Fluorescent Amplified Fragment Length Polymorphism Analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* Soil Isolates

LAWRENCE O. TICKNOR,[1] ANNE-BRIT KOLSTØ,[2] KAREN K. HILL,[3] PAUL KEIM,[4]
MIRIAM T. LAKER,[3] MELINDA TONKS,[3] AND PAUL J. JACKSON[3]*

Decision Applications Division[1] and Bioscience Division,[3] Los Alamos National Laboratory, Los Alamos, New Mexico 87545; Institute of Pharmacy, University of Oslo, Oslo, Norway[2]; and Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011-5640[4]

We examined 154 Norwegian *B. cereus* and *B. thuringiensis* soil isolates (collected from five different locations), 8 *B. cereus* and 2 *B. thuringiensis* reference strains, and 2 *Bacillus anthracis* strains by using fluorescent amplified fragment length polymorphism (AFLP). We employed a novel fragment identification approach based on a hierarchical agglomerative clustering routine that identifies fragments in an automated fashion. No method is free of error, and we identified the major sources so that experiments can be designed to minimize its effect. Phylogenetic analysis of the fluorescent AFLP results reveals five genetic groups in these group 1 bacilli. The ATCC reference strains were restricted to two of the genetic groups, clearly not representative of the diversity in these bacteria. Both *B. anthracis* strains analyzed were closely related and affiliated with a *B. cereus* milk isolate (ATCC 4342) and a *B. cereus* human pathogenic strain (periodontitis). Across the entire study, pathogenic strains, including *B. anthracis*, were more closely related to one another than to the environmental isolates. Eight strains representing the five distinct phylogenetic clusters were further analyzed by comparison of their 16S rRNA gene sequences to confirm the phylogenetic status of these groups. This analysis was consistent with the AFLP analysis, although of much lower resolution. The innovation of automated genotype analysis by using a replicated and statistical approach to fragment identification will allow very large sample analyses in the future.

Bacillus cereus, *Bacillus thuringiensis*, and *Bacillus anthracis* are gram-positive, rod-shaped, spore-forming bacteria referred to as group 1 bacilli (4). *B. cereus* and *B. thuringiensis* inhabit diverse soil habitats and have many economically important representatives. *B. cereus* can be pathogenic and may cause food poisoning, eye infections, and periodontal disease in humans (10). Spore and crystal toxin preparations from *B. thuringiensis* are used as commercial insecticides (22). The plasmid-encoded crystal proteins produced by *B. thuringiensis* are lethal to coleopteran, dipteran, or lepidopteran insect pests (8, 22). *B. anthracis*, a highly virulent mammalian pathogen, is the causal agent of anthrax. *B. anthracis* spores survive for long periods in the environment, but little evidence exists for saprophytic growth in the soil. Either its obligate pathogenic nature or a recent genetic bottleneck may be partially responsible for it high level of molecular monomorphism (15, 16, 17).

Differentiation of these three closely related species was historically based on biochemical assays, flagellum serotype, and the presence of insecticidal protein toxin crystals (8, 10, 22). Analysis of the 16S rRNA gene showed that these three species share almost identical sequences within and adjacent to this structural RNA gene (2, 3). Carlson et al. (6) examined 24 *B. cereus* and 12 *B. thuringiensis* isolates using pulsed-field electrophoresis (PFGE) and multienzyme electrophoresis (MEE). A high degree of genetic variability was observed within and between the two species. Because neither PFGE

nor MEE grouped the strains by recognized species designations, it was suggested that *B. cereus* and *B. thuringiensis* be considered one species. This is consistent with suggestions by others that *B. thuringiensis* is simply *B. cereus* with crystal-protein encoding plasmids (5). Likewise, Helgason et al. (9) examined 154 *B. cereus* and *B. thuringiensis* environmental isolates by using serotyping and 13 enzymes in MEE assays. These were diverse soil isolates collected from five geographic regions in Norway ranging from coastal to Arctic. This study demonstrated great diversity, showing polymorphism at all 13 enzyme loci, 112 electrophoretic types (ET), and 28 different serotypes. In contrast to the diverse nature of *B. cereus* and *B. thuringiensis*, *B. anthracis* strains showed little molecular diversity. Seventy-eight strains collected worldwide were analyzed by amplified fragment length polymorphism (AFLP), and very few polymorphic fragments were observed (15). However, in this same study AFLP revealed a high degree of polymorphism among three *B. cereus* and *B. thuringiensis* isolates. This is consistent with the Norway MEE and serotype results.

We report here the results of fluorescent AFLP analysis of 154 Norwegian *Bacillus* soil isolates, 2 *B. anthracis* strains (Sterne and Vollum), and 8 *B. cereus* and 2 *B. thuringiensis* reference strains obtained from the American Type Culture Collection (ATCC, Manassas, Va.). To compare and analyze such a large number of fluorescent AFLP profiles, it was necessary to develop and test computational methods for automated AFLP data collection and analysis. While software that assists with such analyses is currently available (see references 23 and 24 for examples), it does not consider the experimental variability of AFLP analysis nor function without human intervention. We have identified sources of experimental variability

---

in the fluorescent AFLP technique by triplicate analysis of the same samples on three different polyacrylamide gels on an ABI377 automated DNA sequencer. The extent of variability among AFLP runs results from inaccurate DNA fragment length determinations or differences in peak heights that are inherent in the PCR or arise from analyzing differing amounts of PCR product. Software that accurately considers and addresses this was developed. We compared the AFLP data analysis with analyses conducted by using MEE and 16S rDNA sequence data.

Our results show extensive genetic diversity among different *B. cereus* and *B. thuringiensis* environmental isolates that is not reflected in the reference strains for these two species. Isolates clustered into five major groups that each contained mixtures of these two species.

MATERIALS AND METHODS

Bacterial isolates. *Bacillus* Norwegian soil isolates were collected, and *B. cereus* reference strains 10987, 4342, and 6464 were provided as previously described (6, 9, 11). *B. cereus* strains 11778, 14579, 31293, 43881, and 53522 and *B. thuringiensis* strains 10792 and 33679 were purchased from the ATCC. DNA from the *B. anthracis* Vollum and Sterne strains was kindly provided by Martin Hugh-Jones and Kimothy Smith of the Department of Epidemiology and Community Health, School of Veterinary Medicine, Louisiana State University, Baton Rouge, La.

DNA isolation and purification. A 5-ml portion of nutrient broth was inoculated with a single *Bacillus* isolate colony, and cultures were incubated overnight with shaking at 28°C. Bacterial cells were collected by centrifugation at 1,000 × *g* for 15 min, and bacterial pellets were subjected to three freeze-thaw cycles. DNA was isolated from disrupted cells by using a QIAamp tissue kit (catalog no. 29306; Qiagen, Inc., Valencia, Calif.) according to the protocol provided by the manufacturer. The DNA quantity and quality were determined by electrophoresis through a 1.0% agarose gel dissolved in a solution containing 10 mM Tris borate (pH 8.3) and 1 mM EDTA. Electrophoresis was for 1 h at 80 V. Gels were stained for 20 min with a solution containing 1 µg of ethidium bromide (Sigma Chemical Co., St. Louis, Mo.) per ml, destained in distilled water, and then visualized and photographed under UV light.

