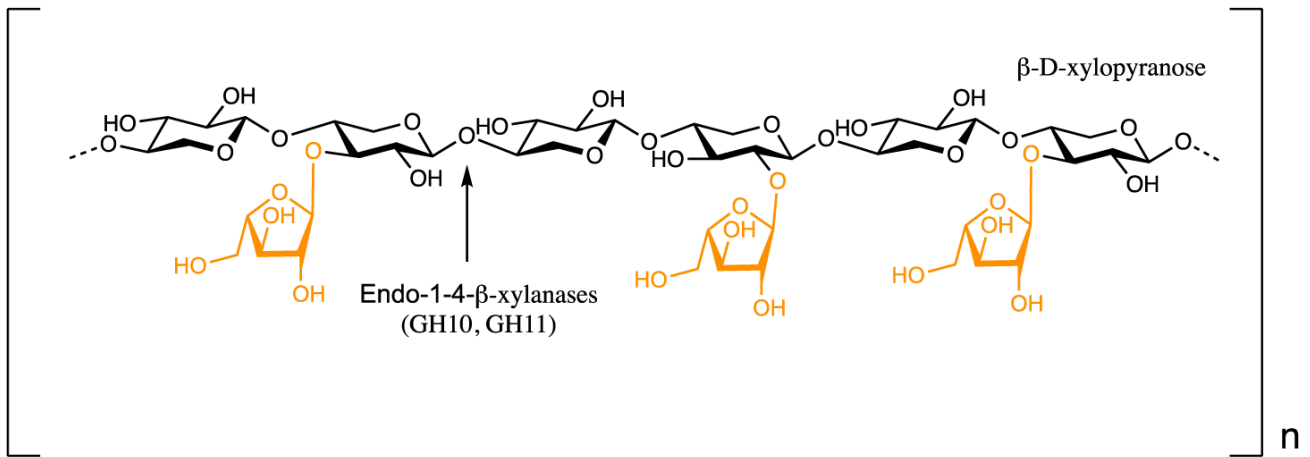


## Supplementary Information

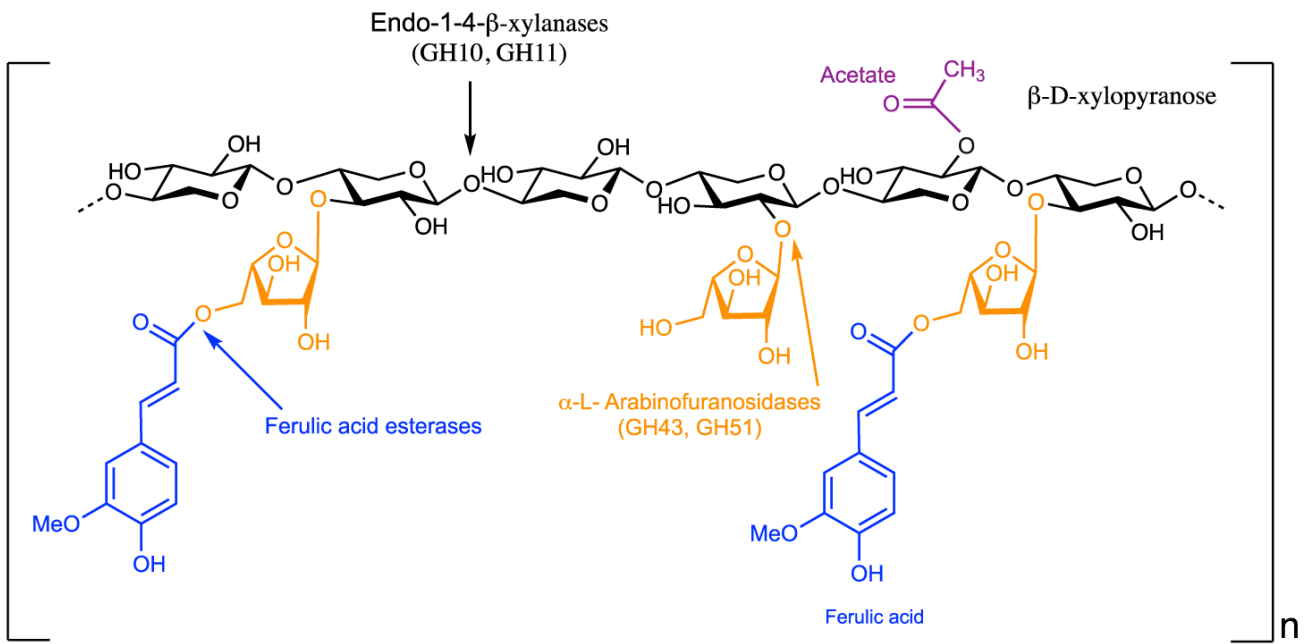
# **Degradation of complex arabinoxylans by human colonic Bacteroidetes**

Pereira *et al.*

# a Soluble wheat arabinoxylan

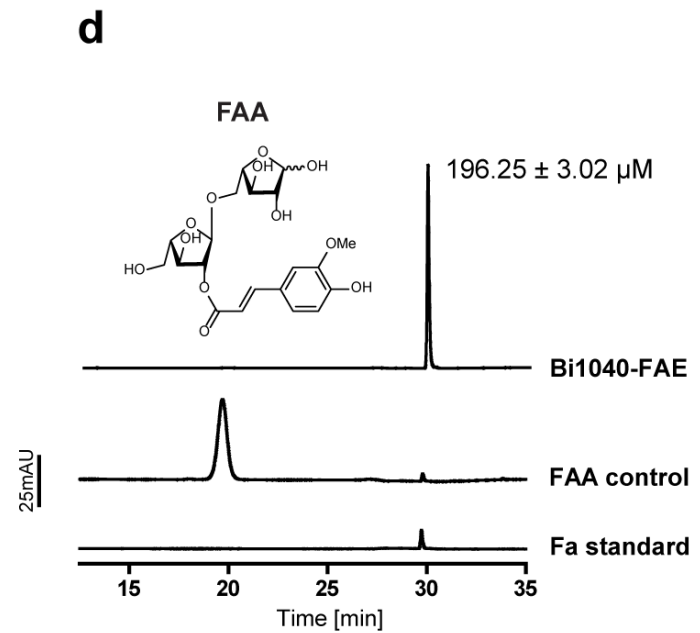
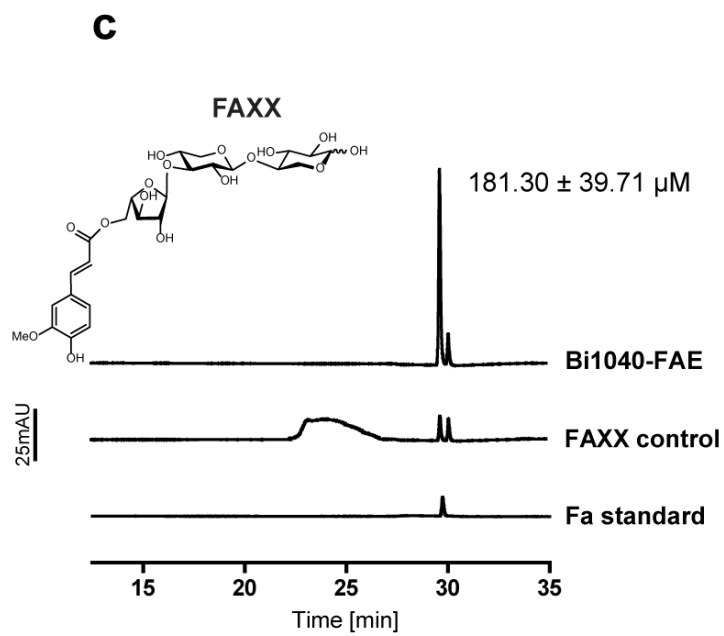
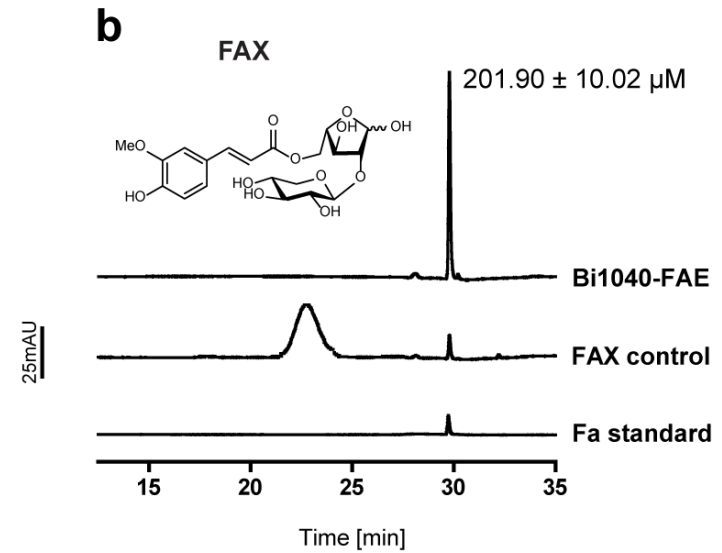
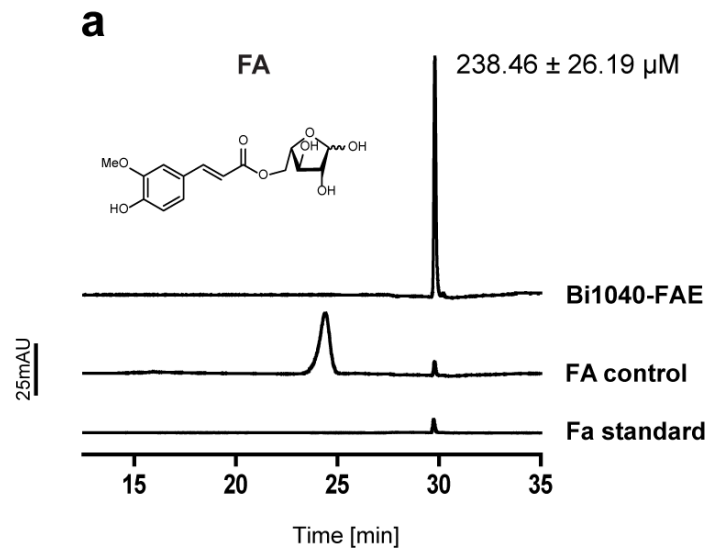


# b Insoluble wheat arabinoxylan

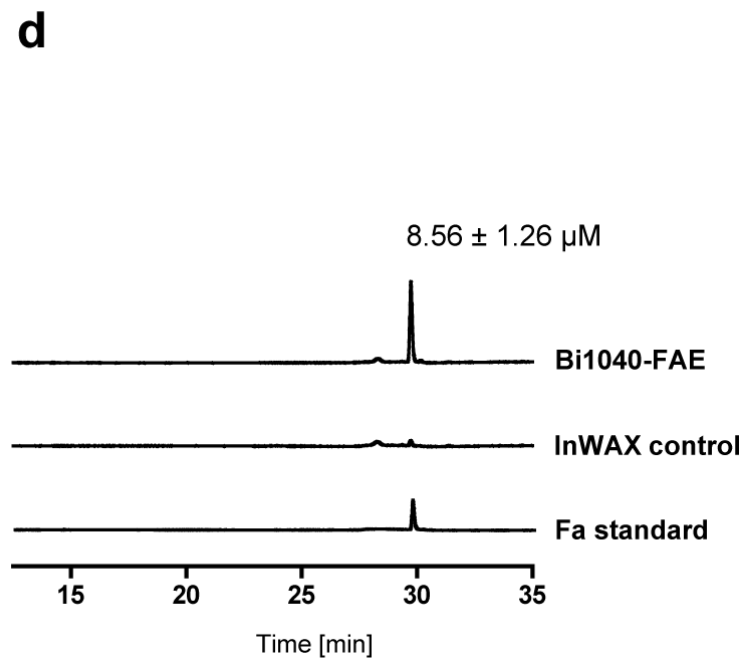
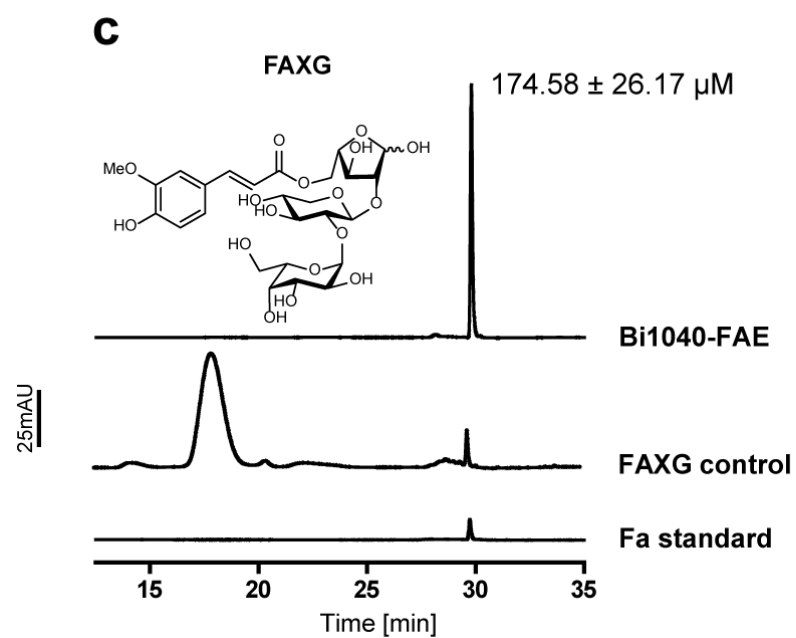
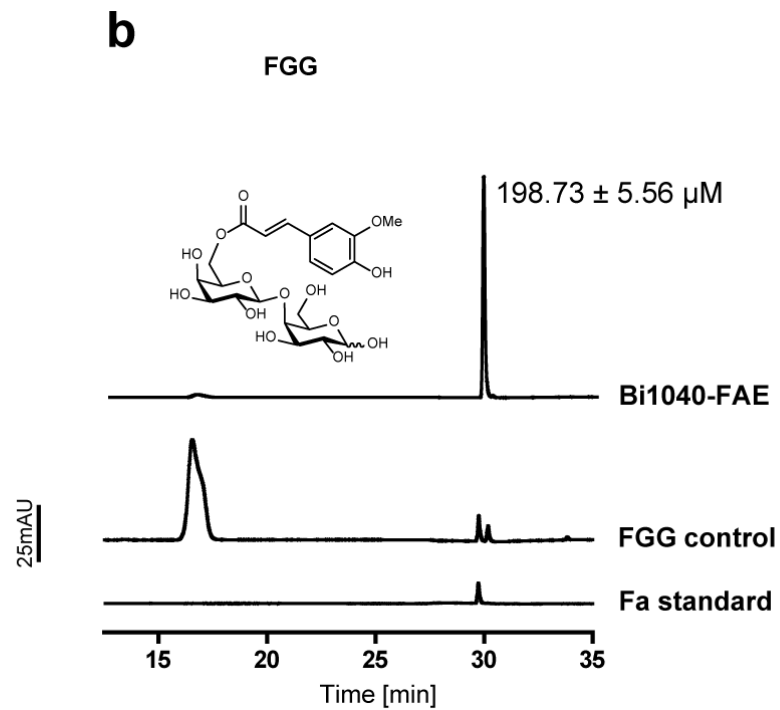
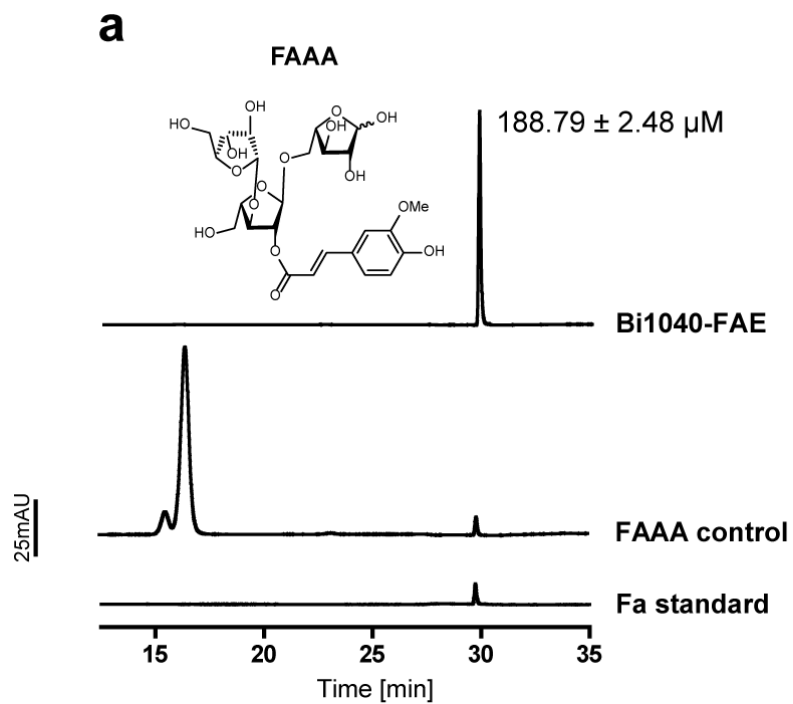


