Description of a new member of the family *Erysipelotrichaceae*: *Dakotella fusiforme gen. nov.,* sp. nov., isolated from healthy human feces

Sudeep Ghimire [1], **Supapit Wongkuna** [1], **Joy Scaria** [Corresp. 1]

[1] Department of Veterinary and Biomedical Sciences, South Dakota State University, Brookings, South Dakota, United States

Corresponding Author: Joy Scaria
Email address: joy.scaria@sdstate.edu

A Gram-positive, non-motile, rod-shaped facultative anaerobic bacterial strain SG502^T was isolated from healthy human fecal samples in Brookings, SD, USA. The comparison of the 16S rRNA gene placed the strain within the family *Erysipelotrichaceae*. Within this family, *Clostridium innocuum* ATCC 14501^T, *Longicatena caecimuris* strain PG-426-CC-2, *Eubacterium dolichum* DSM 3991^T and *E. tortuosum* DSM 3987^T (=ATCC 25548^T) were its closest taxa with 95.28%, 94.17%, 93.25%, and 92.75% 16S rRNA sequence identities respectively. The strain SG502^T placed itself close to *C. innocuum* in the 16S rRNA phylogeny. The members of genus *Clostridium* within family *Erysipelotrichaceae* was proposed to be reassigned to genus *Erysipelatoclostridium* to resolve the misclassification of genus *Clostridium*. Therefore, *C. innocuum* was also classified into this genus temporarily with the need to reclassify it in the future because of its difference in genomic properties. Similarly, genome sequencing of the strain and comparison with its 16S phylogenetic members and proposed members of the genus *Erysipelatoclostridium*, SG502^T warranted a separate genus even though its 16S rRNA similarity was >95% when comapred to *C. innocuum*. The strain was 71.8% similar at ANI, 19.8% [17.4-22.2%] at dDDH and 69.65% similar at AAI to its closest neighbor *C. innocuum*. The genome size was nearly 2,683,792 bp with 32.88 mol% G+C content, which is about half the size of *C. innocuum* genome and the G+C content revealed 10 mol% difference. Phenotypically, the optimal growth temperature and pH for the strain SG502^T were 37^{0}C and 7.0 respectively. Acetate was the major short-chain fatty acid product of the strain when grown in BHI-M medium. The major cellular fatty acids produced were $C_{18:1}$ ω9c, $C_{18:0}$ and $C_{16:0}$. Thus, based on the polyphasic analysis, for the type strain SG502^T (=DSM 107282^T= CCOS 1889^T), the name *Dakotella fusiforme gen. nov.,* sp. nov., is proposed.

Description of a new member of the family *Erysipelotrichaceae*: *Dakotella fusiforme gen. nov.,* sp. nov., isolated from healthy human feces

Sudeep Ghimire[1], Supapit Wongkuna[1,] and Joy Scaria[1]*

[1]Department of Veterinary and Biomedical Sciences, South Dakota State University, Brookings, SD, USA.

*Address correspondence to: Joy Scaria

Email: joyscaria@gmail.com

ABSTRACT

A Gram-positive, non-motile, rod-shaped facultative anaerobic bacterial strain SG502^T was isolated from healthy human fecal samples in Brookings, SD, USA. The comparison of the 16S rRNA gene placed the strain within the family *Erysipelotrichaceae*. Within this family, *Clostridium innocuum* ATCC 14501^T, *Longicatena caecimuris* strain PG-426-CC-2, *Eubacterium dolichum* DSM 3991^T and *E. tortuosum* DSM 3987^T (=ATCC 25548^T) were its closest taxa with 95.28%, 94.17%, 93.25%, and 92.75% 16S rRNA sequence identities respectively. The strain SG502^T placed itself close to *C. innocuum* in the 16S rRNA phylogeny. The members of genus *Clostridium* within family *Erysipelotrichaceae* was proposed to be reassigned to genus *Erysipelatoclostridium* to resolve the misclassification of genus *Clostridium*. Therefore, *C. innocuum* was also classified into this genus temporarily with the need to reclassify it in the future because of its difference in genomic properties. Similarly, genome sequencing of the strain and comparison with its 16S phylogenetic members and proposed members of the genus *Erysipelatoclostridium*, SG502^T warranted a separate genus even though its 16S rRNA similarity was >95% when compared to *C. innocuum*. The strain was 71.8% similar at ANI, 19.8% [17.4-22.2%] at dDDH and 69.65% similar at AAI to its closest neighbor *C. innocuum*. The draft genome size of SG502^T was 2,683,792 bp with 32.88 mol% G+C content, which is about half the size of *C. innocuum* genome and the G+C content revealed 10 mol% difference. Phenotypically, the optimal growth temperature and pH for the strain SG502^T were 37^{0}C and 7.0 respectively. Acetate was the major short-chain fatty acid product of the strain when grown in BHI-M medium. The major cellular fatty acids produced were $C_{18:1}$ ω9c, $C_{18:0}$ and $C_{16:0}$. Thus,

based on the polyphasic analysis, for the type strain SG502^T (=DSM 107282^T= CCOS 1889^T), the name *Dakotella fusiformе gen. nov.,* sp. nov., is proposed.

Key words: *Dakotella fusiforme*, new genus, new species, gut microbiota, taxonomy, microbiome, culturomics

1. INTRODUCTION

The members of family *Erysipelotrichaceae* have been isolated from the intestinal tracts of mammals (Alcaide et al. 2012; Greiner & Backhed 2011; Han et al. 2011) and insects (Egert et al. 2003) and are associated with host metabolism and inflammatory diseases (Cox et al. 2014; Kaakoush 2015). Although metagenome analysis of gut microbiome have revealed the composition and function of the microbiome in the intestine, only a few cultured species available from this family and their function in the gut ecosystem is not yet well understood (Kaakoush 2015).

The family *Erysipelotrichaceae* was originally described by Verbarg et al (Verbarg et al. 2004). The members of this family includes Gram-positive, filamentous rods and were originally described as facultative anaerobes but later amended by Tegtmeier *et al.* (Tegtmeier et al. 2016) to include obligate anaerobes. The members of this family belonged to *Clostridial* Cluster XVI that consists of three major species; *Clostridium innocuum, Eubacterium biforme* and *Streptococcus pleomorphus* (Collins et al. 1994). According to the updated LPSN list of valid bacteria (http://www.bacterio.net), current members of this family includes the following genera: *Allobaculum, Breznakia, Bulleidia, Catenibacterium, Canteisphaera, Coprobacillus, Dielma, Dubosiella, Eggerthia, Erysipelothrix, Faecalibaculum, Faecalicoccus, Faecalitalea, Holdemania, Kandleria, Longibacculum, Longicatena, Solobacterium,* and *Turicibacter.*

Members of this family have low-G+C content and were previously recognized as the "walled relatives" of mycoplasma (Weisburg et al. 1989) and later classified under *Clostridial* cluster XVI (Collins et al. 1994). With major changes in the taxonomy of *Erysipelotrichaceae*, recently some members have been reclassified into new families *Coprobacillaceae* (Collins et al. 1994) and *Turicibacteraceae* (Verbarg et al. 2014) but the placement of cluster XVI is still debated. A few species related to *Clostridium* and *Eubacterium* are also included within *Erysipelotrichaceae* based on the 16S rRNA gene sequence similarity and are considered as misclassified (Verbarg et al. 2014). For the proper classification of these misclassified members of genus *Clostridium*, different genus names were proposed by Yutin and Galperin in 2013 where *C. innocuum* along with *C. cocleatum, C. saccharogumia* and *C. spiroforme* were proposed to be reassigned to a new genus *Erysipelatoclostridium.* As, *C. innocuum* was found to be more distantly related to other members inside *Erysipelatoclostridum* genus, they also highlighted the need of future reclassification of *C. innocuum* (Yutin & Galperin 2013).

The present study describes the isolation, physiology, and genomic characterization of a new member of the family *Erysipelotrichaceae* isolated from healthy human feces. Within the family *Erysipelotrichaceae*, strain SG502^T clustered within clostridial cluster XVI. Also, the strain SG502^T showed less than 97% 16S rRNA gene sequence similarity towards its nearest phylogenetic neighbor *C. innocuum* ATCC 14501^T. Therefore, we performed *in vitro* phenotypic characterization and sequenced the genome for comparative analysis with its neighbors. We found major differences in the genomic and phenotypic characteristics of the strain even though its 16S rRNA gene was >95% similar to nearest neighbor *C. innocuum*. Thus, we propose a novel genus

and species for this strain and propose designating SG502^T as *Dakotella fusiforme* gen. nov., sp. nov within the family *Erysipelotrichaceae*.

