JOURNAL OF BACTERIOLOGY, Jan. 1996, p. 377–384 Vol. 178, No. 2
0021-9193/96/$04.00+0

Identification of a Region of Genetic Variability among *Bacillus anthracis* Strains and Related Species

GARY L. ANDERSEN, JENNY M. SIMCHOCK, AND KENNETH H. WILSON*

Infectious Diseases Section, Department of Veterans Affairs Medical Center, and Division of Infectious Diseases, Duke University Medical Center, Durham, North Carolina

Received 10 July 1995/Accepted 1 November 1995

The identification of a region of sequence variability among individual isolates of *Bacillus anthracis* as well as the two closely related species, *Bacillus cereus* and *Bacillus mycoides*, has made a sequence-based approach for the rapid differentiation among members of this group possible. We have identified this region of sequence divergence by comparison of arbitrarily primed (AP)-PCR "fingerprints" generated by an M13 bacteriophage-derived primer and sequencing the respective forms of the only polymorphic fragment observed. The 1,480-bp fragment derived from genomic DNA of the Sterne strain of *B. anthracis* contained four consecutive repeats of CAATATCAACAA. The same fragment from the Vollum strain was identical except that two of these repeats were deleted. The Ames strain of *B. anthracis* differed from the Sterne strain by a single-nucleotide deletion. More than 150 nucleotide differences separated *B. cereus* and *B. mycoides* from *B. anthracis* in pairwise comparisons. The nucleotide sequence of the variable fragment from each species contained one complete open reading frame (ORF) (designated *vrrA*, for variable region with repetitive sequence), encoding a potential 30-kDa protein located between the carboxy terminus of an upstream ORF (designated *orf1*) and the amino terminus of a downstream ORF (designated *lytB*). The sequence variation was primarily in *vrrA*, which was glutamine- and proline-rich (30% of total) and contained repetitive regions. A large proportion of the nucleotide substitutions between species were synonymous. *vrrA* has 35% identity with the microfilarial sheath protein shp2 of the parasitic worm *Litomosoides carinii*.

Bacillus anthracis, the causative agent of anthrax, inhabits the soil in many temperate regions of the world. Virulent isolates of the organism carry a toxin-encoding plasmid, pXO1, and a capsule-associated plasmid, pXO2, both of which are required to produce disease (7, 18). In addition, chromosomal factors have been implicated as determinants for disease severity (31). Molecular study of the epidemiology and population genetics of *B. anthracis* would be greatly facilitated by the ability to distinguish individual isolates of the pathogen. Although increasingly rare with the advent of an effective animal vaccine, anthrax continues to be endemic in some areas (28).

There are relatively few tests to distinguish isolates of *B. anthracis* from very closely related soil-borne organisms, including *Bacillus cereus* and *Bacillus mycoides*. The primary characteristics used to distinguish *B. anthracis* from these closely related species is the presence of the two virulence plasmids (11). The taxonomy of this group is controversial, given the many similarities, and several workers have suggested that these closely related species should all be grouped as members of *B. cereus* (16, 25). As an example of this similarity, comparison of rRNA between *B. anthracis* and *B. cereus* revealed an identical 16S rRNA sequence and only two differences in the 23S rRNA (1, 2).

The conservation of sequence among isolates of *B. anthracis* has presented an even greater difficulty in the identification of individual strains of this species. To a certain extent, strains can be differentiated by serological methods. The lack of reaction of spores of the Vollum strain of *B. anthracis* to conjugated antibodies raised against the Sterne strain indicated a difference in spore surface antigens that may be used to separate these two isolates (23). Yet it has been more difficult to demonstrate actual differences in DNA sequence between strains. In this laboratory, no difference was found between the Ames and Vollum strains of *B. anthracis*: we found an identical banding pattern with pulsed-field gel electrophoresis and identical DNA sequences of the 16S-23S and the *gyrB-gyrA* intergenic spacer regions (10). Henderson et al. reported no differences in restriction fragment length patterns among 37 isolates of *B. anthracis* for 18 different restriction enzymes (11). However, they did note a slight pattern variation in PCR fragments generated from an M13 bacteriophage-based primer in an arbitrarily primed (AP)-PCR.

We have attempted to differentiate the so-called aberrant or vaccine-resistant Ames strain, a virulent pathogen isolated in the United States (17), from two other strains: the South African strain, Sterne, used for more than 50 years as a live vaccine strain (27), and the Vollum strain, which was isolated in England and lacks a spore surface antigen present in other strains (22). A sequence-based approach for the separation of individual strains of *B. anthracis* has the potential to increase our understanding of the epidemiology of this organism. Also, the identification of a difference in a genomic area apart from the two virulence plasmids would be useful in distinguishing naturally occurring plasmid-free isolates of *B. anthracis* from other members of the *B. cereus* group. Therefore, we designed the present study first to isolate PCR fragment polymorphisms that may arise from an AP-PCR comparison of plasmid-free derivatives of the *B. anthracis* strains. Then, we compared nucleotide sequences corresponding to the polymorphic fragments from strains of *B. anthracis* and closely related species in an effort to identify a region of sequence divergence in the genome. Finally, because sequence divergence has been difficult to find in *B. anthracis*, we attempted to characterize ge-

* Corresponding author. Mailing address: Infectious Diseases Section, VA Medical Center, 508 Fulton St., Durham, NC 27705. Phone: (919) 286-0411, ext. 7308. Fax: (919) 286-6895. Electronic mail address: wilso003@mc.duke.edu.

netically the area where the divergence occurred by searching for homology with known genes.

MATERIALS AND METHODS

Bacterial strains and media. *B. anthracis* strains Ames (Ames, Iowa), Sterne (South Africa), and Vollum (England) were originally isolated from infected cattle. The three strains were cured of virulence plasmids (pXO1$^-$ pXO2$^-$) at the U.S. Army Medical Research Institute of Infectious Diseases, Ft. Detrick, Md., by culturing the strains at 42°C (15). Type strains of *B. cereus* (ATCC 14579) and *B. mycoides* (ATCC 6462) were obtained from the American Type Culture Collection, Rockville, Md. *B. anthracis*, *B. cereus*, and *B. mycoides* were grown in brain heart infusion broth (Becton Dickinson, Cockeysville, Md.) at 37°C. *Escherichia coli* was routinely grown in Luria-Bertani medium at 37°C. Media were supplemented, when required, with ampicillin (100 μg/ml).

