JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 2000, p. 1516–1519 Vol. 38, No. 4
0095-1137/00/$04.00+0

Diversity in a Variable-Number Tandem Repeat from *Yersinia pestis*

D. M. ADAIR,[1] P. L. WORSHAM,[2] K. K. HILL,[3] A. M. KLEVYTSKA,[1] P. J. JACKSON,[3] A. M. FRIEDLANDER,[2] AND P. KEIM[1]*

Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011-5640[1]; U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21702-5011[2]; and Environmental Molecular Biology Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545[3]

Received 23 August 1999/Returned for modification 5 November 1999/Accepted 7 January 2000

We have identified a tetranucleotide repeat sequence, $(CAAA)_N$, in the genome of *Yersinia pestis*, the causative agent of plague. This variable-number tandem repeat (VNTR) region has nine alleles and great diversity (calculated as 1 minus the sum of the squared allele frequencies) (diversity value, 0.82) within a set of 35 diverse *Y. pestis* strains. In contrast, the nucleotide sequence of the *lcrV* (low-calcium-response) gene differed only slightly among these strains, having a haplotype diversity value of 0.17. Replicated cultures, phenotypic variants of particular strains, and extensively cultured replicates within strains did not differ in VNTR allele type. Thus, while a high mutation rate must contribute to the great diversity of this locus, alleles appear stable under routine laboratory culture conditions. The classic three plague biovars did not have single identifying alleles, although there were allelic biases within biovar categories. The antiqua biovar was the most diverse, with four alleles observed in 5 strains, while the orientalis and mediaevalis biovars exhibited five alleles in 21 strains and three alleles in 8 strains, respectively. The CAAA VNTR is located immediately adjacent to the transcriptional promoters for flanking open reading frames and may affect their activity. This VNTR marker may provide a high-resolution tool for epidemiological analyses of plague.

Plague is a disease caused by *Yersinia pestis*, a gram-negative bacterium in the family *Enterobacteriaceae* (10). There have been three major historical plague pandemics, in the 6th, 14th, and 20th centuries. The Justinian pandemic (sixth century) is reported to have killed 100 million persons in Europe, including 40% of the Constantinople (now Istanbul) population (11). The second pandemic began in the 14th century but continued well into the 17th century and resulted in perhaps 44 million European deaths in the 14th century alone. The most recent pandemic arose in China and spread quickly around the world, aided by modern transportation. Cases of plague still occur in association with established reservoirs in wild rodents in primarily rural settings.

Molecular strain discrimination is proving to be a valuable approach to the epidemiological understanding of *Y. pestis* (6, 9). Many commonly used PCR methods, including REP-PCR, AFLPs, and randomly amplified polymorphic DNA techniques, are usually biallelic, a fact which limits their diversity. In contrast, variable-number tandem repeats (VNTRs) are genomic regions with potentially extreme variation and, hence, great strain discrimination capacity (2, 5, 7, 8, 14). Identification of VNTRs has been the limiting factor in the development of such markers, but with an increasing number of partial and complete microbial genome sequences (3), this problem is greatly decreased.

In this study, we have examined the nearly complete *Y. pestis* genome sequence for simple sequence VNTRs and have discovered a tetranucleotide repeat. The repeat is in an intergenic region between two tentatively identified open reading frames (ORFs). PCR primers flanking the repeat motif have been used to characterize the repeat diversity of 35 *Y. pestis* strains and 1 strain of each of the closely related species *Y. pseudotuberculosis* and *Y. enterocolitica*. We have also examined the nucleotide diversity in the V antigen gene (*lcrV*) where, in contrast to the VNTR locus, there was a lack of nucleotide differences. Hence, the tetranucleotide repeat provides great strain discrimination potential for future epidemiological analyses.

* Corresponding author. Mailing address: Department of Biological Sciences, Northern Arizona University, P.O. Box 5640, Flagstaff, AZ 86011-5640. Phone: (520) 523-1078. Fax: (520) 523-7500. E-mail: Paul.Keim@nau.edu.

