The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1

David A. Rasko, Jacques Ravel, Ole Andreas Økstad1,2, Erlendur Helgason1, Regina Z. Cer, Lingxia Jiang, Kelly A. Shores, Derrick E. Fouts, Nicolas J. Tourasse1, Samuel V. Angiuoli, James Kolonay, William C. Nelson, Anne-Brit Kolsto1,2, Claire M. Fraser and Timothy D. Read*

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA, 1The Biotechnology Centre of Oslo, University of Oslo, PB 1125, Oslo N-0316, Norway and 2School of Pharmacy, University of Oslo, PB 1068, Oslo N-0316, Norway

Received November 7, 2003; Revised December 19, 2003; Accepted January 9, 2004 DDBJ/EMBL/GenBank accession nos + AE017194 and AE017195

ABSTRACT

We sequenced the complete genome of *Bacillus cereus* ATCC 10987, a non-lethal dairy isolate in the same genetic subgroup as *Bacillus anthracis*. Comparison of the chromosomes demonstrated that *B. cereus* ATCC 10987 was more similar to *B. cereus* Ames than *B. cereus* ATCC 14579, while containing a number of unique metabolic capabilities such as urease and xylose utilization and lacking the ability to utilize nitrate and nitrite. Additionally, genetic mechanisms for variation of capsule carbohydrate and flagella surface structures were identified. *Bacillus cereus* ATCC 10987 contains a single large plasmid (pBc10987), of ~208 kb, that is similar in gene content and organization to *B. anthracis* pXO1 but is lacking the pathogenicity-associated island containing the anthrax lethal and edema toxin complex genes. The chromosomal similarity of *B. cereus* ATCC 10987 to *B. anthracis* Ames, as well as the fact that it contains a large pXO1-like plasmid, may make it a possible model for studying *B. anthracis* plasmid biology and regulatory cross-talk.

INTRODUCTION

*Bacillus cereus*, *B. thuringiensis* and *B. anthracis* all belong to the *B. cereus* sensu lato group of rod-shaped, Gram-positive, spore-forming bacteria (1). *Bacillus cereus* is the etiological agent of anthrax, an acute fatal animal and human disease that was employed as a bioterror agent in the autumn of 2001 (2). *Bacillus anthracis* shares a very close evolutionary relationship with two other common but much less pathogenic bacterial species: *B. thuringiensis*, a well known biological insecticide (3), and *B. cereus*, often considered at most, a soil-dwelling opportunistic pathogen (1). There are rare, usually non-fatal diseases associated with *B. cereus*, such as endophthalmitis after trauma to the eye (4,5) and two forms of human food poisoning, characterized by either diarrhea and abdominal distress or nausea and vomiting (1,6). However, more serious infections in immunocompromised individuals have been observed (7–10), and some *B. cereus* isolates have been implicated in a lethal infection similar in clinical presentation to *B. anthracis*, posing a potential public health issue (11). Many of the species-specific phenotypes of this group are encoded by plasmid genes, such as the *B. anthracis* lethal toxin complex and poly-D-glutamic acid capsule (plasmids pXO1 and pXO2, respectively) (12,13), and the *B. thuringiensis* insect toxins (14).

*Bacillus cereus*, *B. thuringiensis* and *B. anthracis* are genetically similar to an extent that comparisons of 16S rRNA sequences (15) or 16S–23S rRNA spacer regions (16) cannot adequately distinguish between the members of this group. There is no consensus on whether these bacteria should be separate species or considered specialized variants of a single species (17). Additionally, *B. anthracis* has been shown to be one of the most monomorphic bacterial species (17,18). A number of molecular typing schemes have been applied to distinguish individuals within the group, including pulsed-field gel electrophoresis (PFGE) (19), amplified fragment length polymorphism (AFLP) (20,21), multi-locus enzyme electrophoresis (MLEE) (17,19,22,23) and multi-locus sequence typing (MLST) (18). From this body of work, it is apparent that...
The Genome Sequence of Bacillus Cereus ATCC 10987 Reveals
Metabolic Adaptations and a Large Plasmid Related to Bacillus
Anthracis PXO1

Naval Submarine Medical Research Laboratory
Naval Submarine Base
New London Box 900 Bldg 148, Trout Avenue Groton, CT 06349-5900

Approved for public release, distribution unlimited

The original document contains color images.
a group of *B. cereus* and *B. thuringiensis* isolates are more closely related to *B. anthracis* than strains represented by the *B. cereus* species type strain ATCC 14579 that was sequenced recently (24) (Fig. 1). *Bacillus cereus* ATCC 10987 was chosen for sequencing as it was a widely available strain, has been shown by MLEE and other studies to be closely related to *B. anthracis* (17,22,25) and contained genes similar to those found on pXO1 (26). These features made the strain a useful addition to the comparative genomic analysis of *B. anthracis*.

*Bacillus cereus* ATCC 10987 was isolated from a study on cheese spoilage in Canada in 1930 (27,28). It has been demonstrated to contain putative virulence factors such as phosphatidylinositol-specific phospholipase C (PI-PLC), phosphatidylcholine-preferrring phospholipase C (PC-PLC), sphingomyelinase, non-hemolytic enterotoxin and proteases (29,30), and to express a high level of phospholipase C (A.-B.Kolstø, unpublished data).

