Genome-based reclassification of *Bacillus cibi* as a later heterotypic synonym of *Bacillus indicus* and emended description of *Bacillus indicus*

Samantha J. Stropko,† Shannon E. Pipes† and Jeffrey D. Newman

Biology Department, Lycoming College, Williamsport, PA 17701, USA

While characterizing a related strain, it was noted that there was little difference between the 16S rRNA gene sequences of *Bacillus indicus* LMG 22858\(^\text{T}\) and *Bacillus cibi* DSM 16189\(^\text{T}\). Phenotypic characterization revealed differences only in the utilization of mannose and galactose and slight variation in pigmentation. Whole genome shotgun sequencing and comparative genomics were used to calculate established phylogenomic metrics and explain phenotypic differences. The full, genome-derived 16S rRNA gene sequences were 99.74 % similar. The average nucleotide identity (ANI) of the two strains was 98.0 %, the average amino acid identity (AAI) was 98.3 %, and the estimated DNA–DNA hybridization determined by the genome–genome distance calculator was 80.3 %. These values are higher than the species thresholds for these metrics, which are 95 %, 95 % and 70 %, respectively, suggesting that these two strains should be classified as members of the same species. We propose reclassification of *Bacillus cibi* as a later heterotypic synonym of *Bacillus indicus* and an emended description of *Bacillus indicus*.

*Bacillus indicus* Sd/3\(^\text{T}\) was isolated from arsenic-contaminated aquifer sand and described by Suresh *et al.* (2004). *Bacillus cibi* JG-30\(^\text{T}\) was isolated from the Korean fermented seafood dish jeotgal, and a description was published by Yoon *et al.* (2005). While Yoon *et al.* (2005) used information about reference strains and cited Suresh *et al.* (2004), a comparison with the *B. indicus* Sd/3\(^\text{T}\) 16S rRNA gene sequence was not reported, and *B. indicus* Sd/3\(^\text{T}\) was not used as a reference strain despite tight clustering of the original DNA sequences on the neighbour-joining tree (Fig. 1).

While investigating unknown environmental organisms isolated from a freshwater creek in the undergraduate microbiology course at Lycoming College (Newman, 2000; Strahan *et al.*, 2011; Kirk *et al.*, 2013), strain SJS was found to be closely related to *B. indicus* and *B. cibi*. The type strains *B. indicus* LMG 22858\(^\text{T}\) and *B. cibi* DSM 16189\(^\text{T}\) were obtained from BCCM and DSMZ, respectively, characterized phenotypically and their genomes were sequenced.

Genomic DNA was isolated from *B. indicus* LMG 22858\(^\text{T}\) and *B. cibi* DSM 16189\(^\text{T}\) using the Qiagen DNeasy Blood and Tissue kit according to the manufacturer’s instructions for Gram-positive bacteria. Libraries were prepared and then sequenced on an Illumina MiSeq (V3 2 × 300 base) by the Indiana University Center for Genome Studies as part of a Genome Consortium for Active Teaching NextGen Sequencing Group (GCAT-SEEK) shared run (Buonaccorsi *et al.*, 2011, 2014). Sequencing reads were filtered (median phred score ≥20), trimmed (phred score ≥16), and assembled using the paired-end *de novo* assembly option in NextGENe V2.3.4.2 (SoftGenetics). The assembled genomes were uploaded to the Rapid Annotation with Subsystem Technology web service (Aziz *et al.*, 2008; Overbeek *et al.*, 2014) for analysis, guided contig reordering and assembly improvement.

Approximately 0.9 million reads from the *B. indicus* LMG 22858\(^\text{T}\) genome were assembled into 22 contigs with a total genome length of 4 129 127 bp, a mean coverage of 52 × and a DNA G+C content of 44.4 mol%, significantly higher than the 41.2 mol% reported by Suresh *et al.* (2004). Automated annotation identified 4285 potential protein-coding sequences and 83 RNAs.

