Complete genome sequence of *Bifidobacterium adolescentis* P2P3, a human gut bacterium possessing strong resistant starch-degrading activity

Dong-Hyun Jung¹ · Won-Hyong Chung² · Dong-Ho Seo³ · Ye-Jin Kim⁴ · Young-Do Nam² · Cheon-Seok Park⁴

Received: 25 September 2019 / Accepted: 17 December 2019 / Published online: 7 January 2020
© King Abdulaziz City for Science and Technology 2020

Abstract

Resistant starch (RS) is an important food source from which gut bacteria produce short chain fatty acids, which have beneficial effects for human health. The *Bifidobacterium adolescentis* P2P3, a human gut bacterium possessing a strong RS-degrading activity, was isolated from a healthy Korean adult male. In vitro experiments showed that this bacterium could utilize approximately 63% of high amylose corn starch after forming RS granule clusters. Here we provide the first complete set of genomic information on RS-degrading *B. adolescentis* P2P3. The genome of *B. adolescentis* P2P3 consists of one chromosome (2,202,982 bp) with high GC content (59.4%). Analysis of the protein-coding genes revealed that at least nineteen of the starch degradation-related enzymes were present in the genome. Among those, five genes evidently possess carbohydrate-binding domains, which are presumed to be involved in efficient RS decomposition. The complete set of genomic information on *B. adolescentis* P2P3 could provide an understanding of the role of RS-degrading gut bacteria and its RS degradation mechanism.

Keywords *Bifidobacterium adolescentis* · Gut microbiota · Resistant starch · Whole genome

Introduction

Resistant starch (RS) is incompletely digested in the mammalian duodenum and small intestine due to its digestive enzyme-resistant characteristics, and reaches the large intestine as an intact form (Ashwar et al. 2016). It is eventually

fermented by commensal bacteria in the large intestine, and may be metabolized to produce short chain fatty acids (SCFAs). SCFAs are well known as key signaling molecules between the gut microbiota and host health (Morrison and Preston 2016), providing a positive effect on human health. Since RS is not only an energy source for gut microbiota, but also a metabolic basis for production of SCFAs in the large intestine (Bird et al. 2010), the study of human gut bacteria involved in RS metabolism is very important.

Bifidobacteria belonging to the phylum *Actinobacteria* are Gram-positive, non-motile, non-spore forming, and high GC content bacteria (Ventura et al. 2007). Bifidobacteria are commensal bacteria in the human gut, and known to be the first colonization of bacteria in the infant gut (Turroni et al. 2012). They have no pathogenicity and some strains are widely used in the probiotic industry as a 'generally regarded as safe' (GRAS) strain. Bifidobacteria colonized in human gut may help to maintain a beneficial relationship between gut microbiota and the host. These associations result in positive effects on the host health such as protection against viral infection (Saavedra et al. 1994), stimulation of the immune response (Lee et al. 1993), and prevention of human gastrointestinal disorders, colonic adenomas,

Nucleotide sequence accession number: the complete chromosomal sequence of *B. adolescentis* P2P3 was deposited into Genbank under the Accession Number CP024959.

✉ Cheon-Seok Park
cspark@khu.ac.kr

¹ Bacteria Research Team, Nakdonggang National Institute of Biological Resources, Sangju 37242, Republic of Korea

² Research Group of Healthcare, Korea Food Research Institute, Wanju 55365, Republic of Korea

³ Department of Food Science and Technology, College of Agriculture and Life Sciences, Jeonbuk National University, Jeonju 54896, Republic of Korea

⁴ Graduate School of Biotechnology and Institute of Life Science and Resources, Kyung Hee University, Yongin 17104, Republic of Korea

intestinal infection, and cancer (Picard et al. 2005). Recent studies have suggested that *Bifidobacterium* species were specialized for utilization of starch and starch derivatives such as amylopectin, pullulan, and maltooligosaccharides including host glycan (Ryan et al. 2006; Liu et al. 2015; Duranti et al. 2014).

