



**HAL**  
open science

## Identification and Validation of Specific Markers of *Bacillus anthracis* Spores by Proteomics and Genomics Approaches

Jérôme Chenau, François Fenaille, Valérie Caro, Michel Haustant, Laure Diancourt, Silke Klee, Christophe Junot, Eric Ezan, Pierre L. Goossens, François Becher

### ► To cite this version:

Jérôme Chenau, François Fenaille, Valérie Caro, Michel Haustant, Laure Diancourt, et al.. Identification and Validation of Specific Markers of *Bacillus anthracis* Spores by Proteomics and Genomics Approaches. *Molecular and Cellular Proteomics*, 2014, 13 (3), pp.716-732. 10.1074/mcp.M113.032946 . hal-04403928

HAL Id: hal-04403928

<https://hal.science/hal-04403928v1>

Submitted on 23 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Identification and Validation of Specific Markers of *Bacillus anthracis* Spores by Proteomics and Genomics Approaches\*<sup>§</sup>

Jérôme Chenau‡, François Fenaille‡, Valérie Caro§, Michel Haustant¶, Laure Diancourt§, Silke R. Klee||, Christophe Junot‡, Eric Ezan‡, Pierre L. Goossens‡¶¶\*\*, and François Becher‡‡‡

*Bacillus anthracis* is the causative bacteria of anthrax, an acute and often fatal disease in humans. The infectious agent, the spore, represents a real bioterrorism threat and its specific identification is crucial. However, because of the high genomic relatedness within the *Bacillus cereus* group, it is still a real challenge to identify *B. anthracis* spores confidently. Mass spectrometry-based tools represent a powerful approach to the efficient discovery and identification of such protein markers. Here we undertook comparative proteomics analyses of *Bacillus anthracis*, *cereus* and *thuringiensis* spores to identify proteoforms unique to *B. anthracis*. The marker discovery pipeline developed combined peptide- and protein-centric approaches using liquid chromatography coupled to tandem mass spectrometry experiments using a high resolution/high mass accuracy LTQ-Orbitrap instrument. By combining these data with those from complementary bioinformatics approaches, we were able to highlight a dozen novel proteins consistently observed across all the investigated *B. anthracis* spores while being absent in *B. cereus*/*thuringiensis* spores. To further demonstrate the relevance of these markers and their strict specificity to *B. anthracis*, the number of strains studied was extended to 55, by including closely related strains such as *B. thuringiensis* 9727, and above all the *B. cereus* biovar *anthracis* CI, CA strains that possess pXO1- and pXO2-like plasmids. Under these conditions, the combination of proteomics and genomics approaches confirms the per-

tinence of 11 markers. Genes encoding these 11 markers are located on the chromosome, which provides additional targets complementary to the commonly used plasmid-encoded markers. Last but not least, we also report the development of a targeted liquid chromatography coupled to tandem mass spectrometry method involving the selection reaction monitoring mode for the monitoring of the 4 most suitable protein markers. Within a proof-of-concept study, we demonstrate the value of this approach for the further high throughput and specific detection of *B. anthracis* spores within complex samples. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.M113.032946, 716–732, 2014.

*Bacillus anthracis* is a highly virulent bacterium, which is the etiologic agent of anthrax, an acute and often lethal disease of animals and humans (1). According to the Centers for Disease Control and Prevention, *B. anthracis* is classified as a category A agent, the highest rank of potential bioterrorism agents (<http://www.bt.cdc.gov/agent/agentlist-category.asp>). The infectious agent of anthrax, the spore, was used as a bioterrorism weapon in 2001 in the United States when mailed letters containing *B. anthracis* spores caused 22 cases of inhalational and/or cutaneous anthrax, five of which were lethal (2). These events have emphasized the need for rapid and accurate detection of *B. anthracis* spores.

*Bacillus anthracis* is a member of the genus *Bacillus*, Gram-positive, rod-shaped bacteria characterized by the ability to form endospores under aerobic or facultative anaerobic conditions (3). The genus *Bacillus* is a widely heterogeneous group encompassing 268 validly described species to date (<http://www.bacterio.net/b/bacillus.html>, last accessed on August 9<sup>th</sup> 2013). *B. anthracis* is part of the *B. cereus* group which consists of six distinct species: *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides*, and *B. weihenstephanensis* (4, 5). The latter three species are generally regarded as nonpathogenic whereas *B. cereus* and *B. thuringiensis* could be opportunistic or pathogenic to mammals or insects (5, 6). *B. cereus* is a ubiquitous species that lives in soil but is also found in foods of plant and animal

From the ‡CEA, iBiTec-S, Service de Pharmacologie et d'Immunoanalyse, 91191 Gif-sur-Yvette, France; §Institut Pasteur, Genotyping of Pathogens and Public Health, 75015 Paris, France; ¶Institut Pasteur, Pathogénie des Toxi-Infections Bactériennes, 75724 Paris Cedex 15, France; ||Robert Koch-Institut, Centre for Biological Threats and Special Pathogens (ZBS2), Berlin, Germany; \*\*CNRS URA 2172, 75015 Paris, France

Received August 9, 2013, and in revised form, December 23, 2013  
Published, MCP Papers in Press, December 29, 2013, DOI 10.1074/mcp.M113.032946

Author contributions: J.C., F.F., C.J., E.E., P.L.G., and F.B. designed research; J.C. performed research; V.C., L.D., and S.R.K. contributed new reagents or analytic tools; J.C., F.F., P.L.G., and F.B. analyzed data; J.C., F.F., and F.B. wrote the paper; M.H. assisted with sample preparation; C.J. and E.E. critically reviewed the paper; P.L.G. critically reviewed the paper and provided strains.

origin, such as dairy products (7). Its occurrence has also been linked to food poisoning and it can cause diarrhea and vomiting (6, 8). *B. thuringiensis* is primarily an insect pathogen, also present in soil, and often used as a biopesticide (9).

*B. anthracis* is highly monomorphic, that is, shows little genetic variation (10), and primarily exists in the environment as a highly stable, dormant spore in the soil (1). Specific identification of *B. anthracis* is challenging because of its high genetic similarity (sequence similarity >99%) with *B. cereus* and *B. thuringiensis* (5, 11). The fact that these closely related species are rather omnipresent in the environment further complicates identification of *B. anthracis*. The main difference among these three species is the presence in *B. anthracis* of the two virulence plasmids pXO1 and pXO2 (1), which are responsible for its pathogenicity. pXO1 encodes a tripartite toxin (protective antigen (PA), lethal factor (LF), and edema factor (EF)) which causes edema and death (1), whereas pXO2 encodes a poly- $\gamma$ -D-glutamate capsule which protects the organism from phagocytosis (1). *B. anthracis* identification often relies on the detection of the genes encoded by these two plasmids via nucleic acid-based assays (12–14). Nevertheless, the occasionally observed loss of the pXO2 plasmid within environmental species may impair the robustness of detection (1). In addition, in recent years a series of findings has shown that the presence of pXO1 and pXO2 is not a unique feature of *B. anthracis*. Indeed, Hu *et al.* have demonstrated that ~7% of *B. cereus*/*B. thuringiensis* species can have a pXO1-like plasmid and ~1.5% a pXO2-like plasmid (15). This was particularly underlined for some virulent *B. cereus* strains (*i.e.* *B. cereus* strains G9241, *B. cereus* biovar *anthracis* strains CA and CI) (16–20).

Because of these potential drawbacks, the use of chromosome-encoded genes would be preferable for the specific detection of *B. anthracis*. Such genes (*rpoB*, *gyrA*, *gyrB*, *plcR*, BA5345, and BA813) have been reported as potential markers (21–25), but concerns have also been raised about their ability to discriminate *B. anthracis* efficiently from closely related *B. cereus* strains (26). Ahmod *et al.* have recently pointed out, by *in silico* database analysis, that a specific sequence deletion (indel) occurs in the *yeaC* gene and exploited it for the specific identification of *B. anthracis* (27). Nevertheless, a few *B. anthracis* strains (*e.g.* *B. anthracis* A1055) do not have this specific deletion and so may lead to false-negative results (27).

In the last few years, protein profiling by MS, essentially based on matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF MS), has emerged as an alternative (or a complement) to genotypic or phenotypic methods for the fast and efficient identification of microorganisms (28, 29). Such an approach is based on the reproducible acquisition of global bacterial protein fingerprints/patterns. The combination of MS-based protein patterns and chemometric/bioinformatic tools has been demonstrated to efficiently differentiate members of the *B. cereus* group from other *Bacillus* species (30).

However, the specific discrimination of *B. anthracis* from the closely related *B. cereus* and *B. thuringiensis* remains difficult (30). This study of Lasch and coworkers, performed on vegetative cells, identified a few ribosomal and spore proteins as being responsible for this clustering (30). Closer inspection of the data revealed that *B. anthracis* identification was essentially based on one particular isoform of the small acid-soluble spore protein B (SASP-B)<sup>1</sup> (30–34), which is exclusively expressed in spores, as the samples were shown to contain residual spores. However, the specificity of SASP-B has recently been questioned as the published genomes of *B. cereus* biovar *anthracis* CI and *B. thuringiensis* BGSC 4CC1 strains have been shown to share the same SASP-B isoform as *B. anthracis* (35). Altogether these results underline that the quest for specific markers of *B. anthracis* needs to be pursued.

Mass spectrometry also represents a powerful tool for the discovery and identification of protein markers (36, 37). In the case of *B. anthracis*, this approach has more commonly been used for the comprehensive characterization of given bacterial proteomes. For example, the proteome of vegetative cells with variable plasmid contents has been extensively studied (38–40), as the proteomes of mature spores (41, 42) and of germinating spores (43, 44). Only one recent study, based on a proteo-genomic approach, was initiated to identify protein markers of *B. anthracis* (45). In this work, potential markers were characterized but using a very limited number of *B. cereus* group strains (three *B. cereus* and two *B. thuringiensis*). Moreover, this study was done on vegetative cells, whereas the spore proteome is drastically different. To our knowledge, no study has characterized and validated relevant protein markers specific to *B. anthracis* spores, which constitute the dissemination form of *B. anthracis* and are often targeted by first-line immunodetection methods (46).

Here we report comparative proteomics analyses of *Bacillus anthracis/cereus/thuringiensis* spores, undertaken to identify proteoforms unique to *B. anthracis*. Preliminary identification was performed on a limited set of *Bacillus* species both at the peptide (after enzymatic digestion) and protein levels by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) using a high resolution/high mass accuracy LTQ-Orbitrap instrument. The pertinence of 11 markers was further demonstrated using proteomics and genomics approaches on a representative larger set of up to 55 different strains, including the closely related *B. cereus* biovar *anthracis* CI, CA, and *B. thuringiensis* 9727. Lastly, as a proof-of-concept

<sup>1</sup> The abbreviations used are: SASPs, small acid-soluble spore proteins; SRM, selected reaction monitoring; LC-MS/MS, liquid chromatography coupled to tandem mass spectrometry; IAA, iodoacetamide; RS9, 10, 30S ribosomal protein S9, S10; RL5, 28, 50S ribosomal protein L5, L28; GSP26, General Stress Protein 26; DPS1, DNA protection during starvation protein 1; IS, Internal Standard; MLST, multilocus sequence typing; MLEE, multilocus enzyme electrophoresis; AFLP, amplified fragment length polymorphism.

study, we also report for four *B. anthracis* markers the implementation of a targeted LC-MS/MS method using selected reaction monitoring (SRM), based on the extension of a previous one focused on SASP-B (35). Preliminary results regarding method usefulness for the high throughput and accurate detection of *B. anthracis* spores in complex samples were also obtained and will be reported herein.

### EXPERIMENTAL PROCEDURES

**Chemicals and Reagents**—Brain-heart infusion (BHI) and nutrient broth yeast extract (NBY) media were from Difco (Difco, Detroit, MI) and Radioselectan was from Schering (Renografin 76%; Schering, Lys Lez-Lannoy, France). Sequencing grade modified trypsin was from Promega (Promega, Madison, WI, USA) and endoproteinase Glu-C from Sigma (Sigma Chemical Co.). Synthetic peptides were synthesized by Bachem (Burgdorf, Switzerland). Tween-80, lecithin, histidine and tryptic soy broth (TSB) were from Sigma-Aldrich (Sigma Chemical Co.). Ultrapure water was from a Milli-Q<sub>plus</sub> 185 purifier (Millipore, Bedford, MA), HPLC-grade acetonitrile was from SDS (Peypin, France), and analytical-grade formic acid and trifluoroacetic acid (TFA) were from Sigma-Aldrich (Sigma Chemical Co.).

**Safety Considerations**—*Bacillus anthracis* is a highly virulent bacterium and all strains were handled according to safety considerations applicable at the Institut Pasteur (Biosafety Level 3 laboratory). After protein extraction, complete inactivation and elimination of spores were systematically checked (see below). After this control, protein samples can be handled safely using standard laboratory practice.