Other enzymes:  $\beta$ -xylosidase, acetylxylan esterase

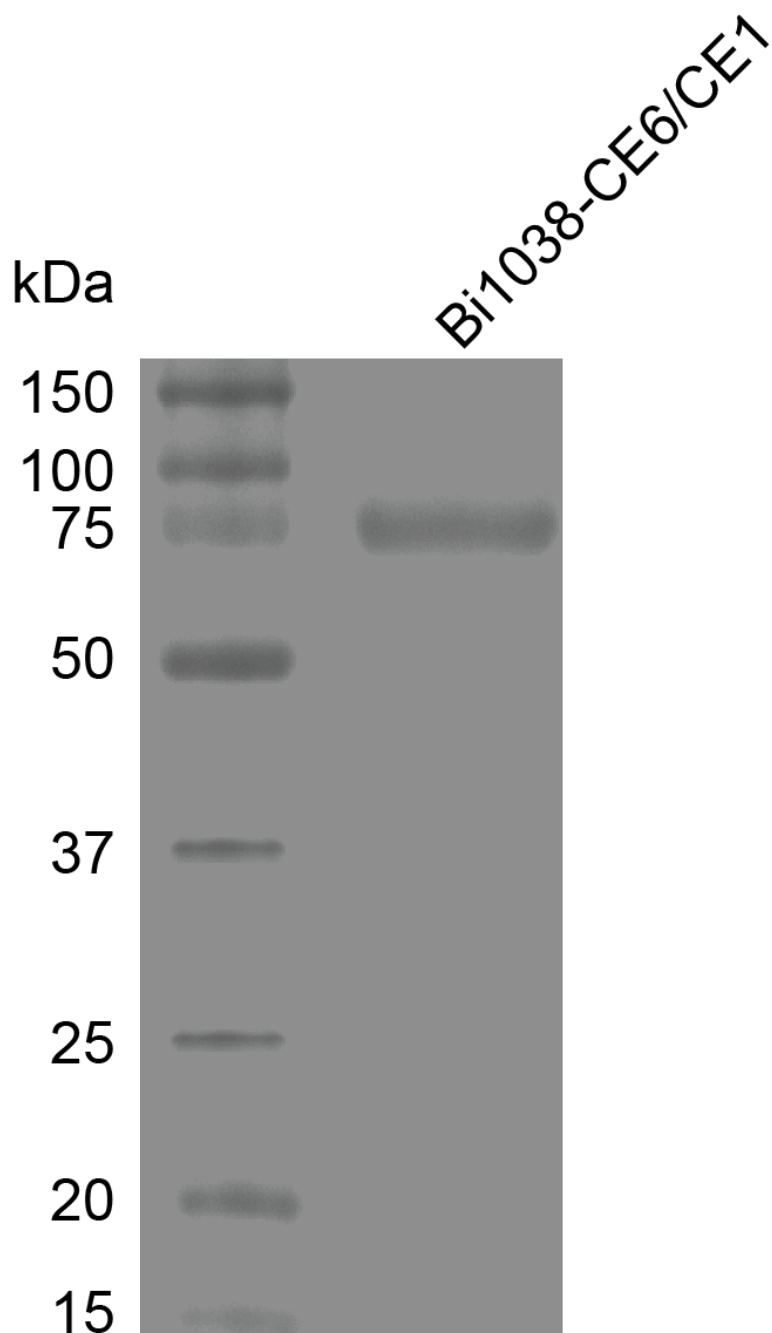
**Supplementary Fig. 1. Schematic showing the differences in soluble and insoluble wheat arabinoxylan.** **a** Schematic structure of the soluble arabinoxylan used in this study showing the  $\beta$ -1,4-linked xylose backbone with arabinose side chains, but absence of ferulic acid. **b** Schematic structure of the insoluble arabinoxylan used in this study showing a similar structure to soluble arabinoxylan but with ferulic acid side chains. Common enzymes known to cleave some of the linkages in arabinoxylan are indicated, including GH10, GH11 families of endoxylanases, GH43, GH51  $\alpha$ -arabinofuranosidases, Ferulic acid esterase, acetyl xylan esterase (catalyzes cleavage of backbone acetyl substituents) and  $\beta$ -xylosidase (cleaves xylo-oligosaccharides into xylose). The n indicates the polysaccharide length is long and variable.



**Supplementary Fig. 2. Cleavage of ferulic acid from feruloylated oligosaccharides by Bi1040-FAE.** The hydrolytic reactions contained 50 nM of enzyme (Bi1040-FAE) with 200  $\mu$ M of the respective feruloylated oligosaccharide (with structure shown). **a** Incubated with FA (ferulic acid linked to arabinose). **b** Incubated with FAX (ferulic acid linked to a disaccharide of arabinose and xylose). **c** Incubated with FAXX (ferulic acid linked to a trisaccharide of arabinose and two xylose residues). **d** Incubated with FAA (ferulic acid linked to arabinobiose). All reactions were carried out for 10 minutes at 37 °C in a buffer of pH 6.5. The end products were analyzed by C18-HPLC, as described in the Methods section. Fa: ferulic acid. The y-axis is absorbance in milli-absorbance units (mAU). The results are presented as the mean  $\pm$  the standard deviation of three independent reactions (n=3), and the traces are representatives of the three different reactions. The source data underlying Supplementary Fig. 2 are provided in the Source Data file.

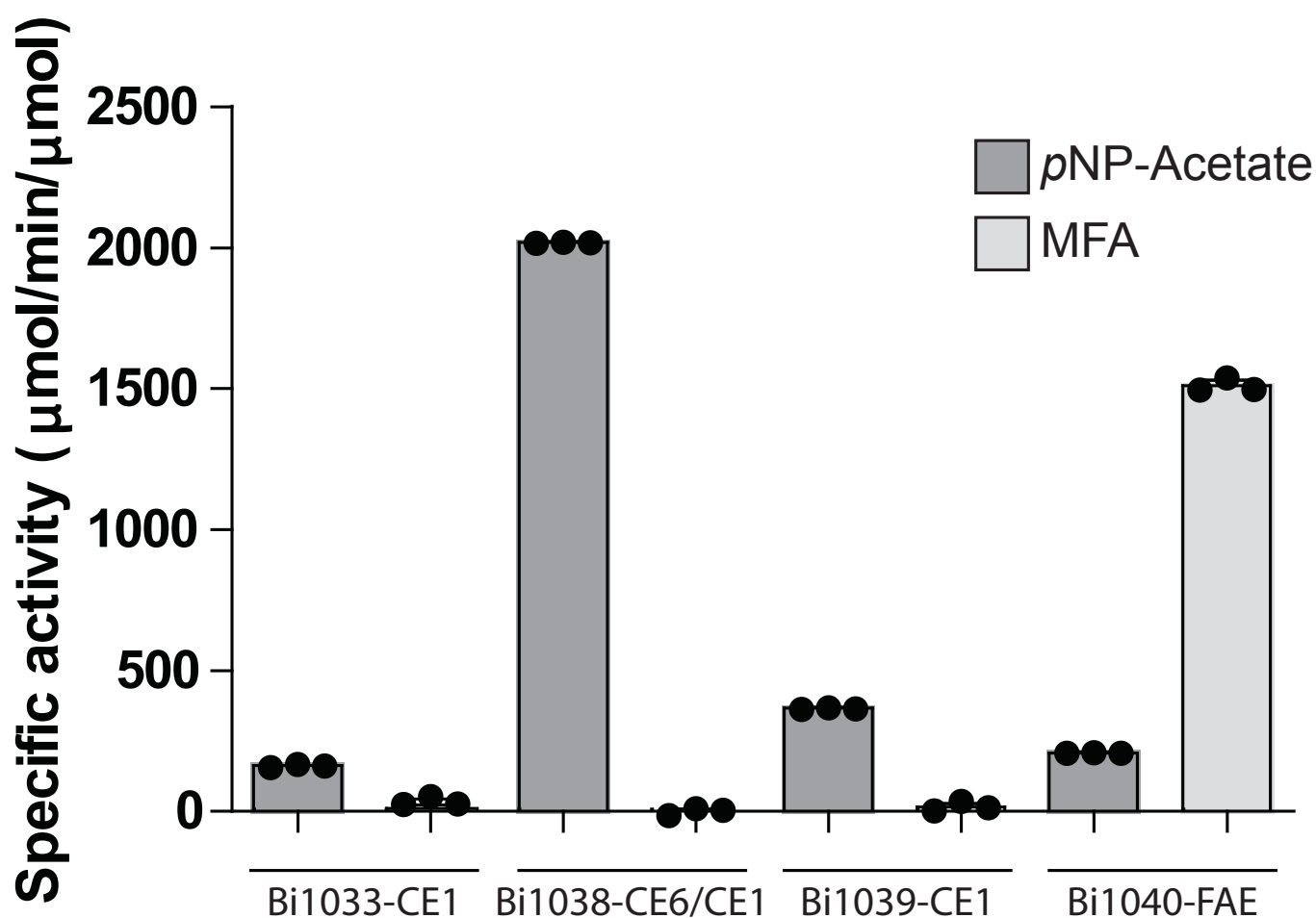


**Supplementary Fig 3. Cleavage of ferulic acid from feruloylated oligosaccharides and InWAX by Bi1040-FAE.** The hydrolytic reactions contained 50 nM of enzyme (Bi1040-FAE) with 200  $\mu$ M of the respective feruloylated oligosaccharide (with structure shown). **a** Incubated with FAAA (ferulic acid linked to arabinotriose). **b** Incubated with FGG (ferulic acid linked to cellobiose or disaccharide of glucose). **c** Incubated with FAXG (ferulic acid linked to a trisaccharide of arabinose, xylose and glucose). **d** Incubated with InWAX (insoluble wheat arabinoxylan). All reactions were carried out for 10 minutes at 37 °C in a buffer of pH 6.5. The end products were analyzed by C18-HPLC, as described in the Methods section. Fa: ferulic acid. The y-axis is absorbance in milli-absorbance units (mAU). The results are presented as the mean  $\pm$  the standard deviation of three independent reactions (n=3), and the traces are representatives of the three different reactions. The source data underlying Supplementary Fig. 3 are provided in the Source Data file.

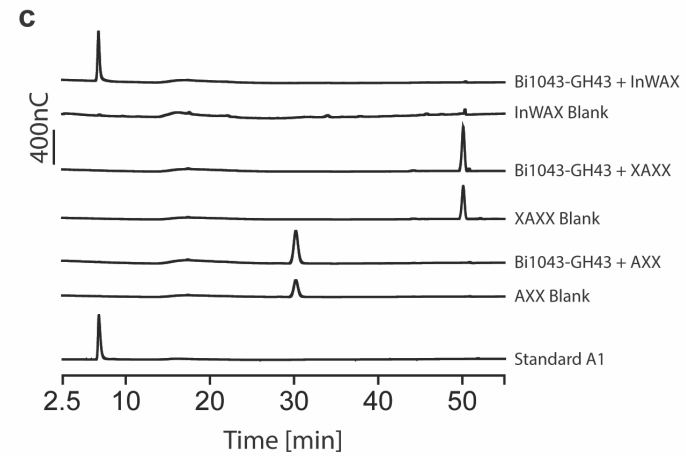
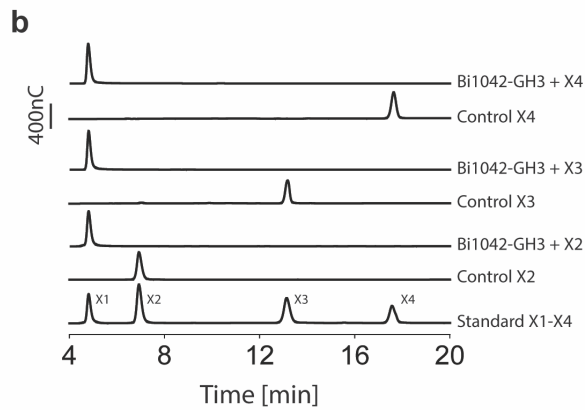
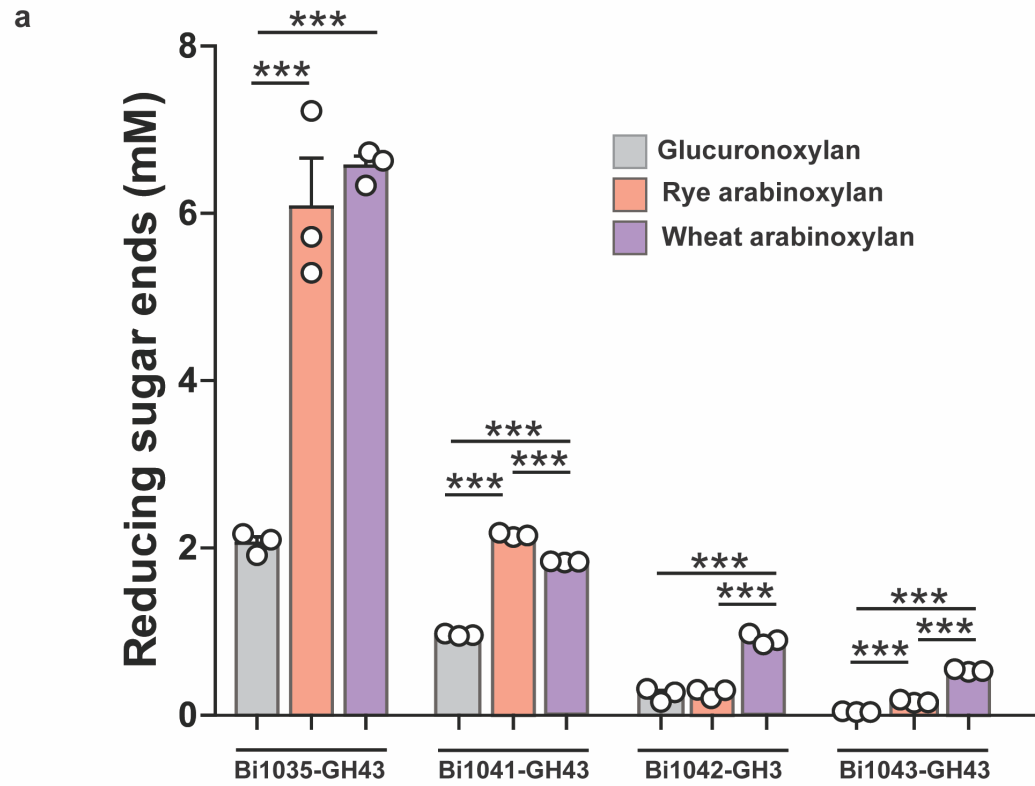


**Supplementary Fig. 4. The serine protease inhibitor, benzamidine HCl, inhibits cleavage of recombinant Bi1038-CE6/CE1 during purification from *E. coli* cells.** A 12% SDS-PAGE showing purified full length Bi1038-CE6/CE1, achievable by adding benzamidine HCl at a concentration of 1 mM to the buffers for re-suspension of recombinant *E. coli* cells and buffers for protein purification. The recombinant protein was resolved on SDS-PAGE during the purification steps to ensure that it migrates according to its predicted molecular mass, and finally a single SDS-PAGE was ran to obtain this image showing migration to the same position relative to the protein molecular mass markers.