To date, only three studies of in vitro RS-degradation experiments about gut bacteria have been described. The reported RS-degrading bifidobacterial strains are *B. adolescentis* L2-32 from infant feces (Ze et al. 2012) and *B. choerinum* FMB-1 from bovine rumen (Jung et al. 2018). In addition, *Ruminococcus bromii* L2-63 from child feces have been reported (Ze et al. 2012). Among these strains, only the genomic information of *R. bromii* L2-63 (Mukhopadhyay et al. 2018) has been reported as a human-originated bacteria which can decompose RS.

Bifidobacterium adolescentis is considered a taxa comprised of human adult-associated species, since it specifically appears in the gut as humans become adults (Turroni et al. 2011). However, despite this importance, genome information for RS-degrading *B. adolescentis* has not been reported yet. We have isolated *B. adolescentis* P2P3 from a healthy Korean adult male and confirmed that it showed the highest RS degradability in vitro among the strains found so far. Here, we provide genome information and bioinformatic analysis of the gene contents of strain P2P3 as the first complete genome sequence of RS-degrading *B. adolescentis*. The genome analysis based on biochemical experiments and its genetic information will provide useful information related to the analysis of RS-degrading gut bacteria and its RS degradation mechanism.

Materials and methods

Bifidobacterium adolescentis strains

Bifidobacterium adolescentis strains (DSM 20083, DSM 20086, DSM 20087, and DSM 24849), were purchased from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, Braunschweig, Germany) and *B. adolescentis* L2-32 was kindly provided from BEI resources (Manassas, VA, USA). All strains were cultivated according to the manual provided by the corresponding suppliers. Meanwhile, *B. adolescentis* P2P3 was isolated from the feces of a healthy Korean adult male (age 29) and deposited to the Korean Agricultural Culture Collection (KACC, Wanju, Korea) as accession number KACC 92235P.

Chemicals and media

RS 2 type starch (S4180, Sigma-Aldrich, St. Louis, MO, USA), an unmodified high amylose corn starch composed

of 70% amylose and 30% amylopectin, was used for RS degradability analysis. De Man, Rogosa and Sharpe (MRS) medium (Difco, Detroit, MI, USA) was employed as a basic growth medium for *B. adolescentis*. MRS medium (20 mL) without glucose (MRSwoG) (MB Cell, Seoul, Korea) was dissolved in serum vials and autoclaved. Then, 0.1 g of sterilized S4180 (Jung et al. 2018) was mixed with sterilized MRSwoG medium, followed by flushing with 99.5% CO₂ gas set up with a 0.2-μm filter.

Decomposition of resistant starch (RS) by *B. adolescentis* strains

To examine RS degradation by *B. adolescentis* strains, each seed culture was anaerobically grown on 5 mL of MRS medium until the optical density reached 0.5. Subsequently, each one (200 μL) was inoculated into 20 mL of MRSwoG medium containing 0.5% S4180. After incubation for 48 h at 37 °C with shaking at 150 rpm, the amount of residual RS was measured via the phenol–sulfuric acid method (Masuko et al. 2005; DuBois et al. 1956). In brief, 5% phenol solution (100 μL) was added to the culture (100 μL), followed by the addition of 99% sulfuric acid (500 μL) with gentle mixing. After incubation for 15 min at 30 °C, the absorbance (490 nm) was measured using an iMark microplate reader (BioRad, Hercules, CA, USA). Experiments were repeated in triplicate.

