Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Comparative genome analysis of Bacillus cereus group genomes with Bacillus subtilis

Permalink

https://escholarship.org/uc/item/1nk4b6nf

Authors

Anderson, Iain
Sorokin, Alexei
Kapatral, Vinayak
et al.

Publication Date

2005-09-14

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

LBNL-58211

Comparative genome analysis of *Bacillus cereus* group genomes with *Bacillus subtilis*

Iain Anderson[1], Alexei Sorokin[2], Vinayak Kapatral[1*], Gary Reznik[‡], Anamitra Bhattacharya[1] Natalia Mikhailova[1], Henry Burd[1], Victor Joukov[1], Denis Kaznadzey[1], Theresa Walunas[1], MarkD'Souza[1], Niels Larsen[¶], Gordon Pusch[1], Konstantinos Liolios[1], Yuri Grechkin[1], Alla Lapidus[††], Eugene Goltsman[††], Lien Chu[1], Michael Fonstein[#], S. Dusko Ehrlich[2], Ross Overbeek[#], Nikos Kyrpides[††], and Natalia Ivanova[††]

[1] Integrated Genomics, 2201 W. Campbell Park Dr., Chicago, Illinois 60612, [2] Génétique Microbienne, CRJ INRA, 78352 Jouy en Josas cedex, France.

* To whom correspondence should be addressed.

Present address: Life Sciences Operation, IITRI, Chicago, Illinois 60616 [‡]
Present address: Aarhus University, Hoegh Guldbergsgade 10, DK 8000 Aarhus C, Denmark [¶]
Present address: Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598 [††]
Present address: Fellowship for Interpretation of Genomes, 15 W
155 81st Street, Burr Ridge, IL 60527 [#]

Address: Vinayak Kapatral
Integrated Genomics Inc
2201. W. Campbell Park Dr Chicago Il 60612.
Phone: 312-491-0846x 326.
Fax: 312-226-9446

Abstract

Genome features of the *Bacillus cereus* group genomes (representative strains of *Bacillus cereus, Bacillus anthracis* and *Bacillus thuringiensis sub spp israelensis*) were analyzed and compared with the *Bacillus subtilis* genome. A core set of 1,381 protein families among the four *Bacillus* genomes, with an additional set of 933 families common to the *B. cereus* group, was identified. Differences in signal transduction pathways, membrane transporters, cell surface structures, cell wall, and S-layer proteins suggesting differences in their phenotype were identified. The *B. cereus* group has signal transduction systems including a tyrosine kinase related to two-component system histidine kinases from *B. subtilis*. A model for regulation of the stress responsive sigma factor σ^B in the *B. cereus* group different from the well studied regulation in *B. subtilis* has been proposed. Despite a high degree of chromosomal synteny among these genomes, significant differences in cell wall and spore coat proteins that contribute to the survival and adaptation in specific hosts has been identified.

Contact: vinayak@integratedgenomics.com.

Key words: Comparative genomics, *Bacillus cereus* group, *B. subtilis*,

Introduction

The *Bacillus cereus* group of bacilli includes *B. anthracis* (causes anthrax in humans and cattle), *B. cereus* (soil borne and food pathogen), *B. thuringiensis* (lepidopteron insect pathogen), *B. mycoides, B. pseudomycoides* and *B. weihenstephanensis*. Although there is significant chromosomal synteny among the *B. cereus* group of genomes, recent studies [1] have demonstrated differences in gene order, chromosomal rearrangements, nucleotide variations, and remnant phages. Phylogenetic analyses have suggested that *B. anthracis* recently diverged from *B. cereus* and *B. thuringiensis* and represents a distinct genetic lineage [1]. Plasmid encoded genes often play significant roles in pathogenesis in these bacteria. Virulence genes encoded by the plasmids of *B. thuringiensis* and *B. anthracis* are well studied, but the role of chromosomal genes in in host adaptation and pathogenesis is less known [1, 2]. The availability of genome sequences of members of the *B. cereus* group include *B. anthracis* A2012 [3], *B. anthracis* Ames [4], and *B. cereus* ATCC 14579 [5], allowing identification of unique metabolism, comparative physiology, sporulation and virulence.

The *B. thuringiensis* bacteria have several sub-species which are classified based on flagellar serotypes and host range [2]. They are widely used in effective biological control of mosquitoes, including those carrying malaria, yellow fever, dengue fever, etc [6]. The *B. thuringiensis* subspecies *isrealensis* genome was used as a representative strain of the *B. thuringiensis* subspecies for comparative analysis. The human pathogenic isolate *B. thuringiensis* serovar Konkukian strain 97-27 sequence was not used in this study (Unpublished; Acc. # NC005957). In addition, the genome sequence of *B. subtilis* was also included in this comparative analysis. Genome sequences of the facultative anaerobe *B. licheniformis*, belonging to the *B. subtilis* group, were not included in this study [7]. Comparative functional analyses allow determination of conserved and unique genes of these closely related bacteria. The unique gene occurrences in each of these species suggest gene sharing by horizontal transfer for host-adaptation and cell metabolism.

Materials and Methods

Bioinformatics tools were used to identify genes and gene families of *B. cereus* group bacteria. Genome sequences of *B. subtilis* 168 (Acc. # NC_000964), *B. anthracis* A2012 (Acc. # NC_003995) and *B. anthracis* Ames (Acc. # AE_016879) and *B. cereus* ATCC 14579, along with draft genome sequence of *B. thuringiensis* were used as representative species for

comparative analysis. *B. thuringiensis* strain ATCC 35646 was obtained from the American Type Culture Collection (Manassas, VA) was used for sequencing. Total DNA was isolated by standard procedure and sheared by sonication into fragments of ~ca 2-3 kb and cloned into plasmid pGEM3 (Promega, Madison, WI) and were maintained in *E. coli* DH5α. End sequencing of plasmids was performed using Applied Biosystems 3700 (PerkinElmer, Foster City, CA) and MegaBACE (Amersham Biosciences, Sunnyvale, CA) DNA sequencers. A total of 67,278 sequencing reactions were performed (~2.5 kb inserts), and 63, 836 reactions of these were assembled into larger contigs. Base calling and sequence assembly were carried out with Phred/Phrap using default parameters. Genome coverage, based on the sequenced DNA, was found to be approximately ~6.2-fold. Contigs smaller than 1500 bp were not included in the functional analysis. Genes were identified with a combination of CRITICA and software developed at Integrated Genomics, as previously described [8, 9, 10]. The *B. thuringiensis* draft genome sequence has been submitted to Genbank under the accession number. The *B. thuringiensis* genome sequence along with annotations is available online at http:// www. ergolight.com.