Recombinant DNA techniques. Large- and small-scale isolations of recombinant plasmids from *E. coli* were performed by alkaline lysis (14). For total genomic DNA isolation, mid-log-phase cells of *Bacillus* species were treated with ampicillin-sulbactam (1 μg/ml) for 30 min (to weaken cell walls) prior to washing in GTE buffer (50 mM glucose, 25 mM Tris-HCl [pH 8.0], 10 mM EDTA). The washed pellet was frozen at −20°C to disrupt cell walls, thawed on ice, resuspended in GTE buffer containing lysozyme (10 mg/ml), and incubated at 37°C for 1 h. After additional treatment with 1% *N*-lauroyl sarcosine, proteinase K (200 μg/ml), and 1% CTAB (hexadecyltrimethyl ammonium bromide), DNA was extracted in phenol-chloroform and chloroform-isoamyl alcohol, ethanol precipitated, and resuspended in TE (10 mM Tris-HCl, 1 mM EDTA) buffer. Digestion of DNA with restriction endonuclease, transformation of competent *E. coli*, and other standard procedures were performed as described by Sambrook et al. (24). DNA fragments were resolved by electrophoresis in 90 mM Tris-borate–2 mM EDTA in gels containing 1.0 to 2.0% (wt/vol) agarose or in 3% (wt/vol) agarose gels consisting of a mixture of 2% (wt/vol) NuSieve GTG (FMC Inc., Rockland, Maine) and 1% (wt/vol) agarose (Sigma). The DNA fragments were purified from agarose gels with the Qiaex Kit (Qiagen Inc., Studio City, Calif.) in accordance with the manufacturer's protocols.

Genomic fingerprinting. Reaction mixtures (100 μl) contained 3.0 mM $MgCl_2$, a single oligonucleotide primer at 1.0 or 3.0 μM, 0.2 mM each deoxynucleoside triphosphate (dNTP), 2.5 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus, Norwalk, Conn.), and 30 or 150 ng of *B. anthracis* genomic DNA. The DNA template and primer concentrations were varied to rule out artifact band polymorphisms. The M13 fingerprinting probe 5′-GAGGGTGGCGGCTCT-3′ (30) for generating a slight pattern variation in the 400- to 500-bp region of *B. anthracis* has been described previously (11). The reaction mixture was denatured for 10 min at 95°C, followed by 40 cycles of 15 s at 96°C (denaturation), 15 s at 36°C (annealing), and 2 min at 72°C (extension) in a Perkin-Elmer Cetus model 9600 thermal cycler. After completion of the cycling, the reaction tubes were held at 72°C for an additional 10 min to minimize single-stranded DNA product.

Cloning procedures. PCR-generated DNA fragments were cloned into the pCRII vector of the TA Cloning System (Invitrogen Co., San Diego, Calif.) following the manufacturer's recommendations. The sense strand primers used to generate amplified products from *Bacillus* genomic DNA were 5′-AGCGTA GTTCACGAACTGC-3′, 5′-TAACTGAAGAAAGGAGG-3′, and 5′-AGGAG GATTTCTCACATG-3′ (*B. cereus* specific). The antisense primers were 5′-TG ACATGGTAACCTTCAGC-3′, 5′-TGATACCATCTTCTTCG-3′, and 5′-ATT GTGACCAATCATACC-3′ (*B. cereus* specific). Restriction endonuclease DNA fragments were blunt-ended by incubation at 37°C with 3.0 U of exonuclease V (US Biochemical Co., Cleveland, Ohio) for 30 min. After inactivation of the enzyme with EDTA and ethanol precipitation, the restriction fragments were resuspended in TE buffer and ligated into *Sma*I-cleaved pBluescript II SK+ vector (Stratagene Inc., La Jolla, Calif.). All resulting transformants were stored at −70°C in 15% glycerol.

PCR amplification and DNA sequencing. For routine PCR amplification of DNA fragments, a 100-μl reaction mixture contained 10 μl of GeneAmp 10× PCR buffer with 15 mM $MgCl_2$ (Perkin-Elmer Cetus), 20 pmol of each primer, 0.2 mM each dNTP, 2.5 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus), and 1 to 10 ng of DNA. The initial denaturation at 95°C for 2 min was followed by 40 cycles of denaturation at 92°C for 30 s, annealing at 55°C (or 45°C low-stringency annealing) for 30 s, and extension at 72°C for 1 min. A small, polymorphic region of the variable sequence was amplified from *Bacillus* genomic DNA with the following primers: EWA2, 5′-ATGGTTCCGCCTTATCG-3′; EWA3, 5′-CAAATGGTTCCGCC-3′ (sense strand); and EWA1, 5′-TATCCTT GGTATTGCTG-3′ (antisense strand). To minimize primer pair interactions, a mixture of the primer pair and dNTP were separated from a mixture of *Taq* polymerase and DNA template by Perkin-Elmer Cetus Ampliwax paraffin beads by the hot-start technique (20). The reaction mixture was amplified for 35 cycles of denaturation at 90°C for 30 s (except for the first two cycles, during which denaturation occurred at 95°C), annealing at 50°C for 2 min, and extension at 75°C for 30 s. Sequencing of amplified products or cloned DNA was performed on an Applied Biosystems 373A DNA sequencer with a *Taq* DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems, Foster City, Calif.). Nucleotide sequences were determined unambiguously on both sides by primer walking and sequencing of multiple clones. Sequences were compiled with the Genetic Data Environment version 2.2 program (University of Illinois, Urbana, Ill.). Computer analysis of nucleotide and protein sequences was performed with the BestFit, PileUp, TFastA, and BLAST (Basic Local Alignment Search Tool) programs supplied with Genetics Computer Group version 8 sequence analysis package (University of Wisconsin, Madison).

Nucleotide sequence accession numbers. The nucleotide sequences of *vrrA* and flanking sequences shown in Fig. 3 have been deposited in GenBank and assigned accession numbers L48552 (Ames), L48553 (Sterne), L48554 (Vollum), L48555 (*B. cereus*), and L48556 (*B. mycoides*).

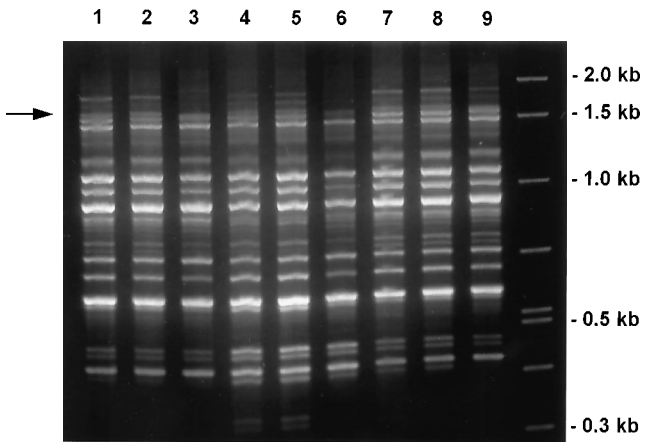


FIG. 1. Genomic fingerprint of three *B. anthracis* isolates at various substrate concentrations. The AP-PCR conditions with the M13 fingerprinting primer are described in Materials and Methods. The following isolates of *B. anthracis* were analyzed: Ames (lanes 1, 4, and 7), Sterne (lanes 2, 5, and 8), and Vollum (lanes 3, 6, and 9). The reaction mixture contained 150 ng of genomic DNA (lanes 1 to 6) or 30 ng of genomic DNA (lanes 7 to 9) and 3 μM primer (lanes 4 to 6) or 1 μM primer (lanes 1 to 3 and 7 to 9). The far right lane contains DNA size markers. The left-side arrow indicates the approximate position of the PCR polymorphism.

