



# Mechanism of Cooperative Degradation of Gum Arabic Arabinogalactan Protein by *Bifidobacterium longum* Surface Enzymes

Yuki Sasaki,<sup>a</sup> Masahiro Komeno,<sup>b</sup>  Akihiro Ishiwata,<sup>c</sup> Ayako Horigome,<sup>d</sup>  Toshitaka Odamaki,<sup>d</sup> Jin-Zhong Xiao,<sup>d</sup> Katsunori Tanaka,<sup>c,e</sup> Yukishige Ito,<sup>c,f</sup> Kanefumi Kitahara,<sup>a,g</sup> Hisashi Ashida,<sup>b,h</sup>  Kiyotaka Fujita<sup>a,g</sup>

<sup>a</sup>The United Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima, Kagoshima, Japan

<sup>b</sup>Graduate School of Biology-Oriented Science and Technology, Kindai University, Kinokawa, Wakayama, Japan

<sup>c</sup>RIKEN Cluster for Pioneering Research, Wako, Saitama, Japan

<sup>d</sup>Next Generation Science Institute, Morinaga Milk Industry Co., Ltd., Zama, Kanagawa, Japan

<sup>e</sup>Department of Chemical Science and Engineering, Tokyo Institute of Technology, Meguro, Tokyo, Japan

<sup>f</sup>Graduate School of Science, Osaka University, Toyonaka, Osaka, Japan

<sup>g</sup>Faculty of Agriculture, Kagoshima University, Kagoshima, Kagoshima, Japan

<sup>h</sup>Faculty of Biology-Oriented Science and Technology, Kindai University, Kinokawa, Wakayama, Japan

**ABSTRACT** Gum arabic is an arabinogalactan protein (AGP) that is effective as a prebiotic for the growth of bifidobacteria in the human intestine. We recently identified a key enzyme in the glycoside hydrolase (GH) family 39, 3-*O*- $\alpha$ -D-galactosyl- $\alpha$ -L-arabinofuranosidase (GAfase), for the assimilation of gum arabic AGP in *Bifidobacterium longum* subsp. *longum*. The enzyme released  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara and  $\beta$ -L-Arap-(1 $\rightarrow$ 3)-L-Ara from gum arabic AGP and facilitated the action of other enzymes for degrading the AGP backbone and modified sugar. In this study, we identified an  $\alpha$ -L-arabinofuranosidase (BIArafE; encoded by BLLJ\_1850), a multidomain enzyme with both GH43\_22 and GH43\_34 catalytic domains, as a critical enzyme for the degradation of modified  $\alpha$ -L-arabinofuranosides in gum arabic AGP. Site-directed mutagenesis approaches revealed that the  $\alpha$ 1,3/ $\alpha$ 1,4-Araf double-substituted gum arabic AGP side chain was initially degraded by the GH43\_22 domain and subsequently cleaved by the GH43\_34 domain to release  $\alpha$ 1,3-Araf and  $\alpha$ 1,4-Araf residues, respectively. Furthermore, we revealed that a tetrasaccharide,  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\beta$ -D-GlcpA-(1 $\rightarrow$ 6)- $\beta$ -D-Galp-(1 $\rightarrow$ 6)-D-Gal, was a limited degradative oligosaccharide in the gum arabic AGP fermentation of *B. longum* subsp. *longum* JCM7052. The oligosaccharide was produced from gum arabic AGP by the cooperative action of the three cell surface-anchoring enzymes, GAfase, exo- $\beta$ 1,3-galactanase (BI1,3Gal), and BIArafE, on *B. longum* subsp. *longum* JCM7052. Furthermore, the tetrasaccharide was utilized by the commensal bacteria.

**IMPORTANCE** Terminal galactose residues of the side chain of gum arabic arabinogalactan protein (AGP) are mainly substituted by  $\alpha$ 1,3/ $\alpha$ 1,4-linked Araf and  $\beta$ 1,6-linked  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\beta$ -D-GlcpA residues. This study found a multidomain BIArafE with GH43\_22 and GH43\_34 catalytic domains showing cooperative action for degrading  $\alpha$ 1,3/ $\alpha$ 1,4-linked Araf of the side chain of gum arabic AGP. In particular, the GH43\_34 domain of BIArafE was a novel  $\alpha$ -L-arabinofuranosidase for cleaving the  $\alpha$ 1,4-Araf linkage of terminal galactose.  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\beta$ -D-GlcpA-(1 $\rightarrow$ 6)- $\beta$ -D-Galp-(1 $\rightarrow$ 6)-D-Gal tetrasaccharide was released from gum arabic AGP by the cooperative action of GAfase, GH43\_24 exo- $\beta$ -1,3-galactanase (BI1,3Gal), and BIArafE and remained after *B. longum* subsp. *longum* JCM7052 culture. Furthermore, *in vitro* assimilation test of the remaining oligosaccharide using *Bacteroides* species revealed that cross-feeding may occur from bifidobacteria to other taxonomic groups in the gut.

**Editor** Haruyuki Atomi, Kyoto University

**Copyright** © 2022 Sasaki et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Kiyotaka Fujita, k4022897@kadai.jp.

The authors declare a conflict of interest. Three authors, A.H., T.O., and J.-Z.X., are employees of Morinaga Milk Industry Co., Ltd.

**Received** 4 November 2021

**Accepted** 22 January 2022

**Accepted manuscript posted online**

2 February 2022

**Published** 22 March 2022

**KEYWORDS** arabinofuranosidase, arabinogalactan protein, *Bacteroides*, bifidobacteria, glycosidase, gum arabic, prebiotics

Arabinogalactan protein (AGP) is one of the highly complex glycoprotein components of plant cell walls. It is widely found in edible plant parts, such as vegetables, fruits, and cereals (1). The carbohydrate component of AGP is composed of type II arabinogalactan (AG) chains comprising a  $\beta$ 1,3-galactan backbone chain with  $\beta$ 1,6-galactan side chains. The side chain structure of AGPs varies among plants and organs. In particular, gum arabic is an AGP with complex sugar modifications at the branch (2) and is reported to act as a prebiotic without digestion by host enzymes and increases bifidobacteria levels in the large intestine (3, 4). In the gum arabic AGP, the terminal galactose residues of the side chain are mainly substituted by  $\alpha$ 1,3-linked Araf,  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara, or  $\beta$ -L-Arap-(1 $\rightarrow$ 3)-L-Ara and  $\alpha$ 1,4-linked Araf and  $\beta$ 1,6-linked  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)-D-GlcP residues (Fig. 1) (5–8).

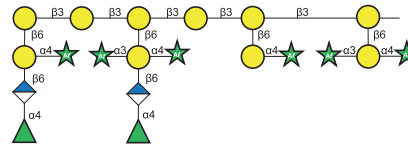
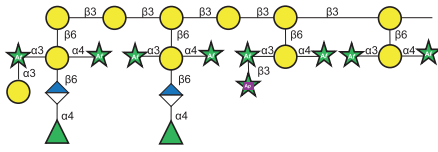
Recently, the mechanisms involved in the assimilation of gum arabic AGP by human gut microbiota have been revealed for *Bacteroides* and bifidobacteria. Regarding *Bacteroides*, Cartmell et al. revealed the function of polysaccharide utilization loci involved in the degradation of gum arabic AGP in *B. thetaiotaomicron* and *B. cellulosilyticus* and reported that the cell surface endo- $\beta$ 1,3-galactanase was a key enzyme for assimilating gum arabic AGP (7). They also characterized many other related enzymes for trimming modified sugar in the periplasm. For the degradation of  $\alpha$ -L-Araf,  $\alpha$ -L-arabinofuranosidase (BT3675) was characterized to hydrolyze  $\alpha$ 1,3-Araf linkage but not  $\alpha$ 1,4-Araf linkage, but  $\alpha$ 1,4-Araf-specific  $\alpha$ -L-arabinofuranosidase has not been detected yet. In bifidobacteria, we found a key enzyme in the glycoside hydrolase (GH) family 39, namely, 3-O- $\alpha$ -D-galactosyl- $\alpha$ -L-arabinofuranosidase (GAFase), for the assimilation of gum arabic AGP in *Bifidobacterium longum* subsp. *longum* JCM7052 (9). The enzyme released  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara and  $\beta$ -L-Arap-(1 $\rightarrow$ 3)-L-Ara from gum arabic AGP. We also found that GAFase released disaccharides as a carbon source and promoted the action of other type II AG-degrading enzymes. However, the combination of these type II AG degradative enzymes (GAFase, exo- $\beta$ 1,3-galactanase [B11,3Gal], and  $\alpha$ -L-arabinofuranosidase [BIArafA]) used in the previous study could not release the same amount of L-arabinose as that released by the bacterial cell fraction used. Therefore, it remains unclear which  $\alpha$ -L-arabinofuranosidases mainly act on  $\alpha$ 1,4-Araf linkage of gum arabic AGP.

Based on the Carbohydrate-Active enZymes (CAZy) database, *B. longum* subsp. *longum* JCM1217 encodes nine  $\alpha$ -L-arabinofuranosidase candidates, of which six are cell surface-anchoring GH43s and three are intracellular GH51s. In GH43s, HypAA (BLLJ\_0213) contained the GH43 subfamily 29 (GH43\_29) domain and was reported to be an  $\alpha$ 1,3-Araf-specific  $\alpha$ -L-arabinofuranosidase for degrading  $\alpha$ -L-Araf-(1 $\rightarrow$ 3)- $\beta$ -L-Araf-(1 $\rightarrow$ 2)- $\beta$ -L-Araf-(1 $\rightarrow$ 2)- $\beta$ -L-Araf-Hyp (10). BLLJ\_1850 to BLLJ\_1854 are arranged in tandem in the gene cluster, and we previously reported that BIArafA (BLLJ\_1854) contained the GH43 subfamily 22 (GH43\_22) domain acting on the  $\alpha$ 1,3-Araf residue of larch AGP (11). BIArafB (BLLJ\_1853) contained the GH43\_22 domain, which hydrolyzes the  $\alpha$ 1,5-Araf linkage of arabinan backbone, and BIArafC (BLLJ\_1852) contained the GH43 subfamily 27 (GH43\_27) domain, which hydrolyzes the  $\alpha$ 1,2- and  $\alpha$ 1,3-Araf linkages on arabinan side chains (12). Recently, it was discovered that BIArafD (BLLJ\_1851) and BIArafE (BLLJ\_1850) cooperatively degrade arabinoxylan (M. Komeno, Y. Yoshihara, J. Kawasaki, W. Nabeshima, K. Maeda, Y. Sasaki, K. Fujita, H. Ashida, unpublished data). BIArafD (BLLJ\_1851) contained GH43\_UC and GH43\_26 domains. GH43\_UC acts on the  $\alpha$ 1,2-Araf linkages of the  $\alpha$ 1,2- and  $\alpha$ 1,3-Araf doubly substituted arabinoxylan, whereas GH43\_26 acts on the arabinan backbone. BIArafE (BLLJ\_1850) contains GH43\_22 and GH43\_34 domains, wherein GH43\_22 acts on the  $\alpha$ 1,3-Araf linkage in arabinoxylan. However, the specificity of GH43\_34 remains unclear.