AFLP analysis of DNA samples. AFLP analysis was accomplished as previously described (15, 27) but was adapted for fluorescent detection as follows. DNA (100 ng) was digested with *Eco*RI and *Mse*I, and the resulting fragments were ligated to double-stranded adapters (5′-CTCGTAGACTGCGTACC-3′ plus 3′-CTGACGCATGGTTAA-5′ and 5′-GACGATGAGTCCTGAG-3′ plus 3′-TACTCAGGACTCAT-5′, respectively). The digested and ligated DNA was then amplified by PCR (30 cycles) by using the *Eco*RI and *Mse*I +0/+0 primers 5′-GTAGACTGCGTACCAATTC-3′ and 5′-GACGATGAGTCCTGAGT AA-3′ in a final volume of 50 µl. The PCR cycling conditions included the cycling profile of 94°C for 30 s, 60°C for 30 s, and 72°C for 60 s repeated for 30 cycles. The +0/+0 PCR product was analyzed by agarose gel electrophoresis. A total of 3 µl was used in subsequent selective amplifications with the +1/+1 primer combination of EcoRI-C (5′-GTAGACTGCGTACCAATTCC-3′) and MseI-G (5′-GACGATGAGTCCTGAGTAAG-3′). Selective amplifications were performed in 20-µl reactions using a cycling profile of 94°C for 30 s, 65°C for 30 s, and 72°C for 1 min for 1 cycle and then lowering the annealing temperature by 1°C each cycle to 56°C (9 cycles), followed by an additional 26 cycles at a 56°C annealing temperature. The EcoRI-C primer was labeled with the fluorescent dye FAM (6-carboxyfluorescein). The resulting AFLP products (0.5 to 1.0 µl) were mixed with 0.75 µl of a solution containing DNA size standards (Genescan-500; Applied Biosystems Inc., Foster City, Calif.; and MapMarker-400; BioVentures, Inc., Murfreesburo, Tenn.), both labeled with TAMARA (*N,N,N,N*-tetramethyl-6-carboxyrhodamine). After a 2-min heat denaturation at 90°C, the reactions were loaded onto a 5% Long Ranger DNA sequencing gel (BioWhittaker Molecular Applications, Rockland, Maine) and visualized on an ABI377 automated fluorescent sequencer (Applied Biosystems, Inc.). Each reaction was analyzed on three different sequencing gels, each time loaded adjacent to different samples. Genescan analysis software (Applied Biosystems, Inc.) was used to determine the length of the sample fragments by comparison to the DNA size standards included. Sample fragments were compared to 24 different DNA standards ranging from 100 to 500 bp in length. Sample fragments of between

100 and 500 bp and with fluorescence above 50 arbitrary units in all three runs on the ABI sequencer were used in the analysis.

AFLP data analysis. AFLP data consisted of the presence or absence of peaks on an electropherogram and the heights of those peaks. The peak location is analogous to fragment size, and the peak height is analogous to the number of fragments of a given size. To compare two or more electropherograms and assign a similarity or distance measure to this comparison, the electropherograms must be aligned to determine which peaks are common and which peaks are different. To determine which peaks are common, a clustering algorithm was used. First, all peak locations for all samples being compared were combined into one vector of data. A hierarchical agglomerative clustering routine using group averages created the clusters (14). A decision rule was added to this clustering routine so that the number of clusters chosen depended on the number of electropherograms being compared and a maximum value for the range of a cluster (a value for what could be considered "the same"). Peaks within a cluster were assigned the average peak value for that cluster so that all peaks in the set being compared that were considered the same have the same peak value.

Because there were triplicate data from three lanes for each sample, the data from the triplicates for a single sample must be combined to create a single record of information for the sample. The set of peaks that was used to represent the combined replicates contained all peaks that were present in each member of the triplicate. This set was called the fingerprint and was used as the description of a sample when similarities among samples were determined. The height of each peak in the fingerprint was the average height of this peak in the triplicates. A matrix that combines all samples and all unique peaks from the sample set being compared was generated. Each row of the matrix corresponded to one sample and contained ones and zeroes showing the presence or absence of a given peak for that sample.

Similarities among samples were described by the Jaccard coefficient for distances. The 40 tallest peaks for each sample were used to calculate the Jaccard coefficient among samples. Dendrograms were produced by using the similarity matrix of Jaccard coefficients and the unweighted pair-group average method (UPGMA) (21).

Data quality was assured by using the triplicates and a DNA control sample that undergoes AFLP analysis and is loaded onto every AFLP analysis gel to provide a standard for comparisons among different data sets. Triplicates and DNA controls from each data set were compared before any sample data were considered for analysis. Triplicates that did not cluster for obvious experimental reasons were removed from the analysis, and the samples were again subjected to AFLP analysis. Similarly, dendrograms were produced by using data generated from control DNA included in all the data sets being compared. If UPGMA analysis did not produce dendrograms for which the DNA controls clustered within the expected uncertainty range, the entire data set tied to that control, one gel in most cases, was discarded and the AFLP analysis was repeated.

MEE data analysis. Thirteen enzymes and their electrophoretic types were previously determined for each *Bacillus* sample collected in Norway (6, 9). The distance among different samples was defined as the fraction of time the electropheretic types were different for the 13 enzymes. A dendrogram based on these distances was constructed by using the same software used for analysis of the AFLP data to minimize differences caused by analysis with different software packages.

The distance matrices for the MEE data and the AFLP data were compared by using the Mantel randomization technique (25) in which the statistic of interest is the sum of the cross-products of the two distance matrices. Ten-thousand randomization trials were performed where in the AFLP distance matrix was resampled and the result was compared to the MEE distance matrix. A "maximum" Mantel value was derived by using the given distances and assuming the dendrograms would be identical if the ranks of the distance matrices were identical.

Principal component analysis of the AFLP data. Principal components for the AFLP fingerprint data were derived (13). The first and second/and the first and third principal components were plotted with characters relating to the five major clusters seen on the UPGMA dendrograms. All statistical data manipulations were done by using codes developed in S-Plus (S-Plus 2000; MathSoft, Seattle, Wash.).

16S ribosomal DNA sequencing. To confirm taxonomic identity and calibrate our results to the 16S rRNA gene, eight *Bacillus* isolates representing the major branches of the AFLP phylogenetic tree were selected for 16S rRNA gene sequencing. The primers srDNA-PA (5′-AGAGTTTGATCCTGGCTCAG-3′) and 16S-R3 (5′-GGAGGTGATCCAACCGC-3′) were used to amplify the full length of the gene. The purified PCR template was then sequenced by using these primers and the internal primers 533F (5′-CCAGCMGCCGCGGTAA-3′), P3MOD (5′-ATTAGATACCCTDGTAGTCC-3′), P3MODrc (5′-GGACT
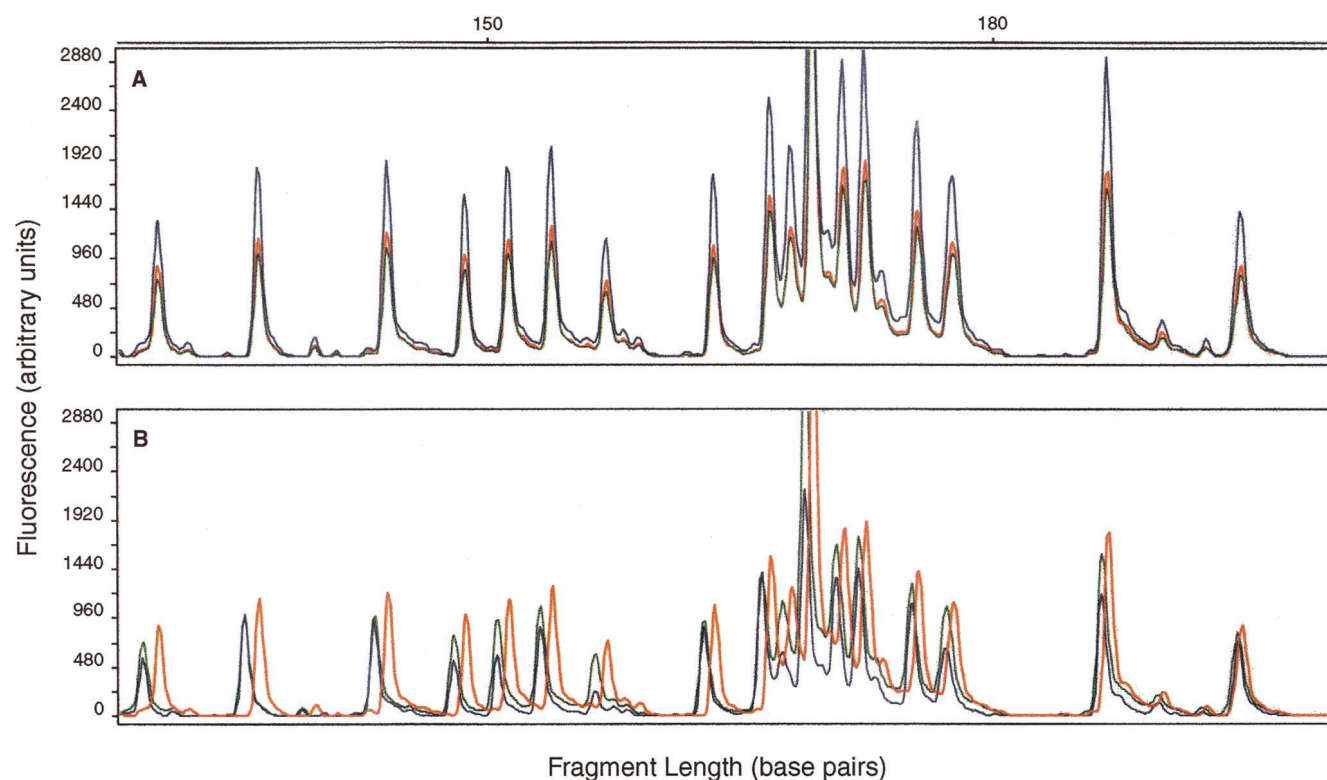
FIG. 1. Triplicate AFLP profiles of *B. anthracis* Vollum. AFLP analysis of *B. anthracis* Vollum was conducted using EcoRI-C and MseI-G primers. Results were analyzed in three different lanes of a single polyacrylamide gel (A) or on three different polyacrylamide gels (B). Analysis was accomplished on an ABI377 automated DNA sequencer. Only a portion of each profile from 100 to 200 bp is shown. The three different colors represent the three different lanes used for analysis.