MATERIALS AND METHODS

Comparative sequencing of the V antigen gene. The complete nucleotide sequence was determined for the V antigen gene (*lcrV*) for most of the unique *Y. pestis* strains listed in Table 1, with the exception of the Indian Isolate, which is *lcrV* negative, and only CO92 from the U.S. isolates. PCR and DNA sequencing primers were designed on the basis of GenBank accession no. M26405. These primers were as follows: 1F, ATTAAGCGTCAGAGGGAGAG (nucleotide positions 390 to 409); 1R, CTCTTTTACTCGCTTGATGC (789 to 770); 2R, ATG GTGCCACTACTAGACAG 3′ (1031 to 1012); 3F, TAGCAAGTTGCGTGA AGA (930 to 947); 4F, GGGGCGTTGGGTAATCTG (1222 to 1239); and 4R, CGTTGAGCATGGCGATAGT 3′ (1561 to 1543). Initially, a 1,171-bp amplicon was generated with the 1F and 4R primers and then used as a template for the DNA sequencing reactions. The six primers described above allowed us to determine the entire nucleotide sequence on both strands of the entire amplicon. Hence, the few differences observed among *Y. pestis* strains were confirmed by at least two sequencing determinations.

Identification of the VNTR marker. A BLASTN (GenBank database) search was conducted on genetic sequence data for *Y. pestis* accessed from the Sanger Centre for Genome Research, Hinxton Hall, United Kingdom, data system. Imported sequences were analyzed for consecutive repetitive elements ranging from 2 to 4 nucleotides long. Once a VNTR locus of at least five repeat units was identified, primer pairs were designed to amplify a 200- to 250-bp region around the marker.

Strains and DNA isolation. All *Yersinia* strains listed in Table 1 were obtained from the U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID) culture collection and were initiated from frozen stocks. Strains were first grown on sheep blood agar plates, and then a single colony was transferred into heart infusion broth. The replicate CO92 cultures 1097, 1098, 1099, 1100, 1108, and 1110 were derived from a single colony and then subcultured on either plates or in liquid media for 10 serial passages at either 28 or 37°C. CO92 variants 1101, 1102, 1103, 1104, 1107, and 1109 differ in phenotypic markers or plasmid composition from the original CO92 stock. For DNA preparation, strains were first cultured in broth and then harvested by centrifugation. DNA was released from the cells by digestion with proteinase K in the presence of sodium dodecyl sulfate. Proteins and other cellular components were removed using hexadecyltrimethylammonium bromide (CTAB; Sigma Chemical Co., St. Louis, Mo.) and then chloroform and phenol-chloroform extractions. DNA was precipitated using isopropanol, followed by an ethanol wash. All DNA samples

TABLE 1. *Yersinia* strains and VNTR marker classification

Strain(s)[a]	Geographical origin[b]	Biovar[c]	VNTR allele[d]	Amplicon size (bp)	No. of CAAA repeats
Angola	Angola	A	C	226	3
Antiqua	DROC	A	L	262	12
Harbin 35	Manchuria	A	H	246	8
Pestoides F and G	FSU	A	D	230	4
Kim10 and variant	Kurdistan	M	K	258	11
Nicholisk 41	Manchuria	M	H	246	8
Pestoides A, B, C, D, Aa, and Bb	FSU	M	I	250	9
Pestoides J	FSU	?	M	266	13
195/P and Indian isolate	India	O	K	258	11
Yp-111 and EV76-Lot 4	Madagascar	O	K	258	11
Java 9	Indonesia	O	H	246	8
La Paz and replicate	Bolivia	O	I	250	9
Russian vaccine	?	O	K	258	11
Stavropol	FSU	O	K	258	11
684	US	O	G	242	7
538, 752, Alexander, Dobson, and Shasta	US	O	H	246	8
242 and A12	US	O	I	250	9
CO92[e], 564, 1171, South Park, and Yreka	US	O	J	254	10
Y. pseudotuberculosis strain PB1/+	US		B	222	2
Y. enterocolitica strain WA	US		NA[f]		

[a] All strains are from the USAMRIID collection.
[b] DROC, Democratic Republic of the Congo; FSU, Former Soviet Union; US, United States.
[c] M, mediaevalis; O, orientalis; A, antiqua.
[d] Allele designations are based on the number of repeats: A, 1; B, 2; and so forth.
[e] Fourteen replicates of CO92 were analyzed, with identical results.
[f] NA, no amplification.

were passed through a 0.22-μm-pore-size filter and checked for sterility prior to transfer to other laboratories.

