A Comparative Genomics Analysis of Numerous *Bacillus cereus* Group Strains Supports Their Reclassification as *Bacillus anthracis*

Marco A. Riojas, PhD, Samantha L. Fenn, MPA, Manzour Hernando Hazbón, PhD

ATCC, Manassas, VA 20110

Abstract

Background. Relative to older methods based on phenotype or 16S rRNA sequencing, whole-genome sequencing (WGS) can provide more accurate and objective results for identification of bacterial strains and their assignment to the correct genus and species. To confirm the identification of several strains deposited as *Bacillus cereus* with BEI Resources and ATCC, we obtained the WGS and performed a variety of genomic analyses. We hypothesize that these strains may be related to previously described atypical *B. anthracis* (ABa) strains and may possibly be ancestral to modern *B. anthracis* (Ba).

Methods. The genomes of the BEI Resources and ATCC strains were compared to the type strain of *B. anthracis* using the Genome-to-Genome Distance Calculator (GGDC). For further comparison, the genomes of these strains were compared with those of strains deposited in GenBank as *B. anthracis*, *B. cereus*, or *B. thuringiensis*. A variety of comparative genomics tools were used to analyze the gene content, protein family distribution, locally colinear block arrangement, and other characteristics of the BEI Resources and ATCC strains, modern Ba strains, a subset of ABa strains, and a representative set of *B. cereus* (Bc) and *B. thuringiensis* (Bt) strains.

Results. Although the genomes of the BEI Resources and ATCC strains are circumscribed by the type strain of *B. anthracis* A0488^T, a series of markers typically associated with Ba (including the plasmid-borne virulence genes) are absent from them. These markers are also absent from the ABa strains analyzed, suggesting they may be related. A two-dimensional cluster analysis shows that the ABa strains and the two strains form a cluster closely related to but distinct from Ba strains and more distant from Bc and Bt strains.

Conclusions. According to the analytical methods employed, the BEI Resources and ATCC strains have characteristics that are more consistent with ABa strains than with *B. cereus*. Taken together, the results from these comparative genomics analyses provide very strong support of their reclassification as nonpathogenic strains of *B. anthracis*.

Introduction

The *Bacillus cereus* Group (BcG) is a group of Gram-positive aerobic rods that are closely related. Recently, many new species were added to the BcG, bringing the current total to 17 species.[1] The three best known members of the BcG are important in both health (*B. anthracis* and *B. cereus*) and biotechnology (*B. thuringiensis*). *B. anthracis* is the etiologic agent of the disease anthrax, while infection with *B. cereus* typically causes food poisoning (either diarrheal or emetic). However, a variety of *B. cereus* strains that have been responsible for anthrax-like disease have been described and have had their whole genome sequenced.[2-14] The primary cause for the significantly increased virulence of these strains is the incorporation of plasmids containing the anthrax toxin genes through natural transformation. Genomic analysis of many of these strains shows that – independent of plasmid content – their chromosome is more closely related to *B. anthracis* than to *B. cereus*.

Two strains deposited as *Bacillus cereus* (ATCC 2 and BEI Resources NR-22161) were recently sequenced at ATCC. The results indicated that both strains were more closely related to *B. anthracis* than to *B. cereus*. The further genomic characterization of these strains is described here. We hypothesize that these strains may be related to previously described atypical *B. anthracis* (ABa) strains[15] and may possibly be ancestral to phylogenetically modern *B. anthracis* (Ba).

Materials and Methods

Whole-Genome Sequencing (WGS). DNA from ATCC 2 and NR-22161 were prepared using the Nextera® XT Library Preparation Kit (Illumina®) and sequenced using Illumina MiSeq® v3 flow cells (2×300). Resultant paired-end reads underwent contamination detection using the One Codex microbial genomics read-based identification algorithm. Read pairs were then adapter trimmed and quality filtered, then used for *de novo* genome assembly using SPAdes 3.12.0. DNA from ATCC 2 was additionally sequenced using the Oxford Nanopore GridION X5. The sequencing reads from both platforms were combined via hybrid assembly using Unicycler.[16]

Selection of Strain Genomes. In addition to the BEI Resources and ATCC strains, NCBI genomes representing modern Ba strains, a subset of ABa strains, *B. cereus* (Bc) and *B. thuringiensis* (Bt) strains, and the type strains from the remaining BcG species were analyzed.

