1    Two Novel α-L-Arabinofuranosidases from *Bifidobacterium longum* subsp. *longum*

2    belonging to Glycoside Hydrolase Family 43 Cooperatively Degrade Arabinan

3

4    Masahiro Komeno,[a] Honoka Hayamizu,[b] Kiyotaka Fujita,[c] Hisashi Ashida[a,b#]

5

6    [a]Graduate School of Biology-Oriented Science and Technology, Kindai University,

7    Kinokawa, Wakayama, Japan

8    [b]Faculty of Biology-Oriented Science and Technology, Kindai University, Kinokawa,

9    Wakayama, Japan

10   [c]Faculty of Agriculture, Kagoshima University, Kagoshima, Kagoshima, Japan

11   [#]Address correspondence to Hisashi Ashida, ashida@waka.kindai.ac.jp.

12

13   Running title: Novel α-L-arabinofuranosidases from *B. longum longum*

1

14      **ABSTRACT** Arabinose-containing poly- or oligosaccharides are suitable carbohydrate

15      sources for *Bifidobacterium longum* subsp. *longum*, though their degradation pathways are

16      poorly understood. In this study, we found that the gene expression levels of *bllj_1852* and

17      *bllj_1853* from *B. longum* subsp. *longum* JCM 1217 were enhanced in the presence of

18      arabinan. Both genes encode previously uncharacterized glycoside hydrolase (GH) family

19      43 enzymes. Subsequently, we cloned those genes and characterized the recombinant

20      enzymes expressed in *Escherichia coli*. Both enzymes exhibited α-L-arabinofuranosidase

21      activity toward synthetic *p*-nitrophenyl glycoside, but the specificities for

22      L-arabinofuranosyl linkages were different. BLLJ_1852 catalyzed the hydrolysis of α1,2-

23      and α1,3-L-arabinofuranosyl linkages found in the side chains of arabinan and arabinoxylan.

24      BLLJ_1852 released L-arabinose 100 times faster from arabinan than from arabinoxylan

25      but did not act on arabinogalactan. BLLJ_1853 catalyzed the hydrolysis of

26      α1,5-L-arabinofuranosyl linkages found on the arabinan backbone. BLLJ_1853 released

27      L-arabinose from arabinan but not from arabinoxylan or arabinogalactan. Both enzyme

28      activities were largely suppressed with EDTA treatment, suggesting that they require

29    divalent metal ions. BLLJ_1852 was moderately activated in the presence of all divalent

30    cations tested, whereas BLLJ_1853 activity was inhibited by $Cu^{2+}$. The GH43 domains of

31    BLLJ_1852 and BLLJ_1853 are classified into GH43 subfamilies 27 and 22, respectively,

32    but hardly share similarity with other biochemically characterized members in the

33    corresponding subfamilies.

34    **IMPORTANCE** We identified two novel α-L-arabinofuranosidases from *B. longum* subsp.

35    *longum* JCM 1217 that act on different linkages in arabinan. These enzymes may be

36    required for efficient degradation and assimilation of arabinan in the probiotic

37    bifidobacteria. The genes encoding these enzymes are located side-by-side in a gene cluster

38    involved in metabolic pathways for plant-derived polysaccharides, which may confer

39    adaptability in adult intestines.

40

41    **KEYWORDS** dietary fiber, hemicellulose, gut microbiota, intestinal microbiota, prebiotics,

42    probiotics

43

## INTRODUCTION

44

45     *Bifidobacterium* spp. are gram-positive, rod-shaped, catalase-negative, strictly

46    anaerobic bacteria. Bifidobacteria mainly inhabit the large intestines of mammals including

47    humans. In humans, bifidobacteria are one of the dominant members of the intestinal

48    microbiota and are believed to exert various health-promoting effects (1, 2). Bifidobacterial

49    flora varies with age. *B. bifidum*, *B. breve*, and *B. longum* subsp. *infantis* are found in infant

50    intestines and *B. adolescentis*, *B. catenulatum*, and *B. pseudocatenulatum* live in adult

51    intestines. *B. longum* subsp. *longum* is found in intestines of both infants and adults (3). *B.*

52    *longum* subsp. *longum* is also added to food products as probiotics because of its wide

53    adaptation and beneficial health effects to the host (1).

54     In the large intestine where bifidobacteria reside, nutrient monosaccharides are

55    limited. Therefore, bifidobacteria possess various glycosidases and sugar transporters to

56    assimilate indigestible polysaccharides, oligosaccharides, and complex carbohydrates.

57    Whole genome sequencing revealed that approximately 8.5 % open reading frames (ORFs)

58    of *B. longum* NCC 2705 are related to glycan degradation and metabolism (4). In the past

4

59     decade, several groups including ours reported that bifidobacteria in infant intestines—*B.*

60     *bifidum*, *B. longum* subsp. *infantis*, and *B. longum* subsp. *longum*—utilize oligosaccharides

61     from the mother's milk (5–11) and glycoprotein glycans secreted from the gastrointestinal

62     tract (12–15). We also reported that *B. longum* subsp. *longum* possesses a unique metabolic

63     pathway to assimilate glycans from extensins that are plant cell-wall glycoproteins (16‑18).

64     It has been known that oligosaccharides derived from plant matrix polysaccharides, such as

65     arabinan, arabinoxylan, and arabinogalactan, could be carbohydrate sources for adult-type

66     bifidobacteria (19,20). However, it is poorly understood how bifidobacteria use such dietary

67     fibers. Therefore, we investigated the mechanism of arabinan utilization in *B. longum* subsp.

68     *longum* inhabiting the intestines of both infants and adults.

69        In the Carbohydrate-Active Enzymes (CAZy) database (http://www.cazy.org),

70     there are 59 putative and experimentally characterized glycosidases belonging to 23

71     glycoside hydrolase (GH) families from *B. longum* subsp. *longum* JCM 1217. Of these GH

72     families, we focused on GH43 that contains hemicellulose-degrading glycosidases

73     including α-L-arabinofuranosidase, β-xylosidase, arabinanase, and xylanase. Nine ORFs

5

74   with eleven GH43 domains occur in the genome of *B. longum* subsp. *longum* JCM 1217.

75   Notably, five previously uncharacterized GH43 ORFs, *bllj*_1850–*bllj*_1854, form a gene

76   cluster. In this study, we cloned and characterized two GH43 enzymes, BLLJ_1852 and

77   BLLJ_1853, and found that these glycosidases cooperatively degrade arabinan by acting on

78   different glycosidic bonds.

79

80   **RESULTS**

81   ***Bifidobacterium longum* subsp. *longum* JCM 1217 utilizes arabinan**

82   **as a carbon source.** Sugar-beet arabinan is composed of $\alpha$1,5-linked

83   poly-L-arabinofuranose as a backbone with branched side chains of $\alpha$1,2- or $\alpha$1,3-linked

84   L-arabinofuranose (21). Arabino-oligosaccharides are considered to be potential prebiotics

85   for bifidobacteria (19,20). However, it is not known whether arabinan is utilized by *B.*

86   *longum* subsp. *longum.* First, we tested whether *B. longum* subsp. *longum* JCM 1217 grows

87   in a medium with arabinan as the sole carbon source (Fig. 1). This strain grew well in the

88   glucose-containing medium and moderately in the L-arabinose-containing one. In the

6

89    arabinan-containing medium, it reached an $OD_{600}$ value similar to that of

90    L-arabinose-containing medium although the growth rate was slightly slower. This result

91    suggests that *B. longum* subsp. *longum* JCM 1217 degrades arabinan into L-arabinose and

92    utilizes that as a carbon source.

93        **Expression of *bllj_1852* and *bllj_1853* was enhanced in the presence**

94    **of arabinan.** To identify the glycosidases responsible for the degradation of arabinan, we

95    focused on the gene cluster consisting of *bllj_1850–bllj_1854*, all of which encode GH43

96    domain-containing proteins. Of these, we had already determined that BLLJ_1854 was an

97    α-L-arabinofuranosidase that preferably acted on arabinogalactan (Fujita *et al.*, submitted).