2. MATERIALS AND METHODS

2.1 Bacterial isolation and culture condition

The strain was isolated from healthy human fecal sample as part of a culturomics study. The collection of the human fecal samples were done with the approval of the Institutional review board (approval #IRB-1709018-EXP) at South Dakota State University, Brookings SD. After transferring the fresh fecal samples into the anaerobic chamber (85% nitrogen, 10% hydrogen and 5% carbon dioxide) within 10 minutes of voiding, the sample was diluted 10 times with anaerobic PBS and stored with 18% DMSO in -80^{0}C. The sample was cultured in modified BHI medium (BHI-M) containing 37g/L of BHI, 5 g/L of yeast extract, 1 ml of 1 mg/mL menadione, 0.3 g of L-cysteine, 1 mL of 0.25 mg/L of resazurin, 1 mL of 0.5 mg/mL hemin, 10 mL of vitamin and mineral mixture,1.7 mL of 30 mM acetic acid, 2 mL of 8 mM propionic acid, 2 mL of 4 mM butyric acid, 100 µl of 1 mM isovaleric acid, and 1% pectin and inulin. After isolation, the strain was subjected to MALDI-ToF (Bruker, Germany). Since MALDI-ToF did not identify a species, 16S rRNA gene sequencing was performed for species identification.

2.2 Phenotypic and chemotaxonomic characterization

For morphological, physiological and biochemical characterization, the strain was cultivated in BHI-M medium in anaerobic conditions at 37^{0}C at pH 6.8±0.2. Colony characteristics were determined after streaking the strain on BHI-M agar plates followed by 48 hours of anaerobic

incubation. Gram staining was performed using a Gram staining kit (BD Difco) according to the manufacturer's protocol. During the exponential growth of the bacterium, cell morphology and flagellation was examined under scanning electron microscopy (SEM). SG502^T was grown separately in aerobic and anaerobic conditions to determine the aerotolerance. Further, the strain was grown at 4, 20, 30, 40 and 55^{0}C to determine the range of growth under anaerobic conditions. The BHI-M media was adjusted to pH levels between 4 and 9 with 0.1N HCl and 0.1N NaOH to determine the growth of the strain at different pH levels. BHI-M medium was supplemented with triphenyltetrazolium chloride (TTC) (Shields & Cathcart 2011

) to determine the motility of the strain.

The phenotypic and biochemical characterizations were performed using AN MicroPlate (Biolog) and API ZYM (bioMerieux) according to the manufacturer's instructions. Also, after growing the strain SG502^T and ATCC 14501^T in BHI-M medium at 37^{0}C for 24 hours, cells were harvested for cellular fatty acid analysis. Fatty acids were extracted, purified, methylated, identified and analyzed using GC (Agilent 7890A) according to manufacturer's instructions (MIDI) (Sasser 1990). Further, short-chain fatty acid (SCFA) production was determined using gas chromatography after cells were grown in BHI-M medium. For SCFA estimation, 800 µl of the bacterial culture was collected and 160 µl of freshly prepared 25% meta-phosphoric acid (w/v) was added before freezing to -80^{0}C. The sample were thawed and centrifuged at >20,000×g for 30 min before injecting 600 µl of the supernatant into the TRACE1310 GC system (ThermoScientific, USA).

2.3 Phylogenetic analysis

Genomics DNA from the strain was isolated using E.Z.N.A bacterial DNA isolation kit (Omega Biotek) following the manufacturer's instructions. The 16S rRNA gene was amplified using universal primer set 27F (5'- AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'- ACCTTGTTACGACTT- 3') and sequenced using a Sanger sequencing chemistry (ABI 3730XL; Applied Biosystems). The sequences were assembled using Genious 10.2.3. The nearly complete 16S rRNA gene sequence obtained was used for a similarity search in EzTaxon-e program (http://www.ezbiocloud.net/) for the valid taxonomic names. The bacterial species that closely resembled the query sequences were then used for alignment and phylogenetic analysis in MEGAX software (Kumar et al. 2016). Initially, the sequences were aligned using MUSCLE (Edgar 2004) and the Neighbor Joining method (Saitou & Nei 1987) was used to reconstruct the phylogenetic tree employing Kimura 2-parameter model (Kimura 1980) with 1000 bootstraps. Phylogenetic trees were also constructed using maximum-likelihood (Felsenstein 1981)and minimum evolution methods (Rzhetsky. A. & M. 1992). *Clostridium butyricum* ATCC 19398^T was used as an out-group.

2.4 Genomic features and comparison

For the whole genome sequencing of SG502^T, we used 0.3ng of the genomic DNA for library preparation. Library was sequenced on an Illumina MiSeq using 2x 250 paired-end V2 chemistry. Genome was assembled from raw fastq files using Unicycler which builds an initial assembly graph from short reads using the de novo assembler SPAdes3.11.1 (Bankevich et al. 2012). Quality assessment for the assemblies was performed using QUAST (Gurevich et al. 2013). Genome

159 annotation was performed using Prokka 1.13 (Seemann 2014). The genome of SG502^T was
160 visualized using DNAplotter (Carver et al. 2009).

161

162 We compared the genome of SG502^T to that of 16S-phylogeny of closely related species *C.*
163 *innocuum* DSM 1286^T, *Longicatena caecimuris* DSM 29481^T, *Eubacterium dolichum* DSM
164 3991^T, *Faecalicoccus pleomorphus* DSM 20574^T, *Faecalitalea cylindroides* ATCC 27803^T,
165 *Holdemanella biformis* DSM 3989^T and *Dielma fastidiosa* DSM 26099^T. In addition, we
166 compared SG502^T with the proposed members of the genus *Erysipelatoclostridium, C. clocleatum*
167 DSM 1551^T, *C. ramosum* DSM 1402^T, *C. saccharogumia* DSM 17460^T and *C. spiroforme* DSM
168 1552^T. We used the average nucleotide identity (OrthoANI) (Lee et al. 2016) and digital DDH
169 (Meier-Kolthoff & Goker 2019) for calculating phylogenomic similarity. We also performed
170 average amino acid identity (AAI) using AAI calculator (Konstantinidis & Tiedje 2005) to
171 determine protein level genome differences.

172

173 **3. RESULTS**

174 The SG502^T strain was isolated from the healthy human fecal sample during the culturomics study
175 of the human gut microbiota. The colonies of the strain appeared white, smooth and convex with
176 entire edges. The cells were initially subjected to MALDI-ToF MS (Figure 1A) which revealed
177 the score <1.70 suggesting no identification. Thus, further phenotypic characterization and genetic
178 based methods were employed for identification of the strain.

179

Morphologically, individual cells of the strain appeared to be gram-positive rods. The cell was observed to be slender with tapering ends with 1.5×0.35μm in dimensions (Figure 1B and Table 1) under SEM. No flagella were observed under SEM suggesting its non-motile nature which was also validated by TTC assay. The strain also lacked endospores, similar to what has been previously reported for the members of *Erysipelotrichaceae* (Verbarg et al. 2004). The strain grew in a pH range of 6.0-7.5 with optimal growth at pH 7.0. It could grow anaerobically over the temperature range of 25-45^{0}C with optimal growth at 37^{0}C. The strain grew well in BHI-M under anaerobic conditions but under aerobic conditions, the growth was comparatively lower and slow confirming that the strain was a facultative anaerobe. Based on the results obtained from a carbon source utilization test (Biolog AN plate), the strain utilizes glucose, sorbitol, maltose, arbutin, D-fructose, L-fucose, palatinose, dextrin, turanose, D-trehalose, L-rhamnose, uridine, pyruvic acid methyl ester, pyruvic acid, 3-methyl-D-glucose, gentiobiose, maltotriose, ducitol, L-phenylalanine, α-ketovaleric acid, N-acetyl-D-glucosamine, N-acetyl-β-D-mannosamine, cellobiose, α-ketobutyric acid, D-galacturonic acid and N-acetyl-D-glucosamine. Also, SG502^T assimilated sorbitol and maltose which were not utilized by its closest neighbor *C. innocuum* ATCC 14501^T. Furthermore, SG502^T was unable to utilize sucrose, salicin, mannitol, lactose, and raffinose when compared to *C. innocuum*. Positive enzymatic activities for leucine arylamidase, cystine arylamidase, α-chymotripsin and acid phosphates were observed for *C. innocuum* differentiating it from SG502^T. Detailed phenotypic and biochemical characteristics of the strain are presented in Table 1. Also, the major fatty acids content identified were $C_{18:1}$ ω9c (29.82%), $C_{18:0}$ (22.55%) and C16:0 (14.7%) compared to *C. innocuum* ATCC 14501^T with $C_{18:1}$ ω9c (14.64%), $C_{18:0}$ (10.56%) and C16:0 (23.7%) (Table 2). The detailed comparison of the fatty acids in SG502^T along with *C. innocuum* 14501^T and *E. dolichum* DSM 3991^T is given in Table 2.