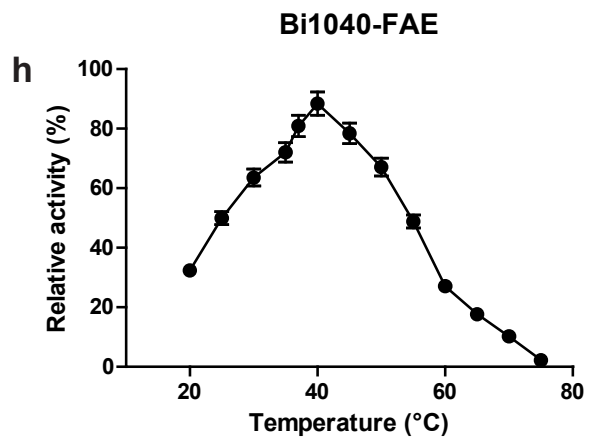
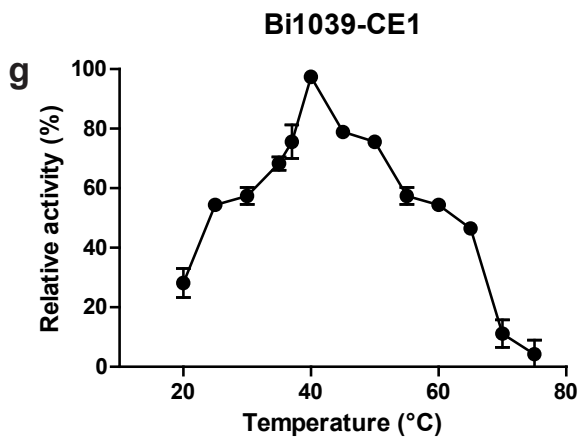
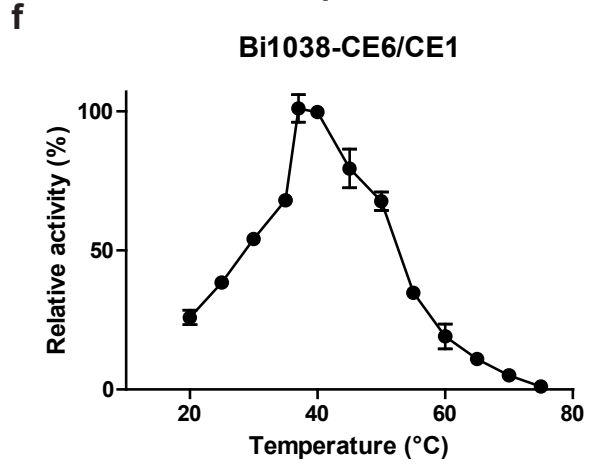
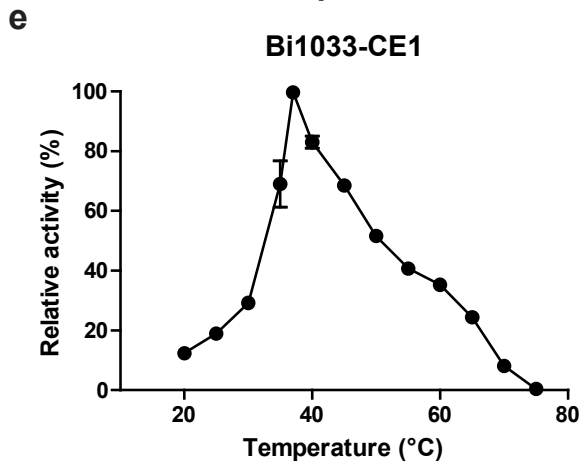
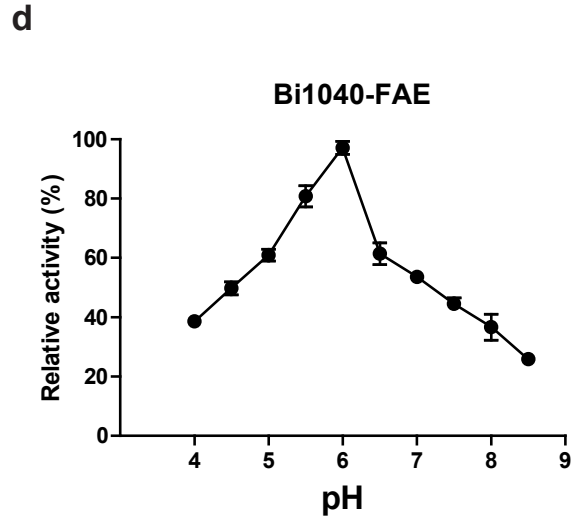
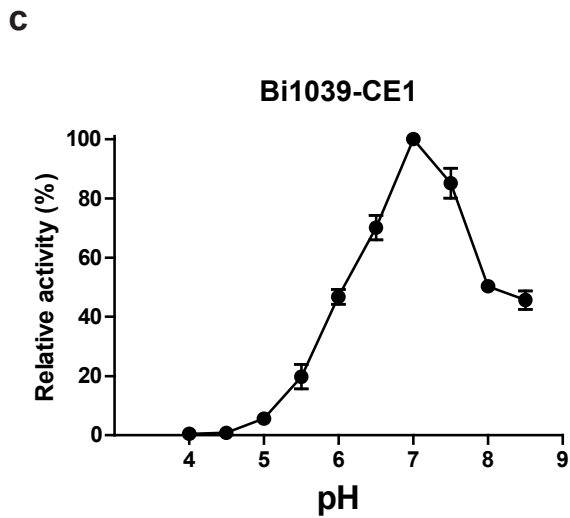
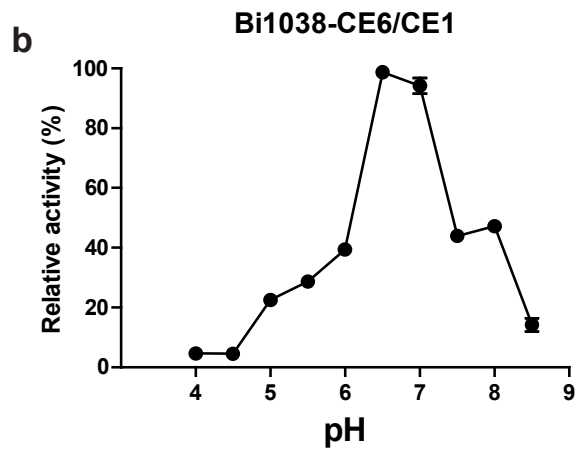
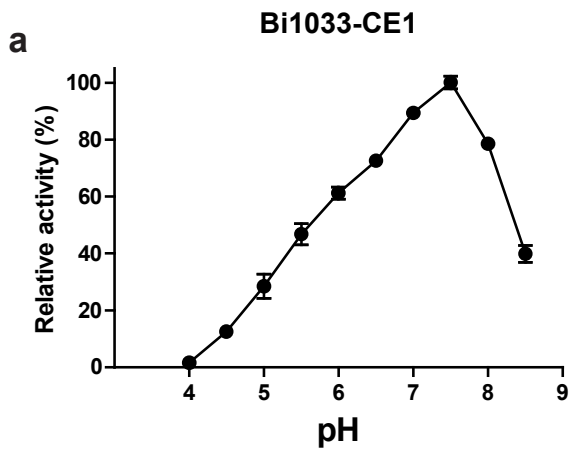




**Supplementary Fig. 5. Hydrolytic activities of the recombinant esterases from the *B. intestinalis* EGE PUL.** The two synthetic substrates methyl ferulate (MFA) and *para*-nitrophenyl acetate (*p*-NP Acetate) were used to analyze for ferulic acid esterase and acetyl xylan esterase activities, respectively, among the recombinant esterases expressed from the *B. intestinalis* EGE PUL. The reactions were carried out in a 0.1 M MOPS buffer at pH 7.5 and at a temperature of 37 °C. The results are the mean specific activity  $\pm$  standard deviation of three independent reactions ( $n = 3$ ). The source data underlying Supplementary Fig. 5 are provided in the Source Data file.



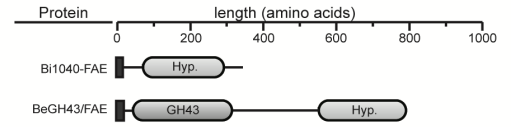
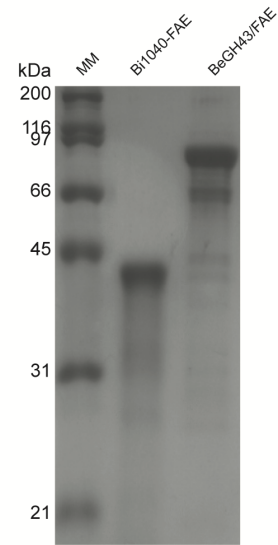
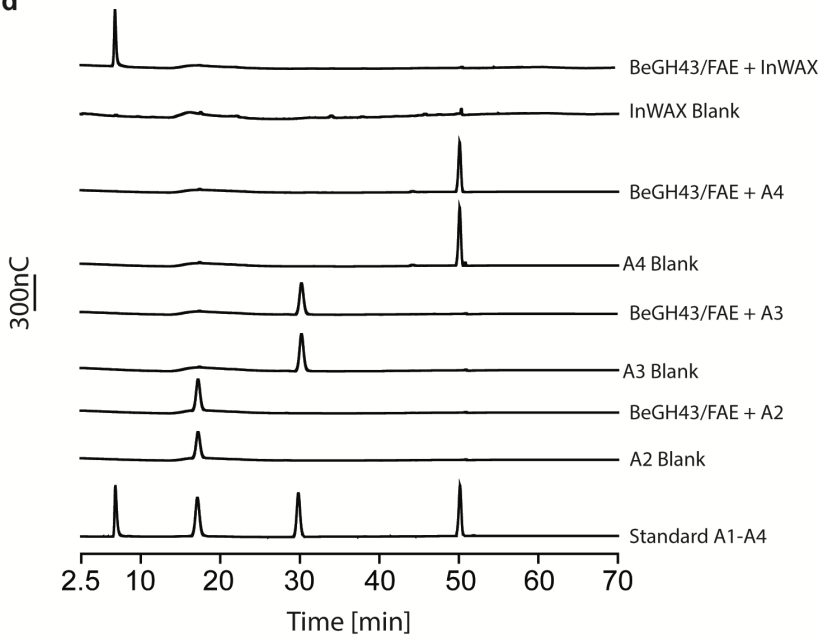
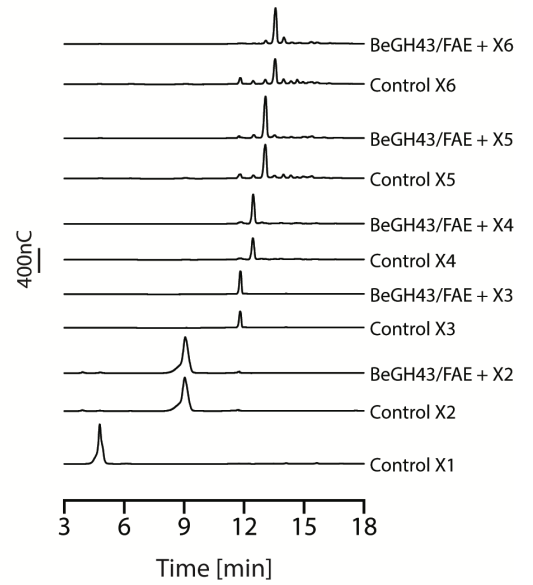
**Supplementary Fig 6. Hydrolytic activities of the glycoside hydrolases encoded in the EGE PUL of *B. intestinalis*.** **a** The reducing ends released from xylan polysaccharide substrates by the putative glycoside hydrolases in the *B. intestinalis* EGE PUL. **b**  $\beta$ -xylosidase activity of Bi1042-GH3 towards xylo-oligosaccharides. **c** Arabinoxylan polysaccharide-dependent  $\alpha$ -arabinofuranosidase activity of Bi1043-GH43. Rye arabinoxylan (Megazyme) has a backbone of xylose linked in  $\beta$ -1,4 glycosidic linkages and arabinose side chains at the O-2 or O-3 positions or both. The ratio of arabinose to xylose is 40:60. Glucuronoxylan (Sigma) has a backbone of xylose linked in  $\beta$ -1,4 glycosidic linkages and glucuronate side chains at the O-2 or O-3 positions or both, which are commonly methylated at position 4. Wheat arabinoxylan (Megazyme) has a backbone of xylose linked in  $\beta$ -1,4 glycosidic linkages and arabinose side chains at the O-2 or O-3 positions or both. The ratio of arabinose to xylose is 38:62. X1: xylose, X2: xylobiose, X3: xylotriose, X4: xylotetraose, AXX: a trisaccharide of arabinose and two xylose residues, XAXX: a tetra-saccharide of xylose, arabinose and two xylose residues, and InWAX: insoluble wheat arabinoxylan (see Supplementary Fig. 1b). In **a**, the bars indicate means  $\pm$  standard deviation of three independent reactions (n=3). Statistical analysis was performed using one-way Analysis of Variance (ANOVA) with multiple comparisons using Tukey's test, \*p<0.033, \*\*p<0.002, \*\*\*p<0.001. The experiments reported in **b** and **c** were carried out in single reactions. The source data underlying Supplementary Fig. 6a are provided in the Source Data file.



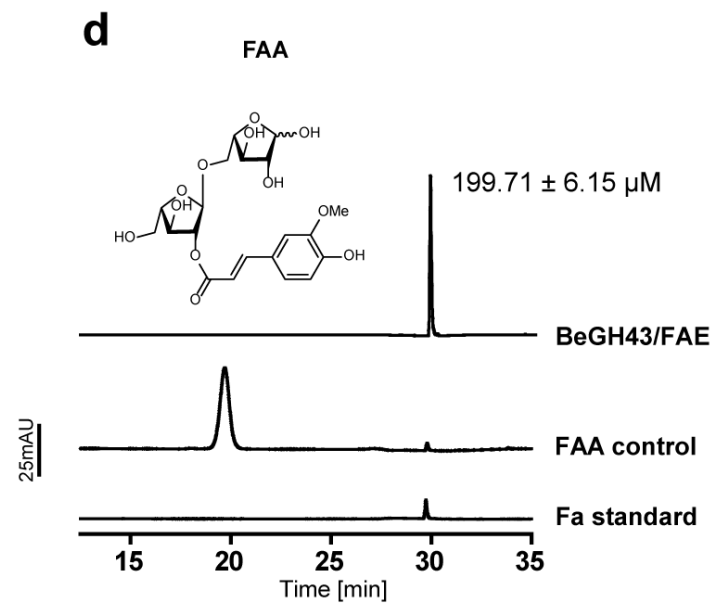
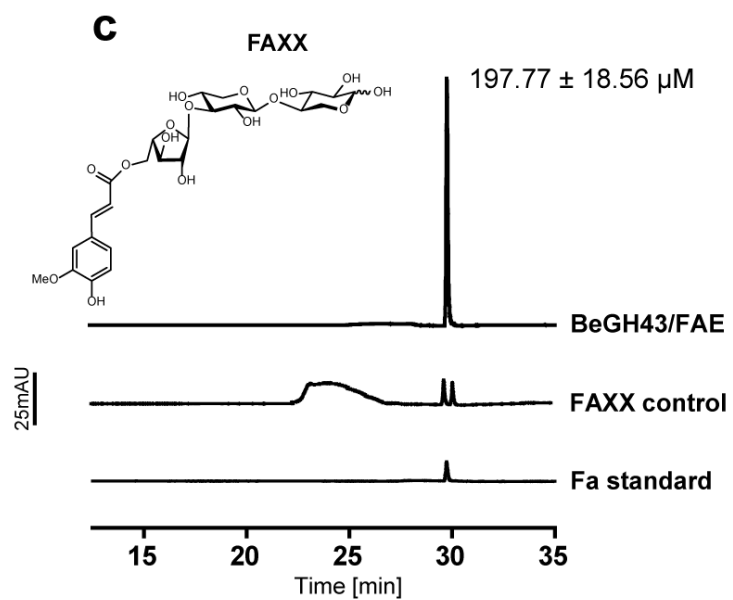
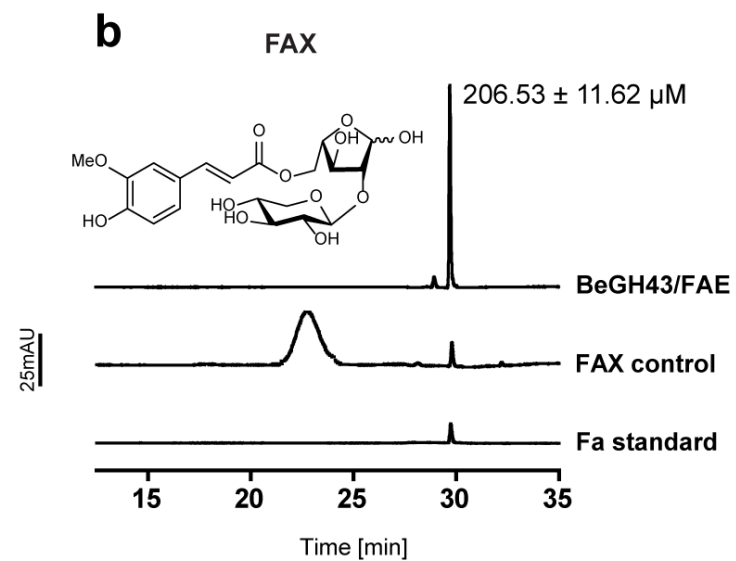
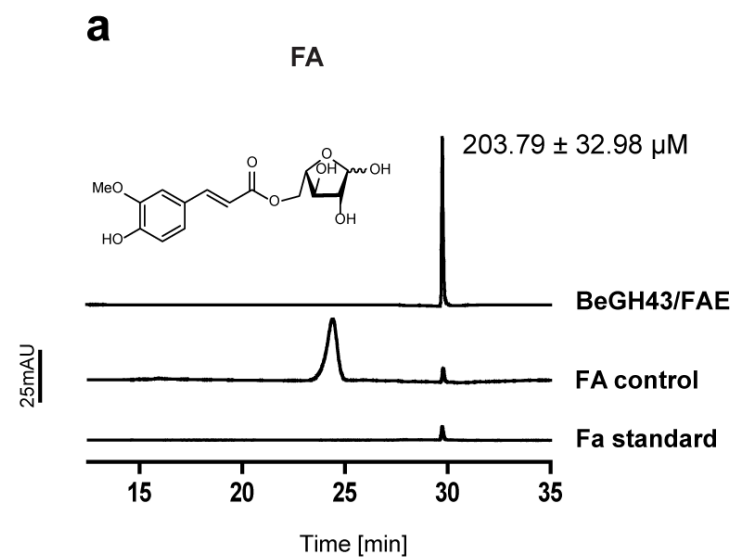
**Supplementary Fig. 7. Determination of the pH and temperature optima of the recombinant esterases from the *B. intestinalis* EGE PUL.** The enzymatic activity of each esterase was determined at different pH at 37 °C using two different buffers and *p*NP-acetate as substrate. The relative activities at different pH were then calculated as the percent of the highest obtained activity (**a**, **b**, **c**, and **d**). The optimal temperatures were then determined by incubating each enzyme at its respective optimal pH at temperatures ranging from 20 °C to 75 °C, and the relative activity was calculated relation to the highest activity obtained (**e**, **f**, **g**, and **h**). The graphs were plotted from the means  $\pm$  standard deviation of three independent reactions (n = 3). The source data underlying Supplementary Figs. 7a-h are provided in the Source Data file.

