

# **The Pathogenomics and Evolution of Anthrax-like *Bacillus cereus* Isolates and Plasmids**

A white paper proposal submitted by:

Geraldine A. Van der Auwera, Ph.D.  
Harvard Medical School

Michael Feldgarden, Ph.D.  
Genomic Sequencing Center for Infectious Diseases  
The Broad Institute of MIT and Harvard

## Executive summary

A key member of the *Bacillus cereus* group, *Bacillus anthracis* is defined by phenotypic and molecular characteristics that are conferred by two large plasmids, pXO1 and pXO2. However the very concept of *B. anthracis* as a distinct species has been called into question by recent discoveries of “intermediate” isolates identified as *B. cereus* and *B. thuringiensis* but possessing features similar to those of *B. anthracis*, including large plasmids that share a common backbone with pXO1 and/or pXO2. Many of these “intermediate” isolates possess potential or demonstrated lethal pathogenic properties and are sometimes called “anthrax-like”, even though they do not meet the strict definition of anthrax-causing *B. anthracis*. We recently showed that pXO1- and pXO2-like plasmids are widely prevalent in environmental isolates of the *B. cereus* group.

Because *B. anthracis*-like isolates do not possess all the molecular hallmarks of typical *B. anthracis*, there is a significant risk that they would escape being flagged as dangerous. Consequently, accidental infection by naturally occurring pathotypes which are not immediately recognized as life-threatening could present a serious health concern. Such cases have already been reported, some with a fatal outcome. The second risk posed by these *B. anthracis*-like isolates could be the intentional use as “stealth anthrax” bioweapon, either in natural form or with genetic modifications that would require only minimal skills and facilities to produce. To address these risks, this proposal focuses on three main questions:

1. What is the functional repertoire of the pXO1- and pXO2-like plasmids and does it include known virulence-related genes?
2. Are pXO1-like and/or pXO2-like plasmids associated with chromosomal genetic features that indicate pathogenic potential and/or similarity with known *B. anthracis*-like strains of *B. cereus*?
3. How does the genomic diversity within this collection of potentially *B. anthracis*-like strains compare to other available *B. cereus* group genomes?

This emerging picture of a virulence-associated plasmid gene pool highlights the inadequacy of the classical *B. cereus* species definitions, at least for operational biosafety purposes. However, the observation that pXO1- and/or pXO2-like plasmids seem to be systematically present in the dangerous pathotypes of the *B. cereus* group provides a valuable starting point to investigate the extent to which such strains occur naturally in the environment.

Here we propose to sequence a panel of 96 isolates representing a variety of plasmid contents and environmental origins to achieve three specific outcomes:

1. Assess the potential for accidental infections;
2. Assess the theoretical feasibility of bioweapons development from naturally occurring bacterial isolates;
3. Generate sequence data suitable for the development of improved diagnostics and clinical countermeasures by specialists in the community.

## Background

The *Bacillus cereus* species complex is a set of ubiquitous rod-shaped Gram-positive soil bacteria with six subspecies that are genetically very similar but have highly specialized lifestyles with distinct virulence spectra. The *B. cereus* species complex includes the opportunistic pathogen *B. cereus sensu stricto*, which is frequently implicated in cases of food poisoning, periodontitis and endophthalmitis, the entomopathogen *B. thuringiensis*, which is widely used as a biopesticide, and the causative agent of anthrax *B. anthracis*. Their main distinguishing phenotypic features, including their respective virulence properties, are directly associated with large plasmids. In *B. anthracis*, the toxin and capsule genes responsible for anthrax disease are located on the 182-kb pXO1 and 95-kb pXO2 plasmids, respectively. Both plasmids also encode regulator genes that control the expression of anthrax virulence.

In recent years, various clinical isolates of *B. cereus* group members have been found with large plasmids that are related to pXO1 and pXO2. Some of these plasmids are known to encode the genetic determinants that confer pathogenic properties to the host strains. For instance, the emetic toxin cereulide, which causes one of two food-poisoning syndromes attributed to *B. cereus s. stricto*, is encoded by a large pXO1-like plasmid called pCER207 (Hoton *et al.*, 2005; Ehling-Schulz *et al.* 2006).