Approximately 1.4 million *B. cibi* DSM 16189\(^\text{T}\) reads were assembled into 24 contigs with a total genome length of 4 072 175 bp, a mean coverage of 71 × and a DNA G+C content of 44.4 mol%, in good agreement with the 45 mol% reported by Yoon *et al.* (2005). Automated annotation identified 4253 potential protein-coding sequences and 85 RNAs. Contigs were ordered and oriented to

†These authors contributed equally to this work.

The GenBank/EMBL/DDBJ accession numbers for the Whole Genome Shotgun projects of *Bacillus indicus* LMG 22858\(^\text{T}\) and *Bacillus cibi* DSM 16189\(^\text{T}\) are JGVU00000000 and JNVC00000000, respectively. The versions described in this paper are versions JGVU02000000 and JNVC02000000.

Two supplementary figures and two supplementary tables are available with the online Supplementary Material.
maximize synteny of the 3964 shared coding sequences (Fig. S1, available in the online Supplementary Material).

The 16S rRNA gene sequences obtained from the sequenced genomes were compared to the partial sequences reported at the time of initial publication using the pairwise alignment (Myers & Miller, 1988) feature implemented on the EzTaxon-e web-based service (Kim et al., 2012). The Bacillus indicus LMG 22858T genome-derived 16S rRNA gene sequence was identical to the Bacillus indicus Sd/3T 16S rRNA gene partial sequence (AJ583158) (Suresh et al., 2004) except for two differences and an insertion within the first 14 bases of the partial rRNA gene sequence, which apparently did not have the amplification primer sequences trimmed. Because the genome-derived sequence is present within a large contig as part of a complete rRNA operon with the 23S and 5S rRNA genes, it is very likely to be the correct sequence. There was one difference between the B. cibi DSM 16189T genome-derived 16S rRNA gene sequence and the B. cibi JG-30T partial sequence (AY550276) (Yoon et al., 2005). The two genome-derived sequences differed at four of 1533 bases giving a similarity of 99.74\%. The four differences corresponded to two pairs of substitutions that maintained base-pairing within stems.

Several phylogenomic metrics have been developed to measure genome similarity for taxonomic determinations. The traditional metric of DNA–DNA hybridization (DDH) (Tindall et al., 2010) has been criticized due to technical difficulty, lack of reproducibility, and inability to incorporate into databases. Meier-Kolthoff et al. (2013) developed the Genome–Genome Distance Calculator to estimate the DNA–DNA hybridization value based on high-scoring segment pairs from genome sequence comparisons. The estimated DDH between the B. cibi DSM 16189T and Bacillus indicus LMG 22858T genomes was 80.3 \% (95 \% confidence interval 78.3–82.3 \%), which exceeds the 70 \% DDH species boundary. Average nucleotide identity (ANI) reflects the similarity of 1 kb sequence fragments using the algorithm described by Goris et al., (2007). The implementation on the ExGenome Web service (http://www.ezbiocloud.net/ezgenome/ani) (Kim et al., 2012) yielded a value of 97.8 \%, while the Kostas Lab implementation (http://enve-omics.ce.gatech.
Average amino acid identity (AAI) reflects the similarity of orthologous proteins. Because amino acid sequences change more slowly than nucleotide sequences and have greater complexity, AAI values are more sensitive over greater evolutionary distances, yet can still be used to distinguish organisms at the species level (Konstantinidis & Tiedje, 2005b). Based on the bidirectional best hits (3964 coding sequences) identified by the sequence-based comparison tool of the SEED viewer (Overbeek et al., 2005), AAI was calculated as 98.33% with the web-based NewmanLab AAI calculator (http://lycofs01.lycoming.edu/~newman/aai/). Again, this is greater than the 95% threshold that correlates to separate species.

