LA-UR-99- -154

Title: GENETIC COMPARISON OF B. *ANTHRACIS* AND ITS CLOSE RELATIVES USING AFLP AND PCR ANALYSIS

Author(s): Paul J. Jackson, Environmental Molecular Biology Group and Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545
Karen K. Hill Environmental Molecular Biology Group and Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545
Miriam T. Laker, Environmental Molecular Biology Group and Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545
Lawrence O. Ticknor, Environmental Molecular Biology Group and Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545
Paul S. Keim, Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 87545

Submitted to: Proceedings of the Third Intercontinal Conference on Anthrax, Plymouth , England

Los Alamos
NATIONAL LABORATORY

Form No. 836 R5
ST 2629 10/91

DISCLAIMER

Genetic comparison of *B. anthracis* and its close relatives using AFLP and PCR analysis.

Authors:

Paul J. Jackson[1], Karen K. Hill[1], Miriam T. Laker[1], Lawrence O. Ticknor[1] and Paul Keim[2]

[1]Environmental Molecular Biology Group and Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA

[2]Department of Biological Sciences, Northern Arizona University, Flagstaff AZ 86011-5640

RUNNING HEAD: Using molecular signatures to identify pathogens.

SUMMARY

Amplified Fragment length Polymorphism (AFLP) analysis allows a rapid, relatively simple analysis of a large portion of a microbial genome, providing information about the species and its phylogenetic relationship to other microbes (Vos, *et al.*, 1995). The method simply surveys the genome for length and sequence polymorphisms. The pattern identified can be used for comparison to the genomes of other species. Unlike other methods, it does not rely on analysis of a single genetic locus that may bias the interpretation of results and it does not require any prior knowledge of the targeted organism. Moreover, a standard set of reagents can be applied to any species without using species-specific information or molecular probes. We are using AFLP's to rapidly identify different bacterial species. A comparison of AFLP profiles generated from a large battery of *B. anthracis* strains shows very little variability among different isolates (Keim, *et al.*, 1997). By contrast, there is a significant difference between AFLP profiles generated for any *B. anthracis* strain and even the most closely related Bacillus species. Sufficient variability is apparent among all known microbial species to allow phylogenetic analysis based on large numbers of genetically unlinked loci. These striking differences among AFLP profiles allow unambiguous identification of previously identified species and phylogenetic placement of newly characterized isolates relative to known species based on a large number of independent genetic loci. Data generated thus far show that the method provides phylogenetic analyses that are consistent with other widely accepted phylogenetic methods. However, AFLP analysis provides a more detailed analysis of the targets and samples a much larger portion of the genome. Consequently, it provides an inexpensive, rapid means of characterizing microbial isolates to further differentiate among strains and closely related microbial species. Such information cannot be rapidly generated by other means.

AFLP sample analysis quickly generates a very large amount of molecular information about microbial genomes. However, this information cannot be analyzed rapidly using manual methods. We are developing a large archive of electronic AFLP signatures that is being used to identify isolates collected from medical, veterinary, forensic and environmental samples. We are also developing the computational packages necessary to rapidly and unambiguously analyze the AFLP profiles and conduct a phylogenetic comparison of these data relative to information already in our database. We will use this archive and the associated algorithms to determine the species identity of previously uncharacterized isolates and place them phylogenetically relative to other microbes based on their AFLP signatures. This study provides significant new information about microbes with environmental, veterinary and medical significance. This information can be used in further studies to understand the relationships among these species and the factors that distinguish them from one another. It should also allow identification of unique factors that contribute to important microbial traits including pathogenicity and virulence.

We are also using AFLP data to identify, isolate and sequence DNA fragments that are unique to particular microbial species and strains. The fragment patterns and sequence information provide insights into the complexity and organization of bacterial genomes relative to one another. They also provide the information necessary for development of species-specific PCR primers that can be used to interrogate complex samples for the presence of *B. anthracis*, other microbial pathogens or their remnants.