98    Therefore, we investigated the expression of *bllj_1850–bllj_1853* in the presence of

99    arabinan by semi-quantitative RT-PCR. The mRNA levels of *bllj_1852* and *bllj_1853* were

100    estimated to be 4.0- and 5.4-fold higher, respectively, in arabinan medium than in glucose

101    medium (data not shown). However, the expression of *bllj_1850* was not significantly

102    enhanced by arabinan (less than 1.2-fold), and mRNA of *bllj_1851* could not be detected in

103    this RT-PCR condition. This result strongly suggests that *bllj_1852* and *bllj_1853* are

7

104     involved in the utilization of arabinan in this strain.

105          **Molecular cloning of BLLJ_1852 and BLLJ_1853.** The ORF of *bllj_1852*

106     consists of 3,741 bp and encodes a polypeptide with 1,246 residues, which is predicted to

107     contain a signal peptide (aa 1–30), LamG domain (aa 95–253), bacterial Ig-like domain (aa

108     669–725), GH43 domain (aa 738–1,087), and transmembrane region (aa 1,217–1,238) (Fig.

109     2A). The ORF of *bllj_1853* consists of 3,294 bp and encodes a 1,097-aa polypeptide, which

110     contains a signal peptide (aa 1–35), LamG domain (aa 81–258), GH43 domain (aa 491–

111     782), bacterial Ig-like domain (aa 823–877), and transmembrane region (aa 1,069–1,090)

112     (Fig. 2A). Therefore, both enzymes are predicted to be extracellular membrane-bound

113     enzymes.

114          We amplified *bllj_1852* and *bllj_1853* without the regions encoding the

115     N-terminal signal peptide and C-terminal transmembrane region using high-fidelity PCR on

116     the genomic DNA from *B. longum* subsp. *longum* JCM 1217 as a template. The PCR

117     products were ligated into pET23b vectors to express C-terminally 6×His-tagged proteins.

118     *Escherichia coli* BL21(λDE3)Δ*lacZ* was transformed with these plasmids, and the

8

119    expression was induced by adding 1.0 mM isopropyl-β-D-thiogalactopyranoside (IPTG).

120    The recombinant proteins were extracted using BugBuster and purified with a $Ni^{2+}$ resin

121    column. Purified proteins migrated as a single protein band around 130 kDa for BLLJ_1852

122    and 115 kDa for BLLJ_1853 on SDS-PAGE (Fig. 2B and 2C, respectively), which are

123    consistent with the predicted molecular masses.

124    **Characterization of recombinant BLLJ_1852 and BLLJ_1853.** The

125    biochemical properties of recombinant BLLJ_1852 and BLLJ_1853 were investigated

126    using pNP-α-L-arabinofuranoside (pNP-α-Ara*f*) as a substrate. The optimum pH of both

127    glycosidases was 6.0; however, the stable pH ranges were quite different: pH 5.0–10.0 for

128    BLLJ_1852 and pH 9.0–10.0 for BLLJ_1853 (Supplementary Fig. 1A and 1B, respectively).

129    The optimum temperature of BLLJ_1852 and BLLJ_1853 were 45°C and 40°C,

130    respectively. BLLJ_1852 was stable bellow 50°C for a 1 hour treatment. In contrast,

131    BLLJ_1853 still active at 60°C for 1 hour (Supplementary Fig. 1C and 1D, respectively).

132    Therefore, we tested the stability of BLLJ_1853 at higher temperatures and found that

133    greater than 50 % activity was maintained below 90°C although it gradually declined

9

134    (Supplementary Fig. 1E). The kinetic parameters of recombinant BLLJ_1852 for

135    pNP-α-Ara*f* were determined: the $K_m$, $k_{cat}$, and $k_{cat}/K_m$ values were 9.8 mM, 19.2 s$^{-1}$, and

136    1.9 s$^{-1}$mM$^{-1}$, respectively. These parameters of BLLJ_1853 could not be determined due to

137    low activity for pNP-α-Ara*f*.

138         **Different effects of divalent metal ions on the activities of BLLJ_1852**

139    **and BLLJ_1853.** Some GH43 glycosidases were reported to have a calcium ion in their

140    active site (22). The activities of the recombinant BLLJ_1852 and BLLJ_1853 for

141    pNP-α-Ara*f* were measured in the presence of various divalent metal ions. BLLJ_1852

142    activity increased approximately 1.5-fold in the presence of all divalent metal cations tested

143    ($Zn^{2+}$, $Mn^{2+}$, $Cu^{2+}$, $Ca^{2+}$, $Co^{2+}$, $Mg^{2+}$, and $Ni^{2+}$). In contrast, BLLJ_1853 activity was not

144    enhanced by these cations and severely inhibited by $Cu^{2+}$ (Fig. 3A). Both glycosidases were

145    inhibited by EDTA, suggesting a requirement for divalent metal ions (Fig. 3A). BLLJ_1852

146    was more sensitive to EDTA than BLLJ_1853 (Fig. 3B). Next, we tested whether these

147    EDTA-treated enzymes could be restored or not by adding cations. BLLJ_1852 and

148    BLLJ_1853 were treated with 100 μM EDTA at 4°C for 16 hours, and then the activities

10

149    were measured in the presence of 1.0 mM $CaCl_2$. The activity of BLLJ_1853 was fully

150    restored by the addition of $Ca^{2+}$, whereas that of BLLJ_1852 was not restored (Fig. 3C).

151    **Substrate specificities of BLLJ_1852 and BLLJ_1853.** As described above, both

152    enzymes hydrolyzed pNP-α-Ara*f*, indicating that they are exo- α-L-arabinofuranosidases.

153    We tested the activities of these enzymes for other pNP-glycosides, such as

154    pNP-α-L-arabinopyranoside, pNP-β-xylopyranoside, and pNP-β-galactopyranoside, but no

155    hydrolysis was observed (Fig. 4A). Therefore, these glycosidases have a strict specificity

156    for α-linked L-arabinofuranose. Next, we tested substrate specificities for natural substrates.

157    Of the natural substrates tested, BLLJ_1852 released L-arabinose from arabinan and

158    arabinoxylan, and BLLJ_1853 released L-arabinose from arabinan only (Figs. 4B and 4C,

159    respectively). There are α1,2- and α1,3-arabinofuranosyl side-chins on the

160    α1,5-poly-L-arabinofuranosyl backbone of arabinan. The same α1,2- and/or

161    α1,3-arabinofuranosyl side-chins are on the xylan backbone of arabinoxylan. Next, we

162    tested linkage specificities using synthetic arabinobioses: Ara*f*α1,2Ara*f*α-OMe,

163    Ara*f*α1,3Ara*f*α-OMe, and Ara*f*α1,5Ara*f*α-OMe. BLLJ_1852 acted on the α1,2- and

11

164    $\alpha$1,3-linkages, whereas BLLJ_1853 acted only on the $\alpha$1,5-linkage (Fig. 4D). These results

165    strongly suggest that BLLJ_1852 hydrolyzes the side chains of arabinan and arabinoxylan,

166    and BLLJ_1853 hydrolyses the backbone of arabinan. Finally, we quantified the release

167    rate of L-arabinose from five substrates (Table 1). Both enzymes showed the highest

168    activity with arabinan. BLLJ_1852 also acted on arabinoxylan, but the rate was only 0.6 %

169    of that for arabinan. Synthetic glycosides, pNP-$\alpha$-Ara*f* and 4-methylumbelliferyl-$\alpha$-Ara*f*

170    (4MU-$\alpha$-Ara*f*), were very slowly hydrolyzed by both enzymes. These results indicate that

171    both enzymes have strict aglycone and linkage specificities.

172

**DISCUSSION**

174        GH43 is a large family composed of more than 10,000 sequences and currently

175    classified into 37 subfamilies (23). Nine enzymes with eleven GH43 domains from *B.*

176    *longum* subsp. *longum* JCM 1217 are listed in the CAZy database. They are highly

177    diversified with ten domains belonging to eight different subfamilies and one remaining

178    unclassified. Four GH43 enzymes, including BLLJ_1852 and BLLJ_1853, have been

12

179 characterized. BLLJ_0213, belonging to subfamily 29 of GH43 (GH43_29), is a HypAA

180 α-L-arabinofuranosidase specific for α1,3-linked Ara*f* in

181 Ara*f*α1,3Ara*f*β1-2Ara*f*β1-hydroxyprolin of plant glycoproteins (24). BLLJ_1840,

182 belonging to GH43_24, is an exo-β1,3-galactanase that acts on the backbone of type II

183 arabinogalactan (25). In this study, we found that BLLJ_1852 (GH43_27) and BLLJ_1853

184 (GH43_22), which are exo-α-L-arabinofuranosidases, cooperatively degrade arabinan.

