**Bacillus velezensis** is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and ‘*Bacillus oryzicola*’ are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics

Christopher A. Dunlap, 1 Soo-Jin Kim, 1,2 Soon-Wo Kwon 2 and Alejandro P. Rooney 1

1Crop Bioprotection Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, United States Department of Agriculture, Peoria, IL, USA
2Korean Agriculture Culture Collection (KACC), National Institute of Agricultural Science, Rural Development Administration, Wanju-gun, Jeollabuk-do, Republic of Korea

*Bacillus velezensis* was previously reported to be a later heterotypic synonym of *Bacillus amyloliquefaciens*, based primarily on DNA–DNA relatedness values. We have sequenced a draft genome of *B. velezensis* NRRL B-41580 T. Comparative genomics and DNA–DNA relatedness calculations show that it is not a synonym of *B. amyloliquefaciens*. It was instead synonymous with *Bacillus methylotrophicus*. ‘*Bacillus oryzicola*’ is a recently described species that was isolated as an endophyte of rice (*Oryza sativa*). The strain was demonstrated to have plant-pathogen antagonist activity in greenhouse assays, and the 16S rRNA gene was reported to have 99.7% sequence similarity with *Bacillus siamensis* and *B. methylotrophicus*, which are both known for their plant pathogen antagonism. To better understand the phylogenetics of these closely related strains, we sequenced the genome of ‘*B. oryzicola*’ KACC 18228. Comparative genomic analysis showed only minor differences between this strain and the genomes of *B. velezensis* NRRL B-41580 T, *B. methylotrophicus* KACC 13015 T and *Bacillus amyloliquefaciens* subsp. *plantarum* FZB42 T. The pairwise in silico DNA–DNA hybridization values calculated in comparisons between the strains were all greater than 84%, which is well above the standard species threshold of 70%. The results of morphological, physiological, chemotaxonomic and phylogenetic analyses indicate that the strains share phenotype and genotype coherence. Therefore, we propose that *B. methylotrophicus* KACC 13015 T, *B. amyloliquefaciens* subsp. *plantarum* FZB42 T, and ‘*B. oryzicola*’ KACC 18228 should be reclassified as later heterotypic synonyms of *B. velezensis* NRRL B-41580 T, since the valid publication date of *B. velezensis* precedes the other three strains.

*Bacillus velezensis* was originally described by Ruiz-García et al. (2005), after discovery in a screen of environmental isolates for novel lipopeptides. In that study, *B. velezensis* was shown to be closely related to *Bacillus subtilis* and *Bacillus amyloliquefaciens*. However, it was subsequently declared a later heterotypic synonym of *B. amyloliquefaciens* by Wang et al. (2008) based on the results of DNA–DNA relatedness studies. Our laboratories have recently been working on whole-genome sequencing of type strains in the *Bacillus subtilis* group to resolve outstanding problems in their phylogenetic systematics (Dunlap, 2015; Dunlap et al., 2015). Recently, we reported that the names of two important biological control strains, *Bacillus methylotrophicus* and *Bacillus amyloliquefaciens* subsp. *plantarum*, were synonymous (Dunlap et al., 2015). However, in a preliminary analysis of the draft

**Abbreviations:** ANI, Average Nucleotide Identity; DDH, DNA–DNA hybridization; SNP, single nucleotide polymorphism.
Phylogenomic analysis of *Bacillus velezensis*

The genome of type strain of *B. velezensis* (Jeong *et al.*, 2015), we found that it was nearly identical to *B. methylotrophicus*.

The recent description of ‘*Bacillus oryzicola*’, a root endophyte of rice (*Oryza sativa*) (Chung *et al.*, 2015), has further complicated matters. *B. oryzicola* YC7010 was reported to be a plant growth promoter, a plant pathogen antagonist, and an inducer of systemic resistance in rice (Chung *et al.*, 2015). *B. methylotrophicus* KACC 13015T (=CBMB205T) was also isolated from the rice rhizosphere (Korea) and has been shown to possess plant growth promotion and antagonism towards plant pathogens (Ma *et al.*, 2013; Madhaiyan *et al.*, 2010). Interestingly, *B. oryzicola* YC7010 was reported to have a very high level of 16S rRNA gene sequence similarity to *B. methylotrophicus* as well as *B. sianensis* (99.7 % for both) (Chung *et al.*, 2015). In addition, the strain was reported to have a DNA G+C content of 50.5 mol%, which is much higher than *B. methylotrophicus* with 46.4 mol% (Dunlap *et al.*, 2015). When examined collectively, all of the above observations suggest that ‘*B. oryzicola*’ may be a later heterotypic synonym of *B. methylotrophicus*.