RESULTS

Genomic fingerprinting of *B. anthracis*. An AP-PCR primed by the M13 fingerprinting probe yielded a polymorphism in a 1.5-kb fragment that differentiated the Vollum strain of *B. anthracis* from the Ames and Sterne strains. This band, which migrated slightly faster in Vollum, was present under all reaction conditions tested, including primer concentrations of 1.0 and 3.0 μM and DNA concentrations of 30 and 150 ng in a 100-μl reaction volume (Fig. 1). Other polymorphisms among several faint bands were observed which correlated with reaction conditions and could not separate the *B. anthracis* isolates (including a polymorphic 0.8-kb fragment with 150 ng of genomic DNA and 1 μM primer and a polymorphic 1.7-kb fragment with 30 ng of genomic DNA and 3 μM primer [Fig. 1]). The polymorphic 1.5-kb PCR fragments from strains Sterne and Vollum were cloned into vector pCRII, and the resulting transformants, containing plasmids pST1 and pVO1, respectively, were digested with the restriction endonuclease *Nla*III. Clones of both strains had restriction fragments of 670, 271, 190, 154, and 37 bp (data not shown). The difference was localized to a 174-bp fragment from pST1 (150 bp in pVO1) that was subcloned into pBluescript II SK+. The resulting plasmids, designated pST2 (Sterne) and pVO2 (Vollum), were sequenced. The difference in size between pVO2 and pST2 was traced to a deletion of two consecutive CAATATCAACAA repeats.

Amplification of a variable region in several *Bacillus* species. The 24-base deletion in Vollum was directly detected from genomic DNA with primers derived from sequence near the end of the cloned inserts in pST2 and pVO2 (Fig. 2). The

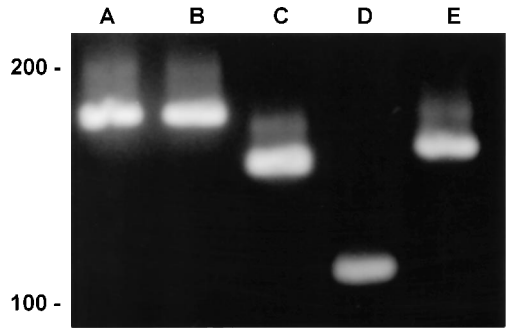


FIG. 2. Differentiation of *Bacillus* isolates with a PCR fragment length polymorphism. The primers EWA1 and EWA2 (see text) were used to generate DNA fragments from (lane A) Ames isolate of *B. anthracis*, (lane B) Sterne isolate of *B. anthracis*, (lane C) Vollum isolate of *B. anthracis*, (lane D) *B. cereus* ATCC 14579, and (lane E) *B. mycoides* ATCC 6462. The relative migration positions of DNA size standards are given (in base pairs) on the left.

167-bp PCR products from Ames and Sterne could not be differentiated from each other but differed from Vollum as indicated. We were able to rapidly differentiate the closely related species *B. cereus* (approx. 100 bp) and *B. mycoides* (approx. 150 bp) from *B. anthracis* by their unique sizes in the variable region with the primer pair EWA1 and EWA2 (Fig. 2). The primer pair EWA1 and EWA3 also amplified a unique PCR product in Vollum to differentiate it from Sterne and Ames but did not amplify the variable region in *B. cereus* and *B. mycoides* (data not shown).

Nucleotide sequence of the variable region. The nucleotide sequence of genomic DNA corresponding to the approximately 1.5-kb polymorphic AP-PCR fragment revealed that the three strains of *B. anthracis* could be differentiated from each other on the basis of sequence divergence in this region. There was a deletion of a single adenine residue in a poly(A) region of the 1,479-bp Ames fragment as well as a 24-bp deletion in the 1,456-bp Vollum fragment to differentiate these two strains from the 1,480-bp Sterne fragment (Fig. 3). The deletion in Vollum occurred in two of the four CAATATCAA CAA repeating units of DNA present in Ames and Sterne (Fig. 3). A related region of DNA was found in *B. cereus* and *B. mycoides* by low-stringency PCR amplification with primers derived from the end of the polymorphic AP-PCR fragment. The 1,444-bp *B. cereus* fragment differed from the Sterne fragment at 158 positions, including 62 transitions, 48 transversions, and 48 deletions or insertions (Fig. 3). The 1,434-bp *B. mycoides* fragment differed from the Sterne fragment at 170 positions and from the *B. cereus* fragment at 154 positions (Fig. 3).

Analysis of the predicted amino acid sequence. Three putative open reading frames (ORFs) were predicted from the nucleotide sequence of the variable region of DNA in the three strains of *B. anthracis* as well as *B. cereus* and *B. mycoides* (Fig. 3). The predicted length of the only completely sequenced ORF (designated *vrrA*, for variable region with repetitive sequence) in Sterne was 256 residues, encoding a 30-kDa protein. Additionally, two partial ORFs were predicted from sequence flanking *vrrA*, including a carboxy-terminal 60 residues and an amino-terminal 123 residues, designated *orf1* and *lytB*, respectively (see Fig. 3). Potential −35 and −10 regions similar to the consensus sequence of the housekeeping σ^A-dependent promoter as well as Shine-Dalgarno sequences approximately 10 nucleotides upstream from the initiation codons were observed in *vrrA* and *lytB*, consistent with expressed *Bacillus* genes (19). The deletion in Vollum corresponded to two units of a repeating QYQQ sequence in VrrA that was also deleted in *B. cereus* (Fig. 4). A truncated VrrA in Ames, lacking the amino-terminal 24 residues, was predicted from its single-nucleotide deletion. There was no difference in the amino acid sequence of the partial ORFs *orf1* and *lytB* among the three strains of *B. anthracis* and only minor differences compared with those of *B. cereus* and *B. mycoides*.