Furthermore, *B. cereus s. stricto* has been implicated in anthrax-like diseases with clinical presentation of severe pneumonia, affecting humans in North America and great apes in Africa (Hoffmaster *et al.*, 2004 and 2006; Klee *et al.*, 2006; Avashia *et al.*, 2007). Most of these isolates possess a pXO1-like plasmid encoding the anthrax toxin genes and some are capable of producing the typical protective poly-D-glutamic acid capsule that shields the infecting bacterium from the immune system. However, several produce a unique capsule that is polysaccharide-based, which is not at all typical of *B. anthracis* (Hoffmaster *et al.*, 2004 and 2006). Such functional substitution could allow *B. anthracis*-like pathogens to evade detection by existing diagnostic tools. Functional substitutions affecting the antigenic determinants that are recognized by anthrax vaccines have not yet been reported, but if they were to occur (whether naturally or by genome engineering) they would pose a threat to exposed personnel such as troops and biological threat response professionals. Indeed, vaccination can only protect individuals against pathogens that bear the same antigenic determinants used to produce the vaccine. Anthrax variants presenting different antigens would therefore not be adequately targeted by the immune system of vaccinated individuals.

Genomic analyses of the available complete genome sequences have shown that many of these strains are significantly different from the bulk of the *B. cereus* group, but do not belong to the lineage of typical *B. anthracis*. They do however belong to the same top-level clade as *B. anthracis*, and are generally referred to as the anthrax-like *B. cereus* pathotypes or as “close-neighbors” of *B. anthracis*.