Additionally, the major SCFAs metabolite identified for SG502^T was acetate in BHI-M medium. Low but detectable amounts of propionate and butyrate were produced by the strain SG502^T. The utilization of such broad substrates and production of SCFAs can be ecologically effective trait against pathogen colonization in the gut.

As the strain was not identified using MALDI-ToF, 16S rRNA sequence was amplified to obtain a continuous stretch of 1338 bp gene which was searched against the Eztaxon 16S rRNA gene database for identification. The closest species identified were all from the *Erysipelotrichaceae* family that included *C. innocuum* ATCC 14501^T, *L. caecimuris* strain PG-426-CC-2, *E. dolichum* DSM 3991^T and *E. tortuosum* ATCC 25548^T with 95.28%, 94.17%, 93.25%, and 92.75% sequence identities respectively. Currently, the cut off for the species and genus level classification of the bacteria based on 16S rRNA gene is <98.7 % (E. & J. 2006) and <94.5% identity (Yarza et al. 2014) respectively. Thus, the strain SG502^T and *C. innocuum* were suggested to fall within same genus but different species. The phylogenetic analysis also revealed that the isolate belonged to *Erysipelotrichaceae* family where the strain SG502^T was closely associated to *C. innoccum* ATCC 14501^T but further from *L. caecimuris* strain PG-426-CC-2*, E. dolichum* DSM 3991^T and *E. tortuosum* ATCC 25548^T which altogether formed a larger clade (Figure 2). The separation of these four species from the strain SG502^T did not depend on the phylogenetic algorithm and was supported by an 100% bootstrap value. To further differentiate the strain, we sequenced the whole genome of the strain and is visualized in Figure 3. The draft genome of the strain SG502^T was 2,683,792 bp with 32.88 mol% G+C content. The largest contig was of 154,144 bp and N_{50} was 52,214. The total number of predicted coding sequences, tRNAs, rRNAs, and tmRNAs was 2654, 49, 2 and 1 respectively.

226

227 *C. innocuum* was the nearest neighbor of SG502^T based on 16S rRNA phylogeny. *C. innocuum*
228 along with *C. cocleatum, C. saccharogumia*, *C. ramosum*, and *C. spiroforme* were suggested to
229 be reclassified previously into genus *Erysipelatoclostridium* with *C. innocuum* needing further
230 reclassification (Yutin & Galperin 2013). Therefore, we checked for the 16S identity of SG502^T
231 with the other members of this proposed genera *C. cocleatum, C. saccharogumia*, *C. ramosum*,
232 and *C. spiroforme* in NCBI. These species were found to be 84.45%, 84.49%, 85.10% and
233 85.014% identical respectively which demonstrated that SG502^T should not be placed into same
234 genera with these species. We also compared the genomic properties of the strain with its 16S
235 rRNA based phylogenetic neighbors along with *C. innocuum* and the members of formerly
236 proposed genus *Erysipelatoclostridium, C. cocleatum, C. saccharogumia*, *C. ramosum*, and *C.*
237 *spiroforme*. The genomic sizes and G+C content of the members of the neighbors of the strain
238 were found to vary as shown in Table 3. *C. innocuum* was 4,772,018 bp in length with 43.4
239 mol% G+C content, while for SG502^T, genome length was 2, 683,792 bp with 32.88 mol%
240 G+C content. The genome sizes and G+C content of the neighboring species were highly
241 variable compared to SG502^T (Table 3). Because of such high differences in the genome
242 properties of SG502^T, we performed further comparison for classifying SG502^T as a novel genus.
243 Hence, the genome of the strain was compared to its neighbors using OrthoANI as shown in
244 Figure 4. The strain SG502^T was 71.8% similar to its nearest neighbor *C. innocuum* and had
245 lower similarities with other neighbors. (Figure 4A, 4B). The proposed cut-off for OrthoANI for
246 a new species is 95-96% (Kim et al. 2014) (Lee et al. 2016). The dDDH was only 19.8%
247 between SG502^T and *C. innocuum* (Table 4). One of the major methods to demarcate the genus
248 is to calculate average amino acid identity (AAI) between the genomes with the possibility of

249 novel genus if the AAI values are in the range of 65-72% (Konstantinidis & Tiedje 2007). The
250 strain SG502^T showed highest AAI with *C. innocuum* (69.65%) followed by *L. caecimuris*
251 (63.45%) and *E. dolichum* (63.02%) (Figure 5) supporting the designation of strain SG502^T in a
252 novel genus.

253

254 **4. DISCUSSION**

255 Recently, next generation sequencing and high-throughput culturing methods has been
256 employed for large scale culture of the unknown gut microbiota. This new approach termed as
257 "culturomics" has evolved as a tool to culture previously uncultured bacteria (Browne et al.
258 2016; Lagier et al. 2016). However, such culture independent studies have also highlighted that
259 the diverse population of gut bacteria are yet to be cultivated (Almeida et al. 2019; Lagier et al.
260 2012). The pure culture of the bacteria is essential to elucidate the role of these organisms in
261 health and diseases for both experimental model and therapeutics purposes (Daillere et al. 2016;
262 Kobyliak et al. 2016; Vetizou et al. 2015). In this study, we report the culturing and
263 characterization of a previously uncultured bacterium SG502^T from the healthy human fecal
264 samples that belongs to a new genus and species. Also, we employed taxono-genomics approach
265 (Fournier & Drancourt 2015) to determine the phenotypic and genetic properties of the taxon.

266

267 16S rRNA based gene sequence homology is the widely used method to determine the novelty of
268 the prokaryotic organism with varying threshold values at distinct taxonomic levels (Clarridge
269 2004; Kim et al. 2014) . Therefore, we performed the 16SrRNA based phylogenetic analysis of
270 the strain SG502^T which showed it as a member of *Erysipelotrichaceae* family. Under this

family, it clustered together with *Clostridium innoccum, Longicatena caecimuris, Eubacterium dolichum* and *Eubacterium tortuosum* with *C. innocuum* as a closest member. *C. innocuum* along with other members of misclassified *Clostridia* under *Erysipelotrichaceae* family were proposed to be reclassified into gen. nov. *Erysipeloclostridium*. The members of this proposed genus *Erysipelatoclostridium* are gram positive, nonmotile, obligately anaerobic straight or helically curved rods which rarely forms spores. The G+C content is lower and varies from 27-33 mol% (Yutin & Galperin 2013). However, *C. innocuum* was identified to be a distantly related member of *Erysipelatoclostridium* with higher G+C content of 43-44% with need of reclassification (Yutin & Galperin 2013). In this context, we also searched for the 16S based identity of the strain SG502^T with the proposed members of genus *Erysipelatoclostridium*. Nevertheless, the proposed members of genus *Erysipelatoclostridium* were < 86% similar at 16S sequence level, suggesting the uniqueness of SG502^T.

Phenotypically, SG502^T revealed several differences in carbon sources utilization, enzymatic activity and fatty acid when compared to its phylogenetic neighbors (Table 1 and Table 2). In addition, whole genome sequence comparison revealed its distinctiveness with respect to 16S phylogenetic members and the proposed members of *Erysipelatoclostridium* genus (Table 3 and 4). Furthermore, OrthoANI based genomic comparison with 16S phylogenetic neighbors showed as high as 71.54% similarity with *L. caecimuris* DSM 29481^T. Also, the genome of SG502^T was only 82.21% similar with *C. cocleatum* DSM 1551^T which is a member of *Erysipelatoclostridium* genus (Figure 4). Major differences were evident in dDDH and amino acid composition comparison as well (Table 4, Figure 5). Finally, the genome size of the nearest neighbor *C. innocuum* was nearly twice that of SG502^T and the difference of G+C content was

294 comparatively high (>10 mol%) suggesting that the strain is not close to *C. innocuum* genetically
295 which means that SG502^T require the placement in a separate genus.

296

297 **5. CONCLUSION**

298 Despite 95.15% 16S rRNA similarity of SG502^T with its nearest neighbor *C. innocuum*, the
299 differences in its physiological, biochemical, and whole genome sequence suggest its placement
300 in a novel genus. Hence, we propose creation of a novel genus *Dakotella* under family
301 *Erysipelotrichaceae* and classification of SG502^T under new genus *Dakotella as Dakotella*
302 *fusiforme* SG502^T.