Genomic Analysis. Genomic distance based on digital DNA-DNA hybridization (dDDH) was calculated with the Genome-to-Genome Distance Calculator (GGDC) v2.1 using the recommended Formula 2.[17,18] The species delineation thresholds used were ≥70% via dDDH; a dDDH distance of ≥70-79.9% was considered to represent different subspecies of the same species, whereas ≥80% was considered to represent the same subspecies of the same species (or no subspecies in the case of species without multiple subspecies).[19] The calculated dDDH values were used as the basis for a phylogenetic tree as described previously.[20] The genomic characteristics of the genomes were analyzed using a sequence feature search (BLAST), analysis of locally colinear block arrangements (Mauve), and protein family distribution (using PATRIC[21]).

Results

Species Identification. The genomes of ATCC 2 and NR-22161 are circumscribed by the type strain of *B. anthracis* A0488^T (dDDH: 81.3 and 79.6%, respectively), whereas their distances from *B. cereus* ATCC 14579^T are far greater (dDDH: 45.1 and 44.9%, respectively). This confirms that these strains were erroneously deposited as *B. cereus*. The identity of these strains should be updated to reflect their proper taxonomy.

Global Genomic Organization. Based upon a locally colinear block analysis, the genome of ATCC 2 appears to share slightly more genomic organization with *B. anthracis* Ames Ancestor than with *B. cereus* ATCC 14579^T or *B. thuringiensis* CT-43 (**Figure 1**). However, the genomes share an overall very similar organization that would be expected from species in such a closely related group.

Phylogenomics. A phylogenomic tree of ATCC 2, NR-22161, the type strains of all 17 BcG species, 15 Ba strains, 15 ABa strains, 18 Bc strains, 8 Bt strains, and 2 strains of *B. paranthracis* is shown in **Figure 2** (at right).

Genomic Markers. The 77 genomes in the phylogenomic tree were queried for markers known to be common to typical strains of Ba, including the *sspE* repeat[22,23], the *plcR* nonsense mutation[24], and four prophages.[25] The results are shown as a trinary (present/partial/absent) trait alongside the phylogenomic tree in **Figure 2**. Importantly, neither ATCC 2 nor NR-22161 harbor the anthrax toxin genes from pXO1 or the poly-γ-d-glutamic acid capsule genes from pXO2 (data not shown). Thus, both of these strains most likely represent nonpathogenic strains of *B. anthracis*.

Protein Family Distribution. Using the Protein Family Sorter (PLfams) on PATRIC, ATCC 2 was compared to 4 Ba strains, 8 ABa strains, 4 Bc strains, and 3 Bt strains. The resulting heatmap illustrates distinct clusters of similar proteomic content (**Figure 3**). Notably, ATCC 2 clearly clusters with the ABa strains.

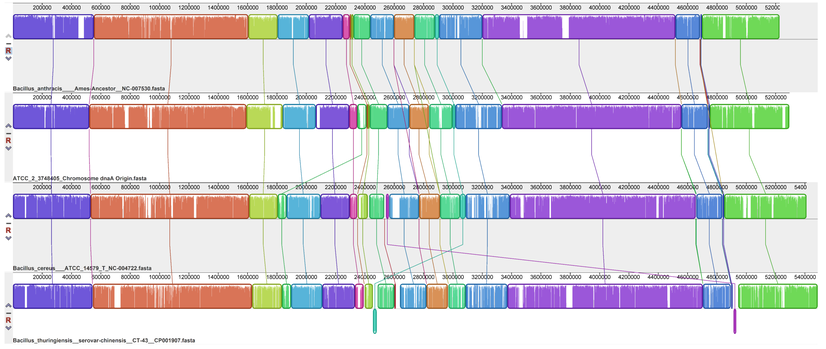

Figure 1. Mauve genomic alignment demonstrating that ATCC 2 displays a genomic organization more consistent with *B. anthracis* than with either *B. cereus* or *B. thuringiensis*.

