Characterization and description of *Faecalibacterium butyricigenerans* sp. nov. and *F. longum* sp. nov., isolated from human faeces

Yuanqiang Zou[1, 2, 3, 4§*], Xiaoqian Lin[1, 5§], Wenbin Xue[1],Ying Dai[1], Karsten Kristiansen[1, 2, 4], Liang Xiao[1, 3, 4*]

[1] BGI-Shenzhen, Shenzhen 518083, China

[2] Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Universitetsparken 13, 2100 Copenhagen, Denmark

[3] Shenzhen Engineering Laboratory of Detection and Intervention of human intestinal microbiome, BGI-Shenzhen, Shenzhen, China

[4] Qingdao-Europe Advanced Institute for Life Sciences, BGI-Shenzhen, Qingdao 266555, China

[5] School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510006, China

[*]**Corresponding authors**:

Liang Xiao.

Address: BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China.

E-mail: xiaoliang@genomics.cn

Yuanqiang Zou.

Address: BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China.

E-mail: zouyuanqiang@genomics.cn

† Yuanqiang Zou and Xiaoqian Lin contributed equally to this work.

Running title: *Faecalibacterium butyricigenerans* sp. nov. and *Faecalibacterium longum* sp. nov.

Contents Category: New Taxa –*Firmicutes* and related organisms

Keywords: *Faecalibacterium*, *Faecalibacterium butyricigenerans* sp. nov., human faeces, taxonomy, genome sequencing, phylogenetic analysis, average nucleotide identity

Abstract

Exploiting a pure culture strategy to investigate the composition of human gut microbiota, two novel anaerobes, designated strains AF52-21^T and CM04-06^T, were isolated from faeces of two healthy Chinese donors and characterized using a polyphasic approach. The two strains were Gram-stain-negative, non-motile, and rod-shaped. Both strains grew optimally at 37°C and pH 7.0. Phylogenetic analysis based on 16S rRNA gene sequences revealed that the two strain clustered with species of the genus *Faecalibacterium* and were most closely related to *Faecalibacterium prausnitzii* ATCC 27768^T with sequence similarity of 97.18% and 96.87%, respectively. The two isolates shared a 16S rRNA gene sequence identity of 98.69%. Draft genome sequencing was performed for strains AF52-21^T and CM04-06^T, generating genome sizes of 2.85 Mbp and 3.01 Mbp. The calculated average nucleotide identity values between the genomes of the strains AF52-21^T and CM04-06^T compared to *Faecalibacterium*

prausnitzii ATCC 27768^T were 83.20% and 82.54%, respectively, and 90.09% when comparing AF52-21^T and CM04-06^T. Both values were below the previously proposed species threshold (95%), supporting their recognition as novel species in the genus _Faecalibacterium_. The genomic DNA G+C contents of strain AF52-21^T and CM04-06^T calculated from genome sequences were 57.77 mol% and 57.51 mol%, respectively. Based on the phenotypic, chemotaxonomic and phylogenetic characteristics, we conclude that both strains represent two new _Faecalibacterium_ species, for which the names _Faecalibacterium butyricigenerans_ sp. nov. (type strain AF52-21^T = CGMCC 1.5206^T = DSM 103434^T) and _Faecalibacterium longum_ sp. nov. (type strain CM04-06^T = CGMCC 1.5208^T = DSM 103432^T) are proposed.

Introduction

The human gastrointestinal[1] tract harbours complex microbial communities[2], dominated by bacteria from the phyla *Bacteroidetes* and *Firmicutes*[3-6]. The composition and diversity of the gut microbiota are affected by numerous factors, including host genetics[7], long-term diet[8,9], drugs[1,10,11] and several other environmental factors[12]. Evidence suggests that the composition of the microbiota is associated with the development of obesity[4,13-15], diabetes[16,17], inflammatory bowel disease[18,19], colorectal cancer[20,21], and non-alcoholic fatty liver disease[22,23]. Therefore, the composition and function of the microbial species living in our gut are crucial importance for maintenance of health. Short-chain fatty acids (SCFAs), produced by fermentation of dietary fibre by several abundant genera of the intestinal microbiota, including *Roseburia*, *Eubacterium* and *Faecalibacterium*[24], have been reported to elicit beneficial effects on energy metabolism and for

prevention of colonization of pathogens[25]. The genus *Faecalibacterium* as an abundant butyric acid-producing bacterium colonizing the human gut displays anti-inflammatory effects and may be used as a potential probiotics for treatment of gut inflammation[26,27].

The genus *Faecalibacterium*, belonging to the family *Ruminococcaceae* within the order *Clostridiales*, comprises only one validated species, *Faecalibacterium prausnitzii*[28], and two non-validly published species, *Faecalibacterium moorei*[29] and *Faecalibacterium hominis*[30], all originally isolated from human faeces. *F. prausnitzii* is a gram-negative non-spore-forming and strictly anaerobic rod-shaped bacterium. The genomic G+C content of genus *Faecalibacterium* ranges from 47% to 57%[31]. The fermentation products from glucose are butyrate, D-lactate and formate. In the present study, we describe two novel species of the genus *Faecalibacterium* by using polyphasic taxonomy along with whole genome sequence analysis.

Results and discussion

Phenotypic and Chemotaxonomic Characterization

Both strains (AF52-21^T and CM04-06^T) were obligate anaerobic, Gram-stain-negative, non-spore-forming, non-motile and rod-shaped bacteria **(Fig. 1)**. After incubation on MPYG agar at 37°C for 2 days, the colonies appeared 1.0-2.0 mm in diameter, round, creamy white to yellowish, convex and opaque with entire margins for AF52-21^T and 2.0 mm in diameter, round, yellowish, slightly convex and opaque with entire margins for CM04-06^T. The growth temperature was 20-42°C (optimum 37°C) for AF52-21^T and 30-45°C (optimum 37°C) for CM04-06^T. Growth was observed at pH 6.0-7.5 (optimum 7.0-7.5) for AF52-21^T and pH 5.0-8.0 (optimum 7.0-7.5) for CM04-06^T. Strains AF52-21^T and CM04-06^T grew with 0-1% and 0-3% NaCl, respectively. Both

88 strains were catalase-negative. The major metabolic end products for strains AF52-21^T and
89 CM04-06^T were acetic acid, formic acid, butyric acid and lactic acid. Differential physiological
90 and biochemical characteristics of strains AF52-21^T and CM04-06^T with the closest related species
91 of genus *Faecalibacterium* are listed in the species description and in **Table 1**.
92
93 **Table 1. Differential phenotypic characteristics of strains AF52-21^T, CM04-06^T, and the**
94 **related species *F. prausnitzii* ATCC 27768^T.**
95 Strains: 1, *F. butyricigenerans* AF52-21^T; 2, *F. longum* CM04-06^T; 3, *F. prausnitzii* ATCC
96 27768^T. +, positive; w, weakly positive; –, negative.

Phenotypic features	**1**	**2**	**3**
Growth:			
Temperature range (optimum) (°C)	20-42 (37)	30-45 (37)	20-42 (37)
pH range	6.0-7.5	5.0-8.0	6.0-7.5
Salt tolerance (%)	1	3	3
Hydrolysis of:			
Aesculin	+	−	+
Gelatin	−	+	−
Acid from (API 20A and API 50CHL):			
Cellobiose	+	−	w
D-Fructose	w	−	+
D-Fucose	w	−	w
D-Galactose	w	−	−
D-Glucose	w	−	+
D-Lactose	+	−	−
D-Maltose	+	+	w
D-Fannitol	+	−	−
D-Fannose	w	−	−
D-Mannose	+	+	−
D-Raffinose	−	w	−
D-Trehalose	+	w	w
Gluconate	−	−	+
Glycogen	+	−	−

Inositol	w	−	−
Inulin	+	−	+
Methyl-*β*-D-Xylopyranoside	w	−	−
Enzyme activity (API ZYM):			
N-acetyl-*β*-Glucosaminidase	−	w	−
Naphthol-AS-BI-Phosphohydrolase	+	−	+
α-Glucosidase	−	−	+
β-Galactosidase	−	−	w
β-Glucosidase	+	−	−
β-Glucuronidase	+	w	+
DNA G+C (mol %)	57.77	57.51	52−57

All data are from this study.

The result of cellular fatty acid profiles of strain AF52-21^T, CM04-06^T and related species are shown in **Table 2**. The major components of fatty acids (constituting >5% of the total) present in strain AF52-21^T were $C_{14:0}$ (5.9%), $C_{16:0}$ (16.3%), $C_{18:1}$ $\omega 7c$ (8.1), $C_{18:1}$ $\omega 9c$ (39.0%) and iso-$C_{19:0}$ (12.9%). The profiles including $C_{16:0}$ (25.5%), $C_{18:1}$ $\omega 7c$ (7.5%), $C_{18:1}$ $\omega 9c$ (32.5%), iso-$C_{19:0}$ (5.9%) and iso-$C_{17:1}$ I/anteiso B (9.7%) were detected as the predominant fatty acids for strain CM04-06^T. The highest levels of fatty acids, including $C_{16:0}$ and $C_{18:1}$ $\omega 9c$, were similar, but not identical comparing strain AF52-21^T, CM04-06^T and ATCC 27768^T. Furthermore, strains AF52-21^T, CM04-06^T and ATCC 27768^T could be differentiated by less abundant fatty acids, such as $C_{18:1}$ 2OH, anteiso-$C_{15:0}$, anteiso-$C_{17:0}$, $C_{13:0}$ 3OH/Iso-$C_{15:1}$ I, $C_{16:1}$ $\omega 7c$/$C_{16:1}$ $\omega 6c$ and antei-$C_{18:0}$ /$C_{18:2}$ $\omega 6$, $9c$ (**Table 2**).