The above situation involving *B. velezensis*, *B. amyloliquefaciens* subsp. *plantarum*, *B. methylotrophicus* and ‘*B. oryzicola*’ creates a particularly confounding taxonomic problem (if not nightmare) for this species group. In order to resolve this problem, we conducted a phylogenomic study of the aforementioned taxa by using complete genome sequence data of both type and representative strains in order to determine the extent to which their genomes vary from one another and to reconstruct their phylogenetic relationships.

*‘B. oryzicola’* KACC 18228 (=YC7010) and *B. methylotrophicus* KACC 13015T (=CBMB205T) were obtained from the Korean Agriculture Culture Collection. *B. velezensis* NRRL B-41580T (=CR-502T=BCRC 17467T=KCTC 13102T=LMG 22478T) and *B. amyloliquefaciens* subsp. *plantarum* NRRL B-59544 (=FZB42T) were obtained from the Northern Regional Research Laboratory in Peoria, IL, USA. Temperature growth studies were conducted from 4 to 60 °C on tryptone-glucose-yeast extract agar and evaluated at 48 h. NaCl tolerance (%) was conducted in 2 % increments from 0 to 20 % (w/v) in Luria–Bertani (LB) media at 30 °C and evaluated at 48 h. All strains were cultured overnight on Biolog universal growth plates and prepared according to manufacturer’s instructions for the GEN III MicroPlate test panel using protocol A (Biolog); the experiment was run in triplicate.

The previously published phenotypic data for *B. velezensis* NRRL B-41580T, *B. methylotrophicus* KACC 13105T, *B. amyloliquefaciens* subsp. *plantarum* FZB42T and ‘*B. oryzicola*’ KACC 18228 are consistent with the strains belonging to the same species (Borriess *et al.*, 2011; Chung *et al.*, 2015; Madhaiyan *et al.*, 2010; Ruiz-Garcia *et al.*, 2005). The previously published fatty acid methyl ester data are slightly variable, but the primary fatty acid components are the same. In addition, we conducted a Biolog analysis of the four strains and found no notable differences (Fig. S1, available in the online Supplementary Material). We also conducted temperature and NaCl tolerance experiments (data not shown), with no notable differences identified. The only notable deviations found when comparing the previously published phenotypic, physiological, or chemotaxonomic properties of these taxa are (1) DNA G+C content and (2) 16S rRNA gene sequence similarity (Chung *et al.*, 2015; Madhaiyan *et al.*, 2010).

The genomes of *B. velezensis* NRRL B-41580T and ‘*B. oryzicola*’ KACC 18228 were sequenced using a MiSeq DNA sequencer (Illumina) and the MiSeq V3 2 ×300 sequencing kit (Illumina) following the manufacturer’s suggested protocols. The resulting reads were quality trimmed to the Q30 confidence level. The draft genome was assembled using CLCbio Genomics Workbench 8.0 (Qiagen) using default parameters. Genome comparisons and alignments for phylogenetic trees were made using BIGSdb software (Jolley & Maiden, 2010). The digital DNA–DNA hybridizations (DDH) were determined online at http://ggdc.dsmz.de/distcalc2.php using the Genome-to-Genome Distance Calculation (GGDC) version 2.0 as described in Meier-Kolthoff *et al.* (2013). The estimated DDH values were calculated using formula two at the GGDC website, originally described in Auch *et al.* (2010) and updated in Meier-Kolthoff *et al.* (2013). Average Nucleotide Identity (ANI) was performed as previously described (Goris *et al.*, 2007), with the following options; minimum length 700 bp, minimum identity 70 %, minimum alignment 50 %, BLAST window size 1000 bp and step size of 200 bp. Phylogenetic analyses were conducted using MEGA software version 6.06 (Tamura *et al.*, 2013). Neighbour-joining trees were reconstructed using the Tamura–Nei model (Tamura & Nei, 1993) with a gamma correction (alpha value=0.5); this model was chosen on the basis of the likelihood test implemented in MEGA 6.06. A total of 1500 bootstrap pseudoreplications were performed in order to gauge support for internal nodes.