**a**

BeGH43/FAE	MKKQKTLTLLCNLFCCILLPLSAQKPATNPVIYADAPDMSMLRVGDTYYMSSTTMHMSPG	60
Bi1040-FAE	-----	0
BeGH43/FAE	VPIMKSNLDLVNWKLVNAYDTLANIPTMNLDDGKNTYGRGSWASCLRYHEGVVYLSTFAQ	120
Bi1040-FAE	-----	0
BeGH43/FAE	TTGKTYFYTTKLEKGPWKCTEFSPAYHDHSFFDEDDGHYIMYVGNKFLAEELKPDLSG	180
Bi1040-FAE	-----	0
BeGH43/FAE	VKPGTERTVLIENASAPAGDNIMLGAEGSQLFKVNGKYLLFNITWPRGGVRTVIIVHRADKI	240
Bi1040-FAE	-----	0
BeGH43/FAE	TGPYEGRVVFQDRGIAQGGVLDTPDGRWFAYLFEDCGAVGRIPYLVVVEWKDGVVFLGVN	300
Bi1040-FAE	-----	0
BeGH43/FAE	GRAPAKLELPDSRGLIPGIVASDDFNRRKGERALPLVWQWNHNPDNALWLSARKGYLRL	360
Bi1040-FAE	-----	0
BeGH43/FAE	TTGRMETSFTQAKNLTQRTIGPVCTGVSVMDSVGMKEGDFAGLSLFQRKYQVGVKVTD	420
Bi1040-FAE	-----	0
BeGH43/FAE	GKKYIVMVNGENETPAEVEKVLNQVVFKAACDFRNKDKGYFYSLDGSNWKATGNV	480
Bi1040-FAE	-----MSLRFLNCLLFLNLRN-----	19
BeGH43/FAE	LKMQYTPPHFMGYRFALFNATKEVGGADFDIKIEDKISDCRMBDICYADKLEGHKL	540
Bi1040-FAE	-----LTFGVICALFVSPMSFGQITQWTDINYNANSLLEGHKL-----	58
BeGH43/FAE	DIYLPDMDLPSYKVVVLIYGSAWFANNMKQAAFOVFGKSLLDKGFVAVVSINHRSSGDAKF	600
Bi1040-FAE	DIYLPDGGTEYKVVVLIYGSAWFANNMKQAFQAMGKPLLDGFAVAVVSINHRSSGDAKF	118
BeGH43/FAE	PAQINDVKAARFIRANAAKYRLDTSFIGITGFSSGGHSLASLAGTTNGVKSPTCAKTD	660
Bi1040-FAE	PAQINDVKAARFIRAHADRYRLDTSFIGITGFSSGGHSLASLAGTTNGVKSPTCAKTD	178
BeGH43/FAE	TEGNVGLYPSFSSRVDAVVWFGPIDMTRMENCNTTKGANSPPEALIGGVPAADNDMLAL	720
Bi1040-FAE	TEGNVGDCTSFSSRVDAVVWFGPIDMTRMENCNTTKGANSPPEALIGGTPAHDNDMLAL	238
BeGH43/FAE	LNPITYIDKNDPKFVIHGGADTVVPCQSTFFSALRAQGRLEEFISVPGGQHGPNFTFN	780
Bi1040-FAE	LNPITYIDEKDPKFIHGGADTVVPCQSVFFKQTLSPKGRLEEFISVPGGQHGPNFTFN	298
BeGH43/FAE	ENTLKKMIDFFAREAGTYRDLNR-----RPIKCDIK-----	813
Bi1040-FAE	ENTLKKMIDFFRKAAMDLCPANITLVPHIRPGSKKPPQKQFYTYL	345

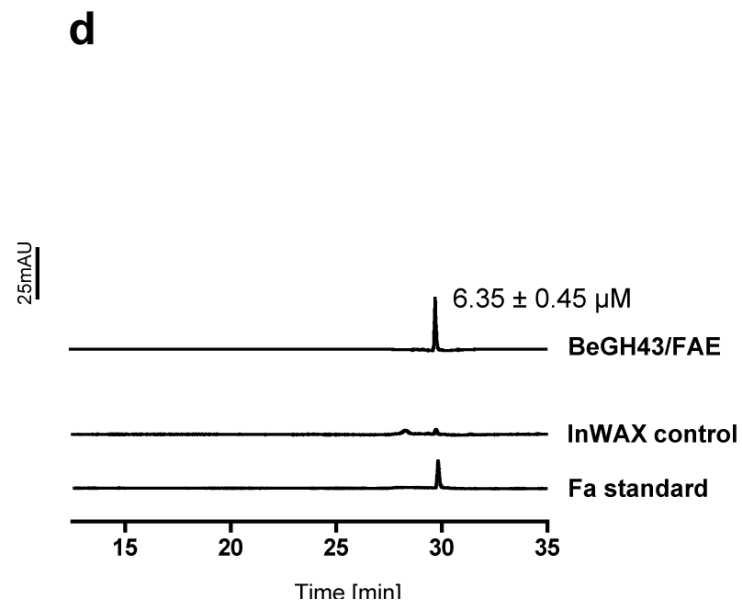
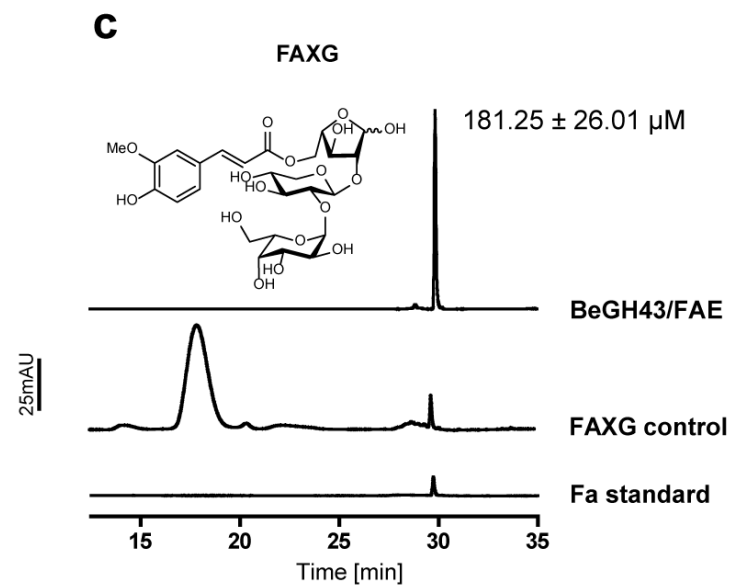
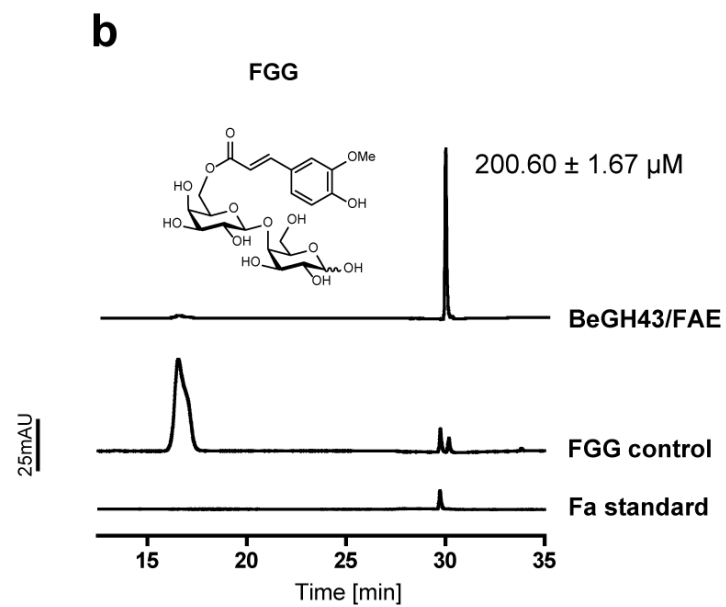
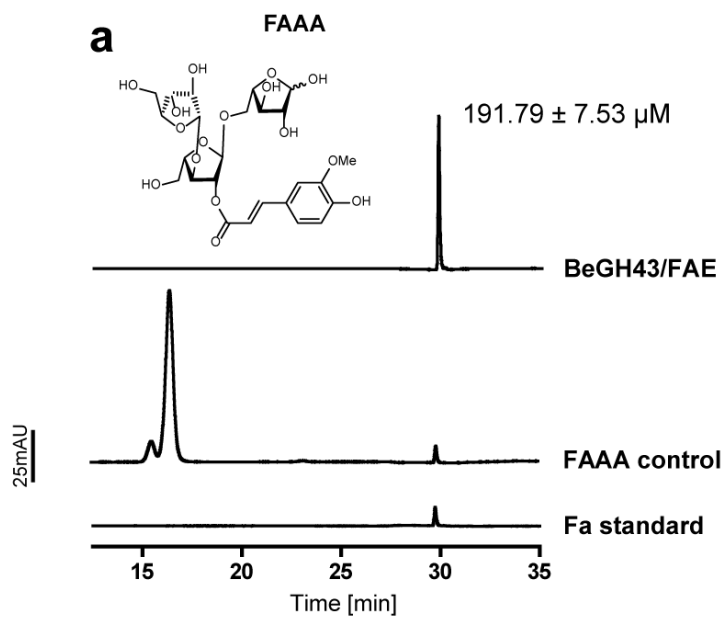
**b****c****d****e**

**Supplementary Fig. 8. Comparison of Bi1040-FAE and BeGH43/FAE demonstrate high conservation of the FAE module.** **a** The polypeptide sequence of the hypothetical protein (Bi1040), demonstrated as a ferulic acid esterase (FAE), aligned with BeGH43/Hyp showed high conservation of the C-terminal region of BeGH43/Hyp with Bi1040-FAE leading to designation as BeGH43/FAE. **b** A schematic showing the modular architectures of Bi1040-FAE and BeGH43/FAE. **c** A 12% SDS-PAGE analysis showing purified recombinant Bi1040-FAE and BeGH43/FAE. **d** HPLC chromatograms showing that BeGH43/FAE cleaves arabinose off insoluble arabinoxylan (InWAX) and not from arabino-oligosaccharides (A2, A3 or A4). **e** HPLC chromatograms showing that BeGH43/FAE does not hydrolyze xylo-oligosaccharides (X2, X3, X4, X5, and X6). The experiments in **d** and **e** were from single reactions. InWAX: insoluble wheat arabinoxylan; A2, A3, A4 are straight chain arabinobiose, arabinotriose and arabinotetraose, respectively; X2, X3, X4, X5, X6, xylobiose, xylotriase, xylo-tetraose, xylo-pentaose, and xylo-hexaose, respectively.

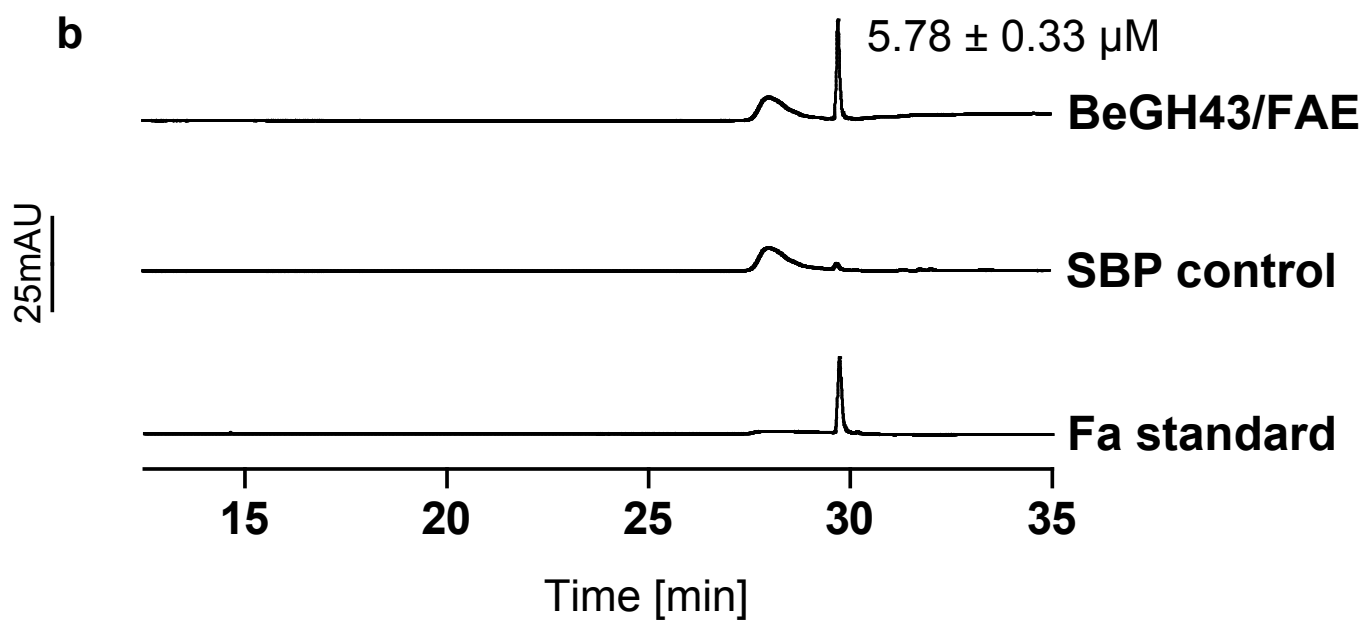
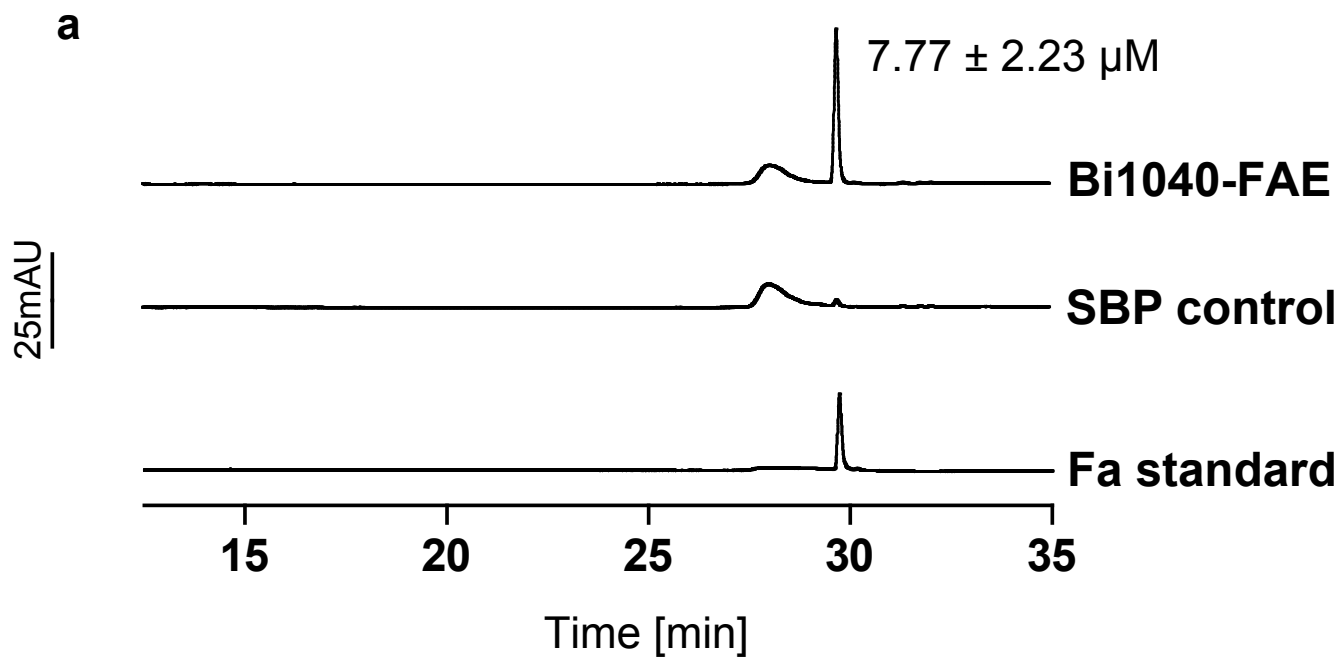




**Supplementary Fig. 9. Cleavage of ferulic acid from feruloylated oligosaccharides by BeGH43/FAE.** The hydrolytic reactions contained 50 nM of enzyme (BeGH43/FAE) with 200  $\mu$ M of the respective feruloylated oligosaccharide (with structure shown). **a** Incubated with FA (ferulic acid linked to arabinose). **b** Incubated with FAX (ferulic acid linked to a disaccharide of arabinose and xylose). **c** Incubated with FAXX (ferulic acid linked to a trisaccharide of arabinose and two xylose residues). **d** Incubated with FAA (ferulic acid linked to arabinobiose). All reactions were carried out for 10 minutes at 37 °C in a buffer of pH 6.5. The end products were analyzed by C18-HPLC, as described in the Methods section. Fa: ferulic acid. The y-axis is absorbance in milli-absorbance units (mAU). The results are presented as the mean  $\pm$  the standard deviation of three independent reactions (n=3), and the traces are representatives of the three different reactions. The source data underlying Supplementary Fig. 9 are provided in the Source Data file.