The present study compares the chromosomes of *B. cereus* ATCC 10987, *B. cereus* ATCC 14579 and *B. anthracis* Ames, and reveals a number of metabolic pathways not identified previously in the *B. cereus* group of organisms, such as urease and xylose utilization, as well as potential mechanisms for antigenic variability of surface structures including capsule and flagella. Additionally, we identify a single large plasmid in *B. cereus* ATCC 10987 that is similar to the *B. anthracis* pXO1 plasmid, and encodes a number of unique potential pathogenicity and resistance factors as well as conserved regulatory proteins.

**MATERIALS AND METHODS**

**Sequencing of the *B. cereus* ATCC 10987 genome**

The random shotgun method, and cloning, sequencing and assembly were as described previously (26). Large (10–12 kb) and small (2.5–3.5 kb) insert random sequencing libraries were sequenced for this genome project with success rates of 84 and 87% and average high-quality read lengths of 666 and 683 nt, respectively. The completed genome sequence contained 23 042 and 57 171 reads from the large and small libraries, respectively, achieving an average of 10.4-fold sequence coverage per base. After assembly, gaps between contigs were closed by editing, walking library clones and linking assemblies by PCR. The Glimmer Gene Finder (31) was utilized to identify potential coding regions, and annotation was completed as described previously (32). The sequences of *B. cereus* ATCC 10987 genome and plasmid can be accessed using the GenBank accession nos AE017194 and AE017195, respectively. An estimate of the copy number of the plasmid was obtained by dividing the coverage depth of the plasmid by the coverage depth of the chromosome.

**BLAST score ratio analysis (BSRA)**

The BSRA is a modification of the technique described by Read et al. (33). For each of the predicted proteins of *B. cereus* ATCC 10987, we obtained a BLASTP raw score (34) for the alignment against itself (REF_SCORE) and the most similar protein (QUE_SCORE) in each of the genomes of *B. cereus* ATCC 14579 (24) and *B. anthracis* Ames (26). These scores were normalized by dividing the QUE_SCORE obtained for each query genome protein by the REF_SCORE. Proteins with a normalized ratio of less than 0.4 were considered to be non-homologous. The normalized BLAST score ratio of 0.4 is generally similar to two proteins being ~30% identical over their entire length.

BLAST score ratios were plotted as x, y coordinates as shown in Figure 3. Each protein in the reference genome (*B. cereus* ATCC 10987) was grouped according to its scores in each of the query genomes, and colored accordingly: yellow, unique to the reference; red, common to all three; cyan, common between *B. anthracis* Ames and the reference, but absent in *B. cereus* ATCC 14579; blue, common between *B. cereus* ATCC 14579 and the reference, but absent in *B. anthracis* Ames.

**PCR screening of a set of *B. cereus* group organisms**

PCR was used to screen a 23 strain set of *B. cereus* group organisms (Supplementary table S1 available at NAR Online) for the presence or absence of integral genes of urease, capsule or xylose pathways using the primers described in Supplementary table S2.
RESULTS

Whole genome analysis shows that *B. cereus* ATCC 10987 is more closely related to *B. anthracis* Ames than *B. cereus* ATCC 14579

*Bacillus cereus* ATCC 10987, the third complete genome sequence in the *B. cereus* group, has broad similarities to the *B. anthracis* Ames (26) and *B. cereus* ATCC 14579 (24) genome sequences (Table 1). The *B. cereus* ATCC 10987 chromosome shares a high degree of synten (conserved gene order) with the *B. anthracis* Ames and *B. cereus* ATCC 14579 chromosomes (Fig. 2A and B). Direct comparison of the complete nucleotide sequences using NUCmer (35) reveals that *B. anthracis* and *B. cereus* ATCC 10987 are 93.94% identical whereas *B. cereus* ATCC 14579 and *B. cereus* ATCC 10987 are 90.94% identical. Additionally, the proteins of *B. cereus* ATCC 10987, when analyzed by BRSA, are more similar to those of *B. anthracis* Ames than those of *B. cereus* ATCC 14579 (Fig. 3). This close relationship between *B. anthracis* Ames and *B. cereus* ATCC 10987 is confirmed in a phylogenetic tree based on seven partially sequenced genes used in MLST analysis of the *B. cereus* group (18) (Fig. 1) and is in line with the results of previous MLEE studies (17).

We identified *B. cereus* ATCC 10987 proteins without significant homology (BLAST score ratio less than 0.4) in the other two *B. cereus* group proteomes using BRSA. These proteins will be referred to as ‘novel’ herein. The relative chromosomal location of the novel proteins is shown in Figure 4A. There are also a significant number of proteins that have homologs in only two of the three chromosomes [i.e. in *B. cereus* ATCC 10987 and *B. cereus* ATCC 14579 but not *B. anthracis* Ames (Fig. 4A, Table 2)], suggesting a history of insertion and/or deletion in the evolution of the *B. cereus* group. In many cases, genes found at a specific position in one genome are replaced with others at the corresponding locus in another (for examples see Figs 5 and 6, and Supplementary fig. S1). Since we do not want to make untestable assumptions about the history of these events, we will use the neutral term ‘replacement’ to describe them herein. Table 2 provides a list of the replacements in each of the three genomes. These loci are often associated with strain-specific phenotypes, and several examples will be discussed in the following analysis.

