Published online February 11, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 3 ***977–988***
DOI: 10.1093/nar/gkh258

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 Økstad[1,2], Erlendur Helgason[1], Regina Z. Cer, Lingxia Jiang, Kelly A. Shores, Derrick E. Fouts, Nicolas J. Tourasse[1], Samuel V. Angiuoli, James Kolonay, William C. Nelson, Anne-Brit Kolstø[1,2], Claire M. Fraser and Timothy D. Read*

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

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.anthracis* 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.anthra*cis plasmid biology and regulatory cross-talk.

INTRODUCTION

Bacillus cereus, *Bacillus thuringiensis* and *Bacillus anthracis* all belong to the *B.cereus sensu lato* group of rod-shaped, Gram-positive, spore-forming bacteria (1). *Bacillus anthracis* 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

*To whom correspondence should be addressed. Tel: +1 301 984 4973; Fax: +1 301 984 4979; Email: ReadT@NMRC.NAVY.MIL
Present addresses:
Timothy D. Read, Biological Defense Research Directorate, Naval Medical Research Center, 503 Robert Grant Avenue, Silver Spring, MD 20910 USA and Henry M. Jackson Foundation for the Advancement of Military Medicine, 1401 Rockville Pike, Rockville, MD 20851 USA

[+]AE017194 and AE017195

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.

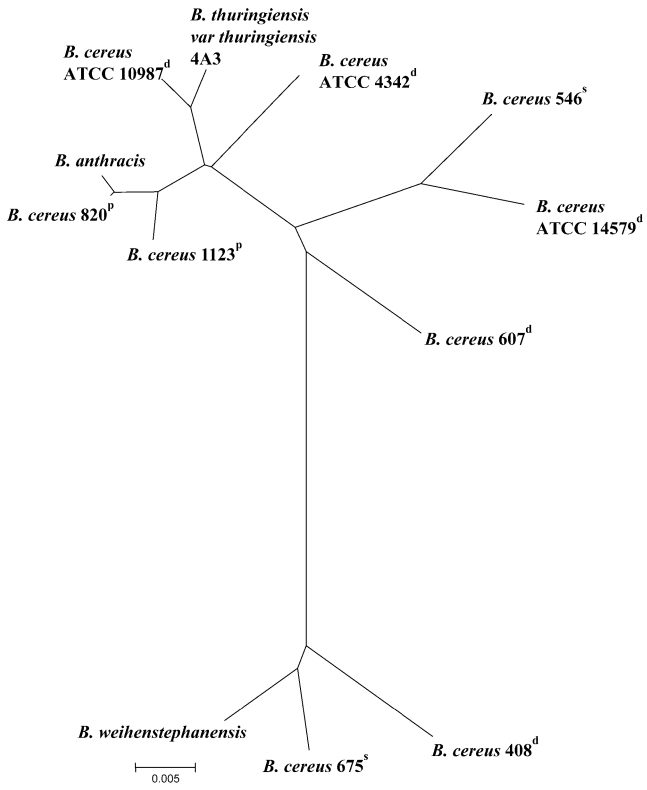

Figure 1. Phylogenetic tree of members of the *B.cereus sensu lato* group. The tree was based on partial nucleotide sequences of seven house-keeping genes totaling 2977 nt (18) and was built using the neighbor-joining method (77) applied to a matrix of pairwise differences among sequences. The scale bar is an average number of nucleotide differences per site. Strains labeled with a 'p' were obtained from human patients, 's' were isolated from soil, and 'd' indicates a dairy origin. All *B.anthracis* strains analyzed ($n = 5$) fall in the same location on the tree and thus no strain has been designated. *Bacillus cereus* clinical isolates were previously described (18,22).

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.

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

Feature	*B.cereus* ATCC 10987		*B.anthracis* Ames[a]			*B.cereus* ATCC 14579[b]	
	Chromosome	Plasmid pBc10987	Chromosome	Plasmid pXO1	Plasmid pXO2	Chromosome	Plasmid pBClin15
Size (bp)	5 224 283	208 369	5 227 293	181 677	94 829	5 426 909	15 100
No. of genes	5642	242	5508	217	113	5366	21
Percentage coding	85	80.9	84.3	77.1	76.2	84	84
Average gene length (nt)	787	696	800	645	639	820	820
%G + C content	35.6	33.5	35.4	32.5	33	35.3	38
rRNA operons	12	0	11	0	0	13	0
tRNAs	98	0	95	0	0	108	0
sRNAs	3	0	3	0	0	3	0
Putative phage	3	0	4	0	0	6	Yes
Genes with assigned function	2802	94	2762	65	38	3839	4
Conserved hypothetical genes[c]	1074	74	1212	22	19	1481	4
Genes of unknown function[c]	672	12	657	8	5	142	6
Hypothetical genes[c]	1094	62	877	122	51	?	11

[a]The genome data for *B.anthracis* are derived from the complete genome sequence of *B.anthracis* Ames (26) and the plasmid sequences of *B.anthracis* str. A2012 (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 frameshifts.

[b]*Bacillus cereus* ATCC 14579 data were obtained from Ivanova *et al.* (24).

[c]For 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.

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 synteny (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 BSRA. 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 BSRA. 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).

