Phylogenomic analysis shows that ‘Bacillus vanillea’ is a later heterotypic synonym of Bacillus siamensis

Christopher A. Dunlap

Correspondence
Christopher A. Dunlap
chris.dunlap@ars.usda.gov

‘Bacillus vanillea’ XY18 (=CGMCC 8629=NCCB 100507) was isolated from cured vanilla beans and involved in the formation of vanilla aroma compounds. A draft genome of this strain was assembled and yielded a length of 3.71 Mbp with a DNA G+C content of 46.3 mol%. Comparative genomic analysis with its nearest relatives showed only minor differences between this strain and the genome of the Bacillus siamensis KCTC 13613T (=BCC 22614T=KACC 16244T), with a calculated DNA–DNA hybridization (DDH) value of 91.2 % and an average nucleotide identity (ANI) of 98.9 %. This DDH value is well above the recommended 70 % threshold for species delineation, as well as the ANI threshold of 95 %. In addition, the results of morphological, physiological, chemotaxonomic and phylogenetic analyses indicate that the type strains of these two taxa are highly similar with phenotype coherence. A core genome multi-locus sequencing analysis was conducted for the strains and the results show that ‘Bacillus vanillea’ XY18 clusters closely to the type strain of Bacillus siamensis. Therefore, it is proposed that the species ‘Bacillus vanillea’ XY18 (=CGMCC 8629=NCCB 100507) should be reclassified as a later heterotypic synonym of Bacillus siamensis KCTC 13613T (=BCC 22614T=KACC 16244T). An emended description of Bacillus siamensis is provided.

‘Bacillus vanillea’ XY18 was obtained directly from the laboratory of Dr F. Gu, Spice and Beverage Research Institute, CATAS, Wanning, China (Chen et al., 2015). B. siamensis KCTC 13613T was obtained from the Korean Agricultural Culture Collection as B. siamensis KACC 16244T. ‘B. vanillea’ XY18 and B. siamensis KACC 16244T were compared using identical conditions. Temperature growth studies were conducted from 4–60 °C on tryptone-glucose-yeast extract agar and evaluated at 48 h. The NaCl tolerance range was investigated at 0–20 % (w/v) in 2 % increments in Luria–Bertani (LB) media at 30 °C and evaluated at 48 h. Acid production from specific carbohydrates was tested using phenol red broth (Remel) with 2 % (w/v) carbohydrate (trehalose, melibiose, lactose) at 30 °C and evaluated at 48 h. All assays were conducted in triplicate. ‘B. vanillea’ XY18 and B. siamensis KACC 16244T were cultured overnight on Biolog universal growth plates and prepared according to the manufacturer’s instructions for the GEN III MicroPlate test panel using protocol A (Biolog); the experiment was run in triplicate. ‘B. vanillea’ XY18 was sequenced using an Ion torrent personal genome machine with a Ion PGM 400 bp Hi-Q sequencing kit following the manufacturer’s suggested protocols. The resulting reads were quality trimmed to the Q20 confidence level. The draft genome was assembled using CLCbio Genomics Workbench 7.1 (Qiagen) with default

Abbreviations: ANI, average nucleotide identity; DDH, DNA–DNA hybridization; GGDC, Genome-to-Genome Distance Calculator

The GenBank/EMBL/DDBJ accession number for the draft genome of ‘Bacillus vanillea’ XY18 is LAGT00000000.

A supplementary figure is available with the online Supplementary Material.
A draft genome of *B. vanillae* XY18 was assembled and yielded 63 contigs with a total length of 3,714,066 bp at 52× coverage and 46.3 mol% G+C content. The draft genome of *B. vanillae* was used to evaluate its similarity to *B. siamensis* using an in silico DDH determination and ANI analysis. The results show the strains are highly similar with a DDH value of 91.2% (Table 1), which is well above the threshold of 70% for species delineation (Wayne et al., 1987). The ANI results show the strains are highly similar with an ANI value of 98.9% (Table 1). The recommended cut-off point of 70% DDH for species delineation corresponds to approximately 95% ANI (Goris et al., 2007). The comparisons of *B. vanillae* and *B. siamensis* with other closely related taxa (*B. amyloliquefaciens* and *Bacillus methylotrophicus*) are comparable (Tables 1), and below the recommended cut-off point for species delineation for both DDH and ANI.

The previously published phenotypic data for *B. vanillae* and *B. siamensis* are consistent with the strains belonging to the same species (Chen et al., 2015; Sumpavapol et al., 2010). The previously published fatty acid methyl ester data are slightly variable, but the primary fatty acid components are the same. The other notable deviations found when comparing the previously published phenotypic, physiological or chemotaxonomic properties of these two taxa are: (1) mol% G+C content of the DNA, (2) acid production from lactose, trehalose and melibiose, and (3) temperature and NaCl growth ranges (Chen et al., 2015; Sumpavapol et al., 2010).

