Differentiation of *Bacillus anthracis*, *Bacillus cereus*, *Bacillus thuringiensis* based on *csaB* gene reflect their hosts’ source

Jinshui Zheng¹, Donghai Peng¹, Xiaoling Song¹, Lifang Ruan¹, Jacques Mahillon² and Ming Sun¹*

¹State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China, and
²Laboratory of Food and Environmental Microbiology, Université catholique de Louvain, B-11348 Louvain-la-Neuve, Belgium.

*Corresponding author: Sun Ming, E-mail: m98sun@mail.hzau.edu.cn; Tel: 86-27-87283455; Fax: 86-27-87280670.

Running title: *csaB*- based discrimination among Bc group strains
Abstract

csaB gene analysis clustered 198 strains of B. anthracis, B. cereus and B. thuringiensis into two groups related to mammalian and insect hosts, respectively. Mammal-related group I strains also have more SLH protein genes than group II strains. This indicates that csaB-based differentiation reflects selective pressure from animal hosts.

Keywords: Bacillus anthracis, Bacillus cereus, Bacillus thuringiensis, csaB, SLH protein
Differentiation among members of the *Bacillus cereus* group has been conducted by many research groups based on single or multiple gene markers (1-10). For most of these typing analyses, *B. anthracis* could be discriminated from the others, whereas *B. cereus* and *B. thuringiensis* strains tended to intermingle with each other (4, 11).

One reason for the difficulty in discriminating within the *B. cereus* group is that the selected marker genes are highly conserved among the tested strains. Consequently, genes or sequences with higher levels of genetic diversity should be screened and selected to better discriminate among the *B. cereus* group strains and isolates.

The surface layer (S-layer) is the outmost cell structure of many archaea and bacteria, and consists of protein(s) or glycoprotein(s) (12). All S-layer proteins in *B. anthracis*, *B. cereus* and *B. thuringiensis* display three S-layer homology domains (SLH, PF00395). In *B. anthracis*, there are more than 20 SLH proteins, and each protein has three copies of the SLH domain (13). Product of the *csaB* gene is involved in the addition of a pyruvyl group to a peptidoglycan-associated polysaccharide fraction, a modification necessary for the binding of SLH proteins to the secondary cell wall polymer (SCWP) in some Gram-positive bacteria (14). In this study, we evaluated *csaB* as a marker for the identification and/or discrimination of *B. anthracis*, *B. cereus* and *B. thuringiensis* strains.

When focusing on the 122 strains (21 *B. anthracis*, 79 *B. cereus* and 22 *B. thuringiensis*, Table S1) whose genome sequences are available from GenBank, we found that the *csaB* gene locates on each chromosome as single copy. The *csaB* sequences show more diversity (nucleotide sequence identity range/average:...
75-100%/84%) than other marker genes, (e.g. 93%~100%/97% for groEL, 90%~100%/97% for sodA and 86%~100%/92% for gyrB). The major topology of the phylogenetic tree based on csaB sequences was similar to those of groEL, sodA and gyrB for the 122 selected strains (data not shown).

In addition, we amplified another 76 csaB fragments from B. thuringiensis strains (Table S2) with primers (P1: 5’-GTGCAGTTAGTCTTATCAGGAT-3’ and P2: 5’-CTTTCGCATCCCAATAMCKYACACT-3’) and sequenced the amplicons in both directions (Text S1). An unrooted phylogenetic tree was constructed based on the above 198 sequences using the neighbor-joining algorithm implemented in MEGA5 (15) after alignment by CLUSTALW (16). All strains could be assembled into two groups, I and II (Fig. 1). Members of Group I were subdivided into two subgroups, Ia and Ib (Fig. 1). Subgroup Ia contains 21 B. anthracis, 28 B. cereus strains including 9 higher-animal illnesses associated isolates, and 31 B. thuringiensis strains. Subgroup Ib contains 1 B. thuringiensis strain (4BQ1) and 15 B. cereus strains. All the 6 emetic B. cereus strains with available csaB genes are located in Group I; four (AH187, AND1407, H3081.97 and NC7401) in Ia and two (CER057 and CER074) in Ib. Emetic B. cereus produces cereulide, which causes nausea and vomiting after ingestion. Two subgroups of Group II were supported by high bootstrap values. Subgroup IIa includes 4 B. cereus strains (MM3, R309803, BAG6X1-1and BAG2X1-2) and B. thuringiensis subsp. konkukian str. 97-27 which was previously shown to be more closely related to B. anthracis (6). Higher levels of genetic conservation of the csaB were observed within subgroup IIb. Thirty out of 79
tested *B. cereus* and 66 out of 98 tested *B. thuringiensis* strains are distributed on 35 branches, indicating that *B. thuringiensis* strains are predominately found in subgroup IIb. Among the 96 strains in this subgroup, only 5 (172560W, AH1134, B4264, F65185 and G9842) were reported as human pathogens. *B. cytotoxicus* NVH 391-98 which was isolated from an outbreak of food poisoning (17) and *B. cereus* BAG5X-1 which was isolated from soil are not related to the two groups. Other methods also identified the former strain as outlier (18).