INTRODUCTION

Amplified Fragment Length Polymorphisms (AFLP's) provide a low cost, relatively rapid and thorough survey of a microbial genome. Purified genomic DNA is digested with two restriction endonucleases and linkers containing a DNA sequence that can be used to prime DNA amplification are ligated to the fragments. Fragment sub-populations are systematically amplified using a combination of labeled PCR primers that selectively amplify subsets containing a manageable number of the total fragments produced. This results in a survey of the microbial genome including any plasmids that might be present. Analysis of samples harboring any plasmids can be compared to those that do not, to identify which DNA fragments were amplified from the plasmids. This method rapidly screens the entire genome for length polymorphisms and evaluates approximately 0.4% of the genome for changes in DNA sequence. Analysis of the

amplified DNA fragments is completed by electrophoresis of the DNA through polyacrylamide DNA sequencing gels. Analysis on an automated DNA sequencer easily resolves the DNA fragments to the single base level (Figure 1).

The data generated can be analyzed manually. However, this is very labor intensive and subject to operator error. The very large numbers of AFLP profiles that can be generated on a single DNA sequencing gel require extensive manual analysis. The time required for such analyses is problematic. Moreover, application of this technology to answer questions about different microbial species requires that previously generated profiles can be compared to newly produced data. Presently there is no efficient, accurate method of conducting this comparison. A rapid method of accurately scoring the profiles is needed. This must be accompanied by mathematically and statistically valid automated methods to phylogenetically analyze the results and compare them to previously obtained archived information. In this manner, AFLP signatures for different species and strains can be used to identify previously unknown species by comparing their profiles to those that are already well studied. This is analogous to comparing an unknown DNA sequence to archived sequences but allows comparison of a much larger percentage of the microbial genome.

Such profiles will also identify those regions of the genome that are conserved among different closely related species. Such regions may contain sequences for different "housekeeping" genes. In groups of closely related pathogens they may also contain information that will identify genes specifically required for virulence, pathogenicity or other traits specific to the pathogens. AFLP profiles will also allow identification and isolation of species-specific DNA fragments. Those fragments unique for a particular microbe will provide valuable diagnostic tools to identify the organism or its remnants in complex environmental or forensic samples.

MOLECULAR DIVERSITY AMONG DIFFERENT BACILLUS SPECIES.

A comparison of a portion of AFLP profiles from five different closely related Bacillus species is shown in Figure 1. Traditional hybridization methods that compare microbial genomes suggest that *B. anthracis* and *B. cereus* share 98% DNA sequence homology. However, comparison of AFLP profiles demonstrates that these genomes show marked organizational differences. A comparison of all of the AFLP fragments produced using sixteen different primer combinations showed that only 45% of the fragments found in *B. anthracis* are shared by *B. cereus*. Comparison of *B. anthracis* strains that contain pX01 and pX02, the two plasmids required for virulence, to strains that do not harbor these plasmids readily demonstrates which AFLP fragments represent plasmid-specific sequences (Keim, *et al.,* 1997). Thus, it is a simple matter to differentiate between virulent and non-virulent *B. anthracis* isolates. Subtraction of these fragments from the profiles allows a direct comparison of genomic DNA among the different species. Phylogenetic analysis using the remaining fragments provides insights into the relationships of these different species to one another (Figure 2).

B. anthracis is genetically very monomorphic. AFLP analysis of *B. anthracis* isolates consequently show very few AFLP pattern differences among different isolates. In contrast, there are significant fragment pattern differences among different isolates of other Bacillus species (Figure 3). This may be due to the manner in which the original isolates were identified. All of the *B. anthracis* isolates were identified because they caused a specific disease in an animal or human host. In contrast, other Bacillus isolates were identified using much less stringent criteria. It is possible that there are many *B. anthracis* strains that vary significantly from one another but show no pathogenic properties. Consequently, they are not identified by any method that screens for pathogenic strains. A comparison of AFLP patterns generated for Bacillus species to those generated for other microbial genera shows that the variability found among different isolates of most Bacillus species is shared by other microbes. In some cases, it appears that specific isolates of one species may be more closely related to members of another species than they are to members of their own species. This is most pronounced when comparing *B. cereus* and *B. thuringiensis* isolates. Such differences may result from the use of less stringent criteria to classify an isolate. Many of the phenotypic aspects of *B. cereus* are shared by *B. thuringiensis*. However, the primary criterion for labeling an isolate as *B. thuringiensis* is its ability to produce a crystal toxin. It may therefore be problematic to compare phylogenetic analyses based on phenotypic traits to

those based on molecular analysis. However, molecular analysis including 16S rDNA sequence comparisons, comparisons of specific genes shared among different isolates and closely related species, RFLP and AFLP analyses normally provide consistent results. This suggests that methods based purely on phenotypic traits may be, to some extent, influenced by unknown selection biases and artificial categorization based on observation of poorly defined traits or misinterpretation of apparently similar traits that may not share the same genetic basis.