While examining the utilization of carbon sources on the Biolog GenIII plates, both organisms yielded a strong positive response in the negative control well when using Biolog Inoculating Fluid A, suggesting that they contained stored carbon sources. Consistent with this hypothesis, both organisms contained a glycogen synthesis and utilization operon. A weaker response in the negative control well was observed with inoculating fluid B, which allowed the organisms to obtain carbon from the complex extracellular molecules. Aoki et al. (2013) noted that B. indicus was originally reported as negative for motility and chemotaxis; however, B. indicus was originally described as non-motile (Suresh et al., 2004). Both organisms grew with 4%, 6% and 8% (w/v) NaCl on TSA plates and with 4% and 8% (w/v) NaCl in Biolog GenIII plates, although B. indicus was originally reported as negative for growth under these moderately high NaCl concentrations. Growth with 8% NaCl and motility were also observed for both strains by Khaneja et al. (2010). Neither organism grew at 44°C, in contrast to the positive growth result reported by Yoon et al. (2005) at this temperature. Khaneja et al. (2010) also observed that both strains showed resistance to arsenic and arsenate, although B. cibi DSM 16189T exhibited slightly lower maximum tolerated concentrations.

Table 1. Fatty acid contents (percentages) of B. indicus LMG 22858T and B. cibi DSM 16189T grown on TSBA at 30°C for 24 h

<table>
<thead>
<tr>
<th>Fatty Acid</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>iso-C14:0</td>
<td>7.2</td>
<td>5.9</td>
</tr>
<tr>
<td>C14:0</td>
<td>2.8</td>
<td>2.1</td>
</tr>
<tr>
<td>iso-C15:0</td>
<td>46.4</td>
<td>48.9</td>
</tr>
<tr>
<td>anteiso-C15:0</td>
<td>15.7</td>
<td>14.9</td>
</tr>
<tr>
<td>C16:1ω7c alcohol</td>
<td>4.6</td>
<td>4.9</td>
</tr>
<tr>
<td>iso-C16:0</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>C16:1ω11c</td>
<td>5.6</td>
<td>5.7</td>
</tr>
<tr>
<td>C16:0</td>
<td>4.4</td>
<td>3.2</td>
</tr>
<tr>
<td>iso-C17:ω10c</td>
<td>2.5</td>
<td>3.3</td>
</tr>
<tr>
<td>Summed feature 4*</td>
<td>2.0</td>
<td>2.2</td>
</tr>
<tr>
<td>iso-C17:0</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td>anteiso-C17:0</td>
<td>2.3</td>
<td>2.4</td>
</tr>
</tbody>
</table>

While colonies of B. indicus LMG 22858T and B. cibi DSM 16189T grown on TSBA at 30°C for 24 h appear similar, the pigmentation of B. cibi DSM 16189T colonies is darker by 72 h (Fig S2a). To extract the pigments, 5–10 mg of cell mass was scraped from a plate, spread around the inside of a 1.5 ml microcentrifuge tube and 1 ml was acetone added. The tube was then incubated in a shaker-incubator at 250 r.p.m. and 28°C for 1 h. The acetone extract was decanted into a 2 ml sample vial and analysed on an Agilent 1200 HPLC with an Agilent Zorbax Eclipse XDB C18 column (4.6 mm × 250 mm) at 40°C and a diode array detector. Separation was achieved using 50 mM NaH2PO4 (pH 4.5) as solvent A; methanol containing 0.1% (v/v) glacial acetic acid as solvent B; a flow rate of 1 ml min−1; and USA. On: Sat, 22 Dec 2018 07:46:12
the following gradient: 0–5 min at 0%; increase to 75% solvent B at 10 min; increase to 100% solvent B at 30 min; remain at 100% solvent B until 45 min. The elution profiles revealed a variety of compounds with different spectra that absorbed light in the visible range (Fig. S2b, c). The optimal wavelength to detect all of the peaks was 452 nm (Fig. 2a). There were three clusters of peaks that could be distinguished based on absorbance spectrum (Fig. 2b); the 29–32 min peaks had spectra consistent with 4,4'-diaponeurosporene, the 36–40 min peaks had spectra consistent with 4,4'-diaponeurosporenoate, and the 33–36 min peaks had spectra consistent with staphyloxanthin. Analysis of the two strains’ genomes revealed several operons with homologues of the Staphylococcus aureus crtOPQMN genes that are responsible for the synthesis of the golden pigment staphyloxanthin (Pelz et al., 2005). Some of the different peaks in each region are very likely to be different geometric isomers (Melendez-Martinez et al., 2013). Both organisms contain all of the genes necessary to produce staphyloxanthin, but B. indicus LMG 22858T contains less of this pigment (Fig. 2b), which is apparent in the more yellow, less golden colour (Fig. S2a).