Results and Discussion

Global genome comparisons. The comparative genome features of *B. thuringiensis*, *B. anthracis*, *B. cereus*, and *B. subtilis* are presented in Table 1. In general, the *B. cereus* group genomes are about 25% larger than the *B. subtilis* genome and have lower GC content. The genomes differ in the total number of nucleotides and size of extra-chromosomal elements, with the exception of *B. cereus* which has one small phage-like element on a linear contig [5]. The two plasmids carrying virulence genes of *B. anthracis*, pX01 and pX02, contribute 276, 500 bp (5%) of the total DNA sequence. In previous studies, *B. thuringiensis* was shown to have eight plasmids and one linear plasmid-like element with a total of 630, 000 bp (10%) of the total DNA sequence [11]. Several plasmids from the *B. thuringiensis* genome have already been sequenced: the toxin-carrying plasmid pBtoxis [12], pTX141 (unpublished, Acc. # NC_002091), pTX142 (unpublished, Acc. # NC _ 0043 34), and pTX143 [13]. The *B. cereus* linear plasmid pBClin15 contains 21 putative coding sequences (CDSs) [5]. The gapped *B. thuringiensis genome* contains a contig with 15 CDSs similar to those present in pBClin15, suggesting a close relationship with the linear extra chromosomal DNA

of *B. cereus*.

In the *B. cereus* and *B. anthracis* A2012 genomes, 71% of the CDSs were assigned with function compared to 69% in the *B. anthracis* Ames strain and *B. thuringiensis* genomes. The number of predicted CDSs in each genome with no sequence similarity to other genes (unique genes) in the ERGO database varied, with 1.7% in *B. cereus*, 3.0% in *B. anthracis* A2012, and 6.8% in *B. thuringiensis* genomes. The disparity is due to the difference in DNA content of plasmids and prophages, which often have higher percentages of CDSs without similarity to known proteins. CDSs with unknown (hypothetical proteins) functions varied from 24-29% in these genomes. In general, the functional categories between the *B. cereus* group and *B. subtilis* are similar in all genomes except for information processing, signal transduction, virulence and transport subsystems. Genes belonging to core metabolism, amino acids, lipid, nitrogen, phosphorus, and sulfur metabolism, and chemotaxis, did not differ significantly between the *B. cereus* group and *B. subtilis* genomes.

Protein clusters. The protein clusters between *B. thuringiensis, B. cereus*, *B. anthracis* A2012 and *B. subtilis* genomes were calculated using the protein clustering WorkBench tool with a cut-off score of 10^{e-20}. Each cluster refers to the number of protein families present in each group of genomes. A combined total of 5,896 clusters, with a core of 1,381 proteins clusters common to all four genomes, were identified. Among the *B. cereus* group, an additional 933 clusters were identified. *B. cereus* had 291 unique clusters whereas *B. anthracis* and *B. thuringiensis* had 606 and 940 clusters, respectively. Similarly, within the *B. cereus* group genomes (*B. thuringiensis* , *B. cereus*, *B. anthracis*), a total of 5,092 families were identified of which 2,411 were common to all three organisms, while each individual organism contained a substantial number of protein families not found in the other (Figure 1).

Comparative signal transduction system analysis Bacteria living in different environments use both chemical and physical cues to regulate metabolism, development, and stress responses. Signal transduction systems, extra-cytoplasmic function (ECF) sigma factors, regulators of sporulation and the σ^B stress response sigma factor were studied in the *B. cereus* group genomes. A comparison of signal transduction proteins is given in Table 2. Rap family aspartate phosphatases inhibit the action of response regulators such as Spo0F, and many of them are regulated by secreted Phr peptides [14]. The *B. subtilis* genome encodes 11 Rap family aspartate phosphatases and 7 Phr peptides [15], while the members in the *B. cereus*

group genomes have fewer genes for both phosphatases and Phr peptides. In the *B. thuringiensis* genome sequence, an N-terminal fragment of a Rap family phosphatase was found on one contig whereas the C-terminal fragment was found encoded on another contig. In *B. subtilis,* a family of aspartate phosphatases induced by different environmental condition (Spo0E, YnzD, YisI proteins) inhibits sporulation by interacting with the Spo0A transcription factor [16]. The *B. cereus* group genomes contain a varying number of Spo0E-related phosphatases, ranging from three in *B. anthracis* to six in *B. thuringiensis*. One of the *B. thuringiensis* phosphatases (BTH08314) has no homolog in either *B. cereus* or *B. anthracis*, whereas two others are found in *B. thuringiensis* and *B. cereus*, but neither is in *B. anthracis*. One of the most striking features is the presence of a large number of orphan histidine kinases in the *B. cereus* group genomes. In the *B. subtilis* genome, five of the six orphan histidine kinases are involved in sporulation initiation [17], but in the *B. cereus* group the larger number of orphan kinases suggests their role in other functions. Some may be involved in a proposed new model of regulation of the stress response sigma factor σ^B. There are more than 120 σ^B-regulated genes in *B. subtilis*, including the stress induced ribosomal Ctc protein [18] and glucose starvation inducible GsiB protein. Both genes *ctc* and *gsiB* are absent from the *B. cereus* group genomes. During unstressed conditions, σ^B is bound to the anti-sigma factor RsbW, while the anti-anti-sigma factor RsbV is phosphorylated and unable to bind RsbW (Figure 2A) [19]. Metabolic and environmental stresses activate two distinct phosphatases, RsbP and RsbU [20], which dephosphorylate RsbV, allowing it to bind RsbW and free σ^B to activate stress-related genes. RsbP is thought to sense metabolic stress directly through a PAS domain (<u>P</u>er: period circadian protein, <u>A</u>rnt; Ah receptor nuclear translocator protein, <u>S</u>im; single-minded protein), and the RsbU phosphatase activity is regulated by a cascade of factors including RsbX, RsbT, RsbS, RsbR, and a family of RsbR--related proteins. While the *sigB* operon in *B. subtilis* contains many of these regulatory factors, in *B. anthracis, B. thuringiensis* and *B. cereus* only the genes for RsbV, RsbW and σ^B were found [21]. The *B. cereus* group organisms lack all other components of the σ^B regulatory pathways from *B. subtilis*. Instead, they possess one phosphatase (BA_1562 in *B. anthracis* A2012) distantly related to RsbP and RsbU, which contains a response regulator receiver domain at its N-terminus, suggesting a distinct regulatory mechanism for σ^B in the *B. cereus* group. We propose that the phosphatase activity depends

on the phosphorylated state of the N-terminal receiver domain, which could be phosphorylated by various orphan histidine kinases (Figure 2B). One of the orphan sensor kinases found in the three *B. cereus* group genomes appears to be an auto-phosphorylating tyrosine kinase rather than a histidine kinase. Sequence alignment of this putative tyrosine kinase with the closely related histidine kinase is shown in Figure 3. The 50 amino acids region surrounding the phosphorylated histidine residue is highly conserved (41% amino acid identity), except that histidine is replaced by tyrosine. Two copies of a histidine kinase fused to a response regulator were found in the *B. cereus* group genomes but not in the *B. subtilis* genome [17]. One of them is found in all three genomes but it lacks a DNA-binding domain.