IDENTIFICATION AND CHARACTERIZATION OF SPECIES-SPECIFIC DNA TARGETS.

The Polymerase Chain Reaction (PCR) is increasingly being used for diagnostic and forensic purposes. However, the technique is limited by the specificity of the PCR primers that are used to amplify targeted sequences. Often unwarranted assumptions are made about the specificity of the DNA sequences targeted. It was once thought that targeting selected virulence genes would provide an unambiguous identification of a microbial pathogen within a complex sample. However, as more DNA sequence and experimental data become available, it is increasingly clear that many sequences thought to be target-specific are shared by other, sometimes unrelated species. This problem is compounded because portions of particular virulence genes are shared among pathogenic and non-pathogenic species. Sometimes they encode a gene with similar functions. However, often the implications of the shared sequences are not understood. It therefore becomes critical to understand whether the targeted DNA sequences are, indeed, species- or strain-specific.

A comparison of AFLP profiles from a battery of closely related microbial species quickly identifies those fragments that are shared among many of the isolates and those that are unique to individual species. Further comparisons to more distantly related microbes will identify batteries of fragments that are characteristic of a group of genetically related species. Isolation of such fragments and development of molecular probes from their sequences will provide tools to determine the genetic basis of shared characteristics among close microbial relatives. Those fragments shared by a large number of species may represent portions of genes that encode traits required by all members of the group. These are likely to include genes encoding "housekeeping" functions, DNA replication and other essential features. Those shared among a small group of close relatives may encode factors that are the essence of the group. Such primers may allow molecular studies of these shared traits. If such studies focus on groups of closely related pathogens, information may be gained about virulence factors and other traits that influence the success of these microbes.

AFLP fragments unique to a particular microbial species or strain are good candidates to contain DNA sequences that are unique to one or a very small number of species. Such sequences are extremely valuable for developing species- or strain-specific PCR primers. Isolation of these DNA fragments from gels, selective amplification of the fragment above the unlabeled fragment background, and sequencing of the amplified product produces a source of DNA sequence for PCR primer design. Thus far, this method has been used to identify, isolate and sequence strain-specific DNA fragments from *B. anthracis* (Keim, *et al. 1999*) and *B. thuringiensis* (our own unpublished results). Such primers must still be tested experimentally to determine their specificity. However, the number of PCR primers that fail the experimental tests due to lack of species specificity is significantly reduced using this method.

AFLP analysis also provides insights into the differences in genome organization among different species. It is quite remarkable that there are such large differences in genome organization among microbes that share such high DNA sequence homology as measured by DNA hybridization methods. This suggests that microbial genomes are very dynamic and that there must be some factor(s) that, at some point, effectively isolate two similar species with a common ancestor from one another.

AUTOMATED AFLP ANALYSIS.

AFLP analysis of microbial DNA generates fragments whose number is influenced by the enzymes used to digest the DNA, the A/T content of the genome, secondary modification of the DNA (i.e., methylation of restriction sites) and the genome size. Thus, digestion with enzymes

that recognize specific short DNA sequences followed by simple scoring of the number of fragments provides information about the microbial genome size and A/T content. AFLP analysis of *B. anthracis* using sixteen different "+1" AFLP primers (Keim, *et al.*, 1997) generates 388 different fragments that must be compared from one isolate to another. Comparison of a limited number of microbial isolates is straightforward. However, this soon becomes problematic when the number of isolates increases significantly. Moreover, in the absence of any other prior indicators, it is difficult to compare the profile produced by an unknown isolate to a "control" profile analyzed on a different polyacrylamide gel. It is therefore necessary to compile an archive of AFLP profiles for a large number of microbes and to develop mathematical algorithms that will compare archived profiles to those generated for new samples. Similar algorithms for comparison of phylogenetic information are already available (Swofford. 1998), but must be integrated with the initial fragment analysis software.