The  $\alpha$ -L-Araf structure presented  $\alpha$ 1,2/1,3/1,5-linkages in some components of the plant cell wall (13), such as arabinan (14, 15), arabinoxylan (16), arabinoxyloglucan (17, 18), AG (6, 19, 20), hydroxyproline-linked  $\beta$ -L-arabinooligosaccharides (21), and

**Gum arabic AGP**

**$\alpha$ -D-Galp-(1→3)-L-Ara-free gum arabic AGP**



**Gum arabic AGP-related oligosaccharides**

Substitution on Gal at O6	$\alpha$ -L-Rhap-(1→4)- $\beta$ -D-GlcA-(1→6)-	-	
at O4	$\alpha$ -L-Araf-(1→4)-	-	$\alpha$ -L-Araf-(1→4)-
at O3			
$\alpha$ -D-Galp-(1→3)- $\alpha$ -L-Araf-(1,3)-	<b>S5-GA</b> 	<b>S4-GA</b> 	<b>S3-GA</b> 
$\alpha$ -L-Araf-(1,3)-	<b>S5-A</b> 	<b>S4-A</b> 	<b>S3-A</b> 
1	<b>S5</b> 	<b>S4</b> 	<b>S3</b> 

● D-Galactose

★ L-Arabinofuranose

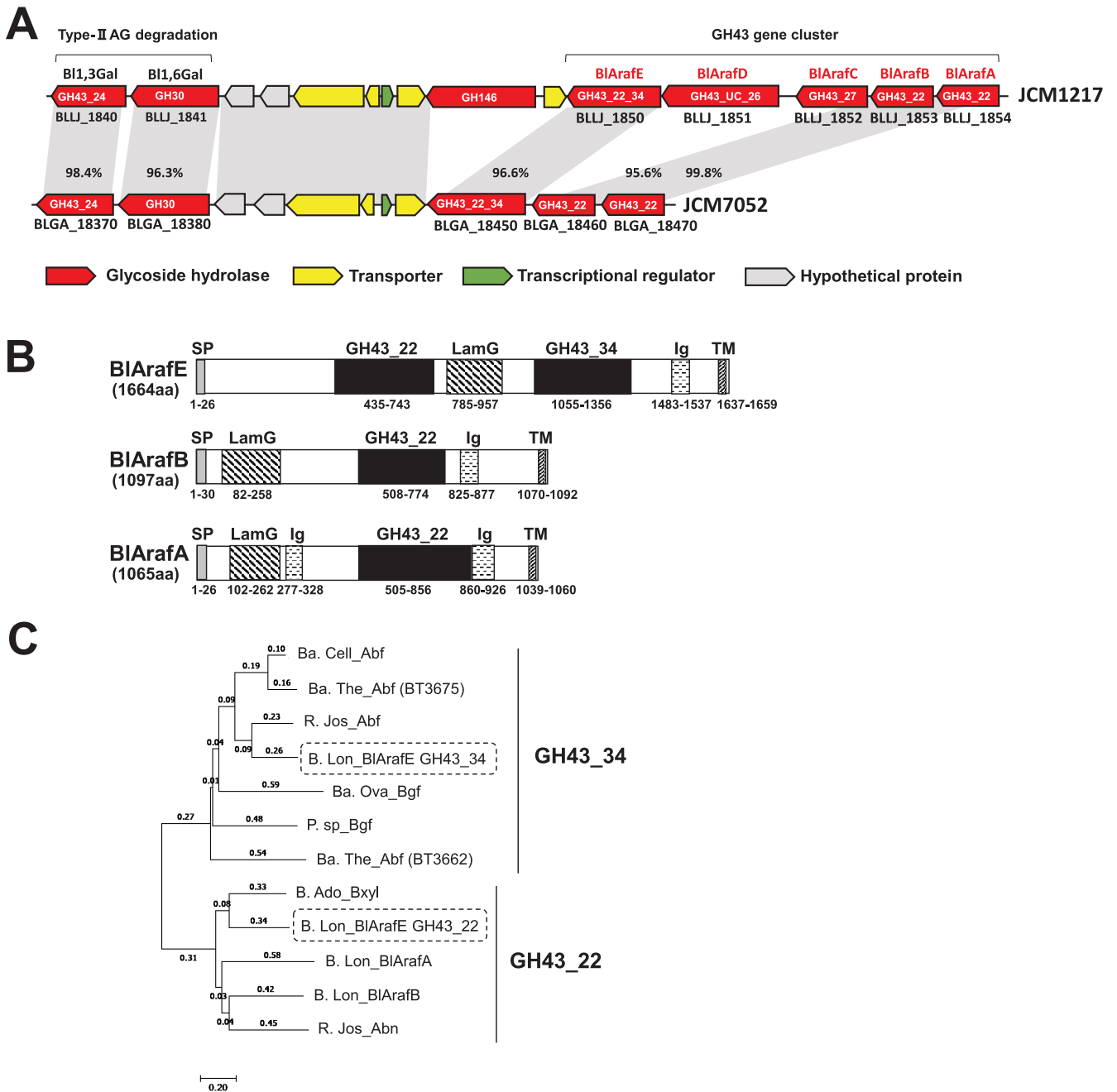
☆ L-Arabinopyranose

◊ D-Glucuronic acid

▲ L-Rhamnose

**FIG 1** Schematic structures of gum arabic AGP,  $\alpha$ -D-Galp-(1→3)-L-Ara-free gum arabic AGP, and gum arabic AGP-related oligosaccharides used in this study.

rhamnogalacturonan (RG) II (22). Conversely, the  $\alpha$ 1,4-linkage of L-arabinose was found in the RG-II structure of pectin in the pyranose form (22), and to the best of our knowledge, the  $\alpha$ 1,4-Araf structure was only found in gum arabic AGP among plant-derived glycans (8, 23). Hence,  $\alpha$ 1,4-Araf-specific  $\alpha$ -L-arabinofuranosidase has not been characterized so far in any organism due to this limited localization. The removal of the  $\alpha$ 1,3-linkage by GAfase was expected to remove the steric hindrance for an enzyme acting on the  $\alpha$ 1,4-Araf linkage connected to the terminal galactose. Therefore, we tried to identify an  $\alpha$ -L-arabinofuranosidase specific for  $\alpha$ 1,4-Araf linkage. In our previous report,  $\alpha$ -D-Galp-(1→3)-L-Ara-free gum arabic AGP was fermented by *B. longum* subsp. *longum* JCM1217 and JCM7052 (9). The common  $\alpha$ -L-arabinofuranosidase genes in the GH43 gene cluster found in the two strains were *BlArafA*, *BlArafB*, and *BlArafE* (Fig. 2A).



**FIG 2** Structural features of the  $\alpha$ -L-arabinofuranosidase candidates. (A) Gene clusters in *B. longum* subsp. *longum* JCM1217 and JCM7052 involved in the degradation of type II AG. The arrowheads (with names below) indicate genes annotated in the Kyoto Encyclopedia of Genes and Genomes database. The glycoside hydrolases (GHs) in the Carbohydrate-Active enZymes (CAZy) database are shown inside the red arrowheads. Light gray bars indicate orthologous regions (>95% identity). (B) Domain structures of BIArafE, BIArafB, and BIArafA. Domain structures were predicted using SignalP5.0 (<http://www.cbs.dtu.dk/services/SignalP/>) and InterPro (<https://www.ebi.ac.uk/interpro/>) servers. The domains are indicated as follows: SP, signal peptide; LamG, laminin G; Ig, bacterial Ig-like domain; TM, transmembrane region. (C) Phylogenetic tree of the GH43\_34 and GH43\_22 domains of BIArafE. The phylogenetic tree was constructed using the neighbor-joining method and the aligned sequences. For constructing the tree, the program MUSCLE was implemented in MEGA7 software. GH43\_34 and GH43\_22 domains of BIArafE are enclosed in the dashed-line box. The characterized enzymatic activities or protein names are shown alongside the abbreviated names of the organisms as follows: Ba. Cell\_Abf, *B. cellulosilyticus*  $\alpha$ -L-arabinofuranosidase (GenPept accession no. ALJ58905.1); Ba. The\_Abf (BT3675), *B. thetaiotaomicron*  $\alpha$ -L-arabinofuranosidase encoded by BT3675 (GenPept accession no. AAO78780.1); R. Jos\_Abf, *Ruminiclostridium josui* exo- $\alpha$ 1,5-arabinofuranosidase (GenPept accession no. BBA94052.1); Ba. Ova\_Bgf, *B. ovatus*  $\beta$ -D-galactofuranosidase (GenPept accession no. ALJ48250.1); P. sp\_Bgf, *Paenibacillus* sp. strain  $\beta$ -D-galactofuranosidase (ACS99115.1); Ba. The\_Abf (BT3662), *B. thetaiotaomicron*  $\alpha$ -L-arabinofuranosidase encoded by BT3662 (GenPept accession no. AAO78780.1); B. Ado\_Bxyl, *B. adolescentis*  $\beta$ -xylosidase (GenPept accession no. BAF40308.1); B. Lon\_BIArafA, *B. longum* subsp. *longum* JCM1217 BIArafA (GenPept accession no. BAJ67518.1); B. Lon\_BIArafB, *B. longum* subsp. *longum* JCM1217 BIArafB (GenPept accession no. BAJ67517.1); and R. Jos\_Abn, *R. josui* arabinanase (GenPept accession no. BBA94052.1).

**TABLE 1** Primers used for plasmid construction and site-directed mutagenesis<sup>a</sup>

Primer name	Sequence of oligonucleotide primers
BLLJ_1850_for	5'-CCCAAGCTTGATACCACCGATTTCATCGGCCGC-3'
BLLJ_1850_rev	5'-GTGCCGCTCGAGGGAGATGACGGCACCCGGCTTCTTG-3'
BLLJ_1850_D571A_for	5'-ACCACGATGATTAAGGCCGA-3'
BLLJ_1850_D571A_rev	5'- <u>TGCGATAACGGACTTGAC</u> -3'
BLLJ_1850_D1197A_for	5'-CCGTCCATCTTCACCGACC-3'
BLLJ_1850_D1197A_rev	5'- <u>TGCGATGGCCTGTCCAAC</u> -3'

<sup>a</sup>Positions of the mutated sequences are underlined.

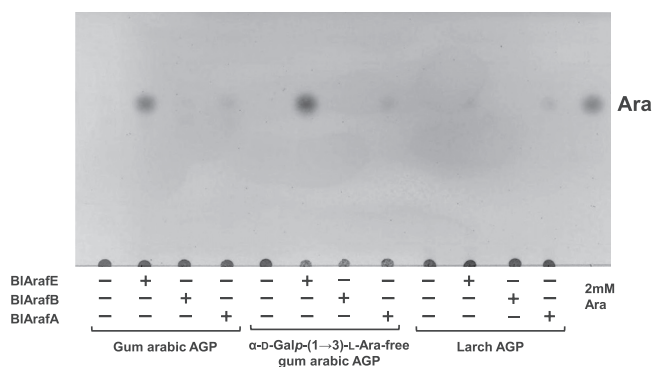
Among them, we focused on the  $\alpha$ -L-arabinofuranosidase candidate BIArafE, containing two catalytic domains, GH43\_22 and GH43 subfamily 34 (GH43\_34).

This study characterized each GH43\_22 and GH43\_34 domain of BIArafE that acted to  $\alpha$ 1,3/ $\alpha$ 1,4-Araf continuously and determined the structure of the remaining oligosaccharide in gum arabic AGP fermentation of *B. longum* subsp. *longum* JCM7052. We also evaluated the ability to assimilate the limit product from *B. longum* in some *Bacteroides* species.

## RESULTS

**Structural features of  $\alpha$ -L-arabinofuranosidase BIArafE.** *B. longum* subsp. *longum* JCM7052 possessed *BIArafE*, *BIArafB*, and *BIArafA* homologous genes with high sequence identity (>95%) but did not possess *BIArafC* and *BIArafD* homologous genes (Fig. 2A). *BIArafE*, *BIArafB*, and *BIArafA* produced cell surface-anchoring enzymes with a signal peptide at the N terminus and a transmembrane region at the C terminus and conserved GH43\_22, laminin G, and bacterial Ig-like domains (Fig. 2B). BIArafE was a multidomain enzyme comprising two catalytic domains of  $\alpha$ -L-arabinofuranosidase candidates, GH43\_22 and GH43\_34. The GH43\_22 and GH43\_34 domains of BIArafE were compared using a molecular phylogenetic tree with the previously characterized GH43 enzymes (Fig. 2C). The GH43\_34 domain of BIArafE had the highest homology (57% identity) with exo- $\alpha$ 1,5-arabinofuranosidase from *Ruminiclostridium josui* (24), and the GH43\_22 domain had the highest homology (51% identity) with  $\beta$ -xylosidase from *Bifidobacterium adolescentis* (25). The GH43\_22 domain of BIArafE showed 35% and 33% identity with BIArafA and BIArafB, respectively.