A unique orphan kinase was found in the *B. cereus* genome (BC5455). Among the *B. cereus* group genomes, only *B. thuringiensis* has an ortholog pair for ComP-ComA, a two-component system similar to that found in *B. subtilis*. Four two-component systems were found in the *B. anthracis* genome are absent from the *B. thuringiensis and B. cereus* genomes. One CDS is adjacent to an ABC transporter (BA_43704373) highly similar to those involved in immunity to the lantibiotics subtilin and mutacin. These groups of genes probably play a role in resistance to antimicrobial peptides, but whether *B. anthracis* produces the corresponding peptide is not known. Other genes (BA_35153516) similar to the *fsrA* and *fsrC* virulence regulators, which control the expression of gelatinase and serine protease in *Enterococcus faecalis* [22], have been identified, but the gene similar to *fsrB* was not found. A *B. subtilis* type two-component system *ykoGH* was found in all *B. cereus* group genomes.

The other group of proteins involved in response to extracellular signals leading to changes in transcriptional regulation is the ECF sigma factors. These are more prevalent in the *B. cereus* group genomes than in the *B. subtilis* genome. They regulate many physiological functions including stress responses, cell wall modification, drug resistance and iron transport. Often they are encoded in an operon with an anti-sigma factor, and many anti-sigma factors are membrane-spanning proteins with an extracellular sensing domain and an intracellular sigma factor binding/inactivation domain. *B. anthracis* has a large number of ECF-type sigma factors compared to the *B. cereus* and *B. thuringiensis genomes* (Table 2). Of the ECF sigma factors found in the *B. subtilis* genome, the salt response sigma factor σ^M [23] has orthologs in the *B. cereus* group genomes. The two CDSs downstream of *sigM*, *yhdL* and *yhdK*, are involved in regulation of σ^M activity. The *yhdL* gene was found in all the *B. cereus*

group genomes, but not the *yhdK* gene (a small protein with two predicted membrane-spanning domains).

Comparative cell wall protein analysis. Surface structures of bacteria are important targets for rapid species detection by serological methods. For instance, a two-component immuno-fluorescence assay using antibodies specific to the cell wall and capsule antigens is recommended for detection of vegetative cells of *B. anthracis*. Vegetative cells of *B. anthracis* have a complex cell wall structure made of the poly γ-D-glutamic acid capsule, S--layer and carbohydrates [24]. *B. cereus* and *B. thuringiensis strains* do not have a capsule, but have several proteins with S-layer motifs. The *B. cereus* ATCC14579 strain lacks S-layer proteins, however it is not clear if *B. thuringiensis* has it. In the *B. cereus* group genomes, S-layer proteins are encoded by a well conserved locus along with the operon *csaAB,* which codes for cell surface anchor for S-layer proteins [24]. In the *B. thuringiensis* genome there is an S-layer protein, with similarity to peptidoglycan hydrolase and S-layer homology (SLH) domain at the N-terminal. Both *B. cereus* and *B. thuringiensis* genomes have nine additional proteins with SLH domains; but none of them has the crystallization domain necessary for S-layer protein polymerization. Six CDSs encode putative peptidoglycan hydrolases and one of them has a leucine-rich repeat (LRR) domain similar to that of internalins found in *Listeria spp.* In the *B. anthracis* genomes, there are two copies of genes for a protein translocase, SecA, which is missing from the *B. cereus* and *B. thuringiensis* genomes.

The composition, structure, and biosynthetic pathways for cell wall carbohydrates are well studied in *B. subtilis,* compared to the *B. cereus* group bacteria. In *B. subtilis,* anionic polymers are important components of the cell wall as they act as a sink for protons that are generated during respiration and are also a major site of metal deposition [25]. D-alanyl esterification of anionic polymers decreases electro-negativity of the cell walls, leading to modulation of autolysis and enhancing folding and stability of secreted proteins [26]. Depending on the availability of phosphate, different *B. subtilis* strains produce either phosphate-free teichuronic acid such as polyglucuronyl-N-acetyl-glucosamine, or teichoic acids, such as polyglycerolphosphate, poly-ribitolphosphate or polyglucosyl-N-acetyl-galactosamine-1-phosphate. The carbohydrate composition of bacterial cell walls in the *B. cereus* group is different from that of *B. subtilis*. Anionic polymers play a less prominent role

in *B. anthracis* pathogenesis, since neither polyribitolphosphate, nor polyglycerolphosphate is detected in their cell walls. The absence of teichoic acids was therefore suggested as a means to differentiate *B. subtilis* and *B. cereus* group bacteria. However, the strain *B. cereus* AHU 1030 is reported to contain polyglycerolphosphate [27]. The genes for ribitol-teichoic acid and glycerol-teichoic acid biosynthesis were studied in two strains of *B. subtilis* [28]. The genes, *tagO* (polyprenylphosphate α-GlcNAc transferase) and *tagA* (β-N-acetyl-mannosaminyl transferase), are required for biosynthesis of teichoic acid linkage unit. The *tagGH* operon encodes an ABC transporter responsible for translocation of teichoic acids through the membrane [29]. Orthologs of *tagO* and UDP-GlcNAc 2-epimerase *mnaA* (*yvyH*) were found in all three *B. cereus* group bacteria (BA_0288 and BA_0286, respectively, in the case of *B. anthracis*, and homologs of *tagA* were found in *B. cereus* and *B. anthracis* (BA_0528) while in the *B. thuringiensis* genome it may be located in the non-sequenced region. CDSs with similarity to *tagG* and *tagH* are present in the *B. anthracis* genome (BA_0360 and BA_0361, respectively), but not in *B. cereus* or *B. thuringiensis* . However, the "hallmark" gene for teichoic acid biosynthesis is the gene for TagF protein that is necessary for polymerization of a polyolphosphate chain which is present only in *B. subtilis*. None of *B. cereus* group bacteria have the gene *tagF*, suggesting they do not have the capability to synthesize teichoic acids. The enzymes involved in D-alanylation of teichoic acids and lipoteichoic acids are encoded by the *dltABCD* genes in *B. subtilis* [30]. Orthologs of the *dltABCD* operon are found in all members of the *B. cereus* group genomes in spite of the absence of the teichoic acid polymerization gene. However, it may be involved in D-alanylation of lipoteichoic acids.