AFLP profiles generated using different DNA sources of the same isolate, different extraction techniques and different reagents are very similar (Figure 4). Manual scoring of such profiles is labor intensive but straightforward and reproducible. However, electronic AFLP profile scoring is very challenging. There are several factors that influence this. Peak positions may vary slightly from one gel to the next. The currently available software that "calls" fragment size can be confused by samples that are probably identical but appear to be slightly different based on minor differences in fragment migration. For example, a DNA fragment for one isolate may be called as being only 0.2 bp different in length from a fragment from another isolate. However, the software may "call" the smaller fragment as a lower whole integer and the larger fragment as a higher whole integer. Such fragments would be scored as different during the phylogenetic analysis while they are, in fact, the same.

More serious scoring errors occur when the amount of DNA loaded on the analysis gel varies sufficiently. A majority of the "small" peaks on the gel represent real AFLP fragments and are reproducible. However, experimental results suggest that any fragments scored below 50 arbitrary units introduce the majority of apparent variability into the analyses. Consequently, peaks with an area below 50 are excluded from the analysis. If sufficiently different amounts of reaction product are analyzed on the gel, then different numbers of marginal peaks will be present. Subsequent phylogenetic analysis of the results will introduce artificial variability into the analysis. This issue is addressed experimentally by analysis of three separate reaction aliquots. Variability among results from reactions known to be identical provides an indication of the scoring and analysis variability. Mathematical methods of identifying and compensating for minor apparent fragment shifts on the gels are also used to reduce artificial variability.

AFLP analysis provides a rapid, relatively inexpensive technique for phylogenetic analysis of microbial genomes. Unlike many methods that conduct analyses on single genes that may be subjected to selective pressure by unknown factors, AFLP analysis is based on a survey of a significant portion of the entire microbial genome. Results of such analyses are consistent with other molecular and phenotypic methods (Harrell, *et al.*, 1995; Henderson, *et al.*, 1995). However, they provide significantly more information and sample multiple genetic loci. Therefore they may be more statistically accurate than other contemporary methods. Moreover, unlike other phylogenetic methods, AFLP rapidly differentiates among all species so far tested and among different strains of even *B. anthracis*, the most genetically monomorphic species so far characterized. Epidemiological and forensic analyses are greatly enhanced by information beyond the species level. Therefore AFLP analysis is directly relevant to advanced studies in these fields.

We are now generating an archive of AFLP profiles for a large battery of microorganisms. However, there are still technical problems associated with comparing a large number of different AFLP profiles. We are therefore developing computational methods of handling and comparing large amounts of AFLP data to rapidly generate a phylogenetic analysis of samples that will assist in identifying previously uncharacterized or unknown samples.

REFERENCES

- Harrell LJ, Andersen GL, Wilson KH (1995) Genetic variability of *Bacillus anthracis* and related species. *Clin Microbiol* **33**:1847-1850

- Keim P, Kalif A, Schupp J, Hill K, Travis SE, Richmond K, Adair DM, Hugh-Jones M, Kuske CR, Jackson P (1997) Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. *J Bacteriol* **179**:818-824

- Henderson I, Dongzheng, Y, Turnbull PCB (1995) Differentiation of *Bacillus anthracis* and other 'Bacillus cereus group' bacteria using IS231-derived sequences. *FEMS Microbiol Lett* **128**:113-118.

- Swofford, DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

- Vos, P, Hogers, R., Bleeker, M., Reijans, M., vander Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**:4407-4424.