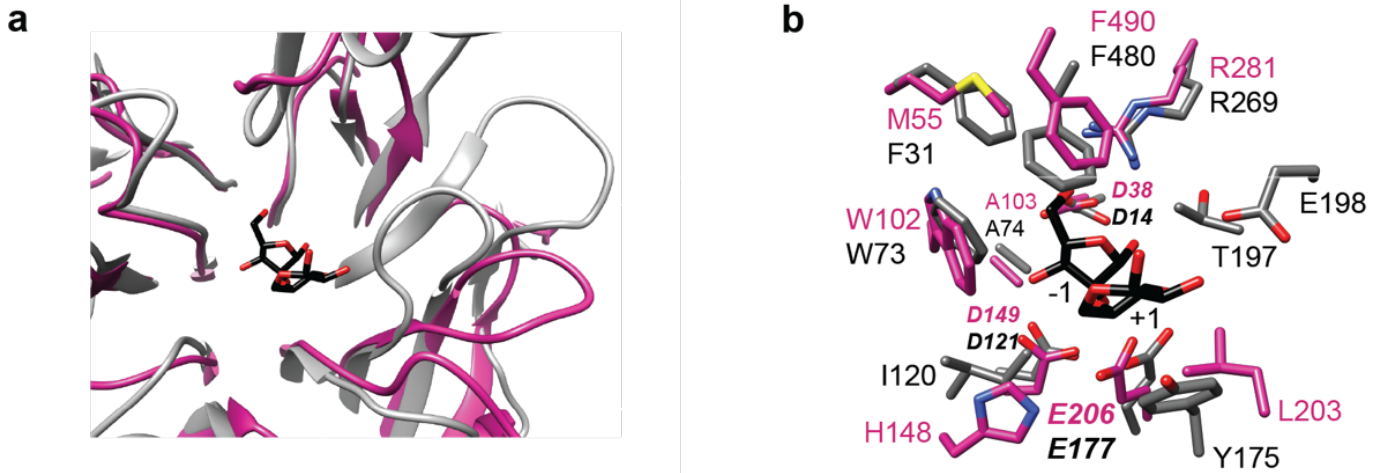


**Supplementary Fig. 10. Cleavage of ferulic acid from feruloylated oligosaccharides and InWAX by BeGH43/FAE.** The hydrolytic reactions contained 50 nM of enzyme (BeGH43/FAE) with 200  $\mu$ M of the respective feruloylated oligosaccharide (with structure shown). **a** Incubated with FAAA (ferulic acid linked to arabinotriose). **b** Incubated with FGG (ferulic acid linked to cellobiose or disaccharide of glucose). **c** Incubated with FAXG (ferulic acid linked to a trisaccharide of arabinose, xylose and glucose). **d** Incubated with InWAX (insoluble wheat arabinoxylan). All reactions were carried out for 10 minutes at 37 °C in a buffer of pH 6.5. The end products were analyzed by C18-HPLC, as described in the Methods section. Fa: ferulic acid. The y-axis is absorbance in milli-absorbance units (mAU). The results are presented as the mean  $\pm$  the standard deviation of three independent reactions (n=3), and the traces are representatives of the three different reactions. The source data underlying Supplementary Fig. 10 are provided in the Source Data file.



**Supplementary Fig. 11. Cleavage of ferulic acid from a natural substrate (sugar beet pulp).**

In substrates such as wheat bran, the ferulic acid is esterified to the C-5 position of arabinose in the arabinoxylyan, a  $\beta$ -1,4-D-xylan to which  $\alpha$ -L-arabinofuranosyl residues attached at position 2 or 3. By contrast, the feruloyl groups in sugar beet pulp are linked to the arabinofuranosyl residues of the main core of  $\alpha$ -1,5 linked arabinan chains and to the galactopyranosyl residues of the main core of the  $\beta$ -1,4-linked type I galactan chains<sup>1</sup>. The ferulic acid linkages in sugar beet pulp is therefore more complex than in wheat bran or wheat arabinoxylyan. Here, we assessed the versatility of the new ferulic acid esterase for the capacity to release ferulic acid from the more complex substrate sugar beet pulp. The reaction was carried out by incubating 50 nM of Bi1040-FAE or BeGH43/FAE with 0.5% of a naturally occurring substrate, sugar beet pulp, for 2 hours at 37 °C in a buffer of pH 6.5. The end products of hydrolysis were analyzed by C18-HPLC. **a** ferulic acid release by Bi1040-FAE. **b** Ferulic acid release by BeGH43/FAE. Fa: ferulic acid. SBP: sugar beet pulp. The y-axis is absorbance in milli-absorbance units (mAU). The results are presented as the mean  $\pm$  the standard deviation of three independent reactions (n=3), and the traces are representatives of the three different reactions. The source data underlying Supplementary Fig. 11 are provided in the Source Data file.



**Supplementary Fig. 12. Overlay of the dual function BeGH43/FAE N-terminal domain with the  $\beta$ 1,4-xylosidase (GH43) from *Geobacillus thermoleovorans*.** **a** Close up view of the overlay of the N-terminus of the BeGH43/FAE bifunctional enzyme (pink ribbon) with the structure of the  $\beta$ 1,4-xylosidase (GH43) from *Geobacillus thermoleovorans* Xyl enzyme (grey, PDB 5z5i, RMSD= 2.1Å for 454 residues) with bound arabinose and xylose. **b** Close up view of the residues coordinating arabinose (-1 subsite) and xylose (+1 subsite) in the *G. thermoleovorans* Xyl structure (grey and black). Residues within 3.5Å of the substrate are displayed for both structures.

a

SEH83175.1	--MKTILSVMCC--LSAMAMAQPAGGFGG-FQAPQVKLETSSQEWKDVNYAGDDQAYHTC	56
WP_006281679.1	--MKKIALEMLLLLAGVTAQAQMPAGGFGG-FQMPQVKLETSSQEWKDVNYAGDNKTYHTC	57
WP_100993215.1	--KNHGVFIAIILGL--LAQRVLA-----QPPDYASRAIQSSKSWLDLYVVDGIIGHKL	59
WP_020527766.1	--KTSWIIITFVGLILPITLSAQH-----LQNDYKSSSEIISLKHWDLDYVVDGRHIGHRL	61
WP_099150634.1	--MRDLVLLL--FCL-ALGTGLS-----AQDQYQRMAIESSRAWLDIYVVDGRHIGHRL	49
WP_013768612.1	--MKKLAFAAFLASINLI GQAQT-----SPMDYNKTAIQSSKYWLDVYVVDGKIVGHRL	52
WP_114463085.1	--MKTLT--FFLT F ISFLAMAQP-----KEINA EKTAIVSSKHWLDIYVVDGMMGHRL	50
WP_015029998.1	--MKK-TFILLFFCIKIFAQPKP-----DFAKSAIEASKSWLDIYVVDGIIGHKL	48
WP_055151666.1	--MKL-IYTLPLFLLLLFMAKAQN-----TPIPGPVESKAWLDLDYVVDGNHIGHKL	48
WP_006799192.1	-----MKN-----YILIVLFFLSLNCFGQEYSQSWKDVSYAGDSAIIYHRL	40
WP_027452672.1	--MKKLILSFMLLAMTIADYAQ-APG-----TMPFAMPQTNPDFKDVNYAGDTLEAHCL	51
KQB43445.1	-----MKK-----ITLILIFLCFNSFAQNANKKWLVDVNYANDAQVYHNL	40
WP_071144936.1	-----MR-----KLLL--I LMLIPAVCSFGQEABWLDIYVVDGTEGHKL	38
WP_100615198.1	-----MKK----NL-----LILILSALCFNSGFTQHAGKQWLDLNYANDEHTEHKL	42
WP_123396545.1	-----MKK----QLATLL----LLI---GA-MTIQAQSSAESQWLDIYVVDNTEGHKL	42
WP_117741097.1	LKMQYTMPHFMGYRFALFNATKETE GYVDFDYFKIEDKISDCRWADV CYADDELEGHKL	540
<b>WP_050793236.1</b>	LKMQYTMPHFMGYRFALFNATKEVGGYADF DYFKIEDKISDCRWEDICYADDKLEGHKL	540
WP_024996568.1	-----I-----LIF---AALFICAGSF-AQTQWTVNSYADDDLEGHKL	38
WP_018709923.1	-----MMAALASFGQTRPTDVTYANDTLVGHKL	29
<b>EDV05955.1</b>	-----K-----LVFGFV CALFYSFMSFGQITQWTDIN YANDSLEGHKL	58
WP_007216415.1	-----K-----LVFGFV CALFYSFMSFGQITQWTDIN YANDSLEGHKL	39

SEH83175.1	DIMLPKKEQASYPVVVLIYGSAWFSNNSKGMADLGTIVKSLLDAGFAVVCNHRSSMDAK	116
WP_006281679.1	DIMLPKPKQKNYPVVVLIYGSAWFSNNSKGMADLGTIVKSLLDAGYAVVCNHRSSSDAS	117
WP_100993215.1	DIHLPLSGKAPFPVLIICYGSAWLANNAKGAFTDGLGQRLIKEGFVAVAINHRSSGDTL	119
WP_020527766.1	DIHLPDKESDKYPVVVLSIYGSAWFSNNSKGAFAVIGQALIKAGYAVVTINHRASDAK	121
WP_099150634.1	DIHLPAEGEGPEPPIVLSIYGSAWFSNNSKATTFSTGLGQKLLAGGFVAVSINHRASDAL	109
WP_013768612.1	DIHLPKKGRGYPYPIIVAIYGSAWFSNASKANIFQEGLGQALLINGFAVVSINHRSSDAK	112
WP_114463085.1	DIHLPSVGKAPFPVIVVAIYGSAWFSNAAKGTVEFDGLGQTLINGFAVVSINHRSSDAK	110
WP_015029998.1	DIHLPNEGKGPFPVVVLIYGSAWFSNNTSKATCFNDGLGQTLIKNGFAVVSINHRSSDAI	108
WP_055151666.1	DIHLPKKEGKGPFPVIVVLIYGSAWFSNNSKSGQLNDEGQTLIRNGYAVVSINHRSSDAI	108
WP_006799192.1	DIMLPKVERNNYPVVVLIYGSAWFSNNGKADMT-TIGKALLDAGFAVVTINHRSSDAK	99
WP_027452672.1	DIMLPKKEGKAPYKVVIVAIYGSAWFANMMKPFAM-SLGKALTDAGYAVVSINHRSSDAK	110
KQB43445.1	DIHLPKAEKTSYKVVIVVLIYGSAWFANMMKMAFQ-SMGKPLLDAGFAVVSINHRSSDAK	99
WP_071144936.1	DIHLPPTGQSSYKAVVLIYGSAWFANMMKQIAFQ-AMGKTLIDGFAVVSINHRASMDAR	97
WP_100615198.1	DIHLPSEVKPKYKAVVLIYGSAWFANMMKQMAFQ-ALGKPLLDGFAVVSINHRSSGDAM	101
WP_123396545.1	DIMLPKSVKSDSYPAVVVLIYGSAWFANNAKMAFD-SMGKQLLDAGFAVVSINHRASGAK	101
WP_117741097.1	DIMLPDTEGKSSHKVVVLIYGSAWFANMMKQNAFQ-VFGRSLLDKGFVAVSINHRSSDAK	599
<b>WP_050793236.1</b>	DIMLPDMDEPSYKVVVLIYGSAWFANMMKQAAAFQ-VFGKSLLDKGFVAVSINHRSSDAK	599
WP_024996568.1	DIMLPDINQPSYKAVVLIYGSAWFANMMKQMAFQ-AMGKPLIDGFAVVSINHRSSDAK	97
WP_018709923.1	DIMLPDGNREKYKVVVLIYGSAWFANMMKQMAFQ-TMGRPLLDAGFAVVCANHRSSDAK	88
<b>EDV05955.1</b>	DIMLPDGGQTEYKVVVLIYGSAWFANMMKQMAFQ-AMGKPLLDGFAVVSINHRSSDAK	117
WP_007216415.1	DIMLPDGGQTEYKVVVLIYGSAWFANMMKQMAFQ-AMGKPLLDGFAVVSINHRSSDAK	98

(Trp563)

SEH83175.1	WPAQIHDIKAVIRFVRGEAKKYKFDTKFIATSGFSSGGHLASTAATTSGTKQTKVGVTDI	176
WP_006281679.1	WPAQIHDIKAVIRFVRGEAKKYKFDPSFVATSGFSSGGHLASTAATTSGTKATKVGSDTI	177
WP_100993215.1	FPAQIHDVKAAIRFVRANAATESVTDQF IGITGWSGGHLASLAGTINHVKTKITIDGIEV	179
WP_020527766.1	FPAQIQDVKAAIRFVRANALAESLDTSF IGITGWSGGHLASAFAGTANNITTFEFGNTI	181
WP_099150634.1	FPAQIQDVKAAIRFVRANAPSEDLDPAF IGITGWSGGHLASALTGTNNGIDSVSTEGEV	169
WP_013768612.1	FPAQIQDVKAAIRFVRANAPENLAPDF IGITGWSGGHLASALTGTANNVRKEVTQGEV	172
WP_114463085.1	FPAQIQDVKAAIRFVRANAAKESMDDSF IGITGWSGGHLASTALTGTNTTQTHSIHGLEV	170
WP_015029998.1	WPAQIHDIKAAIRFVRANASVESLDTSF IGITGFSGGHLSIMAGVTSIGIKSTTINHLP	168
WP_055151666.1	WPAQIHDIKAAIRFVRANADQESLDRFLGIGYSSGGHLSIMAGVTSIGLEEFVIGGLEI	168
WP_006799192.1	FPAQINDVKAAIRFVRANASVYRLDTEFVIGISGGGNMAALAGTSRFAKQCTIIGNATV	159
WP_027452672.1	YPAQINDVKAAIRFVRANAKAYQLDTSF IGITGFSGGHLSSVAGVTNNLPKKTIDHGVSI	170
KQB43445.1	YPAQINDVKAAIRFVRANAKKYNLDSSF IGITGFSGGHLASLAGTNGITNNTYVKKTI	159
WP_071144936.1	YPAQINDVKAAIRFVRANADKYHIDASFVIGITGFSGGHLSSLAGTNNVKEFTVGNVTL	157
WP_100615198.1	YPAQINDVKAAIRFVRANADYINIDASF IGITGFSGGHLASLAGTNGVKTFTVGEKTV	161
WP_123396545.1	FPAQIQDVKAAIRFVRANADKYHIDPSF IGITGFSGGHLSSLAGTNGVKTFTVGDVTV	161
WP_117741097.1	FPAQINDVKAAIRFVRANAAYKLDTSF IGITGFSGGHLASLAGTNGVKSYYTIGDKTV	659
<b>WP_050793236.1</b>	FPAQINDVKAAIRFVRANAAYKLDTSF IGITGFSGGHLASLAGTNGVKSYYTIGAKTV	659
WP_024996568.1	FPAQINDVKAAIRFVRANAKEYKIDASF IGITGFSGGHLSSLAGATNGVKAYTVGNTTV	157
WP_018709923.1	FPAQINDVKAAIRFVRANAHAGEYRLDTMFGITGSSGGHLAALAGTNDVGVYRVDGSEM	148
<b>EDV05955.1</b>	FPAQINDVKAAIRFVRANAHADYRLDTSF IGITGFSGGHLSSLAGTNGVKVYKVGDTM	177
WP_007216415.1	FPAQINDVKAAIRFVRANAHADYRLDTSF IGITGFSGGHLSSLAGTNGVKVYKVGDTM	158

(Ser634)