**Bacterial Strains Included in the Study and Preparation of Spores**—An overview of the 55 strains from the *B. cereus* group used throughout this study is provided in Table I. Presence or absence of pXO1 and pXO2 plasmids was evaluated by PCR (47). Most of the bacterial strains originated from the strain collection of the laboratory of Pathogénie des Toxi-Infections Bactériennes (Institut Pasteur, Paris, France). *B. cereus* biovar *anthracis* CI and CA were from the Robert Koch Institute (17, 19). All strains were handled according to safety considerations applicable at the Institut Pasteur (Biosafety Levels 2 and 3). *B. anthracis* 9602R, RA3R, 7611R, and 6183R are derivatives of the virulent natural isolates 9602, RA3, 7611, and 6183, cured of plasmid pXO2 and therefore do not produce a capsule but still produce toxins (*i.e.* phenotypic equivalent of the Sterne strain) (48). The CAR-P strain is derived from the virulent natural isolate CA, cured of plasmid pXO2 and deleted of the pagA gene. Strains were grown in BHI medium and spore production was performed for 48 strains (see Table I) on NBY medium as previously described (49). Sporulation could not be achieved for the seven remaining strains, which were therefore only used for the genomic approach. Spores from these 48 strains were then purified on Radioselectan (Renografin 76%) using previously described methods, and were stored at 4 °C in sterile deionized water until use (49).

**Preparation of Protein Extracts From Bacillus Spores**—Proteins were extracted from the spores by using the trifluoroacetic acid (TFA) protocol for highly pathogenic microorganisms described by Lasch *et al.* (50). Briefly, TFA solution was added to  $\sim 5 \cdot 10^7$  spores (to a final concentration of 80% TFA) and the mixture was vortexed for 10 min. The total spore extract was then centrifuged at  $14,000 \times g$ , 4 °C for 15 min. Finally, the supernatant was transferred into Millipore's Ultra-free MC filter tubes of 0.22  $\mu\text{m}$  pore size (Millipore, Billerica, MA) and spun at  $10,000 \times g$  for 5 min. The filtrate containing the total protein extract was stored at  $-20$  °C until use. The total inactivation and elimination of spores were systematically checked after neutralization of TFA with 9% Tween-80, 0.9% Lecithin, 3% Histidine in TSB and verification of the absence of viable bacteria in each extract after

cultivation on BHI agar during 3 days, as described by Lasch *et al.* (50). The authors demonstrated that spores of 14 strains were reliably sterilized in each of the 67 separate inactivation replicates performed with the TFA inactivation protocol. We have confirmed the full efficiency of this protocol for spore inactivation in the course of experiments performed on 55 strains. One may consider that spore extracts treated with the current protocol could be handled safely in further experiments (and potentially without any systematic control by cultivation).

Extracted samples could be then handled safely outside the Biosafety Level 3 laboratory. Each protein extract was dried by vacuum centrifugation to eliminate TFA. Proteins were then resuspended in 20  $\mu\text{L}$  of 50 mM ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ) buffer, pH 8.0. Each sample was then reduced by the addition of 5  $\mu\text{L}$  of 45 mM dithiothreitol (DTT) solution (in 50 mM  $\text{NH}_4\text{HCO}_3$  buffer) at 55 °C for 30 min and alkylated by the addition of 5  $\mu\text{L}$  of 100 mM iodoacetamide (IAA) solution (in 50 mM  $\text{NH}_4\text{HCO}_3$  buffer) at room temperature for 45 min. Finally, enzymatic digestion was performed by the addition of 1.5  $\mu\text{L}$  of 1  $\mu\text{g}/\mu\text{L}$  trypsin or Glu-C (in 50 mM  $\text{NH}_4\text{HCO}_3$  buffer) with overnight incubation at 37 °C.

**Marker Discovery by Untargeted Bottom-up Approach**—Enzymatically digested samples were diluted in 0.1% formic acid (1:5, v/v) and injected into the LC-MS/MS system. LC-MS experiments were performed using an Accela HPLC system coupled to an LTQ-Orbitrap Discovery mass spectrometer, both from Thermo Scientific (San Jose, CA). Chromatographic separation was performed on a Zorbax SB-C18 column ( $150 \times 2.1$  mm i.d., 5  $\mu\text{m}$  particle size, 300 Å porosity) from Agilent Technology (Palo Alto, CA). Peptides were eluted from the column at a flow rate of 200  $\mu\text{L}/\text{min}$  using water as mobile phase A and acetonitrile as mobile phase B, both containing 0.1% formic acid. A linear gradient from 5 to 60% B in 40 min was performed. Then the column was washed for 10 min at 95% B. The equilibration time before the next analysis was set at 10 min, so that the time between two experimental cycles was 60 min. Column temperature was maintained at 40 °C. The column effluent was directly introduced into the electrospray source of the LTQ-Orbitrap Discovery mass spectrometer. Analyses were performed in the positive ion mode, and sequential MS2 experiments, based on collision-induced dissociation (CID), were performed using the data-dependent acquisition method. The five most abundant ions (threshold 500 counts, charge states higher than +1) were selected and CID fragment ions were analyzed in the linear ion trap. The source conditions were as follows: capillary temperature, 275 °C; sheath gas flow, 35 arbitrary units; auxiliary gas flow, 20 arbitrary units; capillary voltage, 35 V; ESI spray voltage, 5 kV. The MS survey scan was performed from  $m/z$  300–1500 in the Orbitrap, using a resolution set at 30,000 (full width at half maximum, at  $m/z$  400), and the ion population was held at  $1 \times 10^6$  through the use of automatic gain control. A normalized collision energy of 35% was used, with an activation  $q$  of 0.25 and an activation time of 30 ms. For tandem mass spectrometry in the linear ion trap, the ion population was set to  $1 \times 10^4$  and the precursor isolation width was set to 2  $m/z$  units. FTMS data were collected in the profile mode and ion trap MS/MS data in the centroid mode. Data were acquired using the XCalibur software (version 2.0.7).

Raw files were converted to Mascot compatible .mgf files using Bioworks browser version 3.3.1 SP1 (Thermo, San Jose, CA, USA). Database searches were performed using Mascot server version 2.3.01 through Proteinscape platform version 3.0 (Bruker Daltonics, Bremen, Germany), using the following protein databases downloaded from UniProtKB/TrEMBL (updated July 19, 2010): *B. anthracis* [taxonomy 1392] (55 030 entries); and *B. cereus* [taxonomy 1396] (239 595 entries) and *B. thuringiensis* [taxonomy 1428] (109 966 entries). The parameters for searching were: either trypsin or Glu-C (V8-DE) as digestion enzyme, two missed cleavages allowed, charges states 2+,

3+, fixed modification of cysteine residues (carbamidomethyl), variable oxidation of methionine residues, deamidation of asparagine and glutamine residues, Gln→pyro-Glu (for N-terminal glutamine residues) and loss of N-terminal methionine (-131.04 Da), parent ion tolerance of 10 ppm and fragment mass tolerance of 0.7 amu. Protein identifications were validated if proteins were identified with at least two distinct peptides with a score > 20 and a protein score > 40.

**Marker Discovery by Untargeted Top-down Approach**—For intact protein analysis, each protein extract was diluted in 0.1% formic acid (1:10, v/v) and analyzed in reverse-phase liquid chromatography coupled to an LTQ-Orbitrap Discovery mass spectrometer (LC-MS) from Thermo Scientific. The source conditions were as follows: capillary temperature, 275 °C; sheath gas flow, 30 arbitrary units; auxiliary gas flow, 3 arbitrary units; capillary voltage, 49 V; ESI spray voltage, 4.5 kV. The MS survey scan was performed in the positive ion mode from  $m/z$  500–2000 using a resolution set at 30,000 at  $m/z$  400 and FTMS data were collected in profile mode. Top-down MS/MS experiments were also performed to further confirm protein identity. The precursor isolation width was set to 2  $m/z$  units. A normalized collision energy of 35% was used, with an activation  $q$  of 0.25 and an activation time of 30 ms. Linear ion trap MS/MS data were collected in the centroid mode. These data were manually interpreted using the XCalibur software (version 2.0.7).

**Multiplex SRM Assay**—For each spore sample, trypsin and Glu-C digests (obtained as described above, 20  $\mu$ l of each) were mixed together. Stable isotope-labeled peptides corresponding to each targeted marker were synthesized to be used as internal standards (supplementary Table S1). They were mixed and diluted in 10% acetonitrile, 0.1% formic acid to prepare Internal Standard (IS) stock mixture. The N- and C terminus peptides of SASP-B protein were respectively at 100 and 400 ng/ml and the peptides from the other proteins at 500 ng/ml. Ten microliters of IS stock mixture and 50  $\mu$ l of 0.1% formic acid solution were added for a final sample volume of 100  $\mu$ l (IS mixture at a final concentration of 10 ng/ml for N terminus peptide of SASP-B, 40 ng/ml for C terminus peptide of SASP-B and 50 ng/ml for the 5 other targeted peptides). Thirty microliters of this sample mixture were used for LC-MS/MS analysis.

LC-MS/MS experiments were performed using an HP 1100 HPLC system from Agilent (Palo Alto, CA, USA) coupled to a triple quadrupole TSQ Quantum Ultra mass spectrometer (Thermo Scientific, San Jose, CA, USA). Chromatographic separation was performed on a Zorbax SB-C18 column (150  $\times$  2.1 mm i.d., 5  $\mu$ m particle size, 300 Å porosity) from Agilent Technology. Mass spectrometry and HPLC conditions were optimized using the synthetic peptides of the targeted proteins. The column temperature was maintained at 60 °C. Peptides were eluted from the column at a flow rate of 200  $\mu$ l/min using water as mobile phase A and acetonitrile as mobile phase B, both containing 0.1% formic acid. A first linear gradient from 5 to 24% B in 13 min was performed followed by a rapid gradient from 13 to 13.5 min to reach 38% B. The third linear gradient ranged between 13.5 and 18 min to reach 43% B. The column was washed for 4 min at 95% B and the equilibration time before the next analysis was set at 8 min, so that the time between two cycles was 30 min. The column effluent was directly introduced into the electrospray source of the mass spectrometer. Analyses were performed in the positive ion mode. The electrospray voltage and the capillary voltage were set at 3.9 kV and 35 V, respectively. The sheath, ion sweep, and auxiliary gas flow rates (nitrogen) were optimized at 40, 20, and 15 (arbitrary units), respectively, while the drying gas temperature was set at 350 °C. The SRM method had the following parameters: 0.100  $m/z$  scan width, 0.1 s scan time, 0.7 Q1, 0.7 Q3, 1.5 mTorr Q2 pressure (using argon as the collision gas). Two SRM transitions were monitored for each peptide to improve detection specificity. The fragment ions for each transition were selected to carry the amino acids char-

acteristic of the *B. anthracis* proteoform with the aim of increasing method specificity. Moreover, when possible, product ions were selected with an  $m/z$  ratio higher than that of the parent ion. A signal was considered positive when peak intensity was three times above the noise level for both selected transitions. The SRM acquisition time was divided into five segments. For a given peptide, the retention time observed for the two transitions monitored and the ratio between their peak areas must match those obtained with their corresponding IS peptides. Supplementary Table S1 reports the SRM transitions designed and monitored for each peptide and the optimized analytical parameters (tube lens and collision energy values). Data were acquired and analyzed using the XCalibur software (version 1.4 SR1).

**PCR and Sequencing of Markers**—Total genomic DNA of the 55 strains was extracted from vegetative cells using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. The design of the primers was based on the nucleotide sequence alignment obtained with different strains from the *Bacillus cereus* group available in GenBank, by targeting conserved regions framing each gene of interest and in order generally to amplify the complete gene. If necessary, some primers were degenerated to optimize the amplification (supplementary Table S2). PCR was performed using TaqDNA polymerase (Invitrogen, Carlsbad, CA). Amplification involved a first denaturation step at 94 °C for 2 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s and elongation at 72 °C for 1 min, and by a final cycle of elongation at 72 °C for 2 min. PCR fragments were purified by ultrafiltration before sequencing (Millipore, Molsheim, France). Sequencing reactions were performed using the same PCR primers and the Big Dye Terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and purified by ethanol precipitation. Sequence chromatograms were obtained on automated sequence analyzer ABI3730XL (Applied Biosystems). All amplicons were sequenced on both strands. For sequence analysis, the software BioNumerics version 6.5 (Applied-Maths, Sint-Martens-Latem, Belgium) was used to generate contig assembly, sequence alignments and the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) dendrogram.

## RESULTS

### Discovery of Candidate Protein Markers Specific to *B. anthracis* Spores

Marker discovery was first done on a limited but representative panel of 12 strains (5 *B. anthracis* versus 5 *B. cereus* and 2 *B. thuringiensis*), selected to reflect appropriately the natural diversity and relatedness of the *B. cereus* group (Table I and Fig. 1). The corresponding spore lysates were studied both at the peptide and protein levels using proteomics approaches to highlight *B. anthracis*-specific components (Fig. 2A).

**Identification of *B. anthracis*-specific Peptides Using a Bottom-up Proteomics Approach**—Protein extract from each spore sample was divided into two equal parts and subjected to either trypsin or Glu-C digestion. We postulated that using two enzymes with different specificities would increase both the number of peptides/proteins identified as well as the likelihood of observing discriminating peptides incorporating mutated amino acid residues.