**Characterization of BIArafE.** Recombinant BIArafE protein was designed without the N-terminal signal peptide (amino acids [aa] 1 to 26) and the C-terminal transmembrane region (aa 1637 to 1659) using the primers shown in Table 1. The recombinant molecule was highly expressed as a soluble protein at 25°C and was purified using a C-terminal His-tag. The purified recombinant BIArafE was found to migrate as a band with an apparent molecular mass of 175 kDa on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), corresponding to its calculated molecular mass (see Fig. S1 in the supplemental material). BIArafA and BIArafB were also expressed and purified according to the method described in previous studies (11, 12). Gum arabic AGP, larch AGP, and  $\alpha$ -D-Galp-(1→3)-L-Ara-free gum arabic AGP were treated with the recombinant enzymes, and the released sugars were analyzed by thin-layer chromatography (TLC) (Fig. 3).  $\alpha$ -D-Galp-(1→3)-L-Ara-free gum arabic AGP was prepared from gum arabic AGP by GAfase treatment, and the structural model is presented in Fig. 1. As a result, BIArafE and BIArafA released L-arabinose from all the tested polysaccharides, but BIArafB showed no activity. Notably, BIArafE released the largest amount of L-arabinose from gum arabic AGP and  $\alpha$ -D-Galp-(1→3)-L-Ara-free gum arabic AGP among the three enzymes. In the current study, several oligosaccharides constituting the side chains of gum arabic AGP were prepared (Fig. 1 and Fig. S2) to evaluate the substrate specificities of these enzymes in more detail, and each enzyme was incubated with oligosaccharides S5-GA, S5-A, S5, S4-A, S3-GA, S3-A, and S3 (Fig. 1). The released sugars were analyzed using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Fig. 4A and Fig. S3). As a result, we found that only BIArafE showed  $\alpha$ 1,4-Araf cleavage activity on the oligosaccharide

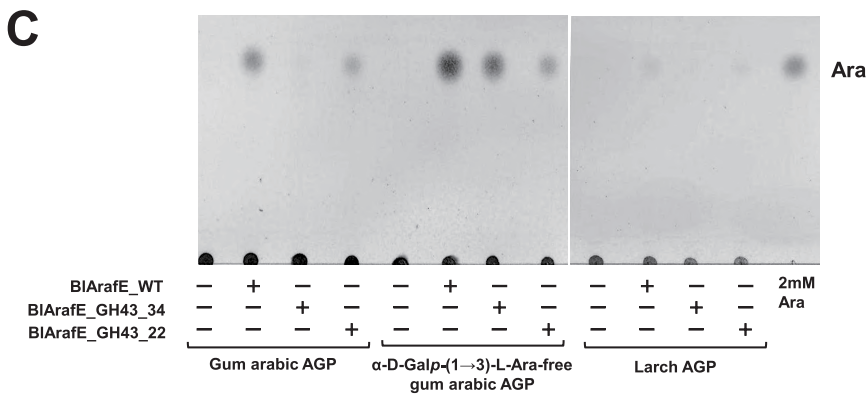
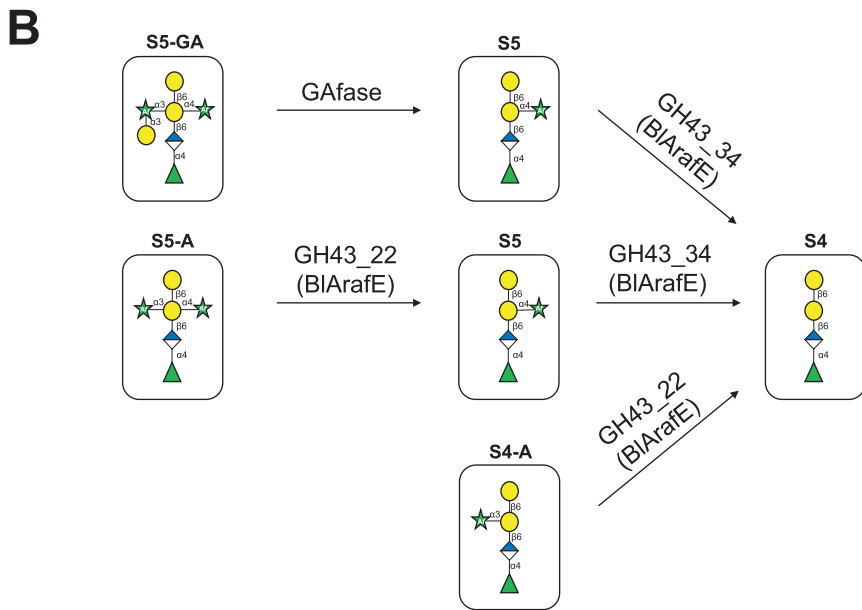
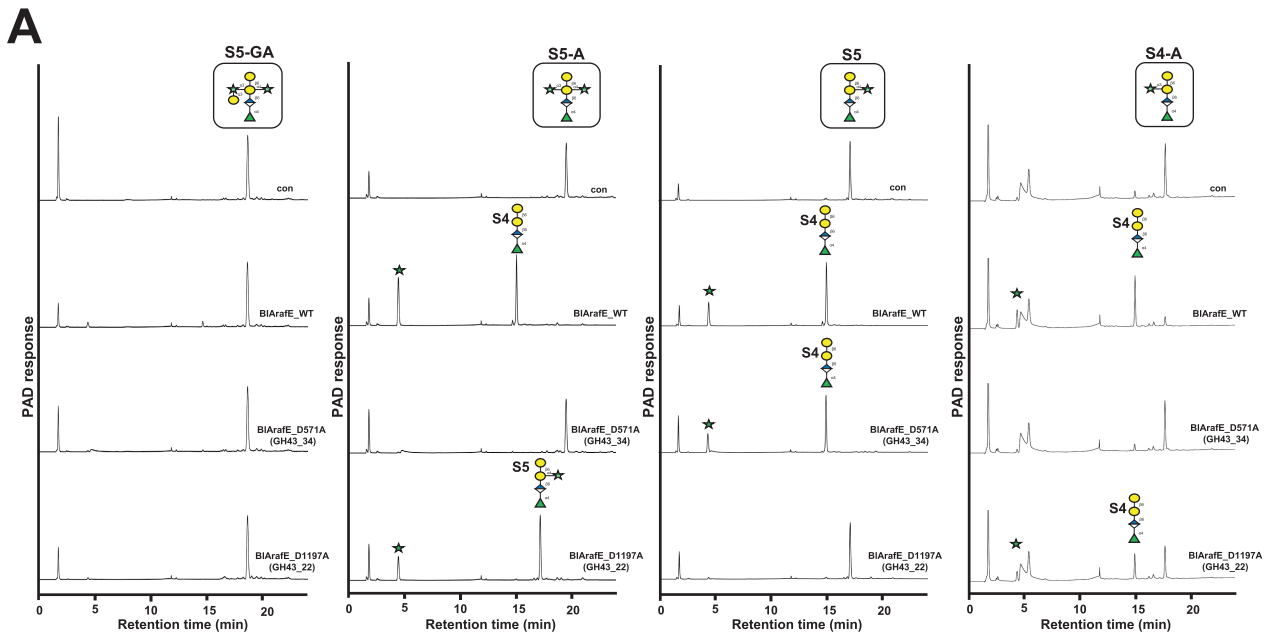


**FIG 3** TLC analysis of the reaction products of  $\alpha$ -L-arabinofuranosidases with AGPs. Gum arabic AGP,  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP, and larch AGP were incubated with either BIArAfE, BIArAfB, or BIArAfA at 37°C for 16 h.

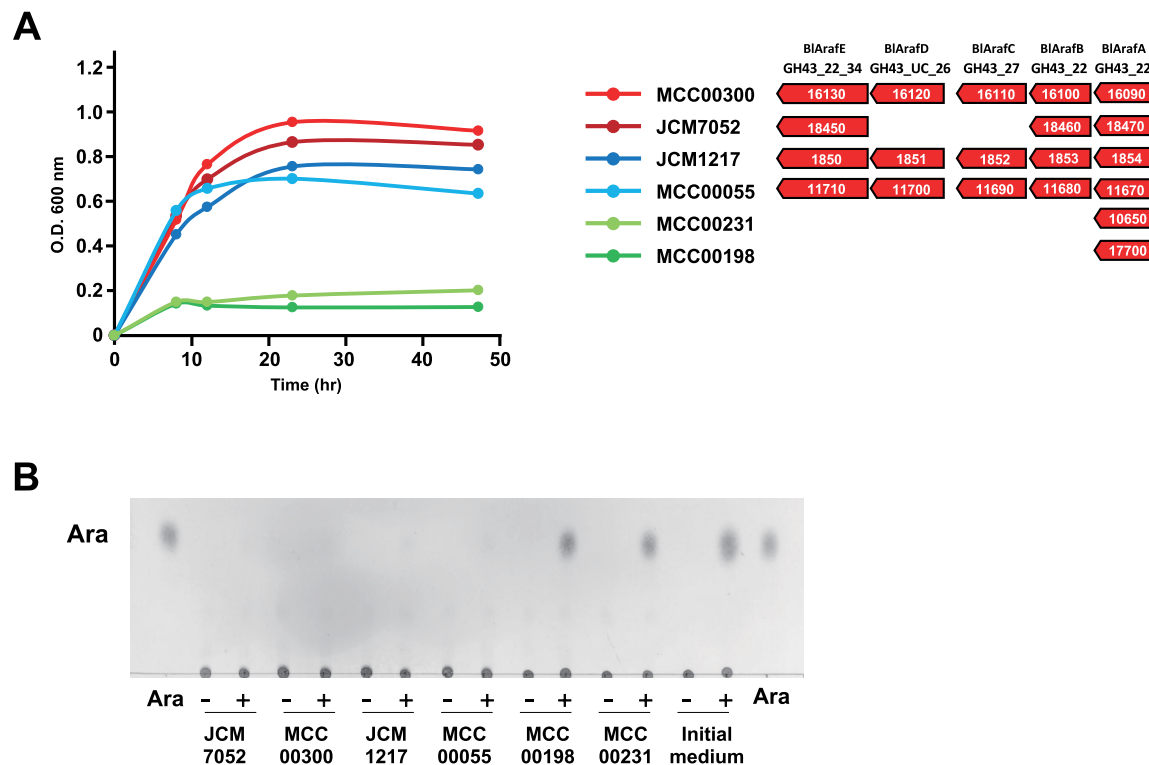
S5, but not BIArAfB and BIArAfA. Furthermore, BIArAfE cleaved S5-A into S4 and L-arabinose, indicating that BIArAfE can act on both  $\alpha$ 1,4-Araf and  $\alpha$ 1,3-Araf. Interestingly, when  $\alpha$ 1,3-Araf in S5-A was capped with  $\alpha$ 1,3-Gal, such as S5-GA, BIArAfE could not act on  $\alpha$ 1,4-Araf, indicating that eliminating the  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara or  $\alpha$ -D-Galp is essential for BIArAfE activity. Conversely, BIArAfA showed activity against  $\alpha$ 1,3-Araf on S4-A but did not act on S5-A, in which  $\beta$ -D-Gal was double substituted with  $\alpha$ -L-Araf at O3 and O4. BIArAfB could not hydrolyze any substrates with  $\alpha$ -L-Araf at O3 and O4. Similarly, we evaluated the enzyme reactivities for S3-GA, S3-A, and S3, in which  $\beta$ -D-Gal was not substituted by  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\beta$ -D-GlcpA at O6 (Fig. S3). As a result, we found that BIArAfE acted on S3-A and S3, suggesting that BIArAfE exhibited activity regardless of whether  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\beta$ -D-GlcpA was present or not.