Teichuronic acids of varying composition are found in the cell walls of many Gram-positive bacteria and their biosynthesis is well characterized in *B. subtilis* [31]. The *tuaABCDEFGH* operon encodes enzymes required for biosynthesis of the teichuronic acid monomers, GlcUA-GlcNAc, its export and polymerization machinery, and UDP-glucuronate. However, the functions of many CDSs in this operon are unknown. There is one copy of the gene for UDP-glucose dehydrogenase in *B. cereus*, whereas both *B. anthracis* and *B. thuringiensis* genomes have two copies each. All are more similar to the *B. subtilis* YwqF protein than to TuaD (teichuronic acids-specific enzyme). In the *B. subtilis* genome, the *tuaABCDEFGH* operon includes four CDSs (*tuaA*, *tuaC*, *tuaG*, *tuaH*) that belong to the

glycosyl transferase family proteins. All the *B. cereus* group genomes have CDS for *tuaA* and *tuaG* but not for the *tuaC*, *tuaH* genes. In the *B. thuringiensis* genome there are two homologs of *tuaA* (BTH06178 and BTH03001, 53% and 44% amino acid identity to *B. subtilis tuaA*, respectively) and two homologs of *tuaG* (BTH06174 and BTH05336, 55% and 48% identity to *tuaG*, respectively). The two genes have higher similarity to the *tuaA* and *tuaG* genes of *B. subtilis* and belong to a chromosomal cluster that also codes for UDP-glucose pyro-phosphorylase, UDP-glucose dehydrogenase (a homolog of *B. subtilis* bi-functional UDP-glucose / UDP-N-acetyl glucosamine 4-epimerase) [31], Wzx family oligosaccharide translocase and Wzz family polysaccharide polymerase. With the exception of *tuaH*, the chromosomal cluster encodes functionality similar to that of the *tua* operon and may be responsible for biosynthesis of teichuronic acids in *B. thuringiensis.* Although there are *tua* operons in the *B. cereus* and *B. anthracis* genomes, they contain two genes that code for UDP-glucose dehydrogenase family proteins. These genes might participate in the synthesis of nucleotide-sugar precursors for teichuronic acid structurally different from that of *B. subtilis.* One of the UDP-glucose dehydrogenases in *B. anthracis* (BA5512) has 63% identity to *Staphylococcus spp* UDP-N acetylmannosamin-uronate dehydrogenase Cap5O [32] and is responsible for biosynthesis of UDP-N-acetyl mannosaminuronic acid. This gene is located in a chromosomal cluster (BA5519-BA5512) with two more genes homologous to teichoic acid export ABC transporter, glycosyl transferase genes [33]. This chromosomal cluster was found in one of the nineteen *B. cereus*, *B. thuringiensis* and *B. weihenstephanensis* strains studied by comparative genomic hybridization [4], which may code for a galactose-GlcNAc, a neutral cell wall polysaccharide that is mostly found in *B. anthracis* strains [34]. However, the presence of UDP-N-acetyl mannosaminuronate dehydrogenase and heparinase orthologs in this cluster suggests a probable role in the biosynthesis of uronic acid-containing acidic polymer. The cell wall associated polysaccharides extracted from *B. anthracis* were separated into three fractions by ion-exchange chromatography [35]. The first two fractions represented the non-pyruvylated and pyruvylated galactose-GlcNAc polysaccharide, whereas the composition and structure of the minor acidic fraction III remains unknown and could represent an uronic acid-containing polysaccharide. The vegetative forms of *B. anthracis* and *B. cereus* had distinct carbohydrate profiles. In *B. anthracis,* high levels of galactose and low quantities of N-acetylgalactosamine were found, while vegetative forms of *B. cereus*

had high levels of *N*-acetylgalactosamine and low amounts of galactose [36]. In addition, *B. cereus* spores had two sugars, 2 *O*-methyl rhamnose and fucose, that were absent from spores of *B. anthracis*. Both D- and L- fucose can be found in bacterial poly- saccharides, and biosynthesis of both forms proceeds from D-fructose 6-phosphate via GDP-D-mannose and GPD-4-keto-6-deoxy D-mannose intermediates. Enzymes catalyzing the biosynthesis of GDP-4-keto-6-deoxy D-mannose, GDP -mannose pyrophosphorylase and GDP-mannose dehydratase, are well conserved in all three kingdoms, but no CDSs with similarity to GDP-mannose pyrophosphorylase and GDP-mannose 4, 6 dehydratase were found in any of the *B. cereus* group genomes. These genes could be replaced by a non-orthologs variant or alternatively fucose could be produced by an alternate pathway. Interestingly, two CDSs in *B. cereus* have similarity to CDP-D-glucose synthase: CDS BC3514 has 43% identity to the StrQ protein of *S. glauscens* [30], and BC3358 has 63% identity to AscA protein of *Y. pseudotuberculosis* [37]. Homologs of BC3514 are found in *B. thuringiensis (*BTH01275) and in *B. subtilis* (*yfnH*) and it is the first gene in a probable operon, which also contains a homolog of CDP-glucose 4, 6-dehydratase (BC3517, 52% identical to the DdhB protein of *Y. pseudotuberculosis*), a putative glycosyltransferase (BC3515) and an CDS BC3516 belonging to the NAD-dependent epimerase/dehydratase family. The latter CDS is identical to GDP-6-deoxy-D-xylo 4-hexulose reductase of *Aneurinbacillus thermoaerophilus* [38], which catalyses the last step in biosynthesis of GDP-D-rhamnose, thus the CDS BC3516 could also code for NDP-6-deoxy-4-ketohexose reductase. The second putative CDP-glucose synthase, BC3358, is surrounded by a chromosomal cluster, which also encodes a CDP-glucose 4, 6-dehydratase homolog (BC3359, 45% identical to DdhB protein of *Y. pseudotuberculosis*), a probable NDP -hexose 3-C-methyltransferase (BC3360) similar to the SnogG2 protein of *S. nogalater* [39] and another NAD-dependent epimerase/dehydratase family protein (BC3361), which, like BC3516, could be a NDP-6-deoxy-4-ketohexose reductase. The CDSs from the chromosomal cluster BC3514-BC3516 has orthologs in *B. thuringiensis* while the chromosomal cluster BC3358-BC3361 is unique for *B. cereus* genome. Although, usually CDP-D-glucose serves as a precursor for biosynthesis of 3, 6 di deoxyhexoses, such as CDP-abequose, CDP-ascarylose, CDP-paratose, and CDP-tyvelose, no homologs of the enzymes catalyzing 3-deoxygenation were found in *B. cereus*. Thus, it is possible that an unusual pathway proceeding from CDP-glucose rather than from GDP

mannose to produce 6-deoxyhexoses.