Marker discovery was first initiated through the in-depth analysis of the proteomics data from the five *B. anthracis* strains, obtained in a rather shotgun manner oriented toward major spore proteins using a fast LC gradient going from 5 to

## Identification of Novel Markers of *B. anthracis* Spores

TABLE I

List of *Bacillus* strains included in the study. Dendrograms were obtained with BioNumerics software version 6.5 (Applied-Maths, Sint-Martens-Latem, Belgium). ATCC: American Type Culture Collection. *Bcbva* strains CA and CI are from the collection of the Robert Koch Institute. Other strains are from the collection of the Laboratoire Pathogénie des Toxi-Infections Bactériennes at the Pasteur Institute

Species <sup>a</sup>	pXO1/pXO2 <sup>b</sup>	Strain <sup>c</sup>
<i>B. anthracis</i> (n = 18)	+/+	9602, RA3, 7611, 6183, A210, 5725, 7412, 10517, 9066
	+/-	7702, 9602R, RA3R, 7611R, 6183R, Vollum
	-/+	ATCC 6602, ATCC 4229
	-/-	A3
<i>B. cereus</i> (n = 28)	-/-	ATCC 10987, ATCC 14579, ATCC 10876, AH259, AH819, 9708, 9709, 9724, 9725, 9726, 9804, 9806, 9812, 9813, 9814, B05010, B05011, B06006, B06008, B06009, B06016, B06018, B06020, B06026
	pBCXO1/-	G9241
	pBCXO1/pBCXO2	CI, CA
	pBCXO1/-	CAR-P
<i>B. thuringiensis</i> (n = 9)	-/-	9711, 9712, 9714, 9715, 9717, 9720, 9721, 9727, 407

<sup>a</sup> For each species, the total number of strains studied is given in parentheses.

<sup>b</sup> For each strain, the presence of pXO1 and pXO2 plasmids, or the closely related pBCXO1 and pBCXO2, is noted.

<sup>c</sup> Strains used for marker discovery are underlined.

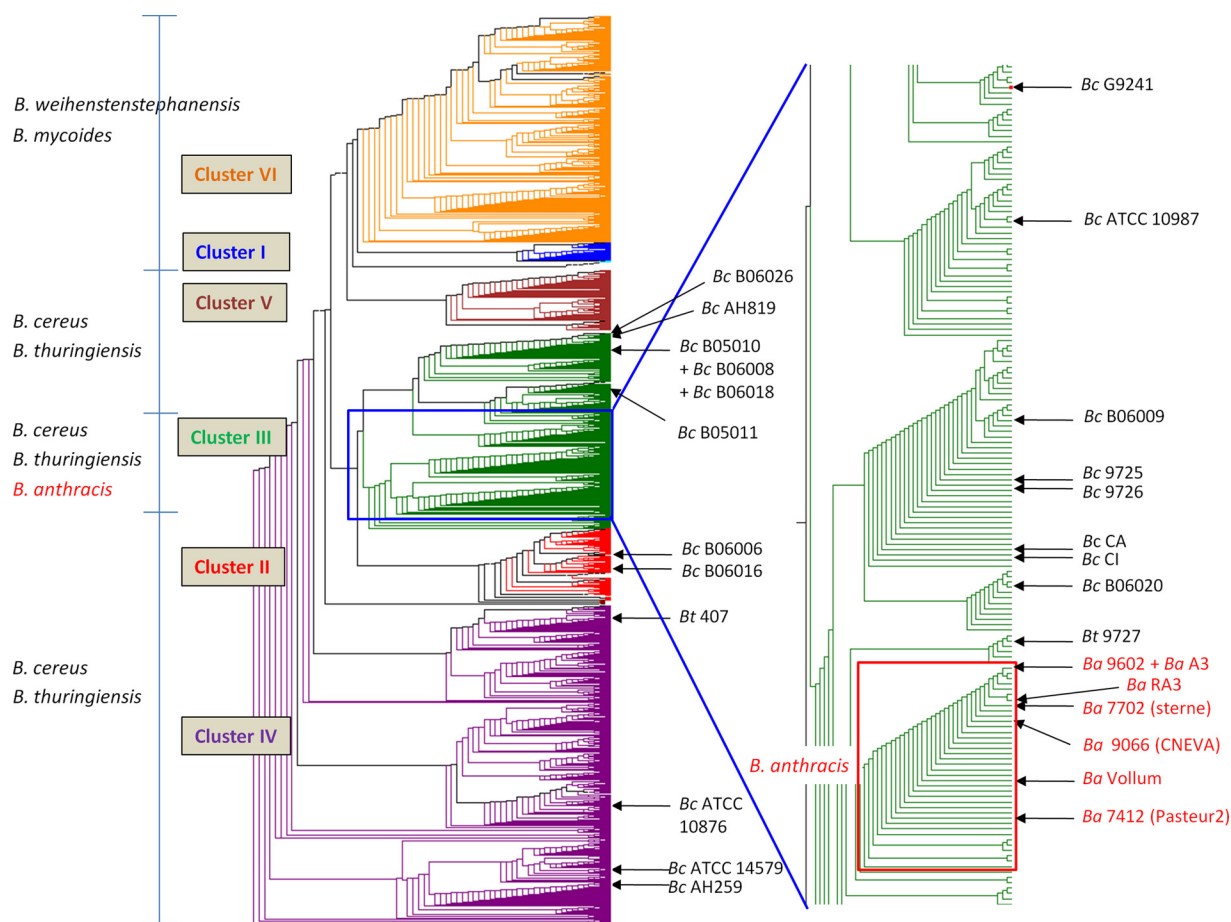


FIG. 1. Mapping of strain clusters constituting the *B. cereus* group. Twenty-eight strains included in this study are highlighted. The freely available HyperCAT algorithm, developed by Tourasse *et al.* (56, 57) (<http://mlstoslo.uio.no>), was used for this classification. The database contains multilocus sequence typing (MLST), multilocus enzyme electrophoresis (MLEE), and amplified fragment length polymorphism (AFLP) phylogenetic data for the *Bacillus cereus* group of bacteria.

60% acetonitrile in only 40 min and a data-dependent “top 5” MS method (Fig. 2A). The first step of the analytical process was to perform a Mascot search against a database including

solely protein sequences from *B. anthracis*. On average for the 5 *B. anthracis* strains, this database search led to the identification of 1252 tryptic and 580 Glu-C peptides per strain with

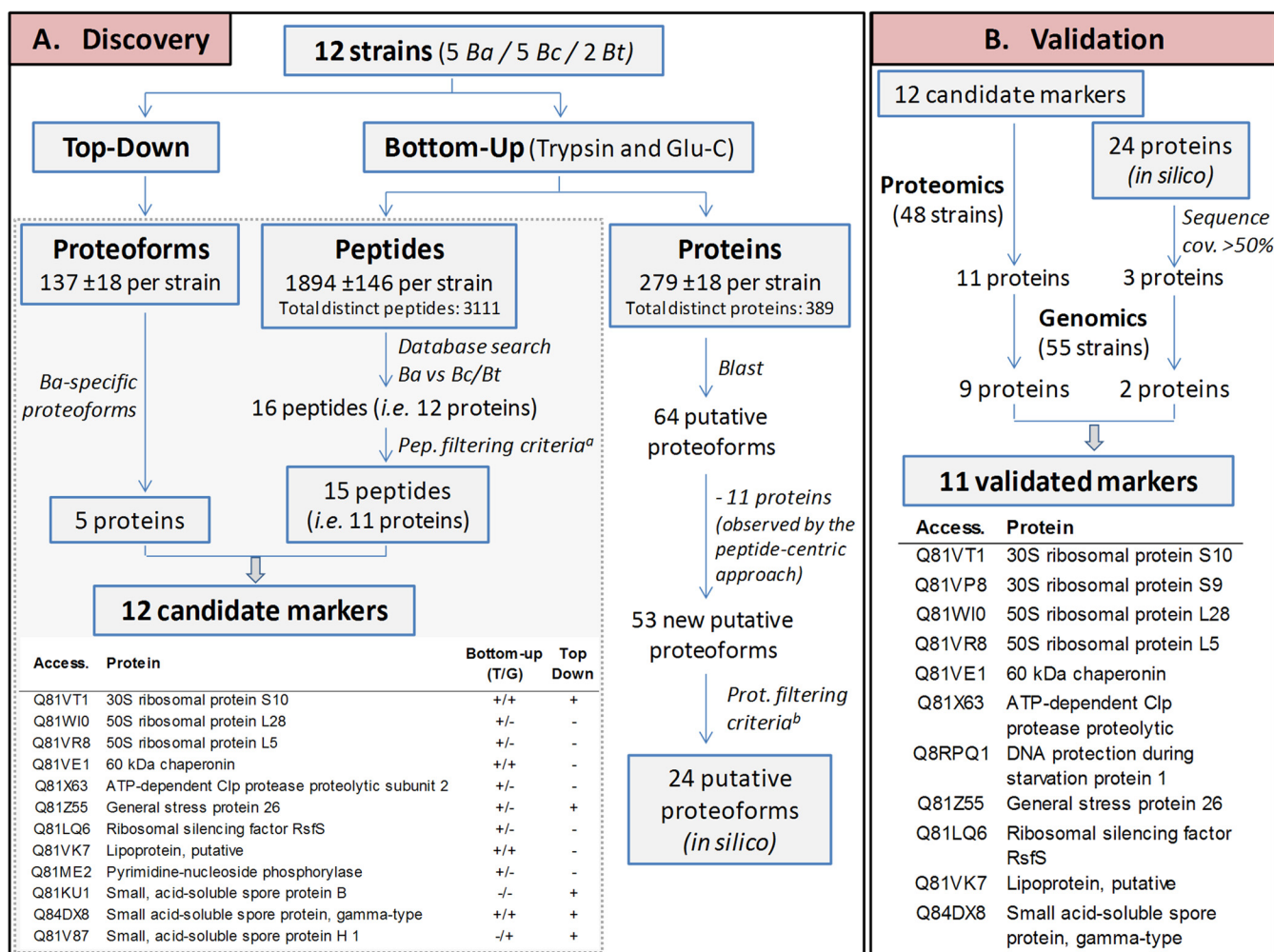


FIG. 2. General workflow used for the identification of reliable *B. anthracis* spore markers. (A) Marker discovery phase. <sup>a</sup>Peptide filtering criteria: all *B. anthracis* strains (5/5) and no *B. cereus*/*B. thuringiensis* (0/7) must have the specific peptide, and all *B. cereus*/*B. thuringiensis* (7/7) must have an ortholog peptide. <sup>b</sup>Prot. filtering criteria: a common proteoform for all *B. anthracis* bearing at least one amino acid mutation specific to *B. anthracis* must be observed. (B) Marker validation phase.

an average mass accuracy better than 2 ppm for the parent ions. These peptide numbers further correspond to 228 and 122 proteins, respectively, which results in 279 distinct proteins (Fig. 2A). The number of peptide identifications remained rather constant for each strain varying only between  $\pm 5$  and 10%. Nevertheless, only a  $\sim 50\%$  overlap in the peptides identified was observed between two given *B. anthracis* strains, as a consequence of biological variability or more probably of the well-known variability of peptide identifications among LC-MS/MS replicates (51). For instance, Tabb *et al.* observed similar  $\sim 45\%$  overlaps in the peptides identified between pairs of replicates from both standard protein mixtures of rather low complexity as well as yeast extracts independently of the mass spectrometer used (different LTQ and LTQ-Orbitrap instruments were used in this interlaboratory study) (51). This means that even rather simple mixtures yield more peptide ions than can be sequenced during an LC-MS/MS run. To encompass this limitation, peptides identified

from the 5 different *B. anthracis* strains were merged into a single list of 3111 peptides (2053 tryptic and 1058 Glu-C peptides, overall corresponding to 389 distinct spore proteins) that was then used for the rest of the marker discovery process (Fig. 2A and supplementary Tables S3 and S4).

Peptides specifically belonging to *B. anthracis* must not be found in any *B. cereus* or *B. thuringiensis* strains. Therefore the compiled peptide list was further refined by excluding the peptides also returned by a Mascot search against a *B. cereus/thuringiensis* database. Only 16 peptides (11 tryptic and 5 Glu-C peptides) were consistently observed across all the five *B. anthracis* strains while not being detected in the digests of the seven *B. cereus/thuringiensis* strains, as confirmed by the thorough manual interrogation of the corresponding LC-MS data (Fig. 2A).