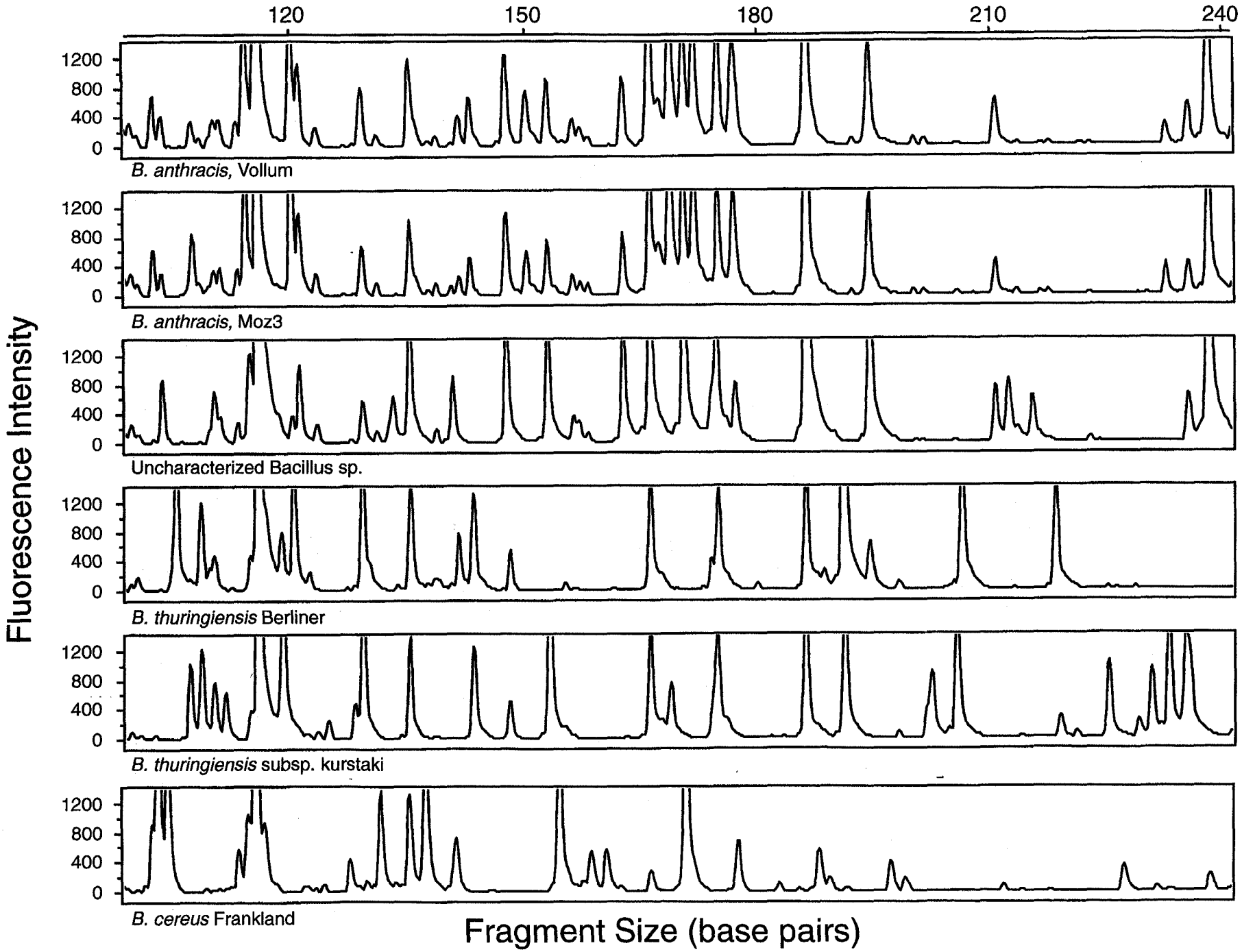
Figure Legends.