**Functional analysis of each of the catalytic domains in BIArAfE.** BIArAfE conserved three critical putative acidic amino acid residues for catalytic reaction, “general acid,” “general base,” and “pKa modulator of general acid,” based on the previously characterized GH43 enzymes (26). Herein, the putative pKa modulator of general acid residues in BIArAfE GH43\_22 (Asp571) and BIArAfE GH43\_34 (Asp1197) were mutated to Ala to inactivate the respective domains. Similar to wild-type BIArAfE, the mutants were expressed, and the purified recombinant BIArAfE mutants were found to migrate as a single band with an apparent molecular mass of 175 kDa on SDS-PAGE, corresponding to their predicted molecular masses (data not shown). The oligosaccharide substrates (S5-GA, S5-A, S5, and S4-A) were used to evaluate the substrate specificity of GH43\_22 and GH 43\_34 domains. The BIArAfE\_D1197A mutant (GH43\_22 active form) exhibited enzymatic activity toward  $\alpha$ 1,3-Araf in  $\alpha$ 1,3- and  $\alpha$ 1,4-Araf double-substituted structures in S5-A and single  $\alpha$ 1,3-substituted structures in S4-A, but could not act on  $\alpha$ 1,4-Araf in S5 (Fig. 4A). S4-A was incompletely degraded by GH43\_22 under the described conditions, and the active form of GH43\_22 likely preferred double-substituted S5-A over single-substituted S4-A. Conversely, the BIArAfE\_D571A mutant (GH43\_34 active form) acted on  $\alpha$ 1,4-Araf in S5 and degraded it to S4, but it did not act on  $\alpha$ 1,3-Araf in S4-A, which had a single substitution of  $\alpha$ -L-Araf at the O3 of the terminal Gal. The GH43\_34 active form also showed no activity when  $\beta$ -D-Gal of the side chain was double substituted with  $\alpha$ -L-Araf at O3 and O4 in S5-A. S5-GA could not be hydrolyzed by either wild-type BIArAfE or both mutants. In summary, the degradation of S5-A to S4 by BIArAfE was first mediated via the digestion of  $\alpha$ 1,3-Araf by GH43\_22, followed by the digestion of  $\alpha$ 1,4-Araf by GH43\_34 (Fig. 4B).

Gum arabic AGP,  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP, and larch AGP were treated with GH43\_22 and GH43\_34 active forms, providing insights into the structural model of gum arabic and larch AGPs. The GH43\_22 active form released L-arabinose from gum arabic AGP (Fig. 4C), indicating that there are some  $\alpha$ 1,3-Araf residues without  $\alpha$ -D-Gal capping in gum arabic AGP (Fig. S4). Furthermore, the GH43\_34 active



**FIG 4** Substrate specificities of GH43\_22 and GH 43\_34 domains of BIARafE. (A) HPAEC-PAD analysis of the reaction products with gum arabic AGP-related oligosaccharides. S5-GA, S5-A, S5, and S4-A were incubated with either wild-type BIARafE (WT), D571A mutant BIARafE (GH 43\_34 (Continued on next page)



**FIG 5** *In vitro* assimilation test performed using  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP. (A) Growth profile of *B. longum* strains cultured in a medium containing  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP ( $n = 2$ , cell culture biological replicates). The gene clusters of GH43  $\alpha$ -L-arabinofuranosidases in *B. longum* strains are shown in the figure. (B) Residual L-arabinose in  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP after 48 h of culturing *B. longum*. TLC analysis of the reaction products by BIAraFE with the ethanol precipitates of the initial medium or cultured media of JCM7052, MCC00300, JCM1217, MCC00055, MCC00198, and MCC00231 at 37°C for 16 h. Reaction products (+) and control samples without BIAraFE (-) were analyzed by TLC.

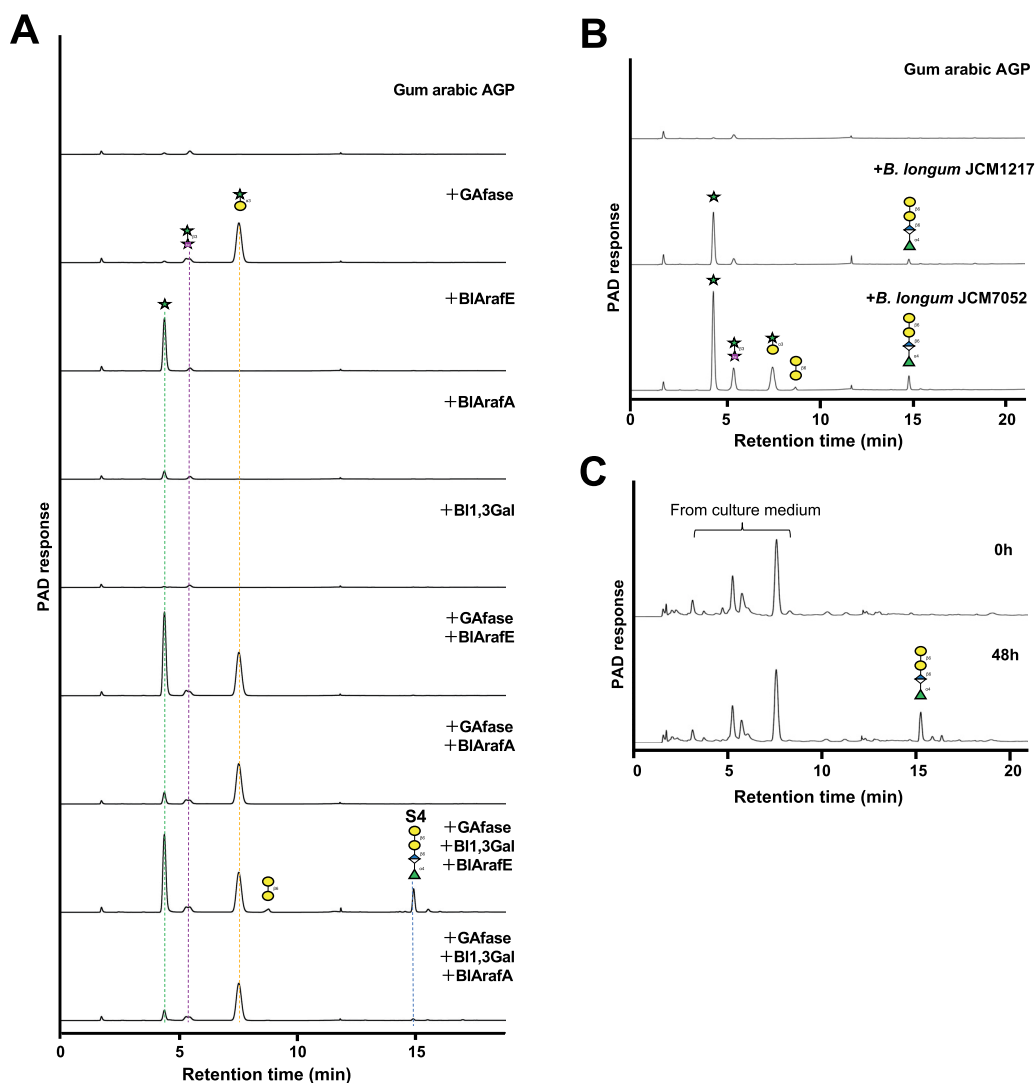
form released many L-arabinose residues from  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP but only a few from intact gum arabic AGP. These results indicated that S5-GA is one of the major components of the side chains of gum arabic AGP, rather than S5 or S3 structures with the single substitution of  $\alpha$ -L-Araf at O4, which was consistent with a previous study (7). Moreover, the GH43\_34 active form could not release L-arabinose from larch AGP, indicating that there is no single-substituted  $\alpha$ 1,4-Araf in larch AGP.

***In vitro* assimilation test of gum arabic AGP using BIAraFE carrier strains.** The gene cluster of GH43  $\alpha$ -L-arabinofuranosidases (BLLJ\_1850 to BLLJ\_1854 homologs) has various conservation patterns in different strains of *B. longum* (12). The BIAraFE homologous gene (identity  $\geq 50\%$ , coverage  $\geq 50\%$ ) was conserved in 54.5% of *B. longum* subsp. *longum* MCC strains (67/123 strains) in the National Center for Biotechnology Information (NCBI) database, which was higher than GAfase (4.88%, 6/123 strains) (9). We performed an *in vitro* assimilation test using BIAraFE carrier strains of *B. longum* and noncarrier strains in a medium containing  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP as the sole carbon source (Fig. 5A). BIAraFE carrier strains, *B. longum* subsp. *longum* JCM7052, MCC00300, JCM1217, and MCC00055, grew in the medium, whereas noncarrier strains, MCC00198 and MCC00231, exhibited no growth. Further, intact gum arabic AGP was only assimilated in the GAfase-carrier strains JCM7052 and MCC00300 (data not shown), which was in agreement with our previous study (9). Subsequently, BIAraFE was

**FIG 4** Legend (Continued)

active form), or D1197A mutant BIAraFE (GH 43\_22 active form) at 37°C for 16 h. (B) Schematic model of the mode of action of GH43\_22 and GH 43\_34 domains of BIAraFE. (C) TLC analysis of the reaction products with AGPs. Gum arabic AGP,  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP, and larch AGP were incubated with either wild-type BIAraFE, GH 43\_34 active form BIAraFE, or GH 43\_22 active form BIAraFE at 37°C for 16 h.





**FIG 6** Cooperative degradation of gum arabic AGP by the cell surface-anchoring enzymes in *B. longum*. (A) HPAEC-PAD analysis of the combination reactions of type II AG degradative enzymes (Gafase, BI1,3Gal, BIArAfA, and BIArAfE) with gum arabic AGP. Gum arabic AGP was incubated at 37°C for 24 h. Data from one representative experiment out of three independent experiments is shown. (B) HPAEC-PAD analysis of the reaction products by the bacterial cell fraction of *B. longum* subsp. *longum* JCM1217 or JCM7052 cells with gum arabic AGP. (C) HPAEC-PAD analysis of the initial medium (0 h) and cultured medium (48 h) of *B. longum* subsp. *longum* JCM7052 culture) with 1% gum arabic AGP as the sole carbon source.

incubated with the ethanol precipitate of the medium in order to examine whether L-arabinose remained to be attached to  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP after culture. It was found that BIArAfE released L-arabinose from the residual polysaccharides in noncarrier strains but not carrier strains (Fig. 5B). In general, *B. longum* strains are known to utilize L-arabinose as a carbohydrate source (27, 28). Therefore, these results suggested that BIArAfE carrier strains utilized L-arabinose released from  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP during their growth. Gafase noncarrier strains could not grow well on intact gum arabic AGP, even though these strains carried BIArAfE. Hence, it is necessary to remove the steric hindrance of BIArAfE by the action of Gafase to facilitate good growth.

**Cooperative degradation of gum arabic AGP.** Gum arabic AGP was incubated with different combinations of Gafase, BI1,3Gal, and  $\alpha$ -L-arabinofuranosidases (BIArAfE and BIArAfA) to assess its cooperative degradation (Fig. 6A). The amount of L-arabinose