185 BLLJ_1852 hydrolyzed Ara*f*α1,2Ara*f*α-OMe and Ara*f*α1,3Ara*f*α-OMe, suggesting that it

186 removes α1,2- and α1,3-linked Ara*f* side chains from arabinan. BLLJ_1853 specifically

187 hydrolyzed Ara*f*α1,5Ara*f*α-OMe, indicating that it processively degrades α1,5-linked

188 L-arabinose of the arabinan backbone from the non-reducing terminus. BLLJ_1852 also

189 acted very slowly on arabinoxylan, which contains Ara*f*α1,2/3Xyl*p*β1-R, but not on

190 arabinogalactan, which contains Ara*f*α1,3Gal*p*β1-R. Since Ara*f*, Xyl*p*, and Gal*p* have

191 equatorial hydroxyl groups at the C2 and C3 positions but only Gal*p* has an axial hydroxyl

192 group at the C4 position, BLLJ_1852 may have a strict recognition of equatorial hydroxyl

193 groups at the C4 position in aglycone.

13

194    Our preliminary results suggest that BLLJ_1850 (GH43_22_34), BLLJ_1851

195    (GH43_unclassified subfamily_26), and BLLJ_1854 (GH43_22) are also

196    α-L-arabinofuranosidases. BLLJ_1850 rapidly acts on arabinoxylan and weakly on

197    arabinogalactan, and BLLJ_1851 rapidly acts on arabinan and weakly on arabinoxylan. The

198    details will be reported elsewhere (Komeno *et al*., in preparation). BLLJ_1854 rapidly acts

199    on arabinogalactan and weakly on arabinan (Fujita *et al*., submitted). Except for

200    BLLJ_0213, these GH43 enzymes form a gene cluster, and there is almost no functional

201    redundancy (Fig. 5A).

202    GH43_27 is a small subfamily whose members are restricted in bacterial enzymes.

203    So far only Abf43B α-L-arabinofuranosidase from *Paenibacillus* sp. E18 (GenBank

204    accession number AFC38437) (26) and GbtXyl43B β- D-xylosidase from *Geobacillus*

205    *thermoleovorans* IT-08 (ABD48561) (27) have been experimentally characterized. The

206    former hydrolyzed pNP-α-L-Ara*f* but neither arabinan nor arabinoxylan, and the latter

207    hydrolyzed pNP-β-D-Xyl*p* and xylan. Since BLLJ_1852 is an arabinan-degrading

208    exo-1,2-1,3-α-L-arabinofuranosidase, the substrate specificity of BLLJ_1852 is quite

14

209    different from those of Abf43B and GbtXyl43B. The amino acid sequence of the GH43_27

210    domain (aa 738–1,087) of BLLJ_1852 had only 10 and 9 % identity with the corresponding

211    domains of Abf43B and GbtXyl43B, respectively. Adult-type bifidobacteria have putative

212    homologs with higher identities of 45–50 %: BAD_0149 from *B. adolescentis* ATCC 15703,

213    BBCT_0647 from *B. catenulatum*, BBPC_0158 from *B. pseudocatenulatum*, and

214    BBSC_0258 from *B. scardovii*. However, within the closely related strains of *B. longum*

215    whose genomic data are available, the BLLJ_1852 homolog was lost in several strains, such

216    as *B. longum* 35264, *B. longum* BG7, and *B. longum* NCC2705 (Fig. 5B), and is also

217    missing in *B. longum* subsp. *infantis*.

218         In comparison with GH43_27, the subfamily GH43_22 is relatively larger and

219    more widely distributed in eukaryote (several fungi) and archaea in addition to bacteria. In

220    this subfamily, we could not find an experimentally characterized enzyme. In our strain,

221    there are three enzymes belonging to this subfamily, BLLJ_1850 (GH43_22_34),

222    BLLJ_1853, and BLLJ_1854, which exhibit distinct substrate preference for arabinoxylan,

223    arabinan, and arabinogalactan, respectively, as described above. Within *B. longum* strains,

15

224     the homologs of BLLJ_1853 and BLLJ_1850 are conserved expect for *B. longum*

225     NCC2705, and those of BLLJ_1854 are conserved in all strains (Fig. 5B). BLLJ_1853 is

226     arabinan-degrading exo-1,5-α-L-arabinofuranosidase. The enzymes with similar substrate

227     specificity are classified in the subfamily GH43_26: Araf43A (BAL68753) from

228     *Streptomyces avermitilis* MA-4680 (28, 29), and Abf43K (ACE82749) and Abf43L

229     (ACE84379) from *Cellvibrio japonics* Ueda107 (30). However, BLLJ_1853 shares very

230     low identities with the amino acid sequences of Araf43A (13%), Abf43K (11%), and

231     Abf43L (15%), suggesting that the evolutionary origin may be different. The

232     exo-1,5-α-L-arabinofuranosidase that belongs to the other families of GH51 was also

233     reported: AbfA (CAA99595) from *Bacillus subtilis* strain 168 (31).

234          It is known that *B. longum* subsp. *longum* is widely distributed in the intestines of

235     infants, adults, and elderly hosts. A recent study on the comparative genome analyses of

236     human isolates revealed that this GH43 gene cluster is enriched in strains from elderly

237     subjects (3). These findings suggest that this gene cluster is generally important for the

238     colonization of *B. longum* subsp. *longum* in the intestines of adults who consume

16

239    vegetables and grains.

240

241    **MATERIALS AND METHODS**

242        **Bacterial strains and culture.** *B. longum* subsp. *longum* JCM 1217 was

243    obtained from the Japan Collection of Microorganisms (RIKEN Bioresource Center, Japan).

244    The strain was cultured in GAM broth (Nissui Pharmaceutical, Japan) or sugar-restricted

245    basal medium (1.0 % Bacto peptone, 0.5 % Bacto yeast extract, 0.5 % sodium acetate

246    trihydrate, 0.2 % di-ammonium hydrogen citrate, 0.08 % L-cysteine hydrochloride

247    monohydrate, 0.02 % magnesium sulfate heptahydrate, 1.36 % L-ascorbic acid, and 0.44 %

248    sodium carbonate anhydrous) at 37°C under anaerobic conditions using the AnaeroPack

249    Anaero (Mitsubishi Chemical, Japan). *Escherichia coli* DH5α and *E. coli*

250    BL21(λDE3)Δ*lac*Z (a gift from Prof. T. Katayama, Kyoto University, Kyoto, Japan) were

251    cultured in Luria-Bertani (LB) broth (BD, USA) at 37°C under aerobic conditions.

252        **Chemicals and substrates.** *p*-Nitrophenyl (pNP)-glycosides were purchased

253    from Sigma-Aldrich, MO, USA; sugar-beet arabinan and wheat arabinoxylan were

17

254    purchased from Megazyme, Ireland; larch wood arabinogalactan was purchased from

255    Tokyo Chemical Industry, Japan; and L-arabinose was purchased from Wako Chemicals,

256    Japan. Synthetic 1-*O*-methylated arabinobioses (Ara$f\alpha$1,2Ara$f\alpha$-OMe, Ara$f\alpha$1,3Ara$f\alpha$-OMe,

257    and Ara$f\alpha$1,5Ara$f\alpha$-OMe) were gifted from Prof. S. Kaneko, Ryukyu University, Japan (32).

258    Other chemicals were obtained from Nacalai Tesque Inc., Japan.

259        **Growth assay of *B. longum* subsp. *longum*.** *B. longum* subsp. *longum*

260    JCM 1217 was inoculated in sugar-restricted basal medium supplemented with 1.0 %

261    L-arabinose and cultured at 37°C under anaerobic conditions until the optical density at 600

262    nm (OD$_{600}$) reached 1.0. A 10 % volume seed culture was added to basal medium

263    containing 2.0 % arabinan, 2.0 % L-arabinose, or 2.0 % glucose. The wells of a 96-well

264    plate were filled with these mixtures followed by sealing with Microseal B (Bio-Rad, USA).