PCR amplification. PCR amplification of the CAAA VNTR region was accomplished using two locus-specific primers: primer 1, GGTTAGGTAGGGTG TTGAAG; and primer 2, AAAGAGGCTAAGTGGCAA. PCR mixtures contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl, 0.001% gelatin, 0.2 mM each deoxynucleotide triphosphate, 1 μM R110-dUTP (Roche Molecular Systems, Inc., Branchburg, N.J.), 5 U of AmpliTaq DNA polymerase (Roche), and 20 pmol of each primer per 100 μl of total reaction volume. PCR conditions began with initial denaturation of the DNA at 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1 min. A final extension at 72°C for 5 min was followed by a 4°C soak. Because the PCR amplicons were labeled with fluorescent dUTP, analysis could be performed on an ABI377 automated DNA sequencer. This protocol labels both strands; hence, two fragments are observed in denaturing gels. Sizing of the amplicons was accomplished using Genescan software (Applied Biosystems).

Regions flanking the VNTR. ORFs flanking the VNTR region were identified using GeneQuest software (DNASTAR Inc., Madison, Wis.). BLASTN searches of GenBank were used to identify homologs of these two ORFs.

RESULTS

Comparative sequencing of the V antigen gene. To understand the genetic variation among *Y. pestis* strains and closely related species, we determined the complete *lcrV* gene (V antigen) sequence for 22 samples. We found that the *lcrV* gene had very little nucleotide variation and, with the exception of two strains, all sequences were identical to the previously reported sequence (12). The Angola strain differed at three single nucleotide positions from the others (GenBank accession no. AF167310), while the Pestoides F strain differed by a 16-nucleotide deletion (15); the Pestoides F difference involved a single deletion involving two direct repeats at the C terminus of the protein (GenBank accession no. AF167309). The *lcrV* gene diversity (1 minus the sum of the squared allele frequencies) based upon the three allele frequencies was low, at 0.17, for these strains. Overall, the lack of *lcrV* differences among these strains suggests that little evolutionary distance separates them from a common ancestor.

VNTR marker identification. In contrast to the lack of *lcrV* gene diversity, we have discovered a region within the *Y. pestis* genome that is highly polymorphic. The basis of this diversity is a tetranucleotide repeat sequence discovered by performing BLASTN searches of the partially completed *Y. pestis* genome sequence (Sanger Centre for Genome Research; http://www.sanger.ac.uk/Projects/Y_pestis). We found a CAAA tandem array with eight repeats in an intergenic region of CO92 (Fig. 1). The genomic regions flanking the CAAA repeats were analyzed for ORFs and regulatory sequences. ORFs were identified on both the 5′ and the 3′ sides of the repeats and were oriented with their 5′ transcribed regions adjacent to the tandem array (i.e., both promoters must be near the tandem array). The 3′ ORF (designated *Orf-1*) was 420 nucleotides long and separated by only 56 nucleotides from the CAAA tandem array. *Orf-1* had a strong similarity (72.2%; BLAST probability, 2.9×10^{-68}) to an *Escherichia coli* ORF of unknown function (5). The second ORF (designated *Orf-2*) was 560 nucleotides long and located 130 nucleotides 5′ of the CAAA repeats (Fig. 1). BLASTN searches of *Orf-2* against the GenBank database showed only a weak correlation (similarity, 50.8%; BLAST probability, 0.92) to the *Bacillus subtilis tagO* gene, encoding undecaprenyl-phosphate-*N*-acetylglucosaminyltransferase (GenBank accession no. AJ004803). Thus, the region 5′ to the VNTR appears to be a monocistronic operon similar to genes encoding teichoic acid synthesis in gram-positive bacteria. Putative ribosome binding sequences and transcriptional promoters were identified for both of the ORFs.