Table 2. Fatty acid profile of strains AF52-21^T, CM04-06^T and the closest related species *F. prausnitzii* ATCC 27768^T.

Numbers represent percentages of the total fatty acids. –, not detected (<1%).

Fatty acids composition	*F. butyricigenerans* AF52-21^T	*F. longum* CM04-06^T	*F. prausnitzii* ATCC 27768^T
$C_{12:0}$	1.5	1.8	1.9
$C_{13:1}$	–	–	1.25
$C_{14:0}$	**5.9**	4.6	**11.8**
$C_{16:0}$	**16.3**	**25.5**	**21.1**
$C_{17:1}$ $\omega 8c$	1.3	–	1.1
$C_{18:1}$ $\omega 7c$	**8.1**	**7.5**	**5.7**
$C_{18:1}$ $\omega 9c$	**39.0**	**32.5**	**31.4**
$C_{18:0}$	4.5	3.5	4.1
$C_{18:1}$ 2OH	2.9	–	2.0
Iso-$C_{19:1}$ I	1.2	1.1	2.1
Iso-$C_{19:0}$	**12.9**	**5.9**	–
Anteiso-$C_{15:0}$	–	2.6	–
Anteiso-$C_{17:0}$	–	2.1	–
$C_{13:0}$ 3OH/ Iso-$C_{15:1}$ I	–	–	2.1
$C_{16:1}$ $\omega 7c$/ $C_{16:1}$ $\omega 6c$	1.5	1.9	4.0
Iso-$C_{17:1}$ I/anteiso B	4.7	**9.7**	**7.6**
Antei-$C_{18:0}$ /$C_{18:2}$ $\omega 6$, $9c$	–	1.9	1.3

Phylogenetic Analysis

The almost complete 16S rRNA gene sequences of strains AF52-21^T and CM04-06^T, comprising 1,382bp and 1,374bp, respectively, were obtained. BLAST analysis of the 16S rRNA gene sequences against the EzBioCloud server showed that the two strains grouped in the genus *Faecalibacterium* within the family *Ruminococcaceae* and were most closely related to *F. prausnitzii* ATCC 27768^T, which is the sole valid species of the genus *Faecalibacterium*, with similarity values of 97.18% and 96.87%, respectively. *Faecalibacterium hominis* 4P-15^T, an unrecognized species of the genus *Faecalibacterium*, was also used as a related taxa for 16S rRNA gene analysis. Strains AF52-21^T and CM04-06^T share 16S rRNA gene sequence similarity of 98.65% and 97.68% with *F. hominis* 4P-15^T. The 16S rRNA gene sequence similarity between

124 strains AF52-21^T and CM04-06^T was 98.69% (**Table 3**), both these values were lower than the
125 recommended thresholds (98.7%) for classification of human-associated bacterial isolates at the
126 species level [32]. Phylogenetic analysis based on the neighbour-joining, maximum-likelihood and
127 minimum-evolution (**Fig. 2**, **Supplementary Fig. S1** and **Fig. S2**, respectively) confirmed the
128 result of affiliation of the novel isolates with the species of the genus *Faecalibacterium* and
129 revealed that the two isolates formed a distinct lineage with *F. prausnitzii* ATCC 27768^T.

130

131 **Table 3. Levels of 16S rRNA gene sequence similarity and ANI values (in percentages) based**
132 **on BLAST for strains AF52-21^T, CM04-06^T and the phylogenetically related species *F.***
133 ***prausnitzii* ATCC 27768^T and the unrecognized species *Faecalibacterium hominis* 4P-15^T.**

134 Taxa: 1, *F. butyricigenerans* AF52-21^T; 2, *F. longum* CM04-06^T; 3, *F. prausnitzii* ATCC
135 27768^T; 4, *Faecalibacterium hominis* 4P-15^T.

Strain	**Accession no.**	1	2	3	4
16S rRNA gene sequence similarity (%)					
AF52-21^T	KX146426	100	98.69	97.18	98.65
CM04-06^T	KX150462	98.69	100	96.87	97.68
ATCC 27768^T	AJ413954	97.18	96.87	100	98.35
4P-15^T	NMDCN000012L	98.65	97.68	98.35	100
ANI values (%)					
AF52-21^T	CNA0017730	100	90.01	83.16	85.72
CM04-06^T	CNA0017731	90.19	100	82.53	85.40
ATCC 27768^T	CNA0017732	83.32	82.58	100	85.7985
4P-15^T	NMDC60014083	85.72	85.40	85.7985	100

136

Genome Analysis and function annotation

The assembled draft genomes of strains AF52-21^T and CM04-06^T comprised total lengths of 2,851,918bp and 3,011,178bp with 73 and 47 scaffolds, respectively (**Table 4**). The G+C contents calculated from genome sequences were 57.77% and 57.51%, which were slightly higher than the range reported previously for the genus *Faecalibacterium* (47-57 mol%) [28]. CheckM analysis of the genomes showed high completeness (>90%) and low contamination (<5%) (**Table 4**), indicating these are high-quality genomes sequences. The genome comparison between strains AF52-21^T, CM04-06^T, ATCC 27768^T and 4P-15^T showed ANI values ranged from 82.53% to 90.19% (**Table 3**), which were significantly below the proposed cutoff value of 95–96% for delineating bacterial species, indicating that strains AF52-21^T and CM04-06^T represented novel species in the genus *Faecalibacterium*. Circular maps of the two strains AF52-21^T and CM04-06^T are shown in **Fig. 3** and **Fig. 4**.

Table 4. Genome properties of *F. butyricigenerans* AF52-21^T and *F. longum* CM04-06^T.

Feature	AF52-21^T	CM04-06^T
Accession No.	CNA0017730	CNA0017731
Approximate Genome Size (bp)	2,851,918	3,011,178
G+C content (mol%)	57.77	57.51
DNA scaffolds	73	47
N50 Length	191,233	119,299
Completeness	100	99.32
Contamination	0	0
Genes total number	2,291	2,506
Gene average length (bp)	939	920
rRNAs (5S, 16S, 23S)	4	5

tRNAs	60	61
sRNA	0	0
Genes assigned to COGs	2029	2164

For genome annotation, the distributions of the genes into clusters of orthologous groups (COGs) functional categories are depicted in **Fig. 5** and **Table S1**. Both strain strains AF52-21^T and CM04-06^T share identical COGs functional categories. The abundant categories comprise amino acid transport and metabolism (E), carbohydrate transport and metabolism (G), cell wall/membrane/envelope biogenesis (M), energy production and conversion (C), general function prediction only (R), replication, recombination and repair (L), signal transduction mechanisms (T), transcription (K) and translation, ribosomal structure and biogenesis (J). Annotated genes associated with synthesis of diaminopimelic acid, teichoic and lipoteichoic acids and lipopolysaccharides, and metabolism of polar lipids and polyamines by RAST annotation, comparing strains AF52-21^T, CM04-06^T with ATCC 27768^T, are shown in **Table 5** and **Table S2**.

Table 5. Number of genes associated with biosynthetic pathway from whole genome sequences of strain *F. butyricigenerans* AF52-21^T and *F. longum* CM04-06^T and *F. prausnitzii* ATCC 27768^T identified by RAST.

Taxa: 1, AF52-21^T; 2, CM04-06^T; 3, *F. prausnitzii* ATCC 27768^T. Data are for type strains. Numbers of genes identified for lipopolysaccharides and mycolic acids were zero for all taxa studied.

Genes responsible for biosynthesis	**1**	**2**	**3**

Diaminopimelic acid	11	12	12
Polar lipids	18	19	22
Polyamines	12	13	11
Quinones	14	16	15
Teichoic and lipoteichoic acids	3	2	3

The annotation showed that AF52-21^T, CM04-06^T, and ATCC 27768^T contained a complete acetyl-CoA to butyrate synthesis pathway, but possessed butyryl-CoA:acetate CoA-transferase activity only in the final step (**Fig. 6**), as discussed previously [33,34]. Prophages were identified using the PHAST software, and the results are shown in **Fig. 7**. Two incomplete phage sequences were detected in the AF52-21^T genome, one of which encodes the Phd_YefM protein, an antitoxin component. Three incomplete phage sequences and two intact prophages were detected in the CM04-06^T genome, encoding the Phd_YefM protein, relaxase/mobilisation nuclease domain, bacterial mobilisation protein (MobC) /ribbon-helix-helix protein, helix-turn-helix, and predicted transcriptional regulators. Moreover, the antibiotic resistance analysis indicated that AF52-21^T contained macrolide antibiotic, lincosamide antibiotic, and streptogramin antibiotic genes, while CM04-06^T and ATCC 27768^T contained aminoglycoside antibiotic genes (**Fig. 8**).

Discussion

The 16S rRNA genes phylogenetic, physiological results and genome description showed that the two new isolates AF52-21^T and CM04-06^T represent two novel species. The ANI values between AF52-21^T, CM04-06^T and the closest related species ATCC 27768^T were 82.54% and 90.09%, respectively, which is in support of a new species delineation. The result of biochemical and

187 genomic functional analyses showed that both AF52-21^{T} and CM04-06^{T} are butyric
188 acid-producing bacteria.