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

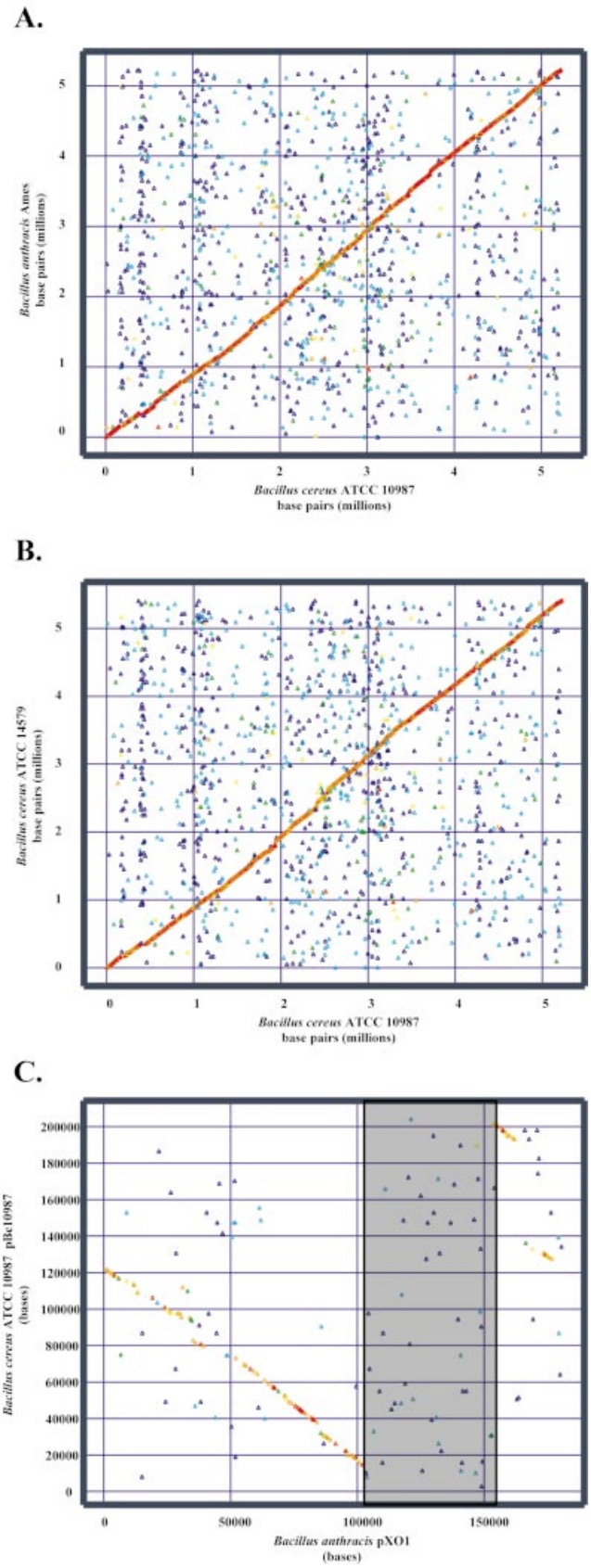

Figure 2. Plot of BSRA based on genomic location. Each triangle indicates a single protein plotted on the 5′ end of the coding region in the reference genome and the best match in the query organism. They are color coded as follows: normalized score ≥0.95, red; normalized score ≥0.80 and <0.95, orange; 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.

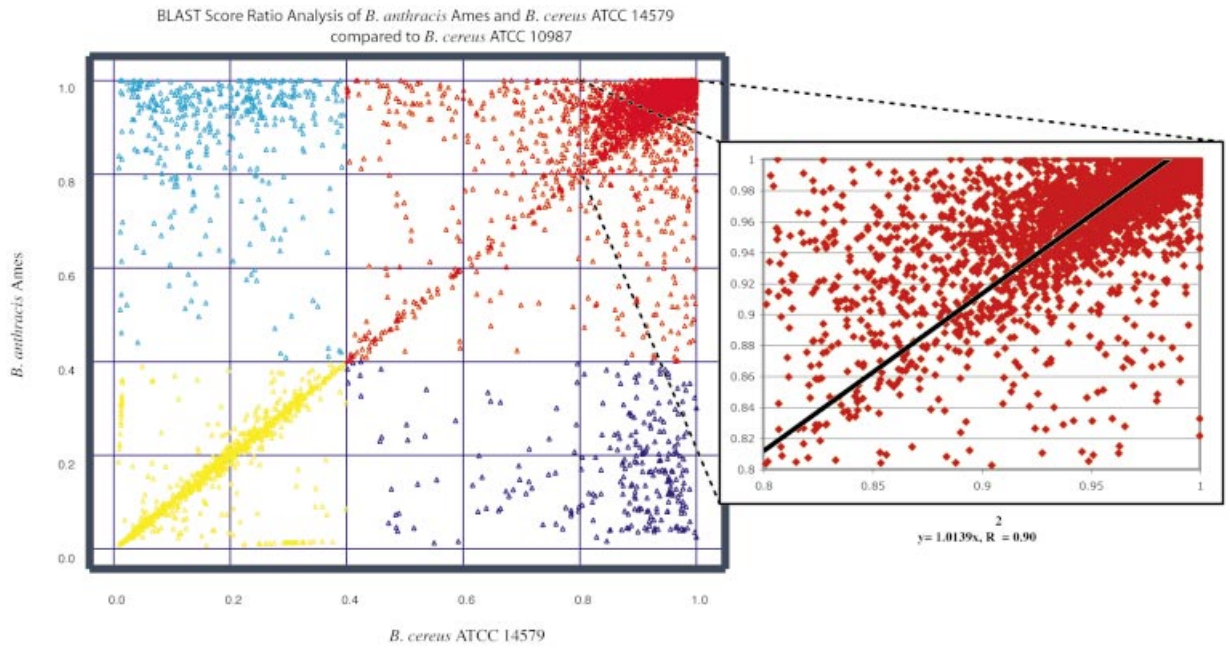

Figure 3. BLAST score ratio analysis between the three genomes. Each point represents a single protein. Proteins that cluster in the upper right hand corner (red) are indicative of coding regions that are similar between all three organisms. Proteins in the lower right (blue) are those that are conserved among *B.cereus* ATCC 10987 and *B.cereus* ATCC 14579 but are not shared with *B.anthracis*; upper left (cyan) are similar between *B.cereus* ATCC 10987 and *B anthracis*; and lower left are unique to *B.cereus* ATCC 10987 (yellow). Proteins that are between these extreme regions are thought to be evolving. The inset evaluates the BLAST score ratio of the genes that are highly similar between the three *B.cereus* group genomes. The slope of the linear regression is 1.0139, which indicates that *B.cereus* ATCC 10987 proteins are generally more similar to *B.anthracis* Ames proteins than *B.cereus* ATCC 14579.

