June 21, 2019
FRIDAY – EEB-488

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, VA 20110

ATCC®
Credible leads to Incredible™

Abstract

Background. 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 species were added to the BcG, bringing the current total to 18 species. With this recent expansion of the BcG, it is useful to revisit the species classification of existing BcG strains to determine whether their species assignments require realignment with the most current taxonomy of the group. 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 observations.

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 18 BcG type strains present in GenBank were determined by using the Genome-to-Genome Distance Calculator (GGDC); genomes were compared in a pairwise manner. These GGDs were used as the basis for inferring phylogeny via FastME 2.0.

Results. The pairwise analysis of the 18 type strain genomes shows that 16 of these are correctly identified as independent species. Two of the type strains fall within the circumscription of other type strains. *B. weihenstephanensis* NBRC 101238^T falls within the circumscription of *B. mycoides* ATCC® 6462^T with a GGD of 78.7%, which indicates that it represents a subspecies of *B. mycoides*. Additionally, *B. thuringiensis* ATCC® 10792^T falls within the circumscription of *B. cereus* ATCC® 14579^T with a GGD of 71.1%, which indicates that it may represent a subspecies of *B. cereus*.
Of the 35 ATCC/BEI Resources genomes examined, 17 were confirmed as correctly belonging to the assigned species. Of the remaining strains, 13 fell within the circumscription of species other than their assigned species. Five strains did not fall within the circumscription of any existing species (but are closely related to each other at GGDs ≥91.1%), suggesting that they may represent a novel species within the BcG.

Conclusions. The phylogenomic analysis described here illustrates the importance of reexamining the identity of existing strains via the most recent tools and taxonomic information. Particularly with items deposited decades before our modern understanding of genotypic characterization, such phylogenomic reexaminations enable the taxonomic reassessment of strains to ensure their accurate alignment with the most current taxonomy.

Introduction

The assignment of a newly isolated strain to a species is necessarily temporally linked to the existing species at the time. The properties of the new isolate are compared to those of the existing type strains to determine whether it falls within the circumscription of an existing species. If it does, it can be assigned to that species. If it does not, it can be proposed as a novel species. For the species to be described and distribution to the scientific community, isolated strains may be deposited into one or more of the various culture collections around the world. These strains are assigned to a species based upon the declaration by the depositor and/or input from the culture collection. However, because new species are discovered and the circumscription of older species are emended, it is possible that the species names for strains in culture collections may reflect the taxonomy contemporaneous to the time of deposit and may no longer reflect the current taxonomy.

Before the adoption of molecular techniques, the assignment of a strain to a species was based upon phenotypic characteristics such as cellular and colony morphology, growth conditions, or biochemical assays. While use of such criteria as the basis for species differentiation may have been the best method available at the time, the genomic and bioinformatic tools available today offer a much greater analytical resolution and allow the identification of bacterial strains to be based upon their whole genome. Although the gold standard for the genomic circumscription of bacterial species has been considered DNA-DNA hybridization (DDH), this wet-lab technique suffers from significant limitations, specifically that it is prone to error, is labor-intensive, and struggles with reproducibility.[1,2] Due to the current preponderance of genome sequences, a more useful sequence-based adaptation of DDH is digital DDH (dDDH).[2,3] This technique has been applied to better align the taxonomy of many taxa with their evolutionary histories, including *Escherichia*, the *Bacillus cereus* Group, *Aeromonas*, and the *Mycobacterium tuberculosis* complex[4], among many others.

Taken together, the continual evolution of both bacterial taxonomy and of the criteria and tools used for differentiation suggest that the species names assigned to some organisms in culture collections (particularly older items) may be incorrect with respect to modern bacterial taxonomy. In the current study, we used whole-genome sequencing and modern phylogenomic analyses to determine whether the species names assigned to a subset of *Bacillus* items within the ATCC and BEI Resources collections correspond to current taxonomic definitions.

Materials and Methods

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

Whole-Genome Sequencing (WGS). DNA was 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.

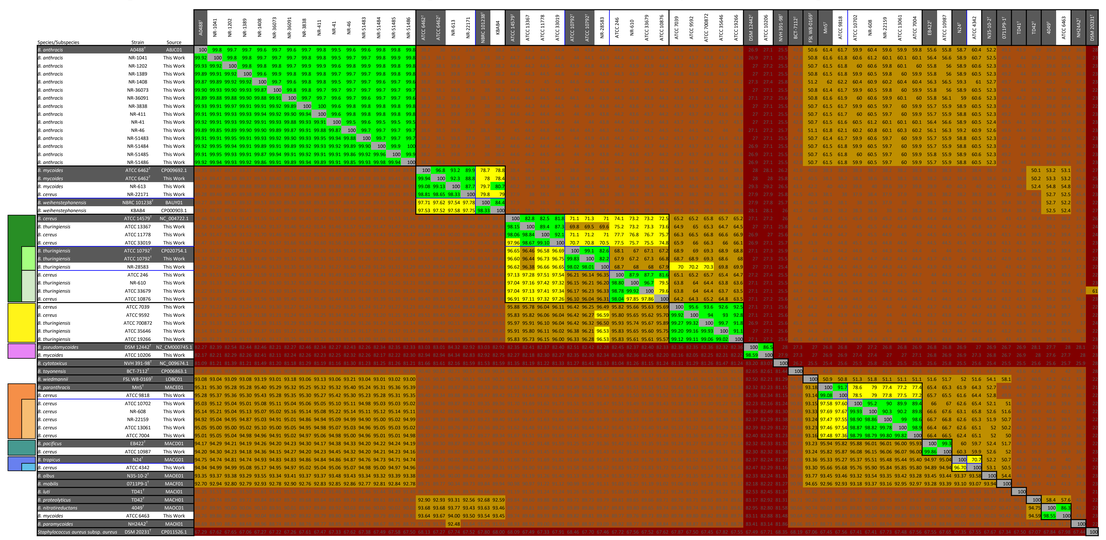
Calculation of Genomic Distances. For independent corroboration of the results, two algorithmic approaches were used. Genomic distance based on digital DNA-DNA hybridization (dDDH) was calculated with the Genome-to-Genome Distance Calculator (GGDC) v2.1 using the recommended Formula 2.[2,3] Average nucleotide identity (ANI) was calculated using OrthoANIu.[5] The species delineation thresholds used were ≥70% via dDDH and ≥96% via ANI.[6,7] 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).[8] 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.[4]