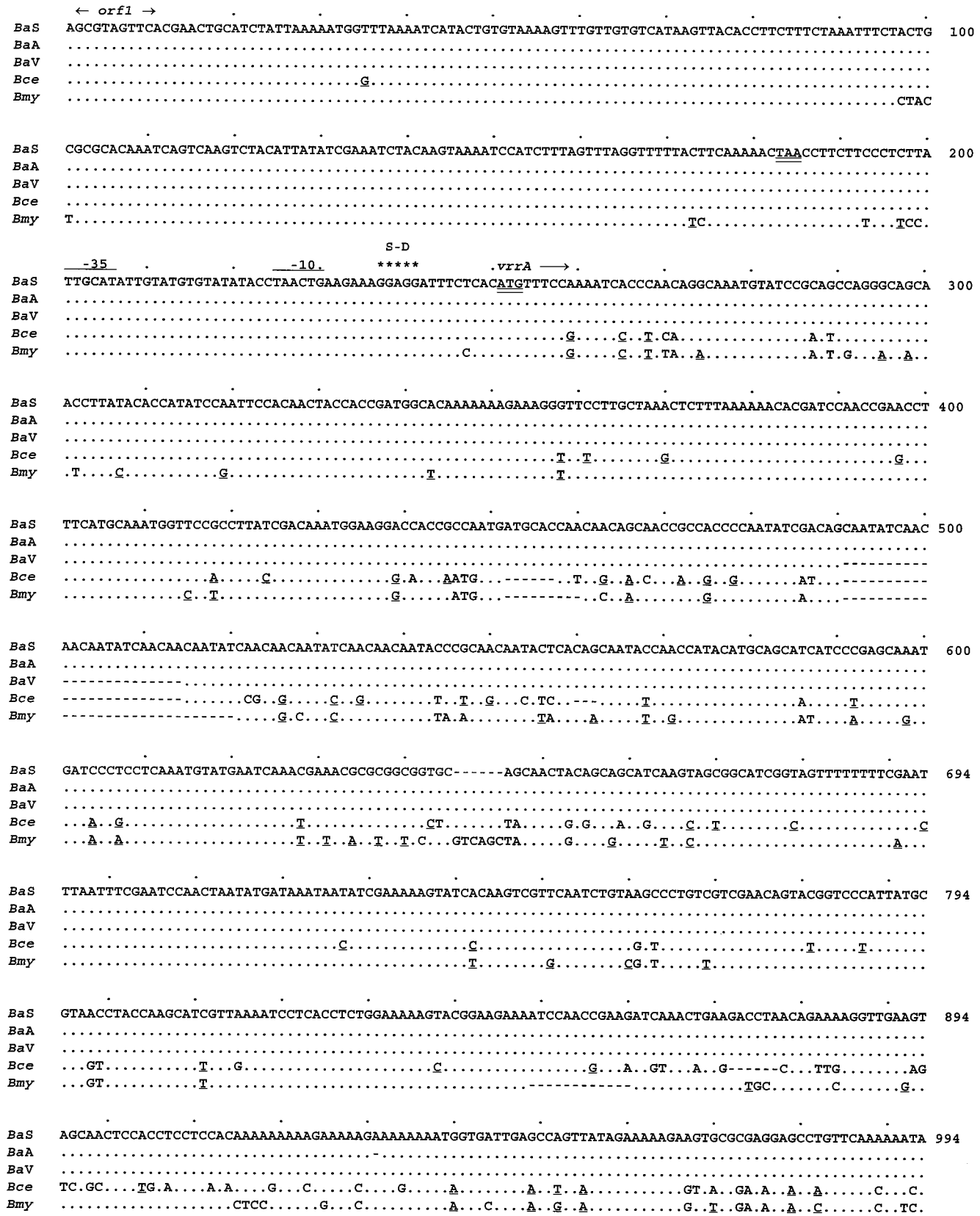
The amino acid sequence encoded by *vrrA* was most similar to that of the *shp2* genes of the parasitic nematodes *Litomosoides carinii*, *Brugia malayi*, and *Brugia pahangi*, encoding a microfilarial sheath protein, when compared with protein sequence databases at the National Center for Biotechnology Information (Bethesda, Md.) by the BLAST program (Fig. 5). The partially sequenced amino terminus of the downstream ORF has regions of significant codon similarity to *lytB*, a gene involved in penicillin tolerance and control of the stringent response in *E. coli* and *Pseudomonas fluorescens*. The overall identity of conserved and identical residues from the *P. fluorescens lytB* product to the partial 123 residues of the *B. anthracis lytB* product is 51% (data not shown), while *E. coli lytB* (*orf316*) is 56% identical (Fig. 6). A comparison of the carboxy terminus of an upstream ORF with published sequences by the BLAST and TFastA algorithms did not reveal any significant areas of homology.

DISCUSSION

Detailed studies of the population genetics of *E. coli* and *Salmonella* species have shown a great deal of clonal diversity and worldwide distribution of clones (4, 21). However, because these organisms colonize humans and humans continually mix geographically, it is not clear that these conclusions apply to most bacteria. *B. anthracis* differs ecologically from the abovementioned organisms in that there is thought to be a significant environmental reservoir and that the human host is unimportant to its spread. Furthermore, studying its population genetics presents a major challenge because its chromosome is highly uniform (26). To develop a basis for a better definition of the genetic diversity of *B. anthracis*, we have identified a region of sequence divergence that allows differentiation of the three geographically distinct isolates examined. This region of DNA differs significantly in the closely related species *B. cereus* and *B. mycoides*, and we found that in all cases, the divergence lay within an ORF.

AP-PCR has been well documented as a method to fingerprint subgroups of closely related organisms (32, 34). Possible explanations for sources of polymorphisms include insertions and deletions in the genome and particular mismatches in the primer-genomic DNA duplex (34). Polymorphic AP-PCR fragments are potentially useful for pinpointing sequence differences. However, small differences in substrate concentration with AP-PCR have the potential to cause variations in the observed banding patterns (33). Although the polymorphism found in our laboratory was present at several different substrate concentrations, it did not correspond to the one described by Henderson et al. (11). The type of thermal cycler used and the temperature profile of the amplification reaction may have contributed to the differences in banding pattern observed between the two laboratories. Despite the drawbacks to this method, the sequence derived from the polymorphic PCR fragment was useful in defining unambiguous differences between strains.