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 Tn*554* element encoded by BCE3147–BCE3149 and BCEA0242, BCEA0001–2. The Tn*554* element is composed of three essential proteins, TnsABC, whose closest relative is the Tn*554* 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 Tn*554* 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 *gerY* 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

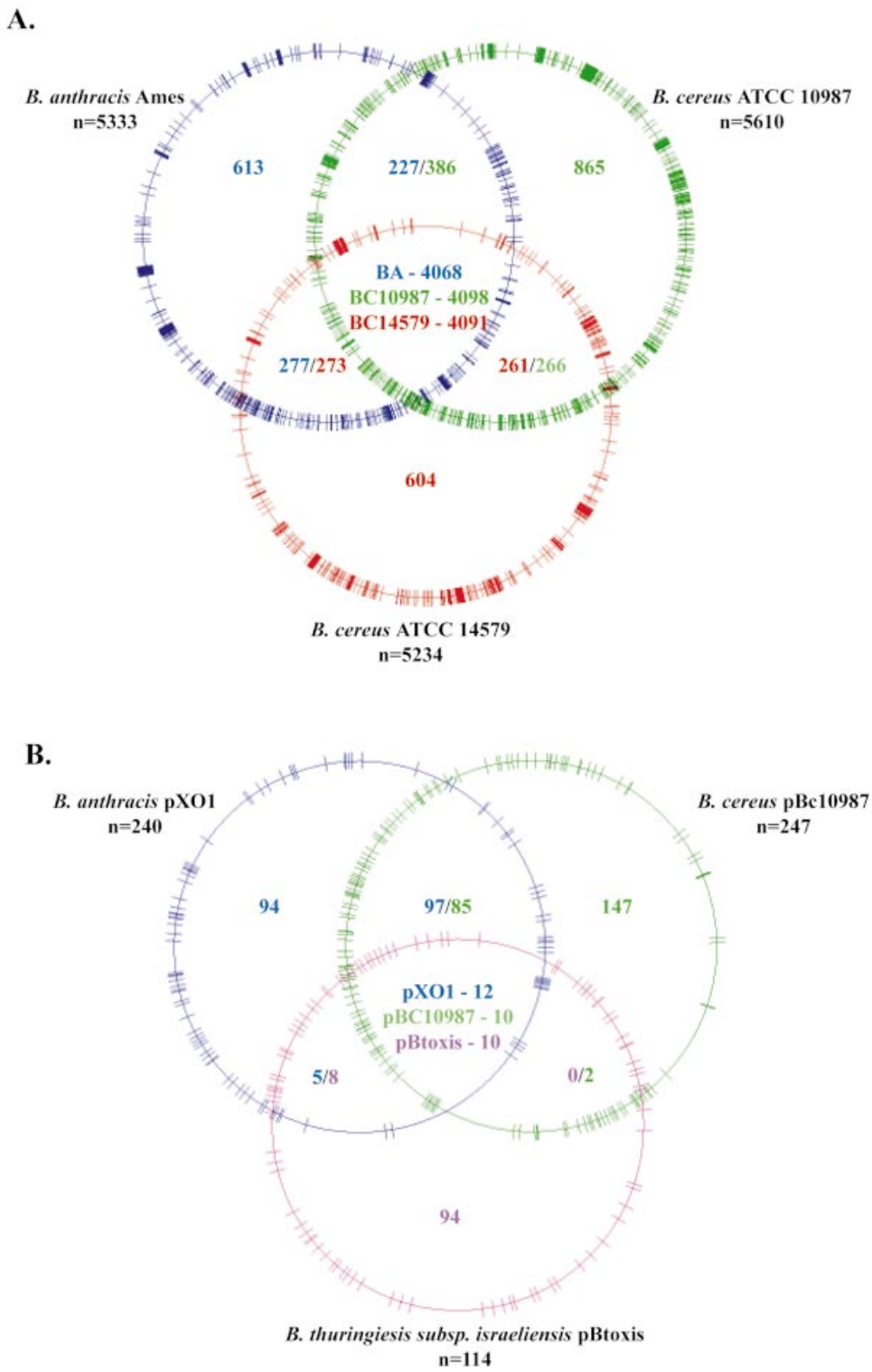

Figure 4. Venn diagram illustrating the number of putative proteins associated with each organism and the number shared with the intersecting organism. Tick marks that are on each circle represent the location of the unique proteins (BLAST score ratio less than 0.4) on the genome. (**A**) Chromosomal 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. *israeliensis* pBtoxis.

