

Journal Pre-proof

Comparative genomics and gene-trait matching analysis of *Bifidobacterium breve* from Chinese children

Rui Liu, Bo Yang, Catherine Stanton, R. Paul Ross, Jianxin Zhao, Hao Zhang, Wei Chen



PII: S2212-4292(19)30650-9

DOI: <https://doi.org/10.1016/j.fbio.2020.100631>

Reference: FBIO 100631

To appear in: *Food Bioscience*

Received Date: 25 June 2019

Revised Date: 4 May 2020

Accepted Date: 4 May 2020

Please cite this article as: Liu R., Yang B., Stanton C., Paul Ross R., Zhao J., Zhang H. & Chen W., Comparative genomics and gene-trait matching analysis of *Bifidobacterium breve* from Chinese children, *Food Bioscience* (2020), doi: <https://doi.org/10.1016/j.fbio.2020.100631>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

Author Contribution Statement

Rui Liu: Methodology, Validation, Investigation, Data Curation, Writing-Original Draft
Preparation

Bo Yang: Conceptualization, Validation, Investigation, Data Curation, Writing-Original Draft
Preparation, Writing-Review & Editing, Visualization, Supervision

Catherine Stanton: Software, Formal Analysis, Writing-Review & Editing

R. Paul Ross: Formal Analysis, Writing-Review & Editing

Jianxin Zhao: Methodology, Software, Validation

Hao Zhang: Conceptualization, Supervision, Writing-Review & Editing, Project Administration

Wei Chen: Conceptualization, Supervision, Writing-Review & Editing, Funding Acquisition

1 **Comparative genomics and gene-trait matching analysis of *Bifidobacterium breve***
2 **from Chinese children**

3 Rui Liu ^{1,2}, Bo Yang ^{1,2,3,*}, Catherine Stanton ^{3,4,5}, R. Paul Ross ^{3,5}, Jianxin Zhao ^{1,2,6},
4 Hao Zhang ^{1,2,6,7}, Wei Chen ^{1,2,6,8}

5 ¹ State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi,
6 Jiangsu, China

7 ² School of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu, China

8 ³ International Joint Research Center for Probiotics & Gut Health, Jiangnan University,
9 Wuxi, Jiangsu, China

10 ⁴ Teagasc Food Research Centre, Moorepark, Fermoy, Cork, Ireland

11 ⁵ APC Microbiome Ireland, University College Cork, Cork, Ireland

12 ⁶ National Engineering Research Center for Functional Food, Jiangnan University,
13 Wuxi, Jiangsu, China

14 ⁷ Wuxi Translational Medicine Research Center and Jiangsu Translational Medicine
15 Research Institute Wuxi Branch, Wuxi, Jiangsu, China

16 ⁸ Beijing Innovation Center of Food Nutrition and Human Health, Beijing Technology
17 and Business University (BTBU), Beijing, China

18

19 Running title: *B. breve* comparative genomics

20

21 *Correspondence: Bo Yang

22 School of Food Science and Technology, Jiangnan University, 1800 Lihu Avenue,
23 Wuxi 214122, Jiangsu, China.

24 E-mail: bo.yang@jiangnan.edu.cn

25 Tel/Fax: +86 510 8591 2155

26 **Abstract**

27 *Bifidobacterium breve* is one of the dominant Bifidobacterial species in children.
28 In the current work, 46 strains of *B. breve* isolated from fecal samples of Chinese
29 children were analyzed using whole-genome sequencing and comparative genomics
30 to explore their genetic diversity, as well as genotype and phenotype analysis for
31 carbohydrate utilization and antibiotic tolerance. The phylogenetic tree was
32 independent of region, age and feeding mode, and without any regularity in the
33 clustering of carbohydrates and antibiotics at the genetic level. Based on
34 genotypic-phenotypic correlation analysis, the diversity of glycosyl hydrolases and
35 the ability of strains to metabolize carbohydrates corroborated the predominance of *B.*
36 *breve* in the children's intestines. Simultaneously, the sensitivity of strains to
37 antibiotics increased the understanding of its genetic features and provided a potential
38 basis for safety evaluation.

39 **Keywords:** *Bifidobacterium breve*; comparative genomics; carbohydrate metabolism;
40 antibiotic resistance; genomic diversity

41 1. Introduction

42 *Bifidobacterium* is one of the commensal microorganisms in the colon and one of the
43 first bacterial colonizers to settle in the human intestine after birth, which has been
44 related to maintaining intestinal health (Arboleya et al., 2016). The beneficial effects
45 of Bifidobacteria include intestinal barrier reinforcement, pathogen inhibition,
46 immune system modulation and nutrient supplementation as well as host metabolism
47 enhancement and expansion (Round and Mazmanian, 2009; Tojo et al., 2014; Ventura
48 et al., 2012). In infants especially in neonates, Bifidobacteria, particularly
49 *Bifidobacterium breve*, *B. longum* and *B. bifidum*, made up a large proportion of gut
50 microbes (Turroni et al., 2018).

51 *B. breve* is not only a commonly encountered species in the feces of infants, and also
52 has been frequently isolated from human milk and the vagina. The species represents
53 one of the most extensively researched Bifidobacteria in terms of comparative and
54 functional genomics investigations. Since genome sequence of *B. breve* UCC2003
55 was published first among the species in 2011, a number of other *B. breve* genomes
56 have been sequenced and publically available (Jin et al., 2019).

57 The successful adaptation of *B. breve* to ecological niches is related to its
58 saccharolytic catabolism. *B. breve* could utilize plant-saccharides, polysaccharides,
59 oligosaccharides, and monosaccharides, which reflects flexibility of carbohydrate
60 utilization (Pokusaeva et al., 2011a). Carbohydrate utilization may be correlated with
61 the niche of some strains, e.g., some *B. breve* strains isolated from breast-fed infants
62 showed the ability to metabolize 2'-fucosyllactose (2'-FL), which is the dominant

63 glycan component among human milk oligosaccharides (HMOS) (Castanys-Muñoz et
64 al., 2013). In addition, Bifidobacteria are used in functional foods and medicines as a
65 probiotic. Safety issues related to antibiotic resistance are of concern (Francino, 2015).
66 Gene-trait matching could serve as an effective approach to elucidate genes
67 responsible for a specific phenotype using a combination of comparative genome
68 analysis and experimental data, such as carbohydrate metabolism and antibiotic
69 resistance (Bottacini et al., 2018). The aim of the current study was to analyze the
70 genomic diversity of *B. breve* isolated from Chinese children, and to further compare
71 their carbohydrate utilization and antibiotic resistance both genotypically and
72 phenotypically. Those analyses may help better understand the physiological
73 properties and metabolic characteristics of different *B. breve* strains to guide the
74 screening and application of particular strains (Boesten et al., 2011).

75 **2. Materials and methods**

76 **2.1 Isolation and 16S rRNA sequence analysis**

77 Eighty children's stool samples from different regions of China were collected (Table
78 1). One g of each stool sample was blended with 9 mL sterile physiological saline
79 (Bottacini et al., 2018). Serial dilution and plating were done using an anaerobic
80 workstation (AW400TG, Electrotek Scientific Ltd., Shipley, West Yorkshire, UK). For
81 selection of Bifidobacteria, 100 μ L of diluent was continuously plated on de
82 Man-Rogosa-Sharpe agar (Qingdao Hope Bio-Technology Co. Ltd., Qingdao,
83 Shandong, China) plus 0.05% (v/w) L-cysteine hydrochloride (Sangon Biotech Co.,
84 Ltd., Shanghai, China) (MRSC) plus 100 mg/L mupirocin (Sangon Biotech Co., Ltd.)

85 and 50 U/mL nystatin (Sangon Biotech Co., Ltd.). Agar plates were cultured in the
86 anaerobic workstation flushed with 80% N₂, 10% CO₂ and 10% H₂ at 37 °C for 72 h
87 (Bottacini et al., 2018). For each sample, colonies on MRSC plates were counted.
88 Colonies were selected randomly and re-streaked onto MRSC agar for purity. The
89 final pure culture was cultured in MRSC and preserved in 30% glycerol (Sangon
90 Biotech Co., Ltd.) at -80 °C (Bottacini et al., 2018). Each of the possible
91 Bifidobacteria isolates was identified by a 16S rRNA sequence using the bacterial
92 universal primers (27F: 5'-AGA GTT TGA TCC TGG CTC AG-3' and 1492R:
93 5'-ACG GCT ACC TTG TTA CGA CTT-3'). PCR amplification conditions were:
94 95 °C for 8 min; 95 °C for 40 s, 60 °C for 40 s, 72 °C for 40 s, 30 cycles; 72 °C for 8
95 min (Boesten et al., 2011). The PCR product was sequenced by BGI (Shenzhen,
96 Guangdong, China). All the strains were compared with the NCBI BLAST database
97 (<http://www.ncbi.nlm.nih.gov/BLAST/>) to assign a particular species.

98 **2.2 Sequencing and draft genome assembly**

99 Draft genomes of all the 46 *B. breve* strains were sequenced using an Illumina HiSeq
100 × 10 platform (Majorbio BioTech Co., Ltd., Shanghai, China), with the use of 2 × 150
101 bp paired-end libraries (average read length of ~400 bp). The assembly was done
102 using SOAPdenovo v2.04 software (<https://omictools.com/soapdenovo-tool>), and the
103 partial gap was filled by GapCloser
104 (<https://sourceforge.net/projects/soap-denovo2/files/GapCloser/>) (Liu et al., 2013;
105 Zerbino and Birney, 2008).

106 **2.3 General features prediction**

107 Bacterial gene prediction and open reading frames (ORFS) were done using Glimmer
108 3.02 software (<http://ccb.jhu.edu/software/glimmer/index.shtml>), GeneMarkS v4.30
109 software (<https://www.genemarks.com/>) (Rice et al., 2000) and the predictor Prodigal
110 v2.0 software (<http://prodigal.ornl.gov>). BLASTX v2.2.26 alignment
111 (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST>) was carried out for all the
112 genomes analyzed (Patel and Jain, 2012). The tRNA genes in the genome were
113 predicted using tRNAscan-SE v1.3.1 software (<https://omictools.com/trnascan-se-tool>)
114 (Schattner et al., 2005).

115 **2.4 Comparative genome analysis**

116 All-versus-all BLASTP alignment (50% identity; E -value $1e^{-4}$ cut-off) was done for
117 protein sequences extracted from each strain (Enright et al., 2002). BLAST outputs
118 served as inputs to cluster into families of proteins sharing the same function using the
119 Markov Cluster Algorithm (MCL) with an inflation index of 2.5
120 (<https://micans.org/mcl/>) (Goris et al., 2007). The gene families obtained were divided
121 into core genome or dispensable genome, based on their existence in all or in a subset
122 of the strain investigated. To calculate ANI values for each pair of genomes, an ANI
123 Perl script was implemented (<https://github.com/chjp/ANI/blob/master/ANI.pl>) (Qin
124 et al., 2014).

125 **2.5 Pan-genome and core-genome analysis**

126 To predict possible dynamic changes in the genome, the size of the pan-genome, core
127 genome and unique genes were calculated. Computation of pan-genome and
128 core-genome was done using PGAP v1.2.1 software

129 (<https://sourceforge.net/projects/pgap/files/PGAP-1.2.1>) (Zhao et al., 2012) in
130 accordance with the Heap's law pan-genome model (Zhao et al., 2012). Among them,
131 amino acids had 50% pairwise identity and a 0.7 length difference cut-off threshold
132 (Harris et al., 2017). Then the Venn diagram was drawn to show the relationships
133 between samples.

134 **2.6 Phylogenetic analyses**

135 The homologous genes for all of the 46 strains sequenced, in addition to 9 publicly
136 available *B. breve* genomes, as well as two *B. bifidum* and one *B. longum* genomes
137 from the NCBI GenBank database were analyzed based on Orthomcl v2.0.9 software
138 (<http://orthomcl.org/common/downloads/software/v2.0/>) (Liu and Warnow, 2014).
139 After extraction of orthologous genes, MAFFT alignment was done using
140 MAFFT-7.313-with-extensions (<https://mafft.cbrc.jp/alignment/software>) (Divakar
141 and Crespo, 2015), and finally phylogeny was used to generate a phylogenetic tree.

142 **2.7 Genotype and phenotype of carbohydrate utilization of *B. breve* species**

143 All the genomes were annotated using the HMMSCAN software (hmmer.org) (Tong
144 et al., 2017) in combination with BLASTP. The carbohydrate active enzyme gene
145 profiles were analyzed and compared using the carbohydrate-active enzymes (CAZy)
146 database (<http://www.cazy.org/>). Cluster analysis was done using HEMI software
147 (hemi.biocuckoo.org).