ACHAGGGTATCTAAT-3′), and BAC281 (5′-CTCAGGTCGGCTACGCAT C-3′). Full-length sequence data were also obtained from the *B. cereus* reference strains 11778, 14579, 31293, 43881, and 53522, the *B. thuringiensis* reference strains 10792 and 33679, and the *B. anthracis* Vollum and Sterne strains to provide the sequence information necessary for comparison. The sequences for the different strains have been submitted to GenBank under accession numbers AF290545 through AF290562 for the 16S rDNA sequences of isolates ATCC 10792, ATCC 11778, ATCC 14579, ATCC 31293, ATCC 33679, ATCC 43881, ATCC 53522, *B. anthracis* Sterne, *B. anthracis* Vollum, AH521, AH527, AH533, AH540, AH648, AH665, AH678, and AH526, respectively.

RESULTS

Development of procedures for automated fluorescent AFLP analysis. The fluorescent AFLP analysis generated over 40 fragments of between 100 and 500 bp for each sample by using just one set of "+1" primers. This study required development of automated analysis methods to handle 166 samples analyzed in triplicate. Development of a fully automated system required an understanding of the sources of variability within the analysis process. AFLP data are generated as peak height and peak location. Each peak represents one or more DNA fragments, and the location or relative migration of the fragment through the gel is directly related to the DNA fragment size. The peak locations of fragments within the polyacrylamide gel were compared to the migration of internal DNA size standards included in each lane of the same gel to determine the length of the DNA fragment represented by each peak. Slight variations in the migration of fragments within the gels and comparisons to different molecular mass standards generate variations in apparent fragment lengths. Fluorescence peak height is influenced by the amount of product in the peak. This is influenced by changes in relative product concentrations within different AFLP reactions, by the number of fragments with the same molecular mass, and by the amount of sample loaded into a lane of the gel. Within a sample, the fragments that were scored to generate the AFLP profile range in fluorescence from 51 to 3,000 arbitrary units. Minor differences in the volumes loaded onto a gel can result in variation in peak height (fluorescence) among replicates. If these differences change the number of fragments with fluorescence values between 51 and 3,000 U, this will influence scoring of the AFLP profile.

Experimental variability in fluorescent AFLP analysis. To determine the extent of fluorescent AFLP variability and its sources, an experiment was designed using a single control sample (*B. anthracis* strain Vollum). AFLP reactions for this sample were analyzed by electrophoresis through different lanes of a single polyacrylamide gel, and the same sample was analyzed on three different polyacrylamide gels. Figure 1A presents a comparison of three different analyses of *B. anthracis* Vollum on the same polyacrylamide gel versus Fig. 1B, where in three profiles of the same samples are compared on different gels. This visual representation clearly illustrates the higher variability across different gels. For these data, obtained using analysis techniques presented here, we found that variation within a gel ranged from 2 to 6%, while variation across gels ranged from 8 to 14%. In both cases, fragment identity was
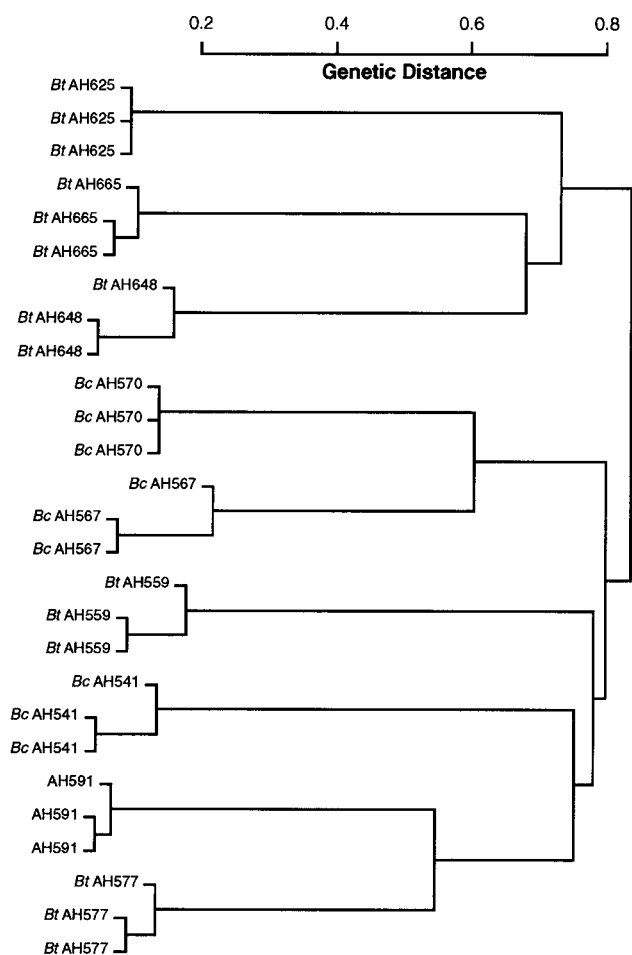
FIG. 2. Phylogenetic analysis of AFLP triplicate samples from different Norway isolates. To demonstrate the reproducibility of sample analysis, AFLP samples for nine different Norway isolates were analyzed on three different polyacrylamide gels. The resulting profiles were then used as the basis for a phylogenetic analysis. AFLP fragments were analyzed, and dendrograms were generated as described in Materials and Methods. The results demonstrate that variability within a sample is far less than AFLP profile differences among different samples.

dependent upon reproducible migration of the DNA size standards in each lane. Figure 1B clearly shows that migration differences are greater among gels than within a single gel.

AFLP products were analyzed in triplicate on polyacrylamide gels. Each sample was analyzed on three different polyacrylamide gels, adjacent to different samples in each gel. Each gel also included AFLP fragments from a common sample (*B. anthracis* Vollum) to allow standardization among the gels. All replicates were analyzed and plotted on a single dendrogram to determine the extent of variability among the different replicates (Fig. 2). This allows identification and removal of unusual replicates, thereby increasing confidence that the data reflects actual differences or similarities among the samples. Common peaks in the AFLP profiles from triplicate lanes were then used to generate DNA signatures or fingerprints for each sample. These composite fingerprints were used to produce dendrograms from the similarity matrix of Jaccard coefficients and the UPGMA method (21).

Norwegian *Bacillus* isolate analysis. AFLP data were collected in a digital format from 166 different *Bacillus* samples, including 154 strains isolated from five different sites in Norway (Table 1) (6, 9). These were used as the basis for a phylogenetic analysis of these different samples. A dendrogram based on the AFLP data is shown in Fig. 3. The results of the AFLP analysis showed great genetic diversity among the *Bacillus* isolates with five discrete groups identified. The diversity does not appear to be geographically based since almost all members of the bottom four groups (Fig. 3) were collected in Moss, Norway. However, these samples from Moss were collected from diverse environments, including leaf tissue, grass compost, a strawberry and a cabbage field, and a beech grove. The remaining samples were collected from other sites in Norway and do not cluster based on geographic origin either. Six of the *B. cereus* reference strains and the two *B. thuringiensis* reference isolates included in the analysis clustered together within one branch of the tree (the bottom branch in Fig. 3). This cluster includes the type strains from *B. cereus* (ATCC 14579) and *B. thuringiensis* (ATCC 10792). Only 16 of the Norwegian soil isolates cluster with these reference strains. The remaining Norwegian isolates populate the entire phylogenetic tree, suggesting that these two species are much more polymorphic than represented by their type strains. Similar results were obtained for *B. cereus* and *B. thuringiensis* isolates from other sources (P. J. Jackson, L. O. Ticknor, and K. K. Hill, unpublished data), suggesting that this is not peculiar to the Norwegian isolates. In contrast to the lack of species grouping by *B. cereus* and *B. thuringiensis* isolates, the two *B. anthracis* isolates (Vollum and Sterne) are very closely related to each other. This is consistent with the previously described monomorphic nature of this species (15, 16, 17). *B. anthracis* strains cluster more closely to several Norway soil isolates than to the *B. cereus* and *B. thuringiensis* type strains. The two reference strains most closely related to *B. anthracis* are *B. cereus* ATCC 10987 and ATCC 4342. The latter is also closely related to *B. cereus* isolates that cause periodontal disease in humans (10). Other *Bacillus* isolates implicated in food poisoning and serious infections also cluster close to *B. anthracis* (Jackson et al., unpublished).