265    The plate was maintained at 37°C, and the OD$_{600}$ was monitored every 3 hours using

266    Powerscan HT (Dainippon Sumitomo Pharma, Japan).

267        **Semi-quantitative RT-PCR.** *B. longum* subsp. *longum* JCM 1217 was cultured

268    at 37°C under anaerobic conditions in sugar-restricted basal medium with either 2.0 %

18

269    arabinan or 2.0 % glucose. RNA extraction was carried out using the TRIzol Reagent

270    (Thermo Fisher Scientific, MA, USA) and purification used the PureLink RNA Mini Kit

271    (Thermo Fisher Scientific). Reverse transcription was performed using PrimeScript Reverse

272    Transcriptase (Takara Bio Inc., Japan). Primers for RT-PCR were designed as follows:

273    5'-CCCAAGCTTGATACCACCGATTCATCGGCCGC and

274    5′-GTGCCGCTCGAGGGAGATGACGGCACCCGGCTTCTTG for *bllj_1850*,

275    5'-CGCGGATCCGGACGACGCGACGCCTGCGGTG and

276    5′-CCCAAGCTTAGAGGACGGCAGACCGGAGTCTGC for *bllj_1851*, 5′-

277    GGAATTCCATATGGATACCGTTCCGACCAATAATCTCATC and 5′-

278    CCCAAGCTTAGACAGACCGAGCTTGTTGCCCG for *bllj_1852,* and 5′-

279    CGGAATTCGGAGAGCGCATCGCCAATCGATG and 5′-

280    GTGCCGCTCGAGGGTGTTGGACAGAGCGCTGCCCGG for *bllj_1853*. RT-PCR was

281    performed using TaKaRa Ex Taq HS (Takara Bio Inc.). PCR products were separated by

282    electrophoresis on a 0.8 % agarose gel. The gel staining used SYBR Gold Nucleic Acid Gel

283    Stain (Thermo Fisher Scientific). The densities of the bands were quantified using the

19

284    ImageJ program.

285        **Cloning, expression, and purification of BLLJ_1852 and BLLJ_1853.**

286    The DNA fragment for base pairs 106-3645 of *bllj_1852* (GenBank accession number

287    BAJ67516) without the coding sequences of the N-terminal signal peptide and C-terminal

288    transmembrane region was amplified by high-fidelity PCR using KOD plus ver. 2 (Toyobo,

289    Japan), genomic DNA from *B. longum* subsp. *longum* JCM 1217 as a template. The same

290    pair of primers for RT-PCR with the NdeI and HindIII sites were used. The amplified

291    fragment was ligated into the NdeI and HindIII sites of the pET23b(+) vector (Novagen,

292    USA). Similarly, the DNA fragment of base pairs 91-3203 of *bllj_1853* (BAJ67517) was

293    amplified using same the pair of primers as the RT-PCR containing the EcoRI and XhoI

294    sites. The product was ligated into corresponding sites of pET23b(+). The nucleotide

295    sequences were confirmed by sequencing. The *E. coli* BL21(λDE3)Δ*lacZ* strain was

296    transformed with these plasmids and cultured in LB liquid medium containing 100 μg/mL

297    ampicillin at 37°C under aerobic conditions until the $OD_{600}$ reached 0.35. Then the flask

298    was put on ice in order to lower the temperature. Next, to induce expression, 0.1 mM

20

299    isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the culture, which was further

300    incubated at 25°C for 15 hours. The cells were harvested by centrifugation and lysed with

301    the BugBuster Protein Extraction Reagent (Novagen). The cell free protein extract was

302    applied to a His GraviTrap (GE Healthcare, USA), and adsorbed proteins were eluted by a

303    stepwise imidazole concentration gradient in 20 mM sodium acetate buffer at pH 6.5

304    containing 500 mM NaCl. The active fraction was dialyzed with a size 36 dialysis

305    membrane (Wako Chemicals) against 0.5 mM sodium acetate buffer at pH 6.5 overnight at

306    4°C.

307    Protein concentration was measured using a Bio-Rad Protein Assay Dye Reagent

308    Concentrate (Bio-Rad).

309        **Glycosidase assay.** α-L-Arabinofuranosidase activity was measured based on

310    the release of pNP from pNP-α-Ara$f$. The reaction mixture containing 1.0 mM pNP-Ara$f$,

311    an appropriate amount of purified enzyme, and 1.0 mM sodium acetate buffer at pH 5.5 to a

312    total volume of 50 μL was incubated at 37°C. After the incubation, 1.5 volumes of 1.0 M

313    $Na_2CO_3$ was added as a stop solution, and the released $p$NP was measured at 405 nm. For

21

314    optimum pH tests, sodium acetate (pH 3.0–6.0), sodium phosphate (pH 6.5–7.5), and

315    Tris-HCl buffers (pH 8.0–10.0) were used. The 50 μL reaction mixture containing 1.0 mM

316    of each pH buffer, 1.0 mM pNP-α-Ara*f*, and 5 μg enzyme was incubated at 37°C for 30

317    min (BLLJ_1852) or overnight (BLLJ_1853). For a stable pH assay, the enzyme was

318    pretreated in 10 mM buffer at 4°C for 24 hours, and then the remaining activity was

319    measured in sodium acetate buffer at pH 5.5. For the metal ion test, 5 mM chloride salts

320    ($ZnCl_2$, $MnCl_2$, $CuCl_2$, $CaCl_2$, $CoCl_2$, $MgCl_2$, and $NiCl_2$) or EDTA 2Na were added to

321    reaction mixtures. For the assays for natural substrates, such as arabinan, arabinoxylan, and

322    arabinogalactan, the reaction mixture containing 0.5 % substrate, 1.0 mM sodium acetate

323    buffer at pH 5.5, and the indicated amount of enzyme was incubated at 37°C for the

324    appropriate period. The reaction mixture was analyzed by TLC or applied to enzymatic

325    measurements of released L-arabinose using an L-arabinose/D-galactose assay kit

326    (Megazyme).

327        **TLC.** TLC analysis was carried out using silica gel 60 plates (Merck, Germany)

328    with 1-butanol:acetic acid:water (2:1:1 by volume) or chloroform:methanol:acetic acid

329    (6:1:1 by volume) as developing solvents. Sugars were visualized by spraying

330    acetone:aniline:diphenyl amine:phosphoric acid (100:1:1:10 by volume) and heating to

331    150°C for 15 min.

332

333    **SUPPLEMENTAL MATERIAL**

334    Supplemental material for this article may be found at ###

335    SUPPLEMENTAL FILE 1, PDF file, ## MB.

336

337    **ACKNOWLEDGMENTS**

342        We declare that there is no conflict of interest.

343

344 **REFERENCES**

345     1. O'Callaghan A, Sinderen D. 2016. Bifidobacteria and their role as members of the

346        human gut microbiota. Front Microbiol 7:925.

347     2. Mitsuoka T. 2014. Establishment of intestinal bacteriology. Biosci Microbiota Food

348        Health 33:99-116.

349     3. Odamaki T, Bottacini F, Kato K, Mitsuyama E, Yoshida K, Horigome A, Xiao J-ZZ,

350        van Sinderen D. 2018. Genomic diversity and distribution of *Bifidobacterium*

351        *longum* subsp. *longum* across the human lifespan. Sci Rep 8:85.

352     4. Schell MA, Karmirantzou M, Snel B, Vilanova D, Berger B, Pessi G, Zwahlen

353        M-CC, Desiere F, Bork P, Delley M, Pridmore RD, Arigoni F. 2002. The genome

354        sequence of *Bifidobacterium longum* reflects its adaptation to the human

355        gastrointestinal tract. Proc Natl Acad Sci USA 99:14422-14427.

356     5. Kitaoka M, Tian J, Nishimoto M. 2005. Novel putative galactose operon involving

357        lacto-*N*-biose phosphorylase in *Bifidobacterium longum*. Appl Environ Microbiol

358        71:3158-3162.

359    6. Wada J, Ando T, Kiyohara M, Ashida H, Kitaoka M, Yamaguchi M, Kumagai H,

360        Katayama T, Yamamoto K. 2008. *Bifidobacterium bifidum* lacto-*N*-biosidase, a

361        critical enzyme for the degradation of human milk oligosaccharides with a type 1

362        structure. Appl Environ Microbiol 74:3996-4004.