The draft genome of *B. velezensis* NRRL B-41580T was assembled and yielded 24 contigs with a total length of 4 034 335 bp at 79 × coverage and 46.3 mol% G+C content. While the draft genome of ‘*B. oryzicola*’ KACC 18228 yielded 25 contigs with a total length of 3 928 423 bp at 40 × coverage and 46.4 mol% G+C content. The draft genomes were used to evaluate similarity of closely related strains using an in silico DDH determination (Table 1) and ANI determination (Table S1). The results show the type strains of *B. methylotrophicus*, *B. amyloliquefaciens* subsp. *plantarum* and ‘*B. oryzicola*’ all have pairwise DDH values well above the recommended threshold of 70 % for species delineation (Wayne *et al.*, 1987). The results contradict the previously reported wet lab DDH values for *B. velezensis* BCRC 17467T (Wang *et al.*, 2008) and ‘*B. oryzicola*’ KACC 18228 (Chung *et al.*, 2015). The ANI values produced a similar result with the type strains of *B. velezensis*, *B. methylotrophicus*, *B. amyloliquefaciens*...
subsp. plantarum and ‘B. oryzicola’ all having pairwise ANI values greater 98% (Table S1). The recommended cut-off point of 70% DDH for species delineation corresponds to approximately 95% ANI (Goris et al., 2007). Our draft genome of B. velezensis NRRL B-41580T was essentially clonal with the recently reported draft genome of B. velezensis KCTC 13012T (Jeong et al., 2015).

The DNA G+C content for ‘B. oryzicola’ KACC 18228 was originally reported at 50.5 mol% (Chung et al., 2015) and 45 mol% for B. methylotrophicus KACC 13015T (Madhaiyan et al., 2010) using wet lab methods for both. While the sequenced genome data provide a G+C content of 46.4 mol% for B. methylotrophicus KACC 13015T (Dunlap et al., 2015) and 46.4 mol% for ‘B. oryzicola’ KACC 18228 in the current study. The discrepancy between the original data generated using the conventional (experimental) method and the whole genome data generated using next-generation sequencing techniques suggests that the conventional method has some experimental error, which is not surprising. This observation is supported by a recent study comparing data from conventional methods and genome-sequencing data, which showed G+C content varies 3–5 mol% within species for conventional methods.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genome-to-genome comparison (regression-based DNA–DNA hybridization, %) with:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1. B. velezensis NRRL-B-41580T</td>
<td>84.9</td>
</tr>
<tr>
<td>2. ‘B. oryzicola’ KACC 18228</td>
<td>84.9</td>
</tr>
<tr>
<td>3. B. methylotrophicus KACC 13105T</td>
<td>84.5</td>
</tr>
<tr>
<td>4. B. amyloliquefaciens FZB42T</td>
<td>85.8</td>
</tr>
<tr>
<td>5. B. amyloliquefaciens DSM 7T</td>
<td>55.5</td>
</tr>
<tr>
<td>6. B. siamensis KCTC-13613T</td>
<td>56.7</td>
</tr>
<tr>
<td>7. B. subtilis NRRL NRS-744T</td>
<td>20.6</td>
</tr>
</tbody>
</table>

Table 1. Results of genome-to-genome distance comparisons of closely related type strains from the B. subtilis group

Fig. 1. Phylogeny of the Bacillus subtilis species group reconstructed from a neighbour-joining analysis of core-genome sequence data (799 genes). Bootstrap values >50%, based on 1500 pseudoreplicates are indicated at branch points. Bar, 0.05 nt substitutions per site.
Fig. 2. Phylogeny of selected strains of species from the *Bacillus velezensis*, *Bacillus amyloliquefaciens* and *Bacillus siamensis* clades reconstructed from a neighbour-joining analysis of core-genome sequence data (2740 genes). Bar, 0.05 nt substitutions per site.
and within 1 % for whole-genome sequencing data (Meier-Kolthoff et al., 2014).