VNTR marker frequency. The CAAA repeat region was found to be a VNTR by PCR amplification of multiple *Y. pestis* strains. Amplicons from each strain were analyzed for size by electrophoresis on denaturing polyacrylamide gels. Among the 35 unique *Y. pestis* strains, there were nine different amplicon lengths varying with a periodicity of 4 nucleotides (Table 1). However, not all possible 4-nucleotide variants of between 1 and 12 repeats were observed. No single-repeat alleles were

```
GTAGCAAAAGCGACTAACTGTTGCTCAAAGCGCGGTAGGTCTGCTAACAAGTCCTGTAATTCAATGATCTCTTCCAGGCC      80
←Orf-2 |      SD          Primer # 1
GTGCATGTGGTTCTCCGTCTGATGGGTTAGGTAGGGTGTTGAAGCGGCAACAATTTACCGTTATTTATGCTCGTCAGCCA      160

GATATAATCTCCCCCTCTTTTTATACCCAACGTCACTGGCGTTGCCGCCAGGCCAACAAACAAACAAACAAACAAACAAA      240
                                                                    | Orf-1 →
CAAACAAACAAACAAATCTCACCGAATTGACATTAAAGATATGATTCGAGTGAGTGAACACCGTTAGCTCTTATGGCTC      320
                    Primer #2
AAGTTCGAAGGGGAGGCGAATAAAATTGCCACTTAGCCTCTTTGGCGATGAGATTCGCCGCAACCACGGGCCTGGTACTG      400
```

FIG. 1. Nucleotide sequence of the VNTR region. The CAAA repeat region that varies among *Y. pestis* strains is underlined. *Orf-1* shows strong similarity to a gene encoding a 190-amino-acid hypothetical protein from *E. coli*. *Orf-2* shows similarity to the *B. subtilis* gene *tagO* involved in teichoic acid synthesis. The putative ribosome binding sequence for *Orf-1* (SD) is underlined. The PCR priming sequences are labeled and underlined.

observed, and the only two-repeat allele was observed in *Y. pseudotuberculosis*. All possible alleles containing between 3 and 13 CAAA repeats were observed within the *Y. pestis* strains, with the exception of the 5- and 6-repeat alleles. The number of repeated CAAA sequences was calculated from the difference in size between the sequences of the observed alleles and the CO92 sequence. In addition, one allele from each size category was completely sequenced to confirm the number of repeats (data not shown). The diversity value for the CAAA repeat region was calculated to be 0.82 for the 35 unique *Y. pestis* strains (Table 1). This value is very high and possible only with multiple allelic loci. Alleles containing between 8 (allele H) and 11 (allele K) repeats were the most common and represented 83% (ca. 20% each) of the 35 unique *Y. pestis* samples (Table 1). The primers flanking the VNTR region did not successfully amplify the single *Y. enterocolitica* strain analyzed.

Allele types and frequencies were different among the three biovar categories (Table 1). Mediaevalis samples were dominated by the nine-repeat allele (I). Only one of the eight unique mediaevalis samples contained a different allele (strain Kim10, 11 repeats, allele K). Orientalis samples were nearly equally represented by four alleles (H to K) but also had one G allele-containing strain. The antiqua samples were diverse, with four different alleles among five strains. The 13 North American strains in this study were also diverse, exhibiting four alleles (G to J), but there were no representatives with the K allele that was common in the orientalis strains.

Allele stability. While the large number of alleles must result from elevated mutation rates, multiple laboratory cultures derived from one strain had identical alleles. This finding was demonstrated after multiple independent serial transfers of strain CO92 under different laboratory conditions (e.g., temperature and media). Fourteen independent cultures were examined, and all contained 10 copies of the CAAA repeats. Likewise, replicates or variants of La Paz, EV76, and Kim10 had the same VNTR alleles. Therefore, it appears that the VNTR mutation rate is not elevated to the extent that strain identity will be lost during laboratory maintenance.