363    7. Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A,

364        Rokhsar DS, Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA. 2008.

365        The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals

366        adaptations for milk utilization within the infant microbiome. Proc Natl Acad Sci

367        USA 105:18964-18969.

368    8. Ashida H, Miyake A, Kiyohara M, Wada J, Yoshida E, Kumagai H, Katayama T,

369        Yamamoto K. 2009. Two distinct α-L-fucosidases from *Bifidobacterium bifidum* are

370        essential for the utilization of fucosylated milk oligosaccharides and

371        glycoconjugates. Glycobiology 19:1010-1017.

372    9. Miwa M, Horimoto T, Kiyohara M, Katayama T, Kitaoka M, Ashida H, Yamamoto

373        K. 2010. Cooperation of β-galactosidase and β-*N*-acetylhexosaminidase from

25

374    bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure.

375    Glycobiology 20:1402-1409.

376    10. Kiyohara M, Tanigawa K, Chaiwangsri T, Katayama T, Ashida H, Yamamoto K.

377    2011. An exo-$\alpha$-sialidase from bifidobacteria involved in the degradation of

378    sialyloligosaccharides in human milk and intestinal glycoconjugates. Glycobiology

379    21:437-447.

380    11. Yamada C, Gotoh A, Sakanaka M, Hattie M, Stubbs K, Katayama-Ikegami A,

381    Hirose J, Kurihara S, Arakawa T, Kitaoka M, Okuda S, Katayama T, Fushinobu S.

382    2017. Molecular insight into evolution of symbiosis between breast-fed infants and

383    a member of the human gut microbiome *Bifidobacterium longum*. Cell Chem Biol

384    24:515-524.e5.

385    12. Fujita K, Oura F, Nagamine N, Katayama T, Hiratake J, Sakata K, Kumagai H,

386    Yamamoto K. 2005. Identification and molecular cloning of a novel glycoside

387    hydrolase family of core 1 type *O*-glycan-specific

388    endo-$\alpha$-*N*-acetylgalactosaminidase from *Bifidobacterium longum*. J Biol Chem

389     45:37415-37422.

390     13. Kiyohara M, Nakatomi T, Kurihara S, Fushinobu S, Suzuki H, Tanaka T, Shoda S,

391     Kitaoka M, Katayama T, Yamamoto K, Ashida H. 2012.

392     α-*N*-Acetylgalactosaminidase from infant-associated bifidobacteria belonging to a

393     novel glycoside hydrolase family 129 is implicated in an alternative mucin

394     degradation pathway. J Biol Chem 1:693-700.

395     14. Wakinaka T, Kiyohara M, Kurihara S, Hirata A, Chaiwangsri T, Ohnuma T,

396     Fukamizo T, Katayama T, Ashida H, Yamamoto K. 2012. Bifidobacterial

397     α-galactosidase with unique carbohydrate-binding module specifically acts on blood

398     group B antigen. Glycobiology 2:232-240.

399     15. Shimada Y, Watanabe Y, Wakinaka T, Funeno Y, Kubota M, Chaiwangsri T,

400     Kurihara S, Yamamoto K, Katayama T, Ashida H. 2015.

401     α-*N*-Acetylglucosaminidase from *Bifidobacterium bifidum* specifically hydrolyzes

402     α-linked *N*-acetylglucosamine at nonreducing terminus of *O*-glycan on gastric

403     mucin. Appl Microbiol Biotechnol 99:3941-3948.

404    16. Fujita K, Sakamoto S, Ono Y, Wakao M, Suda Y, Kitahara K, Suganuma T. 2010.

405        Molecular cloning and characterization of a β-L-arabinobiosidase in

406        *Bifidobacterium longum* that belongs to a novel glycoside hydrolase family. J

407        Biol Chem 7:5143-5150.

408    17. Fujita K, Takashi Y, Obuchi E, Kitahara K, Suganuma T. 2014. Characterization of

409        a novel β-L-arabinofuranosidase in *Bifidobacterium longum*: functional elucidation

410        of a DUF1680 family member. J Biol Chem 8:5240-5249.

411    18. Fujita K, Kitahara K, Suganuma T. 2012. Functional analysis of degradative

412        enzymes for hydroxyproline-linked β-L-arabinofuranosides in *Bifidobacterium*

413        *longum*. Trends Glycosci Glycotech 24:215-224.

414    19. Laere K, Hartemink R, Bosveld M, Schols H, Voragen A. 2000. Fermentation of

415        plant cell wall derived polysaccharides and their corresponding oligosaccharides by

416        intestinal bacteria. J Agric Food Chem 48:1644-1652.

417    20. Holck J, Lorentzen A, Vigsnæs LK, Licht TR, Mikkelsen JD, Meyer AS. 2011.

418        Feruloylated and nonferuloylated arabino-oligosaccharides from sugar beet pectin

419      selectively stimulate the growth of *Bifidobacterium* spp. in human fecal in

420      vitro fermentations. J Agric Food Chem 59:6511-6519.

421   21. Westphal Y, Kühnel S, de Waard P, Hinz SW. A, Schols H. A, Voragen AG. J,

422      Gruppen H. 2010. Branched arabino-oligosaccharides isolated from sugar beet

423      arabinan. Carbohydr Res 345:1180-1189.

424   22. De Sanctis D, Inácio JM, Lindley PF, Sá-Nogueira I, Bento I. 2010. New evidence

425      for the role of calcium in the glycosidase reaction of GH43 arabinanases. FEBS J

426      277:4562-4574.

427   23. Jordan DB, Braker JD, Wagschal K, Lee CC, Chan VJ, Dubrovska I, Anderson S,

428      Wawrzak Z. 2015. X-ray crystal structure of divalent metal-activated β-xylosidase,

429      RS223BX. Appl Biochem Biotechnol 117:637-648.

430   24. Mewis K, Lenfant N, Lombard V, Henrissat B. 2016. Dividing the large glycoside

431      hydrolase family 43 into subfamilies: a motivation for detailed enzyme

432      characterization. Appl Environ Microbiol 82:1686-1692.

433   25. Fujita K, Sakaguchi T, Sakamoto A, Shimokawa M, Kitahara K. 2014.

29

434     *Bifidobacterium longum* subsp. *longum* exo-β-1,3-galactanase is an enzyme for the

435     degradation of type II arabinogalactan. Appl Environ Microbiol 80:4577-4584.

436     26. Shi P, Chen X, Meng K, Huang H, Bai Y, Luo H, Yang P, Yao B. 2013. Distinct

437     actions by *Paenibacillus* sp. E18 α-L-arabinofuranosidases and xylanase on xylan

438     degradation. Appl Environ Microbiol 79:1990-1995.

439     27. Ratnadewi AAI, Fanani M, Kurniasih SD, Sakka M, Wasito EB, Sakka K,

440     Nurachman Z, Puspaningsih NNT. 2013. β-D-Xylosidase from *Geobacillus*

441     *thermoleovorans* IT-08: biochemical characterization and bioinformatics of the

442     enzyme. Appl Biochem Biotechnol 170:1950-1964.

443     28. Ichinose H, Yoshida M, Fujimoto Z, Kaneko S. 2008. Characterization of a modular

444     enzyme of exo-1,5-α-L-arabinofuranosidase and arabinan binding module from

445     *Streptomyces avermitilis* NBRC14893. Appl Environ Microbiol 80:399-408.

446     29. Fujimoto Z, Ichinose H, Maehara T, Honda M, Kitaoka M, Kaneko S. 2010. Crystal

447     structure of an exo-1,5-α-L-arabinofuranosidase from *Streptomyces avermitilis*

448     provides insights into the mechanism of substrate discrimination between exo- and

449     endo-type enzymes in glycoside hydrolase family 43. J Biol Chem 44:34134-34143.

450     30. Cartmell A, McKee L, Peña MJ, Larsbrink J, Brumer H, Kaneko S, Ichinose H,

451     Lewis RJ, Viksø-Nielsen A, Gilbert HJ, Marles-Wright J. 2011. The structure and

452     function of an arabinan-specific α-1,2-arabinofuranosidase identified from

453     screening the activities of bacterial GH43 glycoside hydrolases. J Biol Chem

454     17:15483-15495.

455     31. Inácio JM, Correia IL, de Sá-Nogueira I. 2008. Two distinct arabinofuranosidases

456     contribute to arabino-oligosaccharide degradation in *Bacillus subtilis*. Microbiology

457     154:2719-2729.