Genome sequencing and assembly of *B. adolescentis* P2P3

Bifidobacterium adolescentis P2P3 was grown on MRS medium 37 °C for 18 h. The genomic DNA of *B. adolescentis* P2P3 was extracted and purified using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). The extracted genomic DNA was quantified with a NanoDrop 2000 UV–Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit 2.0 fluorometer (Thermo Fisher Scientific). The whole genome of *B. adolescentis* P2P3 was sequenced with a PacBio RS II (Pacific Biosciences, Menlo Park, CA, USA) sequencing platform. The sequenced reads were assembled using HGAP 3.0 (Chin et al. 2013) with a 2-Mb expected genome size. Chromosome circularization and correction of the genome start position were performed utilizing CGView (Grant and Stothard 2008) and GeneScene (DNASar, Madison, WI, USA). Determination and annotation of the functional genes of *B. adolescentis* P2P3 were carried out with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al. 2016).

Phylogenetic and genome distance analysis between *B. adolescentis* strains

The phylogenetic tree of six *B. adolescentis* strains (P2P3, DSM 20083, DSM 20086, DSM 20087, DSM 24849, and L2-32) was constructed using MEGA7 software (Kumar et al. 2016). Multiple alignment of the 16S rDNA gene (mean length 1473 nt) was performed using EBI-Clustal Omega (Madeira et al. 2019). The neighbor-joining statistical method was applied with bootstrap replications of 1000 datasets. 16S rDNA accessions of each strain are MH828351 (P2P3), NC_008618.1 (DSM 20083), LNKJ01000005.1 (DSM 20086), JNKM01000001.1 (DSM 20087), JDUX01000017.1 (DSM 24849), and AY305304.1 (L2-32). Average nucleotide identity (ANI) values were computed by OrthoANI software (Lee et al. 2016) using NCBI-BLAST+ 2.9.0 program (Zhang et al. 2000) with original ANI option. Six genomes *B. adolescentis* strains, used for the calculation, were downloaded from the NCBI RefSeq assembly data (Accession number of strain P2P3: GCF_003856735.1, DSM 20083: GCF_000010425.1, DSM 20086: GCF_002108135.1, DSM 20087: GCF_000702865.1, DSM 24849: GCF_000741415.1, and L2-32: GCF_000154085.1).

Functional classification of coding sequences in *B. adolescentis* P2P3

The functional classification of protein-coding genes of six *B. adolescentis* strains were predicted using BLASTP search with two criteria: $1e-5$ of *e*-value and 50% minimum coverage from the NCBI Clusters of Orthologous Groups (COG) database. Each gene was classified into COG categories (Tatusov et al. 1997) and unassigned genes were described as 'Not assigned'. The protein data of six *B. adolescentis* are from NCBI RefSeq assembly data, previously sequenced by PGAP.

Analysis of amylolytic enzymes in *B. adolescentis* P2P3

Prediction of signal peptides (Sec- and Tat-secretion) of amylolytic enzymes were performed with CBS-SignalP 5.0 server (Armenteros et al. 2019). The conserved domains of amylolytic enzymes were defined by NCBI-CD search (Marchler-Bauer et al. 2016). Cell-anchored proteins of amylolytic enzymes were detected using the CW-PRED server (Litou et al. 2008).

Results and discussion

Strong resistant starch-degrading human gut bacterium *Bifidobacterium adolescentis* P2P3

Human gut bacterium *B. adolescentis* P2P3 was anaerobically cultured with 0.5% non-gelatinized high-amylose