The 24-nucleotide deletion in Vollum is sufficiently large to be visualized on an agarose gel, allowing direct discrimination from other isolates by PCR amplification of a portion of the variable region (Fig. 2). However, the single-nucleotide dele-


← orf1 →
BaS AGCGTAGTTCACGAACTGCATCTATTAAAAATGGTTTAAAATCATACTGTGTAAAAGTTTGTTGTGTCATAAGTTACACCTTCTTTCTAAATTTCTACTG 100
BaA
BaV
Bce .....G.....
Bmy .....CTAC
BaS CGCGCACAAATCAGTCAAGTCTACATTATATCGAAATCTACAAGTAAAATCCATCTTTAGTTTAGGTTTTTACTTCAAAAACTAACCTTCTTCCCTCTTA 200
BaA
BaV
Bce
Bmy T.....TC.....T...TCC.
-35 -10. S-D ***** .vrrA ⟶
BaS TTGCATATTGTATGTGTATATACCTAACTGAAGAAAGGAGGATTTCTCACATGTTTCCAAAATCACCCAACAGGCAAATGTATCCGCAGCCAGGGCAGCA 300
BaA
BaV
Bce .....G.....C..T.CA.....A.T.....
Bmy .....C.....G.....C..T.TA..A.....A.T.G...A..A..
BaS ACCTTATACACCATATCCAATTCCACAACTACCACCGATGGCACAAAAAAAGAAAGGGTTCCTTGCTAAACTCTTTAAAAAACACGATCCAACCGAACCT 400
BaA
BaV
Bce .....T..T.....G.....G...
Bmy .T....C.....G.....T.....T.....
BaS TTCATGCAAATGGTTCCGCCTTATCGACAAATGGAAGGACCACCGCCAATGATGCACCAACAACAGCAACCGCCACCCCAATATCGACAGCAATATCAAC 500
BaA
BaV .....
Bce .....A.....C.....G.A...AATG...-------..T..G..A.C...A..G..G.....AT...----------
Bmy .....C..T.....G.....ATG...----------..C..A.....G.....A....----------
BaS AACAATATCAACAACAATATCAACAACAATATCAACAACAATACCCGCAACAATACTCACAGCAATACCAACCATACATGCAGCATCATCCCGAGCAAAT 600
BaA
BaV ----------.....
Bce ----------.....CG..G.....C..G.....T..T..G...C.TC..---.....T.....A.....T.....
Bmy ----------.....G.C...C.....TA.A.....TA....A.....T..G.....AT....A.....G..
BaS GATCCCTCCTCAAATGTATGAATCAAACGAAACGCGCGGCGGTGC------AGCAACTACAGCAGCATCAAGTAGCGGCATCGGTAGTTTTTTTTCGAAT 694
BaA
BaV
Bce ...A..G.....T.....CT.....TA.....G.G...A..G....C..T.....C.....C
Bmy ...A..A.....T..T..A..T..T.C...GTCAGCTA.....G....G.....T..C.....A...
BaS TTAATTTCGAATCCAACTAATATGATAAATAATATCGAAAAAGTATCACAAGTCGTTCAATCTGTAAGCCCTGTCGTCGAACAGTACGGTCCCATTATGC 794
BaA
BaV
Bce .....C.....C.....G.T.....T.....T.....
Bmy .....T.....G.....CG.T.....T.....
BaS GTAACCTACCAAGCATCGTTAAAATCCTCACCTCTGGAAAAAGTACGGAAGAAAATCCAACCGAAGATCAAACTGAAGACCTAACAGAAAAGGTTGAAGT 894
BaA
BaV
Bce ...GT.....T...G.....C.....G...A..GT...A..G------C...TTG.....AG
Bmy ...GT.....T.....TGC.....C.....G..
BaS AGCAACTCCACCTCCTCCACAAAAAAAAAGAAAAAGAAAAAAAATGGTGATTGAGCCAGTTATAGAAAAAGAAGTGCGCGAGGAGCCTGTTCAAAAAATA 994
BaA
BaV
Bce TC.GC....TG.A....A.A....G...C.....C....G.....A.....A..T..A.....GT.A..GA.A..A..A.....C...C.
Bmy .....CTCC.....G...C.....A...C....A..G..A.....G..T..GA.A..A..C.....C..TC.