458     32. Kawabata Y, Kaneko S, Kusakabe I, Gama Y. 1995. Synthesis of regioisomeric

459     methyl α-L-arabinofuranobiosides. Carbohydr Res 267:39-47.

460

461

462

463    **FIGURE LEGENDS**

464

465    **FIG 1** Growth of *B. longum* subsp. *longum* JCM 1217 in media containing different carbon

466    sources. The strain was cultured at 37 °C in sugar-restricted basal medium supplemented

467    with 1.0 % of different carbon sources under anaerobic conditions. $OD_{600}$ was measured

468    every 3 hours for 48 hours. Error bars indicate SD (n = 6). Symbols: circle, no sugar;

469    triangle, arabinan; diamond, L-arabinose: square, glucose.

470

471    **FIG 2** Characterization of BLLJ_1852 and BLLJ_1853. (A) Domain structure of

472    BLLJ_1852 and BLLJ_1853. (B) SDS-PAGE of purified BLLJ_1852. (C) SDS-PAGE of

473    purified BLLJ_1853. M, molecular size markers.

474

475    **FIG 3** Effects of divalent metal ions on the activities of BLLJ_1852 and BLLJ_1853. (A)

476    Effects of metal ions on the activities of recombinant enzymes. The divalent metal ions

477    were added at 5.0 mM. (B) Effects of various concentration of EDTA on recombinant

478     enzymes. (C) Restoration of enzyme activity by addition of $Ca^{2+}$ after treatment with EDTA.

479     pNP-α-Ara*f* was used as a substrate in all experiments. Error bars indicate SD (n=3).

480

481     **FIG 4** Substrate specificities of BLLJ_1852 and BLLJ_1853. The substrates were

482     incubated with recombinant BLLJ_1852 or BLLJ_1853 at 37°C overnight and analyzed by

483     TLC. (A) Action on synthetic substrates. α-Ara*f*, pNP-α-L-arabinofuranoside; α-Ara*p*,

484     pNP-α-L-arabinopyranoside; β-Xyl*p*, pNP-β-xylopyranoside; β-Gal*p*,

485     pNP-β-galactopyranoside; Ara, standard L-arabinose; Xyl, standard xylose; Gal, standard

486     galactose. (B) Action of BLLJ_1852 on natural substrates. AX, arabinoxylan; AG,

487     arabinogalactan; AN, arabinan; Ara, standard L-arabinose. (C) Action of BLLJ_1853 on

488     natural substrates. AX, arabinoxylan; AG, arabinogalactan; AN, arabinan; Ara, standard

489     L-arabinose. (D) Action of BLLJ_1852 and BLLJ_1853 on synthetic disaccharide substrates.

490     α2, Ara*f*α1,2Ara*f*α-OMe; α3, Ara*f*α1,3Ara*f*α-OMe; α5, Ara*f*α1,5Ara*f*α-OMe; Ara,

491     standard L-arabinose. The developing solvent used in panels A-C was 1-Butanol:acetic

33

492    acid:water (2:1:1, by volume); chloroform:methanol:acetic acid (6:1:1, by volume) was

493    used for panel D.

494

495    **FIG 5** (A) Map of the hemicellulose-degrading gene cluster in *B. longum* subsp. *longum*

496    JCM 1217. Gray boxes, glycosidases; striped boxes, sugar transporters; black boxes,

497    transcriptional regulators; white boxes, hypothetical proteins. BLLJ_1850 and BLLJ_1850

498    have two GH43 domains, and the N-terminal GH43 domain of BLLJ_1850 belongs to an

499    unclassified (UC) subfamily. (B) Conservation of the GH43 cluster in various *B. longum*

500    strains. Scores indicated above the boxes indicate percentage identity of amino acid

501    sequences with those of *B. longum* subsp. *longum* JCM 1217.

502

503   **TABLE 1** Specific activities of BLLJ_1852 and BLLJ_1853 for various substrates

504

505

| | pNP | 4MU | AN | AX | AG |
|---|---|---|---|---|---|
| BLLJ_1852 | 1.00 | 0.62 | 70.23 | 0.62 | nd |
| BLLJ_1853 | 0.009 | 0.23 | 4.42 | nd | nd |

508   Reaction mixtures were incubated at 37°C for 30 min. The released L-arabinose was

509   determined by the enzymatic measurement method, and the relative values are shown.

510   pNP, pNP-α-L-arabinofuranoside; 4MU, 4-methylumbelliferyl-α-L-arabinofuranoside; AN,

511   sugar-beet arabinan; AX, wheat arabinoxylan; AG, larch wood arabinogalactan; nd, not

512   detected.

513

514    **Fig. S1** General properties of recombinant BLLJ_1852 and BLLJ_1853. pNP-α-Ara*f* was

515    incubated with recombinant enzymes at 37°C. (A) Effect of pH on BLLJ_1852. (B) Effect

516    of pH on BLLJ_1853. (C) Effect of temperature on BLLJ_1852. (D) Effect of temperature

517    on BLLJ_1853. (E) Thermal stability of BLLJ_1853 at high temperatures. Error bars

518    indicate SD (n=3). For stability tests, enzymes were pre-incubated in each pH buffer for 16