Comparative spore coat protein analysis. Spore coat and exosporium proteins of *B. anthracis* and *B. cereus* group bacteria are well characterized. In the *B. cereus* group, the spore coat and exosporium composition is largely conserved, but is different from orthologs of *B. subtilis.* No orthologs of coat protein genes *cotA, cotC, cotG, cotI, cotR, cotS, cotT, cotU* and *cotV* were found in any of the *B. cereus* group genomes. Other genes for coat proteins such as *safA*, *yaaH*, *yabG*, *spoVID*, *cotB, cotD, cotE, cotH, cotJA, cotJB, cotJC, cotM*, and *cotY* are found in all. Unlike in *B. cereus* and *B. thuringiensis* genomes, *B. anthracis* possesses a gene for *cotF* but not for *cotW, cotX* and *cotQ* genes. Genes for exosporium proteins such as *exsB, exsC, exsD, exsE, exsF* and *exsJ* are conserved in all *B. cereus* group organisms [40]. In *B. subtilis*, *spsABCDEFGHIJKL*, *cgeAB* and *cgeCDE* operons are involved in spore coat carbohydrate modification [41]. Deletion of genes in these operons increases in hydrophobic and aggregative properties of the spores and increases binding affinity to nonspecific surfaces. *B. thuringiensis* contains orthologs of all genes of the *spsABCDEFHIIJKL* operon that codes for spore coat polysaccharide biosynthesis protein except for *spsD*. In *B. cereus* and *B. anthracis* genomes, only three CDS for *spsI*, *spsJ*, *spsK* are found and others genes of this operon were missing.

The *cgeAB* operon and *cgeCDE* operon are divergently transcribed in *B. subtilis* and are involved in glycosylation of spores during maturation. Among the *B. cereus* group bacteria, both in *B. anthracis* and *B. cereus* do not have the *cgeAB*, *cgeCDE* operons however like *B. thuringiensis* has all both the operons, Interestingly an ORF in the *B. anthracis* has sequence similarity to *cgeB* gene has been identified. The absence of these genes in *B. cereus* and *B. anthracis* alters hydrophobic and adhesive properties of spores [42].

Comparative membrane transport system analysis. All three *B. cereus* group organisms use phosphotransferase system for carbohydrate transport, and each of them has an HPr protein and catabolite repression protein Crh protein similar to that found in *B. subtilis* [43]. *B. thuringiensis* genome has a second copy of HPr gene located adjacent to a dihydroxy-acetone kinase gene suggesting its role in phosphotransfer similar to that of *E. coli* YcgC protein. *B. thuringiensis* has two putative mannose PTS systems that are not found in *B. cereus* and *B. anthracis* genomes, one of the PTS system is similar to *manP/yjdD* of *B. subtilis* [44] while the other is unique perhaps acquired horizontally from *Enterococcus spp*.

The *B. cereus* group bacteria appear to be well equipped to scavenge lower concentrations of few metals compared to *B. subtilis*. All the three possess genes for a Kdp P-type ATPase for acquiring potassium, but only *B. cereus* and *B. thuringiensis* have genes for Mg^{2+} P-type ATPases whereas *B. subtilis* has none. Similarly, CDSs for ferrous iron transporter are found in the *B. cereus* group (*feoAB*, present in two copies in all three genomes) but absent in *B. subtilis*. One of the *feoB* genes is into two CDS, in all three *B. cereus* group genomes, which is not due to sequencing error but may as well be functional as two subunits. A manganese ABC transporter is found in *B. thuringiensis*, *B. anthracis*, and *B. subtilis* genomes and is absent in *B. cereus*. The Mn^{2+} transporter present in the *B. thuringiensis* and *B. anthracis* genomes are distinct from *B. subtilis* and are more closely related to *Listeria spp*.

The *B. cereus* group and *B. subtilis* genomes possess phosphate transporters and a glycerol-3phosphate/phosphate anti-porter. In addition *B. cereus* group bacteria appear to be able to utilize additional compounds as sources of phosphorus. A phosphoglycerate transporter (*pgtP*), ABC transporter for glycerol-3 phosphate and antiporters is found in all members in the *B. cereus* group. *B. cereus* is capable of using 2 amino-ethyl phosphonate as a phosphate source, and the genes involved in the catabolism of this compound have been identified [45]. This gene is also present in the *B. anthracis* and *B. thuringiensis* genomes, and is located adjacent to an ABC transporter gene specific for 2-aminoethyl-phosphonate. CDSs for S-methyl- transferase transporter for uptake of S-methyl methionine is found in *B. thuringiensis* genome. Only *B. anthracis* genome of the *B. cereus* group possesses a putative nicotinamide mono-nucleotide transporter.

Virulence. The virulence genes in *B. cereus* and *B. anthracis* strains is well described by [1], here virulence genes of *B. thuringiensis* genome is discussed. Several genes corresponding to toxins or toxin-like proteins that were previously unknown were identified. Two CDSs (BTH07769 and BTH07770), which have similarity to crystal protein Cry15Aa1 of *B. thuringiensis sub species thompsoni* [46] and to the crystal protein Cry33Aa1 from *B. thuringiensis* serovar Dakota strain 90 F4514 [47]. The latter strain is non-insecticidal but the toxin exhibits strong cytocidal activity against leukemic T-cells. These CDSs are adjacent and more similar to each other than to the proteins from other *B. thuringiensis* sub species, thus suggesting a recent duplication, in addition these genes are also flanked by transposases,

indicating horizontal acquisition. While the Cry33A1 protein from *B. thuringiensis* strain 90-F4514 has neither insecticidal nor cytocidal activity, Cry15Aa1 from *B. thuringiensis sub spp thompsoni* has been shown to have anti-lepidopteran activity by itself, although the toxicity is higher when combined with a non-insecticidal 40-kD parasporal protein [48] and we did not identify its ortholog in *B. thuringiensis.* A second putative toxin from *B. thuringiensis* (BTH04010) is found to have two domains and perhaps functions as a fusion protein. A domain analysis revealed that the N-terminal domain is encoded as a protein by itself in *S. coelicolor* and *B. halodurans* and has similarity to cell death inducing proteins from *Phytophthora spp, Pythium spp*, and *Fusarium spp* and the C-terminal domain is similar to a mosquitocidal 100-kD toxin from *B. sphaericus* [12, 49], and to HA33 hemeagglutinin from *Clostridium botulinum* which binds to N-acetyl-neuraminic acid or sialo-glycolipids of erythrocytes cells. The presence of a ricin B lectin domain in the C-terminal suggests a carbohydrate recognition function targeted to the insect cell surface.