303

304 **6. DESCRIPTION OF *Dakotella* gen. nov.**

305 *Dakotella* (*Da.ko.tel'la*. M.L. dimin. ending -*ella*; N.L. fem. n. *Dakotella*, from the place of
306 isolation, State of South Dakota, USA).

307 The type strain is elongated spindle shaped. The closest phylogenetic neighbor is *C. innocuum*
308 ATCC 14501^T with corresponding dDDH of 19.8%. The relative genomic G+C difference with
309 *C. innocuum* is 10.48 mol %. The OrthoANI of the isolate and the type species *C. innocuum* is
310 71.8%. *E. dolichum* DSM 3991^T and *Longicatena caecimuris* DSM 20574^T are distantly related
311 with OrthoANI of 70.85% and 71.84% and dDDH of only 20.6% and 20.5% respectively. The
312 corresponding difference in the G+C content is 5.22 % and 4.86% with *E. dolichum* DSM 3991^T
313 and *Longicatena caecimuris* DSM 20574^T respectively. Such differences support the creation of

novel genus to accommodate SG502^T. The G+C content of the genomic DNA of the type strain is 32.88 mol%. The type species is *Dakotella fusiforme*.

7. DESCRIPTION OF *Dakotella fusiforme* sp. nov. SG502^T sp. nov.

Dakotella fusiforme (fu.si.for'me. L. masc. *fusus* spindle): referring to shape

The cells of the bacterium are anaerobic, gram-positive non-motile rods. The average size of the cell is 1.5×0.35 μm. Bacterial colonies on BHI-M agar are white, convex and entire approximately 0.1 cm in diameter. The optimum temperature and pH for the anaerobic growth are 37^{0}C and 7.0 respectively. The strain SG502^T utilizes glucose, sorbitol, maltose, arbutin, D-fructose, L-fucose, palatinose, dextrin, turanose, D-trehalose, L-rhamnose, uridine, pyruvic acid methyl ester, pyruvic acid, 3-methyl-D-glucose, gentiobiose, maltotriose, ducitol, L-phenylalanine, a-ketovaleric acid, N-acetyl-D-glucosamine, N-acetyl-b-D-mannosamine, cellobiose, a-ketobutyric acid, D-galacturonic acid and N-acetyl-D-glucosamine. Positive enzymatic reactions were observed for alkaline phosphatase only. The primary short-chain fatty acid produced by the strain is acetate while small amounts of propionate and butyrate were also noted. The major cellular fatty acids of the strain SG502^T are $C_{18:1}$ ω9c, $C_{18:0}$ and $C_{16:0}$.. The type strain, SG502^T (=DSM 107282^T=CCOS 1889^T), was isolated from a healthy human fecal sample. The genomic size of the strain is 2, 683,792 bp and G+C content of the strain SG502^T is 32.88mol%.

5. PROTOLOGUE

The GenBank accession number for the 16S rRNA gene sequence of the strain SG502^T is MN266902. The GenBank BioProject ID number for the draft genome sequence of the strain SG502^T is PRJNA494608.

AUTHOR STATEMENTS

Funding information

This work was supported by the USDA National Institute of Food and Agriculture, Hatch projects SD00H702-20 and SD00R540-15, and a grant from the South Dakota Governor's Office of Economic Development awarded to Joy Scaria. No additional external funding was received for this study.

Acknowledgments

The authors would like to thank Electron Microscopy Core Facility at the Bowling Green State University, Ohio, USA for assistance with scanning electron microscopy.

Conflicts of interest

The authors declare no conflicts of interest.

ABBREVIATIONS

MALDI-ToF: Matrix Assisted Laser Desorption/Ionization-Time of Flight

ANI: Average Nucleotide Identity

353

354 **REFERENCES**

355 Alcaide M, Messina E, Richter M, Bargiela R, Peplies J, Huws SA, Newbold CJ, Golyshin PN, Simon MA,
356 Lopez G, Yakimov MM, and Ferrer M. 2012. Gene sets for utilization of primary and secondary
357 nutrition supplies in the distal gut of endangered Iberian lynx. *PLoS One* 7:e51521.
358 10.1371/journal.pone.0051521
359 Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, and Finn RD. 2019. A
360 new genomic blueprint of the human gut microbiota. *Nature* 568:499-504. 10.1038/s41586-019-
361 0965-1
362 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S,
363 Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, and Pevzner PA. 2012.
364 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J*
365 *Comput Biol* 19:455-477. 10.1089/cmb.2012.0021
366 Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, and Lawley TD. 2016.
367 Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation.
368 *Nature* 533:543-546. 10.1038/nature17645
369 Carver T, Thomson N, Bleasby A, Berriman M, and Parkhill J. 2009. DNAPlotter: circular and linear
370 interactive genome visualization. *Bioinformatics* 25:119-120. 10.1093/bioinformatics/btn578
371 Clarridge JE, 3rd. 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on
372 clinical microbiology and infectious diseases. *Clin Microbiol Rev* 17:840-862, table of contents.
373 10.1128/CMR.17.4.840-862.2004
374 Collins MD, Lawson PA, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, Cai J, Hippe H, and
375 Farrow JA. 1994. The phylogeny of the genus Clostridium: proposal of five new genera and
376 eleven new species combinations. *Int J Syst Bacteriol* 44:812-826.
377 Cox LM, Yamanishi S, Sohn J, Alekseyenko AV, Leung JM, Cho I, Kim SG, Li H, Gao Z, Mahana D, Zarate
378 Rodriguez JG, Rogers AB, Robine N, Loke P, and Blaser MJ. 2014. Altering the intestinal
379 microbiota during a critical developmental window has lasting metabolic consequences. *Cell*
380 158:705-721. 10.1016/j.cell.2014.05.052
381 Daillere R, Vetizou M, Waldschmitt N, Yamazaki T, Isnard C, Poirier-Colame V, Duong CPM, Flament C,
382 Lepage P, Roberti MP, Routy B, Jacquelot N, Apetoh L, Becharef S, Rusakiewicz S, Langella P,
383 Sokol H, Kroemer G, Enot D, Roux A, Eggermont A, Tartour E, Johannes L, Woerther PL, Chachaty
384 E, Soria JC, Golden E, Formenti S, Plebanski M, Madondo M, Rosenstiel P, Raoult D, Cattoir V,
385 Boneca IG, Chamaillard M, and Zitvogel L. 2016. Enterococcus hirae and Barnesiella
386 intestinihominis Facilitate Cyclophosphamide-Induced Therapeutic Immunomodulatory Effects.
387 *Immunity* 45:931-943. 10.1016/j.immuni.2016.09.009
388 E. S, and J. E. 2006. Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today*
389 33:152-155.
390 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic*
391 *Acids Res* 32:1792-1797. 10.1093/nar/gkh340
392 Egert M, Wagner B, Lemke T, Brune A, and Friedrich MW. 2003. Microbial community structure in
393 midgut and hindgut of the humus-feeding larva of Pachnoda ephippiata (Coleoptera:
394 Scarabaeidae). *Appl Environ Microbiol* 69:6659-6668. 10.1128/aem.69.11.6659-6668.2003
395 Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*
396 17:368-376.

Fournier PE, and Drancourt M. 2015. New Microbes New Infections promotes modern prokaryotic taxonomy: a new section "TaxonoGenomics: new genomes of microorganisms in humans". *New Microbes New Infect* 7:48-49. 10.1016/j.nmni.2015.06.001

Greiner T, and Backhed F. 2011. Effects of the gut microbiota on obesity and glucose homeostasis. *Trends Endocrinol Metab* 22:117-123. 10.1016/j.tem.2011.01.002

Gurevich A, Saveliev V, Vyahhi N, and Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072-1075.

Han I, Congeevaram S, Ki DW, Oh BT, and Park J. 2011. Bacterial community analysis of swine manure treated with autothermal thermophilic aerobic digestion. *Appl Microbiol Biotechnol* 89:835-842. 10.1007/s00253-010-2893-8

Kaakoush NO. 2015. Insights into the Role of Erysipelotrichaceae in the Human Host. *Front Cell Infect Microbiol* 5:84. 10.3389/fcimb.2015.00084

Kim M, Oh HS, Park SC, and Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 64:346-351. 10.1099/ijs.0.059774-0

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111-120.

