APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Apr. 1997, p. 1400–1405 Vol. 63, No. 4
0099-2240/97/$04.00+0


# Characterization of the Variable-Number Tandem Repeats in *vrrA* from Different *Bacillus anthracis* Isolates

PAUL J. JACKSON,[1]* ELIZA A. WALTHERS,[1] ABDULLAHI S. KALIF,[2] KARA L. RICHMOND,[1] DEBRA M. ADAIR,[1] KAREN K. HILL,[1] CHERYL R. KUSKE,[1] GARY L. ANDERSEN,[3] KENNETH H. WILSON,[3] MARTIN E. HUGH-JONES,[2] AND PAUL KEIM[4]

*Environmental Molecular Biology Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545[1]; Department of Epidemiology and Community Health, Louisiana State University School of Veterinary Medicine, Baton Rouge, Louisiana 70803-8404[2]; Infectious Diseases Section, Department of Veterans Affairs Medical Center and Division of Infectious Diseases, Duke University Medical Center, Durham, North Carolina 27705[3]; and Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011-5640[4]*

Received 23 December 1996/Accepted 22 January 1997

**PCR analysis of 198 *Bacillus anthracis* isolates revealed a variable region of DNA sequence differing in length among the isolates. Five polymorphisms differed by the presence of two to six copies of the 12-bp tandem repeat 5′-CAATATCAACAA-3′. This variable-number tandem repeat (VNTR) region is located within a larger sequence containing one complete open reading frame that encodes a putative 30-kDa protein. Length variation did not change the reading frame of the encoded protein and only changed the copy number of a 4-amino-acid sequence (QYQQ) from 2 to 6. The structure of the VNTR region suggests that these multiple repeats are generated by recombination or polymerase slippage. Protein structures predicted from the reverse-translated DNA sequence suggest that any structural changes in the encoded protein are confined to the region encoded by the VNTR sequence. Copy number differences in the VNTR region were used to define five different *B. anthracis* alleles. Characterization of 198 isolates revealed allele frequencies of 6.1, 17.7, 59.6, 5.6, and 11.1% sequentially from shorter to longer alleles. The high degree of polymorphism in the VNTR region provides a criterion for assigning isolates to five allelic categories. There is a correlation between categories and geographic distribution. Such molecular markers can be used to monitor the epidemiology of anthrax outbreaks in domestic and native herbivore populations.**

*Bacillus anthracis* is the causative agent of anthrax. Its spores survive for extended periods in soils throughout many regions of the world. While anthrax can be controlled by vaccination, it continues to be endemic in some areas. *B. anthracis* virulence is determined by information encoded on two large plasmids. The larger plasmid, pX01 (174 kbp), contains the *lef*, *pag*, and *cya* genes. These encode the lethal factor, protective antigen, and edema factor proteins, respectively (4, 16, 20). The smaller plasmid, pX02 (95 kbp), contains the *capA*, *capB*, and *capC* genes (9, 19). Together, these genes encode the information needed to produce a poly-D-glutamic acid capsule. Anthrax virulence is due to the production of this capsule in conjunction with previously mentioned proteins (15, 18, 19). Chromosomally encoded components have also been implicated as determinants of disease severity (21).

Studies of the molecular biology, population genetics, evolution, immunology, and epidemiology of *B. anthracis* are hampered by the availability of few definitive molecular markers to distinguish among different isolates or strains (10, 11, 19). Both plasmids are required for virulence, and their presence or absence can be used as an indication of pathogenicity. However, the vast majority of isolates contain both plasmids (reference 12 and our own unpublished results). Therefore, the presence or absence of these plasmids cannot be used alone to categorize isolates. There are no differences in the 16S rDNA sequences between *B. anthracis* and the closely related *Bacillus cereus* and only two differences, a single nucleotide change and a single base insertion in *B. cereus*, between the 23S rDNA sequences from these two species (2, 3). No differences in the 16S or 23S ribosomal sequences have been detected among different *B. anthracis* strains (1). In an earlier report, Andersen et al. (1) described a 12-bp variable-number tandem repeat (VNTR) and two polymorphic types in the *vrrA* gene of *B. anthracis*. Here we report the analysis of 198 *B. anthracis* isolates and the identification of three additional alleles for a total of five polymorphic types. This VNTR varies in repeat number from two to six copies and is found within an open reading frame (ORF) that encodes a putative 30-kDa glutamine-rich protein. The predicted gross structure of the putative protein varies little whether the VNTR is present in two or six copies. Molecular mechanisms that might lead to variations in this region are suggested.

*B. anthracis* isolates assigned to specific categories are often not randomly distributed spatially. Analysis of the VNTR and other, similar variable repeats found within *B. anthracis* DNA thus provides a valuable tool to characterize the epidemiology of anthrax outbreaks in native and domestic animals, rapidly providing information that helps determine the initial source of the disease.