corn starch (HACS) (S4180) for 48 h, since almost no decomposition occurs after 48 h. Simultaneously, other *B. adolescentis* strains (DSM 20083, DSM 20086, DSM 20087, DSM 24849, and L2-32) whose genomes are completely analyzed were obtained from two microbial culture collections (DSMZ and BEI Resources) and cultured with the same conditions. *Bifidobacterium adolescentis* P2P3 was observed to utilize RS granules (Fig. 1). During incubation for 48 h, the RS granules were decomposed and finally disappeared. An interesting point is that the dispersed RS granules (Fig. 1a) began to form coagulated clusters as soon as the cells were inoculated (Fig. 1b). In addition, many cells were observed around RS granules clusters during RS granule degradation (Fig. 1d, e). RS-degrading ability and RS cluster formation by six *B. adolescentis* strains was compared (Table 1). Four *B. adolescentis* strains (P2P3, DSM 20087, DSM 24849, and L2-32) could utilize RS at levels of 63.3%, 9.9%, 47.3%, and 43.8%, respectively, and they were found to form RS granule clusters. Two strains (DSM 20083 and DSM 20086) did not degrade RS at all and they did not form RS granule clusters. The only reported RS-degrading *B. adolescentis* strain so far is the strain L2-32 mentioned in the *R. bromii* L2-63 study (Ze et al. 2012). In our RS decomposition experiment, strain P2P3 could utilize 20% more RS than strain L2-32. In addition, the RS granule clusters were formed more in strain P2P3 than the strain L2-32. Based on these results, strain P2P3 was regarded as the most efficient RS-utilizing strain. Therefore, the complete genome sequencing of strain P2P3 was performed to provide information about the RS degradation mechanism.

The complete genome of *B. adolescentis* P2P3

The whole genome of *B. adolescentis* P2P3 was composed of one circular sequence, which was a 2,202,982-bp chromosome with 59.4% GC content (Fig. 2 and Table 2). A total of 1848 genes were identified in the genome, including 1778 protein-coding genes, 70 RNA genes, and 60 pseudogenes. The full-length rRNA gene set, including 5S, 16S, and 23S, were placed in the genome; 54 tRNA genes and three non-coding RNA genes were identified. So far, six complete genome sequences of *B. adolescentis* species have been reported to the NCBI genome, including the P2P3 strain. Among them, the strain P2P3 is the first complete genome sequence reported as an RS-degrading *B. adolescentis* species. The six complete genome sequences of *B. adolescentis* species revealed genome sizes ranging from 2,089,645 to 2,389,110 bp and GC content was almost the same at 59%. The number of coding genes was slightly different, ranging from 1634 to 2005. No CRISPR array was identified of the genome of strain

Fig. 1 The degradation of RS by *B. adolescentis* P2P3. Optical microscopic images of **a** dispersed RS granules without cells, **b** clustered RS granules immediately after inoculation of cells, **c** slightly

degraded RS granule clusters after 12 h of incubation, **d**, **e** fairly degraded RS granule clusters after 24 h and 32 h, respectively, and **f** almost degraded RS granule clusters after 48 h

Table 1 The RS-degradation characteristics of six *B. adolescentis* strains

Strain	Origin	RS degradability (%)	Formation of RS clusters
<i>B. adolescentis</i> P2P3	Adult feces in 2017	63.3 ± 3.1	++++
<i>B. adolescentis</i> DSM 20083	Intestine of adult before 1990	2.8 ± 2.1	—
<i>B. adolescentis</i> DSM 20086	Intestine of adult before 1990	2.6 ± 1.3	—
<i>B. adolescentis</i> DSM 20087	Bovine rumen before 1990	9.9 ± 1.9	+
<i>B. adolescentis</i> DSM 24849	Adult feces in 2011	47.3 ± 4.8	+++
<i>B. adolescentis</i> L2-32	Infant feces in 1996	43.8 ± 3.1	++

Formation of RS clusters is designated by —, +, ++, +++, and +++++. The '—' indicates no formation of RS clusters. When the size of RS clusters is between 150 and 200 μm, it is expressed as '+++++'. Similarly, the '+++++' indicates the size of RS clusters of 100 and 150 μm, whereas the '++' displays the size of RS clusters of 50–100 μm. The '+' implies the size of RS clusters of 0–50 μm

P2P3 and DSM 24849, whereas other strains have one or two CRISPR arrays.

Functional classification of the protein-coding genes of *B. adolescentis* P2P3

The protein-coding genes of *B. adolescentis* P2P3 were functionally classified based on the COG categories. Of the 1718 protein-coding genes in strain P2P3, 1289 proteins were classified except for 429 unassigned proteins (Table 3).