189 Most strains in the genus *Faecalibacterium* exhibit a common ability to produce butyric acid,
190 peptides and other anti-inflammatory substances, which have immunomodulatory effects[26,27,35].
191 Some studies have confirmed that the decreased abundance of this genus is related to the
192 occurrence and development of inflammatory bowel diseases[36-38]. Accordingly, *Faecalibacterium*
193 is receiving much attention as one of the candidate next-generation probiotics (NGPs) , which can
194 be used for disease treatment[39,40].

195 Previous studies based on comparative genomics from isolates suggested the wide diversity of this
196 genus, with the presence of at least two phylotypes in *F. prausnitzii*[29]. A recent study analysing the
197 *Faecalibacterium*-like MAGs, proposed that *Faecalibacterium* from the human gut can be divided
198 into 12 clades[40]. These studies have expanded the diversity of *Faecalibacterium* and proposed that
199 different phylotypes have different functions, which results in different contributions to health or
200 diseases.

201 Moreover, as a candidate taxa for the NGPs, the *Faecalibacterium* isolates can be used for in-vitro
202 functional verification and animal model experiments to further explore its probiotic functions,
203 and ultimately expected to be used in clinical disease intervention.

204

205 ## Description of Faecalibacterium butyricigenerans sp. nov.

206 *Faecalibacterium butyricigenerans* (bu.ty.ri.ci.ge'ne.rans. N.L. n. *acidum butyricum* butyric
207 acid; L. part. adj. *generans*, producing; N.L. adj. *butyricigenerans*, butyric acid-producing;
208 referring to its production of butyric acid)

Cells of strain AF52-21^T are Gram-stain-negative, non-motile, non-spore-forming and rod-shaped. Strictly anaerobic and catalase negative. Colonies on PYG agar are round, creamy white to yellowish, convex and opaque with entire margins and colony size is approximately 1.0-2.0 mm in diameter after incubation at 37°C for 2 days. Cells are able to grow at 20-42°C with optimum temperature at 37°C. The pH range for growth is 6.0-7.5 (optimum at 7.0-7.5). Growth occurs at NaCl concentrations 0-1%. The strain is positive for the assimilation of cellobiose, D-lactose, D-maltose, D-mannitol, D-mannose, D-trehalose, glycogen, inulin, D-fructose, D-fucose, D-galactose, D-glucose, inositol and methyl-β-D-xylopyranoside, but negative for amygdalin, arbutin, D-adonitol, D-arabinose, D-arabitol, D-lyxose, D-melezitose, D-melibiose, D-raffinose, D-ribose, D-sorbitol, D-tagatose, D-turanose, dulcitol, D-xylose, erythritol, gentiobiose, gluconate, glycerol, L-arabinose, L-arabitol, L-fucose, L-rhamnose, L-sorbose, L-xylose, methyl-D-glucopyranoside, methyl-α-D-mannopyranoside, *N*-acetyl-glucosamine, salicin, starch, sucrose, xylitol, 2-ketogluconate and 5-ketogluconate. In enzymatic activity tests, strain AF52-21^T is positive for naphthol-AS-BI-phosphohydrolase, β-glucuronidase and β-glucosidase, and negative for alkaline phosphatase, esterase (C4), esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, α-glucosidase, α-mannosidase and β-fucosidase. Indole is not produced. Positive for hydrolysis of esculin and negative for gelatin. Formic acid, acetic acid, butyric acid and lactic acid are the fermentation products. The major fatty acids are $C_{14:0}$, $C_{16:0}$, $C_{18:1}$ $\omega 7c$, $C_{18:1}$ $\omega 9c$ and iso-$C_{19:0}$.

In the result of RAST annotation, 11 genes/proteins are associated with biosynthesis of DAP, including 4-hydroxy-tetrahydrodipicolinate reductase (EC 1.17.1.8),

4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7), aspartate-semialdehyde dehydrogenase (EC 1.2.1.11), aspartokinase (EC 2.7.2.4) (2 copies), diaminopimelate decarboxylase (EC 4.1.1.20), diaminopimelate epimerase (EC 5.1.1.7), L, L-diaminopimelate aminotransferase (EC 2.6.1.83), *N*-acetyl-L, L-diaminopimelate deacetylase (EC 3.5.1.47), UDP-*N*-acetylmuramoylalanyl-D-glutamate-2, 6-diaminopimelate ligase (EC 6.3.2.13) and UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-D-alanyl-D-alanine ligase (EC 6.3.2.10). 18 genes/proteins are associated with biosynthesis of polar lipids, including 1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51) (2 copies), ABC-type multidrug/protein/lipid transport system, ATPase component, acyl carrier protein (3 copies), acyl-phosphate:glycerol-3-phosphate O-acyltransferase PlsY, cardiolipin synthetase (EC 2.7.8.-) (3 copies), CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5), dihydroxyacetone kinase family protein, glycerol kinase (EC 2.7.1.30), glycerol-3-phosphate dehydrogenase [NAD(P)$^{+}$] (EC 1.1.1.94), phosphate:acyl-ACP acyltransferase PlsX, phosphatidate cytidylyltransferase (EC 2.7.7.41) and phosphatidylglycerophosphatase B (EC 3.1.3.27) (2 copies). 12 genes/proteins are associated with biosynthesis of polyamines, including 5'-methylthioadenosine nucleosidase (EC 3.2.2.16), S-adenosylhomocysteine nucleosidase (EC 3.2.2.9), ABC transporter, periplasmic spermidine putrescine-binding protein PotD (TC 3.A.1.11.1), agmatinase (EC 3.5.3.11), arginine decarboxylase (EC 4.1.1.19), arginine/ornithine antiporter ArcD (2 copies), carboxynorspermidine decarboxylase, putative (EC 4.1.1.-), carboxynorspermidine dehydrogenase, putative (EC 1.1.1.-), putrescine transport ATP-binding protein PotA (TC 3.A.1.11.1), spermidine putrescine ABC transporter permease component PotB (TC 3.A.1.11.1), spermidine putrescine ABC transporter permease component potC

(TC..3.A.1.11.1) and spermidine synthase (EC 2.5.1.16). 3 genes/proteins are associated with biosynthesis of teichoic and lipoteichoic acids, including cell wall teichoic acid glycosylation protein gtcA, teichoic acid export ATP-binding protein TagH (EC 3.6.3.40) and membrane protein involved in the export of O-antigen, teichoic acid lipoteichoic acids. 14 genes/proteins are associated with biosynthesis of quinones, including 2-heptaprenyl-1,4-naphthoquinone methyltransferase (EC 2.1.1.163), electron transport complex protein RnfA (2 copies), electron transport complex protein RnfB, electron transport complex protein RnfC, electron transport complex protein RnfD (2 copies), electron transport complex protein RnfE (2 copies), electron transport complex protein RnfG, F420H2:quinone oxidoreductase, heptaprenyl diphosphate synthase component I (EC 2.5.1.30), microsomal dipeptidase (EC 3.4.13.19) and undecaprenyl diphosphate synthase (EC 2.5.1.31). There are no genes responsible for biosynthesis of lipopolysaccharides or mycolic acids. Additional annotations showed that the AF52-21^{T} genome contains a complete butyrate synthesis pathway, two prophage remnants, and three antibiotic genes.

The type strain, AF52-21^{T} (=CGMCC 1.5206^{T} = DSM 103434^{T}), was isolated from human gut. The G+C content of the genomic DNA is 57.77 mol% as calculated from whole genome sequencing.

Description of Faecalibacterium longum sp. nov.

Faecalibacterium longum (lon'gum. L. neut. adj. *longum* long, the shape of the cells)

Cells are Gram-stain-negative, non-motile, non-spore forming, long rod in shape. Strictly anaerobic. Catalase and urease are negative. Colonies are round, yellowish, slightly convex,

opaque with entire margins with 2.0 mm in diameter on PYG agar for incubation at 37°C for 48 h under anaerobic condition. The strain showed growth at 30-45°C (optimum temperature is 37°C). Growth is observed at pH 5.0-8.0 (optimum pH is 7.0-7.5). NaCl is tolerated with concentrations up to 3%. Acid is produced from D-maltose, D-mannose, D-raffinose, D-trehalose and salicin, but not from amygdalin, arbutin, cellobiose, D-adonitol, D-arabinose, D-arabitol, D-cellobiose, D-fructose, D-fucose, D-galactose, D-glucose, D-lactose, D-lyxose, D-maltose, D-mannitol, D-mannose, D-melezitose, D-melibiose, D-raffinose, D-ribose, D-sorbitol, D-sucrose, D-tagatose, D-turanose, dulcitol, D-xylose, erythritol, gentiobiose, gluconate, glycerol, glycogen, inositol, inulin, L-arabinose, L-arabitol, L-fucose, L-rhamnose, L-sorbose, L-xylose, methyl-D-glucopyranoside, methyl-α-D-mannopyranoside, methyl-β-D-xylopyranoside, *N*-acetyl-glucosamine, salicin, starch, sucrose, xylitol, 2-ketogluconate and 5-ketogluconate. In the API ZYM strip, strain showed weakly positive enzyme activities for β-glucuronidase and *N*-acetyl-β-glucosaminidase, but negative for alkaline phosphatase, esterase (C4), esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, α-mannosidase and β-fucosidase. Indole is not produced. Gelatin is hydrolysed, but aesculin is not. Major end products are acetic acid, formic acid, butyric acid and lactic acid. The major fatty acids (constituting >5% of the total) are $C_{16:0}$, $C_{18:1}$ $\omega 7c$, $C_{18:1}$ $\omega 9c$, iso-$C_{19:0}$ and iso-$C_{17:1}$ I/anteiso B.