The 208 kb plasmid in *B. cereus* ATCC 10987 has similarities to *B. anthracis* pXO1

The *B. cereus* ATCC 10987 plasmid, pBc10987, is 208 369 nt in length encoding 242 genes (Table 1). pBc10987 was compared with other sequenced large plasmids of the *B. cereus* group, *B. anthracis* pXO1 (~182 kb), pXO2 (~95 kb) and *B. thuringiensis* subsp. *israeliensis* pBtoxis (~128 kb) (36) using BRSA. pBc10987 and pXO1 show little similarity to the pBtoxis proteome (Fig. 4B) and even less to the pXO2 proteome (only five pXO2 proteins were conserved with either pBc10987, pXO1 or pBtoxis). Comparison of pBc10987 and pXO1 revealed that ~65% of proteins were homologous and ~50% were in a syntenic location (Fig. 2C) and the relative transcriptional direction of many of the pXO1 and pBc10987 genes has been retained, representing a conserved ‘plasmid backbone’. Comparison of the nucleotide sequences of the plasmids using NUCmer (35) reveals that pBc10987 and pXO1 are ~40% identical, whereas pBc10987 and pBtoxis are only ~7% identical. Based on nucleotide and protein similarity, it appears that pBc10987 and pXO1 may be members of a group of low-copy number plasmids (pBc10987 approximately one/cell; pXO1 approximately three copies/cell) that may also include another sequenced plasmid, pBtoxis from *B. thuringiensis* subsp. *israeliensis*, as a distant relative. Lack of knowledge regarding the replication machinery in these plasmids precludes us from conclusively grouping these plasmids together; however, the replication mechanism is different from those employed by *B. anthracis* pXO2 (12), *B. thuringiensis* pAW63 (37) or small *B. thuringiensis* plasmids (38).

Table 1. General features of the *B. cereus* group genomics

<table>
<thead>
<tr>
<th>Feature</th>
<th><em>B. cereus</em> ATCC 10987</th>
<th>Plasmid pBc10987</th>
<th><em>B. anthracis</em> Amesa</th>
<th>Plasmid pXO1</th>
<th><em>B. cereus</em> ATCC 14579b</th>
<th>Plasmid pBClin15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (bp)</td>
<td>5 224 283</td>
<td>208 369</td>
<td>5 227 293</td>
<td>181 677</td>
<td>94 829</td>
<td>5 426 909</td>
</tr>
<tr>
<td>No. of genes</td>
<td>5642</td>
<td>242</td>
<td>5508</td>
<td>217</td>
<td>113</td>
<td>5366</td>
</tr>
<tr>
<td>Percentage coding</td>
<td>85</td>
<td>80.9</td>
<td>84.3</td>
<td>77.1</td>
<td>76.2</td>
<td>84</td>
</tr>
<tr>
<td>Average gene length (nt)</td>
<td>787</td>
<td>696</td>
<td>800</td>
<td>645</td>
<td>639</td>
<td>820</td>
</tr>
<tr>
<td>%G + C content</td>
<td>35.6</td>
<td>33.5</td>
<td>35.4</td>
<td>32.5</td>
<td>33</td>
<td>35.3</td>
</tr>
<tr>
<td>rRNA operons</td>
<td>12</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>tRNAs</td>
<td>98</td>
<td>0</td>
<td>95</td>
<td>0</td>
<td>108</td>
<td>0</td>
</tr>
<tr>
<td>sRNAs</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Putative phage</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>Genes with assigned function</td>
<td>2802</td>
<td>94</td>
<td>2762</td>
<td>65</td>
<td>38</td>
<td>3839</td>
</tr>
<tr>
<td>Conserved hypothetical genesa</td>
<td>1074</td>
<td>74</td>
<td>1212</td>
<td>22</td>
<td>19</td>
<td>1481</td>
</tr>
<tr>
<td>Genes of unknown functionc</td>
<td>672</td>
<td>12</td>
<td>657</td>
<td>8</td>
<td>5</td>
<td>142</td>
</tr>
<tr>
<td>Hypothetical genesd</td>
<td>1094</td>
<td>62</td>
<td>877</td>
<td>122</td>
<td>51</td>
<td>?</td>
</tr>
</tbody>
</table>