There were slight differences among six strains. 'Replication, recombination and repair (L)', and 'Not assigned protein' of strain DSM 24849 and L2-32 were present more than other strains. Only one 'Cell motility (N)' was present in strain DSM 20083, DSM 20086, and DSM 20087 compared to other strains. Also, strain DSM 20087, DSM 24849, and L2-32 had more 'Secondary metabolites biosynthesis, transport, and catabolism (Q)' than other strains. However, as long as six strains belonged to the same species,

Fig. 2 Visualized circular map of chromosome of *B. adolescentis* P2P3. 2.2 Mbp of chromosome consists of seven tracks. From the outermost, track 1 (deep blue): forward-strand coding genes, track 2

(blue): reverse-strand coding genes, track 3 (light blue): tRNAs, track 4 (Orange): rRNAs, track 5 (black): GC content, and track 6 (green and purple): G+C skew

the proportions of other code's distribution were not significantly different.

Phylogenetic and ANI analysis

Figure 3a shows the phylogenetic analysis result of the six *B. adolescentis* strains including P2P3. The phylogenetic relationships between *B. adolescentis* strains were very close to each other. The percent identity between strains range from 98.75 to 99.92%. The highest RS-utilizing strain P2P3 (63.3%) is rather close to strain DSM 20086 and DSM 20083, which cannot utilize RS, whereas the RS-utilizing strain DSM 24849 and L2-32 (47.3% and 43.8%,

respectively) are close to DSM 20087 with low RS utilization (9.9%). As a result, there was no significant correlation between phylogenetic relationships and RS utilization, even if they are in the same species. The genome-wide comparison was analyzed for six *B. adolescentis* strain genomes including P2P3. As shown in Fig. 3b, six genomes of *B. adolescentis* strains were very close to each other with ANI values ranging from a minimum of 97.29% to a maximum of 99.97%. In particular, RS-utilizing strains (Strain P2P3, DSM 24849, and L2-32) appear to be closely related to each other. Although not all *B. adolescentis* strains have been compared, RS utilization strains are assumed to be

Table 2 Genome characteristics of six *B. adolescentis* strains

<i>B. adoles- centis</i>	P2P3	DSM 20083	DSM 20086	DSM 20087	DSM 24849	L2-32
<i>Genome</i>						
Assembly accession	GCF_003856735.1	GCF_000010425.1	GCF_002108135.1	GCF_000702865.1	GCF_000741415.1	GCF_000154085.1
Assembly level	Complete	Complete	Scaffolds	Scaffolds	Scaffolds	Scaffolds
No. of sequences	1 chromosome	1 chromosome	8 contigs	10 contigs	40 contigs	40 contigs
Total size (bp)	2,202,982	2,089,645	2,084,232	2,051,152	2,291,980	2,389,110
GC content (%)	59.4	59.2	59.2	59.4	59.4	59.2
<i>Annotation</i>						
Total genes	1848	1749	1753	1700	1966	2084
Total CDS	1778	1676	1677	1634	1898	2005
Coding genes	1718	1624	1599	1586	1831	1906
Pseudo genes	60	52	78	48	67	99
RNAs	70	73	76	66	68	79
rRNA (5S, 16S, 23S)	5, 4, 4	6, 5, 5	6, 6, 7	4, 3, 2	5, 2, 4	5, 7, 5
Complete	5, 4, 4	6, 5, 5	6, 4, 3	4, 0, 0	5, 2, 0	5, 2, 3
Partial	—	—	0, 2, 4	0, 3, 2	0, 0, 4	0, 5, 2
tRNA	54	54	54	54	54	59
ncRNA	3	3	3	3	3	3
CRISPR repeats	0	1	2	1	0	1

evolutionarily different strains that adapted to specific environments rich in RS.