## MATERIALS AND METHODS

***B. anthracis* isolates and media.** Isolates obtained from different sources (see Tables 2 and 3) were incubated on sheep blood agar for 16 h at 37°C. Single colonies from these plates were inoculated into nutrient broth and incubated without shaking at 37°C. The culture was harvested when the concentration reached $10^8$ to $10^9$ CFU/ml.

**DNA isolation and purification.** A 100-ml liquid culture was centrifuged at 1,350 × *g* for 10 min at 4°C to collect the cells. The supernatant was removed and autoclaved, and the pellet was suspended in 8.5 ml of TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM disodium EDTA) by repeated pipetting. The bacterial suspension was frozen quickly in liquid nitrogen and then thawed by heating in a 65°C

* Corresponding author. Mailing address: Environmental Molecular Biology Group, LS-7, Mail Stop M888, Los Alamos National Laboratory, Los Alamos, NM 87545. Phone: (505) 667-2775. Fax: (505) 665-6894. E-mail: jackson@telomere.lanl.gov.

TABLE 1. PCR primers used in this study

| Primer nomenclature[a] | Oligonucleotide sequence |
|---|---|
| EWA-1 | 5′-TATGGTTGGTATTGCTG-3′ |
| EWA-2 | 5′-ATGGTTCCGCCTTATCG-3′ |
| GPR-1 | 5′-CGTAGTTCACGAACTGCATCT-3′ |
| GPR-2 | 5′-ATGATGTATCTAATGCGGCGT-3′ |
| GPR-4 | 5′-ACAACTACCACCGATGGC-3′ |
| GPR-5 | 5′-TTATTTATCATATTAGTTGGATTCG-3′ |

[a] Primers EWA-1 and EWA-2 are those used by Andersen et al. (1).

water bath. The freezing and thawing were repeated twice more. A 10% (wt/vol) solution of sodium dodecyl sulfate (450 μl) and 45 μl of a freshly prepared proteinase K solution (20 mg/ml) were added to the suspension, which was then mixed thoroughly and incubated at 42°C for 1 h. NaCl (1.5 ml of a 5 M solution) was added to the suspension and mixed thoroughly. A cetyltrimethylammonium bromide (CTAB)-NaCl solution (1.4 ml) (10% [wt/vol] solution of hexadecyltrimethylammonium bromide [Sigma Chemical Co., St. Louis, Mo.] in 0.7 M NaCl) was then added, and the suspension was mixed thoroughly and incubated for 10 min at 65°C. An approximately equal volume of chloroform-isoamyl alcohol (24:1, vol/vol) was added and mixed thoroughly with the sample for 10 min. The organic and aqueous phases were separated by centrifugation for 5 min at 3,700 × *g* and 4°C. The upper, aqueous phase was collected and extracted again for 10 min with an equal volume of buffered phenol-chloroform (1:1, vol/vol). The phases were again separated by centrifugation, and the aqueous phase was collected and extracted with chloroform-isoamyl alcohol as before. Following centrifugation to separate the phases, the aqueous phase was collected and the nucleic acids were precipitated by adding 0.6 volume of isopropyl alcohol. The precipitate was collected by centrifugation for 5 min at 3,700 × *g* and 4°C, and the pellet was washed briefly with ice-cold 70% ethyl alcohol to remove residual CTAB. Following centrifugation at 800 × *g* for 5 min at ambient temperature, the ethyl alcohol was removed, the pellet was dried briefly, and the nucleic acids were dissolved in up to 3 ml of sterile TE buffer. Contaminating RNA was removed by adding DNase-free RNase (prepared as described by Sambrook et al. [17]) to a final concentration of 20 μg/ml and incubating the mixture for 1 h at 42°C. The nucleic acid solution was then extracted once with buffered phenol-chloroform for 10 min, the phases were separated by centrifugation as above, and the aqueous phase was collected. This solution was extracted once more with chloroform-isoamyl alcohol, the phases separated by centrifugation, and the aqueous phase was again collected. DNA was precipitated from solution by adding 1/10 volume of 5 M ammonium acetate (pH 5.5) and 2.5 volumes of ice-cold 95% ethyl alcohol. The precipitated DNA was collected by centrifugation for 10 min at 3,700 × *g*. The pellet was dried briefly and dissolved in up to 3 ml of sterile TE buffer depending on the preparation yield. The purity of the sample and concentration of the DNA were determined by measuring the optical density of a diluted sample at 230, 260, and 280 nm.

**PCR amplification.** PCR amplification of different VNTR DNA fragments was accomplished with the EWA-1 and EWA-2 primers (Table 1) described by Andersen et al. (1). PCR mixtures contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (wt/vol) gelatin, 0.2 mM each deoxynucleoside triphosphate, 20 pmol of each primer, and 5 U of AmpliTaq DNA polymerase (Roche Molecular Systems, Inc., Branchburg, N.J.) in a 100-μl total reaction volume. Template DNA was initially denatured by heating at 94°C for 2 min. This was followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and primer extension at 72°C for 1 min. Incubation for 5 min at 72°C followed to complete the extension.