aThe genome data for *B. anthracis* are derived from the complete genome sequence of *B. anthracis* Ames (26) and the plasmid sequences of *B. anthracis* strain ATCC 2012 (Florida) (13). Note that the number of genes is not the same as the number of proteins used in calculation of Figure 4 due to the presence of framshifts.
b*Bacillus cereus* ATCC 14579 data were obtained from Ivanova et al. (24).
cFor clarification, a hypothetical protein has no other homologous protein in any of the public databases, genes of unknown function have similarity to broad protein families but no specific function can be assigned, whereas a conserved hypothetical protein has one or more homologs in another species; however, no function has yet been attributed to any of these proteins to date.
The genetic basis for replication, maintenance and mobilization of pXO1 is unknown (39,40), suggesting a unique mechanism that may be conserved in pBc10987 based on the level of conservation of these plasmids. pBc10987 BCEA0008–BCEA0073 are most similar in composition and order to pXO1 BXA0064–BXA0120 encoding conserved hypothetical, membrane-associated and conjugative transfer-like proteins of other plasmid systems, such as the TraD/G family protein (BCEA0072). This region also contains proteins that are conserved to a lesser degree in pBtoxis, suggesting that it may be required in the basic maintenance of these plasmids. The similarity between pBc10987 and pXO1 also extends into a number of replication-related proteins including a type I DNA topoisomerase (BCEA0140, BXA0213, respectively) which is thought to aid in the stability of these plasmids (41). Only pBc10987 contains a unique plasmid-encoded DNA polymerase III β subunit, involved in tight association of the template DNA with the polymerase complex (42), which may ensure that the plasmid is replicated at an increased processivity and stability.

The B.anthracis pXO1 plasmid pathogenicity region containing the genes encoding the transcriptional regulator AtxA, lethal factor, protective antigen and edema factor is absent from pBc10987, but this region has been replaced by a B.cereus ATCC 10987 island containing a copper-requiring tyrosinase, amino acid transport system, arsenate resistance gene cluster and regulatory proteins (gray box in Fig. 2C). pBc10987 also includes two novel potential toxins: BCEA0165, a MIP family channel protein, and BCEA0203, a possible metalloprotease. However, the pBc10987 island is not flanked by any mobile genetic elements that are thought to have been involved in the integration of the pathogenicity island on pXO1 (39). The species-specific pathogenicity-associated islands are the most variable portions of the plasmids as none of the proteins in either island are shared between plasmids (Figs 2C and 4B).

Another interesting similarity between pXO1 and pBc10987 that may affect the phenotype is the presence of a plasmid-borne abrB gene. AbrB is a pleiotropic transition state regulatory protein that has been shown to negatively regulate the expression of the lethal toxin genes in B.anthracis (43). Additionally, it has been demonstrated that AbrB in conjunction with Spo0A are the major regulatory factors in the developmental pathways of spores and biofilms in B.subtilis (44). In B.cereus ATCC 10987, as in B.anthracis Ames, there are two divergent chromosomal copies of abrB as well as a plasmid copy of this regulator, whereas B.cereus ATCC 14579

Figure 2. Plot of BSRA based on genomic location. Each triangle indicates a single protein plotted on the 5’ end of the coding region in the reference genome and the best match in the query organism. They are color coded as follows: normalized score ≥0.95, red; normalized score ≥0.80 and <0.95, orange; normalized score ≥0.60 and <0.80, yellow; normalized score ≥0.40 and <0.60, green; normalized score ≥0.20 and <0.40, cyan; normalized score ≥0.0 and <0.20, blue. (A) Comparison of B.cereus ATCC 10987 (reference) and B.anthracis Ames (query). (B) Comparison of B.cereus ATCC 10987 (reference) and B.cereus ATCC 14579 (query). (C) Comparison of B.cereus ATCC 10987 pBc10987 (reference) and B.anthracis Ames pXO1 (query). The shaded region indicates the location of the pathogenicity-associated island in pXO1 that shows no similarity to the proteins of pBc10987.
contains only the two chromosomal copies (Supplementary fig. S2). The two chromosomal copies of *abrB* in *B. cereus* ATCC 10987 (BCE2077 and BCE0035) are each >97% identical to the chromosomal *abrB* orthologs in *B. anthracis* Ames/*B. cereus* ATCC 14579 (BA2000/BC1996 and BA0034/BC0042), but the paralogs are divergent (~67% nucleotide and ~50% amino acid identity; Supplementary fig. S2). Additionally, the plasmid-encoded copies of AbrB are most similar to each other (~80% amino acid identity) and are as similar to the chromosomal copies of this protein as the paralogous chromosomal copies are to each other. It has been previously noted that the *B. anthracis* pXO1 copy of AbrB was truncated by 27 amino acids in comparison with chromosomal copies (43,45). This is not the case for the pBc10987-encoded AbrB. Even though *B. cereus* ATCC 10987 lacks the lethal toxin genes, the conservation of the AbrB homologs on the large plasmids suggests a possible role in expression of plasmid-encoded factors. In *B. anthracis*, the regulatory activity was attributed to a chromosomal encoded copy of AbrB (43,45); however, the role of pBc10987 AbrB is unclear.

*Bacillus cereus* ATCC 10987 may prove to be a convenient non-lethal 'model' organism for studying *B. anthracis* plasmid biology issues such as plasmid replication, maintenance and transfer as well as regulatory cross-talk between chromosome and plasmid.