Putative amylolytic enzymes in *B. adolescentis* P2P3

The amylolytic enzymes of *B. adolescentis* P2P3 were organized and analyzed to provide enzyme information related to RS degradation. Nineteen genes encoding amylolytic enzymes active on α -glucan substrates were found in the genome of *B. adolescentis* P2P3. Details of the enzymes are listed in Table 4. These various amylolytic enzymes are expected to participate directly or indirectly in the degradation of RS granules and small saccharides released from RS granules. Of these enzymes, three putative α -amylases, glycogen debranching enzyme (WP_011743102.1) and pullulanase (WP033499429.1) have carbohydrate-binding modules (CBMs), which may be involved in the attachment of enzyme(s) to the substrate. Both enzymes have a potential to stick to and decompose RS. The rationale behind this prediction was that some α -amylolytic enzymes acting on starch granules commonly possess CBMs as auxiliary modules. Such auxiliary domains help enzymes attach starch granules

and aid the active site of enzymes to locate the right position with precision (Guillén et al. 2010; Rodriguez-Sanoja et al. 2005; Southall et al. 1999). In particular, pullulanase (WP033499429.1) has three CBMs (53 and two 41) and seems to be signaled out of the cell by the Sec signaling. Pullulanase is thought to be attached to the outer cell wall by the LPXTG motif (Navarre and Schneewind 1994; Ton-That et al. 2004). Also, the pullulanase have two catalytic domains of GH13 (α -Amylase and pullulanase). Such amylolytic enzymes contained CBMs (WP_011743102.1 and WP033499429.1) may act as first degrader of RS granules and the small saccharides released from RS can be degraded by other amylases and glucosidases. Three putative α -amylases possess a typical catalytic domain of GH13 with several CBMs. Therefore, these enzymes may be involved in the decomposition of RS. Additional detailed biochemical studies are in progress to support this hypothesis. In conclusion, the whole genome sequence of *B. adolescentis* P2P3 will provide basic information on the RS degradation mechanism in the gut bacteria and later will support the scientific background to develop healthful probiotics along with their biochemical studies.

Table 3 Number of genes in six *B. adolescentis* strains associated with general COG functional categories

Code	Description	P2P3	DSM 20083	DSM 20086	DSM 20087	DSM 24849	L2-32
J	Translation, ribosomal structure and biogenesis	131	130	129	131	132	131
A	RNA processing and modification	1	1	1	1	2	1
K	Transcription	93	93	94	86	97	101
L	Replication, recombination and repair	93	89	85	86	102	105
B	Chromatin structure and dynamics	—	—	—	—	—	—
D	Cell cycle control, cell division, chromosome partitioning	24	21	21	22	23	24
Y	Nuclear structure	—	—	—	—	—	—
V	Defense mechanisms	45	41	40	36	50	49
T	Signal transduction mechanisms	43	43	42	39	41	41
M	Cell wall/membrane/envelope biogenesis	65	63	62	62	63	76
N	Cell motility	7	1	1	1	8	10
Z	Cytoskeleton	—	—	—	—	—	—
W	Extracellular structure	—	—	—	—	—	—
U	Intracellular trafficking, secretion, and vesicular transport	10	12	12	11	13	11
O	Post-translational modification, protein turnover, and chaperones	45	43	43	45	44	47
X	Mobilome: prophages, and transposons	—	—	—	—	—	—
C	Energy production and conversion	42	41	41	40	41	42
G	Carbohydrate transport and metabolism	145	147	143	135	147	166
E	Amino acid transport and metabolism	155	159	158	179	157	156
F	Nucleotide transport and metabolism	53	54	53	56	55	53
H	Coenzyme transport and metabolism	39	39	38	43	41	39
I	Lipid transport and metabolism	37	33	33	35	35	37
P	Inorganic ion transport and metabolism	56	54	54	52	58	59
Q	Secondary metabolites biosynthesis, transport, and catabolism	1	3	3	6	5	5
R	General function prediction only	118	111	111	123	120	118
S	Function unknown	86	86	82	81	93	88
	Not assigned	429	360	353	316	504	547