In the result of RAST annotation, 12 genes/proteins are associated with biosynthesis of DAP, including 4-hydroxy-tetrahydrodipicolinate reductase (EC 1.17.1.8), 4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7) (2 copies), aspartate-semialdehyde

297 dehydrogenase (EC 1.2.1.11), aspartokinase (EC 2.7.2.4) (2 copies), diaminopimelate
298 decarboxylase (EC 4.1.1.20), diaminopimelate epimerase (EC 5.1.1.7), L, L-diaminopimelate
299 aminotransferase (EC 2.6.1.83), *N*-acetyl-L, L-diaminopimelate deacetylase (EC 3.5.1.47),
300 UDP-*N*-acetylmuramoylalanyl-D-glutamate-2, 6-diaminopimelate ligase (EC 6.3.2.13) and
301 UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-D-alanyl-D-alanine ligase (EC
302 6.3.2.10). 19 genes/proteins are associated with biosynthesis of polar lipids, including
303 1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51) (2 copies), ABC-type
304 multidrug/protein/lipid transport system, ATPase component, acyl carrier protein (3 copies),
305 acyl-phosphate:glycerol-3-phosphate O-acyltransferase PlsY, cardiolipin synthetase (EC 2.7.8.-) (2
306 copies), CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5),
307 dihydroxyacetone kinase family protein, glycerate kinase (EC 2.7.1.31), glycerol kinase (EC
308 2.7.1.30), glycerol-3-phosphate dehydrogenase [NAD(P)$^+$] (EC 1.1.1.94), octaprenyl diphosphate
309 synthase (EC 2.5.1.90) / gimethylallyltransferase (EC 2.5.1.1) / (2E,6E)-farnesyl diphosphate
310 synthase (EC 2.5.1.10) / geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29),
311 phosphate:acyl-ACP acyltransferase PlsX, phosphatidate cytidylyltransferase (EC 2.7.7.41) and
312 phosphatidylglycerophosphatase B (EC 3.1.3.27) (2 copies). 13 genes/proteins are associated with
313 biosynthesis of polyamines, including 5'-methylthioadenosine nucleosidase (EC 3.2.2.16) @
314 S-adenosylhomocysteine nucleosidase (EC 3.2.2.9), ABC transporter, periplasmic spermidine
315 putrescine-binding protein PotD (TC 3.A.1.11.1), agmatinase (EC 3.5.3.11), arginine
316 decarboxylase (EC 4.1.1.19), arginine/ornithine antiporter ArcD (3 copies), carboxynorspermidine
317 decarboxylase, putative (EC 4.1.1.-), carboxynorspermidine dehydrogenase, putative (EC 1.1.1.-),
318 putrescine transport ATP-binding protein PotA (TC 3.A.1.11.1), spermidine putrescine ABC

transporter permease component PotB (TC 3.A.1.11.1), spermidine putrescine ABC transporter permease component potC (TC..3.A.1.11.1) and spermidine synthase (EC 2.5.1.16). 2 genes/proteins are associated with biosynthesis of teichoic and lipoteichoic acids, including cell wall teichoic acid glycosylation protein gtcA and teichoic acid export ATP-binding protein TagH (EC 3.6.3.40). 16 genes/proteins are associated with biosynthesis of quinones, including 2-heptaprenyl-1,4-naphthoquinone methyltransferase (EC 2.1.1.163), electron transport complex protein RnfA (2 copies), electron transport complex protein RnfB, electron transport complex protein RnfC, electron transport complex protein RnfD (2 copies), electron transport complex protein RnfE (2 copies), electron transport complex protein RnfG, heptaprenyl diphosphate synthase component I (EC 2.5.1.30), microsomal dipeptidase (EC 3.4.13.19), octaprenyl diphosphate synthase (EC 2.5.1.90) / dimethylallyltransferase (EC 2.5.1.1) / (2E,6E)-farnesyl diphosphate synthase (EC 2.5.1.10) / geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) ubiquinone/menaquinone biosynthesis methyltransferase UbiE (EC 2.1.1.-) @ 2-heptaprenyl-1,4-naphthoquinone methyltransferase MenG (EC 2.1.1.163) (2 copies) and undecaprenyl diphosphate synthase (EC 2.5.1.31). There are no genes responsible for biosynthesis of lipopolysaccharides or mycolic acids. Additional annotations showed that the CM04-06^T genome contains a complete butyrate synthesis pathway, three phage remnants, two intact prophages, and aminoglycoside antibiotic genes.

The type strain, CM04-06^T (=CGMCC 1.5208^T = DSM 103432^T), was isolated from human gut. The G+C content of the genomic DNA is 57.51 mol% as calculated from whole genome sequencing.

Materials and Methods

Origin of bacterial strains

Faeces samples were collected from two healthy donors living in Shenzhen, Guangdong province, China (GPS positioning of the samples collection site is 37°35'37"N/114°15'32"E) and preserved refrigerated and anaerobically until processed. The collection of the samples was approved by the Institutional Review Board on Bioethics and Biosafety of BGI under number BGI-IRB17005-T1. All protocols were in compliance with the Declaration of Helsinki and explicit informed consent was obtained from the participants. 1 g of faecal sample was diluted with 0.1 M PBS (pH 7, supplemented with 0.5% cysteine) and spread onto modified peptone-yeast extract-glucose (MPYG, supplemented with 5g/L sodium acetate in DSMZ 104 medium) agar plates in an anaerobic box (Bactron Anaerobic Chamber, Bactron□-2, shellab, USA). The plates were incubated at 37°C under anaerobic conditions (90% N_2, 5% CO_2 and 5% H_2, v/v) for 3 to 5 days. Single colonies were randomly picked and purified by repetitive subculturing on the new plates containing the same medium and incubated under the same conditions as described above. Among the pure cultures, two isolates, designated as AF52-21^T and CM04-06^T, respectively, were obtained and subsequently maintained in 20% (v/v) glycerol and frozen at −80°C.

Phenotypic characterization

The morphological characteristics of strains AF52-21^T and CM04-06^T were performed on cultures grown on MPYG medium at 37°C. Bacterial cell shape was examined by phase contrast microscopy (Olympus BX51, Japan) during the exponential phase of growth. Cell motility was examined using semi-solid MPYG medium containing 0.5% agar[41]. The Gram reaction was carried out using a Gram-staining kit (Solarbio, China). Spore formation and presence of flagella

were determined by staining using spore stain kit and flagella stain kit supplied by Solarbio (China) following the manufacturer's instructions. Colony morphology was observed following growth of the cultures on PYG agar for 2 days at 37°C. Optimal temperature for growth was determined using growth in MPYG medium at 4, 10, 20, 25, 30, 35, 37, 45 and 50°C for 7 days. The pH range for growth was also measured in MPYG medium covering the range of pH 3.0–10.0 (at interval of 0.5 pH units) at 37°C for 7 days. Growth at various NaCl concentrations (0-6%, in increments of 1.0%) was performed for determining tolerance to NaCl. Bile tolerance was measured at different bile salt concentrations (0-5%, at an interval of 1.0%) in the MPYG medium. Catalase activity was assessed by gas formation after dropping the fresh cells in 3% H_2O_2 solution. Biochemical properties, including utilization of substrates, acid production from carbohydrates, enzyme activities, hydrolytic activities, were determined using the API 20A, API 50CHL and API ZYM systems ((bioMérieux Inc., Marcy-l'Étoile, France) according to the manufacturer's instructions with modification by adding sodium acetate at concentration of 0.5% in all tests. The reference type strain was tested under the same condition with strains AF52-21^T and CM04-06^T.

Chemotaxonomic characteristics

Chemotaxonomic features were investigated by analyses of cellular fatty acids. Biomasses of strains AF52-21^T and CM04-06^T were harvested from cells growing in MPYG at 37°C for 2 days. Whole cell fatty acid methyl esters (FAMEs) were extracted, separated and identified according to the MIDI Microbial Identifications System and performed by CGMGG (China General Microbiological Culture Collection Center, Beijing, China) identification service.

Fermentation products analysis

For analysis the metabolic end products from glucose fermentation, including SCFAs and organic

acids, cells were cultured in MPYG broth at 37°C for 2 days. Supernatant harvested from the cultures centrifuged at 10000g for 10min was used for determining SCFAs and organic acids. SCFAs detection was performed using a gas chromatograph (GC-7890B, Agilent) equipped with a flame ionization detector (FID) and capillary column packed with Agilent 19091N-133HP-INNOWax porapak HP-INNOWax (30m × 0.25mm × 0.25um). Organic acids were analysed by equipping capillary column packed with Agilent 122-5532G DB-5ms (40m × 0.25mm × 0.25um).