DISCUSSION

We have identified a unique VNTR (5′-CAAA-3′) sequence from *Y. pestis* and characterized the repeat variation in 53 *Yersinia* DNAs, 35 of which are from unique *Y. pestis* strains. Nine alleles of the CAAA VNTR were represented within the 35 *Y. pestis* strains tested. This is a highly diverse genetic region (diversity value, 0.82) that may provide great discrimination in

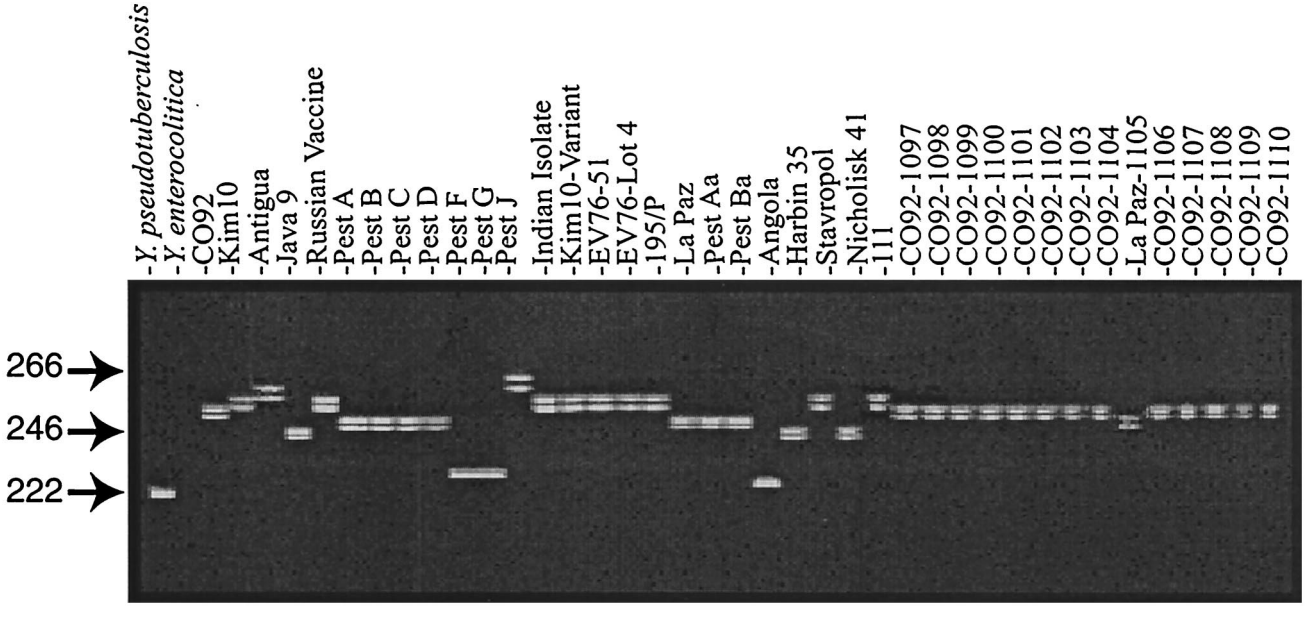

FIG. 2. Gel image of different alleles in the CAAA repeat region. All samples except *Y. enterocolitica* produced an amplicon in reactions containing the PCR primers described in the text. Strains correspond to those listed in Table 1; Pest, Pestoides. Numbers at left are in base pairs.

epidemiological analyses. The diversity observed in the CAAA repeat region was extremely high, in contrast to the nucleotide diversity observed across a 1,171-bp amplicon associated with the *lcrV* gene. Here, only three haplotypes were observed with a low diversity (diversity value, 0.17). The strong bias toward the collection of virulent strains may represent a selective force that results in a narrow definition of a pathogen and low diversity. Alternatively, all *Y. pestis* strains may be derived from a relatively recent ancestral strain and, thus, not have had sufficient evolutionary time to differentiate. The *lcrV* results are consistent with those of Achtman et al. (1), who examined five housekeeping genes in 36 *Y. pestis* strains without detecting even a single nucleotide difference. In contrast to the sequence homogeneity of this pathogen, the CAAA VNTR is highly diverse, with great discriminatory power as a marker among different strains.

While the biovar categorization and VNTR types did not perfectly correlate, there were definite allelic trends. Biovar antiqua (allele A) had the highest diversity of the alleles. This finding is consistent with its supposedly more ancient status. It has been suggested that the antiqua biovar is associated with the Justinian plague, the first recorded pandemic noted in history, around the year 542 A.D. (11). One can speculate that this biovar represents the earliest known strains of the bubonic plague and contains the largest number of VNTR categories. The 8 mediaevalis samples had only three VNTR alleles, while the 21 orientalis samples had five alleles. It has been suggested that the mediaevalis biovar was associated with the Black Death prevalent in 14th-century Europe. Orientalis is known from the current pandemic mainly seen today in animal reservoirs (11). In the 13 North American strains, four different alleles were observed (alleles H to K).