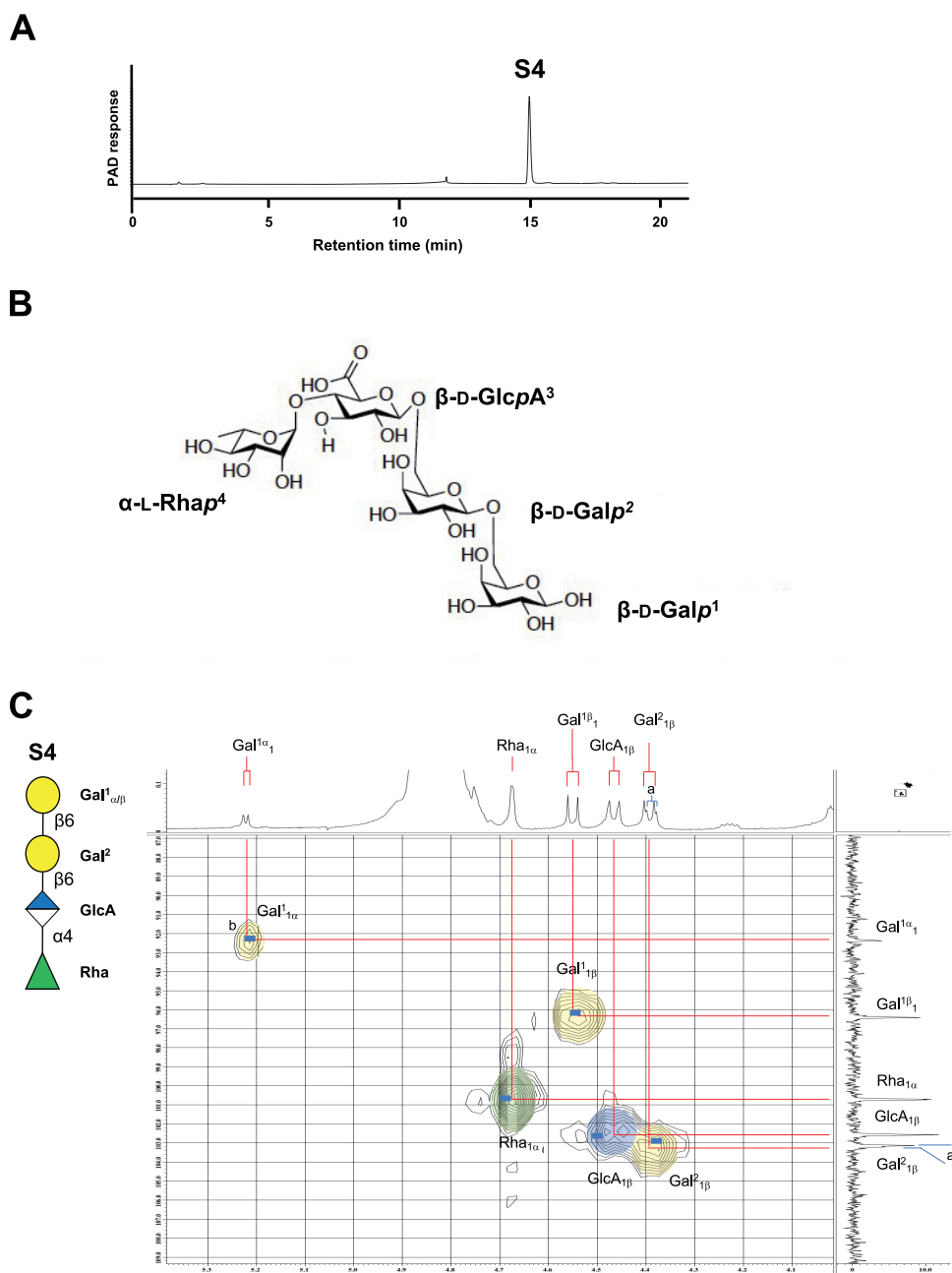
released from gum arabic AGP incubated with the combination of GAfase and BIArafE was 1.64-fold greater than that released from gum arabic AGP incubated with BIArafE alone, indicating that GAfase could remove the steric hindrance and allowed BIArafE to act on  $\alpha$ 1,4-Araf linkage. Furthermore, the peaks of L-arabinose,  $\beta$ -D-Gal-(1 $\rightarrow$ 6)-D-Gal, and S4 were detected by using the combination of GAfase, B11,3Gal, and BIArafE. The amounts of L-arabinose and S4 released by GAfase, B11,3Gal, and BIArafE were 10.5-fold and 13.6-fold higher, respectively, than those released by the combination of GAfase, B11,3Gal, and BIArafA instead of BIArafE. In addition, the peaks of the reaction product obtained by the combination of GAfase, B11,3Gal, and BIArafE were consistent with the reaction product obtained using *B. longum* subsp. *longum* JCM7052 cells as a crude enzyme (Fig. 6B). These results indicated that gum arabic AGP was degraded mainly by the action of three cell surface degradative enzymes (GAfase, B11,3Gal, and BIArafE) present in *B. longum* subsp. *longum* JCM7052. Furthermore, S4 remained in *B. longum* subsp. *longum* JCM7052 culture, but monosaccharides and disaccharides were not detected after 48 h. These results indicated that *B. longum* utilized monosaccharides and disaccharides, but not S4 (Fig. 6C).

**Structural determination of the limit product S4 and utilization by other commensal bacteria.** S4 was purified as a single peak from the limit product of the gum arabic AGP incubated with *B. longum* subsp. *longum* JCM7052 bacterial cells (Fig. 7A). Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) of S4 revealed a high-intensity peak at 687.138 *m/z*, consistent with a sodium adduct of  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\beta$ -D-GlcpA-(1 $\rightarrow$ 6)- $\beta$ -D-Galp-(1 $\rightarrow$ 6)-D-Gal (calculated for  $C_{24}H_{40}O_{21}Na_1$   $[M+Na]^+$  687.196) (Fig. S5A). High-resolution electrospray ionization–time of flight mass spectrometry (ESI-TOF MS) of S4 revealed that the tetrasaccharide structure matched the elemental composition (ESI-TOF high-resolution MS [ESI-TOF HRMS] calculated for  $C_{24}H_{40}O_{21}Na_1$   $[M+Na]^+$  687.1960, found 687.1958). We also identified the chemical structure of S4 (Fig. 7B) using nuclear magnetic resonance (NMR) analysis, including  $^1H$ ,  $^{13}C$ , and heteronuclear multiple quantum correlation (HMQC) spectra (Fig. 7C and Table 2).  $^1H$  and  $^{13}C$  chemical shifts and coupling constants of S4 were  $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\beta$ -D-GlcpA-(1 $\rightarrow$ 6)- $\beta$ -D-Galp-(1 $\rightarrow$ 6)-D-Gal (Table 2), which matched with that determined by the previous reports (7). HMQC spectra showed that the overlap of anomeric regions was consistent with a previous report (7). In addition, the combination of NMR techniques, such as  $^1H$ - $^1H$  correlated spectroscopy (COSY), HMQC, and heteronuclear multiple-bond correlation (HMBC), and nondecoupled HMQC spectra of S4 further increased the validity of the chemical structure of S4 as the peaks assigned in Table 2 (Fig. S5B).

It was reported that some *Bacteroides* species encode degradative enzymes for tetrasaccharide S4 in gum arabic AGP (7). Therefore, we performed *in vitro* assimilation tests using some *Bacteroides* species in a medium containing S4 as the sole carbon source. Six *Bacteroides* species, including *B. thetaiotaomicron*, *B. caccae*, *B. cellulosilyticus*, *B. ovatus*, *B. vulgatus*, and *B. uniformis*, were tested. As a result, *B. vulgatus* was seen to exhibit good growth and S4 utilization (Fig. 8). Although *B. vulgatus* was reported to be a nonutilizer of intact gum arabic AGP (7), it utilized the limit product generated by *B. longum* from gum arabic AGP.

## DISCUSSION

In this study, we revealed that the multidomain enzyme BIArafE encoded in *B. longum* subsp. *longum* JCM1217 is an  $\alpha$ -L-arabinofuranosidase that hydrolyzes  $\alpha$ 1,3/1,4-Araf linkages present on gum arabic AGP side chains. Furthermore, the GH43\_22 domain in BIArafE hydrolyzed  $\alpha$ 1,3-Araf linkages on  $\alpha$ 1,3- and  $\alpha$ 1,4-Araf double-substituted structure and on  $\alpha$ 1,3-Araf single-substituted structure. Conversely, the GH43\_34 domain in BIArafE hydrolyzed  $\alpha$ 1,4-Araf linkages on  $\alpha$ 1,4-Araf single-substituted structure. For degrading S5-GA, one of the main structural components of gum arabic AGP, the removal of terminal  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara or Gal on S5-GA is necessary for BIArafE activity. Although GAfase is a crucial enzyme for the removal of  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara, it was only conserved in 6.84% of the *B. longum* strains according to the NCBI database (9). As



**FIG 7** Structural analysis of S4. (A) HPAEC-PAD analysis of the purified S4. (B) Chemical structure of S4. (C) Expanded HMQC spectra around anomeric region measured at 400 MHz in  $D_2O$  at 21°C. <sup>a</sup>Difference between the anomers of Gal<sup>1</sup> ( $\alpha/\beta$ , approximately 1:2) slightly affected the conformation of Gal<sup>2</sup> that was seen in Gal<sup>2</sup><sub>1 $\beta$</sub> . <sup>b</sup>Assignment of the tetrasaccharide was compared to the HMQC data referenced at Gal<sup>1</sup><sub>1 $\alpha$</sub>  indicated by blue bars, translated from the data reported by Cartmell et al. (7).

54.5% of the *B. longum* strains conserve *BlArafE*, it is likely that there would be an unidentified secretory  $\alpha$ -D-galactosidase present in the microbiome.

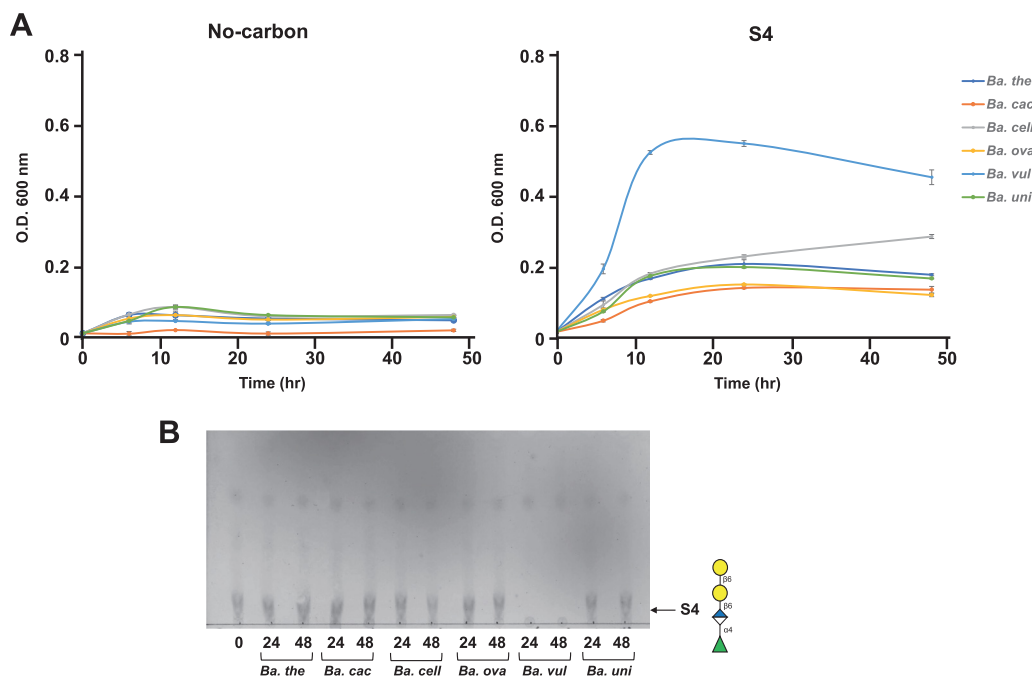
Among GH43 enzymes, there are some proteins containing multiple GH43 catalytic domains belonging to different subfamilies. The previous study based on the CAZy database (29) showed that 82 GH43 proteins possessing the GH43\_22 (first domain from N terminus) and GH 43\_34 (second domain from N terminus) domains were found, which is the most common combination pattern among multiple GH43

**TABLE 2** Assignment of signal in  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of S4 in  $\text{D}_2\text{O}$  at 400 MHz at room temperature<sup>a</sup>