The pBc10987 plasmid sequence allows further identification of genes in flux between *B. cereus* group chromosomes and plasmids

As *B. cereus* group plasmids and phages are identified and sequenced, it is becoming apparent that many genes located on the chromosome are actually homologs of episomal determinants. An example of likely genetic exchange between the plasmid and chromosome discovered in *B. cereus* ATCC 10987 are the two identical copies of a 3605 nt Tn554 element encoded by BCE3147–BCE3149 and BCEA0001–2. The Tn554 element is composed of three essential proteins, TnsABC, whose closest relative is the Tn554 from *Staphylococcus aureus* (46). There are four other potential coding regions that are conserved in association with the transposable element (BCE3150–53/BCEA0003–6). Most interesting in this group of proteins is BclA, which has been shown to be the major spore surface antigen of *B. cereus* and *B. anthracis* (47–49) as well as an exosporium depth determinant in *B. anthracis* (50). *Bacillus anthracis* Ames and *B. cereus* ATCC 14579 each contain only a single chromosomal copy of BclA, whereas *B. cereus* ATCC 10987 contains both a chromosomal copy and a plasmid copy. A BclA-like protein, Bcol (*Bacillus* collagen-like), has been recently identified on small (<12 kb) *B. thuringiensis* plasmids (38), but no functional role could be assigned. It is possible that BclA proteins affecting spore morphology and surface properties may be horizontally transferred among *B. cereus* group bacteria as part of the Tn554 transposable element.

The gerX operon, located on the *B. anthracis* pXO1 plasmid, but on the chromosome of *B. cereus* ATCC 10987, represents another example of genes moving between replicons in the *B. cereus* group. In *B. anthracis*, this operon has been demonstrated to be required for virulence and germination in a mouse anthrax model (51). Comparison of the other germination proteins [*gerH, gerL, gerK, gerP, gerS and gerV* operon gene products; *B. cereus* ATCC 10987 lacks the *B. cereus*-specific gerQ operon (52)] among the three *B. cereus* group bacteria shows a high level of conservation (>90% amino acid identity).
The gerX operon is not present in B. cereus ATCC 14579, whereas the gerX-encoded proteins from B. cereus ATCC 10987 and B. anthracis Ames share a lesser degree of identity than the proteins encoded by other germination operons (~67% amino acid identity). It is unclear whether the gerX operons should be considered as true orthologs, and only

**Figure 4.** Venn diagram illustrating the number of putative proteins associated with each organism and the number shared with the intersecting organism. Tick marks that are on each circle represent the location of the unique proteins (BLAST score ratio less than 0.4) on the genome. (A) Chromosomal comparison: blue, B. anthracis Ames; green, B. cereus ATCC 10987; red, B. cereus ATCC 14579. (B) Plasmid comparison: blue, B. anthracis Ames pXO1; green, B. cereus ATCC 10987 pBc10987; magenta, B. thuringiensis subsp. israelensis pBtox1.
Table 2. Major phenotypic differences and characteristics of B.cereus group genomes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>B.anthracis Ames</th>
<th>B.cereus ATCC 10987</th>
<th>B. cereus ATCC 14579</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid pXO1 (189 kb)</td>
<td>pBc10987 (208 kb)</td>
<td>Linear phage-like (15 kb)</td>
<td></td>
</tr>
<tr>
<td>Plasmid pXO2 (96 kb)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tripartite lethal toxin</td>
<td>Present</td>
<td>Absent</td>
<td>Absent</td>
</tr>
<tr>
<td>rRNA</td>
<td>11 copies</td>
<td>12 copies</td>
<td>13 copies</td>
</tr>
<tr>
<td>Urease gene cluster</td>
<td>Absent</td>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td>Xylose utilization genes</td>
<td>Absent</td>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td>Polysaccharide capsule</td>
<td>Absent</td>
<td>Present</td>
<td>Absent</td>
</tr>
<tr>
<td>Flagellar genes</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Tagatose utilization genes</td>
<td>Absent</td>
<td>Present</td>
<td>Partial</td>
</tr>
<tr>
<td>Functional PlcR</td>
<td>Absent</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Mobile genetic elements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phage</td>
<td>4 (1 degenerate)</td>
<td>6 (1 linear plasmid)</td>
<td></td>
</tr>
<tr>
<td>Tn554</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Tn7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IS505</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>IS231</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Arginine deiminase genes</td>
<td>Absent</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>B.cereus repeat 1 (bcr1)</td>
<td>12 copies</td>
<td>72 copies</td>
<td>52 copies</td>
</tr>
<tr>
<td>Restriction enzymes</td>
<td>Absent</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Competence genes</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
</tbody>
</table>

*aThe capsule biosynthesis region of ~20 kb is partially similar in both B.cereus genomes (Fig. 7).

*bBacillus anthracis Ames contains a non-functional flagellar operon with genes BA1661, BA1670, BA1677 and BA1709 containing frameshifts preventing the assembly of a functional flagellar structure. The flagellar locus also contains one flagellin subunit encoded by gene BA1706.

*cEach of the B.cereus genomes contains an apparent functional flagellar operon. Bacillus cereus ATCC 10987 contains two divergent flagellin subunits, whereas B.cereus ATCC 14579 contains four flagellin subunits, three of which are closely related and one that is divergent (Supplementary fig. S3).

*dThere are similar tagatose utilization genes in B.cereus ATCC 14579; however, they are not assembled into a potential transcriptional unit and not all of the genes in B.cereus ATCC 10987 are represented in B.cereus ATCC 14579.

*e*plcR in B.anthracis Ames contains a frameshift that results in a truncated and non-functional protein. The *plcR* gene in the B.cereus strains is full length and apparently functional and can act as a regulatory protein.