Identification of surrounding ORFs and the proteins that they encode may be one way of understanding the function of the VNTRs. Examples from eukaryotes and prokaryotes indicate that VNTRs can alter gene expression and, in some cases, cause genetic diseases (4, 13). The CAAA repeat region is in proximity to the putative transcriptional promoters of two ORFs. Alteration of transcriptional activity via RNA polymerase binding could be accomplished by changes in repeat length. Alternatively, selection against transcriptional changes could inhibit the formation of particular VNTR alleles. Note that the repeat numbers centered around 8 to 11 CAAA copies, with only a few observations of larger or smaller numbers. This restricted and nonrandom distribution seems to indicate constraints in the generation of VNTR diversity.

ACKNOWLEDGMENTS

We thank Melissa Hunter, Lance B. Price, and James M. Schupp for excellent technical assistance.

This work was supported in part by funds from the U.S. Department of Energy, the National Institutes of Health, and the Cowden Endowment in Microbiology.

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. **Andersen, G. L., J. M. Simchock, and K. H. Wilson.** 1996. Identification of a region of genetic variability among *Bacillus anthracis* strains and related species. J. Bacteriol. **178:**377–384.
3. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1461.
4. **Field, D., and C. Wills.** 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. Proc. Natl. Acad. Sci. USA **495:**1647–1652.
5. **Frothingham, R., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology **144:**1189–1196.
6. **Guiyoule, A., F. Grimont, I. Iteman, P. A. D. Grimont, M. Lefevre, and E. Carniel.** 1994. Plague pandemics investigated by ribotyping of *Yersinia pestis* strains. J. Clin. Microbiol. **32:**634–641.
7. **Jackson, P. J., K. L. Richmond, A. S. Kalif, D. Adair, K. K. Hill, C. R. Kuske, E. Walthers, G. L. Andersen, K. H. Wilson, M. E. Hugh-Jones, and P. Keim.** 1997. Characterization of the variable-number tandem repeats in *vrrA* from different *Bacillus anthracis* isolates. Appl. Environ. Microbiol. **63:**1400–1405.
8. **Keim, P., A. Klevytska, L. B. Price, J. M. Schupp, G. Zinser, R. Okinaka, K. K. Hill, P. Jackson, K. L. Smith, and M. E. Hugh-Jones.** 1999. Molecular diversity in *Bacillus anthracis*. J. Appl. Microbiol. **87:**215–217.
9. **Lucier, T. S., and R. R. Brubaker.** 1992. Determination of genome size, macrorestriction pattern polymorphisms, and nonpigmentation-specific deletion in *Yersinia pestis* by pulsed-field gel electrophoresis. J. Bacteriol. **174:** 2078–2086.
10. **Perry, R. D., and J. D. Fetherston.** 1997. *Yersinia pestis*—etiologic agent of plague. Clin. Microbiol. Rev. **10:**35–66.
11. **Pollitzer, R.** 1954. Plague. WHO Monogr. Ser. **22:**1–698.
12. **Price, S. B., K. Y. Leung, S. S. Barveand, and S. C. Straley.** 1989. Molecular analysis of *lcrGVH*, the V antigen operon of *Yersinia pestis*. J. Bacteriol. **171:**5646–5653.
13. **Richards, R. I., and G. R. Sutherland.** 1997. Dynamic mutation: possible mechanisms and significance in human disease. Trends Biochem. Sci. **22:** 432–436.
14. **Van Belkum, A., S. Scherer, W. Van Leeuwen, D. Willemse, A. Loek, and H. Verbrugh.** 1997. Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. Infect. Immun. **65:**5017–5027.
15. **Worsham, P. L., and M. Hunter.** 1998. Characterization of Pestoides F, an atypical strain of *Yersinia pestis*. Med. Microbiol. **6**(Suppl. II)**:**34–35.