PCR of bacterial 16S rRNA genes and phylogenetic analysis

Total genomic DNA of strains AF52-21^T and CM04-06^T were extracted using the standard phenol:chloroform method as described by Cheng and Jiang[42]. The complete 16S rRNA genes were amplified and sequenced according to the method previously described[43]. Primers used for amplification of 16S rRNA genes were 27f (5'-AGAGTTTGATCATGGCTCAG-3') and 1492r (5'-TAGGGTTACCTTGTTACGACTT-3'). The obtained 16S rRNA gene sequences of strains AF52-21^T and CM04-06^T were compared with the sequences of type strains retrieved from the EzBioCloud database (https://www.ezbiocloud.net/)[44] using the BLAST program to determine the nearest phylogenetic neighbours and 16S rRNA gene sequence similarity values. Phylogenetic trees were reconstructed by using the neighbour-joining method[45], maximum-likelihood method[46] and minimum-evolution method[47] with the MEGA X program package[48], after Clustal W multiple alignment of the sequences. Robustness of the phylogenetic trees was evaluated by using the bootstrap resampling method (1000 resamplings) of Felsenstein[49].

Genome sequencing , assembly, and annotation of isolates

For genome sequences of strains AF52-21^T and CM04-06^T, genome DNA was prepared following

the method described above. The draft genome was sequenced on an Ion Proton Technology (Life Technologies) platform at BGI-Shenzhen (Shenzhen, China) after constructing a paired-end DNA library with insert size of 500 bp. The resulting reads were assembled using the SOAPdenovo 2 package[50]. CheckM (v1.1.2) was used to estimate genome completeness and contamination[51]. Genome assemblies were visualized using CGView Server[52] (http://stothard.afns.ualberta.ca/cgview_server/index.html). Annotation of the assembled genome was performed using the Rapid Annotation Using Subsystem Technology (RAST) server[53] and COG database[54]. The G+C content in genomic DNA was calculated from the whole genome sequence. The genes in known pathways from acetyl-CoA to butyrate were annotated by BLAST (evalue=1e-5, identity≥60%, coverage≥90%)[33]. A search for prophages was performed by PHAST (http://phast.wishartlab.com/)[55]. Antibiotic resistance was analysed using the CARD database[56].

Average nucleotide identities

Genome relatedness was investigated by calculating average nucleotide identity (ANI)[57], with a value of 95-96% proposed for delineating bacterial species, corresponding to the traditional 70% DNA–DNA reassociation standard[58,59]. The ANI values between strains AF52-21^T, CM04-06^T, and closely related species were determined using the FastANI[60].

References

1 Russell, S. L. *et al*. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep* **13**, 440-447, doi:10.1038/embor.2012.32 (2012).

2 Costello, E. K. *et al*. Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694-1697, doi:10.1126/science.1177486 (2009).

3 Robles Alonso, V. & Guarner, F. Linking the gut microbiota to human health. *Br J Nutr* **109 Suppl 2**, S21-26, doi:10.1017/S0007114512005235 (2013).

4 Tagliabue, A. & Elli, M. The role of gut microbiota in human obesity: recent findings and future perspectives. *Nutr Metab Cardiovasc Dis* **23**, 160-168,

433 doi:10.1016/j.numecd.2012.09.002 (2013).
434 5 Ringel-Kulka, T. *et al.* Intestinal microbiota in healthy U.S. young children and adults--a high
435 throughput microarray analysis. *PLoS One* **8**, e64315, doi:10.1371/journal.pone.0064315
436 (2013).
437 6 Voreades, N., Kozil, A. & Weir, T. L. Diet and the development of the human intestinal
438 microbiome. *Front Microbiol* **5**, 494, doi:10.3389/fmicb.2014.00494 (2014).
439 7 Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait
440 shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107**,
441 18933-18938, doi:10.1073/pnas.1007028107 (2010).
442 8 Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly.
443 *Nature* **488**, 178-184, doi:10.1038/nature11319 (2012).
444 9 De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study
445 in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* **107**, 14691-14696,
446 doi:10.1073/pnas.1005963107 (2010).
447 10 Antunes, L. C. & Finlay, B. B. A comparative analysis of the effect of antibiotic treatment and
448 enteric infection on intestinal homeostasis. *Gut Microbes* **2**, 105-108 (2011).
449 11 Wilson, I. D. & Nicholson, J. K. The role of gut microbiota in drug response. *Curr Pharm Des*
450 **15**, 1519-1523 (2009).
451 12 Adlerberth, I. Factors influencing the establishment of the intestinal microbiota in infancy.
452 *Nestle Nutr Workshop Ser Pediatr Program* **62**, 13-29; discussion 29-33,
453 doi:10.1159/000146245 (2008).
454 13 DiBaise, J. K. *et al.* Gut microbiota and its possible relationship with obesity. *Mayo Clin Proc*
455 **83**, 460-469, doi:10.4065/83.4.460 (2008).
456 14 Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes
457 associated with obesity. *Nature* **444**, 1022-1023, doi:10.1038/4441022a (2006).
458 15 Parekh, P. J., Arusi, E., Vinik, A. I. & Johnson, D. A. The role and influence of gut microbiota
459 in pathogenesis and management of obesity and metabolic syndrome. *Front Endocrinol*
460 *(Lausanne)* **5**, 47, doi:10.3389/fendo.2014.00047 (2014).
461 16 Larsen, N. *et al.* Gut microbiota in human adults with type 2 diabetes differs from
462 non-diabetic adults. *PLoS One* **5**, e9085, doi:10.1371/journal.pone.0009085 (2010).
463 17 Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes.
464 *Nature* **490**, 55-60, doi:10.1038/nature11450 (2012).
465 18 Conte, M. P. *et al.* Gut-associated bacterial microbiota in paediatric patients with
466 inflammatory bowel disease. *Gut* **55**, 1760-1767, doi:10.1136/gut.2005.078824 (2006).
467 19 Joossens, M. *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and
468 their unaffected relatives. *Gut* **60**, 631-637, doi:10.1136/gut.2010.223263 (2011).
469 20 Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma-carcinoma
470 sequence. *Nat Commun* **6**, 6528, doi:10.1038/ncomms7528 (2015).
471 21 Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal
472 cancer. *Nat Rev Microbiol* **12**, 661-672, doi:10.1038/nrmicro3344 (2014).
473 22 Aron-Wisnewsky, J., Gaborit, B., Dutour, A. & Clement, K. Gut microbiota and non-alcoholic
474 fatty liver disease: new insights. *Clin Microbiol Infect* **19**, 338-348,
475 doi:10.1111/1469-0691.12140 (2013).
476 23 Gkolfakis, P., Dimitriadis, G. & Triantafyllou, K. Gut microbiota and non-alcoholic fatty liver

disease. *Hepatobiliary Pancreat Dis Int* **14**, 572-581 (2015).

24 Ríos-Covián, D. *et al.* Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Frontiers in Microbiology* **7**, doi:10.3389/fmicb.2016.00185 (2016).

25 Puertollano, E., Kolida, S. & Yaqoob, P. Biological significance of short-chain fatty acid metabolism by the intestinal microbiome. *Curr Opin Clin Nutr Metab Care* **17**, 139-144, doi:10.1097/MCO.0000000000000025 (2014).

26 Quevrain, E. *et al.* Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii, a commensal bacterium deficient in Crohn's disease. *Gut* **65**, 415-425, doi:10.1136/gutjnl-2014-307649 (2016).

27 Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**, 16731-16736, doi:10.1073/pnas.0804812105 (2008).

28 Duncan, S. H., Hold, G. L., Harmsen, H. J., Stewart, C. S. & Flint, H. J. Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. *Int J Syst Evol Microbiol* **52**, 2141-2146, doi:10.1099/00207713-52-6-2141 (2002).

29 Fitzgerald, C. B. *et al.* Comparative analysis of Faecalibacterium prausnitzii genomes shows a high level of genome plasticity and warrants separation into new species-level taxa. *BMC Genomics* **19**, 931, doi:10.1186/s12864-018-5313-6 (2018).

30 Chang Liu, M.-X. D., Rexiding Abuduaini, Hai-Ying Yu, Dan-Hua Li, Yu-Jing Wang, Nan Zhou, Min-Zhi Jiang, Peng-Xia Niu, Shan-Shan Han, Hong-He Chen, Wen-Yu Shi, Linhuan Wu, Yu-Hua Xin, Juncai Ma, Yuguang Zhou, Cheng-Ying Jiang, Hong-Wei Liu, Shuang-Jiang Liu. Enlightening the Taxonomy Darkness of Human Gut Microbiomes With Cultured Biobank. *Research Square*, doi:10.21203/rs.3.rs-74101/v1 (2020).

31 Miquel, S. *et al.* Faecalibacterium prausnitzii and human intestinal health. *Curr Opin Microbiol* **16**, 255-261, doi:10.1016/j.mib.2013.06.003 (2013).

32 Rossi-Tamisier, M., Fournier, P.-E., Benamar, S. & Raoult, D. Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *International Journal of Systematic and Evolutionary Microbiology* **65**, 1929-1934, doi:10.1099/ijs.0.000161 (2015).

33 Vital, M., Howe, A. C. & Tiedje, J. M. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5**, e00889, doi:10.1128/mBio.00889-14 (2014).

34 Zou, Y. *et al.* 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol* **37**, 179-185, doi:10.1038/s41587-018-0008-8 (2019).

35 Zhou, L. *et al.* Faecalibacterium prausnitzii Produces Butyrate to Maintain Th17/Treg Balance and to Ameliorate Colorectal Colitis by Inhibiting Histone Deacetylase 1. *Inflamm Bowel Dis* **24**, 1926-1940, doi:10.1093/ibd/izy182 (2018).

36 Fujimoto, T. *et al.* Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. *J Gastroenterol Hepatol* **28**, 613-619, doi:10.1111/jgh.12073 (2013).

37 Machiels, K. *et al.* A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* **63**, 1275-1283, doi:10.1136/gutjnl-2013-304833 (2014).