**Analysis of PCR amplicons.** The PCR amplicons were analyzed by electrophoresis through 3.5% (wt/vol) Metaphor agarose gels (FMC BioProducts, Inc., Rockland, Maine) dissolved in 0.5× TBE (45 mM Tris-borate [pH 8.3], 1.0 mM EDTA). Electrophoresis was carried out for 4 h at 80 V. The gels were stained for 30 min with a solution containing 1 μg of ethidium bromide per ml, destained in distilled water, and visualized under UV light. Images were captured electronically with an Eagle Eye II still video system (Strategene, Inc., La Jolla, Calif.).

**DNA sequencing.** The GPR-1 and GPR-2 primers (Table 1) were used in PCR to amplify 1.1-kbp DNA fragments of *B. anthracis* chromosomal DNA from strains containing the different VNTR sequences. These PCR amplicons were purified with QIAquick PCR purification kits (QIAGEN, Inc., Chatsworth, Calif.) and then used as the DNA template in subsequent sequencing reactions. GPR-4 and GPR-5 nested oligonucleotides (Table 1; see Fig. 2) were used in the *Taq* Dye Deoxy Terminator (Applied Biosystems, Foster City, Calif.) cycle-sequencing reactions to sequence a 378- to 426-bp internal DNA fragment of the original amplicons containing the VNTR sequence. Unincorporated dyes were removed from these reactions with CENTRI-SEP spin columns (Princeton Separations, Adelphia, N.J.), and sequencing analysis was performed on an Applied Biosystems 373A stretch DNA sequencer. Both strands of each amplicon were sequenced twice, and only DNA sequences confirmed by comparing both DNA strands are included. DNA sequence files were analyzed with Sequence Navigator software (Applied Biosystems).

**DNA sequence analysis.** Dot plot homology analysis was performed with the MEGALIGN subroutine in the LASERGENE software package (DNASTAR Inc., Madison, Wis.). The analysis used a 30-nucleotide window and the 60 or 75% homology requirements.

**Predictive structure of encoded proteins.** The structure of the putative *vrrA* protein encoded by the ORF containing the VNTR repeats was predicted by using the Chou-Fasman (C-F) (6) and Garnier-Robson (G-R) (8) algorithms for alpha, beta, and turn regions. Coil regions were predicted by using the G-R algorithm. Hydrophilicity and surface probability were calculated by using the algorithms of Kyte and Dolittle (14) and Emini et al. (7), respectively.

**Nucleotide sequence accession numbers.** The GenBank accession numbers of $VNTR_2$, $VNTR_3$, $VNTR_4$, $VNTR_5$ and $VNTR_6$ are U63968, U63967, U63966, U63965 and U63964, respectively.

## RESULTS

**Strains and sources.** A total of 198 different *B. anthracis* isolates from different geographic sources (Tables 2 and 3)

TABLE 2. *B. anthracis* isolates used in this study

| VNTR category[a] | No. (%) of isolates | Source identification[b] |
|---|---|---|
| $(5'\text{-CAATATCAACAA})_2$ | 12 (6.1) | 11966, 2PT, 3TE, ASC-3, ASC-6, ASC-43, K5816, MOZ-1, S35, Vollum (plasmid cured), Vollum (USAMRIID), Vollum 1B |
| $(5'\text{-CAATATCAACAA})_3$ | 35 (17.7) | 28, 29, 33, 58, 83, 109, 93/37, 11949A, 14578, A-6, A7, A8, A9, A24, A25, A46, ASC-27, ASC-31, ASC-130, B7, B286/76, B648/82, BA0018, BA1018, BA1035, BA1086, Bekasi, C93022281, C94275042, FRA-1, K88, MOZ-6, MOZ-7, N12, PAK-1 |
| $(5'\text{-CAATATCAACAA})_4$ | 118 (59.6) | 13, 299, 391, 404, 407, 504, 884, 1257, 1259, 1291, 1960B, 4728, 9660, 14185, 62W-8, 67C-174, 72-235C, 73-237L-1, 74-389C-52, 74-402C-A, 74-412C-8, 78-284, 78-306W, 78-345W, 80-77C-4, 81-173C-2, 91-316C-3, 91-323W-3, 91-334W-2, 91-383C, 91-429C-2, 93/33, 93/54, 93/60A, 93/179, 93/192, 93-194C, 93-195C-8, 93-196C-1, 93-212C-2, 93-213C-1, 93-206C, 93/197, 94-188C-1, 20, 25, 26, 37, 38, 39, 47, 70 (Gruinard), 1/6, 2/6, 3/6, 4/6, 5/6, 6/6, 1FG, A-1, A6, A30, A37, A38, A40, A41, A42, A59, A66, A67, A68, Ames (plasmid cured), ASC-28, ASC-29, ASC-30, ASC-129, B1965/77, B7227/83, BA0003, BA0006, BA0052, BA0078/BA1008, BA1007, BA1015, BA1017, BA1021, BA1024, BA1031, BA1033, BA1040, BA1087, D9106771, D9106853, D9106955, D9107191, Dompu, F-1, IRJA, J611, K25, K1143, K6153, K8215, K9810, MOZ-2, MOZ-3, MOZ-8, NMS, Pak-2, Pangkep, SA0573, SA1189, SA1225, SA1265, SA4045, Sterne/MDPH, UNL24673.94, ZIM69 |
| $(5'\text{-CAATATCAACAA})_5$ | 11 (5.6) | 14187, A22, A23, ASC-25, ASC-26, B6273/93, BA0015, BA1009/BA1023, H-1, H-2, K3 |
| $(5'\text{-CAATATCAACAA})_6$ | 22 (11.1) | 179, 1960A, 93/60B, 11949B, ASC-185, G-25, K1, K2, K4, K5, K6, K7, K8, K10, K14, K16, K18, K20, N11, N15, N54, W-21 |

