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

Marco A. Riojas, PhD, Samantha L. Fenn, MPA, Manzouz 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 or species. To confirm the identification of several strains deposited as Bacillus cereus with BEI Resources and ATCC, we have performed whole genome analysis 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 collinear block arrangements, and other characteristics of the BEI Resources and ATCC strains, modern B. cereus 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 ATCC 19792T, a variety of markers typically associated with B. cereus (including the plasmidborne virulence genes) are absent from them. These markers are also absent from the B. cereus strains analyzed, suggesting they may be related. A two-dimensional cluster analysis shows that the ATCC strains and the two strains form a cluster closely related to but distinct from B. cereus strains and more distantly from B. thuringiensis strains.

Conclusions. According to the analytical methods employed, the BEI Resources and ATCC strains have characteristics that are more consistent with B. cereus strains than with B. cereus. Taken together, the results from these comparative genomics analyses provide very strong support for 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 group, 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 their whole genome sequenced.2,3 The primary cause for the significantly increased virulence of these strains is the incorporation of plasmids carrying the anthrax toxin genes through natural transformation. Genomic analysis of many of these strains shows that independent of plasmid content, their chromosomes are 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. Further genomic characterization of these strains is described here. We hypothesize that these strains may be related to previously described atypical B. anthracis strains and may possibly be assigned to a new species of Bacillus, the identity of these strains should be updated to reflect their proper taxonomy.

Materials and Methods

Whole-Genome Sequencing (WGS). DNA from ATCC 2 and NR-22161 were prepared using the Illumina®XT Library Preparation Kit (Illumina®) and sequenced using Illumina MiSeq v4 150 cycle (2x300). Resultant paired-end reads underwent contamination detection using the One Codex microbial genomics read-based identification algorithm. Reads were then adapter trimmed and quality filtered, then used for de novo genome assembly using SPAdes 3.22.1. 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 B. cereus, Bacillus subtilis (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).21 The species designation thresholds used were 70% via dDDH, a dDDH distance of 70-79.5% was considered to represent different subspecies of the same species, whereas 85% was considered to represent the same subspecies of the same species (or no subspecies in the case of species with multiple subspecies).22 The calculated dDDH values were used as the basis for phylogenetic tree as described previously.20 The genomic characteristics of the genomes were analyzed using a sequence feature search (BLAST), analysis of locally collinear block arrangements (Mauve), and protein family distribution (using Phylogenetic).23

Results

Species Identification. The genomes of ATCC 2 and NR-22161 are circumscribed by the type strain of B. anthracis ATCC 19792T (dDDH 81.3 and 79.6%, respectively), whereas their distances from B. cereus ATCC 14579T are far greater (dDDH 45.1 and 44.4%, respectively). This confirms that these strains are 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 colorless block analysis, the genome of ATCC 2 appears to show slightly more genome organization with B. anthracis strains than with B. cereus ATCC 14579T or B. thuringiensis C7-43 (Figure 1). However, the genomes share a very similar organization that would be expected from species in such a closely related group.

Phylogenomics. The phylogenetic trees of ATCC 2, NR-22161, the type strains of all 17 BcG species, 15 Bc strains, 15 Bt strains, and 2 strains of B. thuringiensis are shown in Figure 2 (at right).

Genomic Markers. The 77 genomes in the phylogenetic tree were queried for markers known to be common to typical strains of Bc (including the spoligotype20,23,24, and four prophages25). The results are shown as a ternary (presence/absence/missing) tree alongside the phylogenetic tree in Figure 2. Importantly, neither ATCC 2 nor NR-22161 harbor the anthrax toxin genes from pXO1 or the poly-ε-diglutamic 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 (PLASPAR) on PATRIC, ATCC 2 was compared to 4 Bc strains, 8 Bt strains, and 3 Bt strains. The resulting heatmap illustrates distinct clusters of similar proteomic content (Figure 3). Notably, ATCC 2 clearly clusters with the Bc strains.

Acknowledgements

The following reagent was obtained through BEI Resources, NAID, NIH: Bacillus cereus. Strain 6SP19191, NR-22161

References

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

Marco A. Riojas, Samantha L. Fenn

ATCC, Manassas, VA 20110

© 2018 American Type Culture Collection, Inc. The ATCC trademark and trade name, and any other trademarks listed in this publication are trademarks owned by the American Type Culture Collection unless indicated otherwise. Illumina, MiSeq, and NCBI are registered trademarks of Illumina, Inc.