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

Table 2. Major phenotypic differences and characteristics of *B.cereus* group genomes

Genotype	*B.anthracis* Ames	*B.cereus* ATCC 10987	*B. cereus* ATCC 14579
Plasmid	pXO1 (189 kb) pXO2 (96 kb)	pBc10987 (208 kb)	Linear phage-like (15 kb) pBClin15
Tripartite lethal toxin	Present	Absent	Absent
rRNA	11 copies	12 copies	13 copies
Urease gene cluster	Absent	Present	Absent
Xylose utilization genes	Absent	Present	Absent
Polysaccharide capsule	Absent	Present[a]	Present[a]
Flagellar genes	Present[b]	Present[c]	Present[c]
Tagatose utilization genes	Absent	Present	Partial[d]
Functional PlcR	Absent[e]	Present[e]	Present[e]
Mobile genetic elements			
Phage[j]	4	4 (1 degenerate)	6 (1 linear plasmid)
Tn*554*	0	2[f]	0
Tn*7*	0	2[g]	0
I5605	7	9	5
I5231	3	5	1
Arginine deiminase genes	Absent	Present	Present
B.cereus repeat 1 (bcr1)[k]	12 copies	72 copies	52 copies
Restriction enzymes	Absent	Present[h]	Present[h]
Competence genes	Present[i]	Present[i]	Present[i]

[a]The capsule biosynthesis region of ~20 kb is partially similar in both *B.cereus* genomes (Fig. 7).
[b]*Bacillus 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.
[c]Each 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).
[d]There 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.
[f]Tn*554* 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.
[g]Two similar but unique Tn*7* elements are located on the chromosome.
[h]There are four unique restriction–modification systems in *B.cereus* ATCC 10987 and three in *B.cereus* ATCC 14579.
[i]Many 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).
[j]The 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.

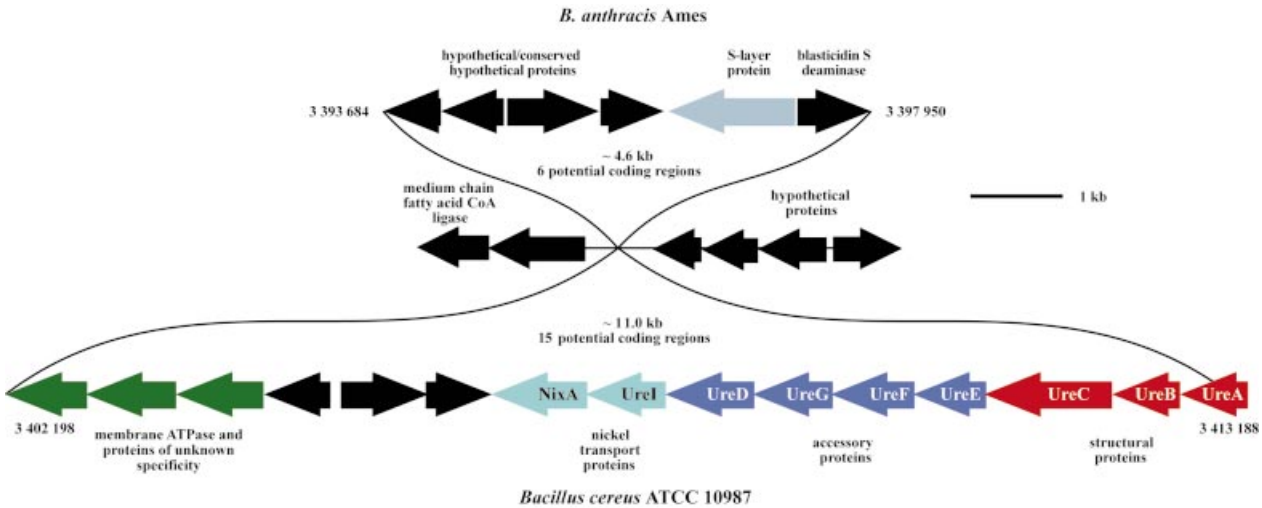

Figure 5. *Bacillus cereus* ATCC 10987 urease operon replacement. The gene cluster contains the three structural genes encoding the urease enzyme (*ureCBA*; BCE3664–BCE3662), genes for four accessory proteins responsible for insertion of Ni^+ into the apo-enzyme (*ureEFGD*; BCE3661–BCE3658) and genes for two putative nickel transport proteins that are responsible for transporting Ni^+ into the cell (*nixA* and *ureI*; BCE3656–BCE3657). 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.cereus*ATCC 14579 genes that encode hypothetical proteins, blasticidin S deaminase and an S layer protein.