SEH83175.1	DIEGNVGNVYLNSSSVNAACDWSGPIDLTAMICGESM---KMGENSPEVDVILNSKLAKEP	233
WP_006281679.1	DIEGNLGOYTQESSQVNAACDWSGPIINLMNMICGHHI---TMGKDSPEDIMLRSKLDEP	234
WP_100993215.1	DIEGHLGPEFTPVSSHVNAVVDWFGPTNFLLMISCDPR-MHHNDPRSPESSLIGGGITQHT	238
WP_020527766.1	DIEGSLGNHTSFSSQVDAVVDWFGPTNFLVMDNCGST-MKHDNDTSPSSLYGGAIQSNK	240
WP_099150634.1	NIEGSLGKFTDSDSQQVAVVDWFGPTDFLIMTICGSS-FSHDKPKSPSSLYGGPIQINP	228
WP_013768612.1	DIEGALCKHLTSSAIDAVVDWFGPTDFLKMDDCGSS-FSHNDAKSPSSLYGGPIQINP	231
WP_114463085.1	DIEGSLGKHTQTSSKVNNAVVDWFGPTDFLIMDCGSS-FAHNEAKSPSSLYGGAIQINP	229
WP_015029998.1	DIEGNIIGKSLGESSNVDAVVDWFGPTDFLLMACGSS-FSHNEAKSPSTLIGGAIQINP	227
WP_055151666.1	SIDNKIGKHPEDSSVDAVVDWFGPTDFLIMISCGSS-FSHDGAISPESTLVGGPIQINP	227
WP_006799192.1	DMEGIGPVTQFSSQVDAVVDWFGPTNMLVMSCGGTDFIHNPNSFAAYIGGGPIQINP	219
WP_027452672.1	DIEGKVGDCLNSSRVDAVVDWFGPTDMAHMKCTIT---NDDKSPSAALIGGDPKPKMS	226
KQB43445.1	NIEGNIIGENTSSASKVDAVVDWFGPVNLALMAACEKP---KDEKSPSAALIGGNPAINL	215
WP_071144936.1	DIEGNLCAVTNYSSRVDAVVDWFGPTDMTMMKCKGV---NDEKSPSAALIGGAPAINL	213
WP_100615198.1	DIEGDLIGNYTQVSSAVDAVVDWFGPTDFTRMENCNTIT---KDDKSPSAALIKGNPAINL	217
WP_123396545.1	DIEGNLGDYTSASSDVNAVVDWFGPTDMSRMENCNTIT---KADSPEAMLIGGAPAINL	217
WP_117741097.1	DIEGNVGEYFSSSRVDAVVDWFGPTDMTRMENCNTIT---KGANSPSAALIGGIPAINL	715
<b>WP_050793236.1</b>	DIEGNVGLYPSFSSRVDAVVDWFGPTDMTRMENCNTIT---KGANSPSAALIGGIPAINL	715
WP_024996568.1	DIEGSLVCAVGSFSSDVDAVVDWFGPTDMTRMENCNTIT---KADSPEAALIGGAPAINP	213
WP_018709923.1	DIEGKVGDCLAHSSRVDAVVDWFGPTDMTRMENCSTIT---KGS DSPSAALIGGNPASHL	204
<b>EDV05955.1</b>	DIEGNVGDCTSFSSRVDAVVDWFGPTDMTRMENCATIT---KADSPEAALIGGTPAASHM	233
WP_007216415.1	DIEGNVGDCTSFSSRVDAVVDWFGPTDMTRMENCATIT---KGV DSPSAALIGGTPAASHM	214

(Glu703)

SEH83175.1	DKYLSLSANTYVDK--NDPPIIFHGEKDNVVPCCQKGAFFETLKAGVKTEATFVEBGS	291
WP_006281679.1	DKYKLSLSATYVDK--NDPPIIFHGEKDNVVPCCQKGEFYEVLLKAGVKTEATFVEKGE	292
WP_100993215.1	SRVELANPIITYYSK--ASPPFLIHGDKDPLVPHCQSEELANKLQEGVKSLLIQVSGGR	296
WP_020527766.1	NKCHLADPVTYYSK--NATPFLIHGDKDPLVPHCQSEYLYEKLNESDIYCEFIIVEGGK	298
WP_099150634.1	LKCLLANPIHYATA--NPPFLLFHGDQDPLVPHCQSEKLLHQIQRRQAPSELVIIVEGGK	286
WP_013768612.1	VKVATANPIISYVK--SNPPFLLFHGDKDPLVPHCQSOLLFEQLQAAGVSSKLVIIIEGG	289
WP_114463085.1	DKCALANPIISYVNA--QPPFLLFHGDKDPLVPHCQSEKLYEALQKNVKSLLIIIEGGG	287
WP_015029998.1	EKVLANPIISYYSK--ATPPFLLFHGDKDPLVPHCSEKLYEKMQKGVKSELIIIEGGG	285
WP_055151666.1	AKVALANPIISYVK--ENPPMLFHGTADPLVPHCSEKLYEAQQKAGAVSRLVIVEGGG	285
WP_006799192.1	DKCLLASPTTYIDS--SDPPFLLFHGDKDRVVPHCQSELLFEALQKAGVPSRFYLVSGGQ	277
WP_027452672.1	EMVSLISPIDYAEAPDCPRFLIHGDSDTVVPHCQSENFAAVLKAGKLVDFIIVEKGG	286
KQB43445.1	DMINLLSPTTYIDK--TDPDFLLIHGDADSVVPYQCSVLFAKSLKKGKLVDFIIVEKGG	273
WP_071144936.1	DMLALLNPMTYIDK--TDPKFLIHGDSDNVVPHCQSEFFSAALKYGLLDDFIIVEAGGQ	271
WP_100615198.1	DMLALLNPMTYIDN--KDPQFLIHGDADNVVPHCQSVFFSNLKDGLNEFIIVEBGGQ	275
WP_123396545.1	DMIKLLNPMTYIDA--NDPKFIVIHGDADPVVPYQSEYFAEALKKNGNLNEFIIVEBGGQ	275
WP_117741097.1	DMLALLNPMTYIDK--NDPKFIVIHGEADTVVPNCQSIFFSEALRAQGRLEEFIVSVEGGQ	773
<b>WP_050793236.1</b>	DMLALLNPMTYIDK--NDPKFIVIHGEADTVVPNCQSIFFSEALRAQGRLEEFIVSVEGGQ	773
WP_024996568.1	EMLALLNPMTYIDV--KDPKFLIHGDADTVVPNCQSIYFSALKKAGLLEEFIVSVEGGQ	271
WP_018709923.1	DMLALLNPMTYIDS--ADPEFLIHGDADPVVPYQSLFFKEALDEKGLAEFIVVEKGE	262
<b>EDV05955.1</b>	DVLTLLNPMTYIDE--KDPKFLIHGDADTVVPHCQSVFFKDTLSAKGRLEEFIVVEQGG	291
WP_007216415.1	DVLTLLNPMTYIDE--KDPKFLIHGDADTVVPHCQSVFFKDALSAKGRLEEFIVVEQGG	272

(Asp742)

SEH83175.1	HGGPAMYVEENIQKMNFLKALL-----	314
WP_006281679.1	HG-MNMYDEENIKKMTDFLNRVRKGG-----	317
WP_100993215.1	HG-PGVLSSESYTKMVSFFTTWKQVKN-----	323
WP_020527766.1	HG-PGVLIPEYIQQMIFFDKIRIDK-----	323
WP_099150634.1	HG-PGVMIPEYDQMVAFQKHAMGK-----	311
WP_013768612.1	HG-PGVLIPEYEQMIQFFKGLKSK-----	314
WP_114463085.1	HG-PGVMIPEYQKMAVFFKGLGKVKK-----	314
WP_015029998.1	HG-PGVMIPEYVQMITFFFSKIKK-----	309
WP_055151666.1	HG-PGVMIPEYDEMVAFFNQKSKNLTR-----	313
WP_006799192.1	HG-PGVHVPENIKLMTDFVSVCKERKYLK-----	305
WP_027452672.1	HG-PITFNENTFKKMVEFFKAEKK-----	310
KQB43445.1	HG-PVTFNEETFKKMTDFFLKAEKK-----	298
WP_071144936.1	HG-PVTFNEETHSKMTEFFFLKAGMQ-----	296
WP_100615198.1	HG-PATFNENTFKKMSDFFLTAKKK-----	300
WP_123396545.1	HG-PVTFNEETFKKMTDFVVEARQVVTGK-----	303
WP_117741097.1	HG-PVTFNEETFKKMTDFFAKAGM-----	797
<b>WP_050793236.1</b>	HG-PVTFNEETFKKMTDFFARAGIYRDLNRIRPKC---DDIK-----	813
WP_024996568.1	HG-PVTFNEATFKKMTDFVVKSK-----	294
WP_018709923.1	HG-PVTFNEQTFKKMTDFFRMAGMD-----	287
<b>EDV05955.1</b>	HG-PITFNEQTFKKMTEFFFRKAAMDLCPANITLVPHIRPGSKKNPPQKQFYTTY	345
WP_007216415.1	HG-PITFNEQTFKKMTEFFFRKAAMDLCPANITLVPHIRPGSKKNPPQKQFYTTY	326

(His774)



**b**

PZX17224.1 -----MIVNNLKIILSM--AIAMVMGNITFAQTAKVNDDFKPSSEVNVQPGKLYFQVN 48  
WP\_117954567.1 -----MAMKRIITLFGAF--ALLTNLNSFSYAQALDFKPSVSNVQIGKAYFQVN 47  
**WP\_007661004.1** -----MKKQILFWSMM--SWMVSVGLPSFAQTVEDFKPSSEVNVQPGKLYFQVN 45  
WP\_007216414.1 -----MKKQILFWSMM--SWMVSVGLPSFAQTVEDFKPSSEVNVQPGKLYFQVN 45  
WP\_123120797.1 -----MNY--KRFIFFAATFLLGASCFA-----QSDGFKPSATNAPGKQYFQVS 42  
WP\_066034169.1 -----MRLSIN--AVIAIFS---LYFQSVNAQSEQIITEDFKPSAANQBERITFQVN 47  
WP\_055090457.1 -----MNYK--RIIYLTFV--CSVTISLAQTKTQEVLEDFKPSVSNVQCGKLYFQVN 47  
WP\_035660712.1 -----MNFK--SILFSVNV--LASICTAQTNQIKVLEDFKPSVSNVQCGKLYFQVN 47  
WP\_083552732.1 -----MNYK--SIYLVSI--MTTIGIGQTNPLKVTEDFKPSVSNVQCGKLYFQVN 47  
WP\_066330097.1 -----MKVN--SIVFVIT--ISNTSFSQSVSSNVVDFVPSVSNVQCGKLYFQVN 47  
WP\_068704550.1 -----MNIKHIFIAFVAA-----AFCQMAPAQTVVDFKPSVSNVQPGKLYFQVN 44  
WP\_020531203.1 MDFLKLNLMTSKINHRYILFLI---GFMLTGVICFAQSEVDFKPSVSNVQPGKLYFQVN 57  
WP\_016778581.1 -----MKYIQSLLVIFALFIGSIATAQTDKVDFKPSVSNVQCGKLYFQVN 46  
WP\_106153941.1 -----MRYIHLLIILF--LLSSNFVTAQTEKVKEDFKPSVSNVQIGKLYFQVN 45  
WP\_053183372.1 -----MKHFLILIALFVLSATNLVLAQTSKAITEDFKPSSEVNVQPGKLYFQVN 46  
WP\_045027037.1 -----MKNTFVLI AALFIG---ICTAQDVVDFKPSVSNVQPGKLYFQVN 42  
WP\_038555069.1 -----MKNIVILIAALFIG---ICTAQDVVDFKPSVSNVQPGKLYFQVN 42

PZX17224.1 SGGCVRVQISAPDAKMQVLDIGGVYTNLVKDGNGLWTGESAPQDEGFHYYQINVDGASVF 108  
WP\_117954567.1 SEGRVRAQIYAPDAKRVQVLDIGGVYDMTKNEQGFWTGESERQDEGFHYYQINVDGASVF 107  
**WP\_007661004.1** SERKVRVQISAPDAKRVQVLDIGGVYDMTKDEKGVWTGESAPQDEGFHYYQINVDGAAVF 105  
WP\_007216414.1 SERKVRVQISAPDAKRVQVLDIGGVYDMTKDEKGVWTGESAPQDEGFHYYQINVDGAAVF 105  
WP\_123120797.1 ADGRVRASIVAPDAHNQVLDIGGVYDMTKDEKGVWTGESLPQDEGFHYYQINIDGVSVF 102  
WP\_066034169.1 SORRVRASISAPNASKVQVLDIGGVYDMRKDKGVWTGESNPQDEGFHYYQINIDGASVF 107  
WP\_055090457.1 SGRVRANILAPDAKRVQVLDIGGVYDMVKDAKGLWTGESLPQDEGFHYYQINIDGASVF 107  
WP\_035660712.1 SERRVRASILAPDAKRVQVLDIGGVYDMVKDEKGLWTGVSENPQDEGFHYYQINIDGASVF 107  
WP\_083552732.1 SGRVRASILAPDAKRVQVLDIGGVYDMVKDKGLWTGVSENPQDEGFHYYQINIDGASVF 107  
WP\_066330097.1 SERRVRASILAPDAKRVQVLDIGGVYDMVKDKGLWTGVSNPQDEGFHYYQINIDGASVF 107  
WP\_068704550.1 SEGRVRSIPAPNAQVQVLDIGGVYDMKDKGVWTGESAPQVEGFHYYQINVDGASVF 104  
WP\_020531203.1 SERRVRASIAAPANLVQVLDIGGVYDMTKDNGVWTGESAPQDEGFHYYQINIDGASVF 117  
WP\_016778581.1 SEGRVVRIEAPDAKRVQVLDIGGVYDMTKDNGVWTGESAPQDEGFHYYQINIDGASVF 106  
WP\_106153941.1 SEGRVVRIEAPGANNVQVLDIGGVYDMTKDNGVWTGESSENPQDEGFHYYQINIDGASVF 105  
WP\_053183372.1 SEGRVRAQISAPDAKRVQVLDIGGVYDMVKDEDGVWTGESAPQDEGFHYYQINIDGASVF 106  
WP\_045027037.1 SEGRVRAQIILAPDAKRVQVLDIGGVYDMVKDENGVWTGESSENPQVGFHYYQINVDGASVF 102  
WP\_038555069.1 SEGRVRAQISAPGANNVRLDIGGVYDMVKDENGVWTGESSENPQVGFHYYQINIDGASVF 102

PZX17224.1 DPGSRVYFYGAGRWGSGTEIPADDSHIFALQDVPHGLVSELNLYFSKHTGLMRRRCFVYTPAG 168  
WP\_117954567.1 DPGTKYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSELNLYSKHTQSWRRRCFVYTPAG 167  
**WP\_007661004.1** DPGTYFYGAGRWGSGTEIPAHDAFYALKDVPHGLLSEMNYSNLTKAWRRRCFVYTPAG 165  
WP\_007216414.1 DPGTYFYGAGRWGSGTEIPAHDAFYALKDVPHGLLSEMNYSNLTKAWRRRCFVYTPAG 165  
WP\_123120797.1 DPGTYFYGAGRWGSGTEIPASDQDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAG 162  
WP\_066034169.1 DPGTYFYGAGRWGSGTEIPSSDQDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAG 167  
WP\_055090457.1 DPGTYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAE 167  
WP\_035660712.1 DPGTYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAN 167  
WP\_083552732.1 DPGTYFYGAGRWGSGTEIPASDQDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAN 167  
WP\_066330097.1 DPGTYFYGAGRWGSGTEIPASDQDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAN 167  
WP\_068704550.1 DPGTYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAE 164  
WP\_020531203.1 DPGTYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAE 177  
WP\_016778581.1 DPGTYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAG 166  
WP\_106153941.1 DPGTKYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAG 165  
WP\_053183372.1 DPGTKYFYGAGRWGSGTEIPAHDFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAG 166  
WP\_045027037.1 DPGTKYFYGAGRWGSGTEIPADDRNFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAE 162  
WP\_038555069.1 DPGTYFYGAGRWGSGTEIPAKMDNFYALKDVPHGLVSENIYFSKHTNSWRRRCFVYTPAE 162

PZX17224.1 YEADANMRYPVLYLQHGSEFDETGWICOGKANLIDLNLIAHKAIPMIIVMDNGYAKPL 228  
WP\_117954567.1 YEVNTERYPVLYLQHGSEFDETGWISOGKANLIDLNLIAAKKAVPMIIVMDNGYAKRPS 227  
**WP\_007661004.1** YGDNKDKRYPVLYLQHGSEFDETGWIFROGKTNLIDLNLIAAGKAVPMIIVMDNGYAKPG 225  
WP\_007216414.1 YGDNKDKRYPVLYLQHGSEFDETGWIFROGKTNLIDLNLIAAGKAVPMIIVMDNGYAKPG 225  
WP\_123120797.1 YEKDAKTRYPVLYLQHGSEFDETGWISOGKANLIDLNLIASGKAVPMIIVMDNGYAKPS 222  
WP\_066034169.1 YNENTKTRYPVLYLQHGSEFDETGWISOGKANLIDLNLIAAKKAVPMIIVMDNGYAKAC 227  
WP\_055090457.1 YFONQDKRYPVLYLQHGSEFDETGWISVOGKANLIDLNLIAAKKANPMIIVMDNGYAKPC 227  
WP\_035660712.1 YNEBSKTRYPVLYLQHGSEFDETGWAVOGKANLIDLNLIASKKAVPMIIVMDNGYAKPC 227  
WP\_083552732.1 YNENTKTRYPVLYLQHGSEFDETGWAVOGKANLIDLNLIASKKANPMIIVMDNGYAKPC 227  
WP\_066330097.1 YNTDKTRYPVLYLQHGSEFDETGWAVOGKANLIDLNLITSKKAVPMIIVMDNGYAKPC 227  
WP\_068704550.1 YDKNPSKRYPVLYLQHGSEFDETGWPTOGKANLIDLNLIAAKKAVPMIIVMDNGYAKAN 224  
WP\_020531203.1 YENDNAKRYPVLYLQHGSEFDETGWISOGKANLIDLNLIAEKKAVPMIIVMDNGYAKPC 237  
WP\_016778581.1 YHRDLKRYPVLYLQHGSEFDETGWISNOGHANLIDLNLIAENKAVPMIIVMDNGYAKAC 226  
WP\_106153941.1 YYEDINKRYPVLYLQHGSEFDETGWISOGKANLIDLNLIAEKKAVPMIIVMDNGYAKAC 225  
WP\_053183372.1 YETSTDKRYPVLYLQHGSEFDETGWPCOGKANLIDLNLIAEKKAVPMIIVMDNGYAKPH 226  
WP\_045027037.1 YTNPKDKRYPVLYLQHGSEFDETGWPCOGKANLIDLNLIAAKKAVPMIIVMDNGYAKPC 222  
WP\_038555069.1 YNANPKQYPVLYLQHGSEFDETGWASOGHANLIDLNLIAAKKAVPMIIVMDNGYAKPC 222