*fTn554 is present in identical copies on the plasmid pBc10987 and the chromosome. The adjacent regions are also duplicated and the total repeat is 8309 bp and contains a bclA gene involved in spore antigen and exosporium determination.

*gTwo similar but unique Tn7 elements are located on the chromosome.

*hThere are four unique restriction–modification systems in B.cereus ATCC 10987 and three in B.cereus ATCC 14579.

*iMany of the required competence genes in B.subtilis have been identified in all of the B.cereus group organisms, but some common genes are absent (Supplementary fig. S4).

*jThe phages are not conserved in sequence or genomic location in the three genomes studied.

*k*bcr1 is an ~160 bp repeated DNA sequence with unknown function (29) overwhelmingly over-represented in intergenic regions of the B.cereus group organisms.

---

**Figure 5.** Bacillus cereus ATCC 10987 urease operon replacement. The gene cluster contains the three structural genes encoding the urease enzyme (*ureCBA;* B3664–B3662), genes for four accessory proteins responsible for insertion of Ni<sup>2+</sup> into the apo-enzyme (*ureEFGD;* B3661–B3658) and genes for two putative nickel transport proteins that are responsible for transporting Ni<sup>2+</sup> into the cell (*mixA and *ureL;* B3656–B3657). The red genes are the structural urease genes, blue are urease accessory proteins, light blue are nickel transport proteins, black are hypothetical, and green are an unknown ABC transporter system. The urease cluster replaces the B.anthracis Ames/B.cereusATCC 14579 genes that encode hypothetical proteins, blasticidin S deaminase and an S layer protein.
experimental evidence can determine if they serve the same function in pathogenesis and germination.

**Gene replacements play a major role in metabolic specialization of *B. cereus* ATCC 10987**

One of the most significant metabolic differences between these three genomes is the presence of a nine-gene urease gene cluster (BCE3656–BCE3664) in *B. cereus* ATCC 10987, the first described in this group of bacteria. The urease gene cluster, consisting of the urease structural (ureABC) and accessory proteins (ureDEFG) as well as two nickel transporters (ureI and nixA), is ordered and oriented in such a way that it may be a transcriptional unit. The gene cluster is part of a larger ~11 kb, 15 gene replacement (Fig. 5). At the corresponding locus, *B. anthracis* Ames and *B. cereus* ATCC 14579 have an ~4.6 kb region containing six potential coding regions of no obvious function (BA3691–BA3696/BC3630–BC3635). The presence of the urease enzyme may increase fitness of *B. cereus* ATCC 10987 in acidic conditions, much the same way that *Helicobacter pylori* urease is required for colonization of the human stomach (53).

Although no regulatory gene has been identified, we have demonstrated that functional urease is produced by *B. cereus* ATCC 10987 by growth and color change of Christensen’s urea agar (54). Additionally, using PCR to screen 23 *B. cereus* group organisms (Supplementary table S1), we have identified the urease genes in one other *B. cereus* isolate (*n* = 10) and five *B. thuringiensis* isolates (*n* = 12), but not in a *B. mycoides* isolate (*n* = 1) (Supplementary table S2).

Another replacement unique to *B. cereus* ATCC 10987 encodes the proteins responsible for the utilization of xylose (27). The five-gene xylose operon (BCE2208–BCE2214) is located on a unique 10.1 kb region that occupies the same relative genomic location as an ~26.6 kb region in *B. anthracis* Ames and *B. cereus* ATCC 14579 (24, 26) (Fig. 6). The *B. anthracis* Ames/*B. cereus* ATCC 14579 region encodes four proteins that are required for nitrate reductase [α, β, δ and γ subunits (BA2125-8/BC2118–BC2121)], a nitrate transporter (BA2139/BC2128), a group of proteins for the synthesis of molybdopterin (BA2133–7/BC2123–7), utilized in nitrogen metabolism, as well as an [NAD(P)H]-requiring nitrite reductase (BA2146–7/BC2136–7). All of the genes contained in the *B. anthracis* Ames/*B. cereus* ATCC 14579 region are absent in *B. cereus* ATCC 10987. The utilization of xylose appears rare as PCR primers specific for the xylose permease (BCE2208) failed to amplify the desired product in any of the other *B. cereus* group bacteria tested (Supplementary table S2).

In *B. subtilis*, the ability to reduce nitrate and nitrite plays a significant role in the energy production under anaerobic or oxygen-limiting conditions, which utilizes nitrate as a terminal electron acceptor during anaerobic respiration (55). *Bacillus subtilis* mutants that lacked a functional nitrate reductase or molybdopterin genes did not survive under anaerobic conditions, but were able to survive in fermentation mixtures containing limited oxygen (56). One other identified nitrite reductase, of a different enzyme class, is present in *B. cereus* ATCC 10987 (BCE1547), but its role in respiration is not clearly delineated. As for nitrogen assimilation, *B. cereus* ATCC 10987 may use ammonia, a breakdown product of urea hydrolysis. The acquisition of the urease operon may be an adaptation to the loss of nitrate and nitrite reduction. Alternatively, the urease activity may allow the organism to survive without the ability to reduce nitrate or nitrite and hence can lose the genes without decreasing fitness.