[a] VNTR categories are designated according to the copy number of the 5′-CAATATCAACAA-3′ repeat.
[b] All isolates were collected and are maintained by Martin Hugh-Jones.

TABLE 3. Geographic origin of the different *B. anthracis* isolates used in this study

| Geographic region | No of isolates of category[a]: | | | | |
|---|---|---|---|---|---|
| | $VNTR_2$ | $VNTR_3$ | $VNTR_4$ | $VNTR_5$ | $VNTR_6$ |
| Argentina | 0 | 0 | 0 | 1 | 0 |
| Asia[b] | 0 | 1 | 6 | 2 | 0 |
| British Isles[c] | 5 | 5 | 7 | 2 | 1 |
| Europe[d] | 2 | 6 | 12 | 2 | 0 |
| Middle East[e] | 0 | 1 | 11 | 0 | 0 |
| North American and Caribbean[f] | 1 | 6 | 46 | 2 | 1 |
| Norway | 0 | 2 | 2 | 1 | 0 |
| Southern Africa[g] | 3 | 15 | 30 | 1 | 20 |
| Unknown | 0 | 0 | 4 | 0 | 0 |
| Total | 11 | 36 | 118 | 11 | 22 |

[a] The number of different *B. anthracis* isolates from different geographic sources are listed. The geographic source is based on the location of the laboratory that provided the isolates unless the isolate is known to have originated from another location.
[b] Indonesia and South Korea.
[c] England, Ireland, and Scotland.
[d] France, Germany, Italy, Slovakia, and Switzerland.
[e] Lebanon, Pakistan and Turkey.
[f] Canada, Jamaica (two cultures), and United States.
[g] Mozambique, Namibia, South Africa, Zambia, and Zimbabwe.

were analyzed to determine the extent of variability in the *vrrA* locus (1). Gel electrophoresis of the resulting PCR amplicons through MetaPhor agarose gels demonstrated the presence of five different fragment length categories (Fig. 1). This represents three new categories in addition to the two identified by Andersen et al. (1). To determine the basis for this size variability, a larger (1.1-kbp) DNA fragment containing the variable region was amplified from the genomic DNA of 25 different *B. anthracis* isolates representing all five size categories. Sequencing reactions with internal primers GPR-4 and GPR-5 (Table 1; Fig. 2) produced sequence information for a 378- to 426-bp portion of each amplicon that included the VNTR sequence (Fig. 2). Sequence comparisons demonstrated that the only differences among the 25 amplicons from the different categories were in the number of copies of a 12-bp repeat, 5′-CAATATCAACAA-3′. This was present two to six times, generating fragments of 378, 390, 402, 414, or 426 bp. There was no other variation in the approximately 400-bp DNA flanking this 12-bp repeated sequence in any of the isolates analyzed.

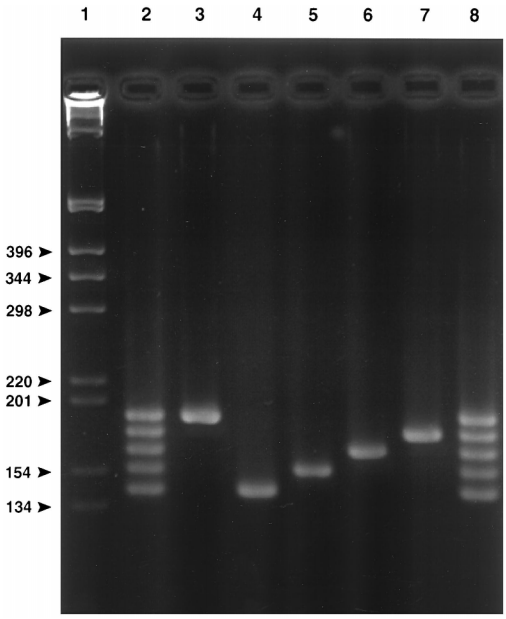


FIG. 1. Gel electrophoresis of amplicons representing the five different VNTR categories. PCR amplicons produced by reactions containing purified *B. anthracis* DNA template from different isolates representing each of the VNTR categories were separated through a 3.5% MetaPhor agarose gel as described in Materials and Methods. Amplification was primed by GPR-4 and GPR-5 (Table 1). Lanes: 1, marker DNA; 2, a mixture of all five VNTR category amplicons; 3, $VNTR_6$ amplicon; 4, $VNTR_2$ amplicon; 5, $VNTR_3$ amplicon; 6, $VNTR_4$ amplicon; 7, $VNTR_5$ amplicon; 8, a mixture of all five VNTR category amplicons. Fragments in lanes 2 and 8 different from adjacent fragments in the same lane by exactly 12 bp.