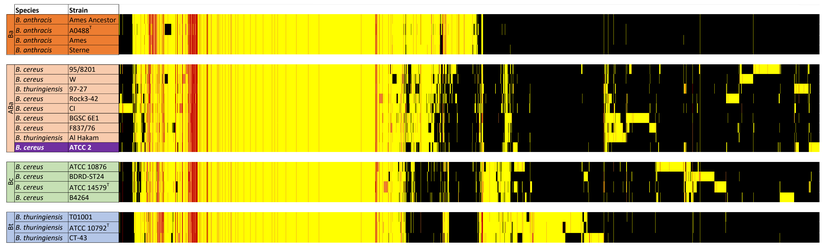

Figure 3. Clustered heatmap of distribution of genus-specific protein families computed using stringent criteria (PLfams). ■0, ■1, ■2, ■3+.

Acknowledgements

The following reagent was obtained through BEI Resources, NIAID, NIH: *Bacillus cereus*, Strain ISP3191, NR-22161.

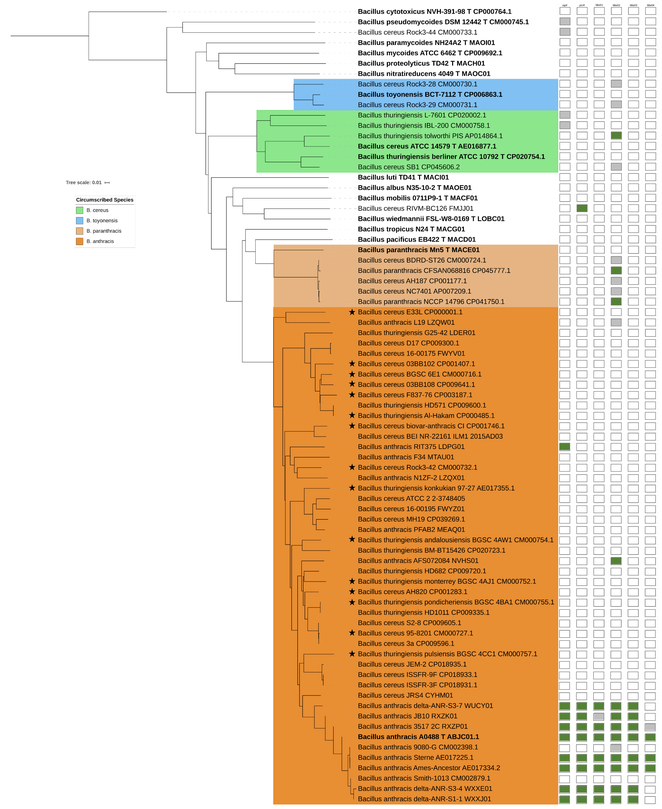

Figure 2. Phylogenomic tree inferred from calculated genomic distances. Genomes are labeled with their existing species classification. Type strains are shown in bold. Strains previously identified as ABa are designated with a star before the label. Circumscribed (dDDH ≥70%) species groups are shown as colored clades. Blocks at right indicate the presence (□absent, ■partial, ■present) of the *sspE* repeat, *plcR* mutation, and the four Ba prophages.