Comparison of AFLP and MEE. Helgason et al. (6, 9) conducted an analysis of these same *Bacillus* samples by using serotyping and MEE as the basis for generating the phylogenetic characters. Thirteen different enzymes representing 112 different ET were used for the analysis. MEE analysis also revealed significant genetic diversity among the different isolates and isolates having the same ET often had different serotypes, suggesting even more complexity than what was revealed by the MEE data. Analysis of the MEE data using the same phylogenetic analysis package used to analyze the AFLP data also divided the isolates into distinct groups (Fig. 4). Of 151 isolates analyzed by both analysis methods, 140 were placed proximal to the same samples by both methods.

The AFLP dendrogram implies that the isolates may be placed into at least five different phylogenetic groups. Principal component analysis of the AFLP data was completed to reduce the dimensionality of the data sets. The first and second and the first and third principal components were plotted with characters relating to the five major clusters seen on the dendrograms. Results of the analysis are shown in Fig. 5. The plots

TABLE 1. *Bacillus* samples analyzed in this study[a]

Origin and isolate	Serotype	ET	Origin and isolate	Serotype	ET	Origin and isolate	Serotype	ET
Moss, Norway site 1			AH692	H15	17	AH693	H3a3b3c	18
AH512	H8	97	Moss, Norway site 1			Moss, Norway site 1		
AH515	H−	91	AH513	H−	67	AH514	H7	98
Moss, Norway site 2			AH516	ND	78	AH517	H35	86
AH518	H−	64	Moss, Norway site 2			Moss, Norway site 2		
AH521	H29	21	AH519	H−	71	AH520	H−	63
AH524	H10	89	AH522	H−	91	AH523	H−	91
AH527	H−	73	AH525	H15	29	AH526	H19	72
AH532	H−	74	AH528	H8	91	AH531	H14	67
AH535	AA	58	AH533	H8a8b	91	AH534	H8	91
AH538	H−	66	AH536	H14	65	AH537	H−	66
AH541	H−	67	AH539	H−	102	AH540	H−	67
Moss, Norway site 3			AH542	H19	69	Moss, Norway site 3		
AH543	H−	77	Moss, Norway site 3			AH545	H−	109
AH546	H8a8b	110	AH544	H−	77	AH548	H10	88
AH549	H−	112	AH547	H−	27	AH551	H−	75
AH552	H5a5b	20	AH550	H29	28	AH554	H−	19
AH555	H−	95	AH553	H−	111	AH557	H−	76
Moss, Norway site 4			AH556	H−	108	Moss, Norway site 4		
AH558	H8a8b	87	Moss, Norway site 4			AH560	H−	69
AH561	H16	106	AH559	H10a10b	88	AH563	H−	86
AH564	H−	80	AH562	H−	93	AH566	H−	96
AH567	H−	82	AH565	H−	67	AH569	H−	84
AH570	H−	85	AH568	H−	82	AH572	H10	68
AH573	H−	107	AH571	H−	79	AH575	H22	92
AH576	H10	70	AH574	H10	68	AH578	H7	99
AH579	H−	81	AH577	H17	100	AH581	H24	86
AH582	H5	104	AH580	H−	83	AH584	H10	105
AH585	H3	101	AH583	H11	94	AH587	H−	90
AH588	10	90	AH586	H7	101	AH590		
AH591			AH589			Singas, Norway		
Singas, Norway			AH592			AH620	H44	55
AH618	H31	1	Singas, Norway			AH623	AA	61
AH621	H34	13	AH619	H−	54	AH626	AA	60
AH624	H34	8	AH622	H5	52	AH630	H25	32
AH627	ND	60	AH625	AA	60	AH633	H−	41
AH631	H34	60	AH629	H−	62	AH636	H11	2
AH634	H−	42	AH632	H34	7	AH639	H4a4c	60
AH637	H6	39	AH635	H34	8	Steigen, Norway		
AH640	H6	39	AH638	H−	59	AH645	H−	53
Steigen, Norway			AH641	AA	39	AH648	H24	48
AH642	H34	12	Steigen, Norway			AH651	H34	48
AH646	H−	39	AH644	H−	50	AH654	H25	31
AH649	AA	40	AH647	H3a3b3c	47	Brekkvasselv, Norway		
AH652	H19	35	AH650	H−	43	AH658	H34	11
AH655	H19	36	AH653	H24	48	AH661	H43	37
Brekkvasselv, Norway			Brekkvasselv, Norway			AH664	ND	37
AH656	H34	10	AH657	H14	49	AH667	ND	57
AH659	H11	37	AH660	H11	37	AH670	H11	38
AH662	H3a3d3e	46	AH663	H34	9	Tromso, Norway		
AH665	H1	37	AH666	H10	56	AH673	H34	16
AH668	H11	77	AH669	H5	41	AH676	H11	25
Tromso, Norway			Tromso, Norway			AH679	H35	23
AH671	H34	3	AH672	H17	103	AH682	H−	49
AH674	H34	16	AH675	H4	16	AH685	H3a3b3c	33
AH677	H34	26	AH678	H10	24	AH688	H−	5
AH680	AA	58	AH681	H3a3b3c	48	AH691	H3a3b3c	15
AH683	H11	34	AH684	H3a3b3c	46	AH694	H34	14
AH686	H3a3b3c	30	AH687	H35	44			
AH689	H−	6	AH690	H16	22			

[a] Isolate names, serotypes, MEE ET (8, 10), and source of all samples are presented. ND, not determined.

of the principal components are labeled with these groups and support the conclusion that five groups of data could be distinguished from the first three principal components. This finding supports the presence of five significant branches within the phylogenetic tree.

A Mantel randomization test from 10,000 randomizations of the AFLP distance matrix was performed to test whether the similarities between the AFLP dendrogram and the MEE den-