Phospholipases are virulence factors of many bacteria, including the bacteria belonging to the *B. cereus* group. In the *B. thuringiensis genome,* virulence-related transcription factor PlcR was identified as a regulator of phospholipase C, however within this genome, there are three closely related and previously unidentified phospholipases (BTH03343, BTH04416, BTH07775) with similarity to the pBtoxin-encoded pseudogene pBt087 [12]. Two of the three phospholipases in *B. thuringiensis* have predicted secretory signal sequences and none have PlcR-binding sites. A gene for cytolysin was found in the chromosomal cluster of *B. thuringiensis* (BTH07389-BTH07373); while none of the genes in the cluster have orthologs in the genomes of other *Bacillus spp*, however several genes in this cluster are similar to genes in the *cyl* operon of *Streptococcus agalactiae.* This group B *Streptococcus spp* demonstrates β-hemolytic activity and produces a yellow orange pigment, with both activities being abolished by insertions in the genes belonging to the *cyl* operon [50]. β--hemolysin of *S. agalactiae* is apparently a cell wall-associated protein, since hemolytic activity is contact dependent and protease sensitive. The *cylE* gene was shown to be both necessary and sufficient to confer β-hemolytic activity to a non-hemolytic *E. coli* strain, so it probably represents a structural gene for β-hemolysin or its precursor. In the *B. thuringiensis* genome, a CDS (BTH07380) with limited identity (15%) to the CylE protein was identified, the N-terminal has hemolysis domain [50]. While β-hemolysin of *S. agalactiae* is a major

virulence factor, the physiological role of the pigment or its molecular structure is not known, however, based on the predicted functions of the CDS in the *cyl* operon, the pigment is more likely to be of polyketide origin, similar to the spore pigments of *Streptomy*ces *spp* and fungal melanin. The c*yl* operon includes three enzymes found in the typeII polyketide synthases (PKS): an acyl carrier protein (ACP), malonyl-CoA: ACP transacylase, and an unusual protein, which resembles a fusion of two subunits of the heterodimeric keto-synthase. The latter is the central component of type II PKS which normally consists of two proteins: the 3-ketoacyl-ACP synthase itself and the chain length factor, which has end-to-end homology with 3-ketoacylACP synthase [51]. The chromosomal cluster BTH07389-BTH07373 in *B. thuringiensis* genome has all three components of type II PKS: BTH07384 codes for malonyl-CoA: ACP transacylase, BTH07382 is an acyl-carrier protein and BTH07377 is a fusion of 3-ketoacyl-ACP synthase and chain length factor. Other *cyl* operon-encoded proteins that have homologs in the *B. thuringiensis* chromosomal cluster include ATPase (BTH07388) and permease components of a putative export ABC transporter (BTH07386), 3ketoacyl-ACP reductase (BTH07383), 3-hydroxyacyl: ACP dehydratase (BTH07381), and a homolog of amino-methyltransferase component of the glycine cleavage complex (BTH07378). The reductase and dehydratase likely participate in biosynthesis of the polyketide starter unit and the homolog of amino-methyltransferase for polyketide modification. Three genes in the *cyl* operon including two hypothetical proteins and a putative glycosyltransferase have no homologs in the *B. thuringiensis* chromosomal cluster; several *B. thuringiensis* genes have no homologs in the cyl operon in *S. agalactiae*. However it is possible the pigment structure and color production by *S. agalactiae* and *B. thuringiensis* are different, like heterologous expression of *B. cereus* UW85 genes in *E. coli* producing orange pigment [52].

Conclusions

A comparative analysis of representative members of *Bacillus* genomes has led to the discovery of common and unique metabolic and virulence capability of each species. The presence of specific genes with functions related to spore coat, exopolysaccharide biosynthesis, and membrane transport has revealed significant differences despite the high level of chromosomal synteny among the *B. cereus* group bacteria. An alternative model for regulation of the stress-responsive sigma factor σ^{B} in the *B. cereus* group was proposed.

Several additional genes encoding toxins similar to Cry15Aa were identified in the *B. thuringiensis* genome.

Acknowledgements

This work was supported in part by DARPA STTR grant DAAH01-99-C-R208 to Integrated Genomics Inc and by INRA France. We thank Dr Robert Haselkorn, University of Chicago for critically reading the manuscript.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36.

References

1. Rasko A., Altherr, M.R., Han, C.S., and Ravel, J. (2005) Genomics of the Bacillus group or organisms. FEMS. Microbiol.29, 303-329.
2. Schnepf, E., Crickmore, N., Van Rie, J., et al., (1998) *B. thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62, 775-806.
3. Read, T. D., Salzberg, S. L., Pop, M., et al. (2000) Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. Science. 296, 2028-2033.
4. Read, T. D., Peterson, S. N., Tourasse, N., et al. (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature. 423 , 81-86.
5. Ivanova, N., Sorokin, A., Anderson, I., et al. (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature. 423, 87-91.
6. Fillinger, U., Knols, B. G. J. and Becker, N. (2003) Efficacy and efficiency of new *B. thuringiensis var. israelensis* and *Bacillus sphaericus* formulations against afro-tropical anophelines in Western Kenya, Trop.Med. Int. Health. 8, 37-47.
7. Rey , M.W., Ramaiya, P, Nelson, B.A. et al.(2004) Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related Bacillus species. Genome Biol. 5(10):R77. Epub.