References

1. Liu Y *et al.* Proposal of nine novel species of the *Bacillus cereus* group. *Int J Syst Evol Microbiol* 67, 2499-2508, doi:10.1099/ijsem.0.001821 (2017).
2. Antonation KS *et al. Bacillus cereus* Biovar Anthracis Causing Anthrax in Sub-Saharan Africa-Chromosomal Monophyly and Broad Geographic Distribution. *PLoS Negl Trop Dis* 10, e0004923, doi:10.1371/journal.pntd.0004923 (2016).
3. Avashia SB *et al.* Fatal pneumonia among metalworkers due to inhalation exposure to *Bacillus cereus* containing *Bacillus anthracis* toxin genes. *Clin Infect Dis* 44, 414-416, doi:10.1086/510429 (2007).
4. Gee JE, Marston CK, Sammons SA, Burroughs MA, and Hoffmaster AR. Draft Genome Sequence of *Bacillus cereus* Strain BcFL2013, a Clinical Isolate Similar to G9241. *Genome Announc* 2, doi:10.1128/genomeA.00469-14 (2014).
5. Hoffmaster AR *et al.* Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes. *J Clin Microbiol* 44, 3352-3360, doi:10.1128/JCM.00561-06 (2006).
6. Hoffmaster AR *et al.* 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, doi:10.1073/pnas.0402414101 (2004).
7. Ishida R *et al.* Fatal community-acquired *Bacillus cereus* pneumonia in an immunocompetent adult man: a case report. *BMC Infect Dis* 19, 197, doi:10.1186/s12879-019-3836-3 (2019).
8. Klee SR *et al.* 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, doi:10.1371/journal.pone.0010986 (2010).
9. Klee SR *et al.* Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon. *J Bacteriol* 188, 5333-5344, doi:10.1128/JB.00303-06 (2006).
10. Leendertz FH *et al.* A new *Bacillus anthracis* found in wild chimpanzees and a gorilla from West and Central Africa. *PLoS Pathog* 2, e8, doi:10.1371/journal.ppat.0020008 (2006).
11. Marston CK *et al.* Anthrax Toxin-Expressing *Bacillus cereus* Isolated from an Anthrax-Like Eschar. *PLoS One* 11, e0156987, doi:10.1371/journal.pone.0156987 (2016).
12. Miller JM *et al.* Fulminating bacteremia and pneumonia due to *Bacillus cereus*. *J Clin Microbiol* 32, 504–507 (1997).
13. Saikia L *et al. Bacillus cereus*-Attributable Primary Cutaneous Anthrax-Like Infection in Newborn Infants, India. *Emerg Infect Dis* 25, 1261-1270, doi:10.3201/eid2507.181493 (2019).
14. Wright AM *et al.* Rapidly progressive, fatal, inhalation anthrax-like infection in a human: case report, pathogen genome sequencing, pathology, and coordinated response. *Arch Pathol Lab Med* 135, 1447-1459, doi:10.5858/2011-0362-SAIR.1 (2011).
15. Liu Y *et al.* Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci Rep* 5, 14082, doi:10.1038/srep14082 (2015).
16. Wick RR, Judd LM, Gorrie CL, and Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13, e1005595, doi:10.1371/journal.pcbi.1005595 (2017).
17. Auch AF, Klenk HP, and Goker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* 2, 142-148, doi:10.4056/sigs.541628 (2010).
18. Meier-Kolthoff JP, Auch AF, Klenk HP, and Goker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14, 60, doi:10.1186/1471-2105-14-60 (2013).
19. Meier-Kolthoff JP *et al.* Complete genome sequence of DSM 30083^T, the type strain (U5/41^T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci* 9, 2, doi:10.1186/1944-3277-9-2 (2014).
20. Riojas MA, McGough KJ, Rider-Riojas CJ, Rastogi N, and Hazbon MH. Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium microti* and *Mycobacterium pinnipedii* are later heterotypic synonyms of *Mycobacterium tuberculosis*. *Int J Syst Evol Microbiol* 68, 324-332, doi:10.1099/ijsem.0.002507 (2018).
21. Wattam AR *et al.* Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* 45, D535-D542, doi:10.1093/nar/gkw1017 (2017).
22. Kim K *et al.* Rapid genotypic detection of *Bacillus anthracis* and the *Bacillus cereus* group by multiplex real-time PCR melting curve analysis. *FEMS Immunol Med Microbiol* 43, 301-310, doi:10.1016/j.femsim.2004.10.005 (2005).
23. Riojas MA, Kiss K, McKee ML, and Hazbon MH. Multiplex PCR for species-level identification of *Bacillus anthracis* and detection of pXO1, pXO2, and related plasmids. *Health Secur* 13, 122-129, doi:10.1089/hs.2014.0056 (2015).
24. Agaisse H, Gominet M, Økstad OA, Kolsto A-B, and Lereclus D. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol Microbiol* 32, 1043-1053, doi:10.1046/j.1365-2958.1999.01410.x (1999).
25. Sozhamannan S *et al.* The *Bacillus anthracis* chromosome contains four conserved, excision-proficient, putative prophages. *BMC Microbiol* 6, 34, doi:10.1186/1471-2180-6-34 (2006).

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209 **Phone:** 800.638.6597 **Email:** sales@atcc.org **Web:** www.atcc.org