Phylogenomic Comparison of *Bacillus cereus* Group Strains to Recently Identified Type Strains Supports the Species Reclassification of Many Strains

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

**ATCC, Manassas, Virginia 20110**

---

**Abstract**

The *Bacillus* cereus Group (BcG) is a group of closely related species that are important in health (e.g., *B. anthracis* and *B. cereus*) and biotechnology (e.g., *B. thuringiensis*). Recently, many new type strains have been assigned to type species (e.g., *B. thuringiensis*). It is useful to revisit the species classification of existing BcG strains to determine whether their species assignments accurately reflect their genetic and physiological relatedness. This is particularly relevant for older strains that may have been assigned to a species at a time when taxonomic differentiation was primarily based upon phenotypic characteristics. BcG includes species that, based on genome-wide data, can be well distinguished from each other. In this study, we compared the genomes of 21 ATCC and 14 BEI Resources strains from BcG species using Illumina MiSeq® v3 flow cells (2×300) for whole-genome shotgun (WGS) sequencing. Resultant paired-end reads underwent contamination detection using the One Codex Manhattan tool. The genome assembly of *B. anthracis* NY9979 was used as the reference for assembly. Provided here is an overview of the phylogenomic characteristics of the BcG strains.

**Methods**

The whole-genome sequences of 21 ATCC and 14 BEI Resources strains from BcG species were obtained via Illumina sequencing. The genome-to-genome distances (GGDs) between the genomes of these strains and that of the 16 BcG type strains present in GenBank were determined using the Genome-to-Genome Distance Calculator (GGDC). The GGDs were compared in a pairwise manner. These GGDs were used as the basis for inferring phylogeny via FastTree 2.0.

**Results**

The pairwise analysis of the 16 type strain genomes shows that 16 of these are correctly identified as independent species. Two of the type strain genomes fall within the circumscriptio of other type strains. *B. weihenstephanensis* NR-610 falls within the circumscriptio of *B. mycoides* ATCC 94422 with a GGD of 71.1%, which indicates that it may represent a subspecies of *B. cereus*. Of the 35 ATCC/BIEI Resources genomes examined, 17 were confirmed as correctly belonging to the assigned species. Of the remaining strains, 13 fell within the circumscriptio of species other than their assigned species. Five strains did not fall within the circumscriptio of any existing species (but are closely related to each other at GGDs of 91.1%), suggesting that they may represent a novel species within the BcG.

**Conclusions**

The phylogenomic analysis described here illustrates the importance of reconsidering the genus-level classification of *Bacillus* strains to ensure their accurate alignment with the most current taxonomy.

---

**Materials and Methods**

**Bacterial Strains and DNA Extraction**

A subset of *Bacillus cereus* Group (BoG) items from ATCC and BEI Resources were selected for this study. Strains were grown in nutrient broth, and DNA was extracted using the QIAGEN™ MagAttract™ High-Molecular Weight (HMW) kit. Additionally, existing genomes from GenBank were also used in the analysis. Together, the genomes from the strains sequenced and those from GenBank comprised the main dataset for the phylogenomic analysis (Table 1).

**Whole-Genome Sequencing (WGS)**

DNA was prepared using the Nester™ XT Library Preparation Kit (Ilumina™) and sequenced using Illumina HiSeq™ 2×300 paired-end reads. 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.

**Calculation of Genomic Distances**

For independent combination of the two algorithmic approaches, were used. The genomic distance based on DNA-DNA hybridization (dDDH) was calculated with the Genome-to-Genome Distance Calculator (GGDC) v2.1 using the recommended Formula 2.1. Average nucleotide identity (ANI) was calculated using QHitch™. The species delineation threshold used were 75% via dDDH and 75% via ANI. A dDDH distance of 70-79.9% was considered to represent different subspecies of the same species, whereas 95% was considered to represent the same subspecies of the same species (or no subspecies in the case of species without multiple subspecies). No subspecies delineation threshold based on ANI values currently exists. The calculated dDDH values were used as the basis for a phylogenetic tree as described previously.

---

**Conclusions**

The results of the current work confirm our hypothesis that the species names assigned to items deposited into culture collections may not correspond to the modern bacterial taxonomy. Our results reiterate the importance of periodic reevaluation of culture collection holdings against the changes that may have occurred in bacterial systematics since the time of deposit. To this end among others, ATCC has shifted its business model from being a passive culture collection to a dynamic and forward-looking biological resource center that anticipates the needs of the scientific community in an effort to be more responsive. This includes the use of phylogenomic analyses for authentication of strains and maintaining a pulse on the changes in bacterial systematics.