*Bacillus cereus* ATCC 10987 also contains a 17.9 kb replacement responsible for the transport and utilization of the carbohydrate tagatose (BCE1896–BCE1912). The corresponding 5.0 kb region in *B. anthracis* Ames/*B. cereus* ATCC 14579 contains genes encoding hypothetical proteins of no described function. *Bacillus cereus* ATCC 10987 was isolated from a study on cheese spoilage (27, 28) where this carbohydrate has been found, hence the tagatose gene cluster may be an adaptation to this carbohydrate-containing environment.

**Comparison of the capsule and flagellar variable surface antigens in *B. cereus* group bacteria**

It has been noted that the *B. cereus* serotyping based on variable surface antigens, flagellum and surface polysacchar-
ides generally does not agree with genotyping schemes that are based on conserved chromosomal markers (23,57–59). Comparative analysis of the three genomes suggests that surface antigen variability may be generated via gene replacements, duplications and deletions.

Ivanova et al. (24) described an ~20 kb region of *B. cereus* ATCC 14579 encoding proteins (BC5263–5279) thought to be involved in polysaccharide capsule synthesis that replaces a *B. anthracis* region encoding proteins similar to teichoic acid synthesis proteins (BA5505–20). *Bacillus cereus* ATCC 10987 also contains a polysaccharide capsule gene cluster at this locus spanning ~20 kb and encoding 21 putative proteins (BCE5380–BCE5400). Approximately half of the 20 kb replacement in *B. cereus* ATCC 10987 contains genes similar to the *B. cereus* ATCC 14579 capsule locus (Fig. 7). Conserved regions highlighted in Figure 7 may have allowed for homologous recombination, resulting in the variation observed in this replacement.

None of the 23 *B. cereus* group strain set produced amplicons by PCR using primers designed to the *B. cereus* ATCC 10987 polysaccharide polymerase (BCE5389) or translocase (BCE5386) (Supplementary table S2). Similarly, the *B. cereus* ATCC 14579 polysaccharide polymerase (BC5268) was identified in only one of the 23 strains tested, whereas BC1588, an additional putative polysaccharide polymerase in this strain, was present in 14 strains (Supplementary table S2). These novel genes in the *B. cereus* ATCC 10987 locus may be responsible for a specialized structure of capsule polysaccharide. It is also possible that these genes may influence flagellar structure; homologs of UDP-galactose phosphate transferase (BCE5393) and an aminotransferase family protein (BCE5394) have been shown in *Campylobacter jejuni* to be involved in the glycosylation of flagella (60).

We can find no record of any description of a complex extracellular polysaccharide capsule produced by *B. cereus* (1), yet the formation of complex biofilms is usually associated with the expression of some type of carbohydrate moiety, and *B. cereus* biofilms are a significant problem in the dairy industry (61). The presence of polysaccharide capsule gene clusters in both *B. cereus* isolates sequenced provides evidence that these structures may be important in environments faced by the *B. cereus* group bacteria.

The flagellar antigens of the *B. cereus* group are another highly variable surface structure, with up to 82 groups being described using serological methods (62–64). *Bacillus anthracis* genome analysis revealed that four essential proteins in the flagellar gene cluster contained point mutations and subsequent frameshifts (26) rendering the flagellum non-functional. Lack of motility is often cited as a distinguishing factor between *B. anthracis* and other *B. cereus* group members (65). In contrast, both *B. cereus* genomes sequenced contained genes encoding full-length proteins (CheA BCE1749/BC1628, CheV BCE1777/BC1654, M-ring protein BCE1766/BC1644 and FlIM BCE1783/BC1662).

Another significant difference between the flagellar gene cluster of *B. cereus* and *B. anthracis* is the number of flagellin subunits present. The *B. anthracis* genome contains only one flagellin gene (BA1706), whereas *B. cereus* ATCC 14579 contains four (BC1656–1659) and *B. cereus* ATCC 10987 has two (BCE1779 and BCE1780). Interestingly, the flagellin genes are transcribed opposite to the orientation of the rest of the genes in the flagellar biosynthetic cluster, and in *B. cereus* the multiple copies are clustered together. Amino acid sequence identities of the flagellin proteins of the *B. cereus* group organisms separate these proteins into two groups; one exclusive to *B. cereus* and one that appears common to all three organisms (Supplementary fig. S3). It is possible that the
different B. cereus ATCC 10987 flagellins are expressed under different conditions resulting in structurally, functionally and antigenically variable flagella.

**Conservation of the PlcR regulon in B. cereus group genomes**

PlcR is a pleiotropic transcriptional regulator that recognizes the palindromic sequence, TATGNAN,TNCATA, and has been implicated in the control of a number of virulence factors in B. cereus and B. thuringiensis (66–68). Slamti and Lereclus (69) demonstrated that PlcR activity is regulated by the presence of a secreted and reimported pentapeptide produced from the processing of the PapR protein C-terminus. The papR gene itself is positively regulated by PlcR, forming a quorum sensing-like system. The 48 amino acid PapR proteins of B. cereus ATCC 10987 and B. anthracis Ames are identical and would produce the same regulatory pentapeptide (VPFEY), whereas the B. cereus ATCC 14579 PapR has four amino acid changes, one of which is present in the secreted pentapeptide (LPFEY).