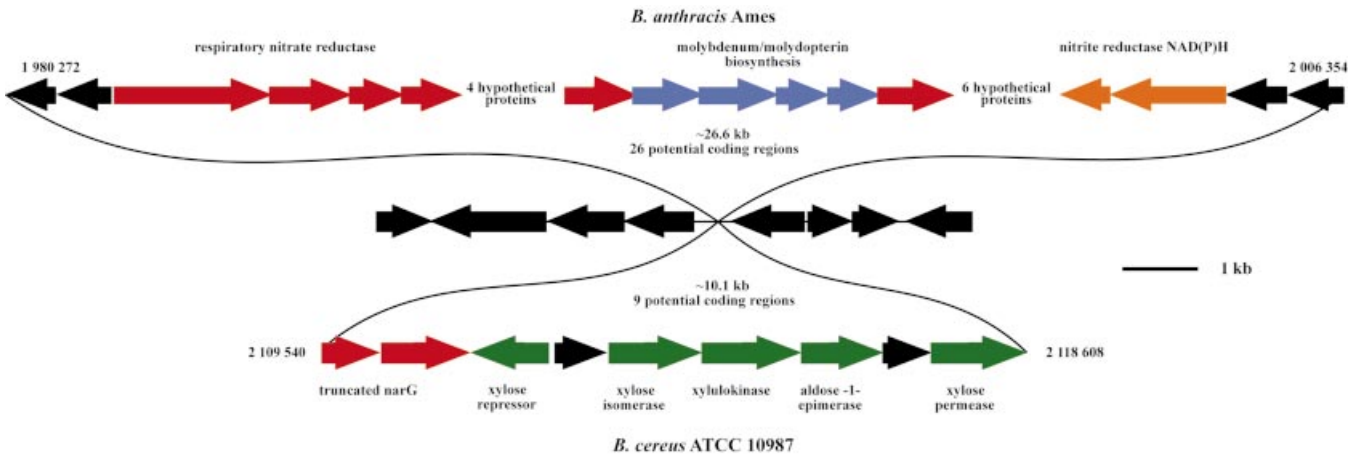

Figure 6. *Bacillus cereus* ATCC 10987 xylose replacement. The xylose utilization operon (green) in *B.cereus* ATCC 10987 replaces the *B.anthracis* Ames/*B.cereus*ATCC 14579 nitrate (red) reductase genes, molybdopterin synthesis genes (blue) and nitrite reductase (orange). Black are hypothetical and conserved hypothetical proteins.

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-

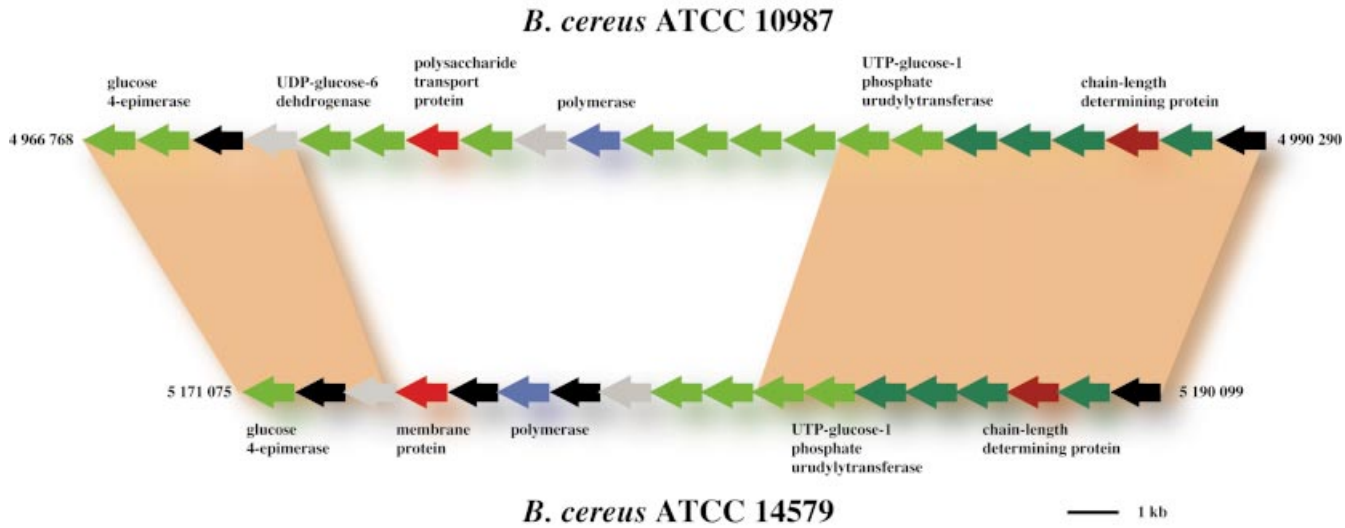

Figure 7. Comparison of the putative *B.cereus* capsule biosynthesis region. The colored areas between the two regions indicate that these regions share a high level of identity (>10^{-5} BLAST, >60% amino acid identity). A conserved region consists of six contiguous genes, including a chain length-determining protein and polysaccharide capsule assembly proteins that are similar to putative polysaccharide capsule biosynthesis operons in *B.subtilis* (75), *Oceanobacillus iheyensis* (78) and *Lactococcus lactis* subsp. *cremoris* (79). However, the central portion of the polysaccharide capsule locus consists of novel genes that encode glycosyltransferases, a polysaccharide polymerase (BCE5389) and a putative translocase (BCE5386). Green arrows are carbohydrate utilization genes; blue, polysaccharide polymerase; dark red, chain length-determining protein; red, translocase (membrane protein); gray, conserved hypothetical; and black, hypothetical.

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$_4$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 MLEE 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, BclA, 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. Tn*554* 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) (Supplemetary 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-AI15447. 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