8. Kapatral, V., Ivanova, N., Anderson, I., et al. (2003) Genome analysis of *F. nucleatum* sub spp *vincentii* and its comparison with the genome of *F. nucleatum* ATCC22586**.** Genome Res. 13, 1180-1189.
9. Overbeek, R., Larsen, N., Walunas, T., et al. (2003) The ERGO genome analysis and discovery system. Nucleic Acids Res. 31, 164-171.
10. Bhattacharyya, A., Stilwagen, S., Reznik, G., et al. (2002) Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights**.** Genome Res. 12, 1556-1563.
11. Gonzalez, J.M., and Carlton, B.C., (1984) A large transmissible plasmid is required for crystal toxin production in *B. thuringiensis* variety israelensis. Plasmid. 11, 28-38.
12. Berry, C., O'Neil, S., BenDov, E., et. al (2002) Complete sequence and organization of pBtoxis, the toxin coding plasmid of *B. thuringiensis* sub spp. *israelensis*, Appl. Environ. Microbiol. 68, 5082-5095.
13. Andrup, L., Damgaard, J., Wassermann, K., Boe, L., Madsen, S.M. and Hansen, F.G. (1994) Complete nucleotide sequence of the *B. thuringiensis* sub spp. *israelensis* plasmid pTX143 and its correlation with biological properties**.** Plasmid. 31, 72-88.
14. Pottathil, M. and Lazazzera, B. A. (2003) The extracellular phr peptide rap phosphatase signaling circuit of *Bacillus subtilis*. Front. Biosci. 8. 32-45.
15. Kunst, F., Ogasawara, N., Moszer, I., et al,. (1997) The complete genome sequence of the Gram positive bacterium *Bacillus subtilis*. Nature. 390, 249-256.
16. Perego, M. (2001) A new family of aspartyl phosphate phosphatases targeting the sporulation transcription factor Spo0A of *Bacillus subtilis*. Mol. Microbiol. 42, 133-143.
17. Fabret, C., Feher, V. A. and Hoch, J. A. (1999) Two component signal transduction in *Bacillus subtilis*: how one organism sees its world. J. Bacteriol. 181, 1975-1983.
18. Schmalisch, M., Langbein, I. and Stulke, J. (2002) The general stress protein Ctc of *Bacillus subtilis* is a ribosomal protein, *J. Mol.* Microbiol. Biotechnol. 4, 495-501.
19. Dufour, A. and Haldenwang, W. G. (1994) Interactions between a *Bacillus subtilis* anti sigma factor (RsbW) and its antagonist (RsbV). J. Bacteriol. 176, 1813-1820.
20. Vijay, K., Brody, M. S., Fredlund, E. and Price, C. W. (2000) A PP2C phosphatase containing a PAS domain is required to convey signals of energy stress to the sigma B

transcription factor of *Bacillus subtilis*. Molecular. Microbiol. 35, 180-188.

21. Fouet, A., Namy, O. and Lambert, G. (2000) Characterization of the operon encoding the alternative sigma (B) factor from *Bacillus anthracis* and its role in virulence**.** J. Bacteriol. 182. 5036-5045.

22. Qin, X., Singh, K.V., Weinstock, G. M. and Murray, B. E. (2000)Characterization of fsr, a regulator controlling expression of gelatinase and serine protease in *Enterococcus faecalis* OG1RF. Infect. Immun. 68, 2579-2586.

23 Horsburgh, M. J. and Moir, A. (1999) Sigma M, an ECF RNA polymerase sigma factor of *Bacillus subtilis* 168, is essential for growth and survival in high concentrations of salt**.** Molecular. Microbiol. 32, 41-50.

24. Mignot, T., Denis, B., CoutureTosl, E., Kolstǿ, A. B., Mock, M. and Fouet, A. (2001) Distribution of S-layers on the surface of *Bacillus cereus* strains: phylogenetic origin and ecological pressure**.** Environ. Microbiol. 3, 493-501.

25. Calamita, H.G., Ehringer, W.D., Koch, A.L. and Doyle, R. J. (2002) Evidence that the cell wall of *Bacillus subtilis* is protonated during respiration**.** Proc. Natl. Acad. Sci. USA. 98, 15260-15263.

26. Chambert, R. and PetitGlatron, M. F. (1999) Anionic polymers of *Bacillus subtilis* cell wall modulate the folding rate of secreted proteins**.** FEMS. Microbiol. Lett. 179, 43-47.

27. Sasaki,Y., Araki,Y. and Ito,E. (1983) Structure of teichoic acid glycopeptide complexes from cell walls of *Bacillus cereu*s AHU 1030**.** Eur. J. Biochem. 132, 207-213.

28. Lazarevic, V., Abellan, F. X., Moller, S. B., Karamata, D. and Mauel, C. (2002) Comparison of ribitol and glycerol teichoic acid genes in *Bacillus subtilis* W23 and 168: Identical function, similar divergent organization, but different regulation**.** Microbiology. 148, 815-824.

29. Lazarevic,V. and Karamata, D. (1995) The *tagGH* operon of *Bacillus subtilis* 168 encodes a two component ABC transporter involved in the metabolism of two wall teichoic acids**.** Mol. Microbiol. 16, 345-355.

30. Perego, M., Glaser, P., Minutello, A., Strauch, M. A., Leopold, K. and Fischer, W. (1995) Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in

Bacillus subtilis. Identification of genes and regulation**.** J. Biol. Chem. 270, 15598-15606.

31. Soldo, B., Scotti, C., Karamata, D. and Lazarevic, V. (2003) The *Bacillus subtilis* Gne (GneA GalE) protein can catalyze UDP-glucose as well as UDP-N-acetylglucosamine 4-epimerisation**.** Gene. 319, 65-69.
32. Portoles, M., Kiser, K. B., Bhasin, N., Chan, K. H. and Lee, J.C. (2001) *Staphylococcus aureus* Cap5O has UDP-ManNAc dehydrogenase activity and is essential for capsule expression**.** Infect. Immun. 69, 217-223.
33. Su, H., Blain ,F., Musil, R.A., Zimmermann, J .J., Gu,K. and Bennet, D.C. (1996), Isolation and expression in *Escherichia coli* of *hpeB* and *hepC*, genes coding for the glycosaminoglycan degrading enzymes heparinase II and heparinase III, respectively, from *Flavobacterium heparinum*. Appl. Environ. Microbiol. 62, 2723-2734.
34. Ezzell, J. W., Jr, Abshire, T. G., Little, S. F., Lidgerding, B.C. and Brown, C. (1990) Identification of *Bacillus anthracis* by using monoclonal antibody to cell wall galactose N-acetylglucosamine polysaccharide**.** *J. Clin. Microbiol*. 28, 223-231.
35. Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000) Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. The EMBO. J. 19, 4473-3384.
36. Fox, A., Stewart, G. C., Waller, L. N., Fox, K.F., Harley,W. M. and Price, R. L. (2003) Carbohydrates and glycoproteins of *Bacillus anthracis* and related bacilli: targets for biodetection**.** J. Microbiol. Methods. 54, 143-152.
37. Thorson, J. S., Kelley, T. M. and Liu, H. W. (1994) Cloning, sequencing, and over-expression in *Escherichia coli* of the alpha-D-glucose1phosphate cytidyltransferase gene isolated from *Yersinia pseudotuberculosis***.** J. Bacteriol. 176, 1840-1849.
38. Kenidinger, B., Graninger, M., Adam, G.,Puchberger M., Kosma, P., and Messner, P (2001) Identification of two GDP6 deoxyD-lyxo4-hexulose reductases synthesizing GDP -D-rhamnose in *Aneurinibacillus thermoaerophilus* L42091T. J. Biol. Chem. 276, 5577-5583.
39. Torkkell, S., Kunnari, T., Palmu, K., Mantsala, P., Hakala, J. and Ylihonko, K. (2001) The entire nogalamycin biosynthetic gene cluster of *Streptomyces nogalater*:

characterization of a 20kb DNA region and generation of hybrid structures. Mol. Genet. Genomics. 266, 276-288.