drogram could have occurred by chance. Since the actual value (summed cross product = 4,320) is much larger than any of the randomization values, it is highly unlikely that the similarities between the AFLP dendrogram (Fig. 3) and the MEE dendrogram (Fig. 4) occurred by chance. The maximum value (4,440) gives an indication of perfection, although its distance to the actual value is somewhat noninformative because small changes in distances can lead to large changes in ranks.
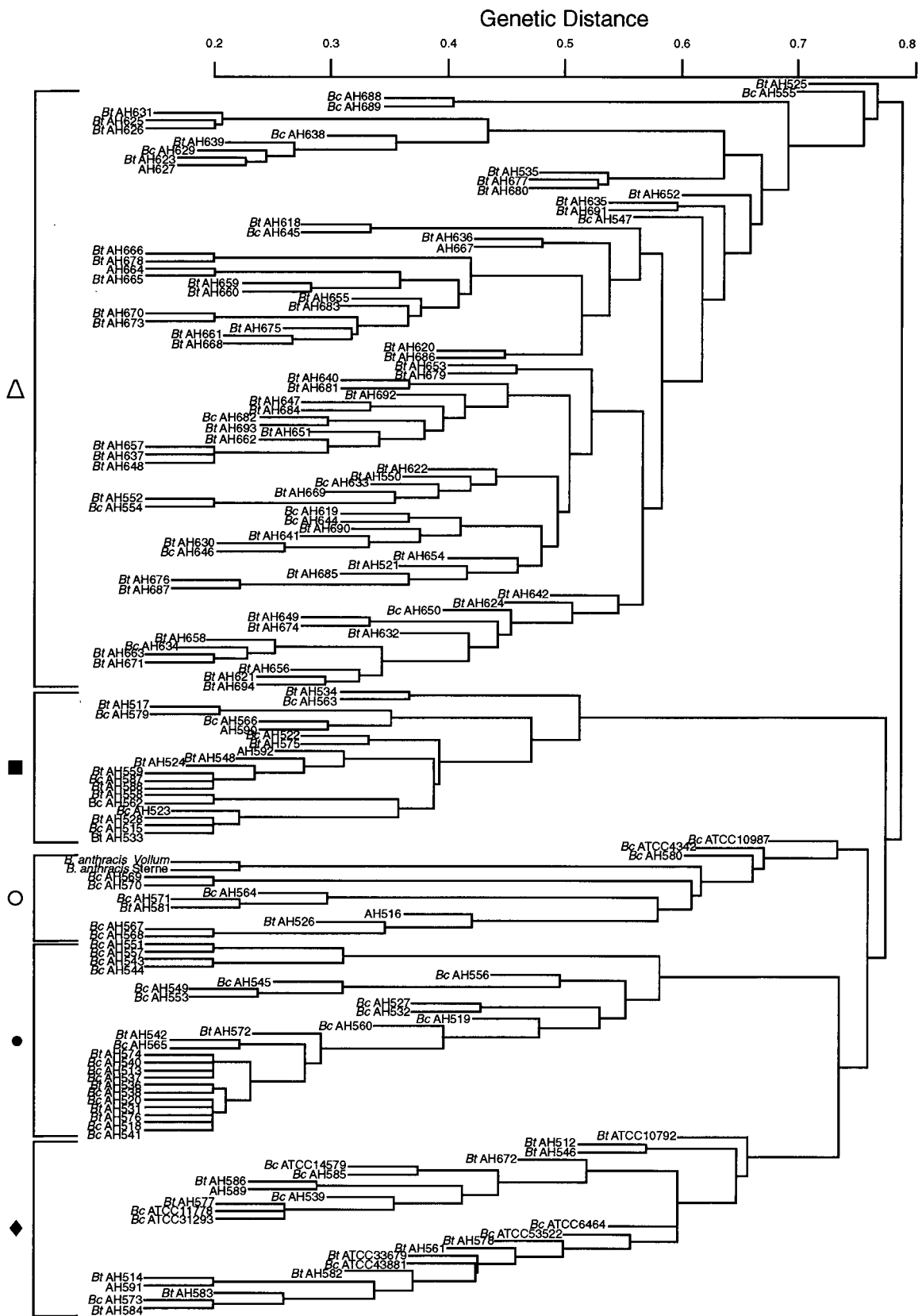
FIG. 3. Phylogenetic dendrograms of different *B. cereus*, *B. thuringiensis*, and *B. anthracis* isolates based on AFLP analysis of the samples with EcoRI-C and MseI-G primers. AFLP markers were used as genetic characters to determine the relationships among different *Bacillus* isolates. AFLP fragments were analyzed and dendrograms were generated as described in Materials and Methods. The distance measure or genetic distance is the fraction of peaks that are different between two samples. The more distance between two nodes of a tree, the more peaks that are different between these two nodes. Isolates are identified as *B. thuringiensis* or *B. cereus* based on the H serotype. Branches and symbols on the left of the figure are for reference to Fig. 4, 5, and 6.
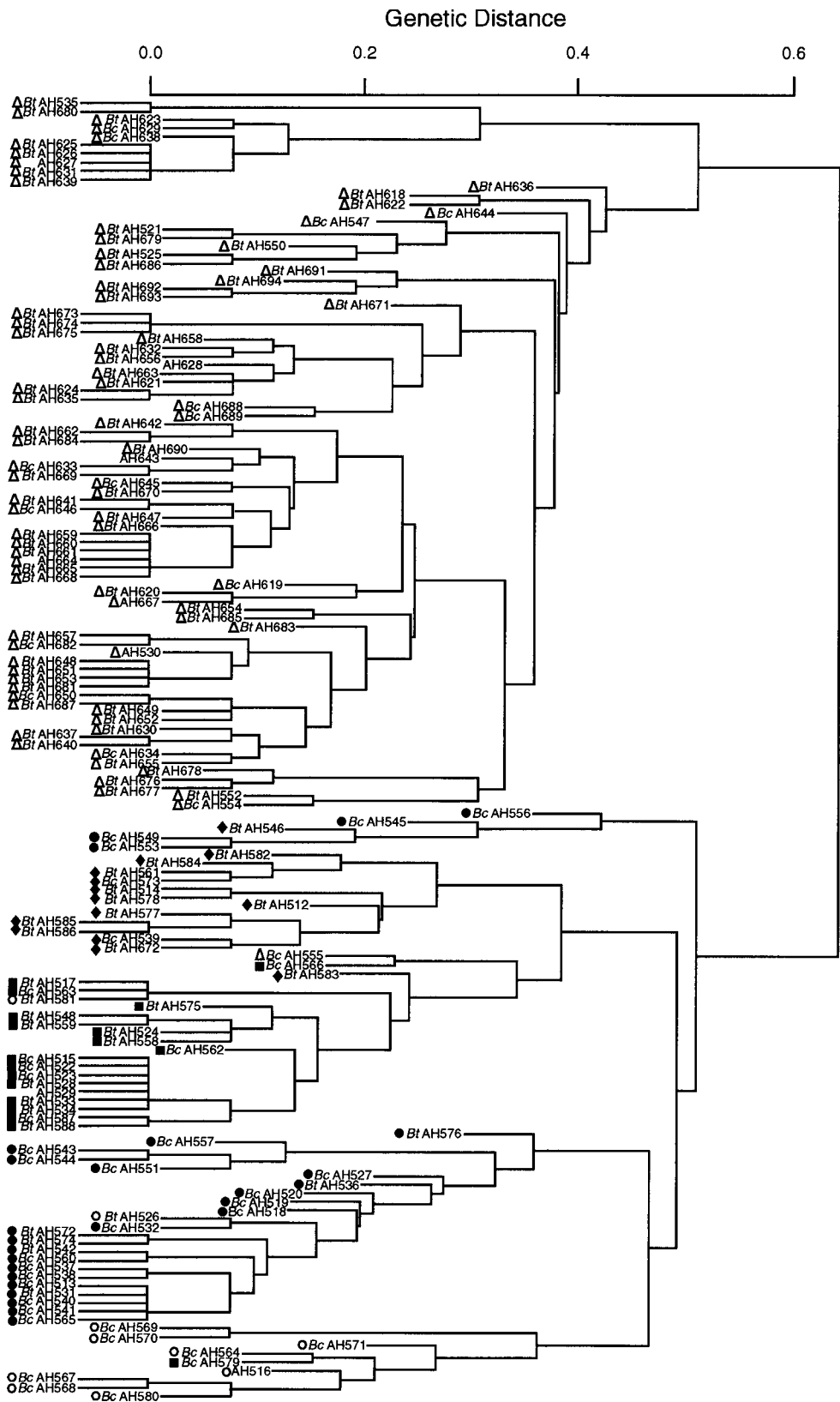
FIG. 4. Phylogenetic dendrogram of different *B. cereus*, *B. thuringiensis*, and *B. anthracis* isolates based on MEE analysis of the samples. MEE data for the different Norwegian isolates generated by Helgason et al. (10) was analyzed by using the same algorithms as used to analyze the AFLP data for these samples. The distance measure or genetic difference is the fraction of the 13 different enzyme alleles that differ among samples. The more distance between two nodes of a tree, the more enzyme alleles are different between those two nodes. Isolates are identified as *B. thuringiensis* or *B. cereus* based on the H serotype. Symbols to the left of each sample identify which branch of the AFLP-based dendrogram the sample occupies.
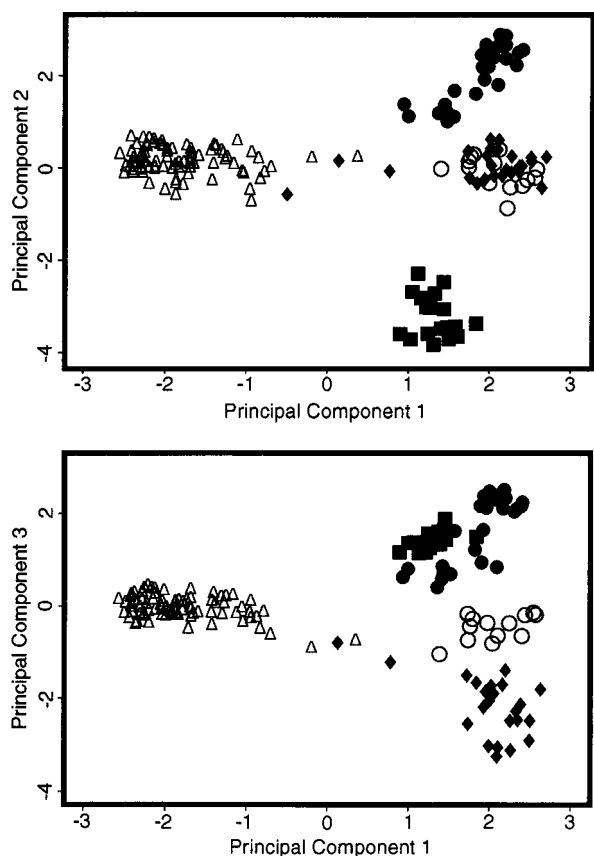
FIG. 5. The first three principal components of the AFLP analysis fingerprint summaries for the Norwegian isolates are presented. Each isolate was placed into one of five groups based on their clustering on the dendrogram shown in Fig. 3. The five groupings from the first three principal components are analogous to the dendrogram results.