**Frequency of the different VNTR types.** Fragment length analysis of 198 different *B. anthracis* isolates demonstrated that the $(5'\text{-CAATATCAACAA-}3')_4$ ($VNTR_4$) was the most frequently found VNTR length (Table 2). A total of 118 (59.6%) of the isolates tested, including the previously reported Sterne and Ames strains (1), fell into this category; 35 isolates (17.7%) fell into the $VNTR_3$ category, and 22 (11.1%) were $VNTR_6$. The $VNTR_2$ category represented 12 (6.1%) of the isolates, while the $VNTR_5$ category contained 11 (5.6%). The $VNTR_2$ category included six different isolates thought to be Vollum strains (the three listed Vollum isolates plus ASC-3, ASC-6, and ASC-43). Vollum strains have previously been assigned to the $VNTR_2$ category (1). Category frequencies are influenced by the number of related isolates collected from different sites. While the 12 members of the $VNTR_2$ category were isolated from five countries on three different continents, 15 (42.9%) of the $VNTR_3$ category were isolated from South Africa and surrounding countries (Table 3). Of the 118 isolates in the $VNTR_4$ category, 46 (39%) originated in North America. Of the 22 $VNTR_6$ isolates, 20 (91%) originated in South Africa or surrounding countries, and 1 of the other 2 isolates, obtained from Great Britain, was obtained from a kudu, an animal native to southern Africa. Information about the host of the other isolate, obtained from a U.S. laboratory, was not provided (Table 3 gives the geographic source of isolates). Our collection is biased with North American isolates. Further collections from other geographic locations might therefore decrease the relative dominance of $VNTR_4$.

**Dot plot homology.** The internal-repeat structure of the *vrrA* variable region is revealed by dot plot homology analysis (Fig. 3). It is apparent that about 100 nucleotides of this region have a unique and strong direct-repeat structure that is lacking in adjacent areas. When dot plot analysis was performed under 75% homology requirements, only the perfect 12-nucleotide repeats were observed as parallel lines to the diagonal line (Fig. 3). These are the same nucleotide repeats that are responsible for the variable nature of this region as their copy number changes among strains. When the same analysis is performed at a slightly reduced homology level (60%), additional direct-repeat structure with a periodicity of 3 nucleotides is observed. This internal homology is not perfect, primarily because of the trinucleotide repeat CAA present within the 12 nucleotides. No significant indirect repeat structure was revealed by this analysis (Fig. 3).

**Predicted structures of encoded proteins.** Amino acid sequence and protein structural characteristics were predicted by using different computer algorithms to understand the effect of nucleotide variation on potential protein structure (Fig. 4). The *vrrA* ORF protein is predicted to contain 248 to 264 amino acids ($VNTR_2$ to $VNTR_6$) and have a mass of 28.6 to 30.8 kDa. It has at least threefold more polar and charged amino acid residues than hydrophobic residues. A comparison of the