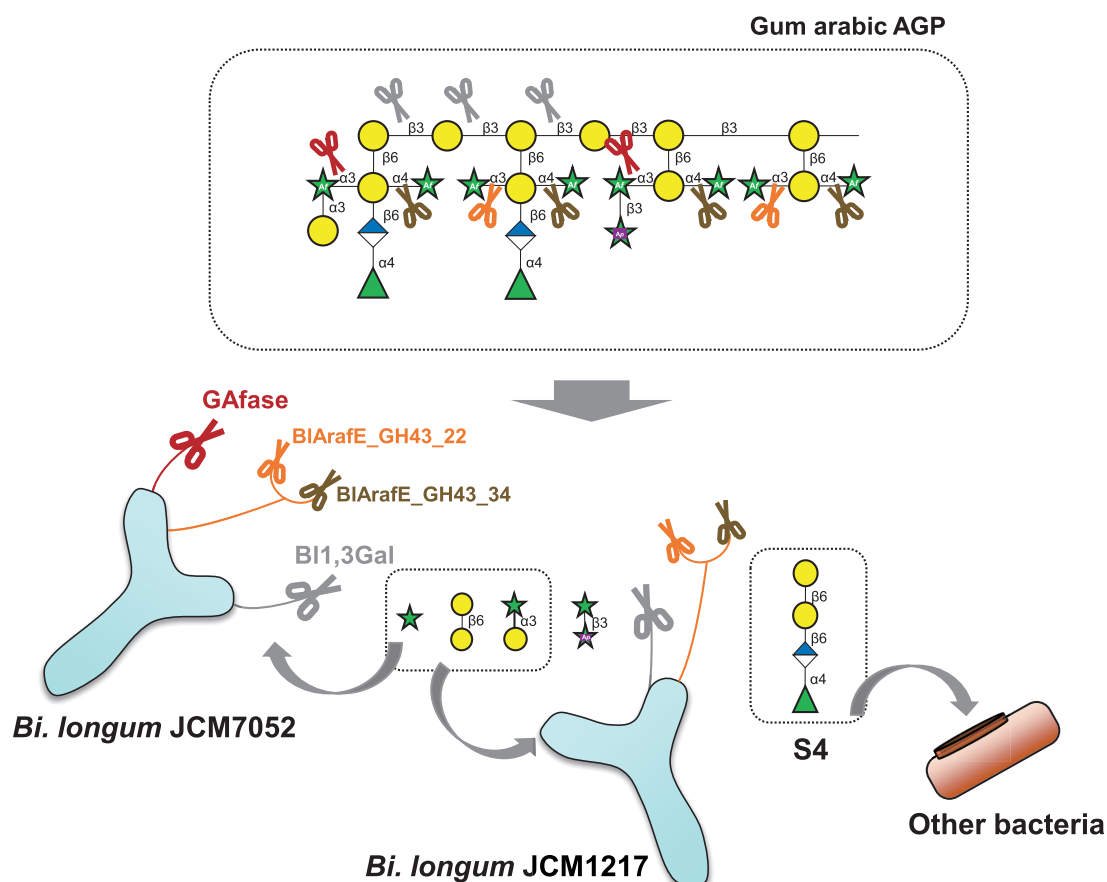
Unit	Value for:											
	H1 ( $^3J_{\text{H-H}}$ )	H2	H3	H4	H5	H6	C1 ( $^1J_{\text{C-H}}$ )	C2	C3	C4	C5	C6
$\beta$ -D-Galp <sup>1</sup>	4.54 (d 8.4 Hz)	3.44	3.62	3.92	3.85	3.85, 4.02	96.4 (164.6 Hz)	71.8	72.6	68.8	73.75	69.3
$\alpha$ -D-Galp <sup>1</sup>	5.22 (d 4.0 Hz)	3.75	3.85	3.98	4.23	3.77, 3.97	92.4 (172.3 Hz)	68.6	68.3	69.3	68.9	69.9
$\beta$ -D-Galp <sup>2</sup>	4.39 (d 8.4 Hz)	3.48	3.62	3.92	3.85	3.85, 4.02	103.1 (169.7 Hz)	70.7	72.5	68.9	73.83	69.3
$\beta$ -D-GlcpA <sup>3</sup>	4.48 (d 8.0 Hz)	3.30	3.53	3.54	3.70		102.6 (171.0 Hz)	73.2	74.2	79.0	76.1	175.3
$\alpha$ -L-Rhap <sup>4</sup>	4.68 (s)	3.88	3.71	3.37	3.97	1.19	100.7 (173.5 Hz)	70.3	70.0	71.9	68.6	16.5

<sup>a</sup> $^1\text{H}$ , 400 MHz;  $^{13}\text{C}$ , 100 MHz. d, doublet.

proteins. In particular, 143 GH43 proteins among 228 multidomain enzymes possess the GH43\_34 domain at the C terminus, suggesting a functional role of this domain (29). Nevertheless, only four proteins have been characterized among the GH43\_34-containing proteins,  $\beta$ -D-galactofuranosidase (Bovatus\_03644) from *B. ovatus* ATCC 8483 (30) and *Paenibacillus* sp. JDR-2 (Pjdr2\_0435) (30) and two  $\alpha$ -L-arabinofuranosidases (BT\_3675 and BT\_3662) from *B. thetaiotaomicron* VPI-5482, based on the CAZY database. BT\_3675 acts on the  $\alpha$ 1,3-Araf linkage present in single  $\alpha$ 1,3-Araf-substituted gum arabic AGP (7), and BT\_3662 acts on  $\alpha$ 1,3-Araf linkage present in grape RG-II (22). However, in this study, the GH43\_34 domain in BIArafE exhibited activity toward  $\alpha$ 1,4-Araf linkage but not toward  $\alpha$ 1,3-Araf in gum arabic AGP. Among GH43\_22-containing proteins, four proteins have been characterized,  $\beta$ -xylosidase (BAD\_1527) from *B. adolescentis* ATCC 15703 (25), two  $\alpha$ -L-arabinofuranosidases (BIArafB and BIArafA) from *B. longum* subsp. *longum* JCM1217, and arabinanase 43A (RjAbn43A) from *R. josui* JCM17888. Based on a previous study, BIArafB acts on the  $\alpha$ 1,5-Araf linkage of arabinan, and BIArafA acts on the  $\alpha$ 1,3-Araf linkage of AGPs (12). Recently, it was found that GH43\_22 in BIArafE acts on  $\alpha$ 1,3-Araf single-substituted arabinoxylan (M. Komeno



**FIG 8** *In vitro* assimilation test of S4 using *Bacteroides* species. (A) Growth of *Bacteroides* species on MM media containing S4 as a sole carbon source (right) and on no-carbon MM medium (left). The absorbance of growth media of *Bacteroides thetaiotaomicron* (*Ba. the*), *Bacteroides caccae* (*Ba. cac*), *Bacteroides cellulosilyticus* (*Ba. cell*), *Bacteroides ovatus* (*Ba. ova*), *Bacteroides vulgatus* (*Ba. vul*), and *Bacteroides uniformis* (*Ba. uni*) was monitored at 0, 6, 12, 24, 48 h ( $n = 3$ , cell culture biological replicates). Error bars indicate SD ( $n = 3$ ). (B) TLC analysis of the residual S4 after *Bacteroides* species culture at 24 h (24) and 48 h (48) and initial medium without any strains (0). The arrow indicates the position of S4. Data are representative of three independent experiments; one representative experiment is shown.



**FIG 9** Schematic model of the cooperative degradation of gum arabic AGP by cell surface enzymes in bifidobacteria.

et al., unpublished data), and RjAbn43A also acts on  $\alpha$ 1,3-Araf linkage on branched arabinan (24). In the present study, we found that GH43\_22 and GH43\_34 catalytic domains of BIARAfE performed a sequential  $\alpha$ 1,3- and  $\alpha$ 1,4-dearabinosylation of the double-substituted side chains of arabinosyl present in gum arabic AGP. Further, the GH43\_34 domain of BIARAfE was the first example of an  $\alpha$ 1,4-Araf-specific  $\alpha$ -L-arabinofuranosidase. Similar to BIARAfE, some multicatalytic enzymes possessing multiple catalytic domains on the same polypeptide have been characterized so far. For example, CelA from *Caldicellulosiruptor bescii* (31) and ChiA from *Flavobacterium johnsoniae* (32) possess exo- and endo-acting enzyme catalytic domains and exhibit intramolecular synergy for degrading recalcitrant crystalline substrates. These enzymes show higher activity when they form multicatalytic enzymes than when their equivalent single catalytic domains are added. Regarding soluble substrates, the multicatalytic enzyme Abf43A-Abf43B-Abf43C (GH43\_22\_26\_34) from *R. josui* was shown to possess GH43\_22 (Abf43A) and GH43\_26 (Abf43B), which cooperatively degraded sugar beet arabinan (24). However, there are no studies reporting catalytic improvement due to the coexistence of multicatalytic domains in a single polypeptide. In the case of BIARAfE, the coexistence of GH43\_22 and GH43\_34 catalytic domains in a single polypeptide appears to increase the encounter rate with substrates and improve the enzymatic activity considering that GH43\_34 cannot act unless trimming is performed by GH43\_22 or GAfase. Nevertheless, further investigation comparing the activity between the full-length enzyme and the truncated single catalytic domain variants is warranted to assess the advantages of multicatalytic enzymes for soluble substrates.

When it became clear that BIARAfE was an  $\alpha$ 1,3/1,4-Araf-specific  $\alpha$ -L-arabinofuranosidase, it was revealed that *B. longum* cooperatively uses three bacterial surface enzymes

to degrade gum arabic AGP (Fig. 9). First,  $\alpha$ -D-Galp-(1→3)-L-Ara was removed from the end of the side chains by GAfase (BLGA\_00340), thereby facilitating the action of BIArafE (BLLJ\_1850, BLGA\_18450), which is widely conserved in *B. longum* strains. Furthermore, the trimming of modified sugars in the side chains facilitated the action of BI1,3Gal (BLLJ\_1840, BLGA\_18370), finally leading to the accumulation of the tetrasaccharide  $\alpha$ -L-Rhap-(1→4)- $\beta$ -D-GlcpA-(1→6)- $\beta$ -D-Galp-(1→6)-D-Gal, which could not be degraded by *B. longum*. The reaction products produced by recombinant enzymes were consistent with those produced using *B. longum* subsp. *longum* JCM7052 bacterial cell fraction, suggesting that these enzymes were used for the assimilation of gum arabic AGP by *B. longum*. However, the  $\beta$ 1,3-galactan main chain of gum arabic AGP was not completely degraded by *B. longum* enzymes, indicating that *B. longum* favors the metabolism of the modified sugar of the side chain such as Ara and  $\alpha$ -D-Galp-(1→3)-L-Ara. This assimilative strategy differs from that of *Bacteroides*, which favors the metabolism of high-degree-of-polymerization (DP) oligosaccharides by cell surface endo- $\beta$ 1,3-galactanase and many accessory enzymes localized in the periplasm (7). However, unlike other colonic *Bacteroides*, *B. vulgatus* is not a versatile utilizer of polysaccharides, and it seems to favor the metabolism of oligosaccharides generated by other bacteria from polysaccharides as per a previous study in which various polysaccharides, including inulin, levan, amylopectin, and gum arabic AGP, were used (7, 33, 34). In this study, *B. vulgatus* exhibited good growth in the limit product S4 generated by *B. longum* subsp. *longum* JCM7052 from gum arabic AGP. pH-controlled batch culture studies and administration tests of gum arabic AGP on human volunteers had shown that the intake of gum arabic AGP increases *Bifidobacterium* and *Bacteroides* (3, 35) and *Bacteroidetes* phyla (36) in the human intestine or the simulation model. Although they could grow on gum arabic AGP using the assimilative strategy described above, cross-feeding may occur between bifidobacteria and *Bacteroides* species. Further studies using cocultures or pH-controlled batch cultures in the presence of key enzymes, such as GAfase, BIArafE, BI1,3Gal, and endo- $\beta$ 1,3-galactanase, are warranted to investigate the symbiosis between bifidobacteria and *Bacteroides* in terms of the assimilation of gum arabic AGP.

## MATERIALS AND METHODS

**Materials.** Gum arabic (*Acacia senegal*) was obtained from Sigma-Aldrich (St. Louis, MO, USA). Larch AGP was purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). Gum arabic AGP and Larch AGP were used after purification using ethanol precipitation.  $\alpha$ -D-Galp-(1→3)-L-Ara-free gum arabic AGP was prepared based on the method described in our previous report (9). Gum arabic AGP (1%) was incubated with GAfase at 37°C in a dialysis membrane (size 36; Wako Pure Chemical Industries Ltd., Osaka, Japan) and dialyzed against 50 mM sodium acetate buffer (pH 5.5). The polysaccharide was then precipitated with 80% ethanol. All other chemicals used in the study were obtained from Wako Pure Chemical Industries Ltd. (Osaka, Japan).