16S rRNA gene analysis. The DNA sequence was determined for almost the full length (1,482 bp) of the 16S rRNA gene for two *B. thuringiensis* type strains, two *B. anthracis* isolates, five *B. cereus* type strains, and eight diverse Norway environmental isolates (Table 2). A comparison of the data for these 17 isolates (Table 2) revealed that only 14 of 1,482 nucleotide positions varied. This illustrates how highly conserved this gene is in the group 1 *Bacilli*. A phylogenetic tree (Fig. 6) based on the 16S ribosomal DNA (rDNA) sequences separated the isolates into discreet groups in a manner consistent with AFLP-based analysis (Fig. 3). However, the limited number of differences within this DNA sequence does not furnish the detailed phylogenetic resolution provided by AFLP analysis. Another example of increased resolution based on AFLP analysis is shown in Fig. 7. Figure 7A shows an AFLP fragment profile for AH648, a Norwegian *B. thuringiensis* isolate; Fig. 7B shows a profile for a different *B. thuringiensis* isolate, AH665, and Fig. 7C shows a profile for a third *B. thuringiensis* isolate, AH678, also from Norway. Figure 7 shows that AH648 is not closely related to the other two isolates and that none of the isolates are identical. However, analysis of the 16S rDNA sequences from these samples (Table 2) demonstrates that they are all *B. thuringiensis* isolates and contain identical 16S rDNA sequences. Phylogenetic analysis based on 16S rDNA sequences suggests that *B. anthracis* is very similar to *B. cereus* and *B. thuringiensis* type strains, in contrast to AFLP analysis, which suggests that they are quite different. The Norwegian *Bacillus* soil isolate AH526 16S rDNA sequence differs from the 16S sequence for *B. anthracis* by, at most, one nucleotide. This is one of only a few *Bacillus* isolates that have such close homology to the *B. anthracis* 16S rDNA sequence.

TABLE 2. Comparison of a 1,482-bp sequence of the 16S rDNA genes from different *Bacillus* isolates[a]

Isolate	Nucleotide at position:														No. of differences
	60	73	75	116	165	175	180	269	998	1002	1013	1129	1406	1445	
33679	A	T	T	G	C	Y	G	A	A	C	G	A	C	T	
14579	A	T	T	G	C	C	G	A	A	C	G	A	C	T	1
11778	A	T	T	G	C	C	G	A	A	C	G	A	C	T	1
31293	A	T	T	G	C	C	G	A	A	C	G	A	C	T	1
43881	A	T	T	G	Y	T	G	A	A	C	G	A	C	T	2
Vollum	A	T	T	G	C	C	G	A	C	C	G	W	C	T	3
Sterne	A	T	T	G	C	C	G	A	C	C	G	W	C	T	3
AH526	A	T	T	G	C	C	G	A	C	C	G	A	C	T	2
10792	G	C	A	G	C	T	G	A	A	C	G	A	C	T	4
53522	G	C	A	G	C	T	G	A	A	C	G	A	M	T	5
AH533	G	C	A	G	Y	T	G	A	A	C	G	A	C	T	5
AH540	R	Y	W	G	C	T	G	A	A	C	G	A	C	A	5
AH527	G	C	A	G	Y	T	G	A	A	C	G	A	C	A	6
AH521	A	T	T	A	T	T	A	G	A	T	A	A	C	A	8
AH648	A	T	T	A	T	T	A	G	A	T	A	A	C	A	8
AH665	A	T	T	A	T	T	A	G	A	T	A	A	C	A	8
AH678	A	T	T	A	T	T	A	G	A	T	A	A	C	A	8

[a] A 1,482-bp amplicon containing a bacterial 16S rDNA gene was amplified from 17 different *Bacillus* isolates, and the amplified DNA fragments were sequenced. Samples analyzed included *B. thuringiensis* subsp. *kurstaki* (ATCC 33679), *B. cereus* ATCC 14579, *B. cereus* ATCC 11778, *B. cereus* ATCC 31293, *B. cereus* ATCC 43881, *B. anthracis* isolates Vollum and Sterne, Norwegian *B. thuringiensis* isolate AH526, *B. thuringiensis* ATCC 10792, *B. cereus* ATCC 53522, six Norwegian *B. thuringiensis* isolates (AH521, AH533, AH648, AH665, and AH678), and two Norwegian *B. cereus* isolates (AH527 and AH540). Sequence results were used to generate a phylogenetic analysis (Fig. 6.) A, C, G, and T refer to the four deoxynucleotides that comprise DNA. M refers to an A or C at that position; R refers to a purine (A or G) at that position; Y refers to a pyrimidine (C or T) at that position; and W refers to an A or T at that position. There are multiple copies of the 16S rDNA gene in these bacteria, and some show differences in sequence among their different alleles. Isolates Vollum and Sterne were provided by the U.S. Army Medical Research Institute of Infectious Diseases, Ft. Detrick, Md. "AH" isolates were provided by the Biotechnology Centre of Oslo, University of Oslo, Oslo, Norway.

FIG. 6. Phylogenetic analysis of different *B. cereus*, *B. thuringiensis*, and *B. anthracis* isolates based on differences in 16S rDNA gene sequences. Differences in 16S rDNA sequences among the different isolates were used as genetic characters to determine the relationships among different *Bacillus* species. DNA sequences were analyzed by using the UPGMA cluster analysis algorithm of the phylogeny analysis using parsimony (PAUP) version 4 software package (26). Analysis was based on 14 different variable nucleotides within a 1,482-bp DNA sequence (Table 2). The numbers on the branches refer to the number of base differences among different isolates.

DISCUSSION

AFLP analysis provides a relatively rapid method of measuring phylogenetic distances among related microbial species and among different isolates of the same species. Such information is important in developing an understanding of the diversity within a microbial species and its relationship to its closest relatives.

AFLP analysis generates a "fingerprint" of DNA fragments (Fig. 1 and 7). The number of fragments generated is only limited by the restriction endonuclease combinations used and the number of different PCR primers used to analyze the resulting DNA restriction fragments. It provides detailed information about the relationships among different microbial species and shows the extent of variation within a species based on analysis of a percentage of the genome sequence and DNA fragment length polymorphisms among different microbes.

Often it is necessary to compare AFLP profiles from a large number of different microbial isolates to one another. In these experiments, fluorescent AFLP analysis of *Bacillus* species generated over 40 fragments between 100 and 500 bp per sample loaded onto three different gels. Comparing these data among 154 samples and controls becomes quite complex. Rapid comparison of a large number of AFLP profiles or

comparison of newly generated profiles to those archived from earlier analyses requires automation of profile scoring and analysis.

There are three aspects to automated AFLP analysis: generating the DNA fragments, analyzing the fragment profiles, and using the information generated by the profiles to conduct accurate phylogenetic analyses. Random and systematic errors add variability to the AFLP profiles as seen in data collected from identical samples (Fig. 1). To increase confidence in the AFLP data analysis, each sample was replicated three times. Plotting all replicates on a single dendrogram allows identification and numeration of the procedure-induced variability and identification of human errors (Fig. 2).