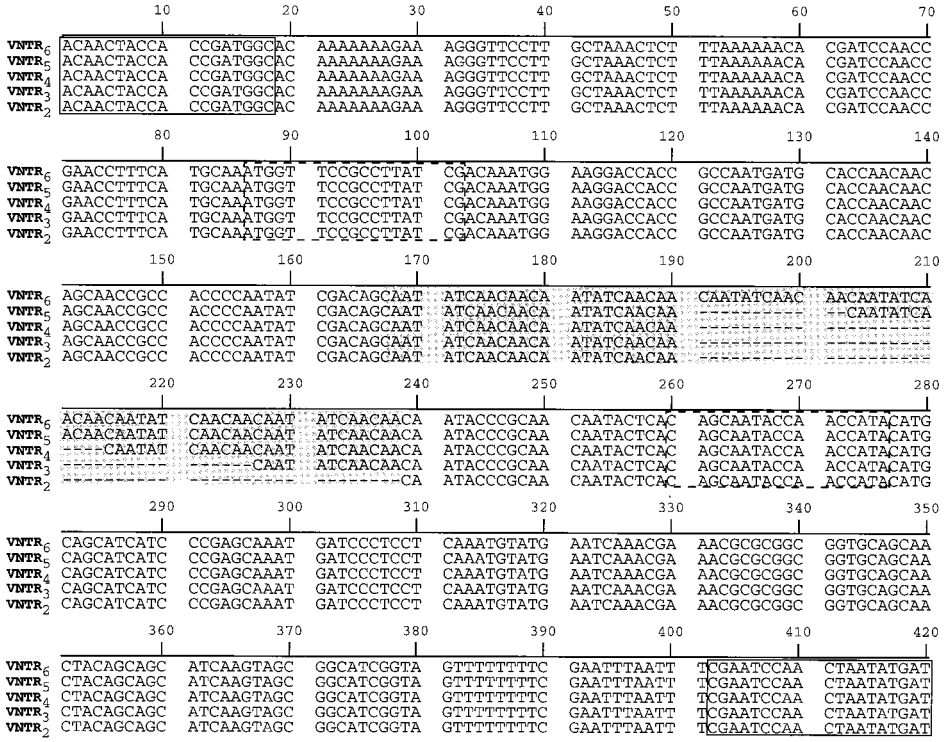

FIG. 2. DNA sequences of the five different VNTR categories. Twenty-five different amplicons representing the five different VNTR categories were sequenced. There was no sequence variability within any of the categories. The shaded area delineates the presence of two to six copies of the sequence 5′-CAATATCAACAA-3′. The solid boxes on either end of the sequences define GPR-4 and GPR-5, the sequences used to prime synthesis of the PCR amplicon. The dashed boxes identify the original EWA-2 and EWA-1 priming sites that generated the fragments shown in Fig. 1. The last 6 bp of each sequence, not included in the figure, are 5′-AAATAA-3′.

$VNTR_2$- to the $VNTR_6$-encoded protein revealed very few differences in predicted structure. Alpha, beta, coil, and turn regions were identical, and the changes in hydrophilicity and surface probability differences were restricted to the variable region. These differences were a contraction of the hydrophilic region and the associated high probability of surface exposure between residues 80 and 104 as the number of repeats decreased.

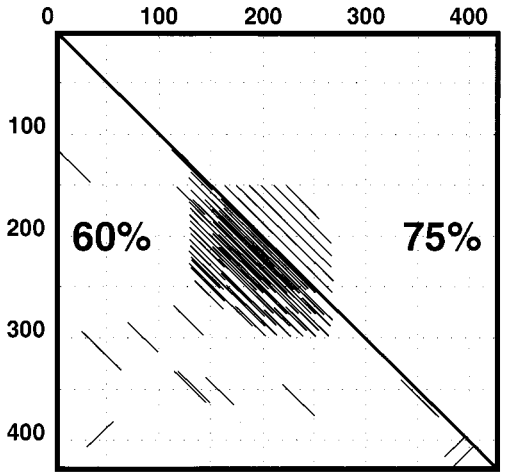


FIG. 3. Dot plot homology analysis of the variable region. Approximately 430 nt from the $VNTR_6$ *vrrA* ORF was examined by dot plot homology analysis. Analysis was performed with a 30-nucleotide window and either 60% or 75% homology requirement, as indicated.

## DISCUSSION

Analysis of 198 different *B. anthracis* isolates demonstrated the presence of a variable DNA sequence that allows classification of the different isolates into categories based on the number of 5′-CAATATCAACAA-3′ tandem repeats present in the VNTR region of *vrrA* (Table 2). $VNTR_4$ and $VNTR_2$ have been reported previously (1) and contain the Sterne and Ames strains and Vollum strain, respectively. We report three additional VNTR categories and the complete monomorphism of the VNTR flanking regions. By contrast, comparison to closely related *Bacillus* species demonstrated many single base substitutions throughout the ORF, including other deletions or additions that affected the overall length (data not shown). Even the closely related *B. cereus* varied at 2.4% of the bases. Moreover, several additional small deletions in the *B. cereus* sequence relative to *B. anthracis* produced a significantly shorter amplicon (1; our unpublished results).

The number of repeats present and the frequency with which they occur within the population are dependent on the extent of sampling and how many times the same isolate or its relatives are collected from a region. However, this frequency, as well as the sequence of the VNTR repeat and flanking DNA sequences, suggests a mechanism by which one VNTR category can be derived from another by polymerase slippage and unequal crossing over.