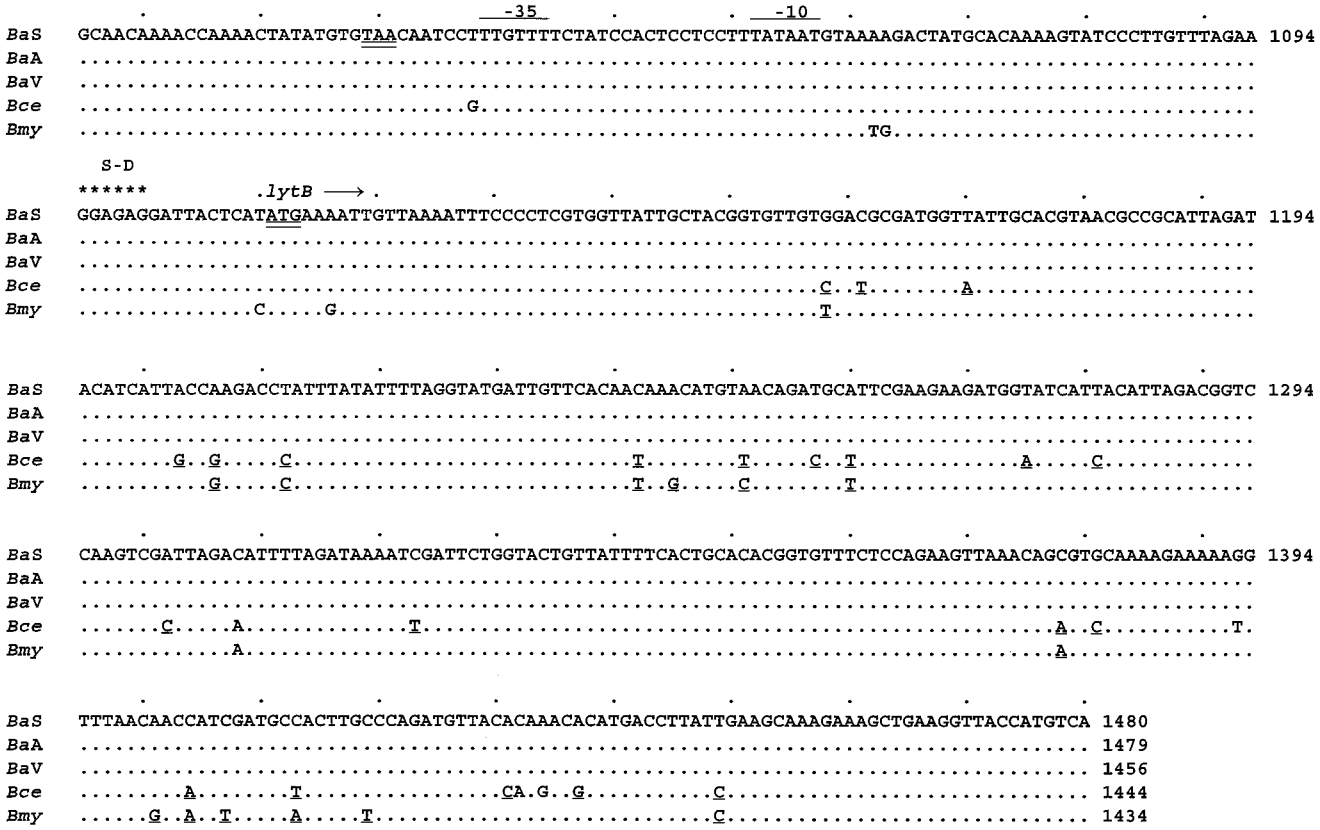


FIG. 3. Alignment of the nucleotide sequence corresponding to a polymorphic PCR fragment containing the *vrrA* ORF and flanking sequence, including a portion of an unidentified upstream ORF (*orf1*) and downstream *lytB* ORF. Sequences were obtained from genomic DNA of *B. anthracis* strain Sterne (*Ba*S), strain Ames (*Ba*A), and strain Vollum (*Ba*V), *B. cereus* (*Bce*), and *B. mycoides* (*Bmy*). The sequences were aligned by using the Genetics Computer Group program PILEUP with default parameters. Nucleotides that are different from those of Sterne at a particular position are indicated; dots indicate identical nucleotides, and a dash shows the deletion of a nucleotide. Underlined nucleotides indicate a synonymous third (3′)-position substitution of a codon in a predicted ORF. Putative start and stop codons are doubly underlined. Numbers at the right represent the position of the last nucleotide base in each line. For Sterne, *orf1* extends 60 codons into the sequenced region to a stop codon at position 183; *vrrA* starts at position 251, extending 256 codons to a stop codon at position 1019; and *lytB* starts at position 1111 and extends beyond the sequenced region. Potential −35 and −10 promoter regions (lines above) and ribosome-binding sites (S-D, asterisks) of *vrrA* and *lytB* are indicated.

tion in Ames is not sufficient for such a rapid, size-based differentiation. Other methods, such as an oligonucleotide probe for the detection of Ames, are not likely to yield useful information because of the long poly(A) region around this deletion (Fig. 3). Thus, determining the sequence of this region appears to be the only way to distinguish Ames from Sterne. By contrast, *B. cereus* and *B. mycoides* varied from the *B. anthracis* isolates by 158- and 170-nucleotide differences, respectively (Fig. 3). Primers derived from these sequences should be useful in rapidly determining whether questionable isolates are *B. anthracis* or closely related members of the *B. cereus* group. Current tests for distinguishing environmental isolates of *B. anthracis* from *B. cereus* are problematic, as variants of the former may lack virulence plasmids or penicillin sensitivity and may possess antigenic cross-reactivity (29). The primer pair EWA1-EWA3 is specific for *B. anthracis*; however, the primer pair EWA1-EWA2 also amplifies DNA fragments in *B. cereus* and *B. mycoides* (Fig. 2). Since there is not a perfect match in the primer-genomic DNA duplex in this region, the amplification may occur at alternative priming sites. One possible binding site for the primer EWA1 in *B. cereus* is CAGCAATAC CAGCAACA (nucleotides 524 to 540 in Fig. 3), with three mismatches corresponding to the 5′ end of the primer. This primer would generate a 101-bp PCR fragment, which would correspond to the product shown in Fig. 2. For *B. mycoides*, the binding of EWA1 with CAGCAATATCCAGAGCAGA (nucleotides 581 to 599 in Fig. 3) would create the observed product of 151 bp.

Our experiments show that the most variable portion of the sequenced region appears to be around the 24-bp deletion in Vollum. Sequence analysis indicated that this area encodes a complete ORF with a potential of creating a 30-kDa protein in Sterne. This region, which we have designated *vrrA*, is rich in glutamine and proline and contains up to seven tandem repeats of QY-hydrophilic residue-Q. *vrrA* is bounded by two partially sequenced ORFs that we have designated *orf1* and *lytB*. The conservation of the amino acid sequence despite divergence of the nucleotide sequence is strong evidence that all three ORFs encode proteins. For *vrrA*, 85 of 103 nucleotide substitutions at the third (3′) positions of codons for Sterne, *B. cereus*, and *B. mycoides* are synonymous. For *lytB*, 29 of 32 nucleotide substitutions at the third (3′) positions of codons are synonymous. Despite the trend towards conservation of particular amino acids in *vrrA*, overall the locus appears to be quite variable, with 61 positions where the amino acid sequence differs among the three species (Fig. 4).

Our finding of the repetitive domain in *vrrA* correlates well with other groups' attempts to differentiate eubacterial strains by sequence polymorphisms. For example, the *Mycobacterium tuberculosis* complex was separated into four distinct groups on

```
BaS  MFPKSPNRQMYPQPGQQPYTPYPIPQLPPMAQKKKGFLAKLFKKHDPTEPFMQMVPPYRQMEGPP   65
BaA  MFPKSPNRQMYPQPGQQPYTPYPIPQLPPMAQKKKGFLAKLFKKHDPTEPFMQMVPPYRQMEGPP   65
BaV  MFPKSPNRQMYPQPGQQPYTPYPIPQLPPMAQKKKGFLAKLFKKHDPTEPFMQMVPPYRQMEGPP   65
Bce  MFPKSPTRQMYPNPGQQPYTPYPIPQLPPMAQKKKGFLAKLFKKHDPTEPFMQMVPPYRQMEG-Q   64
Bmy  MFPKSPIRQMYPNRGQQSYTPYPIPQLPPMAQKKKGFLAKLFKKHDPTEPFMQMVPPYRQMEG-P   64

BaS  PMMHQQQQPPPQYRQQYQQQYQQQYQQQYQQQYQQQYPQQYSQQYQPYMQHHPEQMIPPQMYESNETRG  130
BaA  PMMHQQQQPPPQYRQQYQQQYQQQYQQQYQQQYQQQYPQQYSQQYQPYMQHHPEQMIPPQMYESNETRG  130
BaV  PMMHQQQQPPPQYRQ--------QYQQQYQQQYPQQYSQQYQPYMQHHPEQMIPPQMYESNETRG  122
Bce  PMMHQQPPPP-QYHQ--------QYPQQYQQQYPQQH-PQYQPYMQQHPEQMIPPQMYESNETRG  121
Bmy  PMMHQQ-PPPP-QYQQ----------QQPYQQQYTQQYTQQYQPYMQQYPEQMIPPQMYESNDTRG  118

BaS  GAATT--AASSSGIGSFFSNLISNPTNMINNIEKVSQVVQSVSPVVEQYGPIMRNLPSIVKILTS  193
BaA  GAATT--AASSSGIGSFFSNLISNPTNMINNIEKVSQVVQSVSPVVEQYGPIMRNLPSIVKILTS  193
BaV  GAATT--AASSSGIGSFFSNLISNPTNMINNIEKVSQVVQSVSPVVEQYGPIMRNLPSIVKILTS  185
Bce  -GSTTTAATASSGIGSFFSNLISNPTNMINNIEKVSQVVQSVGPVVEQYGPIMRSLPSIVEILTS  183
Bmy  AASATTAAASSSGIGSFFSNLISNPTNMINNIEKVSQVVQSVGPVVEQYGPIMRSLPSIVKILTS  182

BaS  GKSTEENPTEDQTEDLTEKVEVATPPPPQ--KKRKRKKMVIEPVIEKEVREEPVQKIATKPKLYV  256
BaA  GKSTEENPTEDQTEDLTEKVEVATPPPPQ--KKRKRKKW*                           230
BaV  GKSTEENPTEDQTEDLTEKVEVATPPPPQ--KKRKRKKMVIEPVIEKEVREEPVQKIATKPKLYV  248
Bce  GKSTEEN----QTENVTEPVEVESPPPAPQKKKRKRKKIVIEPVIEKELPKEPVQQTATKPKLYV  244
Bmy  GKSTEEN----QTEDATEQVEVATPPPPPP--KRKRKKIVLEPVIEKEVPKEPVQQSATKPKLYV  241
```