519    hours at 37°C or in sodium acetate buffer (pH 6.0) at each temperature for 1 hour.
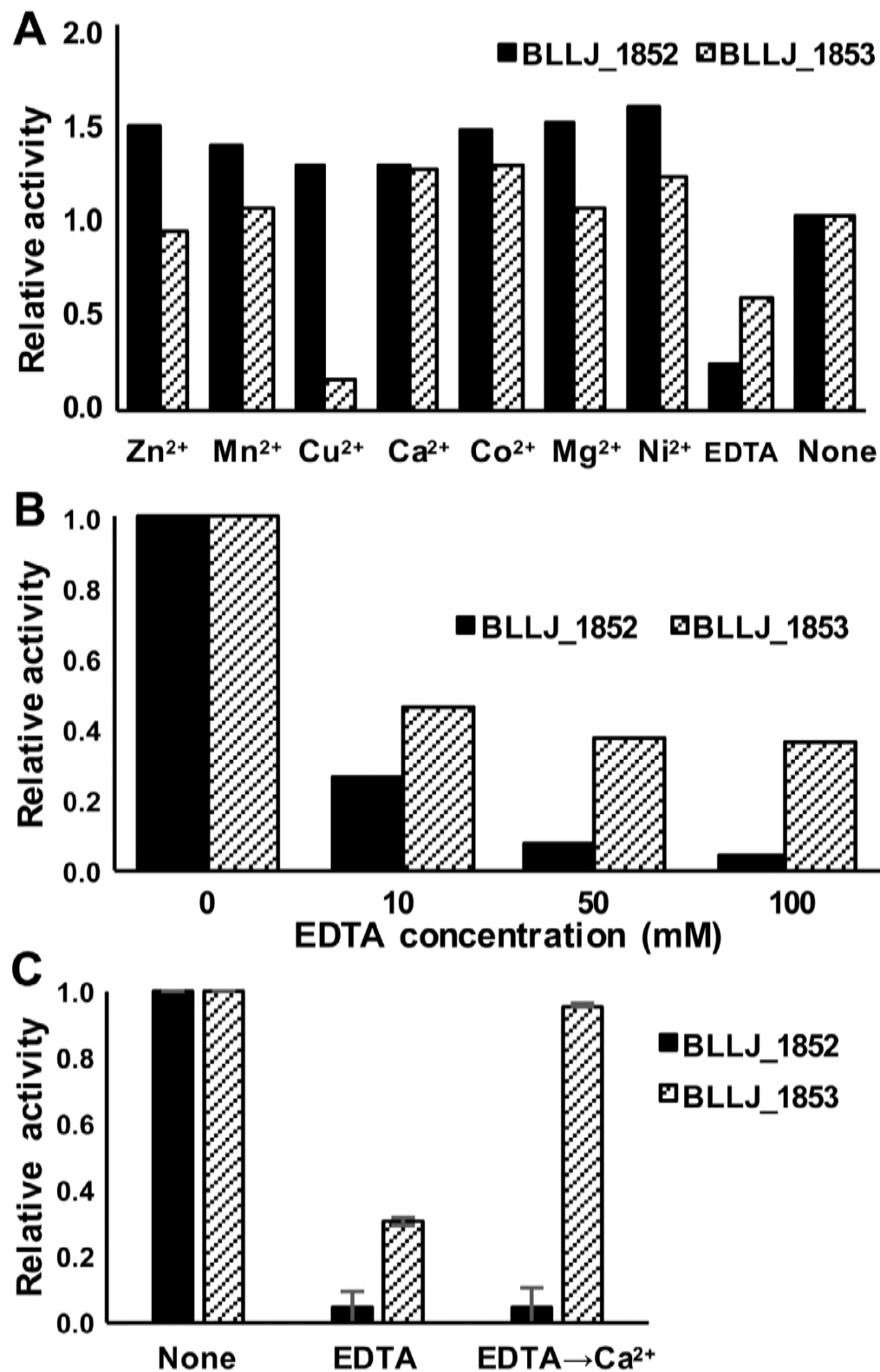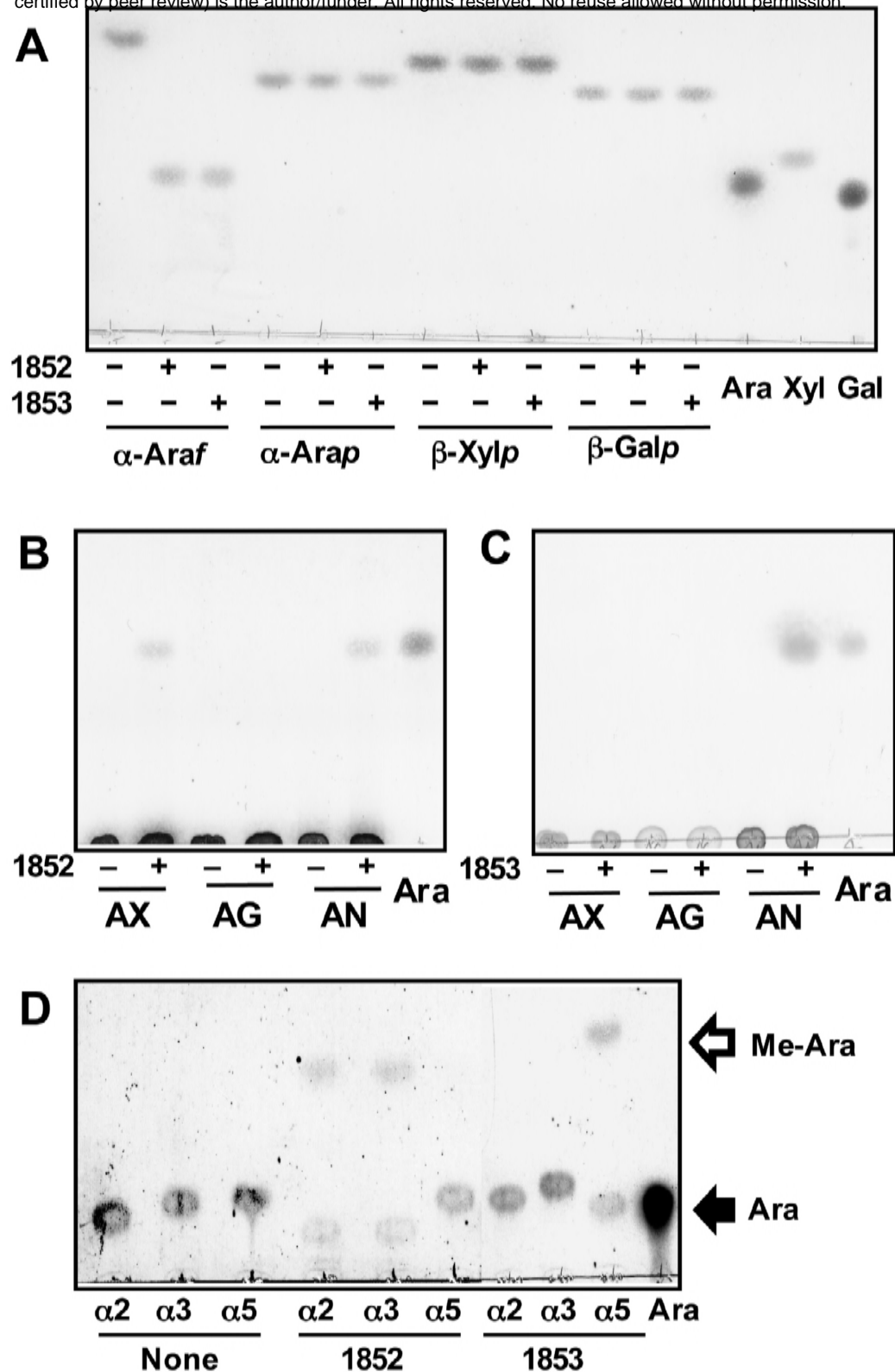
520
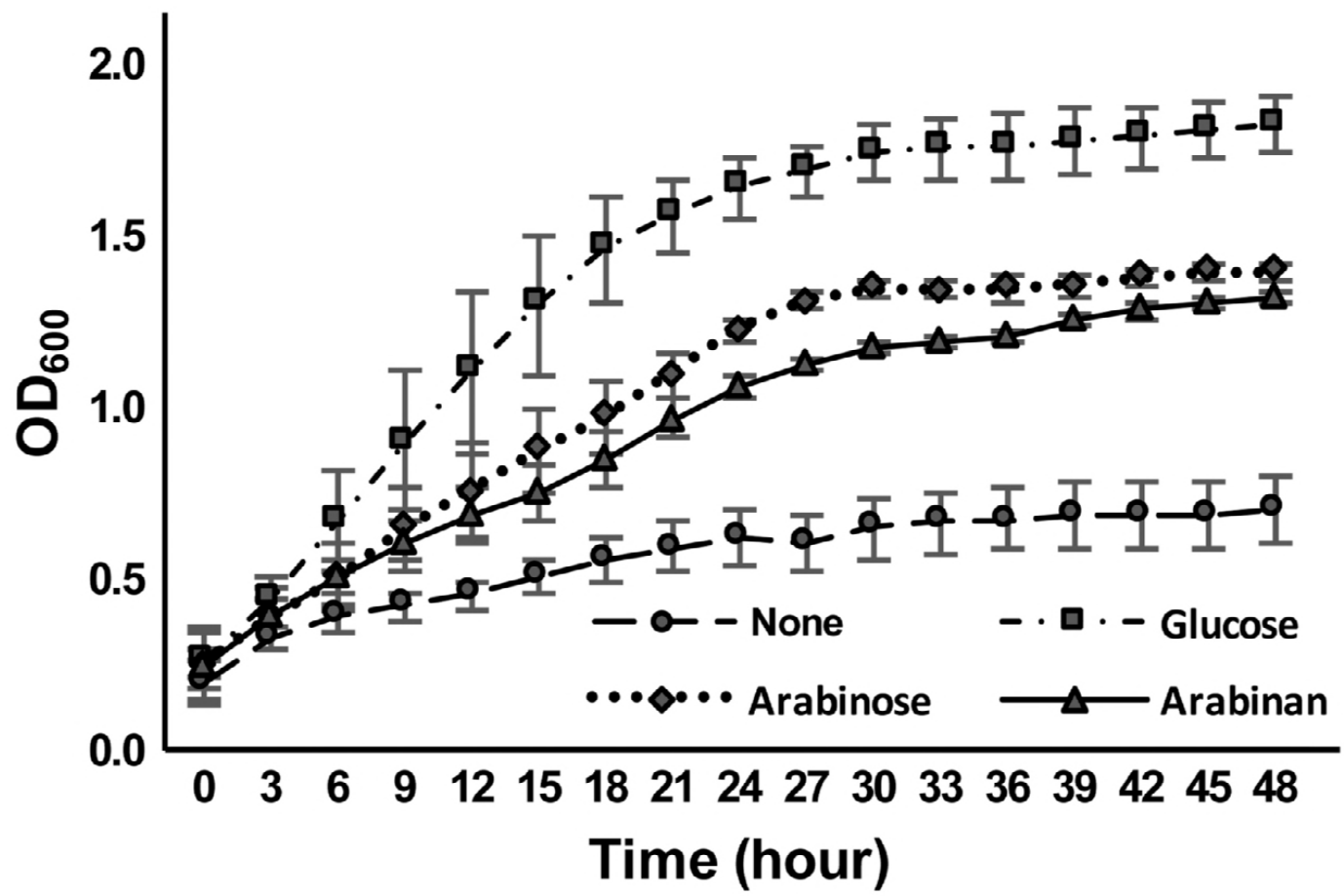
Fig. 1. Komeno et. al.
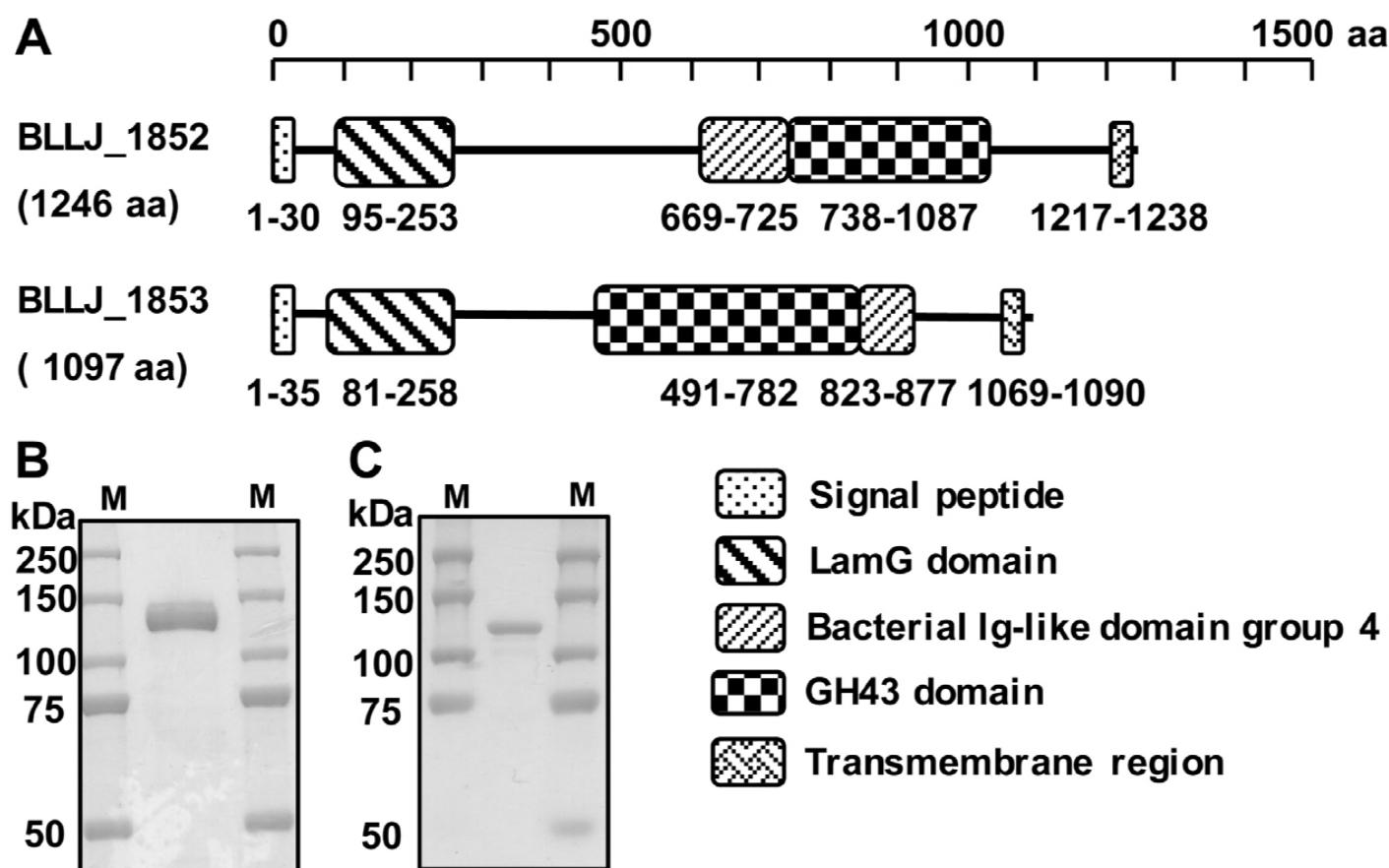
Fig. 2. Komeno et. al.