All cultures grown from single colonies contained only one

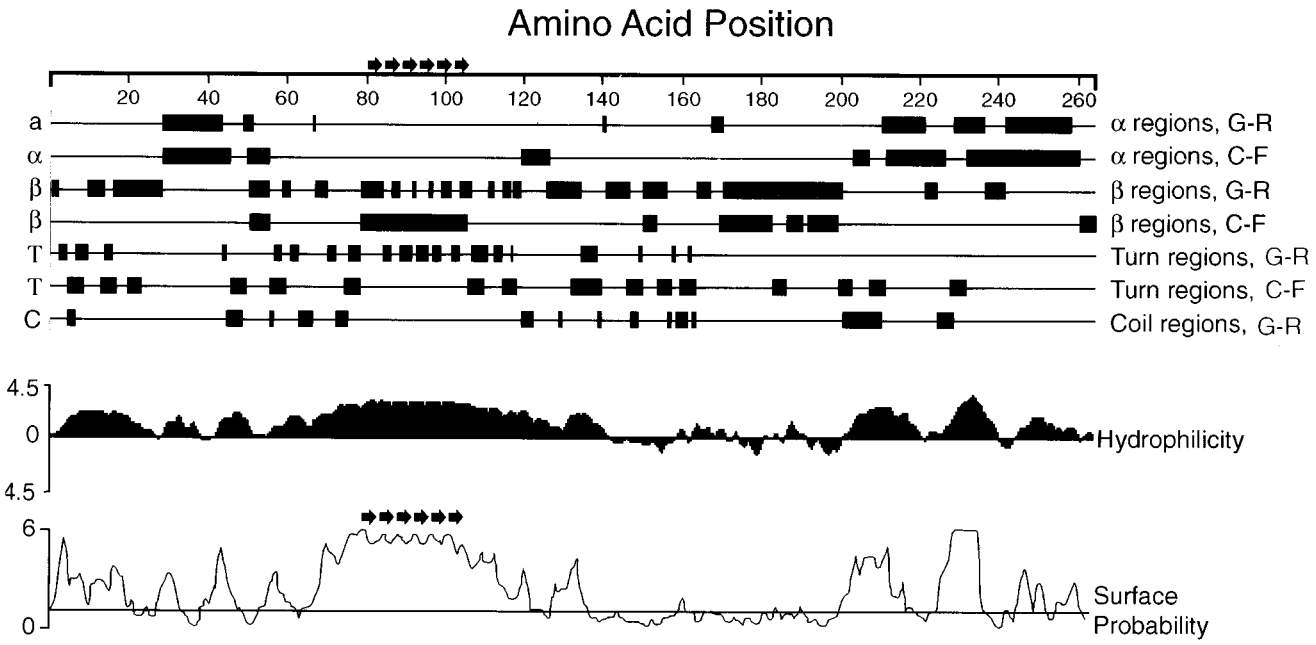


FIG. 4. Protein structural predictions for a putative $VNTR_6$ *vrrA* protein. Protein structural characteristics for the *vrrA* ORF protein with six tandem repeats ($VNTR_6$) were predicted from the C-F and G-R algorithms for alpha, beta, and turn regions. Coil regions were predicted from the G-R algorithm. Hydrophilicity and surface probability were calculated from the algorithms of Kyte and Dolittle and Emini et al., respectively. The 4-amino-acid tandem repeat is indicated by the six tandemly repeated arrows. A similar analysis of a $VNTR_2$ protein was identical except for a contraction of the hydrophilic region associated with the tandem repeat.

VNTR category. Moreover, multiple colonies derived from a single infection source all contained bacteria representative of only one VNTR category. This suggests that most infections arise from a single VNTR type and that changes from one category to another as a result of genetic instability at the locus during the culture or infection process are rare. Such sequence stability suggests the utility of the VNTR region as a marker to determine the source of an infection or outbreak.

There was no host specificity among the different categories. The $VNTR_4$ category was isolated from 17 different mammalian species. The hosts of the other categories represented a similarly wide range of species.

The geographic distribution of the different categories suggests the origins of different anthrax outbreaks. Of 22 members of the $VNTR_6$ category, 20 originated in South Africa or surrounding countries, suggesting a common origin of an ongoing anthrax outbreak in this locale. Forty-six isolates representing the $VNTR_4$ category were isolated from sources in North America. These isolates may therefore also be representative of a single anthrax outbreak and source. The results presented are consistent with other evidence suggesting that this and other variable genetic loci can be used to more thoroughly characterize the epidemiology of this disease. This is more thoroughly discussed by Keim et al. (12).

Andersen et al. (1) showed that the VNTR region is found within an ORF designated the *vrrA* gene, which encodes a putative 30-kDa protein. The variation in the *vrrA* ORF is due to a direct-repeat structure encoding amino acid residues 80 through 104. This region contains the highly homologous 12-nucleotide repeats and trinucleotide repeats of lesser homology (Fig. 3). The highly structured region is in contrast to most sequences and to those immediately flanking the variable region. Polymerase slippage and unequal crossing over are two mechanisms that have been widely proposed to generate similar variable regions. Both mechanisms could account for the results of this study, and distinguishing between them is problematic. Both of these mechanisms are thought to be high-frequency events that would allow the VNTR regions to evolve more rapidly than nonstructured genomic regions. When direct-repeat sequence variation is found in genes, phenotypic consequences have also been observed (5, 13, 22).

The *vrrA* allele frequencies predict a very high diversity index or polymorphic information content for this locus. This is calculated by summing the squared allele frequencies and subtracting from 1. By using our data, a value of nearly 0.6 is obtained. As a diagnostic tool, the VNTR regions should then discriminate among 60% of future *B. anthracis* isolates. This is in great contrast to the previous monomorphism found in this organism (2, 3, 10, 11). Not only is the great diversity in this locus of diagnostic importance, but also it argues for a function or adaptive role in anthrax biology. The maintenance of great diversity within a protein-coding region is highly suggestive that this variation is functional. Antigenic shifting has been previously suggested as maintaining this apparently balanced polymorphism (1), but other selective forces may be responsible.