Fig. 3 Relationship between *B. adolescentis* strains based on sequence alignment. The number in the box indicates percent identity. **a** Phylogenetic analysis based on 16S rDNA. The bar represents0.0005 substitutions per site. **b** Genome distance analysis (ANI). The number in the tree represents the branch length

Table 4 The genes involved in the hydrolysis of α -1,4 and 1,6-glucosidic linkages in *B. adolescentis* P2P3

Protein ID	Length (AA)	Function	Signal peptide	Anchoring protein	CBMs	Enzyme class
WP_117838111.1	818	Glycogen phosphorylase	–	–	–	GT1
WP_124917072.1	720	4- α -Glucanotransferase	–	–	–	GH77
WP_033499301.1	815	Glycogen debranching enzyme GlgX	–	–	–	GH13
WP_124917091.1	585	α -Amylase	–	–	–	GH13
WP_003809562.1	751	1,4- α -Glucan branching enzyme	–	–	–	GH13
WP_011743102.1	714	Glycogen debranching enzyme GlgX	–	–	48	GH13
WP_021913761.1	545	α -Amylase	–	–	–	GH13
WP_033499429.1	1764	Pullulanase, type I	Sec/SPI	LPXTG motif	53, 41, 41	GH13 $\times$ 2
WP_070122448.1	591	α -Amylase	–	–	–	GH13
WP_124917489.1	1107	Hypothetical protein (putative α -amylase)	Sec/SPI	Chw	26, 26, 25	GH13
WP_124917356.1	1219	Hypothetical protein (putative α -amylase)	Sec/SPI	Chw	26, 25, 25	GH13
WP_124917357.1	1426	Hypothetical protein (putative α -amylase)	Sec/SPI	Chw	26, 25, 25	GH13
WP_124917371.1	752	α -1,4-Glucan-maltose-1-phosphate maltosyltransferase	Tat/SPI	–	–	GH13
WP_124917408.1	684	Pullulanase	–	–	–	GH13
WP_124917409.1	543	α -Amylase	–	–	–	GH13
WP_039775972.1	722	4- α -Glucanotransferase	–	–	–	GH77
WP_003811282.1	582	α -Amylase	–	–	–	GH13
WP_046999836.1	606	α -Glucosidase	–	–	–	GH13
WP_124917425.1	841	α -Glucosidase	–	–	–	GH31

Chw clostridial hydrophobic domain, with a conserved tryptophan

Acknowledgements This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2017R1A2B4004218).

Author contributions C-S P designed and coordinated all the experiments; D-H J performed the experiments, analyzed the genome data and wrote the manuscript; W-H C, Y-D N and D-H S performed the genome sequencing and sequence assembly; Y-J K measured residual RS. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to declare.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of human derivatives were followed. The study of human derivatives was approved by the Institutional Review Board of Kyung Hee University (KHSIRB-17-004).

References

- Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420
- Ashwar BA, Gani A, Shah A, Wani IA, Masoodi FA (2016) Preparation, health benefits and applications of resistant starch: a review. *Starch-Starke* 68:287–301
- Bird A, Conlon M, Christophersen C, Topping D (2010) Resistant starch, large bowel fermentation and a broader perspective of prebiotics and probiotics. *Benef Microbes* 1:423–431
- Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569
- DuBois M, Gilles KA, Hamilton JK, Pt R, Smith F (1956) Colorimetric method for determination of sugars and related substances. *Anal Chem* 28:350–356
- Duranti S, Turrone F, Lugli GA, Milani C, Viappiani A, Mangifesta M, Gioiosa L, Palanza P, van Sinderen D, Ventura MJ (2014) Genomic characterization and transcriptional studies of the starch-utilizing strain *Bifidobacterium adolescentis* 22L. *Appl Environ Microbiol* 80:6080–6090
- Grant JR, Stothard P (2008) The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* 36:W181–W184
- Guillén D, Sánchez S, Rodríguez-Sanoja R (2010) Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol* 85:1241–1249
- Jung D-H, Seo D-H, Kim G-Y, Nam Y-D, Song E-J, Yoon S, Park C-S (2018) The effect of resistant starch (RS) on the bovine rumen microflora and isolation of RS-degrading bacteria. *Appl Microbiol Biotechnol* 102:4927–4936
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
- Lee J, Ametani A, Enomoto A, Sato Y, Motoshima H, Ike F, Kami-nogawa S (1993) Screening for the immunopotentiating activity of food microorganisms and enhancement of the immune response by *Bifidobacterium adolescentis* M101–4. *Biosci Biotechnol Biochem* 57:2127–2132