The DNA G+C content for *B. siamensis* was originally reported at 41.4 mol% (Sumpavapol et al., 2010) and 46.4 mol% for *B. vanillae* (Chen et al., 2015) using wet lab methods for both. While the sequenced genome data provide a DNA G+C content of 46.3 mol% for *B. siamensis* (Jeong et al., 2012) and 46.3 mol% for *B. vanillae* in the current study. The discrepancy between the original data and the whole genome data for *B. siamensis* suggests the conventional wet lab method has some higher variability. This is supported by a recent study comparing data from wet lab methods and genome-sequencing studies, which shows the DNA G+C content varies 3–5% within species for

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genome comparison (%) with:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>B. siamensis</em> KCTC 13613T</td>
<td></td>
<td>--</td>
<td>91.2</td>
<td>56.5</td>
<td>54.7</td>
</tr>
<tr>
<td>2. ‘B. vanillae’ XY18</td>
<td></td>
<td>98.9</td>
<td></td>
<td>56.4</td>
<td>55.0</td>
</tr>
<tr>
<td>3. <em>B. methylotrophicus</em> KACC 13105T</td>
<td></td>
<td>94.6</td>
<td>93.7</td>
<td></td>
<td>55.2</td>
</tr>
<tr>
<td>4. <em>B. amyloliquefaciens</em> DSM 7T</td>
<td></td>
<td>94.3</td>
<td>93.4</td>
<td>94.3</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Genome comparisons of strains of taxa closely related to *B. siamensis*

Numbers in bold type are calculated DDH values; non-bold are calculated ANI values. DDH and ANI values are reported as (%) of the core genomes between strains.

The previously published phenotypic data for *B. vanillae* and *B. siamensis* are consistent with the strains belonging to the same species (Chen et al., 2015; Sumpavapol et al., 2010). The previously published fatty acid methyl ester data are slightly variable, but the primary fatty acid components are the same. The other notable deviations found when comparing the previously published phenotypic, physiological or chemotaxonomic properties of these two taxa are: (1) mol% G+C content of the DNA, (2) acid production from lactose, trehalose and melibiose, and (3) temperature and NaCl growth ranges (Chen et al., 2015; Sumpavapol et al., 2010).

The DNA G+C content for *B. siamensis* was originally reported at 41.4 mol% (Sumpavapol et al., 2010) and 46.4 mol% for *B. vanillae* (Chen et al., 2015) using wet lab methods for both. While the sequenced genome data provide a DNA G+C content of 46.3 mol% for *B. siamensis* (Jeong et al., 2012) and 46.3 mol% for *B. vanillae* in the current study. The discrepancy between the original data and the whole genome data for *B. siamensis* suggests the conventional wet lab method has some higher variability. This is supported by a recent study comparing data from wet lab methods and genome-sequencing studies, which shows the DNA G+C content varies 3–5% within species for

**Table 2. Characteristics that were previously reported to differentiate *B. vanillae* and *B. siamensis***

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>‘<em>B. vanillae</em>’ XY18</th>
<th>This study</th>
<th>‘<em>B. siamensis</em>’ KACC 16244T</th>
<th>This study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chen et al. (2015)</td>
<td>This study</td>
<td>Sumpavapol et al. (2010)</td>
<td>This study</td>
</tr>
<tr>
<td>Temperature range (°C)</td>
<td>20–45*</td>
<td>4–55</td>
<td>4–55</td>
<td>4–55</td>
</tr>
<tr>
<td>NaCl range for growth (% w/v)</td>
<td>0–8</td>
<td>0–10</td>
<td>0–14</td>
<td>0–10</td>
</tr>
<tr>
<td>Acid production from:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lactose</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Melibiose</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Trehalose</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Only tested 20–50 °C.
conventional methods and within 1 % for whole-genome sequencing data (Meier-Kolthoff et al., 2014).

In the original reports, the strains showed three distinct differences in carbohydrates producing acid (Table 2). In the current study, neither strain produced acid from lactose, melibiose or trehalose. In a second assay, Biolog GEN III analysis of the two strains showed no notable differences. The temperature range for growth and NaCl tolerance were found to be slightly different from those reported in the original descriptions. The nature of these discrepancies is unclear, other than the two strains were never tested concurrently under identical conditions.

The phylogenomic tree based on the core genome (889 genes) of the strains showed that ‘B. vanillea’ XY18 and B. siamensis KACC 16244T (=KCTC 13613T) were closely related (Fig. 1). A previous phylogenetic analysis of ‘B. vanillea’ XY18 based only on the 16S rRNA gene sequence showed much more divergence between it and B. siamensis KACC 16244T (Chen et al., 2015). To understand the source of the difference, we aligned the 16S rRNA gene sequences of ‘B. vanillea’ XY18 and B. siamensis KACC 16244T from their original descriptions with the 16S rRNA gene sequences extracted from whole genome data, and 11 Bacillus subtilis group strains. The 16S rRNA gene sequences extracted from the whole genomes was 100 % identical (Fig. S1, available in the online Supplementary Material). However the 16S rRNA gene data provided in the original B. siamensis description contained three mismatches and an insertion relative to the genome data (Chen et al., 2015) (Fig. S1). The nature of these discrepancies is unclear.

The valid publication of B. siamensis predates the publication of ‘B. vanillea’, therefore our results call for the dissolution of the latter. Thus, we propose that the species ‘Bacillus vanillea’ should be reclassified as a later heterotypic synonym of Bacillus siamensis. Based on data obtained in this study, an emended description of Bacillus siamensis is also provided.

Emended description of Bacillus siamensis Sumpavapol et al. 2010

The description is the same as given by Sumpavapol et al. (2010), except for the following trait. The DNA G+C content of the type strain, B. siamensis KACC 16244T (=BCC 22614T=KCTC 13613T) is 46.3 mol%.

Acknowledgements

The authors would like to thank Heather Walker 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