Kobyliak N, Conte C, Cammarota G, Haley AP, Styriak I, Gaspar L, Fusek J, Rodrigo L, and Kruzliak P. 2016. Probiotics in prevention and treatment of obesity: a critical view. *Nutrition & Metabolism* 13:14. 10.1186/s12986-016-0067-0

Konstantinidis KT, and Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 187:6258-6264. 10.1128/JB.187.18.6258-6264.2005

Konstantinidis KT, and Tiedje JM. 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* 10:504-509. 10.1016/j.mib.2007.08.006

Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870-1874. 10.1093/molbev/msw054

Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerce J, Mailhe M, Ricaboni D, Bilen M, Dangui Nieko NP, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahao J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier PE, Levasseur A, and Raoult D. 2016. Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* 1:16203. 10.1038/nmicrobiol.2016.203

Lagier JC, Million M, Hugon P, Armougom F, and Raoult D. 2012. Human gut microbiota: repertoire and variations. *Front Cell Infect Microbiol* 2:136. 10.3389/fcimb.2012.00136

Lee I, Ouk Kim Y, Park SC, and Chun J. 2016. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 66:1100-1103. 10.1099/ijsem.0.000760

Meier-Kolthoff JP, and Goker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. 10.1038/s41467-019-10210-3

Rzhetsky. A., and M. N. 1992. A Simple Method for Estimating and Testing Minimum Evolution Trees *MOl Biol Evol* 9:945-967.

Saitou N, and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425. 10.1093/oxfordjournals.molbev.a040454

Sasser M. 1990. Identification of Bacteria by Gas Chromatography of Cellular Fatty Acids. *MIDI Techincal Note 101*.

444 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068-2069.
445 10.1093/bioinformatics/btu153
446 Shields P, and Cathcart L. 2011

447 Motility Test Medium Protocol. *American Society for Microbiology*.
448 Tegtmeier D, Riese C, Geissinger O, Radek R, and Brune A. 2016. Breznakia blatticola gen. nov. sp. nov.
449 and Breznakia pachnodae sp. nov., two fermenting bacteria isolated from insect guts, and
450 emended description of the family Erysipelotrichaceae. *Syst Appl Microbiol* 39:319-329.
451 10.1016/j.syapm.2016.05.003
452 Verbarg S, Göker M, Scheuner C, Schumann P, and Stackebrandt E. 2014. The Families
453 Erysipelotrichaceae emend., Coprobacillaceae fam. nov., and Turicibacteraceae fam. nov. In:
454 Rosenberg E, DeLong EF, Lory S, Stackebrandt E, and Thompson F, eds. *The Prokaryotes:*
455 *Firmicutes and Tenericutes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 79-105.
456 Verbarg S, Rheims H, Emus S, Fruhling A, Kroppenstedt RM, Stackebrandt E, and Schumann P. 2004.
457 Erysipelothrix inopinata sp. nov., isolated in the course of sterile filtration of vegetable peptone
458 broth, and description of Erysipelotrichaceae fam. nov. *Int J Syst Evol Microbiol* 54:221-225.
459 10.1099/ijs.0.02898-0
460 Vetizou M, Pitt JM, Daillere R, Lepage P, Waldschmitt N, Flament C, Rusakiewicz S, Routy B, Roberti MP,
461 Duong CP, Poirier-Colame V, Roux A, Becharef S, Formenti S, Golden E, Cording S, Eberl G,
462 Schlitzer A, Ginhoux F, Mani S, Yamazaki T, Jacquelot N, Enot DP, Berard M, Nigou J, Opolon P,
463 Eggermont A, Woerther PL, Chachaty E, Chaput N, Robert C, Mateus C, Kroemer G, Raoult D,
464 Boneca IG, Carbonnel F, Chamaillard M, and Zitvogel L. 2015. Anticancer immunotherapy by
465 CTLA-4 blockade relies on the gut microbiota. *Science* 350:1079-1084. 10.1126/science.aad1329
466 Weisburg WG, Tully JG, Rose DL, Petzel JP, Oyaizu H, Yang D, Mandelco L, Sechrest J, Lawrence TG, Van
467 Etten J, Maniloff J, and Woese CR. 1989. A phylogenetic analysis of the Mycoplasmas: Basis for
468 their classificiation. *Journal of Bacteriology* 171:6455-6467.
469 Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R,
470 and Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and
471 archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* 12:635.
472 10.1038/nrmicro3330

473 https://www.nature.com/articles/nrmicro3330#supplementary-information
474 Yutin N, and Galperin MY. 2013. A genomic update on clostridial phylogeny: Gram-negative spore
475 formers and other misplaced clostridia. *Environ Microbiol* 15:2631-2641. 10.1111/1462-
476 2920.12173

477

Table 1(on next page)

Table 1

Differential phenotypic features of the strain SG502^T and its closest phylogenetic neighbor *C. innocuum* ATCC 14501^T and *E. dolichum* DSM 3991^T identified using API ZYM (bioMerieux, France).

Table 1: Differential phenotypic features of the strain SG502^T and its closest phylogenetic neighbor *C. innocuum* ATCC 14501^T and *E. dolichum* DSM 3991^T identified using API ZYM (bioMerieux, France).

Characteristics	SG502^T	ATCC 14501^T	DSM 3991^T (□)
Cell shape	Spindle	Rods	Rods
Gram stain	+	+	+
Growth at 37^{0}C (an.)	+	+	+
Optimal pH	7	7	7
Motility	-	-	-
Size (μ)	1.5×0.35	2.0-4.0 × 0.4-1.0	1.6-6.0 × 0.4-0.6
Carbon sources utilization			
Glucose	+	+	+
Sucrose	-	+	-
Salicin	-	+	-
Mannitol	-	+	-
Lactose	-	-	-
Sorbitol	+	-	-
Maltose	+	±	+
D-Trehalose	+	+	+
Raffinose	-	+	-
Cellobiose	+	+	-
Enzyme activity (API ZYM)			
Alkaline phosphatase	+	+	ND
Esterase (C4)	-	-	ND
Esterase Lipase (C8)	-	-	ND
Lipase (C14)	-	-	ND
Leucine arylamidase	-	+	ND
Valine arylamidase	-	-	ND
Cystine arylamidase	-	+	ND
Trypsin	-	-	ND
α-chymotrypsin	-	+	ND
Acid phosphatase	-	+	ND
Naphthol-As-Bi-phosphopydrolase	-	-	ND
α-galactosidase	-	-	ND
β-galactosidase	-	-	ND
β-glucuronidase	-	-	ND
α-glucosidase	-	-	ND
β-glucosidase	-	-	ND
N-acetyl-β-glucosaminidase	-	-	ND
α-mannosidase	-	-	ND

GC content (%)	32.88	44.5	39

 Data obtained from Moore et al (1976).
"ND" not determined.

1

Table 2(on next page)

Table 2

Cellular fatty acid contents percentages (%) of strain SG502^T compared to its phylogenetic neighbors *C. innocuum* ATCC 14501^T and *E. dolichum* DSM 3991^T. Those fatty acids which were not separated using MIDI system were considered as summed features. Summed feature 5 contains $C_{15:0}$ DMA or $C_{14:0}$ 3-OH; summed feature 8 contains $C_{17:1\ cis\ 9}$ or $C_{17:2}$ and summed feature contains $C_{18:1}$ c11/t9/t6 or UN17.83Q.

Table 2: Cellular fatty acid contents percentages (%) of strain SG502^T compared to its phylogenetic neighbors *C. innocuum* ATCC 14501^T and *E. dolichum* DSM 3991^T. Those fatty acids which were not separated using MIDI system were considered as summed features. Summed feature 5 contains $C_{15:0}$ DMA or $C_{14:0}$ 3-OH; summed feature 8 contains $C_{17:1\ cis\ 9}$ or $C_{17:2}$ and summed feature contains $C_{18:1}$ c11/t9/t6 or UN17.83Q.

Characteristics	SG502^T	ATCC 14501^T	DSM 3991^T (‡)
Straight chain			
$C_{10:0}$	0.31	0.32	1
$C_{12:0}$	1.74	3.28	-
$C_{14:0}$	3.54	8.85	1.6
$C_{16:0}$	**14.7**	**23.7**	**22.8**
$C_{16:0}$ aldehyde	0.42	2.78	-
$C_{17:0}$	1.41	-	1.8
$C_{18:0}$	**22.55**	**10.56**	**17.1**
Dimethyl acetal (DMA)			
$C_{16:0}$ DMA	1.46	9.95	-
$C_{18:0}$ DMA	3.62	2.83	-
Unsaturated			
$C_{16:1}$ ω9c	0.99	6.51	-
$C_{16:1}$ ω7c	**6.47**	**10.59**	-
$C_{18:1}$ ω9c	**29.82**	**14.64**	**33.9**
$C_{18:1}$ ω7c	**10.77**	**3.08**	-
Summed Feature 8	0.54	-	-
Summed Feature 10	10.77	3.08	-

(‡) Data from Paek et al. 2017 (Paek et al. 2017)

1

Table 3(on next page)

Table 3

Genomic comparison of the strain SG502^T with its neighbors

1 **Table 3:** Genomic comparison of the strain SG502^T with its neighbors.

