plague microorganism. Vopr. Med. Khim. **14:**185–190.
9. **Filippov, A. A., P. N. Oleinikov, A. G. Bobrov, V. L. Motin, N. P. Konnov, and G. B. Smirnov.** 1994. Comparative study of the structure and distribution of two IS elements in *Yersinia pestis*. Genetika **30**(Suppl.)**:**166.
10. **Filippov, A. A., P. N. Oleinikov, A. V. Drozdov, and O. A. Protsenko.** 1990. The role of IS-elements of *Yersinia pestis* (Lehmann, Neumann) in the emergence of calcium-independent mutations. Genetika **26:**1740–1748.
11. **Filippov, A. A., P. V. Oleinikov, V. L. Motin, O. A. Protsenko, and G. B. Smirnov.** 1995. Sequencing of two *Yersinia pestis* IS elements, IS*285* and IS*100*. Contrib. Microbiol. Immunol. **13:**306–309.
12. **Filippov, A. A., N. S. Solodovnikov, L. M. Kookleva, and O. A. Protsenko.** 1990. Plasmid content in *Yersinia pestis* strains of different origin. FEMS Microbiol. Lett. **55:**45–48.
13. **Fursov, V. V., and Y. A. Popov.** 1983. Studying of *Yersinia pestis* "vole" strains by agarose gel electrophoresis, p. 326–327. *In* Y. G. Suchkov (ed.), Prophylaxis of natural focal infections. Region Administration Press, Stavropol, Russia.
14. **Gorshkov, O. V., E. P. Savostina, I. A. Popov, O. P. Plotnikov, N. A. Vinogradova, and N. S. Solodovnikov.** 2000. Genotyping *Yersinia pestis* strains from various natural foci. Mol. Gen. Mikrobiol. Virusol. **3:**12–17.
15. **Guiyoule, A., F. Grimont, I. Iteman, P. A. Grimont, M. Lefevre, and E. Carniel.** 1994. Plague pandemics investigated by ribotyping of *Yersinia pestis* strains. J. Clin. Microbiol. **32:**634–641.
16. **Guiyoule, A., B. Rasoamanana, C. Buchrieser, P. Michel, S. Chanteau, and E. Carniel.** 1997. Recent emergence of new variants of *Yersinia pestis* in Madagascar. J. Clin. Microbiol. **35:**2826–2833.
17. **Hu, P., J. Elliott, P. McCready, E. Skowronski, J. Garnes, A. Kobayashi, R. R. Brubaker, and E. Garcia.** 1998. Structural organization of virulence-associated plasmids of *Yersinia pestis*. J. Bacteriol. **180:**5192–5202.
18. **Lucier, T. S., and R. R. Brubaker.** 1992. Determination of genome size, macrorestriction pattern polymorphism, and nonpigmentation-specific deletion in *Yersinia pestis* by pulsed-field gel electrophoresis. J. Bacteriol. **174:** 2078–2086.
19. **Martinevsky, I. L.** 1969. Biology and genetic features of plague and plague-related microbes. Meditsina, Moscow.
20. **McDonough, K. A., and J. M. Hare.** 1997. Homology with a repeated *Yersinia pestis* DNA sequence IS*100* correlates with pesticin sensitivity in *Yersinia pseudotuberculosis*. J. Bacteriol. **179:**2081–2085.
21. **Odaert, M., A. Devalckenaere, P. Trieu-Cuot, and M. Simonet.** 1998. Molecular characterization of IS*1541* insertions in the genome of *Yersinia pestis*. J. Bacteriol. **180:**178–181.
22. **Perry, R. D., and J. D. Fetherston.** 1997. *Yersinia pestis*—etiologic agent of plague. Clin. Microbiol. Rev. **10:**35–66.
23. **Podladchikova, O. N., G. G. Dikhanov, A. V. Rakin, and J. Heesemann.** 1994. Nucleotide sequence and structural organization of *Yersinia pestis* insertion sequence IS*100*. FEMS Microbiol. Lett. **121:**269–274.
24. **Portnoy, D. A., and S. Falkow.** 1981. Virulence-associated plasmids from *Yersinia enterocolitica* and *Yersinia pestis*. J. Bacteriol. **148:**877–883.
25. **Rakin, A., and J. Heesemann.** 1995. The established *Yersinia pestis* biovars are characterized by typical patterns of I-CeuI restriction fragment length polymorphism. Mol. Gen. Mikrobiol. Virusol. **26:**26–29.
26. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. E vol. **4:**406–425.
27. **Samoilova, S. V., L. V. Samoilova, I. N. Yezhov, I. G. Drozdov, and A. P. Anisimov.** 1996. Virulence of $pPst^+$ and $pPst^-$ strains of *Yersinia pestis* for guinea pigs. J. Med. Microbiol. **45:**440–444.
28. **Simonet, M., B. Riot, N. Fortineau, and P. Berche.** 1996. Invasin production by *Yersinia pestis* is abolished by insertion of an IS*200*-like element within the *inv* gene. Infect. Immun. **64:**375–379.
29. **Skurnik, M., A. Peippo, and E. Ervela.** 2000. Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. Mol. Microbiol. **37:**316–330.
30. **Swofford, D. L.** 1998. PAUP. Phylogenic analysis using parsimony (and other methods), version 4. Sinauer Associates, Sunderland, Mass.
31. **Welkos, S. L., A. M. Friedlander, and K. J. Davis.** 1997. Studies on the role of plasminogen activator in systemic infection by virulent *Yersinia pestis* strain C092. Microb. Pathog. **23:**211–223.
32. **Wilmoth, B. A., M. C. Chu, and T. J. Quan.** 1996. Identification of *Yersinia pestis* by BBL crystal enteric/nonfermenter identification system. J. Clin. Microbiol. **34:**2829–2830.
33. **Worsham, P. L., and M. Hunter.** 1998. Characterization of Pestoides F, an atypical strain of *Yersinia pestis*. Med. Microbiol. **6**(Suppl. II)**:**34–35.
34. **Wu, L.-T., J. W. H. Chun, R. Pollitzer, and C. Y. Wu.** 1936. Plague: a manual for medical and public health workers. Shangai, China.