1. Jensen,G.B., Hansen,B.M., Eilenberg,J. and Mahillon,J. (2003) The hidden lifestyles of *Bacillus cereus* and relatives. *Environ. Microbiol.*, **5**, 631–640.
2. CDC (2001) Update: investigation of anthrax associated with intentional exposure and interim public health guidelines, October 2001. *MMWR Morb. Mortal. Wkly Rep.*, **50**, 889–893.
3. Aronson,A.I. and Shai,Y. (2001) Why *Bacillus thuringiensis* insecticidal toxins are so effective: unique features of their mode of action. *FEMS Microbiol. Lett.*, **195**, 1–8.
4. Chan,W.M., Liu,D.T., Chan,C.K., Chong,K.K. and Lam,D.S. (2003) Infective endophthalmitis caused by *Bacillus cereus* after cataract extraction surgery. *Clin. Infect. Dis.*, **37**, 31–34.
5. Callegan,M.C., Kane,S.T., Cochran,D.C., Gilmore,M.S., Gominet,M. and Lereclus,D. (2003) Relationship of *plcR*-regulated factors to *Bacillus* endophthalmitis virulence. *Infect. Immun.*, **71**, 3116–3124.
6. Kotiranta,A., Lounatmaa,K. and Haapasalo,M. (2000) Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microb. Infect.*, **2**, 189–198.
7. Tokieda,K., Morikawa,Y., Maeyama,K., Mori,K. and Ikeda,K. (1999) Clinical manifestations of *Bacillus cereus* meningitis in newborn infants. *J. Paediatr. Child Health*, **35**, 582–584.
8. Hilliard,N.J., Schelonka,R.L. and Waites,K.B. (2003) *Bacillus cereus* bacteremia in a preterm neonate. *J. Clin. Microbiol.*, **41**, 3441–3444.
9. Ginsburg,A.S., Salazar,L.G., True,L.D. and Disis,M.L. (2003) Fatal *Bacillus cereus* sepsis following resolving neutropenic enterocolitis during the treatment of acute leukemia. *Am. J. Hematol.*, **72**, 204–208.
10. Girisch,M., Ries,M., Zenker,M., Carbon,R., Rauch,R. and Hofbeck,M. (2003) Intestinal perforations in a premature infant caused by *Bacillus cereus*. *Infection*, **31**, 192–193.
11. Miller,J.M., Hair,J.G., Hebert,M., Hebert,L., Roberts,F.J.,Jr and Weyant,R.S. (1997) Fulminating bacteremia and pneumonia due to *Bacillus cereus*. *J. Clin. Microbiol.*, **35**, 504–507.
12. Okinaka,R., Cloud,K., Hampton,O., Hoffmaster,A., Hill,K., Keim,P., Koehler,T., Lamke,G., Kumano,S., Manter,D. *et al.* (1999) Sequence, assembly and analysis of pXO1 and pXO2. *J. Appl. Microbiol.*, **87**, 261–262.
13. Read,T.D., Salzberg,S.L., Pop,M., Shumway,M., Umayam,L., Jiang,L.X., Holtzapple,E., Busch,J.D., Smith,K.L., Schupp,J.M. *et al.* (2002) Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science*, **296**, 2028–2033
14. Hofte,H. and Whiteley,H.R. (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.*, **53**, 242–255.
15. Ash,C., Farrow,J.A., Dorsch,M., Stackebrandt,E. and Collins,M.D. (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.
16. Daffonchio,D., Cherif,A. and Borin,S. (2000) Homoduplex and heteroduplex polymorphisms of the amplified ribosomal 16S–23S internal transcribed spacers describe genetic relationships in the '*Bacillus cereus* group'. *Appl. Environ. Microbiol.*, **66**, 5460–5468.
17. Helgason,E., Økstad,O.A., Caugant,D.A., Johansen,H.A., Fouet,A., Mock,M., Hegna,I. and Kolstø. (2000) *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl. Environ. Microbiol.*, **66**, 2627–2630.
18. Helgason,E., Tourasse,N.J., Meisal,R., Caugant,D.A. and Kolstø,A.-B. (2004) A multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. *Appl. Environ. Microbiol.*, **70**, 191–201.
19. Carlson,C., Caugant,D. and Kolstø,A. (1994) Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. *Appl. Environ. Microbiol.*, **60**, 1719–1725.
20. Ticknor,L.O., Kolstø,A.B., Hill,K.K., Keim,P., Laker,M.T., Tonks,M. and Jackson,P.J. (2001) Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. *Appl. Environ. Microbiol.*, **67**, 4863–4873.
21. Radnedge,L., Agron,P.G., Hill,K.K., Jackson,P.J., Ticknor,L.O., Keim,P. and Andersen,G.L. (2003) Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*. *Appl. Environ. Microbiol.*, **69**, 2755–2764.