148 Twenty carbohydrates including D-lactose, D-ribose, glucose, FOS, maltose,
149 D-galactose, raffinose, α -L-fucose, D-sorbitol, D-mannitol, D-fructose, sucrose,
150 cellobiose, soluble starch, L-arabinose, D-glucuronate, XOS, 2'-FL, trehalose and

151 D-xylose were further selected for carbohydrate utilization analysis, and all the sugars
152 were purchased from Sangon Biotech Co. Ltd. A 10% (v/w) fresh solution of those
153 carbohydrates was prepared and filtered through a 0.22 μm sterile membrane filter
154 (Saigon Biotech Co. Ltd.). The utilization assay medium was freshly prepared with
155 the same content as MRSC medium except 2% glucose, and bromcresol purple as an
156 indicator (Sangon Biotech Co. Ltd.) were added into the medium. After autoclaving
157 and cooling, each of the sterile carbohydrates solutions were added into the medium at
158 1% final concentration. To test the utilization capacity of each strain, after
159 sub-culturing twice in MRSC medium, a 1% culture was inoculated into the test
160 growth media which was supplemented with a different sugar instead of glucose, and
161 cultured anaerobically at 37 °C for 48 h (Arboleya et al., 2018). Color changes were
162 observed and growth was measured using a microplate reader (Varioskan Lux,
163 Thermo, Waltham, MA, USA) at $\text{OD}_{600\text{nm}}$. All the tests were done in triplicate.

164 **2.8 Genotype and phenotype of antibiotics resistance among *B. breve* species**

165 The antibiotic-resistant genes were analyzed using the comprehensive antibiotic
166 resistance database (CARD) (<https://card.mcmaster.ca>) to obtain information of
167 predicted antibiotic resistance genes encoded by each genome. Cluster analysis was
168 done with Hemi software (hemi.biocuckoo.org) (Masco et al., 2006).

169 According to ISO10932:2010 standard “Milk and milk products-determination of the
170 minimal inhibitory concentration (MIC) of antibiotics applicable to bifidobacteria and
171 non-enterococcal lactic acid bacteria (LAB)”
172 (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=

173 46434)", the MIC value of 6 antibiotics (streptomycin, erythromycin, clindamycin,
174 chloramphenicol, tetracycline, and trimethoprim) (Saigon Biotech Co. Ltd.) were
175 analyzed for all the strains. Microbial dilutions were incubated under anaerobic
176 conditions at 37 °C for 48 h, and finally the absorbance at OD_{625nm} was measured with
177 the microplate reader.

178

179 **3. Results**

180 **3.1 Strains isolation and their general genome features**

181 A total of 400 isolates were obtained from the Bifidobacterial selective agar, and all of
182 them were identified for species using 16S rRNA sequence. Among them, 93 isolates
183 from 46 samples were confirmed as *B. breve*. Other than FFGZ18I1M1 and
184 FFGZ18I1M6, FFGZ19I1M4 and FFGZ19I1M6, only one *B. breve* strain isolate from
185 each sample was used for genome sequencing (Table 1).

186 The draft genome of those isolates were sequenced using an Illumina Hiseq platform
187 and subjected to further analysis with another 9 publicly available *B. breve* genomes
188 (<https://www.ncbi.nlm.nih.gov/genome/genomes/1273>), corresponding to *B. breve*
189 UCC2003 (O'Connell-Motherway et al., 2011b), *B. breve* 12L, *B. breve* 31L, *B. breve*
190 CECT7263 (Jiménez et al., 2012), *B. breve* NCFB2258, *B. breve* ACS071VSCH8b, *B.*
191 *breve* 689b, *B. breve* S27 and *B. breve* JCM7017 (Bottacini et al., 2015). The number
192 of predicted ORFS of each genome ranged from 1748 for *B. breve* S27 to 2396 for
193 FJSWX26M5. On average, 2024 ORFS were identified/genome. The average
194 genomic size was 2.33 Mb, with FFHNXY43M2 being the smallest genome (2.16

195 Mb), and FJSWX26M5 the largest (2.64 Mb). The sequence and assembly results for
196 those strains generated a number of contigs/genome that ranged from 1 to 521. The
197 average G+C% content of all the *B. breve* was $58.8\pm 0.5\%$, and those strains consisted
198 of 51-75 tRNA genes distributed across the genomes (Table 1).

199 **3.2 Comparative analyses**

200 Comparative genome analyses based on 55 genomes of *B. breve* showed the
201 emergence of shared orthologous genes and unique genes. Using BLASTP-mediated
202 comparative genomics methods in conjunction with the use of the MCL algorithm, the
203 results were able to describe a pool of 1,150 gene families shared by those 55
204 genomes, thereby allowing the determination of the predicted core genome (Fig. 1A).
205 Core genes accounted for ~28% of the total gene families. The remaining were
206 presumed to represent the dispensable genome, which indicated the genomic diversity
207 among *B. breve* species (Fig. 1B, 1C).

208 Additionally, the dispensable-gene results showed the presence of 1,411 unique or
209 strain-specific genes, of which 1,244 genes were present in the 46 *B. breve* draft
210 genomes (Fig. 1A). The number of dispensable genes ranged from 108 genes in
211 FFHNFQ49M1 to three genes in FFGZ19I1M4, and most of the genes observed
212 within the distributable genomes reflected the high diversity among *B. breve* strains.
213 The *in silico* method was applied to predict the average nucleotide identity (ANI)
214 value of *B. breve* genomes, and the correlation of ANI values ranged from 97 to 99%,
215 indicating that all those strains belonged to the same species without any potential
216 subspecies (Fig. 1D).

217 **3.3 Pan- and core-genome of *B. breve***

218 To compute the total number of genes in the *B. breve* genomes, the PGAP pipeline
219 was used to do the pan-genome calculation in two steps. First, the pan-genome of all
220 the 55 *B. breve* genomes was computed, showing that the pan-genome trendline was
221 recognized as the total number of gene families. A pan-genome of *B. breve* consisted
222 of 6,707 gene families (Fig. 2A). The number of new genes obtained by adding
223 genomic sequences decreased from 404-100 gene families in the first 14 genome
224 iterations to 45-41 gene families in the last 9 genome additions, thus showing a
225 relatively saturated trend of the pan-genome within *B. breve* species. Second, the
226 core-genome result tended to be stable after the 22nd genomic iteration, reaching the
227 value of 1,111 gene families in the last iteration. (Fig. 2A). The functions of the core
228 genome of *B. breve* were related to secondary metabolites biosynthesis, transportation,
229 amino acid and carbohydrate metabolism and transports systems (Fig. 2B).

230 **3.4 Phylogenomic analyses of *B. breve***

231 A phylogenomic analysis was done to recognize homologous genes between strain
232 genome sequences belonging to the *B. breve* species and other strains, including
233 publicly available *B. longum* and *B. bifidum* genomes (Fig. 3). *B. breve* and other
234 *Bifidobacterium* species were computed based on 997 orthologues, although they had
235 large homologous genes, and *B. longum* combined with *B. bifidum* as a representative
236 outgroup. No correlation was evident from the evolutionary tree within the sampling
237 regions, neither feeding methods nor age of the donors. FFGZ18I1M1 and
238 FFGZ18I1M6 were isolated from the same sample. However, they were unable to

239 cluster together in the phylogenomic tree, and similar results were found for
240 FFGZ19I1M4 and FFGZ19I1M6, which indicated significantly strain-dependent
241 genetic difference even from the same gastrointestinal tract.

242 **3.5 GH families and carbohydrate utilization phenotype analysis**

243 The saccharolytic enzymes of Bifidobacteria are capable of metabolizing large
244 amounts of carbohydrates, ranging from dietary-derived to host-derived carbohydrates
245 (Watson et al., 2013). All the newly sequenced genomes were predicted using the
246 accordance with the CAZy database *in silico*. The pan-genome of *B. breve* species
247 included 9 glycosyl transferase (GT) families, 9 carbohydrate esterase (CE) families,
248 13 carbohydrate-binding modules (CBM) families, one polysaccharide lyase (PL)
249 family (Fig. 4A) and 35 glycosyl hydrolase (GH) families (Fig. 4B). There was no
250 notable correlations between the sugar metabolism gene representation and the
251 sampling regions through cluster analysis (Fig. 4B). Members of the GH13 and GH3
252 family accounted for the largest percentage of GH predicted in the *B. breve* genome
253 (24 and 8%, respectively) (Fig. 4B), consistent with a previous study on
254 *Bifidobacterium* (Kelly et al., 2006). Members of the GH95 and GH29 families acted
255 as representatives, participating in the utilization of fucose-containing substrates
256 (Matsuki et al., 2016). In the current research, *in silico* analysis showed that genes
257 encoding for GH29 and GH95 were present on the genome of a given strain, such as
258 FFHNFQ4M7, FFHNFQ49M1, FHuNCS1M5, FHuNCS6M1, FJSWX5M4,
259 FJSWX17M1, FJSWX24M2, FJSWX26M5, FJSWX23M8, FZJHZ3M2 and
260 FZJHZ7M2, and those strains were mainly isolated from breastfed babies and those

261 <1 yr.

262 To evaluate and verify the *in silico* analyses, 20 carbohydrates were used as a single
263 carbon source, glucose was used as a positive control for *in vitro* growth assays.
264 Comparing carbohydrate utilization of the 46 strains isolated from Chinese children
265 showed that almost all the strains could utilize D-lactose, maltose and FOS. On the
266 other hand, the ability to utilize other sugars such as sucrose, soluble starch,
267 D-glucuronate, 2'-FL, raffinose, D-ribose, trehalose, L-arabinose, cellobiose,
268 D-xylose, D-galactose, D-fructose, XOS and α -L-fucose varied with strain (Fig. 5).
269 All the 46 strains showed similar utilization for D-mannitol and D-sorbitol, which was
270 consistent with a previous report (Bottacini et al., 2018).

271 In addition, 7 gene clusters related to the utilization of 2'-FL, XOS, D-ribose,
272 cellobiose, D-mannitol/D-sorbitol, sucrose and D-galactose were identified (Fig. 6).
273 According to the experimental results, taking FHuNCS6M1 as an example and *B.*
274 *longum* APC1477 as a reference strain (Arboleya et al., 2018), the gene cluster for
275 2'-FL utilization mainly consisted of substrate-binding protein (SBP) of an ABC
276 transporter, two ABC substrate binding proteins and a fucosidase gene (Fig. 6A).
277 Based on the prediction, the GH43 gene family could encode α -L-arabinofuranosidase
278 and β -1,4-xylosidase, which could hydrolyze XOS (Arboleya et al., 2018). According
279 to the combined analysis of genotype and phenotype, FHuNCS6M1 was chosen as
280 representative to speculate on the mechanism of using XOS (Fig. 6B).

281 The gene cluster corresponding to the growth of the strain on cellobiose included
282 genes encoding a ketol-acid reductoisomerase, β -glucosidase, a transcriptional

283 regulator and a ABC transporter system (Pokusaeva et al., 2011b) (Fig. 6C).
284 Additionally, only 8 strains could utilize D-ribose as a sole carbon source to grow and
285 the growth features allowed identification of a gene cluster encoding for a predicted
286 ribokinase, D-ribose pyranase, a transcriptional regulator, a ribose transporter system,
287 a predicted hypothetical protein and a MFS transporter (Pokusaeva et al., 2010) (Fig.
288 6D). As for D-mannitol/D-sorbitol utilization, an important gene cluster consisted of
289 α -acetolactate decarboxylase, an alcohol dehydrogenase, a hypothetical protein and a
290 transport system which may be involved (Bottacini et al., 2014) (Fig. 6E). Growth on
291 D-galactose corresponded to a gene cluster, which included genes encoding for
292 endo-1,4- β -galactosidase, an ABC transporter system and a transcriptional regulator
293 (O'Connell-Motherway et al., 2011a) (Fig. 6F). With regard to sucrose, a relative gene
294 cluster composed of β -fructosidases (levanase/invertase), an ABC transporter and a
295 solute-binding protein (Ryan et al., 2005) were observed (Fig. 6G).

296 **3.6 Resistance gene and phenotype binding analysis**

297 Six different antibiotics were analyzed including ampicillin, gentamicin, tetracycline,
298 chloramphenicol, ciprofloxacin and vancomycin using gene-trait matching analysis
299 for the 46 *B. breve* strains. At the gene level, there was no significant clustering for
300 the 6 antibiotic-related genes among the strains. All of the strains having a gene
301 associated with the β -lactam antibiotic, but no significant difference in the number of
302 resistance genes associated with other antibiotics were obtained (Fig. 7A). For
303 tetracycline, the *tet* homologous genes were identified in all of the strains. The
304 mechanisms underlying resistance to aminoglycosides (such as gentamicin), in which

305 aminoglycoside phosphotransferase (APH), aminoglycoside adenylyltransferase
306 (ANT) and N-acetyltransferase (AAC)-related genes were found in the CARD
307 comment results. The vancomycin genes predicted were *vanA*, *vanB* and *vanG*, which
308 were seen in most of the strains assessed. In addition, the ampicillin resistance gene
309 encoded protein, PBP2, which was also in all the strains. The common resistance gene
310 for chloramphenicol was *fexA* or *cfrA* (Kehrenberg and Schwarz, 2004; 2006), which
311 was found in almost all of those strains. Resistant genes for fluoroquinolone, *gyrA* and
312 *parC* (Janoir et al., 1996), were found in all those strains.