Bacteria	Contigs	size (bp)	G+C (%)	rRNA	tRNA	tmRNAs	CDSs	Ac
SG502^T	119	2,683,792	32.9	2	49	1	2,654	
Clostridium innocuum DSM 1286^T	18	4,772,018	43.4	13	48	1	4,653	AGYV
Longicatena caecimuris DSM 29481^T	45	2,945,084	37.8	1	43	1	2,708	SMBP
Eubacterium dolichum DSM 3991^T	25	2,190,453	38.1	6	41	1	2,144	ABAW
Faecalicoccus pleomorphus DSM 20574^T	47	1,992,636	39	5	44	0	1,961	ATUT
Faecalitalea cylindroides ATCC 27803^T	143	1,944,726	34.7	0	23	1	1,904	AWVI
Holdemanella biformis DSM 3989^T	161	2,415,920	33.8	4	45	0	2,347	ABYT
Dielma fastidiosa DSM 26099^T	145	3,575,363	40	3	50	1	3,447	CAEN
Clostridium cocleatum DSM 1551^T	88	2,957,106	28.5	8	42	1	2,670	FOIN
Clostridium saccharogumia DSM 17460^T	135	3,141,523	30.2	5	56	1	2,751	JMLH
Clostridium spiroforme DSM 1552^T	16	2,507,485	28.6	13	58	1	2,330	ABIK
Clostridium ramosum DSM 1402^T	16	3,234,795	31.4	9	41	1	3,056	ABFX

2

3

4

Table 4(on next page)

Table 4

Genomic comparison of the strain SG502^{T} with its neighbors 16S rRNA phylogenetic neighbors and proposed *Erysipelatoclostridium* members using TYGS

Table 4: Genomic comparison of the strain SG502^T with its neighbors 16S rRNA phylogenetic neighbors and proposed *Erysipelatoclostridium* members using TYGS.

Bacteria	dDDH (d4, %)	C.I. (d4, %)	G+C difference (%)
Clostridium innocuum DSM 1286^T	19.8	[17.6 - 22.2]	10.48
Longicatena caecimuris DSM 29481^T	20.5	[18.3 - 22.9]	4.86
Eubacterium dolichum DSM 3991^T	20.6	[18.4 - 23.0]	5.22
Faecalicoccus pleomorphus DSM 20574^T	19.6	[17.4 - 22.0]	6.14
Faecalitalea cylindroides ATCC 27803^T	18.9	[16.7 - 21.3]	1.79
Holdemanella biformis DSM 3989^T	19	[16.8 - 21.4]	0.9
Dielma fastidiosa DSM 26099^T	20.1	[17.9 - 22.5]	7.08
Clostridium cocleatum DSM 1551^T	23.5	[21.2 - 26.0]	4.37
Clostridium saccharogumia DSM 17460^T	18.6	[16.4 - 21.0]	2.72
Clostridium spiroforme DSM 1552^T	19.8	[17.6 - 22.2]	4.31
Clostridium ramosum DSM 1402^T	22.4	[20.1 - 24.8]	1.5

Figure 1

Figure 1

A. MALDI-ToF reference spectrum obtained for SG502^{T}. **B.** Scanning electron micrograph of strain SG502^{T}. Cells were imaged after culturing in anaerobic conditions for 24 hours at 37^{0}C in BHI-M medium. Bar, 200 nm.

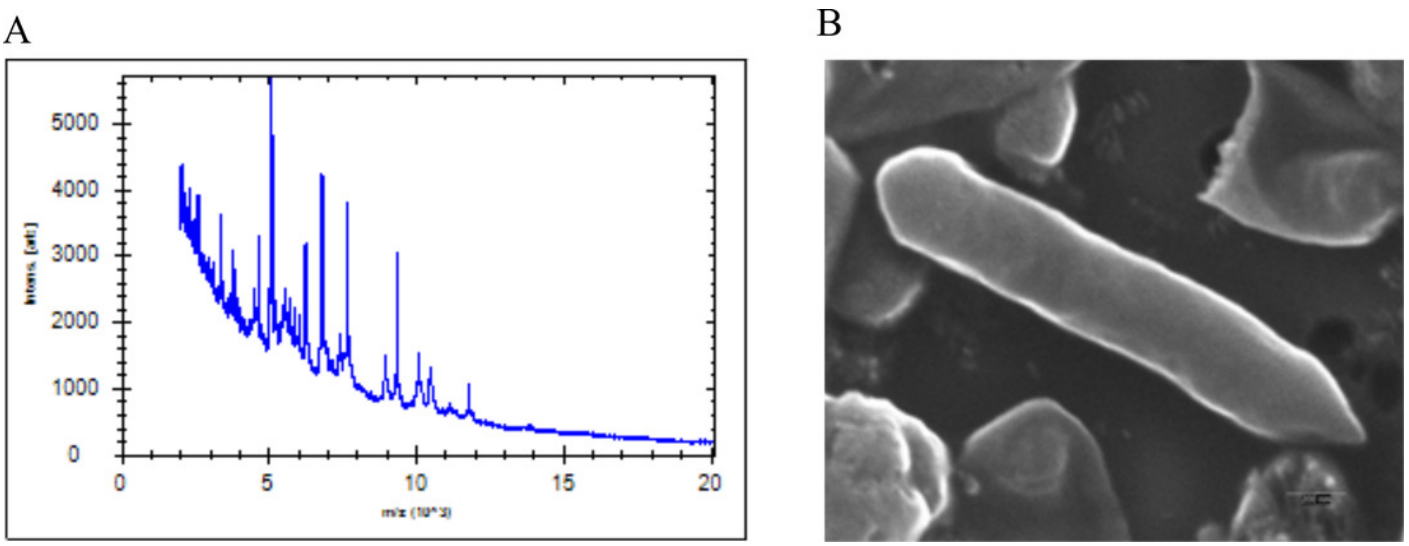


Figure 2

Figure 2

Neighbor Joining tree of 16S rRNA gene sequences of SG502^T with related species under *Erysipelotrichaceae* family. GenBank accession numbers of the 16S rRNA gene sequences are given in parentheses. The sequences were aligned using MUSCLE (Edgar 2004) and the evolutionary distances were computed using Kimura 2-parameter method to obtain the phylogenetic tree in MEGAX (Kumar et al. 2018) after 1000 bootstrap tests (shown as percentages with associated taxa clustered together next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bar, 0.05 substitutions per nucleotide position. *Clostridium butyricum* ATCC 19398^T was used as an outgroup.