As expected, these 16 peptides do not derive from *B. anthracis*-specific proteins but rather from proteins that have orthologs in *B. cereus/thuringiensis* species, with specific se-

## Identification of Novel Markers of *B. anthracis* Spores

TABLE II

*B. anthracis*-specific proteotypic peptides and the corresponding sequences identified for *B. cereus*/*B. thuringiensis*. Ba: *B. anthracis*; Bc: *B. cereus*; Bt: *B. thuringiensis*

Protein Markers	Accession (Ba isoform) <sup>a</sup>	Proteotypic peptide sequence <sup>b</sup>	Number of Strains <sup>c</sup>					
			Discovery (n=12)			Validation (n=36)		
			Ba (n=5)	Bc (n=5)	Bt (n=2)	Ba (n=10)	Bc (n=19)	Bt (n=7)
30S ribosomal protein S10 <sup>d</sup>	Q81VT1	R.ILDQSA <u>D</u> K.I	5/5	-	-	10/10	-	-
	NA	R.ILDQSA <u>E</u> K.I	-	5/5	2/2	-	16/19	7/7
50S ribosomal protein L5	Q81VR8	K. <u>I</u> LDNAVEELTQITGQK.P	5/5	-	-	10/10	-	-
	NA	K. <u>A</u> LDNAVEELTQITGQK.P	-	3/5	2/2	-	11/19	3/7
	NA	K. <u>A</u> LDNAVEELTQ <u>I</u> AQK.P	-	2/5	-	-	8/19	4/7
50S ribosomal protein L28	Q81WI0	R.VC <u>I</u> ITGR.K	5/5	-	-	10/10	-	-
	NA	R.VC <u>A</u> ITGR.K	-	5/5	2/2	-	19/19	7/7
60 kDa chaperonin <sup>d</sup>	Q81VE1	K. <u>L</u> VGGVAVIK.V	5/5	-	-	10/10	-	-
	NA	K. <u>L</u> <u>A</u> GGVAVIK.V	-	5/5	2/2	-	19/19	7/7
ATP-dependent Clp protease proteolytic subunit 2	Q81X63	R.YALPNS <u>E</u> AMHQPLGGAQQGATEIEIAAK.R	5/5	-	-	8/10	-	-
	NA	R.YALPNS <u>V</u> MIHQPLGGAQQGATEIEIAAK.R	-	5/5	2/2	-	19/19	7/7
General stress protein 26	Q81Z55	R.WLL <u>R</u> PEDPNYLIK.I	5/5	-	-	10/10	-	-
	NA	R.WLL <u>G</u> PEDPNYLIK.I	-	5/5	2/2	-	19/19	7/7
Lipoprotein, putative <sup>d</sup>	Q81VK7	K. <u>G</u> TPDQEVIDQVLNVFK.L	5/5	-	-	10/10	-	-
	NA	K. <u>D</u> TPDQEVIDQVLNVFK.L	-	-	-	-	5/19	1/7
	NA	K. <u>D</u> TPDQEVIDQ <u>V</u> NVFK.L	-	-	1/2	-	-	4/7
	NA	K. <u>D</u> T <u>A</u> DQEVIDQ <u>V</u> NVFK.L	-	4/5	-	-	9/19	-
	NA	K. <u>N</u> T <u>A</u> DQEVIDQ <u>V</u> NVFK.L	-	1/5	1/2	-	5/19	2/7
Pyrimidine-nucleoside phosphorylase	Q81ME2	K.HSTGGVGD <u>I</u> ITTLVLGPLVAALGVPVAK.M	5/5	-	-	9/10	-	-
	NA	K.HSTGGVGD <u>I</u> ITTLVLGPLVAALGVPVAK.M	-	2/5	-	-	15/19	4/7
	NA	K.HSTGGVGD <u>I</u> ITTLVLGPLVAAL <u>D</u> VPVAK.M	-	3/5	2/2	-	4/19	3/7
Ribosomal silencing factor RsfS	Q81LQ6	R.WVLVDLGDVVAHV <u>F</u> YK.D	5/5	-	-	8/10	-	-
	NA	R.WVLVDLGDVVAHV <u>H</u> K.D	-	5/5	2/2	-	11/19	3/7
Small, acid-soluble spore protein H 1	Q81V87	E.SVHV <u>H</u> VNALEEK.-	5/5	-	-	10/10	-	-
	NA	E.SVHV <u>N</u> VNALEEK.-	-	-	-	-	13/19	1/7
	NA	E.SVHV <u>D</u> VNALEEK.-	-	5/5	2/2	-	6/19	6/7
Small acid-soluble protein gamma type <sup>d</sup>	Q84DX8	E.AKKAQASGAS <u>IQ</u> STNASYGTE.F	5/5	-	-	10/10	-	-
	NA	E.AKKAQASGAS <u>Q</u> SANASYGTE.F	-	5/5	2/2	-	19/19	7/7

<sup>a</sup> UniProt accession number of *B. anthracis*-specific isoform (NA: not applicable).

<sup>b</sup> Proteotypic peptides specific of *B. anthracis* (top sequence), and the corresponding sequences identified for *B. cereus*/*B. thuringiensis* (other sequences). The specific amino acids are underlined.

<sup>c</sup> Total number of strains for each peptide form for the discovery (12 strains) and the validation (36 strains) steps.

<sup>d</sup> Proteins with the corresponding tryptic/Glu-C peptides also identified.

quence variations. The specificity of 15 among 16 peptides for *B. anthracis* was further ensured by the identification of the corresponding peptides in *B. cereus/thuringiensis* spore extracts harboring particular amino acid substitutions (Table II). As an example, Fig. 3 shows the MS/MS spectra obtained of the Glu-C peptides from SASP-gamma proteoforms characteristic of *B. anthracis* and *B. cereus* species, with a few specific fragment ions. An additional level of specificity was also reached by performing Blast similarity searches against the whole Uniprot database. Thus, this step led to the identification of 15 peptides belonging to 11 distinct proteins, which could be regarded as potential candidate markers: small acid-soluble spore proteins gamma (SASP- $\gamma$ ) and H1

(SASP-H1), 30S ribosomal protein S10 (RS10), 50S ribosomal proteins L5 and L28 (RL5 and RL28, respectively), a putative lipoprotein, 60 kDa chaperonin, general stress protein 26 (GSP26), ribosomal silencing factor RsfS, pyrimidine-nucleoside phosphorylase, and ATP-dependent Clp protease proteolytic subunit 2 (Table II). Table III summarizes the identification data obtained for these proteins.

*Identification of B. anthracis Marker Candidates by a Top-down Approach*—In parallel to experiments performed on peptides, intact protein profiling by LC-MS was evaluated as a complementary approach. Such an approach has already proven valuable for classification of bacteria (52, 53). Under these conditions, we were able to pinpoint ~140 distinct



## SASP-gamma

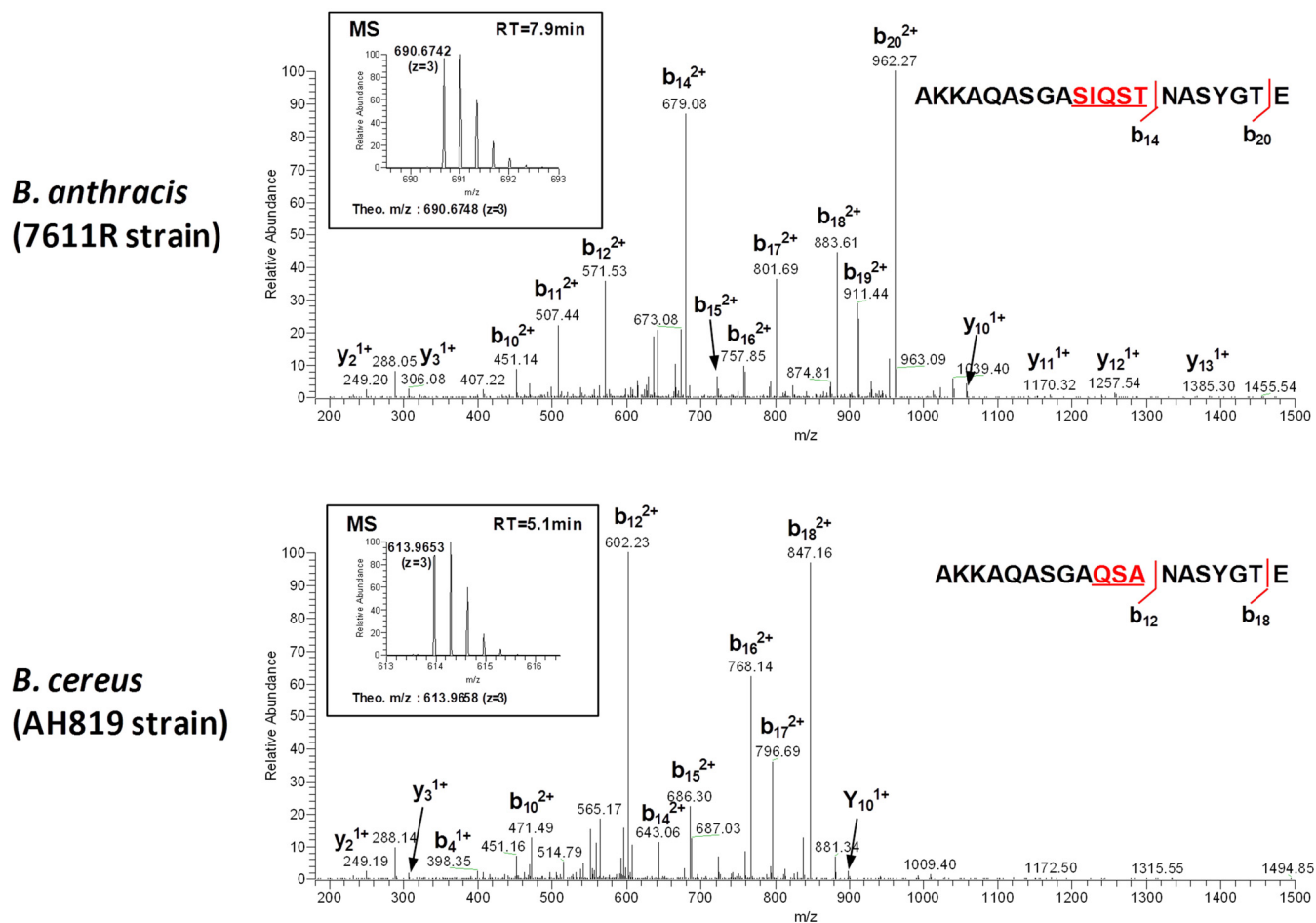


FIG. 3. MS and MS/MS spectra obtained for the proteotypic peptide of SASP-gamma for a given *B. anthracis* (top) and *B. cereus* (bottom) strain. MS spectra of the corresponding precursor ions are shown along with their retention times as insets.

“proteoforms” (54) for each strain (Fig. 2A), and overall detected masses were below 20 kDa. Manual comparison of the protein fingerprints obtained for the five *B. anthracis* and seven *B. cereus/thuringiensis* strains highlighted five proteins as being conserved across all the *B. anthracis* strains: RS10, GSP26, SASP-H1, SASP- $\gamma$ , and SASP-B (Table III). These proteins were first putatively identified by matching experimentally determined masses with theoretical ones, thanks to the high mass accuracy (<10 ppm) obtained under these analytical conditions (55), and further confirmed (when possible) by performing top-down MS/MS experiments. All these proteins, except SASP-B, overlapped the proteins already identified using the peptide-centric data. Three out of these five proteins, the SASP proteins, were successfully confirmed by top-down MS/MS. Among them, we identified SASP-B, a previously characterized marker of *B. anthracis* spores (30–34), which substantially validates the approach (supplementary Fig. S1). The *B. anthracis*-specific version of SASP-B is characterized by the simultaneous presence of two particular

amino acids located at both the N- and C termini (35). Other *B. cereus* group bacteria have only one or none of them, thus explaining why such protein was not highlighted by the analysis performed at the peptide level. Intact protein analysis also gives complementary information on the protein sequence integrity and potential modifications. RS10 and SASP-H1 were observed as predicted from the UniProtKB/TrEMBL database with an N-terminal methionine residue, whereas the GSP26, SASP- $\gamma$  and SASP-B proteins were detected without initial methionine (mass shift of –131.04 Da).

Altogether, the results obtained at both peptides and proteins in a preliminary set of five *B. anthracis* and seven *B. cereus/thuringiensis* strains led to the discovery of 12 candidate protein markers (Fig. 2A).

*Identification of Other Putative B. anthracis Markers by Bioinformatic Analysis of the Protein List Generated by the Bottom-up Approach*—Although very efficient, bottom-up analysis is a peptide-centric marker discovery pipeline that may suffer from several drawbacks. For instance, one may

TABLE III  
Protein markers identified during the discovery step by bottom-up and top-down approaches. Example of the *B. anthracis* RA3R strain<sup>a</sup>

Accession	Protein	MW (kDa)	Trypsin			Glu-C			Total			Intact Protein			
			Score	Pep.	Cov. (%)	RMS90 (ppm)	Score	Pep.	Cov. (%)	RMS90 (ppm)	Pep.	% Cov.	MW theo. (Da)	MW exp. (Da)	Δ (ppm)
Q81VT1	30S ribosomal protein S10	11.7	336.6	5	58.8	1.01	63.9	1	13.7	0.94	6	58.8	11531.3185 <sup>b</sup>	11531.3095	0.78
Q81VP8	30S ribosomal protein S9	14.5	224.3	5	50.8	0.39	70.6	2	15.4	0.58	7	53.8	14471.8867	-	-
Q81WI0	50S ribosomal protein L28 <sup>c</sup>	6.9	19.1	2	21.0	1.17	-	-	-	-	2	21.0	6948.8631	-	-
Q81VR8	50S ribosomal protein L5	20.2	483.1	8	57.5	0.73	42.0	2	11.7	0.28	10	61.5	20182.7958	-	-
Q81VE1	60 kDa chaperonin	57.4	1564.6	25	57.2	2.21	873.9	23	53.7	1.14	48	75.9	57396.0464	-	-
Q81X63	ATP-dependent Clp protease proteolytic subunit 2	21.4	190.2	4	27.5	1.22	72.9	2	11.4	1.48	6	31.1	21378.0447	-	-
Q8RPQ1	DNA protection during starvation protein 1	16.6	345.5	5	51.7	1.39	147.0	3	21.8	0.75	8	51.7	16638.2282	-	-
Q81Z55	General stress protein 26	16.1	69.5	2	21.7	1.57	-	-	-	-	2	21.7	16058.2816	16058.0487	14.50
Q81VK7	Lipoprotein, putative	17.6	573.9	11	60.6	0.68	418.6	9	51.0	0.92	20	71.0	17601.0335	-	-
Q81ME2	Pyrimidine-nucleoside phosphorylase	46.2	131.8	2	11.8	0.92	39.9	1	6.9	1.67	3	18.7	46195.6809	-	-
Q81LQ6	Ribosomal silencing factor RsfS	13.6	82.8	2	26.3	2.81	-	-	-	-	2	26.3	13561.8064	-	-
Q84DX8	Small acid-soluble protein gamma type	9.9	344.4	5	71.6	0.79	380.2	5	90.5	1.10	10	98.9	9731.6745 <sup>b</sup>	9731.7172	4.39
Q81KU1	Small, acid-soluble spore protein B	6.8	926.3	11	89.2	1.44	627.3	9	92.3	2.03	20	98.5	6674.4515 <sup>b</sup>	6674.4606	1.36
Q81V87	Small, acid-soluble spore protein H 1	6.5	-	-	-	-	90.7	2	32.2	1.41	2	32.2	6541.2170	6541.2168	0.03

<sup>a</sup> For each protein, UniProt accession number (Accession) and molecular weight (MW; Da) are represented. Bottom-up experiments are performed using Trypsin and Glu-C enzymes. Mascot protein score (score), number of peptides matched (Pep.), protein sequence coverage (cov.; in percent) and root mean square (RMS90; in ppm) are shown for each experiment. Total number of peptides and sequence coverage are noted. For intact protein analysis, theoretical and experimental protein molecular weight (MW theo. and MW exp. respectively; Da) and mass accuracy (Δ; ppm) are noted.