The *csaB* gene plays a crucial role in the maintenance of a family of proteins covering the whole cell; therefore, its diversity could indirectly be driven by the surrounding environment. We therefore investigated ecological niches and origins of the strains in the two *csaB* clusters. More than 80% of the strains (40 out of 48 strains) isolated from humans and other mammals were clustered in Group I. These strains particularly include most of the strains pathogenic to mammals, i.e. all 21 *B. anthracis* strains, 11 *B. cereus* strains associated with human or animal infection (03BB102, 95/8201, AH1272, AH1273, AH820, E33L, F837/76, G9241, AH1271, IS075 and MSX-A12), and the 8 emetic *B. cereus* strains. Strains isolated from insects were clustered predominantly in group II. Among the 22 *B. thuringiensis* strains that were isolated from insects, only BGSC 4B2 which was isolated from *Malacosoma distria*, is in Group I. Particularly all strains of *B. thuringiensis* displaying high level of toxicity against insects and used as biopesticide fit into Subgroup IIb, including *B. thuringiensis* subsp. *thuringiensis* strains T01001 and ATCC 10792, *B. thuringiensis* subsp. *kurstaki* strains 4D11, HD-1, T03a001 and YBT-1520, and *B. thuringiensis*...
subsp. israelensis strains ATCC 35646 and Bt4Q7. Accordingly, csaB gene sequences show strong correlation to mammalian or insect hosts, indicating convergent evolution driven by host adaptation. However, when the 52 soil-borne strains were considered, they were found to be almost equally distributed between Groups I (24 strains) and II (28 strains). Soil does not appear to exert a major selective pressure for sequence diversity of csaB in strains of the B. cereus group. However, soil is a reservoir for both insect pathogens and strains pathogenic to mammals, and most of the soil isolates are not pathogenic.

CsaB is essential for anchoring the SLH proteins to the SCWP. SLH proteins are located on the outermost layer of bacteria and some of these proteins mediate bacterial adhesion to their host (19, 20). Therefore, the distribution and evolution of all the SLH proteins present in the 122 available B. cereus group strains (Table S1) were investigated. SLH protein sequences were retrieved using HMMER software with hmmsearch command (21) and the SLH domain model PF00395 (Fig. 2 and Table S3). Strains in Group I possess the most abundant number of SLH protein genes, with up to 26 genes in strains B. cereus 03BB108, B. thuringiensis BGSC 4AJ1 and 4CC1, while the Group II strains harbor significantly fewer (p-value < 2.2e-16, Mann-Whitney test). The smallest number (8 SLH protein genes) was found in several strains in Subgroup IIb. For function prediction, all of the SLH proteins were searched against CDD database (22), and could be classified into 13 categories (Table S3). SLH proteins with basal metabolism function, such as those involving in peptidoglycan catabolism, are conserved and distribute in all the strains among Group
I and II. In contrast, SLH proteins not involving in basal metabolism show more diversity and narrower distribution among the two groups. Most SLH proteins which contribute to the bacterial adaptation to higher animal hosts are contained by Group I strains. For instance, S-layer proteins have been reported to mediate bacterial resistance against bactericidal complement activity, to participate in the adhesion to extracellular matrix proteins, and to temper the pro-inflammatory cytokine response (12, 23). The genes encoding these proteins are mainly contained by strains in Group I but not Group II (Table S3). The gene encoding adhesion protein BlsA which mediates adherence of vegetative form of B. anthracis strain to human cells (19) was only found in several B. cereus strains in Group I (G9241, biovar anthracis str. CI and 03BB102) which exhibit B. anthracis-like characteristics (24). Other examples included genes the Ca++ binding domain proteins which are mainly found in Group I and Subgroup IIa strains and the lactamase proteins occurring in strains of Subgroup Ia. Their products are also usually involved in bacterial adaptation to their hosts.

Taken together, these observations indicate that the two csaB-based groups correspond to two distinct lifestyle environments, higher animals and insects for group I and II, respectively.

Acknowledgments

This work was supported by grants from the National High Technology Research and Development Program (863) of China (2011AA10A203), the National Basic Research Program (973) of China (2009CB118902), China 948 Program of Ministry
of Agriculture (2011-G25), and the National Natural Science Foundation of China (39870036 and 31270137). We also want to thank A. Gillis for her critical reading of the manuscript.

Supplemental Material

Table S1 Bacillus strains used in this study whose genome sequences are available from GenBank

Table S2 B. thuringiensis strains used in this study

Table S3 Distribution of SLH protein genes in genomes of B. cereus group strains

Text S1 DNA sequencing for csaB genes
References


Forsberg LS, Choudhury B, Leoff C, Marston CK, Hoffmaster AR, Saile E, Quinn CP, Kannenberg EL, Carlson RW. 2011. Secondary cell wall polysaccharides from Bacillus cereus strains G9241, 03BB87, and 03BB102 causing fatal pneumonia share similar glycosyl structures with the polysaccharides from Bacillus anthracis. Glycobiology.
Fig. 1. Unrooted phylogenetic tree of csaB gene of *B. anthracis*, *B. cereus* and *B. thuringiensis* strains. The number at each branch point represents the percentage of bootstrap support calculated from 1,000 replicates. Only bootstrap values above 50 are shown. *B. a. Ames* represents all the 21 *B. anthracis* strains. *B. c ATCC 14579* represents *B. cereus* strains ATCC 14579, ATCC 10876, AH676, BAG3X2-2, BAG4X12-1, BDRD-Cer4, F65185, VD156 and VD169. *B.t 4AE1*, represents *B. thuringiensis* strains 4AE1, 4AM1, 4AQ1, 4AT1, 4CB1, 4D11, 4G1, 4L1, 4T1, 4W1, 4X1, HD-1, HD73, T03a001 and YBT-1520. *B.t ATCC 35646* represents *B. thuringiensis* strains ATCC 35646, 4AK1, 4AX1, 4BS1, 4BZ1, 4M1, BMB171, Br4Q7, HD-789, T69001.

Fig. 2. Summary of SLH protein genes distribution. I and II refer for the two groups of csaB phylogenetic tree (Fig. 1). In boxes, bold line represents median, and upper and lower boundaries represent 75th and 25th percentiles, respectively. Group I strains have significantly more of SLH protein genes than those from Group II (p-value < 2.2e-16, Mann-Whitney test).