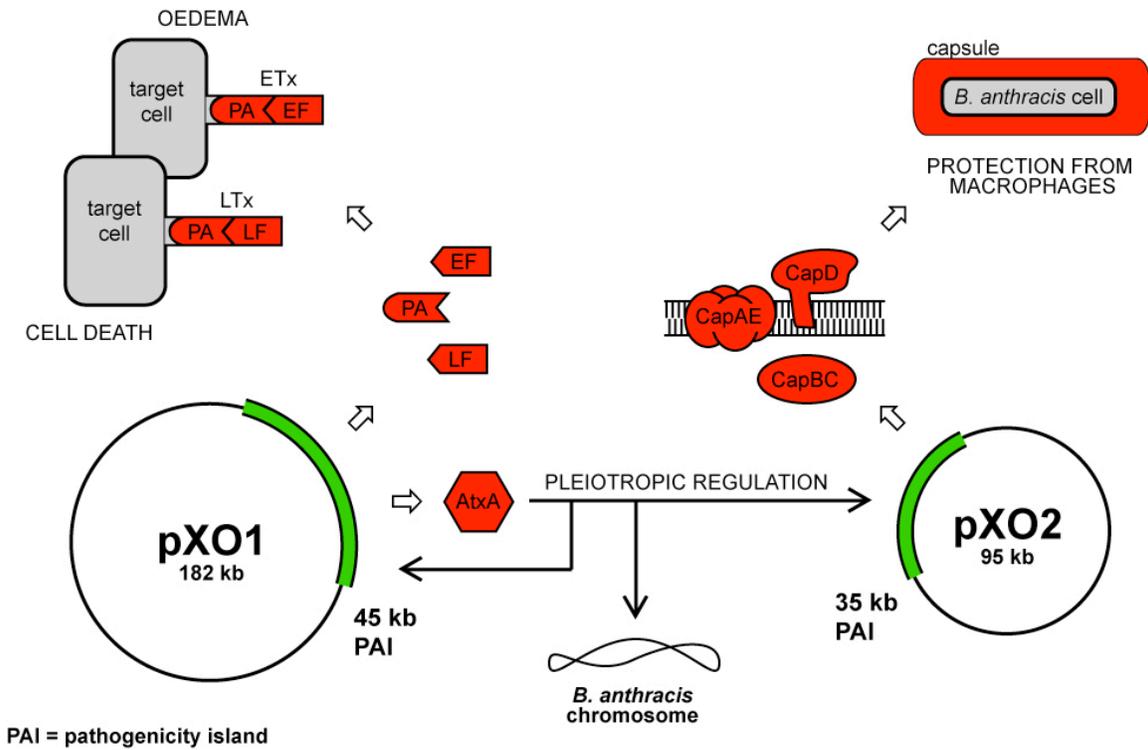


Fig. 1 : Overview of the anthrax virulence factors

The AtxA regulator produced from pXO1 controls the expression of the anthrax toxins components from pXO1 and of the capsule formation components from pXO2. The toxin components EF (oedema factor), LF (lethal factor) and PA (protective antigen) assemble into the toxins ETx (oedema toxin) and LTx (lethal toxin), causing oedema and death, respectively, in targeted host cells. The capsule formation components ABCDE interact at the bacterial cell membrane to produce a poly-gamma-D-glutamic acid capsule that protects *B. anthracis* cells from phagocytic killing during infection.

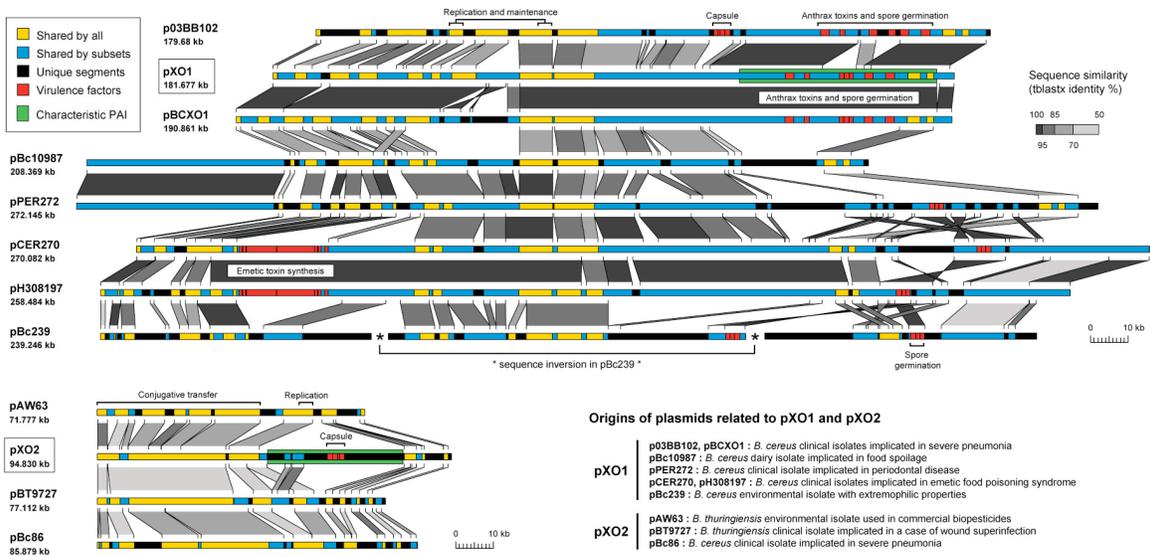


Fig. 2 : Linear alignment maps of fully-sequenced pXO1-like and pXO2-like plasmids

Segments shared by all plasmids or by subsets of plasmids are colored as indicated by the general color key (top left). Genes coding for known virulence factors are represented by red boxes. Shaded areas between plasmid pairs show the average pairwise similarity of the corresponding segments as indicated by the heat color key (top right).

## Analysis and Biological Implications

The genome sequence data will address the pathogenic potential of the strains as well as the genomic diversity and phylogenetic structure of the selected strain collection.