Results

The complete pairwise dDDH/ANI table is shown in Table 1. The phylogenomic tree is shown in Figure 1; proposed taxonomic changes based upon whole-genome comparison are provided. Of specific note are the subspecies-level positioning of *B. weihenstephanensis* within *B. mycoides* and the subspecies-level positioning of *B. thuringiensis* within *B. cereus*. Additional subspecies-level clades were detected within *B. cereus*, *B. paranthracis*, and *B. tropicus*, representing novel subspecies of these species.

Of the 41 ATCC/BEI Resources strains that were sequenced by this work, 28 (68.3%) should be updated to better reflect their taxonomy. This includes 21 (51.2%) strains that should be assigned to a different species other than their current species and 20 (48.8%) that should be assigned to a subspecies for additional taxonomic precision (Figure 1).

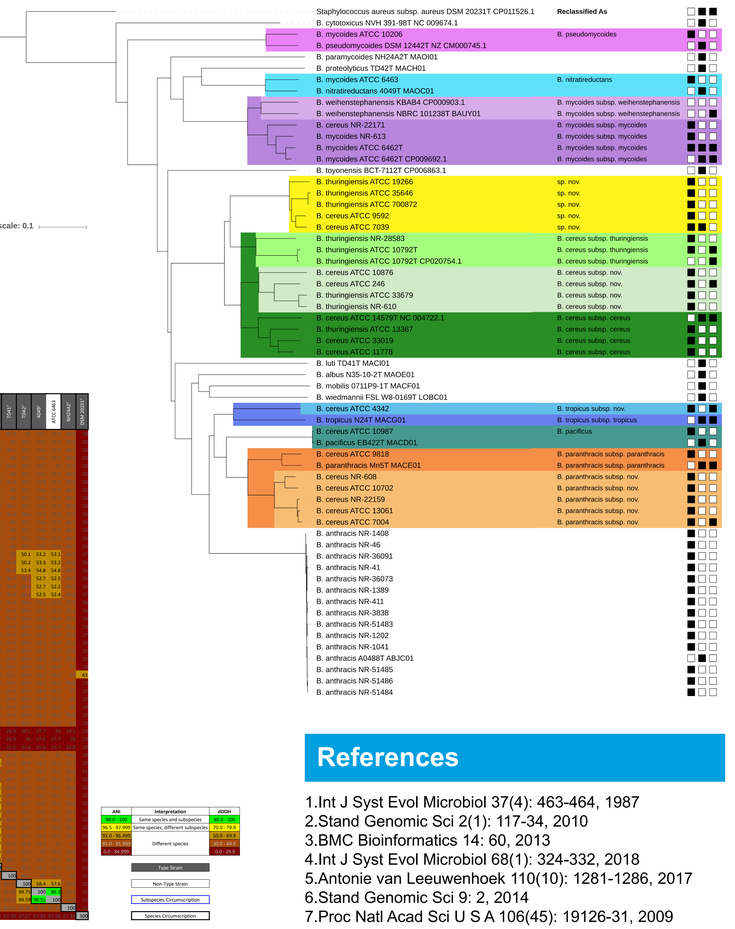
Table 1. Genomic distances between the *Bacillus* strains examined in this work. dDDH values are shown above the self-comparison diagonal; ANI values are shown below the diagonal. Colored blocks on left correspond to colors in Figure 1.



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 reassessment 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.

Figure 1. Phylogenomic tree of the *Bacillus* strains examined. The name on each leaf represents the strain's existing taxonomic identification; at right is the proposed reclassified name.



ANI	Interpretation	dDDH
98.5 - 100	Same species and subspecies	80.0 - 100
96.5 - 97.999	Same species, different subspecies	70.0 - 79.9
92.0 - 96.499	Different species	30.0 - 69.9
91.0 - 91.999		25.0 - 29.9
0.0 - 84.999		0.0 - 24.9

References

1.Int J Syst Evol Microbiol 37(4): 463-464, 1987
2.Stand Genomic Sci 2(1): 117-34, 2010
3.BMC Bioinformatics 14: 60, 2013
4.Int J Syst Evol Microbiol 68(1): 324-332, 2018
5.Antonie van Leeuwenhoek 110(10): 1281-1286, 2017
6.Stand Genomic Sci 9: 2, 2014
7.Proc Natl Acad Sci U S A 106(45): 19126-31, 2009

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209 **Phone:** 703.263.8277 **Email:** SalesRep@atcc.org **Web:** www.atcc.org