<sup>b</sup> Proteins known to have a loss of N-terminal methionine (-131.04 Da), which was taken into account for calculating their theoretical masses.

<sup>c</sup> For 50S ribosomal protein L28, two peptides were identified based on MS data but only one was sequenced by MS/MS.

imagine that even the use of two enzymes with different specificity (namely trypsin and Glu-C) would not exhaustively highlight the specific peptides belonging to proteoforms efficiently discriminating *B. anthracis* from the other members of the *B. cereus* group. In addition, *B. anthracis*-specific peptides may not be observed under the present LC-MS/MS conditions, as a consequence of their sequence or physico-chemical properties (e.g. ionization efficiency, chromatographic behavior). Therefore, as a complementary approach, we also aligned the 389 *B. anthracis* spore protein sequences identified during the bottom-up proteomics approach against the Uniprot database using the BlastP tool to further pinpoint other potential signature proteins. In a first round, 64 candidates were highlighted, including of course the 11 proteins already identified by the peptide-centric approach (Fig. 2A, supplementary Table S5). Such sequence-based analysis also has intrinsic limitations. Indeed, it may not predict accurately the specificity of a given protein because of the potential sequence mistakes existing in the databases, some amino acid sequences being only deduced from DNA. Therefore, additional filters were applied and implied that newly identified proteins should incorporate unique and specific amino acid substitutions and have a unique proteoform clearly identified for all the *B. anthracis* included in the databases. These criteria refined the list to 24 additional putative proteoforms (supplementary Table S5).

**Candidate Marker Validation**—Markers were validated in a two-step procedure using first a proteomics and then a genomics approach. Taking into account the 12 initial strains, 55 strains (18 *B. anthracis*, 28 *B. cereus*, and nine *B. thuringiensis*) were targeted for the whole validation process. However, sporulation could not be achieved for seven out of these 55 strains which were therefore used only for genomics-based validation.

**Validation of the Candidate Markers Exhibiting Specific Peptide/Protein Isoforms by LC-MS(/MS)**—For this first validation step, 36 additional strains, composed of 10 *B. anthracis*, 19 *B. cereus* and seven *B. thuringiensis* strains, were analyzed by both bottom-up and top-down approaches, thus bringing the total of spore extracts analyzed to 48 (Table I, Fig. 2B). Optimal strain selection was achieved thanks to the HyperCAT database (<http://mlstoslo.uio.no>) (56, 57) (Fig. 1). The final strain list notably included the phylogenetically very close *B. thuringiensis* 9727 and *B. cereus* B06020 as well as recently described *B. cereus* var. *anthracis* carrying the plasmids pBCXO1 and/or pBCXO2 close to pXO1 and pXO2 of *B. anthracis* (i.e. *B. cereus* G9241, *B. cereus* biovar *anthracis* CI and CA) (16–20).

Eleven of the 12 candidate markers were confirmed as being *B. anthracis*-specific (Fig. 2B and Table II). As expected, the SASP-B proteoform of *B. anthracis* was also detected in CA, CARP, CI and B06020 strains, and so was eliminated at this step (supplementary Fig. S2). For each of the 11 protein markers, a constant and unique proteotypic peptide se-

quence was observed in all *B. anthracis* strains, except for the ribosomal silencing factor RsfS, the ATP-dependent Clp protease proteolytic subunit 2 and the pyrimidine-nucleoside phosphorylase, for which the unique peptide was not detected in 2/10, 2/10, and 1/10 supplementary *B. anthracis* strains, respectively (Table II). This was probably because of their poor detection sensitivity (as a direct consequence of the low expression levels of the corresponding proteins). It is of note that the corresponding peptides incorporating at least one amino acid modification were also detected in almost all *B. cereus/thuringiensis* strains (Table II).

**Marker Validation by Gene Sequencing**—To further confirm the data obtained at peptide- and protein-levels, as well as the uniqueness of the identified proteoform to *B. anthracis*, PCR amplification followed by DNA sequencing of the 11 candidate markers was performed on 55 strains (Table I). Analyses were successfully performed for every marker, except for SASP-H1 and pyrimidine-nucleoside phosphorylase for which no amplification was achieved. This may be linked to the existence of a higher variability than expected within the flanking regions of these particular genes. To be considered as reliable marker candidates, highlighted proteins had to fulfill both proteomics and genomics validation steps for every strain. For this reason, despite being confirmed by the proteomics approach, SASP-H1 and pyrimidine-nucleoside phosphorylase were excluded from the final list of markers, but this does not completely preclude their potential as markers. One interesting finding is that translation of the experimentally determined nucleotide sequences into amino acids fully confirmed the peptide sequences identified by proteomics for characteristic peptides originating from *B. anthracis*, *B. cereus* and *B. thuringiensis* (supplementary Table S6).

In addition, among the 24 putative proteins highlighted by a sequence similarity search as mentioned above, three supplementary markers were also selected for DNA sequencing: DNA protection during starvation protein 1 (DPS1), 30S ribosomal protein S9 (RS9) and phosphoglycerate kinase (supplementary Table S5, Table III). We reasoned that the protein candidates having the highest chance of success are the most abundant ones, and thus exhibiting the highest sequence coverage (>50%) (Table S-3). Interestingly, the relevance of 2 of these proteins (DPS1 and RS9) was demonstrated by gene sequencing, thus showing the complementary nature of the investigated approaches.

For the sake of clarity, sequencing data are represented as dendrograms as shown in Fig. 4 for SASP-gamma (see supplementary Fig. S3 for the other markers). For each marker, *B. anthracis* strains are grouped within a specific cluster whereas *B. cereus* and *B. thuringiensis* showed a high level of isoform diversity. For example, the 18 *B. anthracis* strains analyzed have exactly the same SASP-gamma nucleotide sequence whereas the 37 *B. cereus/B. thuringiensis* strains are divided into 18 different clusters (Fig. 4). Sequence diversity observed for *B. cereus* and *B. thuringiensis* strains

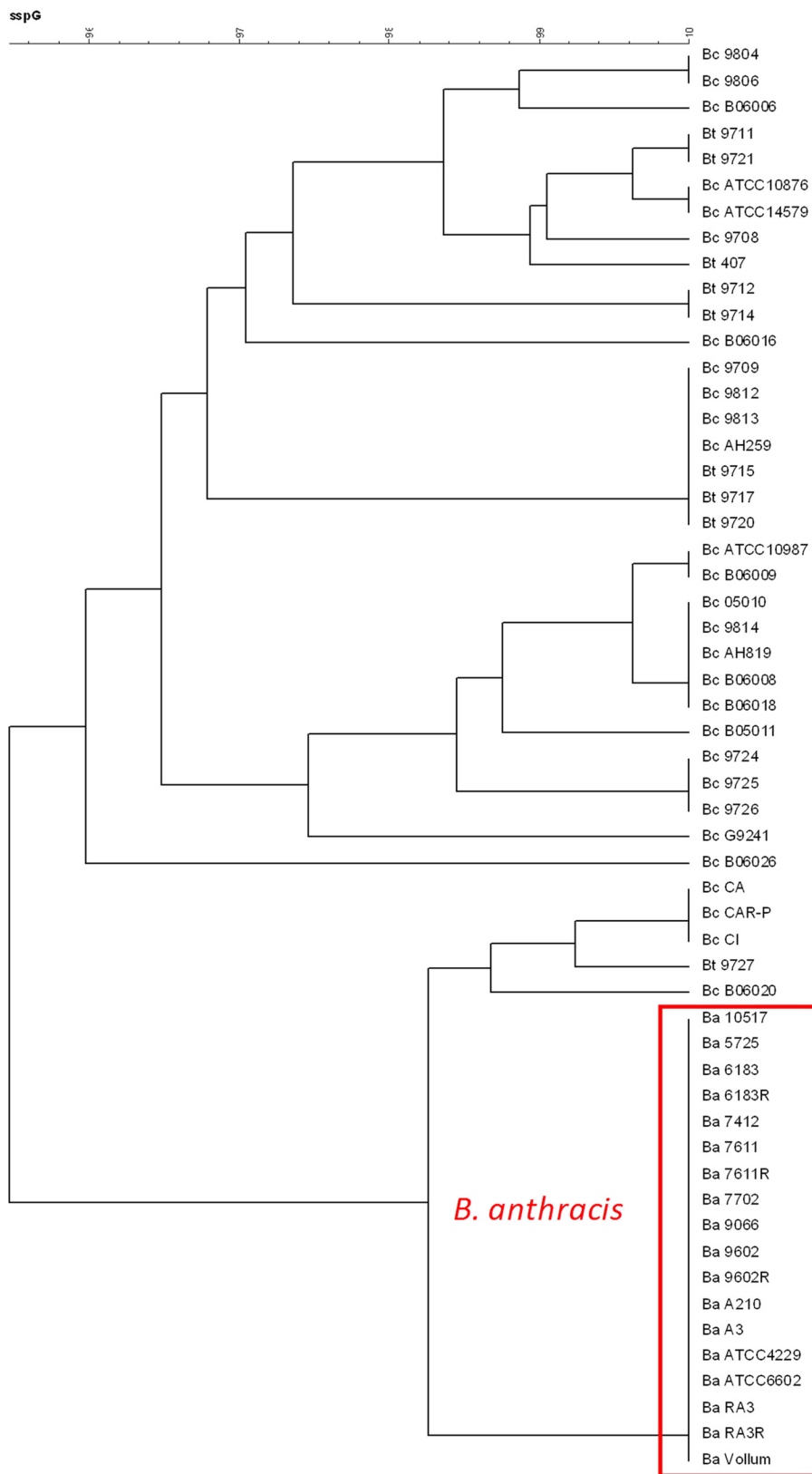


FIG. 4. Dendrogram based on gene sequencing data obtained for SASP-gamma, illustrating the high degree of similarity among *B. anthracis* strains. Dendrograms were obtained with BioNumerics software version 6.5 (Applied-Maths, Sint-Martens-Latem, Belgium).

was reduced when translated into amino acids, as several nucleotide changes coded for silent mutations. For example, the 18 different SASP-gamma gene sequences coded for seven distinct proteoforms, as can also be deduced from the proteomics data (Fig. 4, [supplementary Table S6](#)). The majority of identified sequences matched with proteoforms previously described in databases; however a few new ones were highlighted (noted as “new” in [supplementary Table S6](#)).

Altogether these data demonstrate a perfect match between proteomics and genomics data for the 11 markers.

#### *Development of a Targeted Multiplex SRM Assay for Further Validation and Unambiguous Detection of B. anthracis Spores in Complex Samples*

*Development of the LC-SRM Assay*—Among the 11 characterized markers, only nine exhibit specific trypsin or Glu-C peptides that could be further used for the development of a targeted SRM assay (Table II). Moreover, peptide selection should be refined by considering peptides the most suitable for robust and sensitive LC-SRM detection, that is, containing eight to 25 amino acids, showing good chromatographic behavior and ionization efficiency, while not containing any methionine, cysteine, or tryptophan residues that could undergo oxidation during sample handling. Peptides with two neighboring basic amino acids or RP/KP sequences were also avoided (58). Such rather stringent criteria (especially when dealing with already filtered species-specific peptides) produced a short list of four peptides corresponding to the four following proteins: SASP-gamma, 30S ribosomal protein S10, putative lipoprotein, and 60 kDa chaperonin proteins. The strategy was to monitor in a single analysis the proteotypic peptides corresponding to the *B. anthracis* unique proteoforms. We have recently described an LC-SRM approach targeting SASP-B for the specific detection of *B. anthracis* spores in complex environmental samples, following immunocapture of spores (35). *B. anthracis* spores have a common SASP-B proteoform, which could be targeted by monitoring a combination of two particular proteotypic peptides after a *B. anthracis*-specific immunocapture step (35), whereas *B. cereus* and *B. thuringiensis* share at least one of these different peptides (with some exceptions as above mentioned). Hence, we extended the scope of this first LC-SRM approach by including the four new markers described above. Consequently, the final method involved the monitoring of five distinct proteins for the specific detection of *B. anthracis* spores.