313 In phenotypic experiments, *B. breve* strains showed different sensitivity to those 6
314 antibiotics, and the MIC varied (Table 2). According to the MIC value of *B. longum*
315 ATCC15707, the threshold value of antibiotic tolerance and intolerance was
316 determined (Fig. 7B). For tetracycline, the MIC values ranged from 0.125 to 64
317 µg/mL. For gentamicin, strains showed significantly different degrees of tolerance,
318 thus the MIC ranged from 2 to 256 µg/mL. For vancomycin, a typical glycopeptide
319 antibiotic, vancomycin-sensitive strains accounted for 76%, only FJSWX17M1,
320 FBJCP1M6, FFBJHD5M2 and FAHWH9M5 showed a high tolerance, and the MIC
321 value was 16 µg/mL. Each strain was tested for a gene associated with tolerance to
322 ampicillin, which was a typical β-lactam antibiotic. The MIC ranged from 0.0625 to 8
323 µg/mL, while the MIC for chloramphenicol ranged from 0.5 to 32 µg/mL. For one of
324 the representative fluoroquinolone drugs, the MIC values of ciprofloxacin ranged
325 from 0.25 to 128 µg/mL among all the strains assessed.

326

327 **4. Discussion**

328 *B. breve* is one of the most abundant species of Bifidobacteria in the gastrointestinal
329 tract of breast-fed neonates and infants, and their existence is considered helpful
330 (Bottacini et al., 2018b; Freitas and Hill, 2018). Forty-six strains of *B. breve* isolated
331 from Chinese children were used to carry out the genomic diversity analysis within
332 the species and evaluated their abilities to metabolize different carbohydrates and
333 resist antibiotics. Draft genomes of 46 strains were sequenced and comparative
334 genomic analysis was done with other published *B. breve* genomes. The average G+C%
335 content of *B. longum* and *B. bifidum* was ~60 and 62.7%, respectively (Odamaki et al.,
336 2018; Turrone et al., 2011), while the average G+C% content of *B. breve* in the current
337 study was $58.8 \pm 0.5\%$, which was consistent with that in *B. longum* and *B. bifidum*,
338 but inconsistent with published *B. breve* results (Bottacini et al., 2018). The total
339 predicted ORFS of each genome varied (ranging from 1,748 to 2,396), which might
340 depend on the sequencing, or some of the differences might be explained because *B.*
341 *breve* may have megaplasms or contain large chromosomally integrated elements
342 (~200 Kb) (Bottacini et al., 2015). The pan-genome of 55 strains was determined as a
343 not fully closed but gradually saturated pan-genome, which was consistent with
344 Bottacini et al. (2014, 2018). But the difference between current and previous results
345 was the number of core-genes and total genes, in which the core-gene size previously
346 calculated for 74 *B. breve* strains (1,282 gene families) was just slightly larger than
347 the current 55 representatives (1,111 gene families). On the other hand, the calculated
348 pan-genome size for 74 *B. breve* strains (6,138 gene families) was smaller than those

349 55 strains (6,707 gene families). A possible reason was that the origin of the current
350 strains were only from children (Bottacini et al., 2018).

351 MCL comparative genomic analysis showed that *B. breve* species had high genomic
352 diversity. The strain with the largest difference in gene content was FFHNFQ49M1,
353 which had the highest numbers of truly unique genes (TUGS). Based on comparative
354 genomic analysis of 73 *B. breve* strains, most of the variome (the genetic variability)
355 in the strain composed of TUGS, and the predictive function of TUGS indicated that
356 at least 50% of the genes could encode uncharacterized, hypothetical proteins, or
357 mobile genetic elements (Bottacini et al., 2018a). Species diversity and the ability to
358 metabolize large amounts of carbohydrates were thought to be the results of
359 horizontal gene transfer and gene duplication (Kim et al., 2013). In addition, *B. breve*
360 from the same sample did not congregate according to the phylogenetic tree, such as
361 FFGZ18I1M1 and FFGZ18I1M6, and FFGZ19I1M4 and FFGZ19I1M6, indicating
362 different genes might be involved in carbohydrate metabolism and antibiotic
363 resistance, which may be the reason for the host adaptability of the strain.

364 Bifidobacteria could utilize a diverse range of dietary carbohydrates which were not
365 digested in the upper intestine (O'Callaghan and van Sinderen, 2015;
366 O'Connell-Motherway et al., 2011a). Pan-genome analysis made it possible for 35
367 GH families to participate in the prediction of carbohydrate metabolism. Most of the
368 carbohydrates, from milk and milk-derived foods present in an infant's diet until
369 weaning (Arslanoglu et al., 2008), were the substrates of β -galactosidase (GH42,
370 GH2). In the process of transitioning from a milk-based diet to solid food (around 6

371 months), starch related carbohydrates in mashed potatoes, noodles or rice, might
372 constitute a rich matrix for the growth of Bifidobacteria in the large intestine (Paturi et
373 al., 2012). The current results were also consistent with those circumstances, for
374 instance, the existence of a lot of genes encoding GH13 (α -amylase), GH3
375 (β -glucosidase) and GH38 (α -mannosidase).

376 Oligosaccharides are the third largest component in breast milk after lactose and fat,
377 and more than 200 different types of oligosaccharides in breast milk have been
378 identified in which 2'-FL is the most abundant type of HMOS (Bode, 2012;
379 Vandenplas et al., 2018). The results showed that 36 out of 46 strains showed limited
380 growth with 2'-FL as the sole carbon source. It was reported that two ABC transporter
381 permeases and the fucosidase gene were adjacent to the substrate binding protein
382 (SBP), and gene cluster for 2'-FL utilization which was consistent with previous
383 research (Matsuki et al., 2016). The SBP gene showed a substantial role in
384 determining the strain with fucosyllactose (2'-FL, 3'-FL) degradation ability. The
385 presence of fucosidase genes (GH95 and GH29) confirmed that strains with vigorous
386 growth in HMOS as the sole carbon source had one fucosidase gene. According to the
387 glycoside hydrolase gene sequence of *B. breve*, it was speculated that the glycosidase
388 was located intracellularly, and the gene sequence was homologous to that in *B.*
389 *longum* subsp. *infantis* (Ruiz-Moyano et al., 2013). The metabolic pathway of HMOS
390 by *B. breve* might be similar to that in *B. longum* subsp. *infantis*. The ability to utilize
391 2'-FL by Bifidobacteria might affect the development of intestinal microbes in infants,
392 ultimately producing beneficial effects (Ruiz-Moyano et al., 2013), which might be

393 the reason for the high abundance of *B. breve* in the children's intestine, especially in
394 the breast-fed children.

395 XOS could be directly used in the large intestine by Bifidobacteria, promoting the
396 proliferation of Bifidobacteria and producing a variety of organic acids, thus helping
397 the host to inhibit pathogenic bacteria and having a probiotic role. According to the
398 literature (Bragatto et al., 2013), GH43 could have an important role in hydrolyzing
399 XOS. First, XOS was transported into cells by the *LacI* family transcriptional
400 regulator and substrate binding protein, and then degraded into xylobiose and
401 xylotriose by intracellular β -1,4-xylosidase (GH43) for further metabolism, ultimately
402 promoting bacterial cell proliferation and producing a variety of organic acids. Most
403 strains where GH43 was found were isolated from mixed-fed children, such as
404 FHuNCS6M1, FFJND26M5, FJSWX23M8, FJSWX26M5, FJSWX17M1 and
405 FBJCP1M6. As a plant-derived bifidus factor, XOS could promote the proliferation of
406 *B. breve* in children's gut with the diversification of children's diet.

407 Research on the sensitivity of microbes to antibiotics is mainly to explore the changes
408 of antibiotics in the intestinal microbiota, and the research on drug sensitivity of
409 Bifidobacteria is also the basis for the development of live bacterial preparations. The
410 important mechanism of tetracycline resistance is the result of the active drug efflux
411 system encoded by the *tet* gene. Those strains showed a certain degree of tolerance
412 associated with the existence of the *tet* gene, and MIC values of most strains were 2
413 $\mu\text{g/mL}$. However, the results of Duranti et al. (2016) were 0.5 $\mu\text{g/mL}$. FFJND2M11
414 was isolated from a child who had taken a commercial probiotics product ("Miami",

415 Hanmi Pharmaceutical Co. Ltd., Seoul, Korea), which only contained *Enterococcus*
416 *faecium* and *Bacillus subtilis*. As enterococci are often resistant to tetracycline, the
417 resistance had the potential risk to be a lateral gene transfer to other bacteria *in vivo*.
418 Penicillin-binding protein (PBP) mutations confer resistance to ampicillin
419 (Campedelli et al., 2018), when PBP was changed, β -lactam could not bind to the cell
420 wall of the strain, and the strain finally developed resistance. In the current work, 56%
421 of *B. breve* showed higher tolerance, which was different from a previous report
422 (Fouhy et al., 2013). One of the antibacterial mechanisms for fluoroquinolones is to
423 inhibit DNA gyrase (*gyrA*) and topoisomerase IV (*parC*), disturb DNA replication
424 and kill the bacterial strain. The strains assayed (59%) were sensitive, which was
425 consistent with previous clinical data (Jose et al., 2014). The aminoglycoside
426 modifying enzyme is a passivating enzyme that inactivates aminoglycoside antibiotics,
427 and the inactivated enzymes are AAC, ANT and APH. Among them, APH has the
428 highest resistance to antibacterial drugs (Fouhy et al., 2013). A passivation enzyme
429 produces different drug resistant phenotypes. An aminoglycoside could be inactivated
430 by one or more enzymes, and several aminoglycosides could be simultaneously
431 inactivated by the same enzyme. Forty-six percent of the strains were resistant to
432 gentamicin, which was roughly in line with a previous report (Jose et al., 2014).
433 Chloramphenicol antibiotics act on the 50S subunit of the bacterial ribonucleoprotein
434 and block the synthesis of proteins (Campedelli et al., 2018). In the annotation results
435 for *B. breve*, the ATP-binding cassette ribosome protective proteins and antibiotic
436 target protection were found in predicted strains, consequently, and those *B. breve*

437 strains showed drug resistance to a certain degree. Furthermore, the resistance gene in
438 *Bifidobacterium* may be due to the co-evolution of the microbial-host (Gibson et al.,
439 2015). Although there was little difference in the number of genes involved in the
440 various antibiotics resistance predicted in the strain, phenotypes differed, possibly as a
441 result of host specificity. The current results led to a preliminary understanding of the
442 resistance mechanism of *B. breve*, which provides a research basis for the next step
443 which is the prevention and treatment of drug-induced bacteria.

444

445 **5. Conclusions**

446 Comparative genomics and functional analysis of 46 *B. breve* strains has been
447 completed. Combined with pan-genome and core-genome, functional genome
448 analysis allowed the identification of the genes responsible for strain diversity and
449 host specificity. Different gene families of different strains could serve as the basis for
450 gene-trait matching phenotype studies. The 46 strains of *B. breve* strains isolated from
451 Chinese children represented many different phenotypes, providing further
452 information on the meanings of the genetic diversity in the species, especially the
453 carbohydrate utilization capabilities and antibiotic resistance.

454

455 **Conflict of interest**

456 All authors declared no conflict of interest.

457

458 **Acknowledgement**

459 This research was supported by the National Natural Science Foundation of China

460 (Nos. 31771953, 31820103010, 31801521), National First-Class Discipline Program
461 of Food Science and Technology (JUFSTR20180102), the Fundamental Research
462 Funds for the Central Universities (No. JUSRP11733), the Collaborative Innovation
463 Center of Food Safety and Quality Control in Jiangsu Province, and the Nutrition and
464 Care of Maternal & Child Research Fund Project of Guangzhou Biostime Institute of
465 Nutrition & Care (BINCMCF1201). 2'-FL was kindly provided by FrieslandCampina,
466 the Netherlands.

467

468 **References**

469 Arboleya, S., Bottacini, F., O'Connell-Motherway, M., Ryan, C.A., Ross, R.P., van
470 Sinderen, D., & Stanton, C. (2018). Gene-trait matching across the *Bifidobacterium*
471 *longum* pan-genome reveals considerable diversity in carbohydrate catabolism among
472 human infant strains. *BMC Genomics*, 19, 33.