### Pathogenomic analysis:

- What is the functional repertoire of the pXO1- and pXO2-like plasmids and does it include known virulence-related genes?
- Are pXO1- and/or pXO2-like plasmids associated with chromosomal genetic features that indicate pathogenic potential and/or similarity with known *B. anthracis*-like strains or other clinical strains of *B. cereus*?

### Genomic diversity and phylogenetic structure:

- How does the genomic diversity within this collection compare to other available *B. cereus* group genomes?
- Do strains that were isolated from the same sampling site but have different plasmid content (pXO1 and/or pXO2) share similar chromosomal genomic features? Conversely, do strains that were isolated from different sampling sites but have with similar plasmid content (pXO1 and/or pXO2) share similar chromosomal genomic features?
- How diverse are the pXO1- and pXO2-like plasmids in terms of backbone architecture, modularity and phylogeny? Specifically, what constitutes the universal common backbone of each type of plasmid, how diverse are the accessory segments and how are individual plasmids related to each other and to known chromosomal sequences?
- Does comparing the phylogenies of the plasmids and of their host strains show evidence of horizontal gene transfer?

The general outcome of these analyses will be an improved understanding of the genome diversity and pathogenomic potential of naturally occurring *B. cereus* group strains that carry pXO1- and/or pXO2-like plasmids. More specifically, three practical outcomes will be achieved:

- Assessment of the potential for accidental infections based on the prevalence of environmental strains with pathogenomic potential;
- Assessment of the theoretical feasibility of bioweapons development from naturally occurring bacterial isolates;
- Production of molecular/sequence data suitable for the development of improved diagnostics and clinical countermeasures by specialists in the community.

## Rationale and strain selection

A recent molecular probe-based study performed at the Laboratory of Food and Environmental Microbiology in Louvain-la-Neuve, Belgium revealed that *Bacillus cereus* group isolates containing pXO1- and pXO2-like plasmids are widely distributed across a variety of environmental niches and geographical locations (Hu *et al.*, 2009). Approximately 15 % of a collection of 2,000 environmental isolates contained either a pXO1- or a pXO2-like plasmid, and a minor proportion (0.5 % of total) contained both. None tested positive for the anthrax toxins, but there is evidence that many of these plasmids carry some kind of genetic “payload” in addition to the conserved backbone. The highest overall prevalence of isolates carrying pXO1- and pXO2-like plasmids was found in soil samples (26 %). There was also a high prevalence of pXO1-like plasmids specifically in small wild mammals such as bank voles (16.5 %), which may act as natural reservoirs. Ecological mixing within and between these niches could foster the emergence of dangerous pathotypes of *B. cereus*, notably through horizontal gene transfer events promoted by these large plasmids and other mobile genetic elements.

To explore the genomic diversity and pathotypical potential associated with pXO1- and pXO2-like plasmids, we propose to sequence the genomes of a panel of 96 environmental isolates of *B. cereus* group organisms (listed in Table 1).

At the core of this panel is a set of 74 isolates organized in groups of isolates originating from the same material samples, matched by plasmid content combinations (Table 1A). Specifically, the selection process was performed as follows: for each sample of soil or water that had yielded a pool of isolates some of which contained plasmids of interest, we selected from that pool a subset of strains representing each combination of plasmid content observed in the original sample (including the “no plasmids” case figure), distributed equally and in amounts based on the maximum common number of isolates available for each case figure. To illustrate by an example, one soil sample from Boston, MA yielded 2 isolates containing pXO1-like plasmids and 3 isolates containing pXO2-like plasmids, as well as isolates containing neither plasmid. From this, we selected 2 isolates containing pXO1-like plasmids, 2 isolates containing pXO2-like plasmids and 2 isolates containing neither plasmid.