PZX17224.1	A--SADGNARPASVFEEVMMQEIIPIDSEFRTLVVHRDHRATAGLSMGANQTMRIAMNHL	286
WP_117954567.1	EQT-----PFQITVFEEVLMQEMIPMIDKFKRTLIPNRESRAIAGLSMGANQTMRIAMNHP	282
<b>WP_007661004.1</b>	EKS-----PFAASIFEEVLMNEVIMPIDAKFRTLSGREDRATAGLSMGANQTMRIAMNHP	280
WP_007216414.1	EKS-----PFAASIFEEVLMNEVIMPIDAKFRTLSGREDRATAGLSMGANQTMRIAMNHP	280
WP_123120797.1	DNP-ADSKTRPQFVFEEVMMINEVIMPIDGKFKRTLIPNREHRAIAGLSMGANQTMRIIMNHL	281
WP_066034169.1	EENS-SDKNPKBSVFELVLIITFIMPIDAKFRTIPKRENRAIAGLSMGANQTMRIIMNHL	286
WP_055090457.1	TNG-----ERQBSVFEEVLINEIIPMIDAKFRTLANRENRAIAGLSMGANQTMRIIMNHL	282
WP_035660712.1	SND-----SRQBSVFEEVVIQEIIPMIDTKFKRTLIPNRENRAIAGLSMGANQTMRIIMNHL	282
WP_083552732.1	SGT-----ARQBSVFEEVLINEIIPMIDAKFRTIPNRENRAIAGLSMGANQTMRIIMNHL	282
WP_066330097.1	TNG-----ARQBSVFEEVLINEIIPMIDAKFRTIPNRENRAIAGLSMGANQTMRIIMNHL	282
WP_068704550.1	A---SAPSKSAFFPFEVMMINEIIPMIDASFRTLSDREHRAIAGLSMGANQTITITMNNL	280
WP_020531203.1	NLENADRRGRPASAFEEVMTTEIIPMIDTEFRTLKDKREHRAIAGLSMGANQTMRIIMNHL	297
WP_016778581.1	E--N-GITGRPAMVFEEVMMQEIIPMIDAFRTIPNRENRAIAGLSMGANQTMRIAMNHL	283
WP_106153941.1	D--K-DKGRPAMVFEEVMMQEIIPMIDAFRTIPNRENRAIAGLSMGANQTMRIAMNHL	282
WP_053183372.1	E--NSDCNSRPVSVFEEVMTSEIIPMIDAFRTILDRKHRAIAGLSMGANQTMRIIMNHL	284
WP_045027037.1	T--S--CGGRPAMVFEEVMMINEIIPMIDNRFRTIADREHRAIAGLSMGANQTMRIAMNHL	278
WP_038555069.1	S--S--CGGRPAMVFEEVMMINEIIPMIDKFKFRTIADREHRAIAGLSMGANQTMRIAMNHL	278

(Ser226)

PZX17224.1	DFFAFYGGFSGTANYPSPAQPIVDVTFLDGKFKNGSWNQQLLWLG/GTKKEPVFKS	346
WP_117954567.1	EAFAYYGGFSGTSNYPSTEPIDADTFLGGKFKDGAANKQKSF/FLGLGTSE/PAPFPGVV	342
WP_007661004.1	GHFAYYGGFSGTSNYPSTEPIDATFLNGKFKDAKAVNQEFKVF/FLGLGTAE/PHFPFGVV	340
<b>WP_007216414.1</b>	GHFAYYGGSSGTSNYPSTEPIDATFLNGKFKDAKAVNAQEFKVF/FLGLGTAE/PHFPFGVV	340
WP_123120797.1	DFFSSLCG/FSGTSNYPNDV/NPATFLDGKFNDAAN/KQK/VLFLGLGTKEPNPFGSV	341
WP_066034169.1	DFFAYYGGFSGTANYPSSDAIDASTFLNGKFKDGNAN/KQK/VLFWLGLGTKEPAPFPKSV	346
WP_055090457.1	NFFSYGGFSGTSNYPSSADAI/DVNTFLDGKFKDGSVNEK/KLFWLGLGTKEPSPFPGSV	342
WP_035660712.1	DFFSHYGGFSGTSNYPSSADAI/DVNTFLDGKFKDGNAN/KK/VLFWLGLGTKEPSPFPGSV	342
WP_083552732.1	DFFSHYGGFSGTSNYPSSADAI/DVNTFLDGKFKDGNAN/KK/VLFWLGLGTKEPSPFPGSV	342
WP_066330097.1	DFFSHYGGFSGTSNYPSSADAI/DVNTFLDGKFKDANAN/KK/VLFWLGLGTKEPSPFPGSV	342
WP_068704550.1	DFFSYIAGFSGTSNYPSTPAIDVETFLNGAFKFKDGAAN/KQK/VLFWLGLGTVE/PAPFPGSV	340
WP_020531203.1	DVFASYGGFSGTSNYPSSADDI/DVNTFLGGKFKDGEAL/SKQK/VLFWLGLGTKEPDPFPGSV	357
WP_016778581.1	DFFAFYGGFSGTANYPSSDEIDVETFLDGAFAFDGKVD/EO/KVLWGLGLGTKEPEPFPFESI	343
WP_106153941.1	DFFAFYGGFSGTANYPSSDEIDVKTFLDGAFAFDKVD/EO/KVLWGLGLGTKEPDPFPGSI	342
WP_053183372.1	GVFSNYGGFSGTSNYPSSSEIDPATFLDGKFKDGGKVN/EO/VLFWLGLGTKEPEPFPGSV	344
WP_045027037.1	DFFAFYGGFSGTSNYPSSDEIDVETFLNGAFKFKGKSVND/OK/VLFWLGLGTKEPEPFPGSV	338
WP_038555069.1	DFFAFYGGFSGTSNYPSSDEIDVETFLNGAFKFKGKSVNKQK/VLFWLGLGTKEPEPFPGSV	338

(Glu332)

PZX17224.1	GAFRNMLEQQGIRHVFYYSQGTAEHWLTWRRSLHQYAQLVFK--	388
WP_117954567.1	KAFROMLDKQGIKYTYYESPETAEHWLTWRRALHQYAQLLEQ--	384
<b>WP_007661004.1</b>	KAFROMMDKQGIKYVYYESEPETAHEWLTWRRALNEFAPLLFK--	382
WP_007216414.1	KAFROMMDKQGIKYVYYESEPETAHEWLTWRRALNEFAPLLFK--	382
WP_123120797.1	GAFRNMLEKQGIKYVYYS/PATSHWLTWRRDLKEFAQLLFFK--	384
WP_066034169.1	GAFRTMLEQQGIKYDYYESPETAEHWLTWRRCLHQFASKLFK--	388
WP_055090457.1	GAFRNMLEKQGIKYTYYESAETAHEWLTWRRCLNQFAAKLFK--	384
WP_035660712.1	RAFRRKMLEKQSIKYVYYESEPETAHEWLTWRRCLNQFASKLFK--	384
WP_083552732.1	GVFRTMLEQQGIKYAYYESPKETAHEWLTWRRCLNQFATQLEFK--	384
WP_066330097.1	GAFRTMLEKQGIKYGYYESPETAEHWLWRRCLNQFAAKLEQ--	384
WP_068704550.1	KAFRNMLEKQGIKYTYYESQGTAEHWLTWRRCLNQFAALLFK--	382
WP_020531203.1	GAFRNMIDNIGLDHVYYESEPETAHEWLTWRRSLKEYAALLFK--	399
WP_016778581.1	GFTINMIKKQGIQYEYYESPETAEHWLTWRRSLYQYAQLLFK--	385
WP_106153941.1	GAFINMIKKQGIQYEYYESPETAEHWLTWRRSLYQYAQLLFK--	384
WP_053183372.1	GAFRRKMLEKQGIKYVYYESEPETAHEWLTWRRSLHEYAQR/LFKKN	388
WP_045027037.1	GAFRNMLEKQGIQYVYYESEPETAHEWLTWRRSLHQYAQLLFK--	380
WP_038555069.1	GAFRNMLEKQGIQYVYYESEPETAHEWLTWRRSLHQYAQLLFK--	380

(His364)

C

WP\_062122566.1 --MKNV-VKLMFSLVLI SFMANGQNAPAFPEGTLPNEHNIAGADYPRVGEDGRVHFRVF 57  
 WP\_071144963.1 --MKNRLANVLIALIGFVPIV--IAAAQESFPEGTVPNNEHNIAGADYPRIGVDRRVHFRVH 56  
 WP\_123396861.1 --MKTFLSKLILSTCLVPLA-LGAQQQPAFPAGTVPNEHNIAGADYPRIGEDGRVHFRVF 56  
 WP\_101690191.1 --MKRV-LKLVFFLALPMV-LNAQQQDFPEGTVPNEHNIAGADYPRIGEDGRVHFRVH 56  
 WP\_007660993.1 --MKNV-LKLI FSSLLMF PFT-MNAQQQDFPAGTVPNEHNIAGADYPRIGEDRRVHFRVH 56  
 WP\_082717889.1 MLMKNV-LKLI FSSLLVFPFT-MNAQQQDFPAGTVPNEHNIAGADYPRIGEDRRVHFRVH 58  
 WP\_081743618.1 MLMKNV-LKLI FSSLLALPFT-MNAQQQDFPAGTVPNEHNIAGADYPRIGEDRRVHFRVH 58  
 WP\_004293186.1 --MKNA-LKLI FSSLLALPLT-MYAQQQDFPEGTVPNEHNIAGADYPRIGEDRRVHFRVF 56  
 WP\_022104001.1 --MKNA-LNFI FLFLAVFPLT-MYAQQEFPAGTVPNEHNIAGADYPRIGEDRRVHFRVF 56  
 WP\_026314837.1 --MKNA-LNFI FLFLAALPLT-VYAQQEDFPAGTVPNEHNIAGADYPRIGEDRRVHFRVF 56

WP\_062122566.1 APDAQKVEISFRGEMTKGDDGFWLISDEPEVGFHYYQIINDVSVADPHGCKPFFGMGK 117  
 WP\_071144963.1 APNAQKVEISFRGEMTKADGYSLSVSKAEVGFHYYQIINDGVSAADPNKPPFFGMGK 116  
 WP\_123396861.1 APNAQKVEISFRGEMTKGDDGFWLISVSKPEVWGFHYYQIINDGVSAADPNKPPFFGMGK 116  
 WP\_101690191.1 APNAQKVEISFRGEMTKADGYSLSVSKPEVWGFHYYQIINDGVSAADPNKPPFFGMGK 116  
 WP\_007660993.1 APNAQKVEISFRGEMTKADGYSLSVSKPEVGFHYYQIINDGVSAADPNKPPFFGMGK 116  
 WP\_082717889.1 APNAQKVEISFRGEMTKADGYSLSVSKPEVGFHYYQIINDGVSAADPNKPPFFGMGK 118  
 WP\_081743618.1 APNAQKVEISFRGEMTKADGYSLSVSKPEVWGFHYYQIINDGVSAADPNKPPFFGMGK 118  
 WP\_004293186.1 APNAQKVEISFRGEMTKADGYSLSVSKPEVWGFHYYQIINDGVSAADPNKPPFFGMGK 116  
 WP\_022104001.1 APNAQKVEISFRGEMTKADGYSLSVSKPEVWGFHYYQIINDGVSAADPNKPPFFGMGK 116  
 WP\_026314837.1 APNAQKVEISFRGEMTKADGYSLSVSKPEVWGFHYYQIINDGVSAADPNKPPFFGMGK 116

WP\_062122566.1 WVSGIIEPEKGCDDYKPKQGVPHGLVSWWYYSVRCWRRQVYTPAEYDKNPKKYPV 177  
 WP\_071144963.1 WVSGIIEPEKGVDDYYSIK--NVPHGLVSWWYYSDIRKEWRQVYTPAEYDKNPKKYPV 175  
 WP\_123396861.1 WVSGIIEPEKGVDDYYSIK--NVPHGLVSWWYYSKIRWRRQVYTPAEYDKNPKKYPV 175  
 WP\_101690191.1 WVSGIIEPEKGVDDYYSIK--NVPHGLVSWWYYSQVYSDIRKEWRQVYTPAEYDKNPKKYPV 175  
 WP\_007660993.1 WVSGIIEPEKGVDDYYSIK--NVPHGLISQSWYYSDIRKEWRQVYTPAEYDKNPKKYPV 175  
 WP\_082717889.1 WVSGIIEPEKGVDDYYSIK--NVPHGLISQSWYYSDIRKEWRQVYTPAEYDKNPKKYPV 177  
 WP\_081743618.1 WVSGIIEPEKGVDDYYSIK--NVPHGLISQSWYYSDIRKEWRQVYTPAEYDKNPKKYPV 177  
 WP\_004293186.1 WVSGIIEPEKGVDDYYSIK--NVPHGLISQSWYYSIRWRRQVYTPAEYDKNPKKYPV 175  
 WP\_022104001.1 WVSGIIEPEKGVDDYYSIK--NVPHGLISQSWYYSDIRKEWRQVYTPAEYDKNPKKYPV 175  
 WP\_026314837.1 WVSGIIEPEKGVDDYYSIK--NVPHGLISQSWYYSDIRKEWRQVYTPAEYDKNPKKYPV 175

WP\_062122566.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKVTPIVMDNGNIEVLFKAEEDMNEVFR 237  
 WP\_071144963.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 235  
 WP\_123396861.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 235  
 WP\_101690191.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 235  
 WP\_007660993.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 235  
 WP\_082717889.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 237  
 WP\_081743618.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 237  
 WP\_004293186.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 235  
 WP\_022104001.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 235  
 WP\_026314837.1 LYLQHGGENETSWANQGMNFI MDNLI AEGKAPMIVMDNGNIEVLFKAEEDMNEVFR 235

WP\_062122566.1 KRFGAQPFAILLVNEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 297  
 WP\_071144963.1 KRFGGQFPDILLVKEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 295  
 WP\_123396861.1 KRFGGQFPDILLVKEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 295  
 WP\_101690191.1 KRFGGQFPDILLVKEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 295  
 WP\_007660993.1 KRFGAQPFAILLVNEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 295  
 WP\_082717889.1 KRFGAQPFAILLVNEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 297  
 WP\_081743618.1 KRFGAQPFAILLVNEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 297  
 WP\_004293186.1 KRFGAQPFAILLVNEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 295  
 WP\_022104001.1 KRFGAQPFAILLVNEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 295  
 WP\_026314837.1 KRFGAQPFAILLVNEIIPHIESNFRVLTDRNRAMAGLSWGGLLTFNNTLNNLDKFAYIGG 295

(Ser273)

WP\_062122566.1 FSGAGSIDLNNLDAAVYNGVFSASAFAFNKVKVFFLGVGSEEHPERTKLS DGLKAAGINN 357  
 WP\_071144963.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLKAAGINN 355  
 WP\_123396861.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLKAAGINN 355  
 WP\_101690191.1 FSGAGSIDLNNLDVYGGVFKDRKAFNQQVHVFVFLGIGSEEHPERTKLS DGLKAAGINN 355  
 WP\_007660993.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLQAAGINT 355  
 WP\_082717889.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLQAAGINT 357  
 WP\_081743618.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLQAAGINT 357  
 WP\_004293186.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLKAAGINN 355  
 WP\_022104001.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLKAAGINN 355  
 WP\_026314837.1 FSGAGSIDLNNLDVYGGVFKDRKAFNDKVHVFVFLGIGSEEHPERTKLS DGLKAAGINN 355

(Glu336)

WP\_062122566.1 IYYESPGTAHEFLTWRRLKEFAPLLFK--- 385  
 WP\_071144963.1 IYYESPGTAHEFLTWRRLKEFVPLLVKFK 386  
 WP\_123396861.1 IYYESPGTAHEFLTWRRLKEFAPLLFN--- 383  
 WP\_101690191.1 IYYESPGTAHEFLTWRRLKEFVPLLVTKK 386  
 WP\_007660993.1 IYYESPGTAHEFLTWRRLKEFAPLLFKTK- 385  
 WP\_082717889.1 IYYESPGTAHEFLTWRRLKEFVPLLVKTK- 387  
 WP\_081743618.1 IYYESPGTAHEFLTWRRLKEFAPLLFKTK- 387  
 WP\_004293186.1 IYYESPGTAHEFLTWRRLKEFVPLLVKTK- 384  
 WP\_022104001.1 IYYESPGTAHEFLTWRRLKEFAPLLFK--- 383  
 WP\_026314837.1 IYYESPGTAHEFLTWRRLKEFVPLLVKTK--- 383