473 Arboleya, S., Watkins, C., Stanton, C., & Ross, R.P. (2016). Gut Bifidobacteria
474 populations in human health and aging. *Frontiers in Microbiology*, 7, 1204.

475 Arslanoglu, S., Moro, G.E., Schmitt, J., Tandoi, L., Rizzardi, S., & Boehm, G. (2008).
476 Early dietary intervention with a mixture of prebiotic oligosaccharides reduces the
477 incidence of allergic manifestations and infections during the first two years of life.
478 *Journal of Nutrition*, 138, 1091-1095.

479 Bode, L. (2012). Human milk oligosaccharides: Every baby needs a sugar mama.
480 *Glycobiology*, 22, 1147-1162.

- 481 Boesten, R., Schuren, F., Wind, R.D., Knol, J., & de Vos, W.M. (2011). Analysis of
482 infant isolates of *Bifidobacterium breve* by comparative genome hybridization
483 indicates the existence of new subspecies with marked infant specificity. *Research in*
484 *Microbiology*, 162, 664-670.
- 485 Bottacini, F., Morrissey, R., Esteban-Torres, M., James, K., van Breen, J., Dikareva,
486 E., Egan, M., Lambert, J., van Limpt, K., Knol, J., O'Connell-Motherway, M. & van
487 Sinderen, D. (2018a). Comparative genomics and genotype-phenotype associations in
488 *Bifidobacterium breve*. *Scientific Reports*, 8, 10633.
- 489 Bottacini, F., Morrissey, R., Roberts, R.J., James, K., van Breen, J., Egan, M.,
490 Lambert, J., van Limpt, K., Knol, J., O'Connell-Motherway, M., & van Sinderen, D.
491 (2018b). Comparative genome and methylome analysis reveals
492 restriction/modification system diversity in the gut commensal *Bifidobacterium breve*.
493 *Nucleic Acids Research*, 46, 1860-1877.
- 494 Bottacini, F., O'Connell-Motherway, M., Casey, E., McDonnell, B., Mahony, J.,
495 Ventura, M., & van Sinderen, D. (2015). Discovery of a conjugative megaplasmid in
496 *Bifidobacterium breve*. *Applied and Environmental Microbiology*, 81, 166-176.
- 497 Jin, Z., Li, W., Wang, W., & Sun, B. (2019). Draft genome sequence of
498 *Bifidobacterium longum* ZJ1, isolated from a centenarian in Anhui, China.
499 *Microbiology Resource Announcements*, 8: e00878-19.
- 500 Bottacini, F., O'Connell-Motherway, M., Kuczynski, J., O'Connell, K.J., Serafini, F.,
501 Duranti, S., Milani, C., Turrone, F., Lugli, G.A., Zomer, A., Zhurina, D., Riedel, C.,

- 502 Ventura, M., & van Sinderen, D. (2014). Comparative genomics of the
503 *Bifidobacterium breve* taxon. *BMC Genomics*, 15, 170.
- 504 Bragatto, J., Segato, F., & Squina, F.M. (2013). Production of xylooligosaccharides
505 (XOS) from delignified sugarcane bagasse by peroxide-HAc process using
506 recombinant xylanase from *Bacillus subtilis*. *Industrial Crops and Products*, 51:
507 123-129.
- 508 Campedelli, I., Mathur, H., Salvetti, E., Clarke, S., Rea, M.C., Torriani, S., Ross, R.P.,
509 Hill, C., & O'Toole, P.W. (2018). Genus-wide assessment of antibiotic resistance in
510 *Lactobacillus spp.* *Applied and Environmental Microbiology*, 85(1), e01738-18.
- 511 Castanys-Muñoz, E., Martin, M.J., & Prieto, P.A. (2013). 2'-Fucosyllactose: An
512 abundant, genetically determined soluble glycan present in human milk. *Nutrition*
513 *Reviews*, 71, 773-789.
- 514 Divakar, P.K., & Crespo, A. (2015). Molecular phylogenetic and phylogenomic
515 approaches in studies of lichen systematics and evolution. In: D. Upreti, P. Divakar, V.
516 Shukla, R. Bajpai (Eds.), *Recent Advances in Lichenology*. New Delhi: Springer. p
517 45-60.
- 518 Duranti, S., Lugli, G.A., Mancabelli, L., Turrone, F., Milani, C., Mangifesta, M.,
519 Ferrario, C., Anzalone, R., Viappiani, A., van Sinderen, D., & Ventura, M. (2016).
520 Prevalence of antibiotic resistance genes among human gut-derived Bifidobacteria.
521 *Applied and Environmental Microbiology*, 83, e02894-16.
- 522 Enright, A.J., van Dongen, S., & Ouzounis, C.A. (2002). An efficient algorithm for
523 large-scale detection of protein families. *Nucleic Acids Research*, 30, 1575-1584.

- 524 Fouhy, F., O'Connell-Motherway, M., Fitzgerald, G.F., Ross, R.P., Stanton, C., van
525 Sinderen, D., & Cotter, P.D. (2013). *In silico* assigned resistance genes confer
526 *Bifidobacterium* with partial resistance to aminoglycosides but not to β -lactams. *PLoS*
527 *One*, 8, e82653.
- 528 Francino, M.P. (2015). Antibiotics and the human gut microbiome: Dysbioses and
529 accumulation of resistances. *Frontiers in Microbiology*, 6, 159-160.
- 530 Freitas, A.C., & Hill, J.E. (2018). Bifidobacteria isolated from vaginal and gut
531 microbiomes are indistinguishable by comparative genomics. *PLoS One*, 13,
532 e0196290.
- 533 Gibson, M.K., Crofts, T.S., & Dantas, G. (2015). Antibiotics and the developing infant
534 gut microbiota and resistome. *Current Opinion in Microbiology*, 27, 51-56.
- 535 Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., &
536 Tiedje, J.M. (2007). DNA-DNA hybridization values and their relationship to
537 whole-genome sequence similarities. *International Journal of Systematic and*
538 *Evolutionary Microbiology*, 57, 81-91.
- 539 Harris, H.M.B., Bourin, M.J.B., Claesson, M.J., & O'Toole, P.W. (2017).
540 Phylogenomics and comparative genomics of *Lactobacillus salivarius*, a mammalian
541 gut commensal. *Microbial Genomics*, 3(8), e000115.
- 542 Janoir, C., Zeller, V., Kitzis, M.D., Moreau, N.J., & Gutmann, L. (1996). High-level
543 fluoroquinolone resistance in *Streptococcus pneumoniae* requires mutations in *parC*
544 and *gyrA*. *Antimicrobial Agents and Chemotherapy*, 40, 2760-2764.

- 545 Jiménez, E., Villar-Tajadura, M.A., Marín, M., Fontecha, J., Requena, T., Arroyo, R.,
546 Fernández, L., & Rodríguez, J.M. (2012). Complete genome sequence of
547 *Bifidobacterium breve* CECT7263, a strain isolated from human milk. *Journal of*
548 *Bacteriology*, 194, 3762-3763.
- 549 Jose, N.M., Bunt, C.R., & Hussain, M.A. (2014). Implications of antibiotic resistance
550 in probiotics. *Food Reviews International*, 31, 52-62.
- 551 Kehrenberg, C., & Schwarz, S. (2004). *FexA*, a novel *Staphylococcus lentus* gene
552 encoding resistance to florfenicol and chloramphenicol. *Antimicrobial Agents and*
553 *Chemotherapy*, 48, 615-618.
- 554 Kehrenberg, C., & Schwarz, S. (2006). Distribution of florfenicol resistance genes
555 *fexA* and *cfr* among chloramphenicol-resistant *Staphylococcus* isolates. *Antimicrobial*
556 *Agents and Chemotherapy*, 50, 1156-1163.
- 557 Kelly, E.D., Bottacini, F., O'Callaghan, J., O'Connell-Motherway, M., O'Connell, K.J.,
558 Stanton, C., & van Sinderen, D. (2016). Glycoside hydrolase family 13
559 alpha-glucosidases encoded by *Bifidobacterium breve* UCC2003: A comparative
560 analysis of function, structure and phylogeny. *International Journal of Food*
561 *Microbiology*, 224, 55-65.
- 562 Kim, J.H., An, H.J., Garrido, D., German, J.B., Lebrilla, C.B., & Mills, D.A. (2013).
563 Proteomic analysis of *Bifidobacterium longum* subsp. *infantis* reveals the metabolic
564 insight on consumption of prebiotics and host glycans. *PLoS One*, 8, e57535.

- 565 Liu, Y., Schröder, J., & Schmidt, B. (2013). Musket: A multistage k-mer
566 spectrum-based error corrector for Illumina sequence data. *Bioinformatics*, 29,
567 308-315.
- 568 Liu, K., & Warnow, T. (2014). Large-scale multiple sequence alignment and tree
569 estimation using SATé. *Methods in Molecular Biology*, 1079, 219-244.
- 570 Masco, L., van Hoorde, K., de Brandt, E., Swings, J., & Huys, G. (2006).
571 Antimicrobial susceptibility of *Bifidobacterium* strains from humans, animals and
572 probiotic products. *Antimicrobial Agents and Chemotherapy*, 58, 85-94.
- 573 Matsuki, T., Yahagi, K., Mori, H., Matsumoto, H., Hara, T., Tajima, S., Ogawa, E.,
574 Kodama, H., Yamamoto, K., Yamada, T., Matsumoto, S., & Kurokawa, K. (2016). A
575 key genetic factor for fucosyllactose utilization affects infant gut microbiota
576 development. *Nature Communications*, 7, 11939.
- 577 O'Callaghan, A., & van Sinderen, D. (2016). Bifidobacteria and their role as members
578 of the human gut microbiota. *Frontiers in Microbiology*, 7, 925.
- 579 O'Connell-Motherway, M., Fitzgerald, G.F., & van Sinderen, D. (2011a). Metabolism
580 of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve*
581 UCC2003. *Microbial Biotechnology*, 4, 403-416.
- 582 O'Connell-Motherway, M., Zomer, A., Leahy, S.C., Reunanen, J., Bottacini, F.,
583 Claesson, M.J., O'Brien, F., Flynn, K., Casey, P.G., Munoz, J.A., Kearney, B.,
584 Houston, A.M., O'Mahony, C., Higgins, D.G., Shanahan, F., Palva, A., de Vos, W.M.,
585 Fitzgerald, G.F., Ventura, M., O'Toole, P.W., & van Sinderen, D. (2011b). Functional
586 genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence

587 (Tad) pili as an essential and conserved host-colonization factor. *Proceedings of the*
588 *National Academy of Sciences of the United States of America*, 108, 11217-11222.

589 Odamaki, T., Bottacini, F., Kato, K., Mitsuyama, E., Yoshida, K., Horigome, A., Xiao,
590 J.Z., & van Sinderen, D. (2018). Genomic diversity and distribution of
591 *Bifidobacterium longum* subsp. *longum* across the human lifespan. *Scientific Reports*,
592 8, 85.

593 Patel, R.K., & Jain, M. (2012). NGS QC toolkit: A toolkit for quality control of next
594 generation sequencing data. *PLoS One*, 7, e30619.

595 Paturi, G., Butts, C.A., Stoklosinski, H., & Ansell, J. (2012). Effects of early dietary
596 intervention with a fermentable fibre on colonic microbiota activity and mucin gene
597 expression in newly weaned rats. *Journal of Functional Foods*, 4, 520-530.

598 Pokusaeva, K., Fitzgerald, G.F., & van Sinderen, D. (2011a). Carbohydrate
599 metabolism in Bifidobacteria. *Genes and Nutrition*, 6, 285-306.

600 Pokusaeva, K., Neves, A.R., Zomer, A., O'Connell-Motherway, M., MacSharry, J.,
601 Curley, P., Fitzgerald, G.F., & van Sinderen, D. (2010). Ribose utilization by the
602 human commensal *Bifidobacterium breve* UCC2003. *Microbial Biotechnology*, 3,
603 311-323.

604 Pokusaeva, K., O'Connell-Motherway, M., Zomer, A., MacSharry, J., Fitzgerald, G.F.,
605 & van Sinderen, D. (2011b). Cellodextrin utilization by *Bifidobacterium breve*
606 UCC2003. *Applied and Environmental Microbiology*, 77, 1681-1690.