This core set carries the following advantages:

- 1) The selection covers the diversity of plasmid content combinations observed in soil and water samples from diverse geographical origins (pXO1-like alone, pXO2-like alone, both plasmids, or neither one);
- 2) Each possible combination of plasmid content is represented by enough isolates to allow us to detect correlations between plasmid content and pathogenomic potential.
- 3) Matching plasmid-bearing isolates with non-bearing isolates from the same original samples will provide sample-appropriate negative controls for the correlation analyses. While there are several *B. cereus* genome sequencing projects currently underway elsewhere that will yield some genomes of *B. cereus* without pXO1- or pXO2-like plasmids, none uses the same selection criteria. It is important for the predictive power of the analyses to sequence plasmid-lacking controls isolated and selected with the same methods and criteria as the rest of

the collection.

Furthermore, the core set is complemented by two sets of 12 and 10 isolates, respectively, that will provide additional depth and breadth to the analyses, as follows:

1) We identified a number of strains that originate from samples for which the rest of the pool of isolates was not stored in collection, and could thus not be matched by plasmid content as described above. This was mostly the case of samples taken from mammals, insects and food. We selected 12 of these that have particularly interesting features in terms of plasmid content, *in vivo* phenotype and/or isolation history, which we expect to contribute specific insights of interest to this study (described below and in Table 1B):

- One strain of *B. thuringiensis* that is commonly used as a biopesticide contains a pXO2-like plasmid; it will be in the public interest to know whether that strain possesses any pathogenomic potential.
- Three isolates from wild mammals, combined with three isolates implicated in cases of emetic food poisoning, all carrying pXO1-like plasmids, will help us find out whether the pXO1-like plasmids found in wild mammals, a potential reservoir for these plasmids, are more similar to the pXO1-like plasmids implicated in emetic food poisoning cases or to those isolated from clinical cases of anthrax-like disease (sequenced elsewhere), or whether they belong to yet another category.
- Five strains of various environmental origins, which we obtained individually from collaborators, carry both a pXO1-like and a pXO2-like plasmid, the rarest and potentially most worrying combination; their sequences will greatly enrich the dataset.

2) We also included “outlier” strains that tested negative for the respective pXO1 and pXO2 replication genes, but tested positive for other sequences of known pXO1-like and/or pXO2-like plasmids that appear to be located on other large plasmids or on the chromosome (Table 1C). Sequencing these isolates will provide insight into proximal connections that exist between the pXO1- and pXO2-like plasmids and the rest of the *B. cereus* gene pool, an important aspect of molecular evolution in the *B. cereus* group.

This collection of isolates is currently available at the Kolter Laboratory (Harvard Medical School, Boston) and will be deposited at BEI in accordance with NIAID policy.

**Table 1**

<b>A. Core set</b>					
Distribution of matched isolates arranged by plasmid content combinations					
	pXO1-like	pXO2-like	Both	Neither	<i>Isolates</i>
Type A	X	X		X	24
Type B	X			X	20
Type C		X		X	20
Type D		X	X	X	6
Type E	X	X	X	X	4
<i>Subtotal</i>					74
<b>B. Features of interest of the unmatched isolates</b>					
<i>B. thuringiensis</i> strain used as biopesticide carrying a pXO2-like plasmid					1
<i>B. cereus</i> isolates from wild mammals, carrying a pXO1-like plasmid					3
<i>B. cereus</i> isolates from food, carrying a pXO1-like plasmid					3
<i>B. cereus</i> isolates from various origins carrying both plasmids					5
<i>Subtotal</i>					12
<b>C. Features of interest of the outlier strains</b>					
Other large plasmids sharing non-replicon-related segments with pXO1-like and/or pXO2-like plasmids					7
Chromosomal sequences similar to segments of pXO1-like plasmids					3
<i>Subtotal</i>					10
<b>Total</b>					<b>96</b>

With the exception of one selected strain that is the wild-type of a strain commonly used in commercial biopesticide formulations (*B. thuringiensis* var. *kurstaki* HD73), all strains included in this selection are wild-type isolates originating from studies by the Laboratory of Food and Environmental Microbiology (UCL, Belgium) and direct collaborators. The original samples were collected in Belgium, Denmark, Greenland, Poland, Scotland, Switzerland, Spain, Tunisia, Ivory Coast, Guadeloupe, Martinique, Abu Dhabi (United Arab Emirates), China, Canada, USA, and the Concordia research station in Antarctica. None of these strains has been or is being sequenced elsewhere.