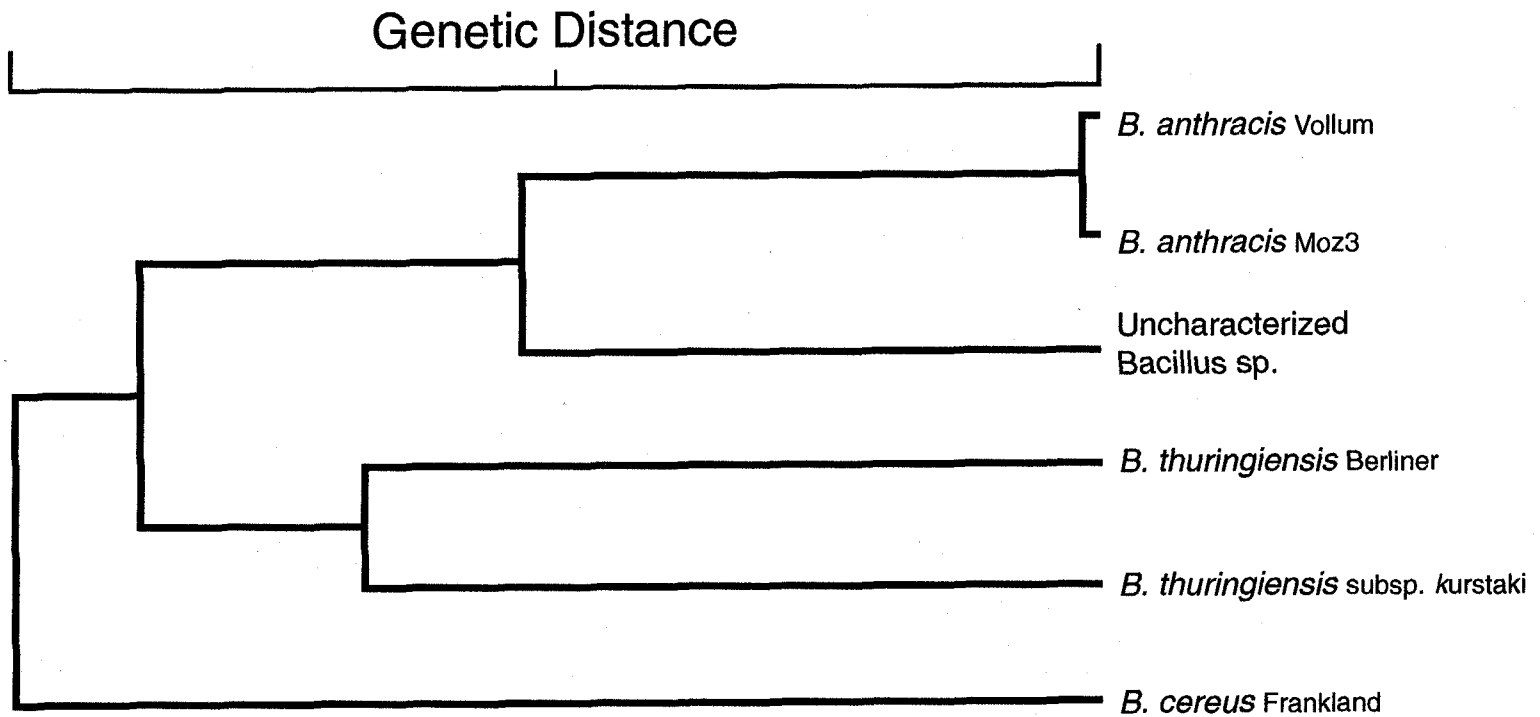
Figure 1. Polyacrylamide gel electrophoresis of DNA fragments produced by AFLP analysis of total microbial genomic DNA using C/G +1 primers. Electronic profiles of DNA fragments were scanned using an Applied Biosystems 377 Automated DNA sequencer and the profiles were analyzed using Applied Biosystems GeneScan software package version 2.0.2. These profiles can be compared to identify similarities and differences among different AFLP profiles. Only the profiles of fragments between 100 and 240 bp are shown.