- 607 Qin, Q., Xie, B., Zhang, X., Chen, X., Zhou, B., Zhou, J., Oren, A., & Zhang, Y.
608 (2014). A proposed genus boundary for the prokaryotes based on genomic insights.
609 *Journal of Bacteriology*, 196, 2210-2215.
- 610 Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European molecular
611 biology open software suite. *Trends in Genetics*, 16, 276-277.
- 612 Round, J.L., & Mazmanian, S.K. (2009). The gut microbiota shapes intestinal
613 immune responses during health and disease. *Nature Reviews in Immunology*, 9,
614 313-323.
- 615 Ruiz-Moyano, S., Totten, S.M., Garrido, D.A., Smilowitz, J.T., German, J.B., Lebrilla,
616 C.B., & Mills, D.A. (2013). Variation in consumption of human milk oligosaccharides
617 by infant gut-associated strains of *Bifidobacterium breve*. *Applied and Environmental*
618 *Microbiology*, 79, 6040-6049.
- 619 Ryan, S.M., Fitzgerald, G.F., & van Sinderen, D. (2005). Transcriptional regulation
620 and characterization of a novel beta-fructofuranosidase-encoding gene from
621 *Bifidobacterium breve* UCC2003. *Applied and Environmental Microbiology*, 71,
622 3475-3482.
- 623 Schattner, P., Brooks, A.N., & Lowe, T.M. (2005). The tRNAscan-SE, snoscan and
624 snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids*
625 *Research*, 33, 686-689.
- 626 Tojo, R., Suárez, A., Clemente, M.G., de los Reyes-Gavilán, C.G., Margolles, A.,
627 Gueimonde, M., & Ruas-Madiedo, P. (2014). Intestinal microbiota in health and

628 disease: Role of Bifidobacteria in gut homeostasis. *World Journal of Gastroenterology*,
629 20, 15163-15176.

630 Tong, C., Chen, Q., Zhao, L., Ma, J., Ibeagha-Awemu, E.M., & Zhao, X. (2017).
631 Identification and characterization of long intergenic noncoding RNAs in bovine
632 mammary glands. *BMC Genomics*, 18, 468.

633 Turrioni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J.H., Zomer, A., Sánchez, B.,
634 Bidossi, A., Ferrarini, A., Giubellini, V., Delledonne, M., Henrissat, B., Coutinho, P.,
635 Oggioni, M., Fitzgerald, G.F., Mills, D., Margolles, A., Kelly, D., van Sinderen, D., &
636 Ventura, M. (2011). Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals
637 metabolic pathways for host-derived glycan foraging. *Proceedings of the National*
638 *Academy of Sciences of the United States of America*, 107, 19514-19519.

639 Turrioni, F., Milani, C., Duranti, S., Ferrario, C., Lugli, G.A., Mancabelli, L., van
640 Sinderen, D., & Ventura, M. (2018). Bifidobacteria and the infant gut: An example of
641 co-evolution and natural selection. *Cellular and Molecular Life Sciences*, 75,
642 103-118.

643 Vandenplas, Y., Berger, B., Carnielli, V.P., Ksiazzyk, J., Lagström, H., Luna, M.S.,
644 Migacheva, N., Mosselmans, J.M., Picaud, J.C., Possner, M., Singhal, A., & Wabitsch,
645 M. (2018). Human milk oligosaccharides: 2'-Fucosyllactose (2'-FL) and
646 lacto-n-neotetraose (LNnT) in infant formula. *Nutrients*, 10, 1161.

647 Ventura, M., Turrioni, F., O'Connell-Motherway M, MacSharry, J., & van Sinderen, D.
648 (2012). Host-microbe interactions that facilitate gut colonization by commensal
649 Bifidobacteria. *Trends in Microbiology*, 20, 467-476.

- 650 Watson, D., O'Connell-Motherway, M., Schoterman, M.H., van Neerven, R.J., Nauta,
651 A., & van Sinderen, D. (2013). Selective carbohydrate utilization by lactobacilli and
652 Bifidobacteria. *Journal of Applied Microbiology*, 114, 1132-1146.
- 653 Zerbino, D.R., & Birney, E. (2008). Velvet: Algorithms for de novo short read
654 assembly using de Bruijn graphs. *Genome Research*, 18, 821-829.
- 655 Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., & Yu, J. (2012). PGAP: Pan-genomes
656 analysis pipeline. *Bioinformatics*, 28, 416-418.
- 657

658 **Figure captions**

659 **Fig. 1 Comparative analysis of *B. breve* genomes.**

660 A) Venn diagram showing core-gene families and the genes specific to the *B. breve*
661 strains.

662 B) Hierarchical clustering heat map suggesting the variability of *B. breve* according to
663 presence or absence of gene families.

664 C) Pie chart displaying the percentage of dispensable and core gene families.

665 D) ANI of 55 strains.

666

667 **Fig. 2 Pan-genome and core-genome of the *B. breve* species.**

668 A) Trend diagram of pan-genome and core-genome of *Bifidobacterium breve*. B)
669 Functional distributions of the *B. breve* core genome.

670 Each letter standing for the following function: [A]: RNA processing and
671 modification, [G]: Carbohydrate transport and metabolism, [C]: Energy production
672 and conversion, [D]: Cell cycle control, cell division, chromosome partitioning, [E]:

673 Amino acid transport and metabolism, [F]: Nucleotide transport and metabolism, [H]:

674 Coenzyme transport and metabolism, [I]: Lipid transport and metabolism, [J]:

675 Translation, ribosomal structure and biogenesis, [K]: Transcription, [L]: Replication,

676 recombination and repair, [M]: Cell wall/membrane/envelope biogenesis, [O]:

677 Post-translational modification, protein turnover, chaperones, [P]: Inorganic ion

678 transport and metabolism, [Q]: Secondary metabolites biosynthesis, transport, [R]:

679 General function prediction only, [S]: Function unknown, [T]: Signal transduction

680 mechanisms, [U]: Intracellular trafficking, secretion, and vesicular, [V]: Defense
 681 mechanisms.

682 **Fig. 3 Phylogenetic analysis of *B. breve*.**

683 Phylogenetic supertree showing the relationship among 55 *B. breve* strains and *B.*
 684 *bifidum* S17, *B. bifidum* PRL2010 and *B. longum* ATCC15697 as outliers. Strains
 685 isolated from Henan (green): FSXR13, FHuNan2016497, FHuNan2016415,
 686 FFHNXY43M2, FFHNXY26M4, FFHNFQ4M7, FFHNFQ49M1, FHeNJZ9M1,
 687 FHeNJZ2M1 and FHeNJZ1M1; Fujian (pink): FFJND6M1, FFJND2M11,
 688 FFND12M6, FFJND14L2 and FFJND26M5; Guangdong (purple): FFGZ3I1M6,
 689 FFGZ23I1M6, FFGZ19I1M6, FFGZ19I1M4, FFGZ18I1M1 and FFGZ18I1M6;
 690 Xinjiang (blue): FCJ951, FCJ1041 and FCJ653; Beijing (orange): FFBJHD5M2,
 691 FFBJCP2M1, FBJSJS1M2 and FBJCP1M6; Anhui (cyan): FAHWH9M5 and
 692 FAHWH21M7; Zhejiang (yellow): FZJHZ13M2, FZJHZ24M9, FZJHZ3M2,
 693 FZJHZ7M2 and FZJHZD20M12; Jiangsu (black): FJSWX5M4, FJSWX4M9,
 694 FJSWX39M4, FJSWX34M6, FJSWX26M5, FJSWX24M2, FJSWX23M8 and
 695 FJSWX17M1; Hunan (brown): FHuNCS6M1, FHuNCS3M4 and FHuNCS1M5.
 696 Hyphen representing other strains from NCBI. Star representing feeding method:
 697 mixed feeding (red), breast-milk (yellow), formula milk (green). Triangle representing
 698 age: ~0-6 months (red), ~6-12 months (yellow), ~1-5 years old (green).

699 **Fig. 4 The predicted glycometabolism gene in *B. breve*.**

700 A) Pie chart showing the predicted carbohydrate-active enzymes.

701 B) Heat map of the predicted GH family.

702 **Fig. 5 Heat map of carbohydrate metabolism in 46 *B. breve* strains based on 20**
703 **different substrates.**

704 Yellow for utilization; Black for non-utilization.

705 **Fig. 6 Predicted carbohydrate utilization clusters**

706 The locus map (A-G) showing that gene clusters may be involved in the utilization of
707 various carbohydrates by certain *B. breve* strains.

708 **Fig. 7 Resistance genes and phenotypes of *B. breve*.**

709 A) The predicted resistance gene(s).

710 B) Tolerance and sensitivity of the 46 strains to antibiotics. Red for sensitivity; Black
711 for tolerance.

Table 1 *Bifidobacterium breve* genomes sequenced and analysed in this study.

Strains	Region/ Province	Source (age; feeding)	Genome Size (Mb)	ORFs	tRN A	G+C (%)	Conti gs	Accession No.
FFJND14L2	Fujian	1yr; MF	2.26	2010	54	58.87	31	SAMN13258891
FFJND2M11	Fujian	5m; MF	2.26	2017	55	58.30	41	SAMN13258892
FFJND6M1	Fujian	7m; MF	2.30	2065	53	58.77	40	SAMN13258893
FFJND12M6	Fujian	6m; MF	2.23	1939	55	58.66	50	SAMN13258894
FFJND26M5	Fujian	8m; MF	2.24	1902	53	58.56	521	SAMN13258895
FFGZ3I1M6	Guangdong	7d; FM	2.28	1982	56	58.61	77	SAMN13258896
FFGZ18I1M6	Guangdong	7d; BM	2.26	2023	53	58.86	36	SAMN13258897
FFGZ18I1M1	Guangdong	7d; BM	2.26	2020	51	58.86	36	SAMN13258898
FFGZ19I1M4	Guangdong	7d; BM	2.26	2022	57	58.86	34	SAMN13258899
FFGZ19I1M6	Guangdong	7d; BM	2.29	2001	54	59.01	31	SAMN13258900
FFGZ23I1M6	Guangdong	7d; FM	2.34	2037	56	59.04	59	SAMN13258901
FZJHZ13M2	Zhejiang	2m; BM	2.26	2033	53	58.87	35	SAMN13258902
FZJHZ24M9	Zhejiang	<1yr; ND	2.27	1979	53	58.63	29	SAMN13258903
FZJHZ7M2	Zhejiang	<1yr; ND	2.30	2018	56	58.77	56	SAMN13258904
FZJHZ3M2	Zhejiang	<1yr; ND	2.38	2099	60	58.95	48	SAMN13258905
FZJHZD20M12	Zhejiang	3yr; ND	2.46	2190	69	58.30	97	SAMN13258906
FAHWH9M5	Anhui	6m; ND	2.31	1993	53	58.91	202	SAMN13258907
FAHWH21M7	Anhui	4m; ND	2.27	1936	54	58.67	416	SAMN13258908
FCJ951	Xinjiang	11m; ND	2.33	2058	53	58.83	53	SAMN13258909
FCJ653	Xinjiang	5yr; ND	2.33	2065	54	58.72	54	SAMN13258910
FCJ1041	Xinjiang	2yr; ND	2.38	2101	52	59.06	62	SAMN13258911
FBJSJS1M2	Beijing	9m; FM	2.35	2053	54	58.93	38	SAMN13258912
FFBJHD5M2	Beijing	10m; BM	2.24	1898	54	58.71	176	SAMN13258913
FBJCP1M6	Beijing	10m; MF	2.46	2190	74	58.72	304	SAMN13258914
FFBJCP2M1	Beijing	5m; BM	2.40	2083	54	58.95	276	SAMN13258915
FJSWX4M9	Jiangsu	10m; BM	2.31	2011	52	59.00	334	SAMN13258916
FJSWX5M4	Jiangsu	11m; MF	2.36	2033	53	58.84	479	SAMN13258917
FJSWX17M1	Jiangsu	6m; MF	2.36	2056	55	58.90	332	SAMN13258918
FJSWX23M8	Jiangsu	10m; MF	2.22	1882	55	58.61	179	SAMN13258919
FJSWX24M2	Jiangsu	6m; MF	2.26	1945	53	58.89	277	SAMN13258920
FJSWX26M5	Jiangsu	6m; MF	2.64	2396	72	59.32	289	SAMN13258921

FJSWX34M6	Jiangsu	5m;ND	2.46	2213	51	58.84	64	SAMN13258922
FJSWX39M4	Jiangsu	3d; ND	2.28	1961	52	58.61	66	SAMN13258923
FHeNJZ1M1	Henan	10m; ND	2.38	2099	54	58.86	54	SAMN13258924
FHeNJZ2M1	Henan	<1yr; ND	2.37	2103	71	58.54	45	SAMN13258925
FHeNJZ9M1	Henan	5m; ND	2.36	2074	55	58.82	41	SAMN13258926
FFHNFQ4M7	Henan	8m; ND	2.27	2025	54	58.93	51	SAMN13258927
FFHNFQ49M1	Henan	8m; ND	2.29	2048	55	58.79	28	SAMN13258928
FFHNXY43M2	Henan	2yr; ND	2.16	1848	55	58.58	33	SAMN13258929
FFHNXY26M4	Henan	1yr; ND	2.28	1979	55	58.78	100	SAMN13258930
FHuNan2016497	Henan	4ys; ND	2.27	2047	57	58.82	34	SAMN13258931
FHuNan2016415	Henan	3yr; ND	2.42	2273	75	58.68	33	SAMN13258932
FSXR13	Henan	1yr; ND	2.57	2395	71	58.48	319	SAMN13258933
FHuNCS1M5	Hunan	<1yr; ND	2.51	2279	69	58.75	66	SAMN13258934
FHuNCS3M4	Hunan	<1yr; ND	2.38	2118	55	58.81	68	SAMN13258935
FHuNCS6M1	Hunan	<1yr; MF	2.58	2348	69	59.40	62	SAMN13258936
UCC2003	Ireland [27]	Infant faeces	2.42	1854	54	58.70	1	SAMN02604112
12L	Unpublished	Human milk	2.24	1765	52	58.90	1	SAMN03081478
31L	Unpublished	Human milk	2.27	1814	53	58.60	4	SAMN02951889
CECT7263	Spain [28]	Human milk	2.33	2053	53	58.90	34	SAMN06473336
NCFB2258	Unpublished	Infant intestine	2.32	1834	53	58.70	1	SAMN03081481
ACS-071-V-Sch8b	Unpublished	Human vagina	2.33	1826	53	58.70	1	SAMN00100758
689b	Unpublished	Infant feces	2.33	1821	53	58.70	1	SAMN03081482
S27	Unpublished	Infant feces	2.29	1748	53	58.70	1	SAMN03081483
JCM7017	[29]	Infant feces	2.29	1770	54	58.70	2	SAMN03081479

MF: mixed feeding (breastmilk-based added to complementary foods or formula milk); BM: breast-milk; FM: formula milk; ND: not determined or unknown.