**Preparation of gum arabic AGP-derived oligosaccharides.** The schematic structures of oligosaccharides and brief purification process are shown in Fig. 1 and Fig. S2. S5-GA, S4-GA, and S3-GA were prepared from the supernatant obtained from ethanol precipitation of gum arabic as described previously (8, 9). Briefly, the ethanol supernatants were dried, dissolved in water, and separated via gel filtration chromatography on a Bio-Gel P-2 column (dimensions [ $\varphi$ ], 25 by 830 mm; Bio-Rad Laboratories, Hercules, CA, USA). Subsequently, the oligosaccharides were purified via using an activated carbon column (Autoprep FiberAC; Showa Denko, Tokyo, Japan), high-pressure liquid chromatography on a Cosmosil Sugar-D column ( $\varphi$ , 4.6 by 250 mm; Nacalai Tesque Inc., Kyoto, Japan), and a Cosmosil PBr column ( $\varphi$ , 4.6 by 250 mm; Nacalai Tesque Inc.). The structures of S5-GA, S4-GA, and S3-GA were confirmed by enzymatic reaction and MS analysis. S5 and S3 were prepared from 0.01 mM S5-GA and 0.05 mM S3-GA by enzymatic reaction with GAfase (2.9 nM and 58 nM, respectively) at 37°C until the reactions were completed. S5-A, S4-A, and S3-A were prepared from 0.01 mM S5-GA, S4-GA, and S3-GA by enzymatic reaction at 37°C with GH36  $\alpha$ -D-galactosidase BIAga3 from *B. longum* subsp. *longum* JCM7052 (0.10  $\mu$ M) (5). The reaction mixture was loaded onto a graphitized carbon (GC) cartridge column (InertSep GC column, 2 g/12 mL; GL Sciences, Tokyo, Japan) for further cleanup and enrichment of S5-A and S5. GC cartridge columns were equilibrated with 10 mL of 0.1% trifluoroacetic acid in 80% acetonitrile and 8 mL of water before sample loading. After loading the samples, oligosaccharides were eluted in 80% (vol/vol) acetonitrile containing 0.5% (vol/vol) trifluoroacetic acid. The dried samples were dissolved using an appropriate amount of water and used as the substrate.

**Expression and purification of recombinant BIArafE.** The genomic DNA of *B. longum* subsp. *longum* JCM1217 (GenBank accession no. [AP010888.1](#)) was used for PCR amplification of the gene encoding BLLJ\_1850 (BIArafE; GenPept accession no. [BAJ67514](#)) as a template. The forward (BLLJ\_1850\_for) and

reverse (BLLJ\_1850\_rev) primers (Table 1) were designed for amplifying the 88 to 4899 residues of *BlAraF*E without the coding sequences of the N-terminal signal peptide (aa 1 to 26) and C-terminal transmembrane region (aa 1637 to 1659). *BlAraF*E was amplified by high-fidelity PCR using KOD plus ver. 2 (Toyobo, Japan), and the amplified fragment was ligated into the EcoRI and XhoI sites of the pET23b(+) vector (Novagen, USA). The resulting plasmid was used to transform *Escherichia coli* BL21 ( $\lambda$ DE3) cells (Genlantis, San Diego, CA, USA), which were then cultured at 25°C using the Overnight Express autoinduction system (Novagen). The cell culture was centrifuged, and the resultant pellet was resuspended in the BugBuster protein-extraction reagent (Novagen). N-terminal His-tagged *BlAraF*E protein was purified using a column containing Talon metal-affinity resin (Clontech Laboratories Inc., Palo Alto, CA, USA). The purified fraction was desalted and concentrated using an ultrafiltration membrane (30-kDa molecular weight cutoff; Millipore Co., Billerica, MA, USA). The recombinant *BlAraF*A, *BlAraF*B, *Gafase*, and *Bl1,3Gal* enzymes used in this study were prepared according to previous studies (11, 12).

**Enzyme assays.** The reactivity of the recombinant enzymes was analyzed using the following substrates: polysaccharides [gum arabic AGP, larch AGP, and  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP] and gum arabic AGP-related oligosaccharides (S5-GA, S5-A, S5, S4-A, S3-GA, S3-A, and S3). In the case of the polysaccharides, each substrate (final concentration, 1.0%) was incubated with 2.2 nM *BlAraF*A, 2.2 nM *BlAraF*B, or 1.4 nM *BlAraF*E in 40  $\mu$ L of 50 mM sodium acetate buffer (pH 6.0) for 16 h. For the gum arabic AGP-related oligosaccharides, each substrate (final concentration, 0.01 mM) was incubated with 0.88 nM *BlAraF*A, 0.90 nM *BlAraF*B, or 0.57 nM *BlAraF*E in 100  $\mu$ L of 50 mM sodium acetate buffer (pH 6.0) for 16 h. The reaction products were analyzed by TLC or HPAEC-PAD. In the TLC analysis, the reaction products were spotted on a silica gel 60 aluminum plate (Merck, Darmstadt, Germany) with a 7:1:2 (vol/vol/vol) 1-propanol/ethanol/water solvent mixture, and the separated sugars were visualized by spraying an orcinol-sulfate reagent on the plates. In HPAEC-PAD analysis, a CarboPac PA-1 column ( $\varphi$ , 4 by 250 mm; Dionex Corp., Sunnyvale, CA, USA) was used with the flow rate of 1.0 mL/min using the following gradient: 0 to 5 min, 100% eluent A (0.1 M NaOH), 5 to 30 min, 0% to 100% eluent B (0.5 M sodium acetate in 0.1 M NaOH), and 30 to 35 min, 100% eluent B.

**Site-directed mutagenesis.** A KOD Plus mutagenesis kit (Toyobo Co., Ltd., Osaka, Japan) was used to introduce D571A and D1197A amino acid substitutions into *BlAraF*E using specific primers (Table 1). The nucleotide sequences were confirmed by sequencing. *BlAraF*E\_D571A and *BlAraF*E\_D1197A mutants were expressed and purified using the same method as that used for the wild-type *BlAraF*E. Enzyme assay of the mutant enzymes was performed using the gum arabic AGP-related oligosaccharides (S5-GA, S5-A, S5, and S4-A) and polysaccharides [gum arabic AGP,  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP, and larch AGP]. For gum arabic AGP-related oligosaccharides, S5-GA, S5-A, and S5 (final concentration, 0.025 mM) were incubated with 0.80 nM *BlAraF*E\_WT, 1.1 nM *BlAraF*E\_D571A, and 0.47 nM *BlAraF*E\_D1197A in 100  $\mu$ L of 50 mM sodium acetate buffer (pH 6.0) for 16 h. S4-A (final concentration, 0.024 mM) was incubated with 2.1 nM *BlAraF*E\_WT, 2.9 nM *BlAraF*E\_D571A, or 1.2 nM *BlAraF*E\_D1197A in 20  $\mu$ L of 50 mM sodium acetate buffer (pH 6.0) for 16 h. Polysaccharides (final concentration, 1.0%) were incubated with 1.0 nM *BlAraF*E\_WT, 1.4 nM *BlAraF*E\_D571A, and 0.57 nM *BlAraF*E\_D1197A in 40  $\mu$ L of 50 mM sodium acetate buffer (pH 6.0) for 16 h. The reaction products were analyzed by TLC or HPAEC-PAD, as mentioned above.

**In vitro assimilation test of  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP using *B. longum* strains.** The following four strains of *B. longum* used in *in vitro* assimilation tests were obtained from stock cultures maintained at the Morinaga Milk Industry Co., Ltd., Zama, Japan: MCC00300, MCC00055, MCC00198, and MCC00231. *B. longum* subsp. *longum* JCM7052 and JCM1217 were obtained from the Japan Collection of Microorganisms (Riken Bioresource Center, Ibaraki, Japan). The strains were precultured at 37°C under anaerobic conditions in MRS broth containing 0.05% L-cysteine hydrochloride using an AnaeroPack system (Mitsubishi Gas Chemical, Tokyo, Japan). The precultured cells were inoculated into MRS medium containing 1.0% gum arabic AGP or  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP as the sole carbon source and then cultured for 48 h at 37°C under anaerobic conditions. The absorbance at 600 nm of each culture medium was measured after 8, 12, 23, and 48 h. The experiment was performed in duplicates. To analyze the residual L-arabinose on  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)-L-Ara-free gum arabic AGP after culture, the polysaccharide fraction was separated using ethanol precipitation. The precipitate was dissolved in water and then incubated with 0.021  $\mu$ M *BlAraF*E in 40  $\mu$ L of 50 mM sodium acetate buffer at 37°C for 16 h. The reaction products were analyzed by TLC, as mentioned above.

**Combination reactions of type II AG degradative enzymes.** Gum arabic AGP (0.25%) was incubated with different combinations of 20 nM *Gafase*, *Bl1,3Gal*, and *BlAraF*E, and *BlAraF*A in 40  $\mu$ L of 50 mM sodium acetate buffer (pH 6.0) at 37°C for 24 h. The reaction products were analyzed by HPAEC-PAD as described above. Data from one representative experiment of three independent experiments are shown. To assay bacterial enzyme activity, the freeze-thawed cell pellets of *B. longum* subsp. *longum* JCM1217 and JCM7052 grown on larch AGP were incubated with 1.0% gum arabic AGP in 50 mM sodium acetate buffer at 37°C for 16 h, and the reaction products were analyzed using HPAEC-PAD as described above.

**Structural identification of oligosaccharide S4.** The oligosaccharide S4 was obtained from gum arabic AGP (final concentration, 5.0%) following incubation with the bacterial cell fraction of *B. longum* subsp. *longum* JCM7052 grown in larch AGP in 400 mL of 50 mM sodium acetate buffer (pH 5.0) for 53 h. The released oligosaccharides were obtained by ethanol precipitation, and the supernatant was evaporated to dryness. S4 was separated via gel-filtration chromatography on a Bio-Gel P-2 column ( $\varphi$ , 25 by 830 mm; Bio-Rad Laboratories, Hercules, CA, USA) equilibrated with water and high-pressure liquid chromatography using a Cosmosil PBr column ( $\varphi$ , 4.6 by 250 mm; Nacalai Tesque Inc.) and was eluted using 20 mM sodium phosphate (pH 2.5). Finally, the sample was desalted by gel filtration chromatography on

a Bio-Gel P-2 column equilibrated with water. The freeze-dried sample was dissolved in deuterium oxide ( $D_2O$ ).  $^1H$  and  $^{13}C$  NMR spectra and two-dimensional spectra ( $^1H$ - $^{13}C$  COSY, HMQC with/without  $^{13}C$ - $^1H$  decoupling, HMBC) were measured in  $D_2O$  at room temperature on a Jeol ECX 400 spectrometer (400 MHz). For MS, MALDI-TOF mass spectra were recorded on Shimadzu Kompact MALDI Axima-CFR spectrometer with 2,5-dihydroxybenzoic acid as the matrix. ESI-TOF mass spectra were recorded on Jeol AccuTOF JMS-T700LCK with  $CF_3CO_2Na$  as the internal standard.

**In vitro assimilation test of *Bacteroides* species on S4.** The following six *Bacteroides* species used in *in vitro* assimilation tests were obtained from the Japan Collection of Microorganisms: *B. thetaiotaomicron* JCM5827, *B. caccae* JCM9498, *B. cellulosilyticus* JCM15632, *B. ovatus* JCM5824, *B. vulgatus* JCM5826, and *B. uniformis* JCM5828. The strains were precultured in Gifu anaerobic modified medium (Nissui Pharmaceutical, Japan) at 37°C under anaerobic conditions. The precultured cells were inoculated into minimal medium (MM) (37) containing S4 as a sole carbon source (final concentration of 0.5% [wt/vol]) and then cultured at 37°C under anaerobic conditions. MM contained 100 mM  $KH_2PO_4$ , 15 mM NaCl, 8.5 mM  $(NH_4)_2SO_4$ , 4 mM L-cysteine, 200  $\mu$ M L-histidine, 1.9  $\mu$ M hematin, 100  $\mu$ M  $MgCl_2$ , 1.4  $\mu$ M  $FeSO_4 \cdot 7H_2O$ , 50  $\mu$ M  $CaCl_2$ , 1  $\mu$ g/mL of vitamin  $K_3$ , and 5 ng/mL of vitamin  $B_{12}$ . The absorbance at 600 nm was measured after 0, 6, 12, 24, and 48 h. The experiment was performed in triplicates. The supernatants were analyzed by TLC as mentioned above to analyze the residual S4 after culturing for 24 and 48 h.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 5.1 MB.