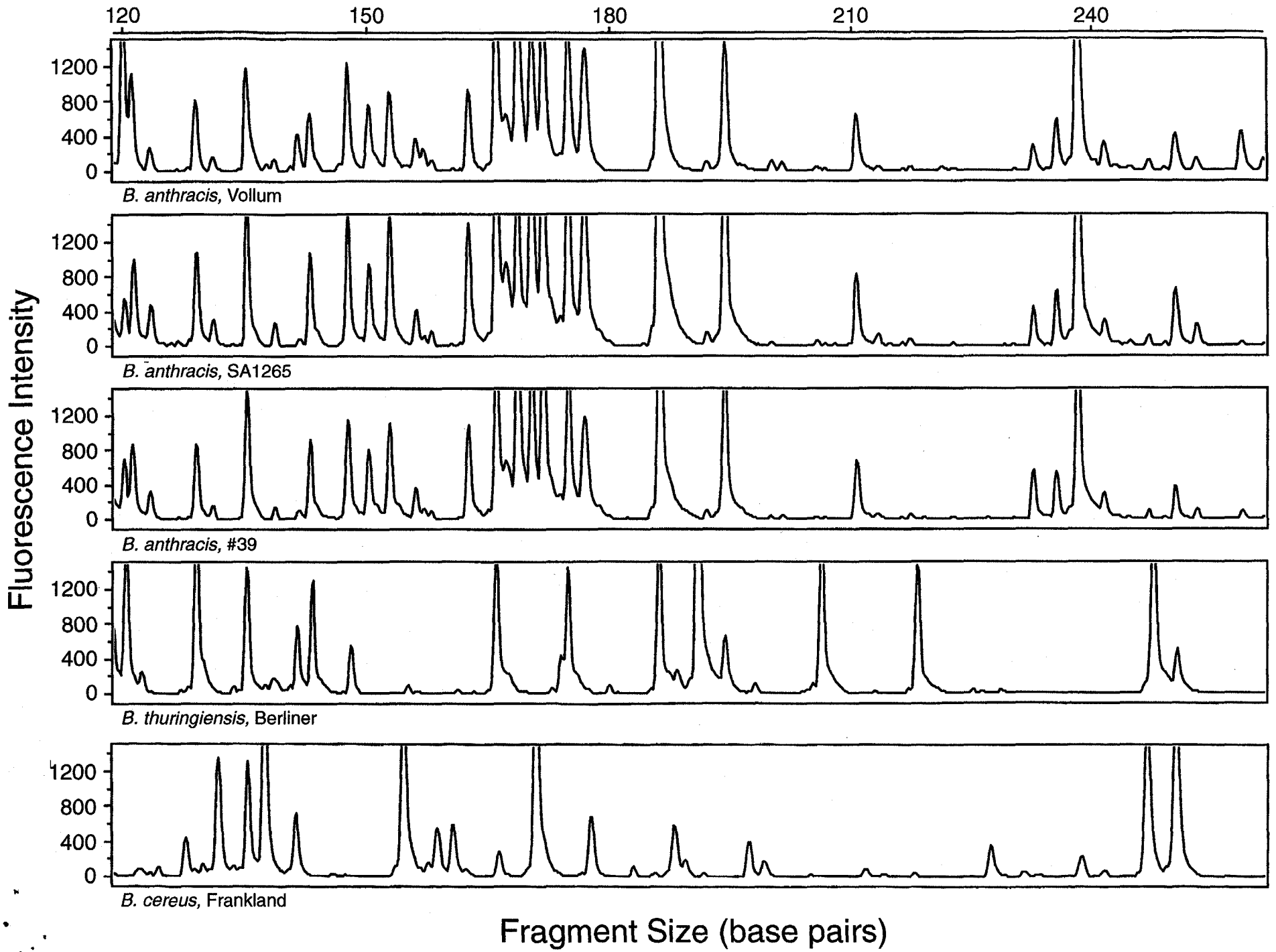
Figure 2. Phylogenetic analysis of *B. anthracis* and related Bacillus species using AFLP fragment data. AFLP markers were used as genetic characters to determine the relationships among different Bacillus species. AFLP fragments were analyzed cladistically by using the UPGMA cluster analysis algorithm of the Phylogeny Analysis Using Parsimony (PAUP) version 4 software package (Swofford, 1998). Analysis was based on 388 AFLP fragments generated from four different "+1" primer sets.