Table 2 MIC of different antibiotics for *B. breve*.

Strain	Concentration ($\mu\text{g/ml}$)					
	Tet	Cm	Cip	Amp	Gm	Van
FJSWX4M9	2	16	32	0.0625	32	4
FJSWX26M5	0.5	2	16	0.0625	64	4
FJSWX39M4	2	2	4	4	16	0.25
FJSWX17M1	0.125	32	8	0.0625	64	16
FJSWX34M6	4	8	8	4	256	1
FJSWX5M4	1	4	4	0.125	4	2
FJSWX23M8	0.5	4	4	0.5	64	4
FJSWX24M2	0.5	8	8	0.125	64	8
FFJND12M6	2	8	32	8	32	0.5
FFJND2M11	2	2	16	0.5	32	0.25
FFJND14L2	0.5	2	16	0.5	64	0.25
FFJND26M5	1	32	64	0.0625	16	0.25
FFJND6M1	4	2	4	2	16	0.25
FFBJCP2M1	2	2	32	0.0625	16	0.25
FBJCP1M6	16	32	128	0.0625	32	16
FFBJHD5M2	0.25	32	128	0.0625	8	16
FBJSJS1M2	64	2	16	8	256	0.25
FFHNFQ4M7	1	8	64	0.0625	8	4
FFHNFQ49M1	8	8	32	0.0625	4	0.25
FAHWH9M5	1	4	16	0.0625	2	16
FAHWH21M7	1	16	8	0.0625	16	4
FHuNan2016497	2	4	32	8	16	0.25
FHeNJZ1M1	2	2	8	8	8	0.25
FHeNJZ2M1	64	4	4	8	8	0.25
FHeNJZ9M1	2	16	8	2	8	0.25
FHuNan2016415	2	16	32	8	16	0.5
FHuNCS1M5	4	0.5	4	4	4	0.25
FHuNCS3M4	1	8	32	8	32	0.25
FHuNCS6M1	2	8	8	4	2	0.25
FFHNXY26M4	0.5	4	8	4	4	0.25
FZJHZ7M2	32	16	32	8	8	0.5
FZJHZ24M9	2	2	8	8	16	0.25
FZJHZ13M2	0.5	2	4	1	16	1
FZJHZ3M2	2	16	4	4	32	1
FZJHZD20M12	4	2	16	8	256	0.25
FCJ653	2	8	4	8	64	0.25
FCJ1041	64	8	4	2	16	0.5
FCJ951	2	4	8	16	16	0.25
FSXR13	4	2	16	8	256	0.25
FFHNXY43M2	0.5	2	0.5	4	256	0.25
FFGZ1811M1	32	16	4	16	128	0.5

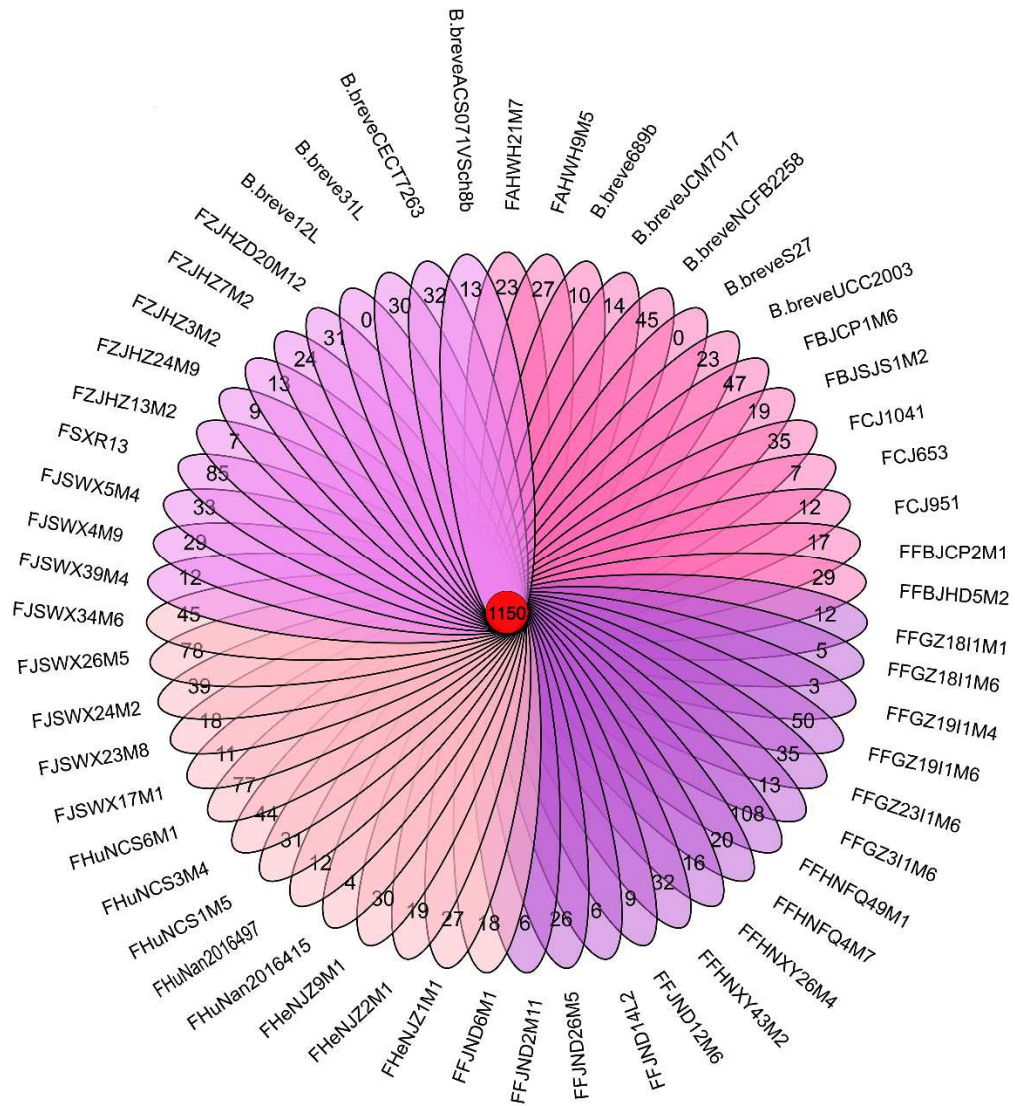
FFGZ1811M6	0.5	2	4	4	128	0.25
FFGZ1911M4	0.5	2	4	4	64	0.25
FFGZ2311M6	2	2	8	16	16	0.25
FFGZ1911M6	0.5	1	4	8	64	0.25
FFGZ311M6	0.5	1	0.25	8	16	0.25

Tet: tetracycline; Cm: chloramphenicol; Cip: ciprofloxacin; Amp: ampicillin; Gm: gentamicin; Van: vancomycin

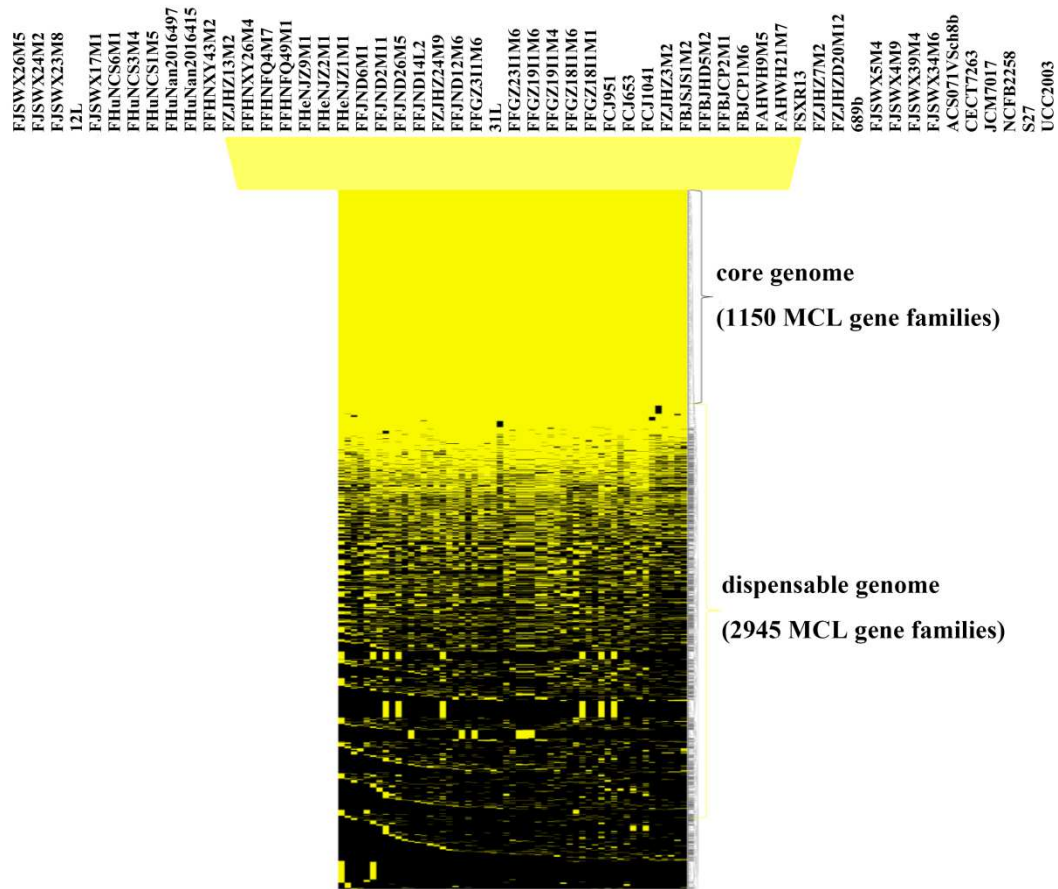
Journal Pre-proof

Fig 1

A)



B)



C)

■ *B. breve* core genome ■ *B. breve* dispensable genome

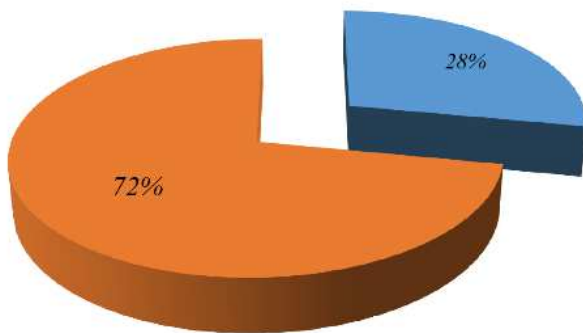
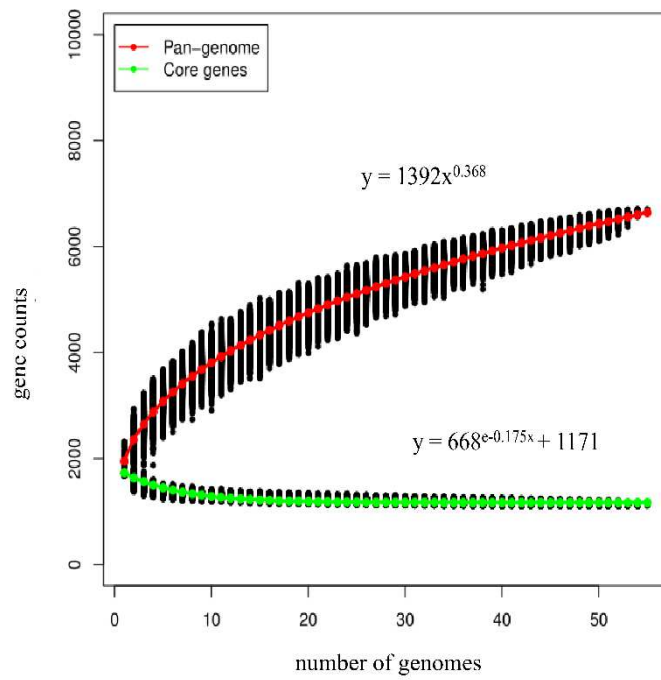