Following optimization of MS and MS/MS conditions on labeled and unlabeled synthetic peptides, two SRM transitions were monitored per peptide for improved specificity ([supplementary Table S1](#)). The final LC-MS/MS method was performed within 30 min and implies the monitoring of 28 different SRM transitions, corresponding to the five different markers ([supplementary Fig. S4](#)). Spore extracts from the 48 strains previously used for the marker discovery phase (Table

I) were screened with this multiplex LC-SRM assay. Typical LC-MS/MS chromatograms obtained for spores from *B. anthracis* (i.e. *B. anthracis* 9602) and a closely related *B. cereus* strain (i.e. *Bcbva* CI) are shown in Fig. 5. As expected, positive signals corresponding to the 4 newly identified markers were observed for *B. anthracis* strains (Fig. 5A). In addition, no signal corresponding to these proteins was detected in any other strains, even those closely related to *B. anthracis* such as *Bcbva* CI (Fig. 5B). All these results confirm the high selectivity of the identified markers. The situation was quite different for SASP-B, because we experimentally confirmed that the closely related *B. cereus* strains (CI, CA and B06020) have the same SASP-B proteoform as *B. anthracis* (Fig. 5B). It is of note that these particular strains were not initially included in our first work focused on SASP-B (35). These data demonstrate that careful strain selection is mandatory for achieving reliable marker identification.

*Application of the Multiplex LC-SRM Assay to the Detection of B. anthracis Spores in Complex Spore Mixtures*—*B. cereus* and *B. thuringiensis* are classically retrieved from environmental samples and the specific identification of some *B. anthracis* spores in complex mixtures containing high amounts of other *Bacillus* species is rendered difficult because of their high phenotypic and genetic similarities. Mixtures composed of various concentrations of 5 different *B. cereus* strains were designed to mimic real-life samples and further confirm the ability of the LC-SRM method to selectively identify *B. anthracis* spores in such complex matrices. As expected, *B. anthracis* spores cannot be discriminated in such complex mixtures by solely monitoring SASP-B ([supplementary Fig. S5A](#)). In contrast, the four new markers clearly discriminated *B. anthracis* spores even when they were 100-fold less abundant than the *B. cereus* ones ([supplementary Fig. S5B](#)). Moreover, SRM detection proved to be linear over a 2-log concentration range, thus demonstrating that matrix components do not significantly impact the method's robustness, as illustrated for the SASP-gamma signature peptide ([supplementary Fig. S6](#)).

#### DISCUSSION

The discovery of protein markers specific to *B. anthracis* is hampered by the high genetic similarity existing within the *B. cereus* group (5). Such genetic relatedness is so high that some authors have raised the possibility that the different members of the *B. cereus* group could originate from one and the same species with varying plasmid contents (the virulence plasmids pXO1 and pXO2 encoding for the toxins and the capsule in *B. anthracis* strains, respectively, and the plasmid encoding for insecticidal toxins in *B. thuringiensis* strains) (59). These similarities further translated at the protein level, with the almost impossible distinction of *B. anthracis* from closely related *B. cereus* and *B. thuringiensis* by MALDI-TOF MS protein profiling (30). Therefore, the identification of specific and reliable markers of *B. anthracis* spores implies that a consistent panel of closely related strains of *B. cereus* and *B.*

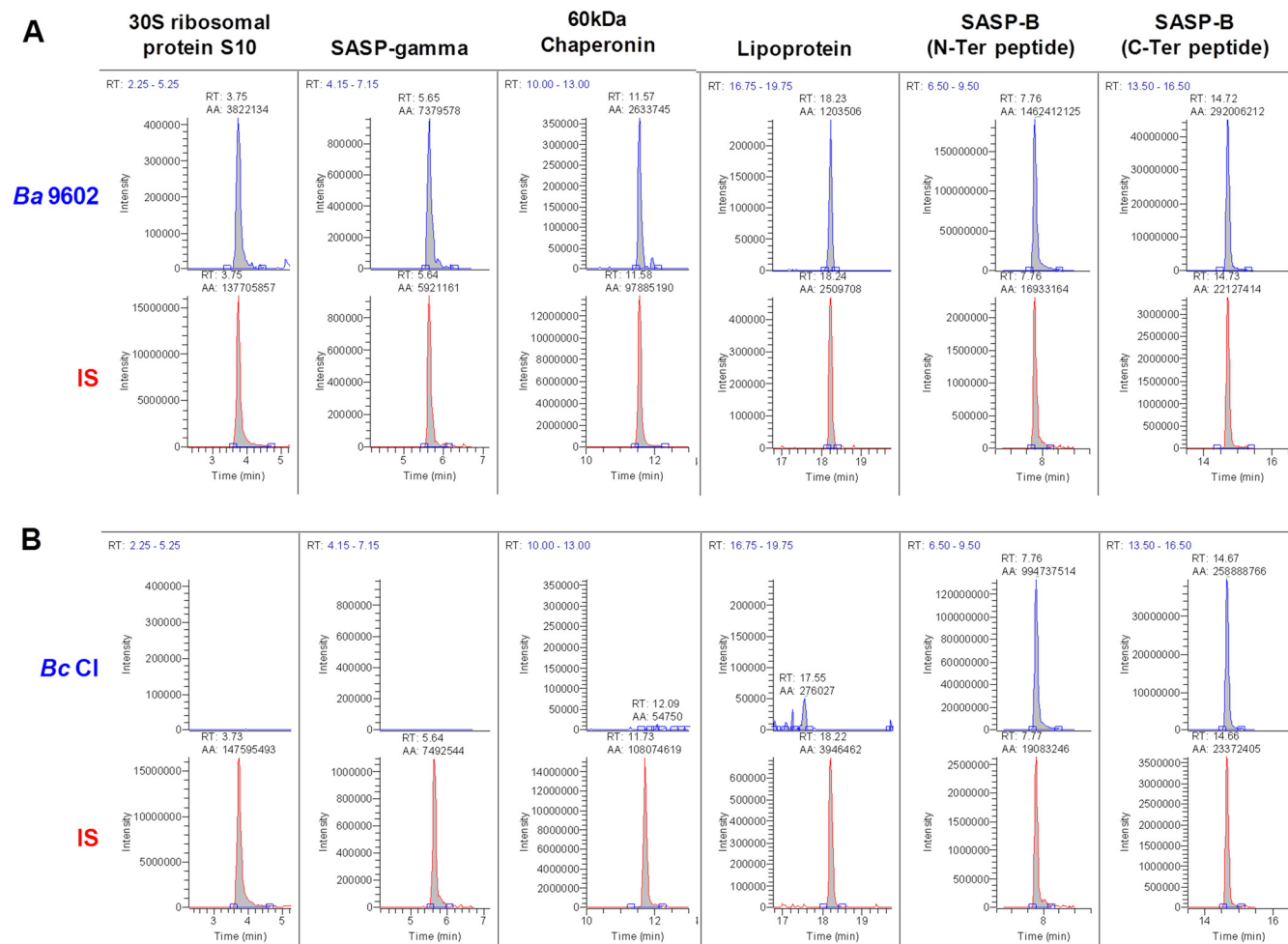


FIG. 5. Typical LC-SRM chromatograms obtained for (A) *B. anthracis* 9602, and (B) *B. cereus* biovar *anthracis* CI strains. Depicted SRM transitions are as follows: 30S ribosomal protein S10 (transition: 445.3 → 663.2), SASP-gamma (690.9 → 962.3), 60 kDa chaperonin (428.3 → 742.4), triosephosphate isomerase (734.7 → 498.3), lipoprotein (601.5 → 724.4), N-terminal (438.7 → 583.9), and C-terminal peptides (759.7 → 536.2) of SASP-B protein. In each case, the signal of the corresponding isotopically labeled peptide is also shown and noted IS (internal standard). These data were obtained using  $10^7$  spores.

*thuringiensis* has to be studied, to accurately reflect the naturally occurring relatedness.

With this aim, we took special care to constitute a panel of 55 relevant strains, to ensure reliable identification of specific markers for the robust detection of *B. anthracis*. Two key features were considered (i) the phylogenetic proximity to *B. anthracis* strains (5), and (ii) the presence of the virulence plasmids (16–18). The HyperCAT database is the most complete phylogenetic dataset for the *B. cereus* group with a total of 2297 bacterial isolates (as of March 17, 2012), confidently and thoroughly characterized by various gene-based approaches: MultiLocus Sequence Typing (MLST), MultiLocus Enzyme Electrophoresis (MLEE), and Amplified Fragment Length Polymorphism (AFLP) (56, 57). Based on these analyses, all the *B. anthracis* strains fall in the same cluster (cluster III according to the HyperCAT database), along with a few closely related *B. cereus* and *B. thuringiensis* strains (Fig. 1) (5). Therefore, our selection primarily targeted this cluster

(seven *B. anthracis*, 14 *B. cereus*, and one *B. thuringiensis*). It is of particular interest that this first set of strains included the *B. cereus* strain G9241, associated with severe pneumonia and carrying an almost complete pXO1-like plasmid (99.6% sequence similarity in shared regions) (16); and above all, the two recently described *B. cereus* biovar *anthracis* strains (CI and CA) recovered from great apes killed by an anthrax-like disease in tropical forests of Ivory Coast and Cameroon, and carrying two plasmids, pBCXO1 and pBCXO2 close to pXO1 and pXO2 of *B. anthracis* (17, 18). These two plasmids carry the anthrax toxin and the poly- $\gamma$ -D-glutamate capsule biosynthesis genes, respectively (17, 18), which underlines the *B. anthracis*-like nature of these strains and the associated risk of false-positive detection if not taken into account. To the best of our knowledge, no similar other in-depth spore proteomics study including these two particular strains has ever been undertaken. Six other *B. cereus* strains belonging to the more distant clusters II and IV were also considered. Last but

not least, 27 other strains from the Institut Pasteur collection (11 *B. anthracis* and 16 *B. cereus/B. thuringiensis*), for which no phylogenetic classification was available, were included. Another point of interest is that the different *B. anthracis* strains included in this study take into consideration the possible loss of either the pXO2 plasmid and/or less commonly the pXO1 plasmid (1). Indeed, among the 18 *B. anthracis* strains, nine harbor pXO1 and pXO2, six only pXO1, two only pXO2, and one neither pXO1 nor pXO2 plasmids (Table I). Overall, the total number of strains studied was 55 (Table I).

Taking into account the high homogeneity within the *B. cereus* group, and based on our first experimental data, it rapidly became obvious that we would not be able to highlight protein expressed exclusively in *B. anthracis* spores, but rather proteins that differ by a few (or even one) amino acid residues from closely related species. TFA-extracted spore proteins were submitted to trypsin and Glu-C digestion, and then analyzed by a single-step LC-MS/MS approach performed on a high resolution/high mass accuracy LTQ-Orbitrap instrument. Although not initially intended to reach comprehensive spore proteome coverage, our shotgun bottom-up proteomics approach identified ~400 distinct proteins (with  $\geq 2$  unique peptides) in *B. anthracis* spores, well exceeding the ~200–250 distinct proteins reported in previous studies using slightly different sample preparation protocols and 2D gel electrophoresis/mass spectrometry (41, 42, 60). The most exhaustive proteomics study described ~750 proteins but involved multidimensional liquid chromatography coupled to tandem mass spectrometry to go deeper into the *B. anthracis* spore proteome and identify less abundant proteins (44). Although less exhaustive than the latter study, our results perfectly meet the initial objective of studying the most abundant spore proteins to highlight easily detectable targets for the most sensitive detection of *B. anthracis* spores.

On a preliminary set of 12 strains (five *B. anthracis*, five *B. cereus*, and two *B. thuringiensis*), LC-MS/MS analyses highlighted 15 peptides (representing 11 distinct proteins) that were consistently observed across all *B. anthracis* while not being detected in *B. cereus/thuringiensis* strains (Fig. 2).

For diverse reasons linked to peptide detectability or enzyme specificity, such bottom-up conditions may not always pinpoint *B. anthracis*-specific peptides. To encompass this limitation, a complementary approach involving the profiling of protein species (with molecular weights up to ~20 kDa) was performed and identified only one additional protein, thus leading to a total of 12 distinct proteins potentially specific for *B. anthracis* (Fig. 2). Moreover, three additional protein candidates were retrieved from a bioinformatics-based screening of all the data gathered from proteomics (DPS1, RS9 and phosphoglycerate kinase).

Marker validation was performed by extending the number of strains to 55. This panel of strains was interrogated by a combination of bottom-up proteomics, intact protein profiling, bioinformatics-based and gene sequencing approaches, to

guarantee the comprehensiveness and robustness of the marker discovery pipeline. This strategy confirmed the relevance of 11 proteins, but excluded phosphoglycerate kinase, pyrimidine-nucleoside phosphorylase, SASP-H1 and SASP-B proteins. Inefficient DNA amplification prevented SASP-H1, pyrimidine-nucleoside phosphorylase and phosphoglycerate kinase from being fully confirmed, but it does not completely preclude them from being regarded as potential *B. anthracis* markers. SASP-B is often used as a marker of *B. anthracis* spores (30–34), but it was clearly invalidated when analyzing 3 closely related *B. cereus* strains (*i.e.* CA, CI, and B06020). The latter clearly demonstrates that marker specificity must be constantly re-evaluated, especially by working on closely related strains, to minimize the false-positive rate.