Structural characteristics of the putative *vrrA* ORF protein with six tandem repeats ($VNTR_6$) and the same protein with two tandem repeats ($VNTR_2$) were predicted by using different algorithms. Based on this comparative analysis, the predicted amino acid changes will affect only surface residues. While this may very well affect protein function and, perhaps, antigenic sites, it is unlikely that the overall structure will be changed significantly. The 12-bp repeat encodes two to six copies of the amino acid sequence Gln-Tyr-Gln-Gln (QYQQ). Therefore, the number of repeats present significantly influences the glutamine content of the encoded region. The *vrrA* gene encodes a putative protein with some homology to the microfilarial sheath protein shp2 of the parasitic worm *Litomosoides carinii* (1). The evolutionary conservation of this

ORF among different *Bacillus* species (1; our unpublished results) suggests that the putative gene encodes a functional protein. However, this function is unknown in *B. anthracis.*

In summary, analysis of a region of the *B. anthracis vrrA* gene in 198 different isolates revealed five categories of this putative gene that differ only in the number of 12-bp DNA sequence repeats that are present. It is not known whether this variability affects any physiological functions of the organism. It does not appear to change the pathogenicity of the organism, since most of the 198 isolates were collected from infected animals or their remains. However, there are few molecular markers that allow characterization of this highly monomorphic bacterial pathogen at the strain level. When used in combination with markers that are specific for *B. anthracis*, the *vrrA* gene is an excellent locus to identify genetic variability among different strains and isolates.

## ACKNOWLEDGMENTS

This research was partially funded by and was conducted under the auspices of the U.S. Department of Energy. P.K. was partly supported by a sabbatical salary grant from Northern Arizona University and the Cowden Endowment for Food Microbiology.

It is a pleasure to acknowledge the generosity of all those who freely provided their *B. anthracis* isolates and accompanying documentation. All isolates were collected and are maintained by Martin Hugh-Jones, Department of Epidemiology and Community Health, Louisiana State University School of Veterinary Medicine, Baton Rouge, LA 70803-8404.

## REFERENCES

1. **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.
2. **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.
3. **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.
4. **Bragg, T. S., and D. L. Robertson.** 1989. Nucleotide sequence and analysis of the lethal factor gene (*lef*) from *Bacillus anthracis*. Gene **81:**45–54.
5. **Cancel, G., N. Abbas, G. Stevanin, A. Durr, H. Chneiweiss, C. Neri, C. Duyckaerts, C. Penet, H. M. Cann, Y. Agid, and A. Brice.** 1995. Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia-3 Machado-Joseph disease locus. Am. J. Hum. Genet. **57:**809–816.
6. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47:**45–148.
7. **Emini, E. A., J. Hughes, D. Perlow, and J. Boger.** 1985. Induction of a hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol. **55:**836–839.
8. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of a simple method for predicting the secondary structure of globular proteins. J. Mol. Biol. **120:**97–120.
9. **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.
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. **Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, and P. Jackson.** 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J. Bacteriol. **179:**818–824.
13. **Kunzler, P., K. Matsuo, and W. Schaffner.** 1995. Pathological, physiological, and evolutionary aspects of short unstable DNA repeats in the human genome. Biol. Chem. Hoppe-Seyler **376:**201–211.
14. **Kyte, J., and R. F. Dolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
15. **Mikesell, P., B. E. Ivins, J. D. Ristroph, and T. M. Dreier.** 1983. Evidence for plasmid-mediated toxin production in *Bacillus anthracis*. Infect. Immun. **39:** 371–379.
16. **Robertson, D. L., M. T. Tippetts, and S. H. Leppla.** 1988. Nucleotide sequence of the *Bacillus anthracis* edema factor gene (*cya*): a calmodulin-dependent adenylate cyclase. Gene **73:**363–371.
17. **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.
18. **Uchida, I., K. Hashimoto, and N. Terakado.** 1986. Virulence and immunogenicity in experimental animals of *Bacillus anthracis* strains harbouring or lacking 110 MDa and 60 MDa plasmids. J. Gen. Microbiol. **132:**557–559.
19. **Uchida, I., T. Sekizaki, K. Hashimoto, and N. Terakado.** 1985. Association of the encapsulation of *Bacillus anthracis* with a 60 megadalton plasmid. J. Gen. Microbiol. **131:**363–367.
20. **Welkos, S. L., J. R. Lowe, F. Eden-McCutchan, M. Vodkin, S. H. Leppla, and J. J. Schmidt.** 1988. Sequence and analysis of the DNA encoding protective antigen of *Bacillus anthracis*. Gene **69:**287–300.
21. **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. Pathog. **14:**381–388.
22. **Willems, P. J.** 1994. Dynamic mutations hit double figures. Nat. Genet. **8:**213–215.