Fig 2

A)



B)

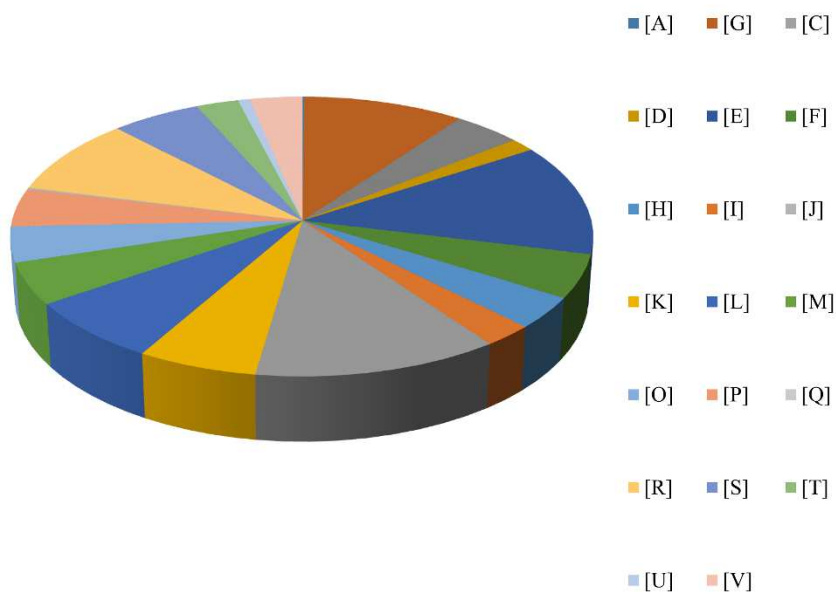


Fig 3

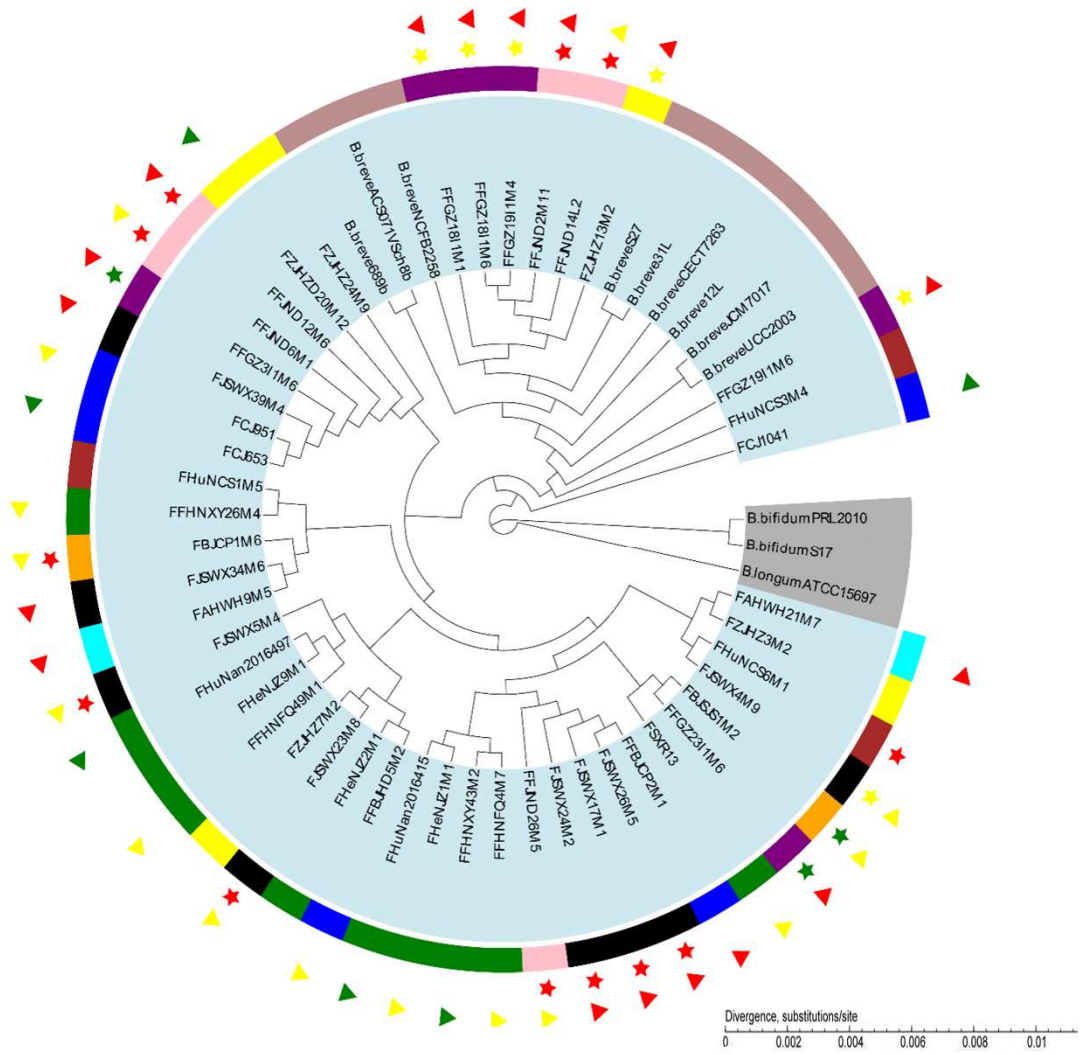
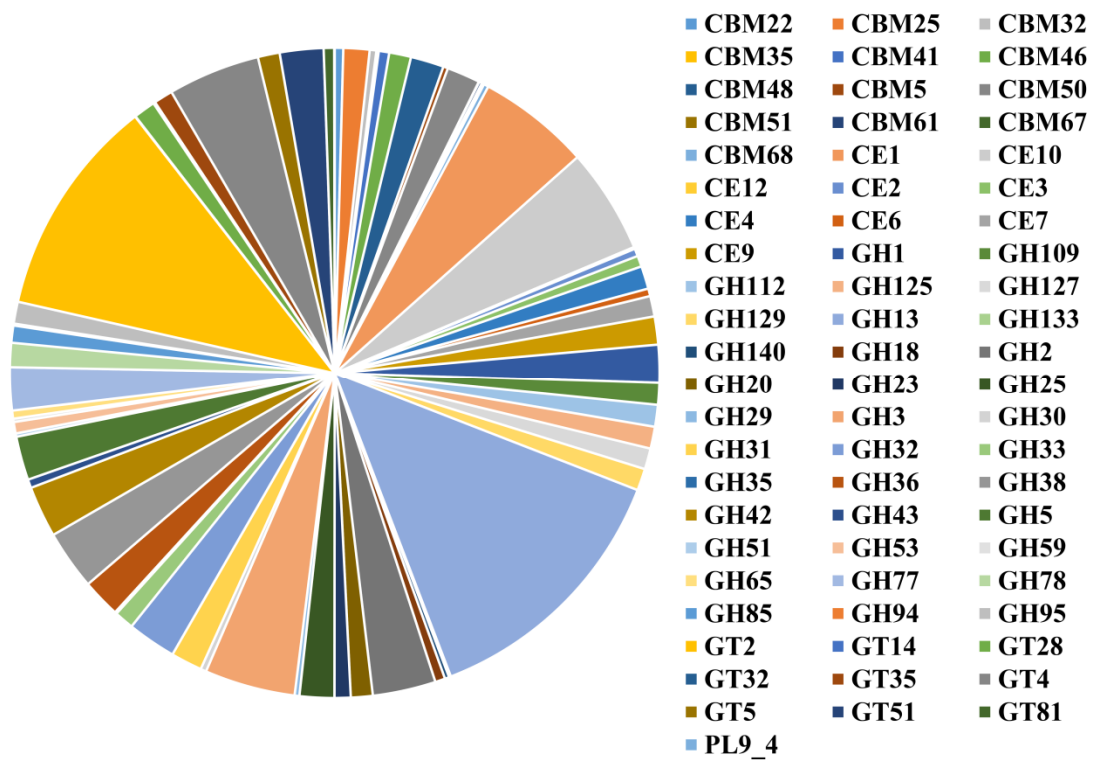
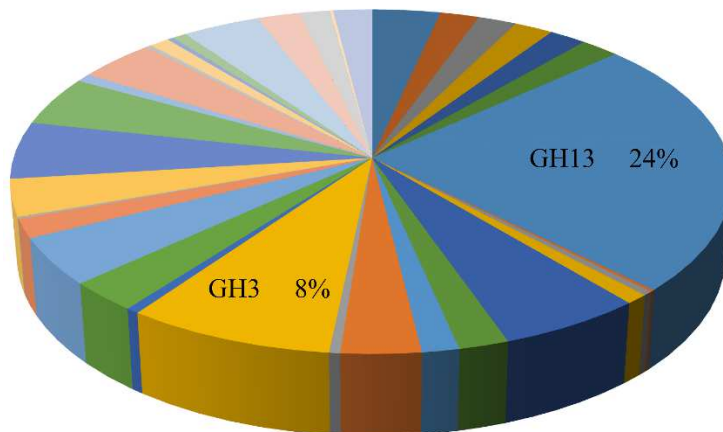


Fig 4

A)



B)



C)

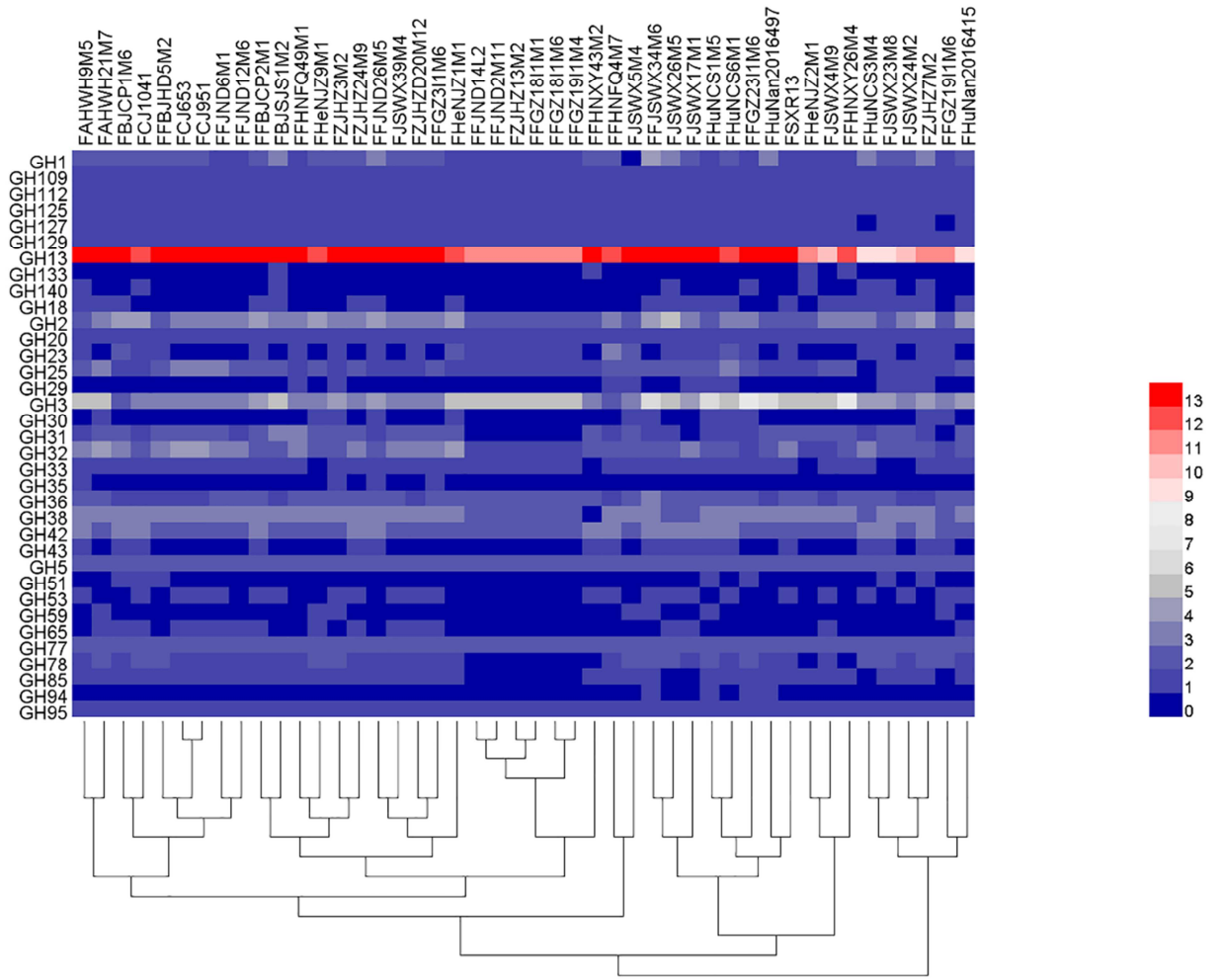
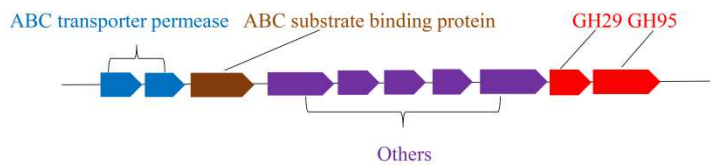