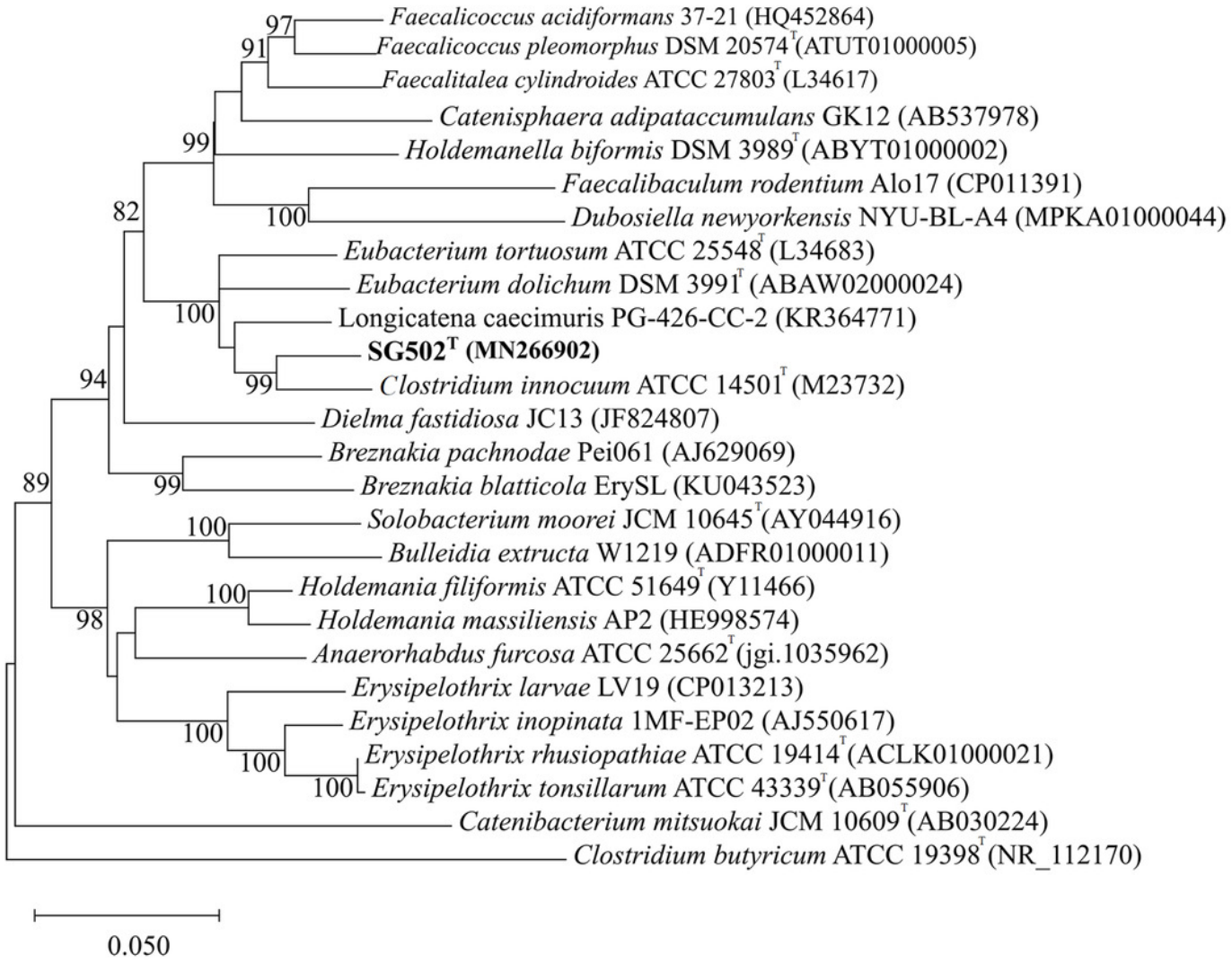

97
91
99
82
100
100
94
99
89
99
100
100
98
100
100
100
Faecalicoccus acidiformans 37-21 (HQ452864)
Faecalicoccus pleomorphus DSM 20574T (ATUT01000005)
Faecalitalea cylindroides ATCC 27803T (L34617)
Catenisphaera adipataccumulans GK12 (AB537978)
Holdemanella biformis DSM 3989T (ABYT01000002)
Faecalibaculum rodentium Alo17 (CP011391)
Dubosiella newyorkensis NYU-BL-A4 (MPKA01000044)
Eubacterium tortuosum ATCC 25548T (L34683)
Eubacterium dolichum DSM 3991T (ABAW02000024)
Longicatena caecimuris PG-426-CC-2 (KR364771)
SG502T (MN266902)
Clostridium innocuum ATCC 14501T (M23732)
Dielma fastidiosa JC13 (JF824807)
Breznakia pachnodae Pei061 (AJ629069)
Breznakia blatticola ErySL (KU043523)
Solobacterium moorei JCM 10645T (AY044916)
Bulleidia extructa W1219 (ADFR01000011)
Holdemania filiformis ATCC 51649T (Y11466)
Holdemania massiliensis AP2 (HE998574)
Anaerorhabdus furcosa ATCC 25662T (jgi.1035962)
Erysipelothrix larvae LV19 (CP013213)
Erysipelothrix inopinata 1MF-EP02 (AJ550617)
Erysipelothrix rhusiopathiae ATCC 19414T (ACLK01000021)
Erysipelothrix tonsillarum ATCC 43339T (AB055906)
Catenibacterium mitsuokai JCM 10609T (AB030224)
Clostridium butyricum ATCC 19398T (NR_112170)
0.050

Figure 3

Figure 3

Circular visualization of genome of SG502^{T}. From outside to inside, “green” circle represents total number of CDS, “blue” represents number of CDS in positive strand, “bluish green” represents number of CDS in negative strand and “red” represents tRNAs position in the genome. The innermost circle represents GC skewness and circle inside to tRNAs represent average GC content. “Magenta” color represents GC at lower level while “Olive green” color represents GC at higher level.

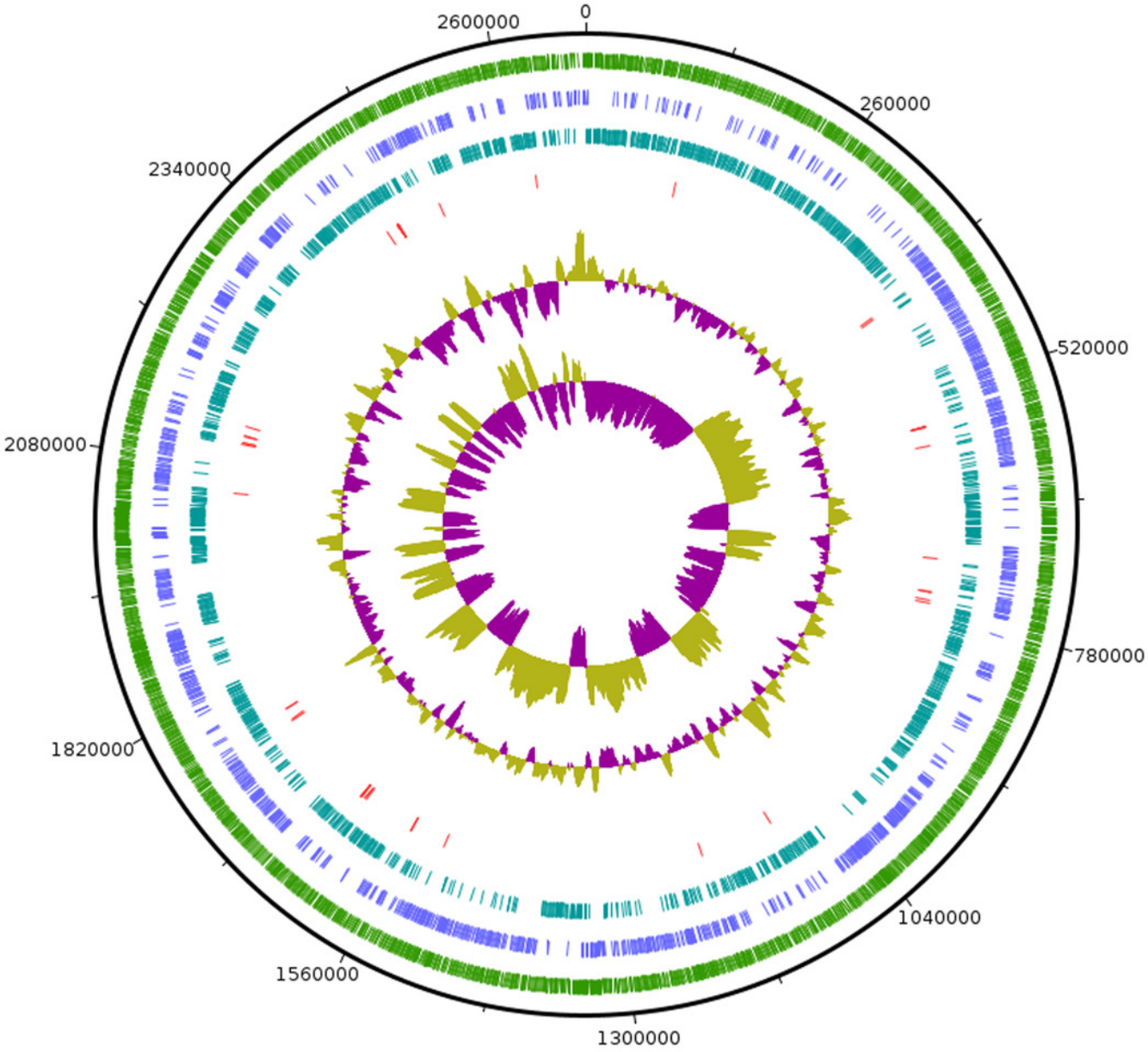

0
260000
520000
780000
1040000
1300000
1560000
1820000
2080000
2340000
2600000

Figure 4

Figure 4

Genomic comparison of SG502^T genome with its neighbors using OrthoANI in OAT software. A. Comparison with 16S phylogenetic neighbors B. Comparison with formerly proposed members of genus *Erysipelatoclostridium*. Color scale indicates % identity between the genomes.