After genome sequencing of both B. velezensis NRRL B-41580T and ‘B. oryzicola’ KACC 18228, we extracted the 16S rRNA gene sequence and used the EzTaxon-e database (Kim et al., 2012) to identify the closest relatives. The analysis identified the nearest neighbour for both strains was B. methylotrophicus (JTKJ01000077) with 100 % sequence similarity. To understand the difference between our findings and the previous report of 99.7 % sequence similarity for ‘B. oryzicola’ KACC 18228 (Chung et al., 2015), we did an alignment of 16S rRNA gene sequences (Fig. S2). The only difference observed between the ‘B. oryzicola’ KACC 18228 16S rRNA gene sequence generated in Chung et al. (2015) and our study was a C to T single nucleotide polymorphism (SNP) at position 146. This difference was not enough to explain the observed differences in reported sequence similarity. After review, the differences arise from different 16S rRNA gene sequences reported for the type strain of B. methylotrophicus at the EzTaxon-e database. In the current study, the 16S rRNA gene sequence of B. methylotrophicus KACC 13105T obtained at the EzTaxon-e database is from GenBank accession number JTKJ01000077, based on whole genome data released 23 March 2015 (Dunlap et al., 2015). Prior to the EzTaxon-e database being updated, the 16S rRNA gene of the type strain of B. methylotrophicus would have been from GenBank accession number EU194897 (Madhaiyan et al., 2010). There are 8 nt differences between these two 16S rRNA gene sequences, which arise from three N’s and three deletions in the first 15 nt, an insert at position 624 and two deletions at positions 1476–1477 of GenBank accession number EU194897. We have recently reported similar discrepancies in other strains from the Bacillus subtilis group 16S rRNA gene sequence data obtained from capillary electrophoresis as well as whole-genome sequence data, which also resulted in part to another synonym being described as a new species (Dunlap, 2015). It is notable that the 16S rRNA gene sequences extracted from the genomes of the four conspecific type strains, only two SNPs were observed.

The phylogenomic tree based on the core genome (799 genes) of the strains of the Bacillus subtilis group showed that B. velezensis NRRL B-41580T, B. methylotrophicus KACC 13105T, B. amyloliquefaciens subsp. plantarum FZB42T and ‘B. oryzicola’ KACC 18228 were closely related (Fig. 1). The results are consistent with B. velezensis NRRL B-41580T, B. methylotrophicus KACC 13105T, ‘B. oryzicola’ KACC 18228 and B. amyloliquefaciens subsp. plantarum FZB42T being conspecific. To further resolve the confusion surrounding these closely related strains, we downloaded and analysed all of the genomes of species in the Bacillus subtilis group present in the GenBank database. Strains belonging to the clade containing B. velezensis and its nearest neighbours B. amyloliquefaciens and B. siamensis were used to reconstruct a tree based on the sequences of genes in their core genomes (Fig. 2). The results highlight the confusion that has arisen regarding the nomenclature and taxonomy of these strains commonly isolated as plant pathogen antagonists and developed as biological control agents (see Dunlap et al., 2015). Because the valid publication of B. velezensis (Ruiz-Garcia et al., 2005) predates the publication of B. methylotrophicus (Madhaiyan et al., 2010), B. amyloliquefaciens subsp. plantarum (Boriss et al., 2011) and ‘B. oryzicola’ (Chung et al., 2015), we propose that B. methylotrophicus, B. amyloliquefaciens subsp. plantarum, and ‘B. oryzicola’ be reclassified as later heterotypic synonyms of B. velezensis.

Acknowledgements

This study was supported by the ‘Research Program for Agricultural Science & Technology Development (project no. PJ011248)’ of the National Institute of Agricultural Science, Rural Development Administration, Republic of Korea. The authors would like to thank Heather Walker of the National Center for Agricultural Utilization Research, Peoria, IL, USA for expert technical assistance. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the US Department of Agriculture. The mention of firm names or trade products does not imply that they are endorsed or recommended by the USDA over other firms or similar products not mentioned. USDA is an equal opportunity provider and employer.

References