On the basis of the minor phenotypic and genomic differences, it is proposed that B. cibi Yoon et al. 2005 should be considered as a later heterotypic synonym of B. indicus Suresh et al. 2004.

**Emended description of Bacillus indicus Suresh et al. 2004**

The description is the same as that given by Suresh et al. (2004) except for the following traits. Cells are motile. Growth is observed on R2A and tryptic soy agars, but not on phenylethanol or mannitol-salt agars. Yellow–golden, non-diffusible, carotenoid-type pigments are produced that are consistent with staphyloxanthin and biosynthetic precursors. The pH range for growth is pH 6–9, and cells grow with up to 8% (w/v) NaCl. Cells exhibit oxidase,

![Figure 2](http://ijs.sgmjournals.org)
catalase, amylase and caseinase activities, but do not liquefy gelatin. On the API ZYM panel, positive for C4 esterase, C8 esterase lipase, leucine arylamidase (weakly positive), z-chymotrypsin, acid phosphatase (weakly positive), naphthol-AS-BI-phosphohydrolase, β-galactosidase and α-glucosidase, and negative for alkaline phosphatase, C14 lipase, valine and cystine arylamidases, trypsin, α-galactosidase, β-gluco- nidiase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase; β-glucosidase activity varies. When using inoculating fluid B and the Biolog GenIII plate (27 °C), positive for utilization of dextrin, maltose, trehalose, cellobiose, gentiobiose, sucrose, turanose, stachyose, raffinose, melibiose, methyl β-D-glucoside, D-salicin, N-acetyl-D-glucosamine, α-D-glucose, D-fructose, glycerol, gelatin, glycyrl-L-proline, L-alanine, L-aspartic acid, L-glutamic acid, L-histidine, L-prolylglutamic acid, L-serine, pectin, D-gluconic acid, D-glucuronic acid, methyl pyruvate, D-lactic acid ester methyl, L-lactic acid, α-ketoglutaric acid, L-malic acid, Tween-40, hydroxy-β-DL-Butyric acid, acetoacetic acid and acetic acid; utilization of D-mannose and galactose varies; negative for the remaining utilization tests. Also on the GenIII plate, tolerant of pH 6, NaCl at 1, 4 and 8 % (w/v), 1 % (w/v) sodium lactate, LiCl, K-tellurite, aztreonam and sodium butyrate; sensitive to the remaining inhibitory conditions.

The type strain is SD/3T (=MTCC 4374T=DSM 15820T=LMG 22858T), the former type strain of B. cibi, IG-30T (=KCTC 3880T=DSM 16189T), is a second strain of B. indicus. The DNA G+C content of strain DSM 16189T LMG 22858T is 44.4 mol%.

Acknowledgements

This work was funded by a Lycoming College Professional Development Grant to J.D.N and a Summer Student Research Grant to S.J.S. GCAT-SEEK has been supported by US National Science Foundation award DBI-1248096: RCN-UBE - GCAT-SEEK: The Genome Consortium for Active Undergraduate Research and Teaching Using Next-Generation Sequencing.

References