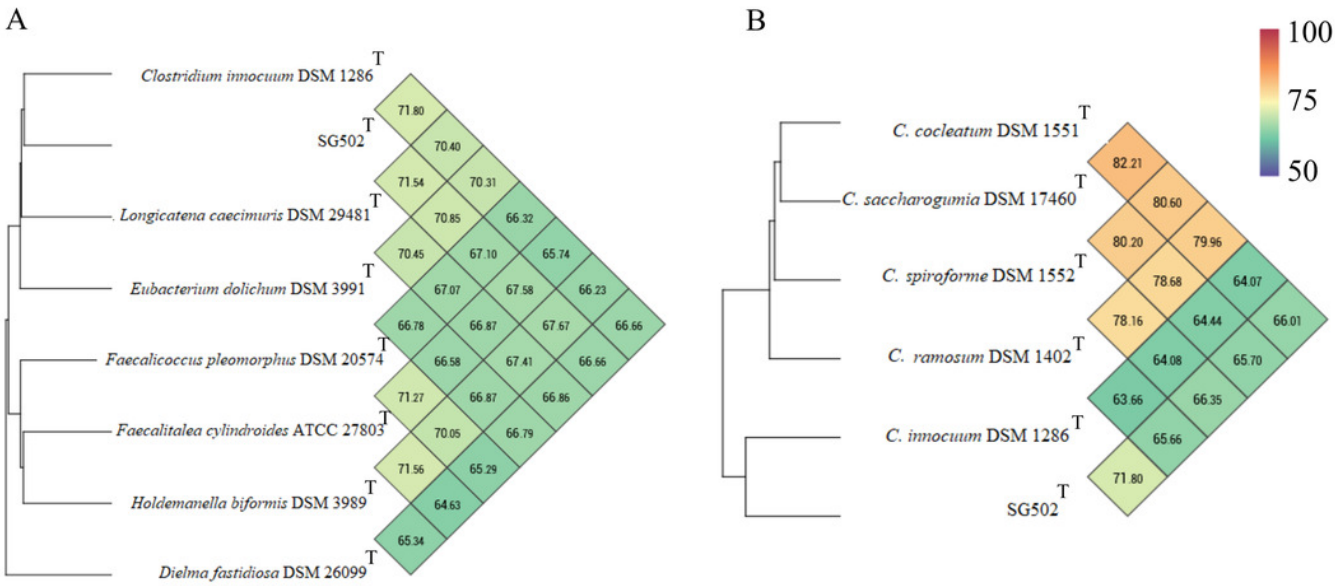
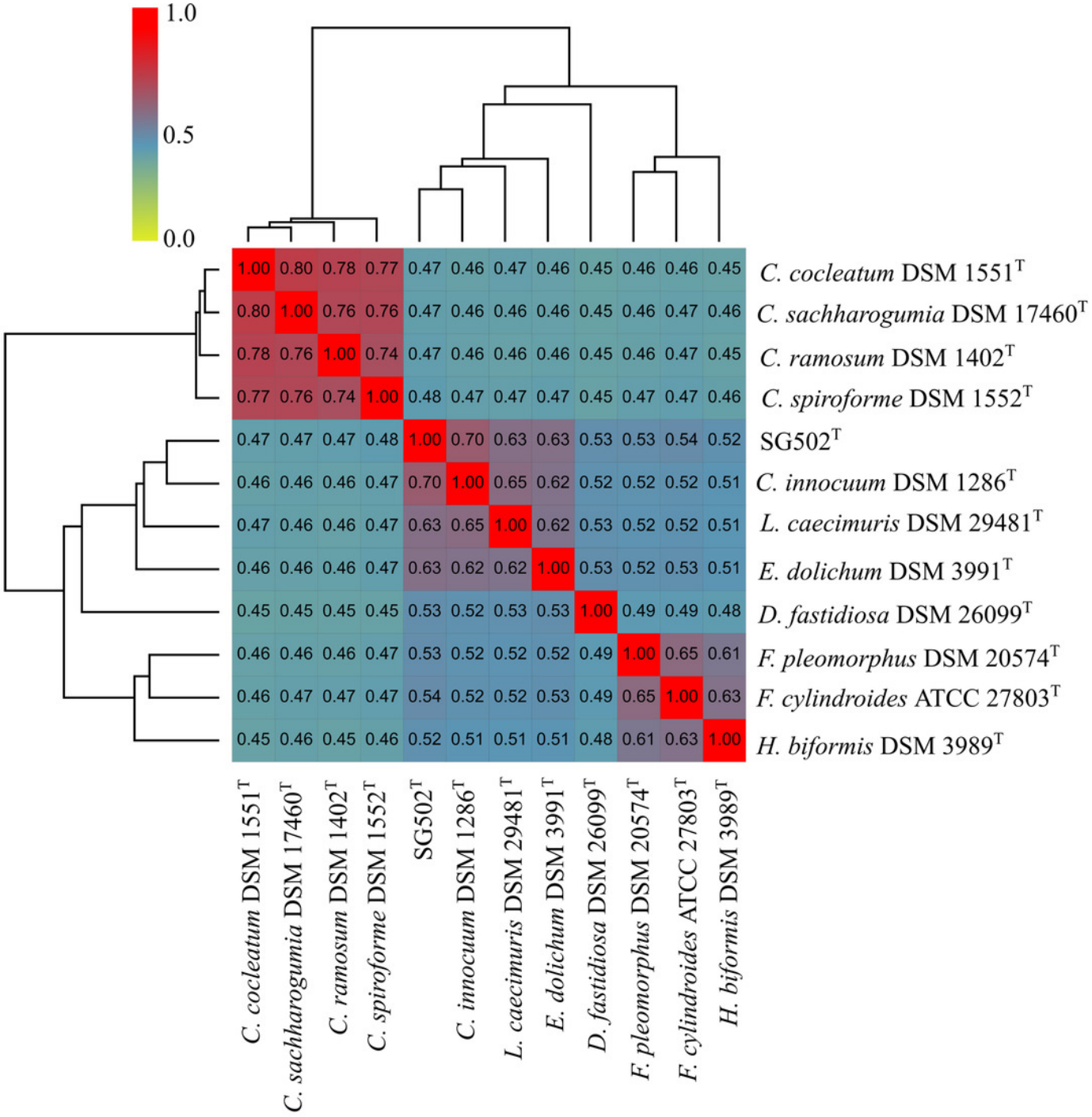


Figure 5

Figure 5

Average Amino acid composition comparison of the strain SG502^{T} with its neighbors 16S rRNA phylogenetic neighbors and proposed *Erysipelatoclostridium* members using AAI calculator


1.0
0.5
0.0
1.00 0.80 0.78 0.77 0.47 0.46 0.47 0.46 0.45 0.46 0.46 0.45
0.80 1.00 0.76 0.76 0.47 0.46 0.46 0.46 0.45 0.46 0.47 0.46
0.78 0.76 1.00 0.74 0.47 0.46 0.46 0.46 0.45 0.46 0.47 0.45
0.77 0.76 0.74 1.00 0.48 0.47 0.47 0.47 0.45 0.47 0.47 0.46
0.47 0.47 0.47 0.48 1.00 0.70 0.63 0.63 0.53 0.53 0.54 0.52
0.46 0.46 0.46 0.47 0.70 1.00 0.65 0.62 0.52 0.52 0.52 0.51
0.47 0.46 0.46 0.47 0.63 0.65 1.00 0.62 0.53 0.52 0.52 0.51
0.46 0.46 0.46 0.47 0.63 0.62 0.62 1.00 0.53 0.52 0.53 0.51
0.45 0.45 0.45 0.45 0.53 0.52 0.53 0.53 1.00 0.49 0.49 0.48
0.46 0.46 0.46 0.47 0.53 0.52 0.52 0.52 0.49 1.00 0.65 0.61
0.46 0.47 0.47 0.47 0.54 0.52 0.52 0.53 0.49 0.65 1.00 0.63
0.45 0.46 0.45 0.46 0.52 0.51 0.51 0.51 0.48 0.61 0.63 1.00
C. cocleatum DSM 1551T
C. sachharogumia DSM 17460T
C. ramosum DSM 1402T
C. spiroforme DSM 1552T
SG502T
C. innocuum DSM 1286T
L. caecimuris DSM 29481T
E. dolichum DSM 3991T
D. fastidiosa DSM 26099T
F. pleomorphus DSM 20574T
F. cylindroides ATCC 27803T
H. biformis DSM 3989T
C. cocleatum DSM 1551T
C. sachharogumia DSM 17460T
C. ramosum DSM 1402T
C. spiroforme DSM 1552T
SG502T
C. innocuum DSM 1286T
L. caecimuris DSM 29481T
E. dolichum DSM 3991T
D. fastidiosa DSM 26099T
F. pleomorphus DSM 20574T
F. cylindroides ATCC 27803T
H. biformis DSM 3989T