22. Helgason,E., Caugant,D.A., Olsen,I. and Kolstø,A.B. (2000) Genetic structure of populations of *Bacillus cereus* and *B.thuringiensis* isolates associated with periodontitis and other human infections. *J. Clin. Microbiol.*, **38**, 1615–1622.
23. Helgason,E., Caugant,D.A., Lecadet,M.M., Chen,Y., Mahillon,J., Lövgren,A., Hegna,I., Kvaloy,K. and Kolstø,A.B. (1998) Genetic diversity of *Bacillus cereus/B.thuringiensis* isolates from natural sources. *Curr. Microbiol.*, **37**, 80–87.
24. Ivanova,N., Sorokin,A. anderson,I., Galleron,N., Candelon,B., Kapatral,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A. *et al.* (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature*, **423**, 87–91.
25. Økstad,O.A., Gominet,M., Purnelle,B., Rose,M., Lereclus,D. and Kolstø,A.B. (1999) Sequence analysis of three *Bacillus cereus* loci carrying PlcR-regulated genes encoding degradative enzymes and enterotoxin. *Microbiology*, **145**, 3129–3138.
26. Read,T.D., Peterson,S.N., Tourasse,N., Baillie,L.W., Paulsen,I.T., Nelson,K.E., Tettelin,H., Fouts,D.E., Eisen,J.A., Gill,S.R. *et al.* (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature*, **423**, 81–86.
27. Smith,N. (1952) Aerobic spore forming bacteria. *US Dep. Agric. Monogr.*, **16**, 1–148.
28. Herron,W.M. (1930) Rancidity in cheddar cheese. Master's Thesis, Queen's University, Kingston, Ontario, Canada.
29. Økstad,O.A., Hegna,I., Lindbäck,T., Rishovd,A.L. and Kolstø,A.B. (1999) Genome organization is not conserved between *Bacillus cereus* and *Bacillus subtilis*. *Microbiology*, **145**, 621–631.
30. Lindbäck,T., Økstad,O.A., Rishovd,A.L. and Kolstø,A.B. (1999) Insertional inactivation of *hblC* encoding the L2 component of *Bacillus cereus* ATCC 14579 haemolysin BL strongly reduces enterotoxigenic activity, but not the haemolytic activity against human erythrocytes. *Microbiology*, **145**, 3139–3146.
31. Salzberg,S.L., Delcher,A.L., Kasif,S. and White,O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.*, **26**, 544–548.
32. Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J. *et al.* (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science*, **293**, 498–506.
33. Read,T.D., Myers,G.S., Brunham,R.C., Nelson,W.C., Paulsen,I.T., Heidelberg,J., Holtzapple,E., Khouri,H., Federova,N.B., Carty,H.A. *et al.* (2003) Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. *Nucleic Acids Res.*, **31**, 2134–2147.
34. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
35. Delcher,A.L., Kasif,S., Fleischmann,R.D., Peterson,J., White,O. and Salzberg,S.L. (1999) Alignment of whole genomes. *Nucleic Acids Res.*, **27**, 2369–2376
36. Berry,C., O'Neil,S., Ben-Dov,E., Jones,A.F., Murphy,L., Quail,M.A., Holden,M.T.G., Harris,D., Zaritsky,A. and Parkhill,J. (2002) Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp *israelensis*. *Appl. Environ. Microbiol.*, **68**, 5082–5095.
37. Wilcks,A., Smidt,L., Okstad,O.A., Kolsto,A.B., Mahillon,J. and Andrup,L. (1999) Replication mechanism and sequence analysis of the replicon of pAW63, a conjugative plasmid from *Bacillus thuringiensis*. *J. Bacteriol.*, **181**, 3193–3200.
38. Andrup,L., Jensen,G.B., Wilcks,A., Smidt,L., Hoflack,L. and Mahillon,J. (2003) The patchwork nature of rolling-circle plasmids: comparison of six plasmids from two distinct *Bacillus thuringiensis* serotypes. *Plasmid*, **49**, 205–232.
39. Okinaka,R.T., Cloud,K., Hampton,O., Hoffmaster,A.R., Hill,K.K., Keim,P., Koehler,T.M., Lamke,G., Kumano,S., Mahillon,J. *et al.* (1999) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J. Bacteriol.*, **181**, 6509–6515.
40. Andrup,L., Jorgensen,O., Wilcks,A., Smidt,L. and Jensen,G.B. (1996) Mobilization of 'nonmobilizable' plasmids by the aggregation-mediated conjugation system of *Bacillus thuringiensis*. *Plasmid*, **36**, 75–85.