40. 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.
41. Driks, A. (1990) *Bacillus subtilis* spore coat. Microbiol. Mol. Biol. Rev. 63, 1-20.
42. Ronner, U., Husmark, U. and Henriksson, A. (1990) Adhesion of *Bacillus* spores in relation to hydrophobicity. J. Appl. Bacteriol. 69, 550-556.
43. Galinier, A., Haiech, J., Kilhoffer, M.C., Jaquinod, M., Stulke, J., Deutscher J, and Martin-Verstraete I. (1997) The *Bacillus subtilis* *crh* gene encodes a HPr like protein involved in carbon catabolite repression. Proc. Natl. Acad. Sci. USA. 94, 8439-8444.
44. Reizer, J., Bachem, S., Reizer,A., Arnaud, M., Saier, M. H., Jr. and Stulke, J. (1999) Novel phosphotransferase system genes revealed by genome analysis the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*. Microbiol. 145, 3419-3429.
45. Baker, A.S., Ciocci, M.J., Metcalf, W. et al. (1998) Insights into the mechanism of catalysis by the PC bond cleaving enzyme phosphono-acetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. Biochemistry. 37, 9305-9315.
46. Brown, K. L. and Whiteley, H. R. (1992) Molecular characterization of two novel crystal protein genes from *B. thuringiensis* sub spp. *thompsoni*. J. Bacteriol. 174, 549-557.
47. Kim, H. S., Saitoh, H., Yamashita, S., et al.,(2003) Cloning and characterization of two novel crystal protein genes from a *Bacillus thuringiensis* serovar dakota strain**.** Curr. Microbiol. 46, 33-38.
48. Rang, C., Lacey, L. A. and Frutos, R. (2000) The crystal proteins from *Bacillus thuringiensis* subspp. *thompsoni* display a synergistic activity against the codling moth, *Cydia pomonell*. Curr. Microbiol. 40, 200-204.
49. Thanabalu, T., Hindley, J., Jackson Y, J. and Berry, C. (1991) Cloning, sequencing, and expression of a gene encoding a 100 kilodalton mosquitocidal toxin from *Bacillus sphaericus* SSII1. J. Bacteriol. 173, 2776-2785.
50. Pritzlaff, C. A., Chang, J. C., Kuo, S. P., Tamura, G. S., Rubens, C.E. and Nizet, V.

(2001) Genetic basis for the beta haemolytic/cytolytic activity of group B *Streptococcus***.** Molecular. Microbiol. 39, 236-247.

51. Dreier, J. and Khosla, C. (2000) Mechanistic analysis of a type II polyketide synthase. Role of conserved residues in the β-ketoacyl synthase chain length factor heterodimer. Biochemistry. 39, 2088-2095.

52. Rondon, M. R., Raffel, S. J., Goodman, R. M. and Handelsman, J. (1999) Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. Proc. Nat. Acad. Sci. USA. 96, 6451- 6455.

Table 1. Comparative genome statistics of the *Bacillus* species.

.	*B. thuringiensis Sub spp israelensis*	*B. anthracis* (A2012)	*B. cereus*	*B. subtilis*
Contigs (Plasmids)	866 (9)	3 (2)	2 (1)	1 (0)
DNA sequence (bp)	5,880,839	5,370,060	5,427,548	4,214,630
G+C content (%)	35.0	35.1	35.3	43.5
Total CDSs	6,451	5,842	5,370	4,106
CDSs with assigned function (%)	4,413 (68)	4,175 (71)	3,835 (71)	3,076 (75)
CDS without similarities (%)	438 (6.7)	175 (3.0)	89 (1.7)	17 (0.4)
Conserved Hypotheticals (%)	1,600 (29)	1,492 (25)	1,446 (27)	1,013 (25)

Table 2. Comparison of gene numbers of selected signal transduction protein families.

Protein families	*B. thuringiensis sub spp israelensis*	*B.anthracis* A2012	*B. cereus*	*B. subtilis*
Rap phosphatases	8	6	5	11
Phr peptides	4	5	5	7
Spo0E phosphatases	6	3	5	3
SK/RR pairs	44	43*	39	30
SK orphans	14	10*	15	6
DNA-binding RR orphans	3	3	3	0
Fused SK/RRs	2	1	2	0
ECF sigma factors	13	16	10	7

SK/RR pairs, SK orphans, DNA binding RR orphans, and fused SK/RRs in *B. subtilis* are from [17]. SK: sensor histidine kinase; RR: response regulator; ECF: extra cytoplasmic function. Spo0A was not included in the DNA binding RR category. * Some of the SK/RR pairs appear to be disrupted by frameshifts. [#] Number includes two ECF sigma factors that were not found in *B. anthracis* A2012 but likely to be within the gaps. In *B. anthracis* A2012 the putative anti-sigma factor genes are present, but they are adjacent to contig ends. BLAST searches of the *B. anthracis* Ames ECF sigma factor amino acid sequences against the *B. anthracis* A2012 genomic DNA sequence identified the N-terminus of one of these sigma factors adjacent to a contig end.

Figure 1. Distribution of protein families calculated by WorkBench among organisms of the *B. cereus* group genomes.

Figure 2. Model for regulation of σ^B in the *B. cereus* group.

A) During unstressed conditions, the phosphatase acting on RsbV is inactive, so RsbV is phosphorylated and incapable of binding to the antisigma factor RsbW. RsbW is free to bind to and inactivate σ^B.

B) During stressful conditions, histidine kinases activate the phosphatase by phosphorylating its response regulator receiver domain. RsbV is dephosphorylated and binds to RsbW, freeing σ^B to activate transcription of stress-regulated genes. Abbreviations: Pase, phosphatase; RR, response regulator receiver domain.

Figure 3. Alignment of proposed autophosphorylating tyrosine kinases from the *B. cereus* group organisms. The residue corresponding to the phosphorus-accepting histidine residue is marked with an arrow above the alignment using ClustalW. Conserved amino acid residues are marked with an asterisk; similar amino acid residues are marked with a colon; more distantly similar amino acids are marked with a period.