38 Lopez-Siles, M. *et al.* Alterations in the Abundance and Co-occurrence of Akkermansia muciniphila and Faecalibacterium prausnitzii in the Colonic Mucosa of Inflammatory Bowel

Disease Subjects. *Front Cell Infect Microbiol* **8**, 281, doi:10.3389/fcimb.2018.00281 (2018).

39 Chang, C. J. *et al.* Next generation probiotics in disease amelioration. *J Food Drug Anal* **27**, 615-622, doi:10.1016/j.jfda.2018.12.011 (2019).

40 De Filippis, F., Pasolli, E. & Ercolini, D. Newly Explored Faecalibacterium Diversity Is Connected to Age, Lifestyle, Geography, and Disease. *Curr Biol*, doi:10.1016/j.cub.2020.09.063 (2020).

41 Tittsler, R. P. & Sandholzer, L. A. The Use of Semi-solid Agar for the Detection of Bacterial Motility. *J Bacteriol* **31**, 575-580 (1936).

42 Cheng, H. R. & Jiang, N. Extremely rapid extraction of DNA from bacteria and yeasts. *Biotechnol Lett* **28**, 55-59, doi:10.1007/s10529-005-4688-z (2006).

43 Zou, Y. *et al.* Lactobacillus shenzhenensis sp. nov., isolated from a fermented dairy beverage. *Int J Syst Evol Microbiol* **63**, 1817-1823, doi:10.1099/ijs.0.041111-0 (2013).

44 Yoon, S. H. *et al.* Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* **67**, 1613-1617, doi:10.1099/ijsem.0.001755 (2017).

45 Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425 (1987).

46 Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368-376 (1981).

47 Rzhetsky, A. & Nei, M. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol* **10**, 1073-1095 (1993).

48 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-1549, doi:10.1093/molbev/msy096 (2018).

49 Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791, doi:10.2307/2408678 (1985).

50 Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18, doi:10.1186/2047-217X-1-18 (2012).

51 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

52 Grant, J. R. & Stothard, P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* **36**, W181-184, doi:10.1093/nar/gkn179 (2008).

53 Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75, doi:10.1186/1471-2164-9-75 (2008).

54 Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* **43**, D261-269, doi:10.1093/nar/gku1223 (2015).

55 Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. PHAST: a fast phage search tool. *Nucleic Acids Res* **39**, W347-352, doi:10.1093/nar/gkr485 (2011).

56 Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* **45**, D566-D573, doi:10.1093/nar/gkw1004 (2017).

57 Richter, M. & Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic

species definition. *Proc Natl Acad Sci U S A* **106**, 19126-19131, doi:10.1073/pnas.0906412106 (2009).

58 Goris, J. *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81-91, doi:10.1099/ijs.0.64483-0 (2007).

59 Kim, M., Oh, H. S., Park, S. C. & Chun, J. 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, doi:10.1099/ijs.0.059774-0 (2014).

60 Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 5114, doi:10.1038/s41467-018-07641-9 (2018).

61 Guo, X. *et al.* CNSA: a data repository for archiving omics data. *Database (Oxford)* **2020**, doi:10.1093/database/baaa055 (2020).

62 Chen, F. Z. *et al.* CNGBdb: China National GeneBank DataBase. *Yi Chuan* **42**, 799-809, doi:10.16288/j.yczz.20-080 (2020).

Acknowledgment

This work was supported by grants from National Key Research and Development Program of China (No. 2018YFC1313801) and Natural Science Foundation of Guangdong Province, China (No. 2019B020230001). We also thank the colleagues at BGI-Shenzhen for sample collection, and discussions, and China National Genebank (CNGB) Shenzhen for DNA extraction, library construction, and sequencing.

Author contributions

Conceived and designed the experiments:Y.Z. and L.X. Performed the experiments: Y.Z., W.X., and Y.D. Analyzed the data: Y.Z., L.X., and X.L. Contributed reagents/materials/analysis tools: Y.Z., W.X.. and Y.D. Wrote the paper: Y.Z. and X.L. Revised the paper: K.K.

Data Availability Statement

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene sequences determined in this study are: AF52-21^{T} (KX146426) and CM04-06^{T} (KX150462). The data of draft genome

sequences have been deposited into CNGB Sequence Archive (CNSA) [61] of China National GeneBank DataBase (CNGBdb) [62] with accession number CNA0017730 and CNA0017731 for strains AF52-21^{T} and CM04-06^{T}, respectively.

Figure Legends

Figure 1. Micrographs of strains AF52-21^T, CM04-06^T after Gram staining.

A, AF52-21^T; B,CM04-06^T.

Figure 2. Neighbour-joining phylogenetic tree based on 16S rRNA gene sequences showing the phylogenetic relationships of strains AF52-21^T, CM04-06^T and the representatives of several other related taxa within the family *Ruminococcaceae*.

Clostridium butyricum DSM 10702^T (AQQF01000149) was used as an out-group. Bootstrap values based on 1000 replications higher than 70% are shown at the branching points. Bar, substitutions per nucleotide position.

Figure 3. Circular map of AF52-21^T.

Innermost circle, GC skew; circle 2, G+C content; circle 3, contigs; circles 4, predicted prophage remnants; circle 5, tmRNA, tRNA and rRNA genes; circles 6, CDS; circles 7-9, homologous genomic segments from CM04-06^T, *F. prausnitzii* ATCC 27768^T and *F. hominis* 4P-15^T.

Figure 4. Circular map of CM04-06^T.

Innermost circle, GC skew; circle 2, G+C content; circle 3, contigs; circles 4, predicted prophage remnants; circle 5, tmRNA, tRNA and rRNA genes; circles 6, CDS; circles 7-9, homologous genomic segments from AF52-21^T, *F. prausnitzii* ATCC 27768^T and *F.hominis* 4P-15^T.

Figure 5. Comparison of COG functional categories of strains AF52-21^T, CM04-06^T and the

620 **closest related species *F. prausnitzii* ATCC 27768^{T}.**

621

622 **Figure 6. The synthesis pathways from Acetyl-CoA to Butyrate. Strains AF52-21^{T},**

623 **CM04-06^{T} and ATCC 27768^{T} were annotated as blue, red, and yellow.**

624 Thl, thiolase; Hdb, β-hydroxybutyryl-CoA dehydrogenase; Cro, crotonase; Bcd, butyryl-CoA

625 dehydrogenase; But, butyryl-CoA:acetate CoA transferase; Ptb, phosphate butyryltransferase; Buk,

626 butyrate kinase.

627

628 **Figure 7. Distribution of prophage in strains AF52-21T and CM04-06T.**

629

630 **Figure 8. Comparison of resistance genes in strains AF52-21^{T}, CM04-06^{T}, and *F.***