The information from several replicates must be combined to generate a DNA profile, a set of DNA fragments for a particular sample that can be used for comparison to other samples. We define the combined peak profile, or AFLP fingerprint, from a set of replicates as the peaks that occur in every replicate. Replicates of the data generated from the Norway *Bacillus* isolates have a maximum Jaccard distance of between 0.1 and 0.2 when a fluorescence threshold level set at 50 is used. Once the replicates are combined, comparisons that show differences of >0.2 almost always indicate actual differences in the DNA fragment profiles. By combining replicates, the noise level in the AFLP fingerprint drops to <0.2. The level of uncertainty is believed to be 0.1 or less. However, the level of uncertainty for these AFLP fingerprints has not been ascertained, so a conservative 0.2 Jaccard distance level was used as the minimum level at which comparisons should be considered reliable for this AFLP data (Fig. 3).

Hierarchical methods were chosen over methods that used a fixed number of clusters to reduce the computing time. However, the use of a hierarchical algorithm can produce a different cluster membership than a fixed cluster size algorithm, such as k-means (14). The clustering algorithms also do not prevent two peaks from the same sample from being placed in the same cluster. Moreover, there is no allowance for peak location errors to be different within a lane of the gel (as shown in Fig. 1B) and no current method that can reliably determine the size of these errors. Our results show that all these types of errors are in the noise level of the analysis.

Distance methods were used to determine similarities for the large data set analyzed because they were computationally fast and easy to implement. The data were distilled down to the presence or absence of a DNA fragment in the profile, and the profile changed with every sample. The Jaccard coefficient is an appropriate measure of distance since it looks only at peaks present in at least one profile.

When all suitable AFLP peaks are kept for analysis there is a problem in correctly comparing two samples that does not occur when a predetermined number of fragments is used. For example, if one sample has fewer peaks above the threshold value because less material was loaded into the polyacrylamide gel lane, then the distance value reflects the different amount of material loaded onto the gel in addition to the true distance between the two samples. There must be some method to standardize the amount of sample per lane before distances are computed. One method of approaching standardization is to assume all samples are replicates of one another, and therefore their tallest peaks should be the same. Standardizing the
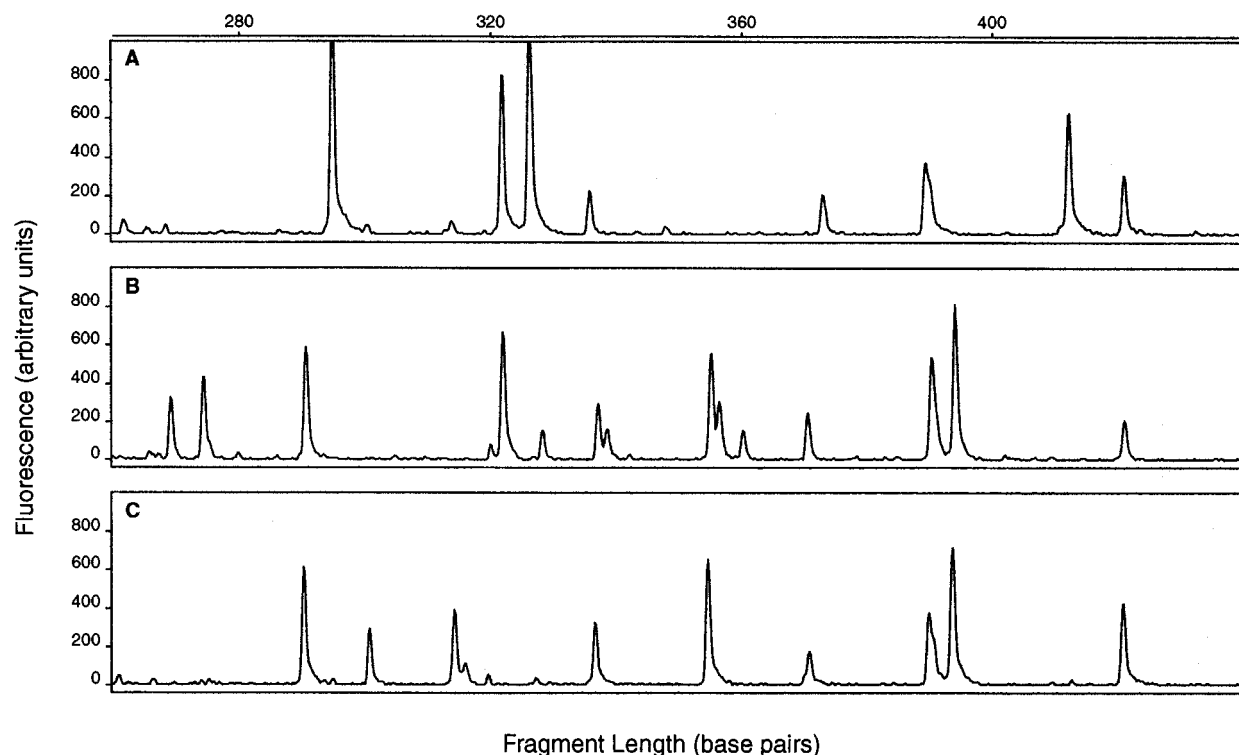
FIG. 7. AFLP profiles of three different *B. thuringiensis* samples sharing the same 16S rDNA gene sequence. AFLP analysis of the samples was conducted using EcoRI-C and MseI-G primers. Resulting DNA fragments were separated on an ABI377 automated DNA sequencer. (A) Profile for *B. thuringiensis* isolate AH648. (B) Profile for *B. thuringiensis* isolate AH665. (C) Profile for *B. thuringiensis* isolate AH678. These three isolates share exactly the same 16S rDNA sequence (Fig. 6 and Table 2). Only a portion of each profile from 260 to 440 bp is shown.

different amounts of material in a lane can be done by selecting the tallest peaks in each lane where the number of peaks selected is determined by the lane with the fewest peaks within the comparison being done. For the Norway AFLP data, the minimum number of peaks in a sample was ca. 40, so the 40 tallest peaks in each sample were used for comparisons. Using the Jaccard similarity measure and a standardized set of peaks simplifies interpretation of the dendrogram. The distance measure is really the fraction of the 40 peaks that are different between two samples. The greater the distance between two nodes within the tree, the more of the 40 peaks that differ between these two nodes.

Agglomerative hierarchical clustering using the similarity matrix of Jaccard coefficients and the UPGMA method (21) was used to give an indication of the relationships of samples among themselves and to simplify comparisons between the AFLP and the MEE data (Fig. 3 and 4). The computer analysis of the fluorescent AFLP data was confirmed by the MEE and 16S rDNA sequence data. Phylogenetic analysis based on the three different methods provided consistent results (Fig. 3, 4, and 6). A comparison of the phylogenetic analysis based on AFLP results (Fig. 3) to an analysis based on MEE data (Fig. 3 and 4) showed that, for most isolates, the members of different phylogenetic branches cluster together using either method. However, 40 fragments/sample were used in the AFLP analysis compared to a fewer number of loci for the MEE and 16S rDNA analysis (13 and 14 datum points, respectively). Therefore, AFLP analysis provided more phylogenetic resolution than the other methods (Fig. 3).

It can be argued that the presence or absence of one or more large plasmids can affect the analysis. *B. anthracis* contains two plasmids and plasmids of similar size have been reported in *B. cereus* and *B. thuringiensis* (7). These plasmids, combined, account for ca. 5% of the *B. anthracis* genome (8, 19, 20). Based upon the plasmids' nucleotide sequence, the AFLP primers chosen for this study generate one fluorescently labeled fragment from pX01 and none for pX02. The presence or absence of the single pX01 fragment among the 39 other fragments does not greatly change its relationship within the phylogenetic tree because the AFLP fingerprints for other *Bacillus* isolates analyzed show significantly more differences.

The phylogenetic analysis of these Norwegian soil isolates (Fig. 3) illustrates the great genetic diversity among the group 1 *Bacilli*. The Norway samples were collected from a comparatively small geographic area, yet they display a high degree of genetic variation. This is in contrast to *B. anthracis*, which in a previous analysis of a global collection of 78 strains showed little variation among the isolates (15).

Phylogenetic analysis using AFLP fragments separates the Norwegian isolates into at least five distinct groups. Almost all members of the smaller groups were collected from diverse environments around Moss, Norway (6, 9). The samples in the largest group (labeled "Δ" in Fig. 3) were collected from four other geographic sites in Norway. AFLP-based phylogenetic analysis was confirmed by sequencing of the 16S rRNA gene.