Gene sequencing confirmed marker identification and allowed us to go deeper into protein sequence characterization. For the 55 strains experimentally analyzed, the uniqueness of the *B. anthracis* proteoforms was demonstrated and perfectly matched with proteomics results. Furthermore, translation of the determined nucleotide sequences complemented the amino acid ones, by shedding light on unobserved peptide sequences in the bottom-up analysis. This is of great interest for the two proteins only retrieved from the sequence-based analysis (*i.e.* DPS1 and RS9), for which no proteotypic peptide was experimentally observed. These results well illustrate the complementarity between gene sequencing and proteomics.

All the data resulting from proteomics and genomics analyses confirmed that *B. anthracis* is highly monomorphic (Tables II and [supplementary Table S6](#)). Conversely, a great diversity in sequences was observed for the *B. cereus* and *B. thuringiensis* strains. For instance, the 18 *B. anthracis* strains have exactly the same nucleic acid sequence for SASP-gamma whereas the 37 *B. cereus* and *B. thuringiensis* strains were divided into 18 different DNA clusters that further translate into seven distinct proteoforms (Fig. 4, [supplementary Table S6](#)). This was in concordance with previous DNA homology studies indicating that *B. anthracis* strains have a similarity greater than 90% and therefore are one of the most molecularly homogenous bacteria known (59, 61). The great homogeneity of *B. anthracis* in the natural environment may be the result of the organism spending the majority of its life as a dormant spore (62). In this particular state, *B. anthracis* is not exposed to DNA-altering events such as the presence of phages and constant replication of DNA (59, 61); this considerably limits the amount of genetic diversity found among isolates of this species (62) compared with *B. cereus*, which more commonly exists in the environment as vegetative cells (63). Regarding the markers highlighted in this study, ~5 distinct *B. cereus* proteoforms exist on average for a given protein versus a single *B. anthracis* proteoform ([supplementary Table S6](#)). This great inter-strain homogeneity for *B. anthracis* is an important point as it intrinsically minimizes the risk of false-negative results (because of the misdetection of a potentially mutated peptide by targeted approaches) and in-

creases confidence in marker robustness in the case of emerging strains.

Marker identification was performed on spores and therefore identified proteins expressed in this particular bacterial state. Among the 11 markers highlighted, we identified SASP-gamma which is encoded by the *sspE* gene and is linked to amino acid storage (64). This protein is part of a protein group (SASPs) present in large amounts in the core of *Bacillus* spores comprising ~15% of the protein content (60), and protects spores from a number of external factors such as chemical and enzymatic cleavage and UV light (65). SASPs are recognized as potential markers for bacterial identification (31). These proteins are produced exclusively at the bacterial sporulation stage and are degraded during spore germination (65). Their concentration is so high in spores that even traces of spores in huge amounts of vegetative cells would result in strong SASPs signals. Indeed, Lasch *et al.* have previously reported that ~50% of their preparations of vegetative cells from 374 *Bacillus* strains contained detectable amounts of SASPs (30). This suggests that SASPs would represent particularly pertinent protein markers when *B. anthracis* detection is envisioned (whatever the targeted form, *i.e.* spores or cells).

We also identified 4 ribosomal proteins as potential markers, which is not so surprising because up to 21% of the cell's overall protein content is ribosomal (66). This is also in good agreement with previous MALDI-TOF studies of intact microorganisms (30, 67). Among the other markers, we found proteins involved in metabolism (putative lipoprotein, ribosomal silencing factor RsfS, CH60), a protease (ATP-dependent Clp protease proteolytic subunit 2), and proteins involved in the response to stress (GSP26, DPS1). One could easily imagine that some of the newly characterized *B. anthracis* markers could also be detected after spore germination, such as ribosomal proteins (RL5, RL28, RS9, and RS10) or other proteins involved in metabolism (CH60 or RsfS), because germination is accompanied by the resumption of cell growth and metabolic activity. For example, Clp-protease and CH60 were detected in similar amounts in dormant and germinating spores by Huang and colleagues (43) whereas CH60 has been previously identified in the proteome of *B. anthracis* vegetative cells (68, 69). In addition, ribosomal proteins, more generally speaking, also represent abundant proteins in vegetative cells.

These 11 specific markers of *B. anthracis* spores are encoded by the chromosome and have never been previously characterized. With a view to accurate identification of *B. anthracis* spores, these markers provide complementary results to the virulence markers encoded in pXO1 and pXO2 plasmids.

In a subsequent step, a multiplex SRM assay was designed to monitor four of the particular proteoforms of *B. anthracis* spores that were selected for their ability to generate a proteotypic peptide on enzymatic digestion. The method confirmed the positive detection of the spore of the 15 *B. anthra-*

*cis* strains and the absence of interfering signals for the 33 *B. cereus/B. thuringiensis* strains tested. This method represents a very efficient tool for the validation of marker specificity by screening hundreds or thousands of *B. cereus/B. thuringiensis* strains, as those used, for example, in the reference phylogenetic classification of Kolsto *et al.* (<http://mlstoslo.uio.no>) (56, 57). Furthermore, it could be used as a specific, sensitive and high-throughput detection method of *B. anthracis* spores in complex environmental samples. Indeed, targeted proteomics on triple quadrupole instruments (SRM acquisition mode) offers an attractive approach to the detection and/or quantification of proteins (70, 71). We have previously demonstrated its high sensitivity in the context of *B. anthracis* spore detection by targeting SASP-B protein without any prior culture step (35). The value of such an approach was illustrated here with the specific multiplexed detection of *B. anthracis* spores in a 100-fold excess of complex spore mixtures containing several *B. cereus* strains.

### CONCLUSION

We present here a proteomics (bottom-up and top-down) methodology combined with bioinformatics and genomics approaches applied to the identification of specific markers of *B. anthracis* spores. Using such an approach, 11 novel specific markers were successfully identified. Marker specificity was further validated by screening a large panel of 55 distinct strains chosen from *B. anthracis* and closely related *B. cereus/thuringiensis* species. Interestingly, the chromosomal location of the genes encoding these new markers also provides the opportunity to develop a robust genomic assay that would constitute an attractive alternative or complement to the existing methods targeting the *B. anthracis* virulence plasmids. In continuity with our previous work focused on SASP-B, we sought to implement a multiplex SRM method by targeting the four most suitable (mainly in terms of MS detection) markers among the 11 identified for the rapid and robust screening of *B. anthracis* spores. In addition, such a method could also be extended to the direct sensitive detection of *B. anthracis* spores in environmental samples without any prior culture step.

\* This project was supported by the joint ministerial program of R&D against Chemical, Biological, Radiological, Nuclear and Explosive (CBRNE) risks.

☒ This article contains supplemental Figs. S1 to S6 and Tables S1 to S4.

✉ To whom correspondence should be addressed: CEA, iBiTec-S, Service de Pharmacologie et d'Immunoanalyse (SPI), 91191 Gif-sur-Yvette, France. Tel.: 33-1-69-08-13-15; Fax: 33-1-69-08-59-07; E-mail: francois.becher@cea.fr; Institut Pasteur, Pathogénie des Toxi-Infections Bactériennes, 75724 Paris Cedex 15. Tel.: 33-1-45-68-83-12; Fax: 33-1-45-68-89-54; E-mail: pierre.goossens@pasteur.fr.

### REFERENCES

1. Mock, M., and Fouet, A. (2001) Anthrax. *Annu. Rev. Microbiol.* **55**, 647–671
2. Bartlett, J. G., Inglesby, T. V., and Borio, L. (2002) Management of anthrax. *Clin. Infect. Dis.* **35**, 851–858



3. Blackwood, K. S., Turenne, C. Y., Harmsen, D., and Kabani, A. M. (2004) Reassessment of sequence-based targets for identification of bacillus species. *J. Clin. Microbiol.* **42**, 1626–1630
4. Bavykin, S. G., Lysov, Y. P., Zakhariyev, V., Kelly, J. J., Jackman, J., Stahl, D. A., and Cherni, A. (2004) Use of 16S rRNA, 23S rRNA, and gyrB gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *J. Clin. Microbiol.* **42**, 3711–3730
5. Kolstø, A. B., Tourasse, N. J., and Økstad, O. A. (2009) What sets *Bacillus anthracis* apart from other *Bacillus* species? *Annu. Rev. Microbiol.* **63**, 451–476
6. Barker, M., Thakker, B., and Priest, F. G. (2005) Multilocus sequence typing reveals that *Bacillus cereus* strains isolated from clinical infections have distinct phylogenetic origins. *FEMS Microbiol. Lett.* **245**, 179–184
7. Vilas-Bôas, G. T., Peruca, A. P., and Arantes, O. M. (2007) Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. *Can. J. Microbiol.* **53**, 673–687
8. Granum, P. E., and Lund, T. (1997) *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol. Lett.* **157**, 223–228
9. Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D. R., and Dean, D. H. (1998) *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* **62**, 775–806
10. Van Ert, M. N., Easterday, W. R., Huynh, L. Y., Okinaka, R. T., Hugh-Jones, M. E., Ravel, J., Zanecki, S. R., Pearson, T., Simonson, T. S., U'Ren, J. M., Kachur, S. M., Leadem-Dougherty, R. R., Rhoton, S. D., Zinser, G., Farlow, J., Coker, P. R., Smith, K. L., Wang, B., Kenefic, L. J., Fraser-Liggett, C. M., Wagner, D. M., and Keim, P. (2007) Global genetic population structure of *Bacillus anthracis*. *PLoS One* **2**, e461
11. 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
12. Ellerbrok, H., Nattermann, H., Ozel, M., Beutin, L., Appel, B., and Pauli, G. (2002) Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR. *FEMS Microbiol. Lett.* **214**, 51–59
13. Hoffmaster, A. R., Meyer, R. F., Bowen, M. D., Marston, C. K., Weyant, R. S., Thurman, K., Messenger, S. L., Minor, E. E., Winchell, J. M., Rasmussen, M. V., Newton, B. R., Parker, J. T., Morrill, W. E., McKinney, N., Barnett, G. A., Sejvar, J. J., Jernigan, J. A., Perkins, B. A., and Popovic, T. (2002) Evaluation and validation of a real-time polymerase chain reaction assay for rapid identification of *Bacillus anthracis*. *Emerg. Infect. Dis.* **8**, 1178–1182
14. Ireng, L. M., Durant, J. F., Tomaso, H., Pilo, P., Olsen, J. S., Rasmussen, V., Mahillon, J., and Gala, J. L. (2010) Development and validation of a real-time quantitative PCR assay for rapid identification of *Bacillus anthracis* in environmental samples. *Appl. Microbiol. Biotechnol.* **88**, 1179–1192
15. Hu, X., Van der Auwera, G., Timmerly, S., Zhu, L., and 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**, 3016–3028
16. Hoffmaster, A. R., Ravel, J., Rasko, D. A., Chapman, G. D., Chute, M. D., Marston, C. K., De, B. K., Sacchi, C. T., Fitzgerald, C., Mayer, L. W., Maiden, M. C., Priest, F. G., Barker, M., Jiang, L., Cer, R. Z., Rillstone, J., Peterson, S. N., Weyant, R. S., Galloway, D. R., Read, T. D., Popovic, T., and Fraser, C. M. (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**, 8449–8454
17. Klee, S. R., Ozel, M., Appel, B., Boesch, C., Ellerbrok, H., Jacob, D., Holland, G., Leendertz, F. H., Pauli, G., Grunow, R., and Nattermann, H. (2006) Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *J. Bacteriol.* **188**, 5333–5344
18. Klee, S. R., Brzuszkiewicz, E. B., Nattermann, H., Brüggemann, H., Dupke, S., Wollherr, A., Franz, T., Pauli, G., Appel, B., Liebl, W., Couacy-Hymann, E., Boesch, C., Meyer, F. D., Leendertz, F. H., Ellerbrok, H., Gottschalk, G., Grunow, R., and Liesegang, H. (2010) The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS One* **5**, e10986
19. Leendertz, F. H., Ellerbrok, H., Boesch, C., Couacy-Hymann, E., Mätz-Rensing, K., Hakenbeck, R., Bergmann, C., Abaza, P., Junglen, S., Moebius, Y., Vigilant, L., Formenty, P., and Pauli, G. (2004) Anthrax kills wild chimpanzees in a tropical rainforest. *Nature* **430**, 451–452
20. Leendertz, F. H., Yumlu, S., Pauli, G., Boesch, C., Couacy-Hymann, E., Vigilant, L., Junglen, S., Schenk, S., and Ellerbrok, H. (2006) A new *Bacillus anthracis* found in wild chimpanzees and a gorilla from West and Central Africa. *PLoS Pathog.* **2**, e8
21. Ko, K. S., Kim, J. M., Kim, J. W., Jung, B. Y., Kim, W., Kim, I. J., and Kook, Y. H. (2003) Identification of *Bacillus anthracis* by rpoB sequence analysis and multiplex PCR. *J. Clin. Microbiol.* **41**, 2908–2914
22. Qi, Y., Patra, G., Liang, X., Williams, L. E., Rose, S., Redkar, R. J., and DelVecchio, V. G. (2001) Utilization of the rpoB gene as a specific chromosomal marker for real-time PCR detection of *Bacillus anthracis*. *Appl. Environ. Microbiol.* **67**, 3720–3727
23. La Duc, M. T., Satomi, M., Agata, N., and Venkateswaran, K. (2004) gyrB as a phylogenetic discriminator for members of the *Bacillus anthracis*-*cereus*-*thuringiensis* group. *J. Microbiol. Methods* **56**, 383–394
24. Rasmussen, V., Patra, G., Vaissaire, J., and Mock, M. (1999) The Ba813 chromosomal DNA sequence effectively traces the whole *Bacillus anthracis* community. *J. Appl. Microbiol.* **87**, 224–228
25. Hurtle, W., Bode, E., Kulesh, D. A., Kaplan, R. S., Garrison, J., Bridge, D., House, M., Frye, M. S., Loveless, B., and Norwood, D. (2004) Detection of the *Bacillus anthracis* gyrA gene by using a minor groove binder probe. *J. Clin. Microbiol.* **42**, 179–185
26. Rao, S. S., Mohan, K. V., and Atreya, C. D. (2010) Detection technologies for *Bacillus anthracis*: prospects and challenges. *J. Microbiol. Methods* **82**, 1–10
27. Ahmad, N. Z., Gupta, R. S., and Shah, H. N. (2011) Identification of a *Bacillus anthracis* specific indel in the yeaC gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J. Microbiol. Methods* **87**, 278–285
28. Sogawa, K., Watanabe, M., Sato, K., Segawa, S., Ishii, C., Miyabe, A., Murata, S., Saito, T., and Nomura, F. (2011) Use of the MALDI BioTyper system with MALDI-TOF mass spectrometry for rapid identification of microorganisms. *Anal. Bioanal. Chem.* **400**, 1905–1911
29. Welker, M., and Moore, E. R. (2011) Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst. Appl. Microbiol.* **34**, 2–11
30. Lasch, P., Beyer, W., Nattermann, H., Stämmler, M., Siegbrecht, E., Grunow, R., and Naumann, D. (2009) Identification of *Bacillus anthracis* by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. *Appl. Environ. Microbiol.* **75**, 7229–7242
31. Hathout, Y., Setlow, B., Cabrera-Martinez, R. M., Fenselau, C., and Setlow, P. (2003) Small, acid-soluble proteins as biomarkers in mass spectrometry analysis of *Bacillus* spores. *Appl. Environ. Microbiol.* **69**, 1100–1107
32. Castanha, E. R., Vestal, M., Hattan, S., Fox, A., Fox, K. F., and Dickinson, D. (2007) *Bacillus cereus* strains fall into two clusters (one closely and one more distantly related) to *Bacillus anthracis* according to amino acid substitutions in small acid-soluble proteins as determined by tandem mass spectrometry. *Mol. Cell Probes* **21**, 190–201
33. Swatkoski, S., Russell, S. C., Edwards, N., and Fenselau, C. (2006) Rapid chemical digestion of small acid-soluble spore proteins for analysis of *Bacillus* spores. *Anal. Chem.* **78**, 181–188
34. Pribil, P. A., Patton, E., Black, G., Doroshenko, V., and Fenselau, C. (2005) Rapid characterization of *Bacillus* spores targeting species-unique peptides produced with an atmospheric pressure matrix-assisted laser desorption/ionization source. *J. Mass Spectrom.* **40**, 464–474
35. Chenau, J., Fenaille, F., Ezan, E., Morel, N., Lamourette, P., Goossens, P. L., and Becher, F. (2011) Sensitive detection of *Bacillus anthracis* spores by immunocapture and liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **83**, 8675–8682
36. Brewis, I. A., and Brennan, P. (2010) Proteomics technologies for the global identification and quantification of proteins. *Adv. Protein Chem. Struct. Biol.* **80**, 1–44
37. Wang, Q., Chaekady, R., Wu, J., Hwang, H. J., Papadopoulos, N., Kopolovich, L., Maitra, A., Matthaehi, H., Eshleman, J. R., Hruban, R. H., Kinzler, K. W., Pandey, A., and Vogelstein, B. (2011) Mutant proteins as cancer-specific biomarkers. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2444–2449
38. Park, S. H., Oh, H. B., Seong, W. K., Kim, C. W., Cho, S. Y., and Yoo, C. K. (2007) Differential analysis of *Bacillus anthracis* after pXO1 plasmid curing