### Sequencing Strategy

We propose to use 454 XLR-based sequencing to generate the data for this project. High quality plasmid genomes are critical to the success of this project. Previous experience suggests that, in the absence of complete finishing, obtaining and identifying high quality draft plasmid genomes is difficult. To assess the optimal strategy for sequencing plasmid genomes, the sequencing plan will occur in two phases.

#### Phase 1:

The goal of phase 1 is to optimize plasmid genome sequencing. We will choose five

strains, three that have both plasmids (pXO1 and pXO2), one that has pXO1, and one that has pXO2. These five strains will undergo standard genome sequencing. These five strains will also have plasmid DNA isolated separately, and these separate plasmid preparations will be sequenced separately. We will determine if sequencing plasmids by themselves significantly improves plasmid assembly and coverage. We will also assess the extent to which informatic or manual finishing improves plasmid genomes.

### **Phase 2:**

Based on the results of phase 1, in phase 2, we will sequence the remaining strains using a strategy that optimizes the tradeoff between plasmid genome quality and resources. Not only will this sequencing strategy aid the goals of this project, but it will also yield methods that may improve plasmid genome sequencing for other projects.

### **Community Interest**

The project was conceived to create a community resource that will provide the foundation for a broad range of research interests including, but not limited to, the identification and characterization of bacilli with pathogenomic potential such as atypical *B. anthracis*-like bacilli. We have assembled a Bacillus Virulence Plasmid Consortium which has provided input into the specific strains to be sequenced ensuring that they are selected to meet the broadest range of needs. In addition to advising strain selection, several individuals will bring their specific areas of expertise to this project. Contributors to this proposal and those who will contribute to the analysis of the data are listed below.

### **Data Release Policy**

All sequences and trace files (chromatograms) generated under this proposal will be submitted to the Trace Archive at NCBI/NLM/NIH on a weekly basis. These data will also include information on templates, vectors, and quality values for each sequence.

Genome assemblies will be made available via GenBank and the Broad web site (<http://www.broad.mit.edu/seq/msc/>). Assembled contigs and scaffolds will be deposited in the Whole Genome Shotgun (WGS) section of GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/wgs.html>, within 45 calendar days of completing the shotgun or high-throughput sequencing. When finishing work is completed, assembled contigs and scaffolds will be deposited in GenBank within 45 days of completing the finishing work. If it is determined that the final assembly can be significantly improved, an updated record will be deposited in the appropriate part of GenBank when complete.

Annotation data will be made available via GenBank and the Broad web site (<http://www.broad.mit.edu/seq/msc/>) after consistency checks and quality control have been completed by the MSC and collaborators. Assuming no significant errors are detected during the validation process, annotation data will be released within 45 calendar days of being generated.

Candidate polymorphisms identified by comparison of new genome sequences to a reference will be deposited in dbSNP at NCBI and released through the Broad MSC's web site. Prior to public release, polymorphisms will be made available to collaborating scientists for a one week period for quality control purposes. Candidate polymorphisms

will then be released barring any quality issues discovered in the QC process.

### **Bacillus Virulence Plasmid Consortium**

**[might modify description of specific expertise/contributions in final version]**

Fergus Priest (Heriot-Watt University, UK) and Per Einar Granum (NORVSCI, Norway) are leaders in the study of the phylogenetic structure of the *B. cereus* complex. They will provide guidance on how to best explore the diversity of the strains collection and where it fits within the wider context of *B. cereus*.

Anne-Brit Kolstø (University of Oslo, Norway) is a leading expert in the field of *B. cereus* phylogenomics as well as the study of *B. cereus* mobile genetic elements. She will contribute to the comparative genome analysis of the collection.