There are 52 putative PlcR-binding motifs in the B. anthracis genome (24), 56 in B. cereus ATCC 14579 (26) and 57 in B. cereus ATCC 10987 which potentially regulate over 100 genes in each isolate. Comparative analysis reveals that there is a conserved core of putative PlcR-regulated proteins present in all three genomes. However, a number of potentially PlcR-regulated proteins are present only in both B. cereus strains, including cytotoxin K (BCE1209), non-hemolytic enterotoxin C subunit (BCE1970), the methyl-accepting chemotaxis protein (BCE0638) and ribonucleotide-diphosphate reductase, whereas an aromatic compound degradation pathway is present only in B. cereus ATCC 10987 (BCE2151–BCE2161).

**DISCUSSION**

Based on synteny (Fig. 2), overall protein and nucleotide similarity (Fig. 3), phylogeny (Fig. 1) and shared novel genes (Fig. 4), B. cereus ATCC 10987 is more closely related to B. anthracis Ames (26) than it is to another dairy-isolated strain, B. cereus ATCC 14579 (27,28). Although this may seem initially a surprising finding, it confirms recent MLTE and other studies that point to the phylogenetic intermingling of species in the B. cereus group (16–18,70). B. cereus ATCC 10987 contains a number of characterized virulence factors such as the non-hemolytic enterotoxin complex, phospholipase C, sphingomyelinase and cytotoxin K, and thus has pathogenic potential. Additionally, the large (~208 kb) plasmid pBc10987 shares a conserved backbone with B. anthracis pXO1 (Fig. 2C), which may contain as yet unidentified conserved plasmid replication and maintenance functions. pBc10987 also has some other intriguing parallels with pXO1, such as the presence of a transition state regulator homolog, AbRB, and the spore coat determinate, Bc1A, both of which have been demonstrated to play a role in pathogenesis.

Although B. cereus ATCC 10987 overall has much genetic similarity to the other two B. cereus group genomes, there are clear differences in gene content that point to metabolic specializations (e.g. xylose utilization and urease genes; Figs 5 and 6) and surface structure variation (capsule and flagellum genes; Fig. 7 and Supplementary fig. S3). Some gene movements appear to have been mediated by insertion of genes via phages or insertion elements [i.e. Tn554 element or novel phage (Table 2)], previously observed in bacteria where several closely related genomes have been sequenced (71–74). However, many replacements in B. cereus ATCC 10987 do not appear to be associated with mobile genetic elements, suggesting that either the insertion has taken place through homologous recombination of flanking DNA or the mobile elements are no longer identifiable due to sequence divergence or deletion. A recent MLST study on 77 B. cereus group organisms demonstrated that recombination in seven housekeeping genes was occurring at a low level in the B. cereus group (18); however, the frequency of horizontal transfer among genes required for adaptation to new environments may well be much higher.

This concept raises a number of intriguing questions that should be subjected to further analysis. How does DNA enter the cell in a natural situation? Why are multiple gene clusters found at similar loci (e.g. xylose genes in B. cereus ATCC 10987 and nitrate reductase genes in B. anthracis and B. cereus ATCC 14579): are these hotspots for recombination? What are the roles of the restriction–modification systems and other potential barriers to the flow of genetic information? Each of the B. cereus group organisms contains a number of unique restriction–modification systems (Table 2). These bacteria can be genetically manipulated by electroporation, transconjugation and other methods with much effort, but may have a natural mechanism for acquiring DNA.

A sample of three genomes of the B. cereus group, common across the globe and adapted to numerous specific environments, is not sufficient to begin to understand the dynamics of genome evolution or to even make any generalized statements with great conviction. For instance, it is interesting to note that when the B. anthracis Ames genome was completed, some traits were labeled as B. anthracis-specific but now with the two B. cereus genome sequences need to be considered as B. cereus group-specific. An example of this is genetic competence: all three B. cereus group organisms lack similar genes that have been shown in B. subtilis to be required for full genetic competence (75,76) (Supplementary fig. S4). Yet are we really sure that these homologs are not present in some, unsequenced, B. cereus group strains or present with low-level homology and currently labeled as hypothetical? These questions can be addressed with techniques such as suppressive subtractive hybridization, plasmid and phage sequencing (which often contain novel, niche-specific genes) and comparative genomic hybridizations using microarrays. Inevitably, however, whole-genome sequencing of key strains in phylogenetically relevant subgroups of B. cereus sensu lato, such as pathogenic B. cereus from periodontal, neonatal or immunocompromised patient sources, is going to be the workhorse of discovery in the near future.

**ACKNOWLEDGEMENTS**

This project was supported in part by a grant from the Office of Naval Research (N00014-96-1-0604) and Federal funds from the National Institute of Allergy and Infectious Disease, National Institutes of Health, under Contract No. N01-AI-15447. O.A.O, E.H. and N.J.T were supported by grants to A.B.K. from the Norwegian Research Council (NRC).
N.J.T. also received support from the European Union TMR programme.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

REFERENCES