Fig 6

A)

2'-FL utilization cluster

B. longum APC 1477 (reference strain)

FHuNCS6M1(+): gene1737-gene1745

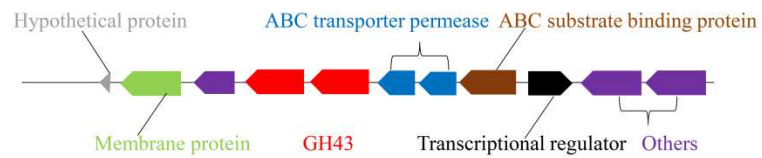


FHNXY43M2(-): gene1344-gene1347



B)

XOS utilization cluster

B. longum APC 1464 (reference strain)

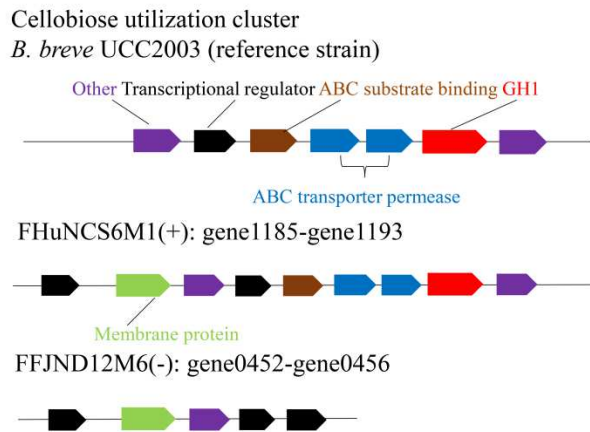
FHuNCS6M1(+): gene1296-gene1302



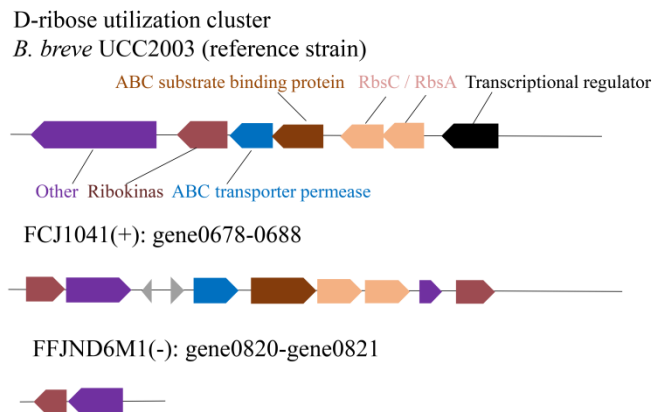
FJSWX39M4(-): gene0437-gene0439



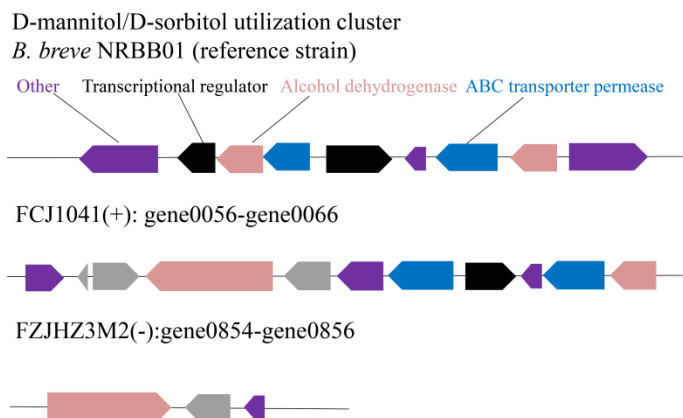
C)



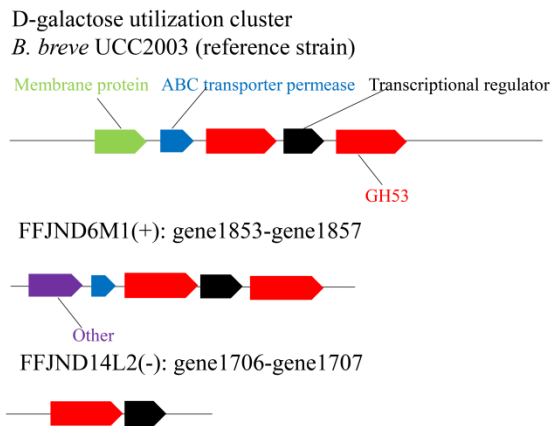
D)



E)



F)



G)

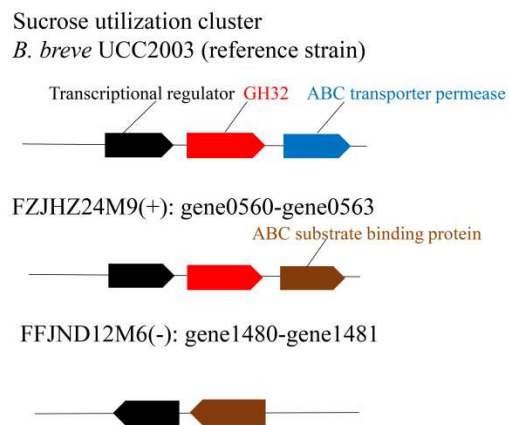
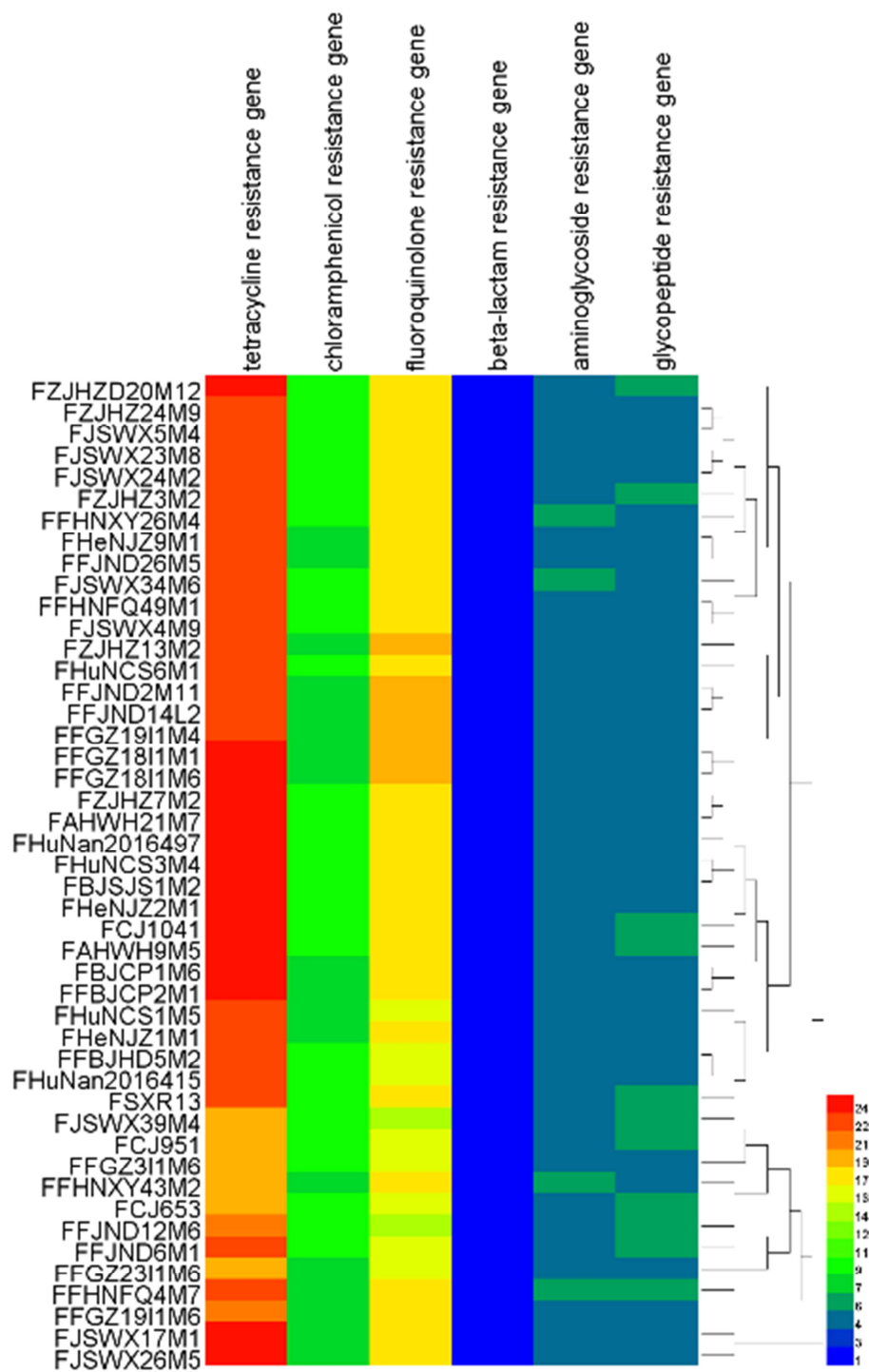
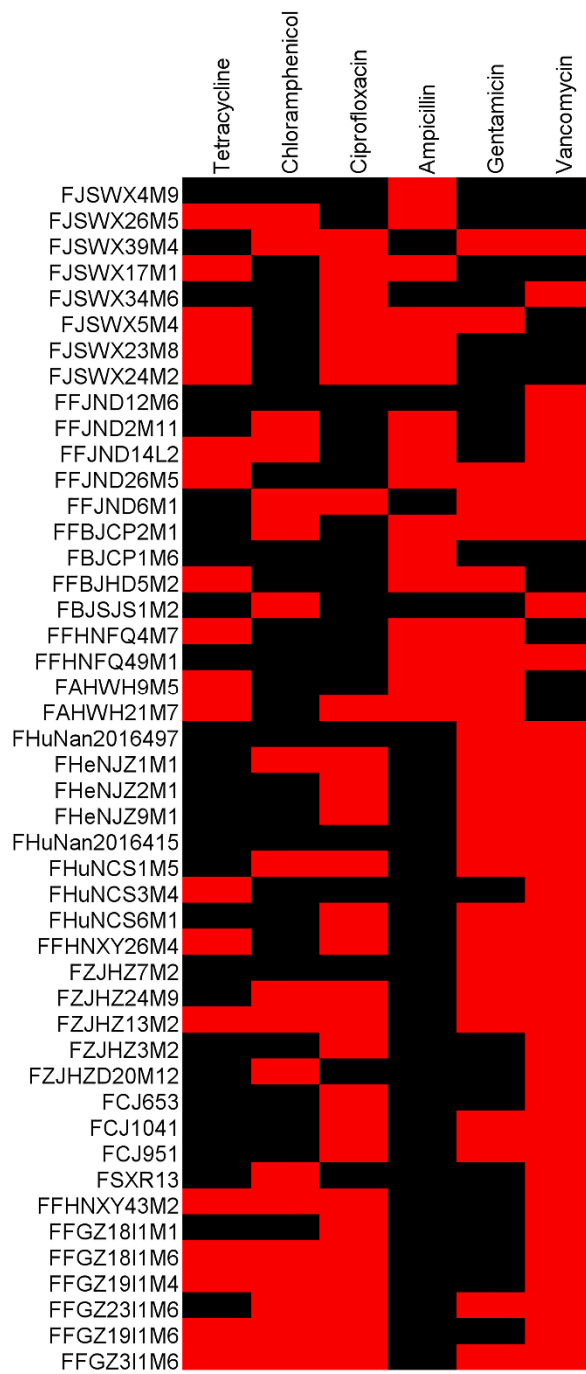


Fig 7

A)



B)



Conflict of interest

All authors declared no conflict of interest.

Journal Pre-proof