631 ***prausnitzii* ATCC 27768^{T}.**

632

633

634

635

636

637

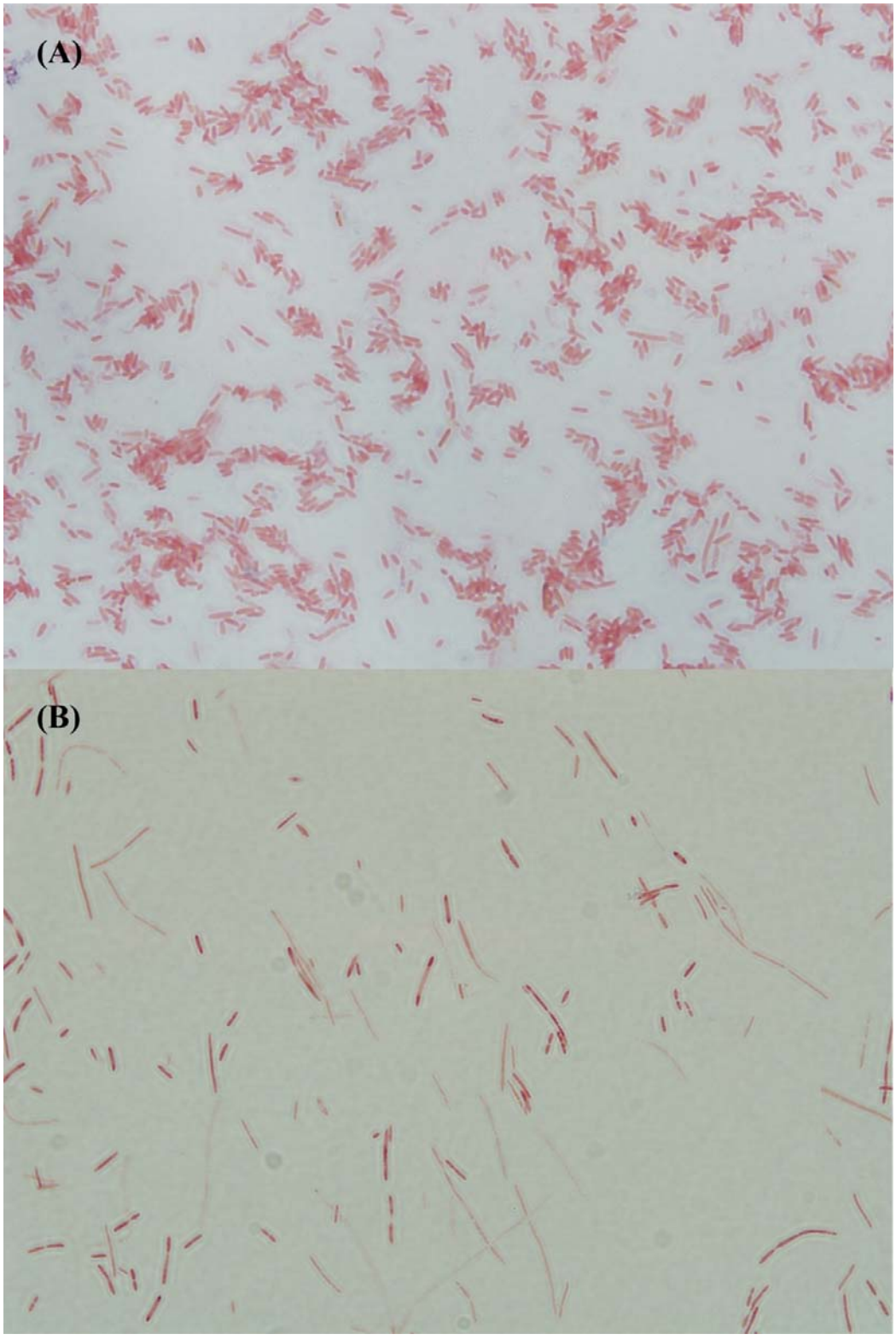
638

639

640

641

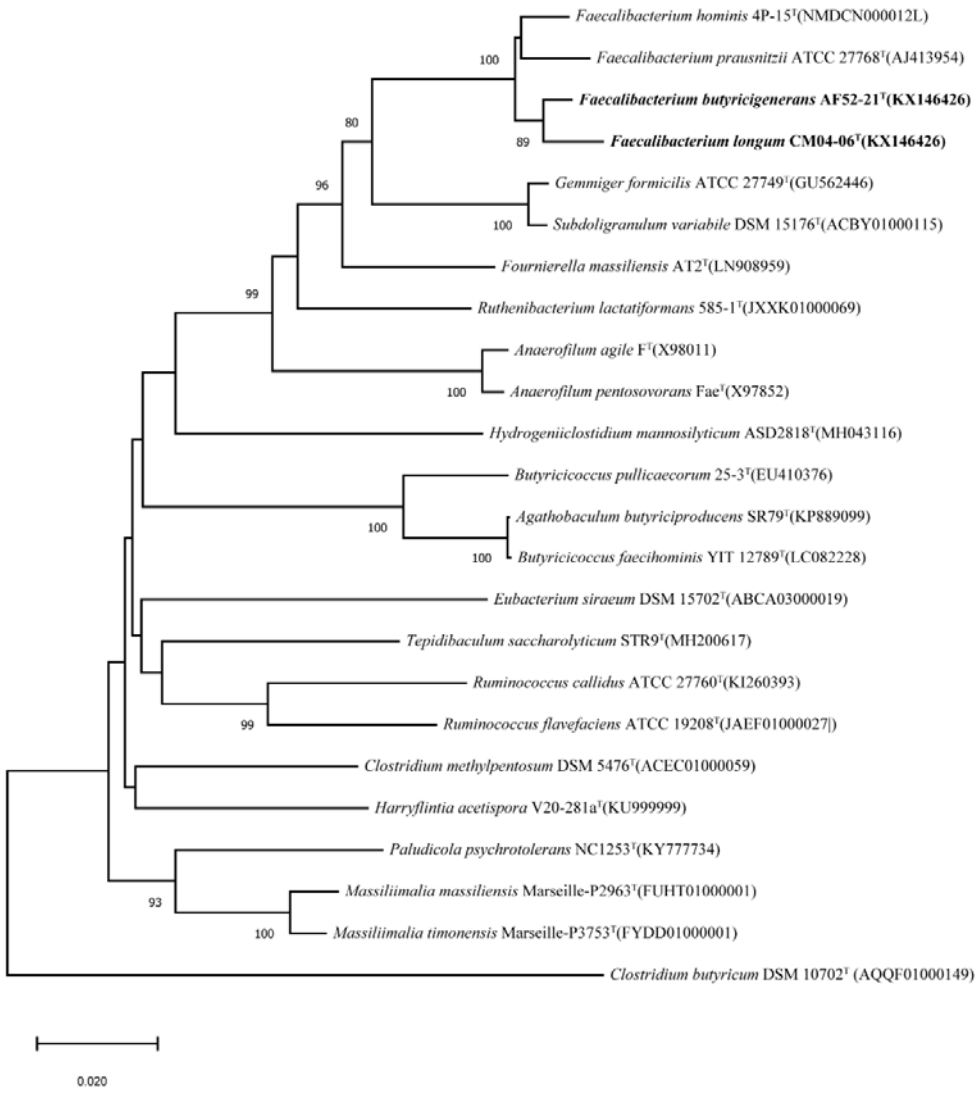
642 **Figure 1**



643

644

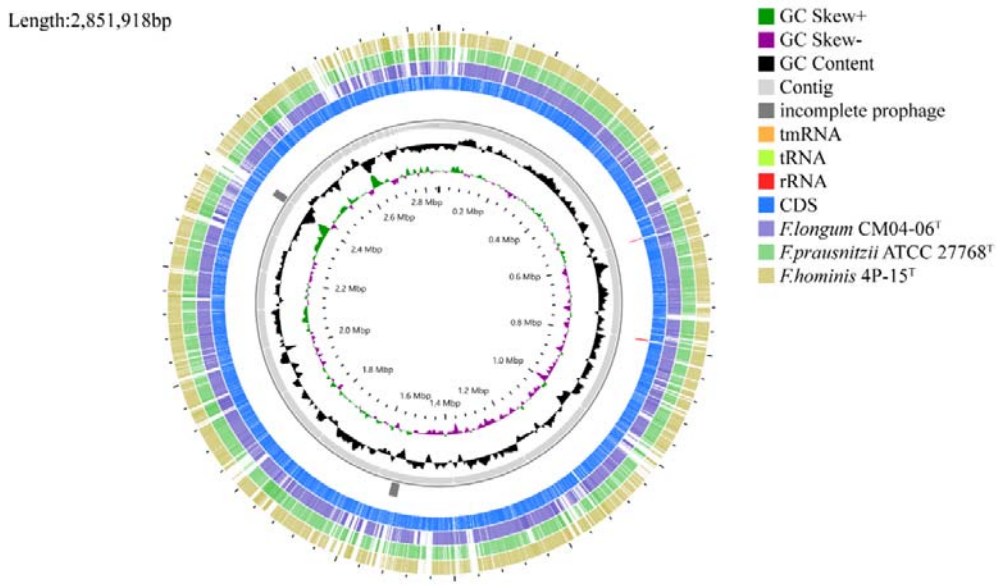
645 **Figure 2**



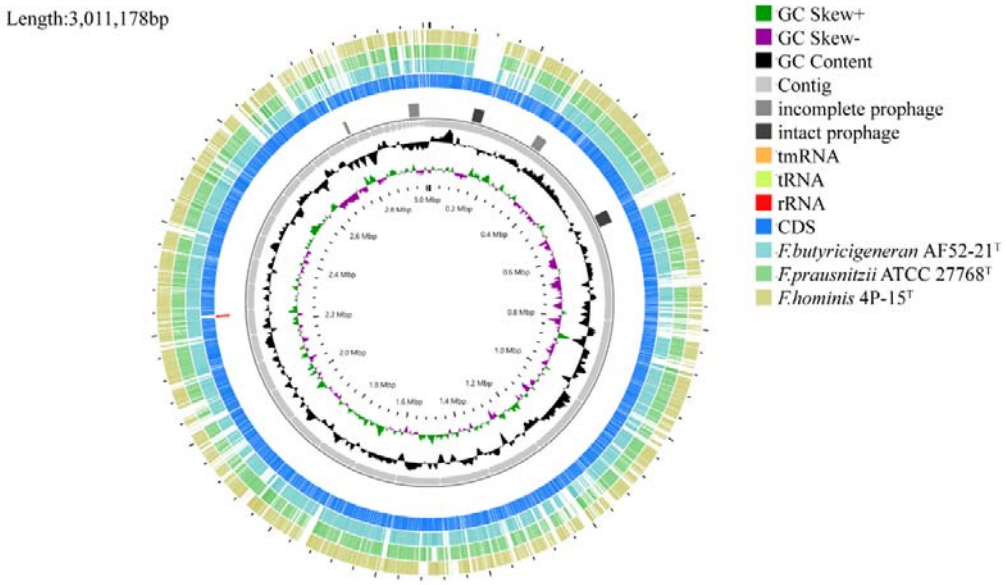
646

647

Figure 3

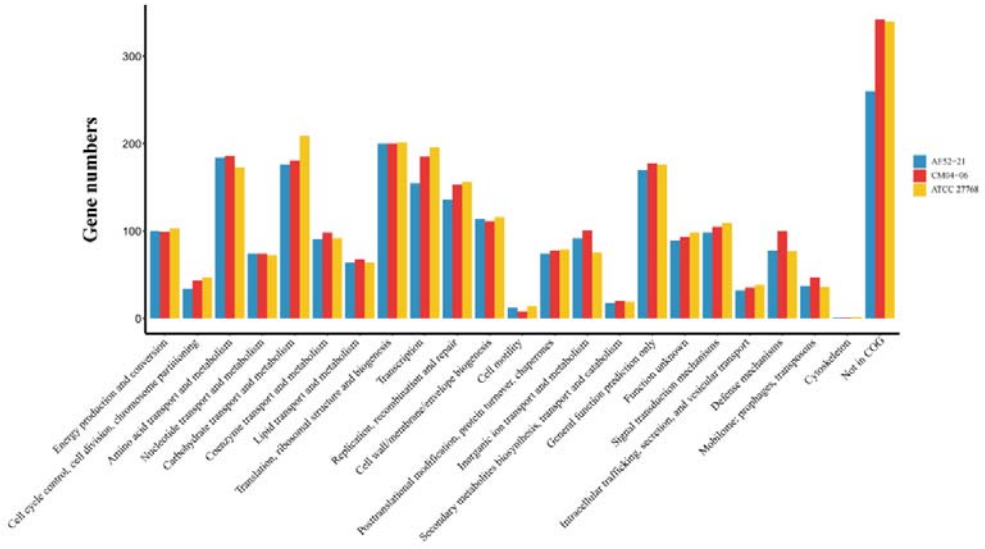


651 **Figure 4**



652

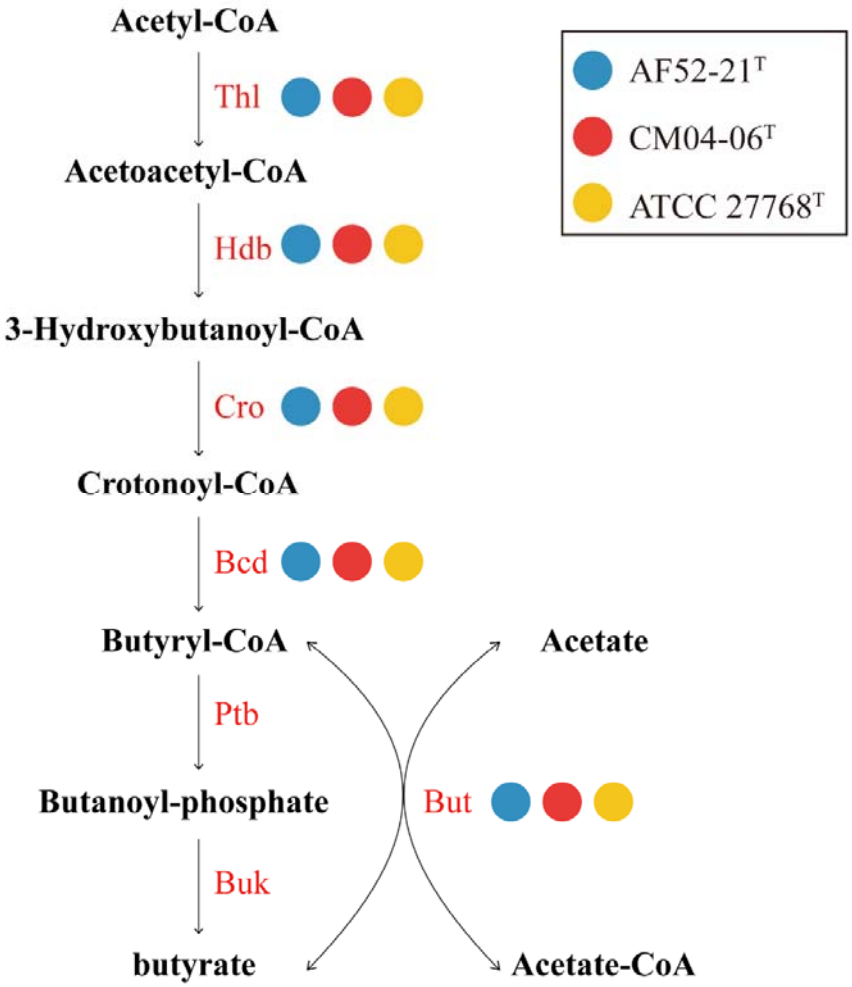
653 **Figure 5**



654

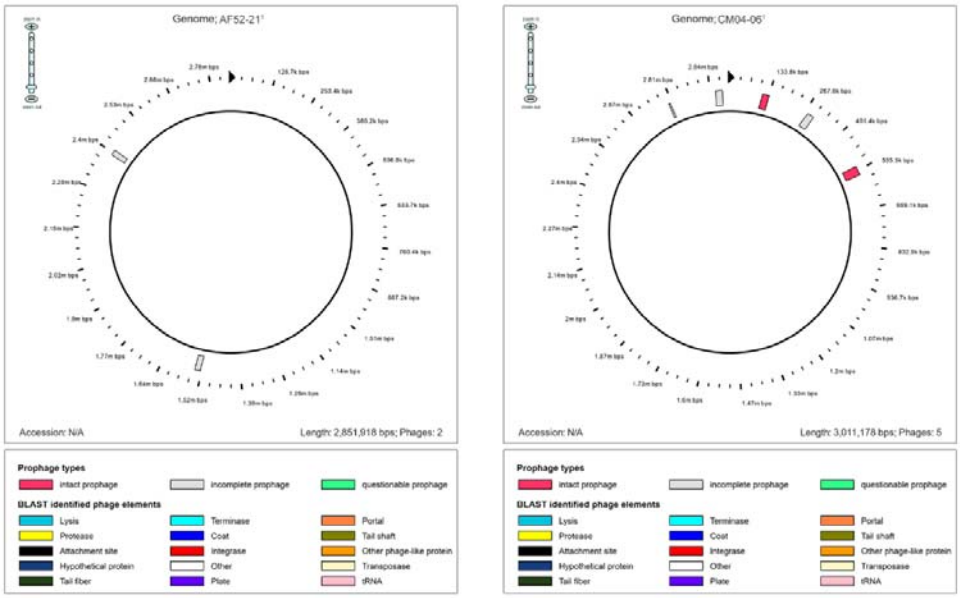
655

656 **Figure 6**



657
658

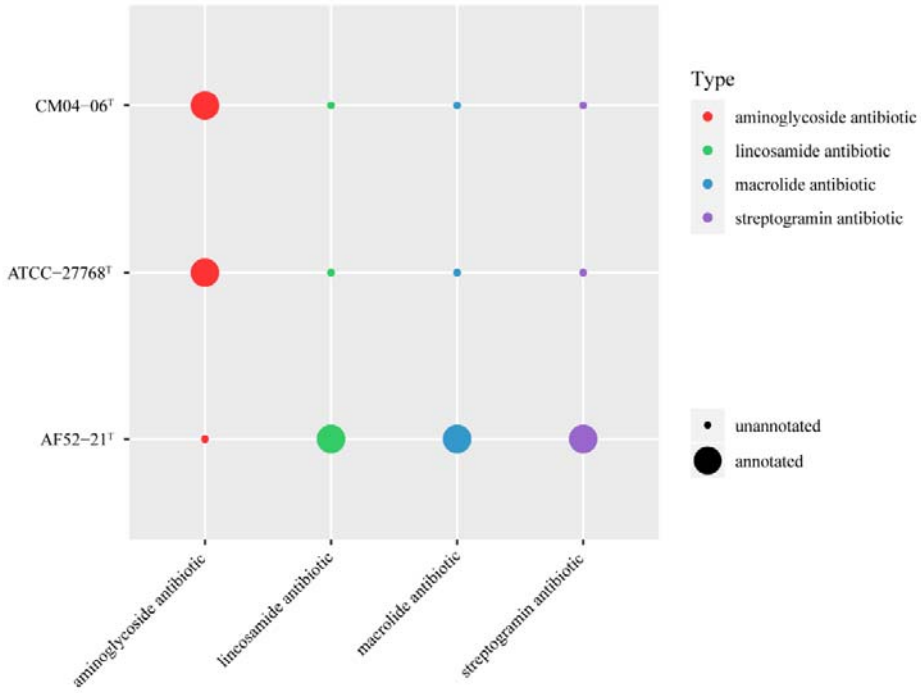
659 **Figure 7**



660

661

662 **Figure 8**



663

664

Supplementary Material

Supplementary Table S1. Number of genes associated with general COG functional categories in the genome of *F. butyricigenerans* AF52-21^T and *F. longum* CM04-06^T.

Supplementary Table S2. The specific genes/protein related to biosynthesis of DAP, polar lipids, polyamines and lipoteichoic and teichoic acids and their positions in the genome in comparation of strains AF52-21^T, CM04-06 ^{T} and related organism, ATCC 27768^T identified by Rapid Annotation Subsystem Technology (RAST).

Supplementary Figure S1. Maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences showing the phylogenetic relationships of strains AF52-21^T, CM04-06^T and the representatives of related taxa. *Clostridium butyricum* DSM 10702^T (AQQF01000149) was used as an out-group. Bootstrap values based on 1000 replications higher than 70% are shown at the branching points. Bar, substitutions per nucleotide position.

Supplementary Figure S2. Minimum-evolution phylogenetic tree based on 16S rRNA gene sequences showing the phylogenetic relationships of strains AF52-21^T, CM04-06^T and the representatives of related taxa. *Clostridium butyricum* DSM 10702^T (AQQF01000149) was used as an out-group. Bootstrap values based on 1000 replications higher than 70% are shown at the branching points. Bar, substitutions per nucleotide position.

Supplementary Figure S3. Certification. Deposit certification of AF52-21^T in CGMCC.

686 **Supplementary Figure S4. Certification. Deposit certification of AF52-21^{T} in DSMZ.**

687 **Supplementary Figure S5. Certification. Deposit certification of CM04-06^{T} in CGMCC.**

688 **Supplementary Figure S6. Certification. Deposit certification of CM04-06^{T} in DSMZ.**