## ACKNOWLEDGMENTS

We thank Hiroyuki Koshino (Molecular Structure Characterization Unit, Riken CSRS) and his staff for ESI HRMS measurements.

This work was supported in part by JSPS Kakenhi Grant-in-Aid, grant numbers 19J20806 (to Y.S.), 19K05816 (to K.F.), 18K05345 (to A.I.), and 16H06290 (to Y.I. and A.I.).

## REFERENCES

- Jermyn M, Yeow YM, Woods E. 1975. A class of lectins present in the tissues of seed plants. *Functional Plant Biol* 2:501–531. <https://doi.org/10.1071/PP9750501>.
- Fujita K, Sakaguchi T, Sakamoto A, Shimokawa M, Kitahara K. 2014. *Bifidobacterium longum* subsp. *longum* exo- $\beta$ -1,3-galactanase, an enzyme for the degradation of type II arabinogalactan. *Appl Environ Microbiol* 80: 4577–4584. <https://doi.org/10.1128/AEM.00802-14>.
- Calame W, Weseler AR, Viebke C, Flynn C, Siemensma AD. 2008. Gum arabic establishes prebiotic functionality in healthy human volunteers in a dose-dependent manner. *Br J Nutr* 100:1269–1275. <https://doi.org/10.1017/S0007114508981447>.
- Daguet D, Pinheiro I, Verhelst A, Possemiers S, Marzorati M. 2016. Arabinogalactan and fructooligosaccharides improve the gut barrier function in distinct areas of the colon in the Simulator of the Human Intestinal Microbial Ecosystem. *J Funct Foods* 20:369–379. <https://doi.org/10.1016/j.jff.2015.11.005>.
- Sasaki Y, Uchimura Y, Kitahara K, Fujita K. 2021. Characterization of a GH36  $\alpha$ -D-galactosidase associated with assimilation of gum arabic in *Bifidobacterium longum* subsp. *longum* JCM7052. *J Appl Glycosci* 68:47–52. [https://doi.org/10.5458/jag.jag.JAG-2021\\_0004](https://doi.org/10.5458/jag.jag.JAG-2021_0004).
- Goodrum LJ, Patel A, Leykam JF, Kieliszewski MJ. 2000. Gum arabic glycoprotein contains glycomodules of both extensin and arabinogalactan-glycoproteins. *Phytochemistry* 54:99–106. [https://doi.org/10.1016/S0031-9422\(00\)00043-1](https://doi.org/10.1016/S0031-9422(00)00043-1).
- Cartmell A, Munoz-Munoz J, Briggs JA, Ndeh DA, Lowe EC, Basle A, Terrapon N, Stott K, Heunis T, Gray J, Yu L, Dupree P, Fernandes PZ, Shah S, Williams SJ, Labourel A, Trost M, Henrissat B, Gilbert HJ. 2018. A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation. *Nat Microbiol* 3:1314–1326. <https://doi.org/10.1038/s41564-018-0258-8>.
- Tischer CA, Gorin PAJ, Iacomini M. 2002. The free reducing oligosaccharides of gum arabic: aids for structural assignments in the polysaccharide. *Carbohydr Polym* 47:151–158. [https://doi.org/10.1016/S0144-8617\(01\)00173-4](https://doi.org/10.1016/S0144-8617(01)00173-4).
- Sasaki Y, Horigome A, Odamaki T, Xiao J-Z, Ishiwata A, Ito Y, Kitahara K, Fujita K. 2021. Novel 3-O- $\alpha$ -D-galactosyl- $\alpha$ -L-arabinofuranosidase for the assimilation of gum arabic arabinogalactan protein in *Bifidobacterium longum* subsp. *longum*. *Appl Environ Microbiol* 87:e02690-20. <https://doi.org/10.1128/AEM.02690-20>.
- Fujita K, Kitahara K, Sugauma T. 2012. Functional analysis of degradative enzymes for hydroxyproline-linked  $\beta$ -L-arabinofuranosides in *Bifidobacterium longum*. *Trends Glycosci Glycotechnol* 24:215–224. <https://doi.org/10.4052/tigg.24.215>.
- Fujita K, Sakamoto A, Kaneko S, Kotake T, Tsumuraya Y, Kitahara K. 2019. Degradative enzymes for type II arabinogalactan side chains in *Bifidobacterium longum* subsp. *longum*. *Appl Microbiol Biotechnol* 103:1299–1310. <https://doi.org/10.1007/s00253-018-9566-4>.
- Komeno M, Hayamizu H, Fujita K, Ashida H. 2019. Two novel  $\alpha$ -L-arabinofuranosidases from *Bifidobacterium longum* subsp. *longum* belonging to glycoside hydrolase family 43 cooperatively degrade arabinan. *Appl Environ Microbiol* 85:e02582-18. <https://doi.org/10.1128/AEM.02582-18>.
- Kotake T, Yamanashi Y, Imaizumi C, Tsumuraya Y. 2016. Metabolism of L-arabinose in plants. *J Plant Res* 129:781–792. <https://doi.org/10.1007/s10265-016-0834-z>.
- Levigne SV, Ralet M-CJ, Quémener BC, Pollet BN-L, Lapiere C, Thibault J-FJ. 2004. Isolation from sugar beet cell walls of arabinan oligosaccharides esterified by two ferulic acid monomers. *Plant Physiol* 134:1173–1180. <https://doi.org/10.1104/pp.103.035311>.
- Westphal Y, Kühnel S, de Waard P, Hinze SW, Schols HA, Voragen AG, Gruppen H. 2010. Branched arabinoligosaccharides isolated from sugar beet arabinan. *Carbohydr Res* 345:1180–1189. <https://doi.org/10.1016/j.carres.2010.03.042>.
- Virkki L, Maina HN, Johansson L, Tenkanen M. 2008. New enzyme-based method for analysis of water-soluble wheat arabinoxylans. *Carbohydr Res* 343:521–529. <https://doi.org/10.1016/j.carres.2007.10.029>.
- Vincken J-P, Wijsman AJ, Beldman G, Niessen WM, Voragen AG. 1996. Potato xyloglucan is built from XXGG-type subunits. *Carbohydr Res* 288:219–232. [https://doi.org/10.1016/S0008-6215\(96\)90801-0](https://doi.org/10.1016/S0008-6215(96)90801-0).
- York WS, Kumar Kolli VS, Orlando R, Albersheim P, Darvill AG. 1996. The structures of arabinoxyloglucans produced by solanaceous plants. *Carbohydr Res* 285:99–128. [https://doi.org/10.1016/S0008-6215\(96\)90176-7](https://doi.org/10.1016/S0008-6215(96)90176-7).
- Goellner EM, Utermoehlen J, Kramer R, Classen B. 2011. Structure of arabinogalactan from *Larix laricina* and its reactivity with antibodies directed against type-II-arabinogalactans. *Carbohydr Polym* 86:1739–1744. <https://doi.org/10.1016/j.carbpol.2011.07.006>.
- Tryfona T, Liang HC, Kotake T, Kaneko S, Marsh J, Ichinose H, Lovegrove A, Tsumuraya Y, Shewry PR, Stephens E, Dupree P. 2010. Carbohydrate structural analysis of wheat flour arabinogalactan protein. *Carbohydr Res* 345: 2648–2656. <https://doi.org/10.1016/j.carres.2010.09.018>.



21. Lamport DTA, Miller DH. 1971. Hydroxyproline arabinosides in the plant kingdom. *Plant Physiol* 48:454–456. <https://doi.org/10.1104/pp.48.4.454>.
22. Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, Venditto I, Briggs J, Zhang X, Labourel A, Terrapon N, Buffetto F, Nepogodiev S, Xiao Y, Field RA, Zhu Y, O'Neil MA, Urbanowicz BR, York WS, Davies GJ, Abbott DW, Ralet M-C, Martens EC, Henrissat B, Gilbert HJ. 2017. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 544: 65–70. <https://doi.org/10.1038/nature21725>.
23. Nie SP, Wang C, Cui SW, Wang Q, Xie MY, Phillips GO. 2013. A further amendment to the classical core structure of gum arabic (*Acacia senegal*). *Food Hydrocoll* 31:42–48. <https://doi.org/10.1016/j.foodhyd.2012.09.014>.
24. Sakka M, Yamada K, Kitamura T, Kunitake E, Kimura T, Sakka K. 2019. The modular arabinolytic enzyme Abf43A-Abf43B-Abf43C from *Ruminiclostridium josui* consists of three GH43 modules classified in different subfamilies. *Enzyme Microb Technol* 124:23–31. <https://doi.org/10.1016/j.enzmictec.2019.01.011>.
25. Amaretti A, Bernardi T, Leonardi A, Raimondi S, Zanoni S, Rossi M. 2013. Fermentation of xylo-oligosaccharides by *Bifidobacterium adolescentis* DSMZ 18350: kinetics, metabolism, and  $\beta$ -xylosidase activities. *Appl Microbiol Biotechnol* 97:3109–3117. <https://doi.org/10.1007/s00253-012-4509-y>.
26. Vandermarliere E, Bourgois TM, Winn MD, Van Campenhout S, Volckaert G, Delcour JA, Strelkov SV, Rabijns A, Courtin CM. 2009. Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem J* 418:39–47. <https://doi.org/10.1042/BJ20081256>.
27. Pastell H, Westermann P, Meyer AS, Tuomainen P, Tenkanen M. 2009. *In vitro* fermentation of arabinoxylan-derived carbohydrates by bifidobacteria and mixed fecal microbiota. *J Agric Food Chem* 57:8598–8606. <https://doi.org/10.1021/jf901397b>.
28. Rivière A, Moens F, Selak M, Maes D, Weckx S, De Vuyst L. 2014. The ability of Bifidobacteria to degrade arabinoxylan oligosaccharide constituents and derived oligosaccharides is strain dependent. *Appl Environ Microbiol* 80:204–217. <https://doi.org/10.1128/AEM.02853-13>.
29. Mewis K, Lenfant N, Lombard V, Henrissat B. 2016. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl Environ Microbiol* 82:1686–1692. <https://doi.org/10.1128/AEM.03453-15>.
30. Helbert W, Poulet L, Drouillard S, Mathieu S, Loiodice M, Couturier M, Lombard V, Terrapon N, Turchetto J, Vincentelli R, Henrissat B. 2019. Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. *Proc Natl Acad Sci U S A* 116: 6063–6068. <https://doi.org/10.1073/pnas.1815791116>.
31. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang S-J, Resch MG, Adams MWW, Lunin VV, Himmel ME, Bomble YJ. 2013. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* 342:1513–1516. <https://doi.org/10.1126/science.1244273>.
32. Mazurkewich S, Helland R, Mackenzie A, Eijsink VG, Pope PB, Brändén G, Larsbrink J. 2020. Structural insights of the enzymes from the chitin utilization locus of *Flavobacterium johnsoniae*. *Sci Rep* 10:1–11. <https://doi.org/10.1038/s41598-020-70749-w>.
33. Rakoff-Nahoum S, Coyne MJ, Comstock LE. 2014. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr Biol* 24:40–49. <https://doi.org/10.1016/j.cub.2013.10.077>.
34. Rakoff-Nahoum S, Foster KR, Comstock LE. 2016. The evolution of cooperation within the gut microbiota. *Nature* 533:255–259. <https://doi.org/10.1038/nature17626>.
35. Wyatt GM, Bayliss CE, Holcroft JD. 1986. A change in human faecal flora in response to inclusion of gum arabic in the diet. *Br J Nutr* 55:261–266. <https://doi.org/10.1079/bjn19860033>.
36. Terpend K, Possemiers S, Daguet D, Marzorati M. 2013. Arabinogalactan and fructo-oligosaccharides have a different fermentation profile in the Simulator of the Human Intestinal Microbial Ecosystem (SHIME). *Environ Microbiol Rep* 5:595–603. <https://doi.org/10.1111/1758-2229.12056>.
37. Martens EC, Chiang HC, Gordon JI. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 4:447–457. <https://doi.org/10.1016/j.chom.2008.09.007>.