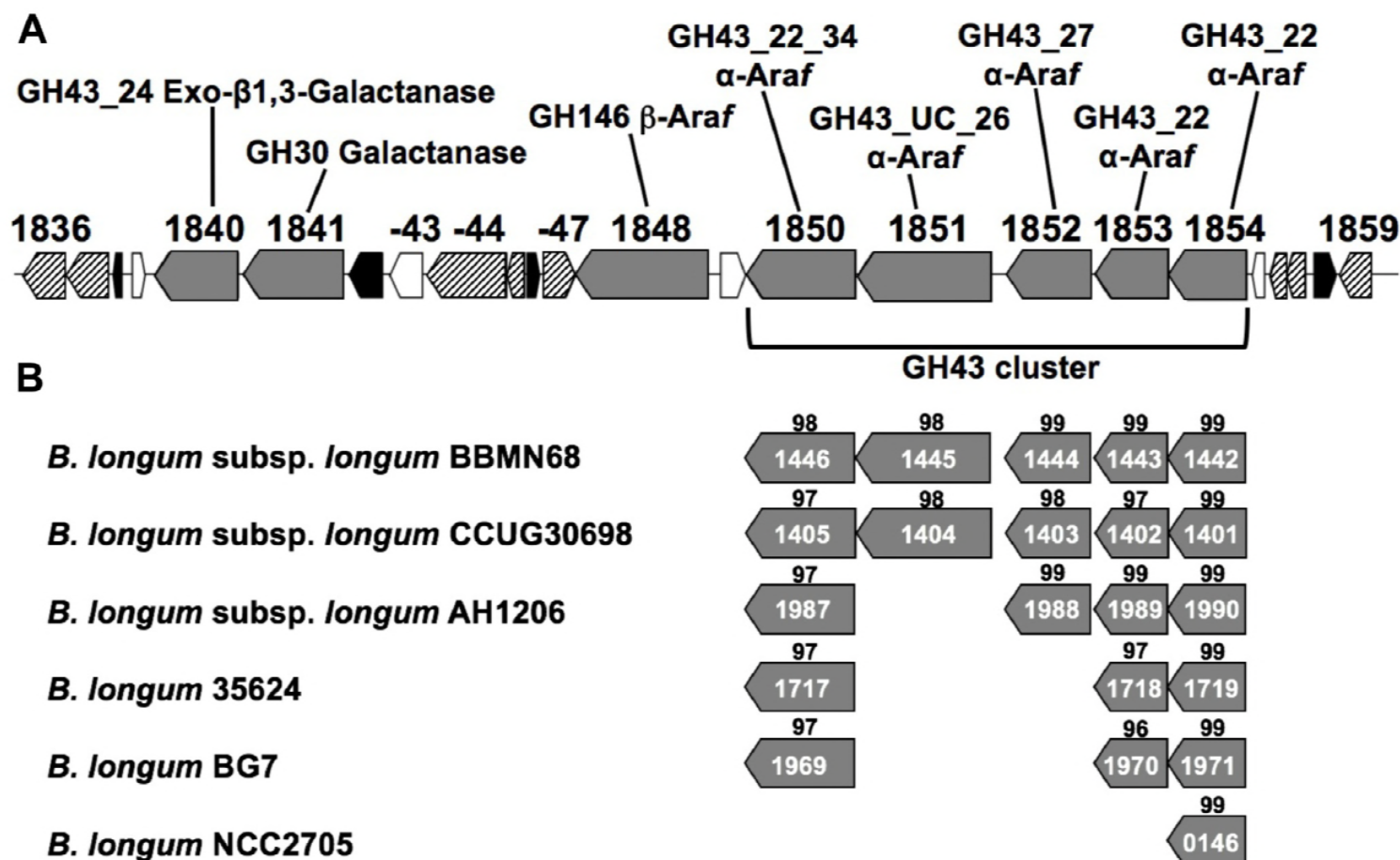Fig. 3. Komeno et. al.

Fig. 4.  Komeno et. al.

Fig. 5. Komeno et. al.