FIG. 4. Alignment of the putative VrrA protein sequences. The amino acid sequences of *B. anthracis* strains Sterne (*Ba*S), Ames (*Ba*A), and Vollum (*Ba*V), *B. cereus* (*Bce*), and *B. mycoides* (*Bmy*) were aligned by using the Genetics Computer Group program PILEUP with default parameters. Residues that are different between sequences are in boldface. The underlined residues correspond to the repeating sequence QY-hydrophilic residue-Q. The asterisk indicates the stop codon created by the single-nucleotide deletion at position 1031 in Ames.

the basis of the DNA sequence divergence of a major polymorphic tandem repeat in a 517-amino-acid ORF adjacent to a 65-kDa antigen gene (6). Geographically distinct strains of the *M. tuberculosis* complex have also been differentiated by sequence analysis of a coding region of unknown function, designated the DR cluster, containing direct repeats and an insertion element (8). Also, by examining a repetitive region of the surface protein A gene (*spa*) of *Staphylococcus aureus*, another laboratory was able to distinguish epidemic, rapidly spreading isolates from nonepidemic isolates (5). In general, segments of repetitive DNA appear to be promising regions to search for sequence variability, as these areas stimulate recombinations, deletions, and inversions (22).

We found that the *vrrA* locus most closely resembles a microfilarial sheath protein gene, *shp2*, of the filarial parasites *L. carinii*, *B. malayi*, and *B. pahangi* (Fig. 5). The repetitive sequence QYPQ has been implicated in the covalent cross-linking of multiple subunits of this extracellular sheath protein (12). In addition, microfilarial components, including shp2, are recognized by antibodies from infected rats (3).

We also found that an ORF downstream from *vrrA* (designated *lytB*) has strong homology to the *lytB* genes from *E. coli*

```
VrrA     1  MFPKSPNRQMYPQPGQQPYTPYPIPQLPPMAQKKKGFLAKLFKKHDPTEPFMQMVPPYRQME  62
            ||  |   : |..||.| .: | |. :: :    .  |    . .|..|  . .||  . :
Shp2    37  QFPFLPPFCLPPKQGQAPAPAAPAPPAAAPPPAAAPAPAAPAAPAAPAAPAAPPPVPQQ  98

VrrA    63  GP----------PPMMHQQQQPP----------PQYRQQYQQQYQQQYQQ-QYQQQYPQQYS 103
            ..          ||| :|||||.|          |||.|||.|. .|.|||  ||.|||||||.
Shp2    99  PQQFLQTKTWWCPPMYQQQQQQPSYPPRPPVQYPQYQQQYAQPPKQPYQQMQYPQQYPQQYP 160

VrrA   104  QQYQPYMQHHPEQMIPPQMYESNETRGGAATTAASSSGIGSFFSNLISNPTNMINNIEKV   163
            |||..  |::|:. |:.|       .:|:|. .:|     |::: : .:...::...:..
Shp2   161  QQYPQQQQQQPQRQIAGQQ------PAGGAGLVGS-----SLYGMFGEDSDSLLKSYDHF   209
```

FIG. 5. Comparison of the deduced amino acid sequences of VrrA from *B. anthracis* strain Sterne (VrrA) and the microfilarial sheath protein (Shp2) from *L. carinii* (GenBank accession number Z35443). Identical amino acids are indicated by vertical lines. Dots indicate similarity between amino acids by analysis with the University of Wisconsin Genetics Computer Group BESTFIT program. Dashes indicate gaps to facilitate alignment. The overall amino acid identity between the two sequences was 35%, and the amino acid similarity between the two sequences was 50%.

```
Ban    1 MKIVKISPRGYCYGVVDAMVIARNAALDTSLPRPIYILGMIVHNKHVTDAFEEDGIITLDGP   62
Eco    1 MQILLANPRGFCAGVDRAISIVENAL--AIYGAPIYVRHEVVHNRYVVDSLRERGAIFIE--   58

Ban   63 SRLDILDKIDSGTVIFTAHGVSPEVKQRAQEKGLTTIDATCPDVTQTH-DLIEAKKAEGYHV  123
Eco   59 ---QISEVPDGAILIFSAHGVSQAVRNEAKSRDLTVFDATCPLVTKVHMEVARASRRGEESI  117
```

FIG. 6. Comparison of the deduced amino acid sequences of the amino-terminal protein from *B. anthracis* (*Ban*) and the LytB protein (SWISSPROT accession number P22565) from *E. coli* (*Eco*). Other parameters are as in Fig. 5. The overall amino acid identity between the two sequences was 40%, and the amino acid similarity between the two sequences was 56%.

and *P. fluorescens*, involved in penicillin tolerance and control of the stringent response (Fig. 6). In both *E. coli* and *P. fluorescens*, *lytB* is the fifth of five genes in the *ileS-lsp* operon that show a conservation in gene order (9, 13). This conservation did not carry over to *Bacillus*, as we found no homology of any of the other genes in this operon to the *vrrA* locus or to the upstream, unidentified *orf1*. The low degree of variability between *Bacillus* species and lack of variation between strains of *B. anthracis* indicate that this locus is not a good a candidate for strain differentiation.

By comparing the sequences of polymorphic bands in an AP-PCR fingerprint, we were able to locate a variable region of the chromosome despite lacking prior knowledge of its function. Such a method should have general applicability in the identification of sequence polymorphisms when comparing nearly identical genomes of the size of bacterial chromosomes. The *vrrA* locus appears to be an excellent candidate for the differentiation of isolates of *Bacillus*, especially *B. anthracis*, for which three geographically distinct strains were resolved. Further studies should elucidate whether the similarities of this locus to *shp2* are extended to involvement in host-pathogen interactions.

ACKNOWLEDGMENTS

We thank Patty Davidson for technical support, Dianne Bennett for secretarial assistance, and Pat Spivey for assistance in editing.