Results presented here demonstrate that *B. cereus* and *B. thuringiensis* are highly polymorphic species and that simple analysis of a limited number of reference strains is not sufficient to

characterize these species. They also demonstrate that different *B. thuringiensis* and *B. cereus* isolates are interspersed with one another throughout the phylogenetic tree, although the *B. cereus* isolates analyzed tended to cluster in the bottom three branches of the tree and the *B. thuringiensis* isolates were more prevalent in the top branch. Analysis demonstrates the presence of at least five branches in the phylogenetic tree for these species, and principal component analysis confirms this (Fig. 5).

The phylogeny of *B. anthracis*, *B. cereus*, and *B. thuringiensis* is under debate. Helgason et al. (11) suggested that these three species should be considered as one based on phylogenetic studies with MEE results. Interspersion of *B. cereus* and *B. thuringiensis* isolates argues that these are artificial systematic designations and that, based on the high degree of polymorphism among the different isolates, these species may be polyphyletic. At least in the case of *B. thuringiensis*, the traditional classification has been based on the *cry* proteins whose genes are found on large and small plasmids. The wide distribution of this character in chromosomally diverse backgrounds is indicative of its horizontal transfer. *B. cereus*, evidently by default, has been any group 1 *Bacillus* that is not *B. thuringiensis* or *B. anthracis*. AFLP results confirm that there is also significant variability among the different *B. cereus* isolates and that these differences are much greater than those seen for other microbial species, including *B. anthracis* (1, 12, 15, 16, 18, 24, 28; Jackson et al., unpublished). Members of an individual branch share similar profiles and are quite closely related. In contrast, members of other branches may differ extensively, sometimes sharing only a minority of the DNA fragments in their profiles. It may therefore be more accurate to group the different isolates of *B. cereus* and *B. thuringiensis* based on which branch they occupy in the phylogenetic dendrogram. In our opinion, it is important to understand the phylogenetic diversity of *B. cereus* and *B. thuringiensis* when drawing conclusions about relationships among the group 1 bacilli. Since *B. anthracis* is very monomorphic in comparison to *B. cereus* and *B. thuringiensis* isolates, *B. anthracis* should probably be considered as a distinct species from *B. cereus* and *B. thuringiensis* in spite of the close relationship to some strains, as measured by 16S rRNA sequencing and MEE.

REFERENCES

1. **Aarts, H. J. M., L. A. J. T. van Lith, and J. Keijer.** 1998. High-resolution genotyping of Salmonella strains by AFLP-fingerprinting. Lett. Appl. Microbiol. **26:**131–135.
2. **Andersen, G. L., J. M. Simchock, and K. H. Wilson.** 1995. Identification of a region of genetic variability among *Bacillus anthracis* strains and related species. J. Bacteriol. **178:**377–384.
3. **Ash, C., J. A. 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. **Ash, C., J. A. E. Farrow, S. Wallbanks, and M. D. Collins.** 1991. Phylogenetic heterogeneity of the genus Bacillus revealed by comparative analysis of small subunit ribosomal RNA sequences. Lett. Appl. Microbiol. **13:**202–206.
5. **Baumann, L., K. Okamoto, B. M. Unterman, M. J. Lynch, and P. Baumann.** 1984. Phenotypic characterization of *Bacillus thuringiensis* and *B. cereus*. J. Invertebr. Pathol. **44:**329–341.
6. **Carlson, C. R., D. A. Caugant, and A.-B. Kolstø.** 1994. Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. Appl. Environ. Microbiol. **60:**1719–1725.
7. **Carlson, C. R., T. Johansen, M.-M. Lecadet, and A.-B. Kolstø.** 1996. Genomic organization of the entomopathogenic bacterium *Bacillus thuringiensis* subsp. Berliner 1715. Microbiology **142:**1625–1634.
8. **Crickmore, N., D. R. Zeigler, J. Feitelson, E. Schnepf, J. Van Rie, D. Lereclus, J. Baum, and D. H. Dean.** 1998. Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. **62:**807–813.
9. **Helgason, E., D. Caugant, M.-M. Lecadet, Y. Chen, J. Mahillon, A. Lövgren, L. Hegna, K. Kvaløy, and A.-B. Kolstø.** 1998. Genetic diversity of *Bacillus cereus/B. thuringiensis* isolates from natural sources. Curr. Microbiol. **37:**80–87.
10. **Helgason, E., D. A. Caugant, I. Olsen, and A.-B. Kolstø.** 2000. Genetic structure of population of *Bacillus cereus* and *B. thuringiensis* isolates associated with periodontitis and other human infections. J. Clin. Microbiol. **38:**1615–1622.
11. **Helgason, E., O. A. Økstad, D. A. Caugant, H. A. Johansen, A. Fouet, M. Mock, I. Hegna, and A.-B. Kolstø.** 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. Appl. Environ. Microbiol. **66:**2627–2630.
12. **Jackson, P. J., K. K. Hill, M. T. Laker, L. O. Ticknor, and P. Keim.** 1999. Genetic comparison of *Bacillus anthracis* and its close relatives using amplified fragment length polymorphism and polymerase chain reaction analysis. J. Appl. Microbiol. **87:**263–269.
13. **Johnson, R. A., and D. W. Wichern.** 1982. Applied multivariate statistical analysis. Prentice-Hall, Inc., Englewood Cliffs, N.J.
14. **Kaufman, L., and P. J. Rousseeuw.** 1990. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, Inc., New York, N.Y.
15. **Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, C. R. Kuske, and P. J. Jackson.** 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J. Bacteriol. **179:**818–824.
16. **Keim, P., A. M. Klevytska, L. B. Price, G. Zinser, K. L. Smith, M. E. Hugh-Jones, R. Okinaka, K. K. Hill, and P. J. Jackson.** 1999. Molecular diversity in *Bacillus anthracis*. J. Appl. Microbiol. **87:**215–217.
17. **Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. **182:**2928–2936.
18. **McLauchlin, J., G. Ripabelli, M. M. Brett, and E. J. Threlfall.** 2000. Amplified fragment length polymorphism (AFLP) analysis of *Clostridium perfringens* for epidemiological typing. Int. J. Food Microbiol. **56:**21–28.
19. **Okinaka, R., K. Cloud, O. Hampton, A. Hoffmaster, K. Hill, P. Keim, T. Koehler, G. Lamke, S. Kumano, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P. Jackson.** 1999. Sequence assembly and analysis of pX01 and pX02. J. Appl. Microbiol. **87:**261–262.
20. **Okinaka, R. T., K. Cloud, O. Hampton, A. R. Hoffmaster, K. K. Hill, P. Keim, T. M. Koehler, G. Lamke, S. Kumano, J. Mahillon, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P. J. Jackson.** 1999. Sequence and organization of pX01, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. J. Bacteriol. **181:**6509–6515.
21. **Rohlf, F. J.** 1993. NTSYS-PC-numerical taxonomy and multivariate analysis system, version 1.8. Exeter Software, Setauket, N.Y.
22. **Schnepf, E., N. Crickmore, J. Van Rie, D. Lereclus, J. Baum, J. Feitenson, D. R. Zeigler, and D. H. Dean.** 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. **62:**775–806.
23. **Schwarz, G., M. Herz, X. Q. Huang, W. Michalek, A. Jahoor, G. Wenzel, and V. Mohler.** 2000. Application of fluorescent-based semi-automated AFLP analysis in barley and wheat. Theor. Appl. Genet. **100:**545–551.
24. **Sloos, J. H., P. Janssen, C. P. A. van Boven, and L. Dijkshoorn.** 1998. AFLP™ typing of *Staphyloccocus epidermidis* in multiple sequential blood cultures. Res. Microbiol. **149:**221–228.
25. **Sokal, R. R., and F. J. Rohlf.** 1995. Biometry: the principles and practice of statistics in biological research, 3rd ed. W. H. Freeman and Company, New York, N.Y.
26. **Swofford, D. L.** 1998. Paup∗. Phylogenetic analysis using parsimony (∗and other methods), version 4. Sinauer Associates, Sunderland, Mass
27. **Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van der Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau.** 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23:**4407–4424.
28. **Zhao, S. H., S. E. Mitchell, J. H. Meng, S. Kresovich, M. P. Doyle, R. E. Dean, A. M. Casa, and J. W. Weller.** 2000. Genomic typing of *Escherichia coli* O157:H7 by semi-automated fluorescent AFLP analysis. Microbes Infect. **2:**107–113.