Jacques Mahillon (UCL Louvain-la-Neuve, Belgium) is a leading expert on mobile genetic elements and horizontal gene transfer in *B. cereus*. He will assist in the identification, classification and comparative analysis of plasmids, transposons and other mobile genetic elements in the genomes of the collection.

David Rasko (University of Maryland School of Medicine, USA) possesses extensive experience in comparative genomics of pathogens with a focus on plasmids and other mobile elements. He will contribute to the genomic analysis of the plasmids.

Paul Keim (Northern Arizona University, USA) is a leading expert in the field of pathogen genomics, especially as relates to anthrax. He will provide guidance for the identification of pathogenomic content in the collection.

Theresa Koehler (University of Texas, USA) and Agnès Fouet (CNRS/Institut Pasteur, France) are leading experts on *B. anthracis* pathogenesis and anthrax disease and will assist in assessing the pathogenic potential of isolates in the collection.

Didier Lereclus (INRA, France) is a leading expert in the genetics of *B. cereus* and entomopathogenicity of *B. thuringiensis*. He will contribute to the assessment of the pathogenic potential of isolates in the collection.

Jacques Ravel (University of Maryland School of Medicine, USA) **has extensive experience in comparative genomics of pathogens including *B. anthracis* and *B. cereus*. He has contributed to the selection of isolates for this project.**

### **Broad Institute Personnel:**

Michael Feldgarden  
Bruce Birren

### **Project Leader:**

Geraldine A. Van der Auwera (Kolter Lab, Harvard Medical School)

## References

**Avashia SB, Riggins WS, Lindley C, Hoffmaster A, Drumgoole R, Nekomoto T, Jackson PJ, Hill KK, Williams K, Lehman L, Libal MC, Wilkins PP, Alexander J, Tvaryanas A, Betz T. 2007** Fatal pneumonia among metalworkers due to inhalation exposure to *Bacillus cereus* Containing *Bacillus anthracis* toxin genes. *Clin Infect Dis.* 44:414-6.

**Ehling-Schulz M, Fricker M, Grallert H, Rieck P, Wagner M, Scherer S. 2006** Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. *BMC Microbiol.* Mar 2;6:20.

**Hoffmaster AR, Hill KK, Gee JE, Marston CK, De BK, Popovic T, Sue D, Wilkins PP, Avashia SB, Drumgoole R, Helma CH, Ticknor LO, Okinaka RT, Jackson PJ. 2006** Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. *J Clin Microbiol.* 44:3352-60.

**Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, Marston CK, De BK, Sacchi CT, Fitzgerald C, Mayer LW, Maiden MC, Priest FG, Barker M, Jiang L, Cer RZ, Rilstone J, Peterson SN, Weyant RS, Galloway DR, Read TD, Popovic T, Fraser CM. 2004** Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc Natl Acad Sci U S A* 101: 8449-54.

**Hoton FM, Andrup L, Swiecicka I, Mahillon J. 2005** The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. *Microbiology.* 151:2121-4.

**Hu X, Van der Auwera G, Timmerly S, Zhu L, Mahillon J. 2009** Distribution, diversity, and potential mobility of extrachromosomal elements related to the *Bacillus anthracis* pXO1 and pXO2 virulence plasmids. *Appl Environ Microbiol.* 75:3016-28.

**Klee SR, Ozel M, Appel B, Boesch C, Ellerbrok H, Jacob D, Holland G, Leendertz FH, Pauli G, Grunow R, Nattermann H. 2006** Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *J Bacteriol.* 188:5333-44.

**Tourasse NJ, Helgason E, Økstad OA, Hegna IK, Kolstø AB. 2006** The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *J Appl Microbiol.* 101:579-93.

**Tourasse NJ, Kolstø AB. 2008** SuperCAT: a supertree database for combined and integrative multilocus sequence typing analysis of the *Bacillus cereus* group of bacteria (including *B. cereus*, *B. anthracis* and *B. thuringiensis*). *Nucleic Acids Res.* 36:D461-8.