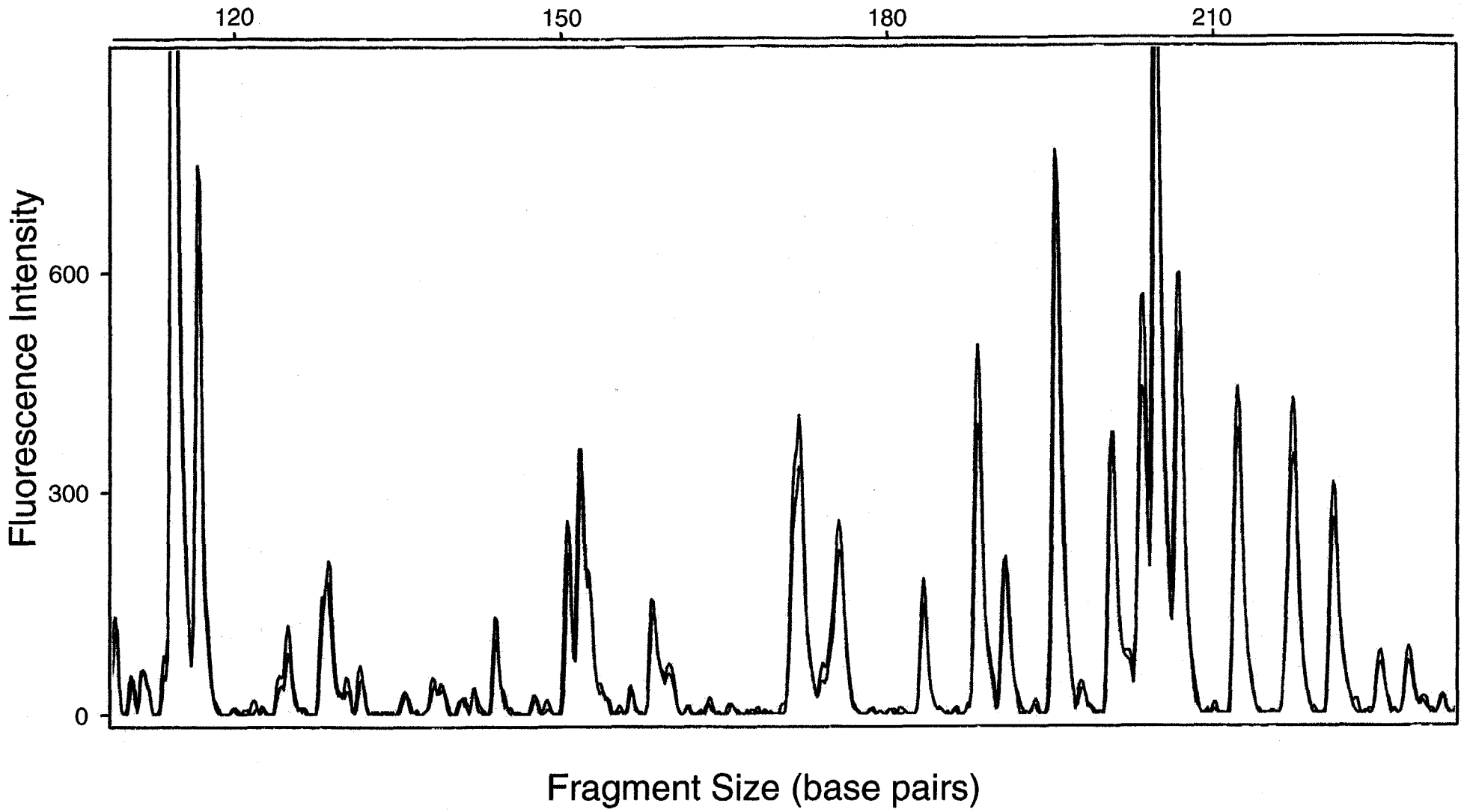
Figure 3. Polyacrylamide gel electrophoresis of DNA fragments produced by AFLP analysis of genomic DNA from different *B. anthracis* strains and closely related Bacillus species. Electronic profiles of DNA fragments were scanned using an Applied Biosystems 377 Automated DNA sequencer and the profiles were analyzed using Applied Biosystems GeneScan software package version 2.0.2. The top three panels show portions of AFLP profiles for three different *B. anthracis* strains while the bottom two panels show profiles for a *B. thuringiensis* strain and a *B. cereus* strain respectively.

Figure 4. AFLP profiles for two different preparations of *B. globigii* DNA. *B. globigii* isolates from two different sources were grown at different times and the DNA from each culture was isolated using different DNA extraction methods. Each sample was independently subjected to AFLP analysis and AFLP profiles were superimposed electronically to demonstrate that AFLP analysis provides very similar analysis results for the two preparations. Red, *B. globigii* obtained from Dugway Proving Grounds, Utah; blue, *B. subtilis* sub-species Niger obtained from ATCC.


120
150
180
210
240
1200
800
400
0
B. anthracis, Vollum
B. anthracis, Moz3
Uncharacterized Bacillus sp.
B. thuringiensis Berliner
B. thuringiensis subsp. kurstaki
B. cereus Frankland
Fluorescence Intensity
Fragment Size (base pairs)


Genetic Distance
B. anthracis Vollum
B. anthracis Moz3
Uncharacterized
Bacillus sp.
B. thuringiensis Berliner
B. thuringiensis subsp. kurstaki
B. cereus Frankland


120
150
180
210
240
1200
800
400
0
B. anthracis, Vollum
B. anthracis, SA1265
B. anthracis, #39
B. thuringiensis, Berliner
B. cereus, Frankland
Fluorescence Intensity
Fragment Size (base pairs)


120
150
180
210
Fluorescence Intensity
600
300
0
Fragment Size (base pairs)