This work was supported by the Department of Veterans Affairs.

REFERENCES

1. **Ash, C., and M. D. Collins.** 1992. Comparative analysis of 23S ribosomal RNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing. FEMS Microbiol. Lett. **94:**75–80.
2. **Ash, C., J. A. E. Farrow, M. Dorsch, E. Stackebrandt, and M. D. Collins.** 1991. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. **41:**343–346.
3. **Bardehle, G., M. Hintz, D. Linder, G. Schares, H. H. Schott, S. Stirm, and H. Zahner.** 1992. *Litomosoides carinii*: extraction of the microfilarial sheath components and antigenicity of the sheath fractions. Parasitol. Res. **78:** 501–508.
4. **Dykhuizen, D. E., and L. Green.** 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. **173:**7257–7268.
5. **Frenay, H. M. E., J. P. G. Theelen, L. M. Schouls, C. M. J. E. Vandenbroucke-Grauls, J. Verhoef, W. J. Van Leeuwen, and F. R. Mooi.** 1994. Discrimination of epidemic and nonepidemic methicillin-resistant *Staphylococcus aureus* strains on the basis of protein A gene polymorphism. J. Clin. Microbiol. **32:**846–847.
6. **Frothingham, R.** 1995. Differentiation of strains in *Mycobacterium tuberculosis* complex by DNA sequence polymorphisms, including rapid identification of *M. bovis* BCG. J. Clin. Microbiol. **33:**840–844.
7. **Green, B. D., L. Battisti, T. M. Koehler, C. B. Thorne, and B. E. Ivins.** 1985. Demonstration of a capsule plasmid in *Bacillus anthracis*. Infect. Immun. **49:** 291–297.
8. **Groenen, P. M. A., A. E. Bunschoten, D. van Soolingen, and J. D. A. van Embden.** 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. Mol. Microbiol. **10:**1057–1065.
9. **Gustafson, C. E., S. Kaul, and E. E. Ishiguro.** 1993. Identification of the *Escherichia coli lytB* gene, which is involved in penicillin tolerance and control of the stringent response. J. Bacteriol. **175:**1203–1205.
10. **Harrell, L. J., G. L. Andersen, and K. H. Wilson.** 1995. Genetic variability of *Bacillus anthracis* and related species. J. Clin. Microbiol. **33:**1847–1850.
11. **Henderson, I., C. J. Duggleby, and P. C. B. Turnbull.** 1994. Differentiation of *Bacillus anthracis* from other *Bacillus cereus* group bacteria with the PCR. Int. J. Syst. Bacteriol. **44:**99–105.
12. **Hirzmann, J., A. Schnaufer, M. Hintz, F. Conraths, S. Stirm, H. Zahner, and G. Hobom.** 1995. *Brugia* spp. and *Litomosoides carinii*: identification of a covalently cross-linked microfilarial sheath matrix protein (*shp2*). Mol. Biochem. Parasitol. **70:**95–106.
13. **Isaki, L., R. Beers, and H. C. Wu.** 1990. Nucleotide sequence of the *Pseudomonas fluorescens* signal peptidase II gene (*lsp*) and flanking genes. J. Bacteriol. **172:**6512–6517.
14. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9:**2989–2998.
15. **Ivins, B. E., J. W. Ezzell, J. Jemski, K. W. Hedlund, J. D. Ristroph, and S. H. Leppla.** 1986. Immunization studies with attenuated strains of *Bacillus anthracis*. Infect. Immun. **52:**454–458.
16. **Kaneko, T., R. Nozaki, and K. Aizawa.** 1978. Deoxyribonucleic acid relatedness between *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*. Microbiol. Immunol. **22:**639–641.
17. **Little, S. F., and G. B. Knudson.** 1986. Comparative efficacy of *Bacillus anthracis* live spore vaccine and protective antigen vaccine against anthrax in the guinea pig. Infect. Immun. **52:**509–512.
18. **Mikesell, P., B. E. Ivins, J. D. Ristroph, and T. M. Dreier.** 1983. Evidence of plasmid-mediated toxin production in *Bacillus anthracis*. Infect. Immun. **39:** 371–376.
19. **Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186:**339–346.
20. **Mullis, K.** 1991. The polymerase chain reaction in an anemic mode: how to avoid cold deoxyribonuclear fusion. PCR Methods Appl. **1:**1–4.
21. **Nelson, K., and R. K. Selander.** 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. **174:** 6886–6895.
22. **Petes, T. D., and C. W. Hill.** 1988. Recombination between repeated genes in microorganisms. Annu. Rev. Genet. **22:**147–168.
23. **Phillips, A. P., and K. L. Martin.** 1988. Investigation of spore surface antigens in the genus *Bacillus* by the use of polyclonal antibodies in immunofluorescence tests. J. Bacteriol. **64:**47–55.
24. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
25. **Seki, T., C. Chang, H. Mikami, and Y. Oshima.** 1978. Deoxyribonucleic acid homology of the genus *Bacillus*. Int. J. Syst. Bacteriol. **28:**182–189.
26. **Somerville, H. J., and M. L. Jones.** 1972. DNA competition studies within the *Bacillus cereus* group of bacilli. J. Gen. Microbiol. **73:**257–265.
27. **Sterne, M.** 1939. The use of anthrax vaccines prepared from avirulent (uncapsulated) variants of *Bacillus anthracis*. Onderstepoort J. Vet. Sci. Anim. Indust. **13:**307–312.
28. **Thorne, C. B.** 1993. *Bacillus anthracis*, p. 113–124. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
29. **Turnbull, P. C. B., R. A. Hutson, M. J. Ward, M. N. Jones, C. P. Quinn, N. J. Finnie, C. J. Duggleby, J. M. Kramer, and J. Melling.** 1992. *Bacillus anthracis* but not always anthrax. J. Appl. Bacteriol. **72:**21–28.
30. **Vassart, G., M. Georges, R. Monsieur, H. Brocas, A. S. Lequarre, and D. Christophe.** 1987. A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. Science **235:**683–684.

31. **Welkos, S. L., N. J. Vietri, and P. H. Gibbs.** 1993. Non-toxigenic derivatives of the Ames strain of *Bacillus anthracis* are fully virulent for mice: role of plasmid pX02 and chromosome in strain-dependent virulence. Microb. Pathogen. **14:**381–388.
32. **Welsh, J., and M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. **18:**7213.
33. **Welsh, J., and M. McClelland.** 1993. Characterization of pathogenic microorganisms by genomic fingerprinting using arbitrarily primed PCR, p. 595–602. *In* D. H. Persing, T. F. Smith, F. C. Tenover, and T. J. White (ed.), Diagnostic molecular microbiology: principles and applications. American Society for Microbiology, Washington, D.C.
34. **Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18:**6531–6535.