- and comprehensive data on *Bacillus anthracis* infection in macrophages and glial cells. *Proteomics* **7**, 3743–3758
39. Shahid, S., Park, J. H., Lee, H. T., Kim, S. J., Kim, J. C., Kim, S. H., Han, D. M., Jeon, D. I., Jung, K. H., and Chai, Y. G. (2010) Comparative proteome analysis of *Bacillus anthracis* with pXO1 plasmid content. *J. Microbiol.* **48**, 771–777
  40. Lamonica, J. M., Wagner, M., Eschenbrenner, M., Williams, L. E., Miller, T. L., Patra, G., and DelVecchio, V. G. (2005) Comparative secretome analyses of three *Bacillus anthracis* strains with variant plasmid contents. *Infect. Immun.* **73**, 3646–3658
  41. Delvecchio, V. G., Connolly, J. P., Alefantis, T. G., Walz, A., Quan, M. A., Patra, G., Ashton, J. M., Whittington, J. T., Chafin, R. D., Liang, X., Grewal, P., Khan, A. S., and Mujer, C. V. (2006) Proteomic profiling and identification of immunodominant spore antigens of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. *Appl. Environ. Microbiol.* **72**, 6355–6363
  42. Mukhopadhyay, S., Akmal, A., Stewart, A. C., Hsia, R. C., and Read, T. D. (2009) Identification of *Bacillus anthracis* spore component antigens conserved across diverse *Bacillus cereus* sensu lato strains. *Mol. Cell. Proteomics* **8**, 1174–1191
  43. Huang, C. M., Foster, K. W., DeSilva, T. S., Van Kampen, K. R., Elmets, C. A., and Tang, D. C. (2004) Identification of *Bacillus anthracis* proteins associated with germination and early outgrowth by proteomic profiling of anthrax spores. *Proteomics* **4**, 2653–2661
  44. Liu, H., Bergman, N. H., Thomason, B., Shallom, S., Hazen, A., Crossno, J., Rasko, D. A., Ravel, J., Read, T. D., Peterson, S. N., Yates, J., and Hanna, P. C. (2004) Formation and composition of the *Bacillus anthracis* endospore. *J. Bacteriol.* **186**, 164–178
  45. Misra, R. V., Ahmod, N. Z., Parker, R., Fang, M., Shah, H., and Gharbia, S. (2012) Developing an integrated proteo-genomic approach for the characterisation of biomarkers for the identification of *Bacillus anthracis*. *J. Microbiol. Methods* **88**, 237–247
  46. Morel, N., Volland, H., Dano, J., Lamourette, P., Sylvestre, P., Mock, M., and Créminon, C. (2012) Fast and sensitive detection of *Bacillus anthracis* spores by immunoassay. *Appl. Environ. Microbiol.* **78**, 6491–6498
  47. Ramière, V., Patra, G., Garrigue, H., Guesdon, J. L., and Mock, M. (1996) Identification and characterization of *Bacillus anthracis* by multiplex PCR analysis of sequences on plasmids pXO1 and pXO2 and chromosomal DNA. *FEMS Microbiol. Lett.* **145**, 9–16
  48. 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
  49. 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
  50. Lasch, P., Nattermann, H., Erhard, M., Stämmler, M., Grunow, R., Bannert, N., Appel, B., and Naumann, D. (2008) MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. *Anal. Chem.* **80**, 2026–2034
  51. Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A. J., Bunk, D. M., Kilpatrick, L. E., Billheimer, D. D., Blackman, R. K., Cardasis, H. L., Carr, S. A., Clauser, K. R., Jaffe, J. D., Kowalski, K. A., Neubert, T. A., Regnier, F. E., Schilling, B., Tegeler, T. J., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Fisher, S. J., Gibson, B. W., Kinsinger, C. R., Mesri, M., Rodriguez, H., Stein, S. E., Tempst, P., Paulovich, A. G., Liebler, D. C., and Spiegelman, C. (2010) Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **9**, 761–776
  52. Wynne, C., Fenselau, C., Demirev, P. A., and Edwards, N. (2009) Top-down identification of protein biomarkers in bacteria with unsequenced genomes. *Anal. Chem.* **81**, 9633–9642
  53. Wynne, C., Edwards, N. J., and Fenselau, C. (2010) Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an Orbitrap. *Proteomics* **10**, 3631–3643
  54. Smith, L. M., Kelleher, N. L., and Proteomics, C. f. T. D. (2013) Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186–187
  55. Contrepois, K., Ezan, E., Mann, C., and Fenaille, F. (2010) Ultra-high performance liquid chromatography-mass spectrometry for the fast profiling of histone post-translational modifications. *J. Proteome Res.* **9**, 5501–5509
  56. Tourasse, N. J., Okstad, O. A., and Kolstø, A. B. (2010) HyperCAT: an extension of the SuperCAT database for global multi-scheme and multi-datatype phylogenetic analysis of the *Bacillus cereus* group population. *Database* 2010, baq017
  57. Tourasse, N. J., Helgason, E., Klevan, A., Sylvestre, P., Moya, M., Haustant, M., Økstad, O. A., Fouet, A., Mock, M., and Kolstø, A. B. (2011) Extended and global phylogenetic view of the *Bacillus cereus* group population by combination of MLST, AFLP, and MLEE genotyping data. *Food Microbiol.* **28**, 236–244
  58. Gallien, S., Duriez, E., and Domon, B. (2011) Selected reaction monitoring applied to proteomics. *J. Mass Spectrom.* **46**, 298–312
  59. Helgason, E., Okstad, O. A., Caugant, D. A., Johansen, H. A., Fouet, A., Mock, M., Hegna, I., and Kolstø, A. B. (2000) *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* **66**, 2627–2630
  60. Jagtap, P., Michailidis, G., Zielke, R., Walker, A. K., Patel, N., Strahler, J. R., Driks, A., Andrews, P. C., and Maddock, J. R. (2006) Early events of *Bacillus anthracis* germination identified by time-course quantitative proteomics. *Proteomics* **6**, 5199–5211
  61. Edwards, K. A., Clancy, H. A., and Baemner, A. J. (2006) *Bacillus anthracis*: toxicology, epidemiology and current rapid-detection methods. *Anal. Bioanal. Chem.* **384**, 73–84
  62. Van Ert, M. N., Easterday, W. R., Huynh, L. Y., Okinaka, R. T., Hugh-Jones, M. E., Ravel, J., Zanecki, S. R., Pearson, T., Simonson, T. S., U'Ren, J. M., Kachur, S. M., Leadem-Dougherty, R. R., Rhoton, S. D., Zinser, G., Farlow, J., Coker, P. R., Smith, K. L., Wang, B., Kenefic, L. J., Fraser-Liggett, C. M., Wagner, D. M., and Keim, P. (2007) Global genetic population structure of *Bacillus anthracis*. *PLoS One* **2**, e461
  63. Papazisi, L., Rasko, D. A., Ratnayake, S., Bock, G. R., Remortel, B. G., Appalla, L., Liu, J., Dracheva, T., Braisted, J. C., Shallom, S., Jarrahi, B., Snesrud, E., Ahn, S., Sun, Q., Rilstone, J., Okstad, O. A., Kolstø, A. B., Fleischmann, R. D., and Peterson, S. N. (2011) Investigating the genome diversity of *B. cereus* and evolutionary aspects of *B. anthracis* emergence. *Genomics* **98**, 26–39
  64. Yuan, K., Johnson, W. C., Tipper, D. J., and Setlow, P. (1981) Comparison of various properties of low-molecular-weight proteins from dormant spores of several *Bacillus* species. *J. Bacteriol.* **146**, 965–971
  65. Setlow, P. (1988) Small, acid-soluble spore proteins of *Bacillus* species: structure, synthesis, genetics, function, and degradation. *Annu. Rev. Microbiol.* **42**, 319–338
  66. Bremer, H., and Dennis, P. P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. *Cell. Mol. Biol.* **2**, 1553–1569
  67. Pineda, F. J., Antoine, M. D., Demirev, P. A., Feldman, A. B., Jackman, J., Longenecker, M., and Lin, J. S. (2003) Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. *Anal. Chem.* **75**, 3817–3822
  68. Antelmann, H., Williams, R. C., Miethke, M., Wipat, A., Albrecht, D., Harwood, C. R., and Hecker, M. (2005) The extracellular and cytoplasmic proteomes of the non-virulent *Bacillus anthracis* strain UM23C1–2. *Proteomics* **5**, 3684–3695
  69. Park, S. H., Oh, H. B., Seong, W. K., Kim, C. W., Cho, S. Y., and Yoo, C. K. (2007) Differential analysis of *Bacillus anthracis* after pXO1 plasmid curing and comprehensive data on *Bacillus anthracis* infection in macrophages and glial cells. *Proteomics* **7**, 3743–3758
  70. Picotti, P., Aebersold, R., and Domon, B. (2007) The implications of proteolytic background for shotgun proteomics. *Mol. Cell. Proteomics* **6**, 1589–1598
  71. Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., and Aebersold, R. (2010) High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **7**, 43–46