- Lee I, Kim YO, Park S-C, Chun J (2016) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 66:1100–1103
- Litou ZI, Bagos PG, Tsirigos KD, Liakopoulos TD, Hamodrakas SJ (2008) Prediction of cell wall sorting signals in gram-positive bacteria with a hidden Markov model: application to complete genomes. *J Bioinform Comput Biol* 6:387–401
- Liu S, Ren F, Zhao L, Jiang L, Hao Y, Jin J, Zhang M, Guo H, Lei X, Sun E (2015) Starch and starch hydrolysates are favorable carbon sources for *Bifidobacteria* in the human gut. *BMC Microbiol* 15:54
- Madeira F, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey A, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47:W636–W641
- Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200–D203
- Masuko T, Minami A, Iwasaki N, Majima T, Nishimura S-I, Lee YC (2005) Carbohydrate analysis by a phenol–sulfuric acid method in microplate format. *Anal Biochem* 339:69–72
- Morrison DJ, Preston T (2016) Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* 7:189–200
- Mukhopadhyay I, Morais S, Laverde-Gomez J, Sheridan PO, Walker AW, Kelly W, Klieve AV, Ouwerkerk D, Duncan SH, Louis P (2018) Sporulation capability and amylose conservation among diverse human colonic and rumen isolates of the keystone starch-degrader *Ruminococcus bromii*. *Environ Microbiol* 20:324–336
- Navarre WW, Schneewind O (1994) Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in Gram-positive bacteria. *Mol Microbiol* 14:115–121
- Picard C, Fioramonti J, Francois A, Robinson T, Neant F, Matuchansky C (2005) *Bifidobacteria* as probiotic agents-physiological effects and clinical benefits. *Aliment Pharmacol Ther* 22:495–512
- Rodriguez-Sanoja R, Ruiz B, Guyot J-P, Sanchez S (2005) Starch-binding domain affects catalysis in two *Lactobacillus* α -amylases. *Appl Environ Microbiol* 71:297–302
- Ryan SM, Fitzgerald GF, van Sinderen D (2006) Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. *Appl Environ Microbiol* 72:5289–5296
- Saavedra JM, Bauman NA, Perman J, Yolken R, Oung I (1994) Feeding of *Bifidobacterium bifidum* and *Streptococcus thermophilus* to infants in hospital for prevention of diarrhoea and shedding of rotavirus. *Lancet* 344:1046–1049
- Southall SM, Simpson PJ, Gilbert HJ, Williamson G, Williamson MP (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett* 447:58–60
- Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631–637
- Tatsova T, DiCuccio M, Badretin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624
- Ton-That H, Marraffini LA, Schneewind O (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim Biophys Acta-Mol Cell Res* 1694:269–278
- Turroni F, Van Sinderen D, Ventura M (2011) Genomics and ecological overview of the genus *Bifidobacterium*. *Int J Food Microbiol* 149:37–44
- Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F (2012) Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE* 7:e36957
- Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen D (2007) Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* 71:495–548
- Ze X, Duncan SH, Louis P, Flint HJ (2012) *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J* 6:1535–1543
- Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203–214