41. Fouet,A., Sirard,J.C. and Mock,M. (1994) *Bacillus anthracis* pXO1 virulence plasmid encodes a type 1 DNA topoisomerase. *Mol. Microbiol.*, **11**, 471–479.
42. Ason,B., Handayani,R., Williams,C.R., Bertram,J.G., Hingorani,M.M., O'Donnell,M., Goodman,M.F. and Bloom,L.B. (2003) Mechanism of loading the *Escherichia coli* DNA polymerase III beta sliding clamp on DNA. Bona fide primer/templates preferentially trigger the gamma complex to hydrolyze ATP and load the clamp. *J. Biol. Chem.*, **278**, 10033–10040.
43. Saile,E. and Koehler,T.M. (2002) Control of anthrax toxin gene expression by the transition state regulator abrB. *J. Bacteriol.*, **184**, 370–380.
44. Hamon,M.A. and Lazazzera,B.A. (2001) The sporulation transcription factor Spo0A is required for biofilm development in *Bacillus subtilis*. *Mol. Microbiol.*, **42**, 1199–1209.
45. Koehler,T.M. (2002) *Bacillus anthracis* genetics and virulence gene regulation. *Curr. Top. Microbiol. Immunol.*, **271**, 143–164.
46. Bastos,M.C. and Murphy,E. (1988) Transposon Tn554 encodes three products required for transposition. *EMBO J.*, **7**, 2935–2941.
47. Todd,S.J., Moir,A.J.G., Johnson,M.J. and Moir,A. (2003) Genes of *Bacillus cereus* and *Bacillus anthracis* encoding proteins of the exosporium. *J. Bacteriol.*, **185**, 3373–3378.
48. Sylvestre,P., Couture-Tosi,E. and Mock,M. (2002) A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Mol. Microbiol.*, **45**, 169–178.
49. Steichen,C., Chen,P., Kearney,J.F. and Turnbough,C.L. (2003) Identification of the immunodominant protein and other proteins of the *Bacillus anthracis* exosporium. *J. Bacteriol.*, **185**, 1903–1910.
50. Sylvestre,P., Couture-Tosi,E. and Mock,M. (2003) Polymorphism in the collagen-like region of the *Bacillus anthracis* BclA protein leads to variation in exosporium filament length. *J. Bacteriol.*, **185**, 1555–1563.
51. Guidi-Rontani,C., Pereira,Y., Ruffie,S., Sirard,J.C., Weber-Levy,M. and Mock,M. (1999) Identification and characterization of a germination operon on the virulence plasmid pXO1 of *Bacillus anthracis*. *Mol. Microbiol.*, **33**, 407–414.
52. Barlass,P.J., Houston,C.W., Clements,M.O. and Moir,A. (2002) Germination of *Bacillus cereus* spores in response to L-alanine and to inosine: the roles of *gerL* and *gerQ* operons. *Microbiology*, **148**, 2089–2095.
53. Tsuda,M., Karita,M., Morshed,M.G., Okita,K. and Nakazawa,T. (1994) A urease-negative mutant of *Helicobacter pylori* constructed by allelic exchange mutagenesis lacks the ability to colonize the nude mouse stomach. *Infect. Immun.*, **62**, 3586–3589.
54. Qadri,S.M., Zubairi,S., Hawley,H.P. and Ramirez,E.G. (1984) Simple spot test for rapid detection of urease activity. *J. Clin. Microbiol.*, **20**, 1198–1199.
55. Hoffmann,T., Troup,B., Szabo,A., Hungerer,C. and Jahn,D. (1995) The anaerobic life of *Bacillus subtilis*: cloning of the genes encoding the respiratory nitrate reductase system. *FEMS Microbiol. Lett.*, **131**, 219–225.
56. Nakano,M.M., Dailly,Y.P., Zuber,P. and Clark,D.P. (1997) Characterization of anaerobic fermentative growth of *Bacillus subtilis*: identification of fermentation end products and genes required for growth. *J. Bacteriol.*, **179**, 6749–6755.
57. Myerowitz,R.L., Gordon,R.E. and Robbins,J.B. (1973) Polysaccharides of the genus *Bacillus* cross-reactive with the capsular polysaccharides of *Diplococcus pneumoniae* type 3, *Haemophilus influenzae* type b and *Neisseria meningitidis* group A. *Infect. Immun.*, **8**, 896–900.
58. Gilbert,R.J. and Parry,J.M. (1977) Serotypes of *Bacillus cereus* from outbreaks of food poisoning and from routine foods. *J. Hyg. (Lond.)*, **78**, 69–74.
59. Taylor,A.J. and Gilbert,R.J. (1975) *Bacillus cereus* food poisoning: a provisional serotyping scheme. *J. Med. Microbiol.*, **8**, 543–550.
60. Szymanski,C.M., Yao,R., Ewing,C.P., Trust,T.J. and Guerry,P. (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.*, **32**, 1022–1030.
61. Oosthuizen,M.C., Steyn,B., Theron,J., Cosette,P., Lindsay,D., von Holy,A. and Brozel,V.S. (2002) Proteomic analysis reveals differential protein expression by *Bacillus cereus* during biofilm formation. *Appl. Environ. Microbiol.*, **68**, 2770–2780.
62. Lecadet,M.M., Frachon,E., Dumanoir,V.C., Ripouteau,H., Hamon,S., Laurent,P. and Thiery,I. (1999) Updating the H-antigen classification of *Bacillus thuringiensis*. *J. Appl. Microbiol.*, **86**, 660–672.
63. Mikami,T., Hiraoka,K., Murakami,T., Boon-Long,J., Matsumoto,T. and Suzuki,M. (1990) Detection of common flagella antigen in *Bacillus cereus* by monoclonal antibody. *Microbiol. Immunol.*, **34**, 709–714.
64. Murakami,T., Hiraoka,K., Mikami,T., Matsumoto,T., Katagiri,S., Shinagawa,K. and Suzuki,M. (1993) Analysis of common antigen of flagella in *Bacillus cereus* and *Bacillus thuringiensis*. *FEMS Microbiol. Lett.*, **107**, 179–183.
65. Turnbull,P.C. (1999) Definitive identification of *Bacillus anthracis*—a review. *J. Appl. Microbiol.*, **87**, 237–240.
66. Lereclus,D., Agaisse,H., Gominet,M., Salamitou,S. and Sanchis,V. (1996) Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase C gene at the onset of the stationary phase. *J. Bacteriol.*, **178**, 2749–2756.
67. Gohar,M., Økstad,O.A., Gilois,N., Sanchis,V., Kolstø,A.B. and Lereclus,D. (2002) Two-dimensional electrophoresis analysis of the extracellular proteome of *Bacillus cereus* reveals the importance of the PlcR regulon. *Proteomics*, **2**, 784–791.
68. Agaisse,H., Gominet,M., Økstad,O.A., Kolstø,A.B. and Lereclus,D. (1999) PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol. Microbiol.*, **32**, 1043–1053.
69. Slamti,L. and Lereclus,D. (2002) A cell–cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. *EMBO J.*, **21**, 4550–4559.
70. Jackson,P.J., Walthers,E.A., Kalif,A.S., Richmond,K.L., Adair,D.,M., Hill,K.K., Kuske,C.R., Anderson,G.L., Wilson,K.H., Hugh-Jones,M. *et al.* (1997) Characterization of the variable-number tandem repeats in *vrrA* from different *Bacillus anthracis* isolates. *Appl. Environ. Microbiol.*, **63**, 1400–1405.
71. Welch,R.A., Burland,V., Plunkett,G., 3rd, Redford,P., Roesch,P., Rasko,D., Buckles,E.L., Liou,S.R., Boutin,A., Hackett,J. *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **99**, 17020–17024.
72. Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C., Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L. *et al.* (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, **397**, 176–180.
73. Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A., Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I., Read,T.D. *et al.* (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc. Natl Acad. Sci. USA*, **99**, 12391–12396.
74. Paulsen,I.T., Banerjei,L., Myers,G.S., Nelson,K.E., Seshadri,R., Read,T.D., Fouts,D.E., Eisen,J.A., Gill,S.R., Heidelberg,J.F. *et al.* (2003) Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science*, **299**, 2071–2074.
75. Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G., Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S. *et al.* (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, **390**, 249–256.
76. Hamoen,L.W., Venema,G. and Kuipers,O.P. (2003) Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology*, **149**, 9–17.
77. Saitou,N. and Nei,M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
78. Takami,H., Takaki,Y. and Uchiyama,I. (2002) Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res.*, **30**, 3927–3935.
79. van Kranenburg,R., van,S., II, Marugg,J.D., Kleerebezem,M. and de Vos,W.M. (1999) Exopolysaccharide biosynthesis in *Lactococcus lactis* NIZO B40: functional analysis of the glycosyltransferase genes involved in synthesis of the polysaccharide backbone. *J. Bacteriol.*, **181**, 338–340.