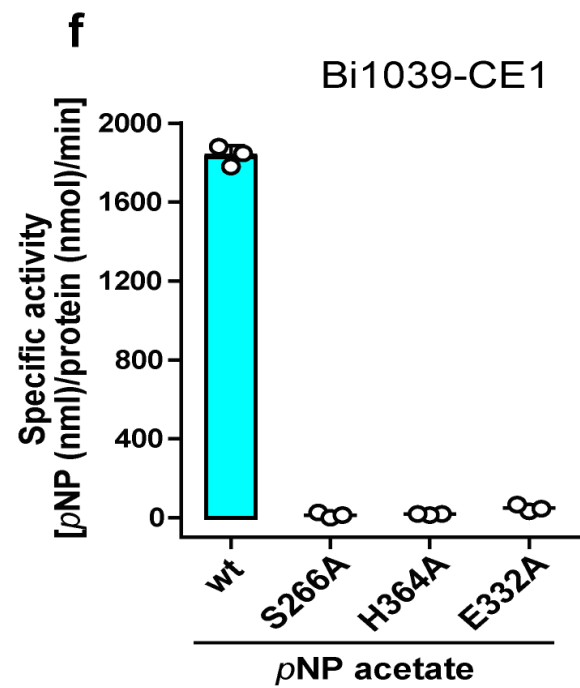
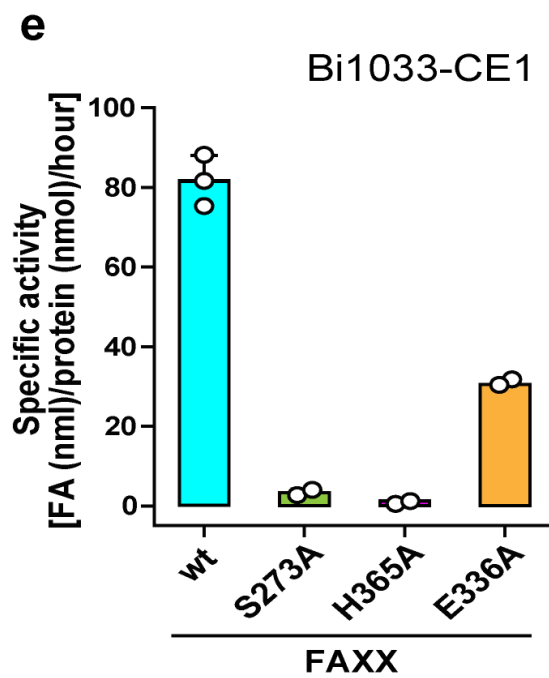
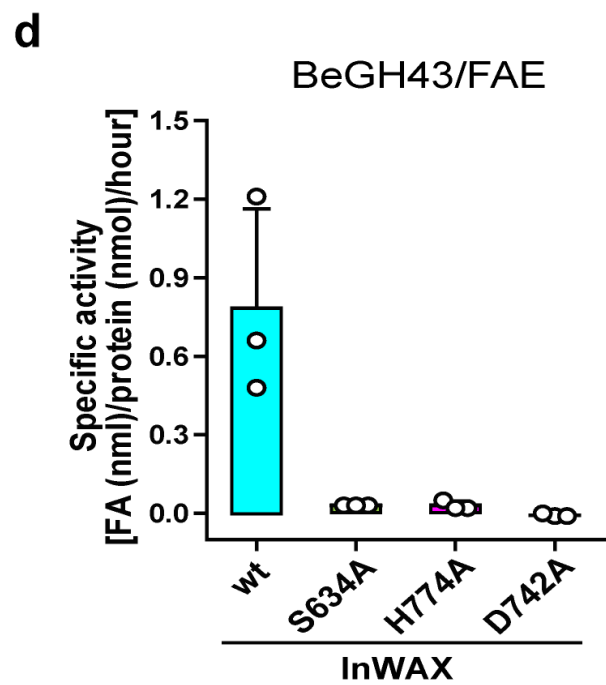
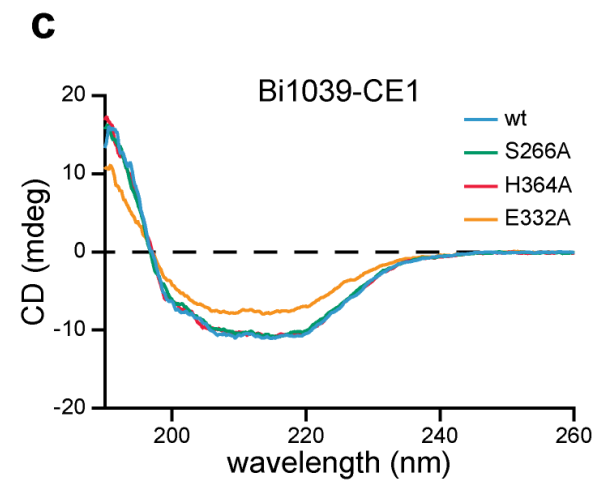
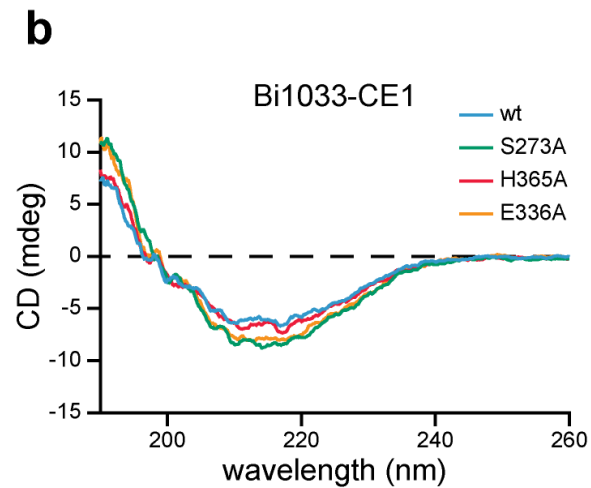
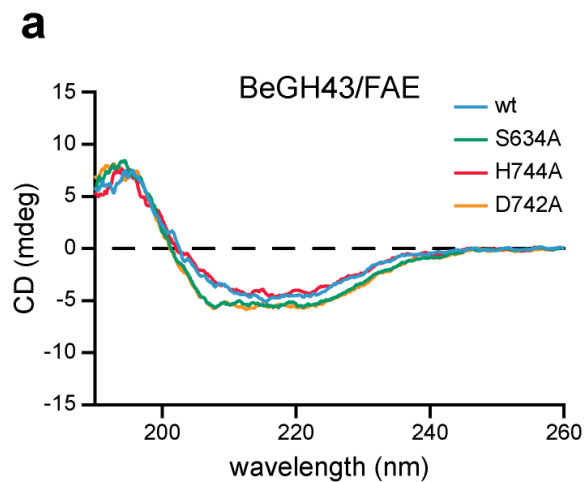
(His365)

**Supplementary Figure 13. Alignments showing that the newly characterized polypeptides associated with the EGE PUL and demonstrated to cleave side-chain ester-linkages from complex arabinoxylans are widely distributed. a**

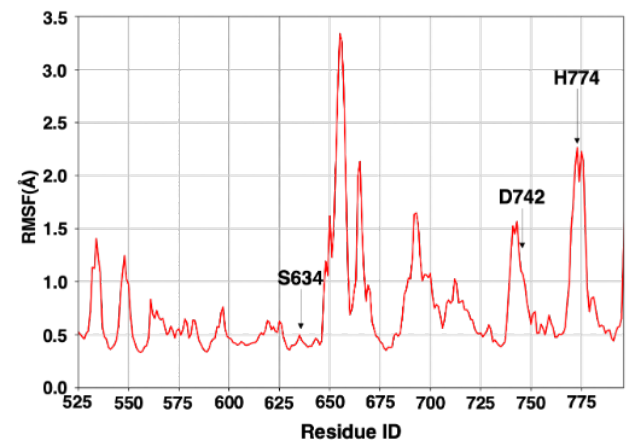
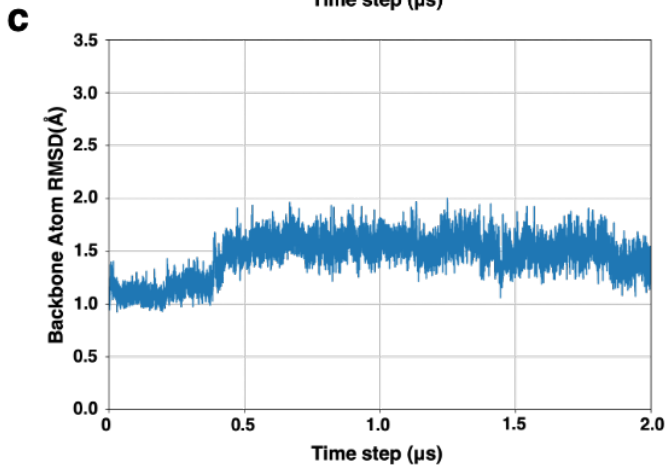
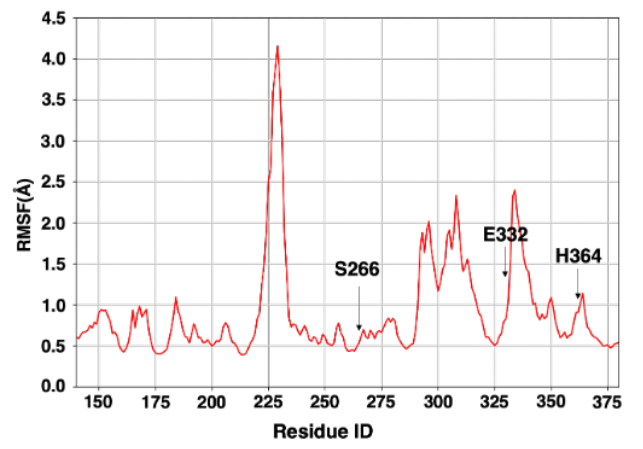
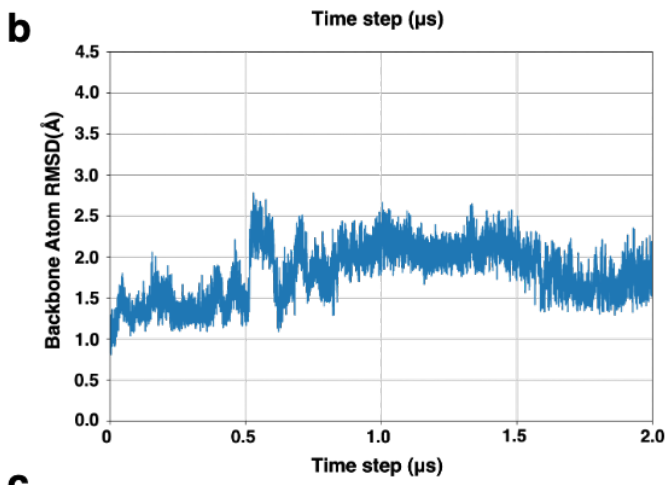
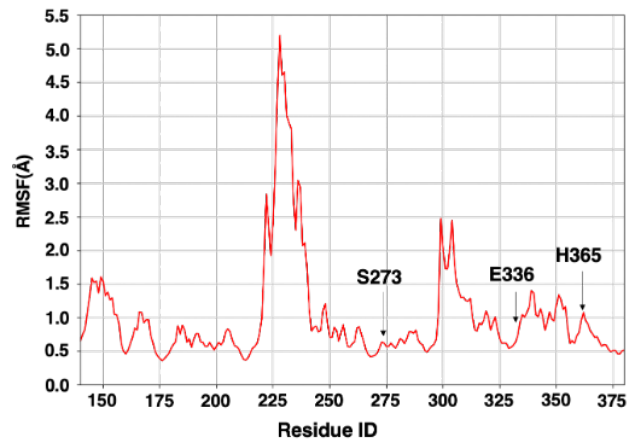
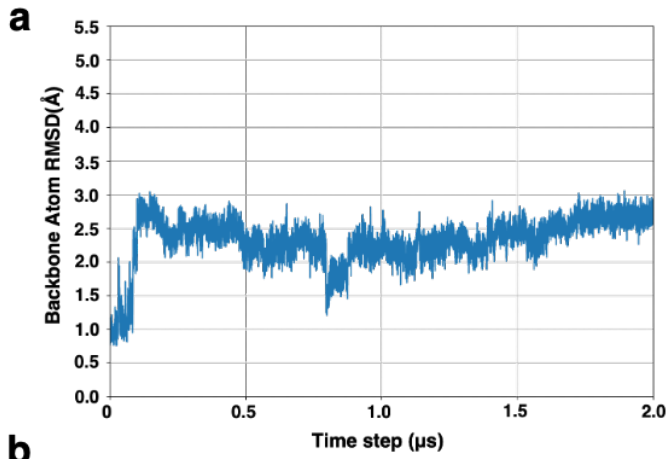
A search was made in the Genbank protein database (<https://www.ncbi.nlm.nih.gov/protein/>) to extract polypeptides with high sequence identities to the *B. intestinalis* homolog (**accession number bolded and italicized-Bi1040-FAE**) and aligned with the CLUSTAL Multiple Alignment program. The Genbank accession numbers and annotations are as follows: SEH83175.1 Acetyl esterase/lipase [*Prevotella ruminicola*]; WP\_006281679.1 alpha/beta hydrolase [*Prevotella bryantii*]; >WP\_100993215.1 alpha/beta hydrolase [*Spirosoma pollinicola*]; WP\_020527766.1 alpha/beta hydrolase [*Flexithrix dorotheae*]; WP\_099150634.1 alpha/beta hydrolase [*Lewinella nigricans*]; WP\_013768612.1 alpha/beta hydrolase [*Haliscomenobacter hydrossis*]; WP\_114463085.1 alpha/beta hydrolase [*Runella* sp. YX9]; WP\_015029998.1 alpha/beta hydrolase [*Emticicia oligotrophica*]; WP\_055151666.1 alpha/beta hydrolase [*Jiulongibacter sediminis*]; WP\_006799192.1 alpha/beta hydrolase [*Dysgonomonas gadei*]; WP\_027452672.1 alpha/beta hydrolase [*Prevotella albensis*]; KQB43445.1 Esterase/lipase [*Flavobacterium daejeonense*]; WP\_071144936.1 alpha/beta hydrolase [*Bacteroides ihuae*]; WP\_100615198.1 alpha/beta hydrolase [*Confluentibacter citreus*]; WP\_123396545.1 alpha/beta hydrolase [*Muribaculaceae* bacterium Isolate-102 (HZI)]; WP\_117741097.1 glycoside hydrolase [*Bacteroides stercoris*]; >WP\_050793236.1 glycoside hydrolase [**bolded-Bacteroides eggerthii**]; WP\_024996568.1 alpha/beta hydrolase [*Bacteroides graminisolvens*]; >WP\_018709923.1 alpha/beta hydrolase [*Bacteroides barnesiae*]; EDV05955.1 hypothetical protein BACINT\_01040-FAE [*Bacteroides intestinalis*]; and WP\_007216415.1 alpha/beta hydrolase [*Bacteroides cellulosilyticus*]. The *B. eggerthii* and *B. stercoris* polypeptides are from the C-terminal esterase module of the bi-functional enzyme (GH43/FAE). The crystal structure of the *B. eggerthii* bi-functional enzyme (**bolded**) was solved in this study.

**b** A search was made in the Genbank protein database to extract polypeptides with high sequence identities to the *B. intestinalis* homolog (**accession number bolded-Bi1039-CE1**) and aligned with the CLUSTAL Multiple Alignment program. The Genbank accession numbers and annotations are as follows: PZX17224.1 enterochelin esterase family protein [*Cytophaga xylanolytica*]; WP\_117954567.1 esterase [*Bacteroides stercoris*]; WP\_007661004.1 esterase [*Bacteroides intestinalis*]; WP\_007216414.1 esterase [*Bacteroides cellulosilyticus*]; WP\_123120797.1 esterase [*Ferruginibacter* sp. BO-59]; WP\_066034169.1 esterase [*Flavobacterium anhuiense*]; WP\_055090457.1 esterase [*Flavobacterium daejeonense*]; WP\_035660712.1 esterase [*Flavobacterium seoulense*]; WP\_083552732.1 esterase [*Flavobacterium flevense*]; WP\_066330097.1 esterase [*Flavobacterium glycinis*]; WP\_068704550.1 esterase [*Paludibacter jiangxiensis*]; WP\_020531203.1 hypothetical protein

[*Flexithrix dorotheae*]; WP\_016778581.1 esterase [*Anaerophaga thermohalophila*]; WP\_106153941.1 esterase [*Marinilabilia salmonicolor*]; WP\_053183372.1 esterase [*Sunxiuqinia dokdonensis*]; WP\_045027037.1 esterase [*Draconibacterium sediminis*]; and WP\_038555069.1 esterase [*Draconibacterium orientale*]. The accession number of the *B. intestinalis* enzyme with the crystal structure solved in this study is bolded. **c** A search was made in the Genbank protein database to extract polypeptides with high sequence identities to the *B. intestinalis* homolog (**accession number bolded-Bi1033-CE1**) and aligned with the CLUSTAL Multiple Alignment program. The Genbank Accession numbers and their annotations are as follows: WP\_062122566.1 esterase [*Geofilum rubicundum*]; WP\_071144963.1 esterase [*Bacteroides ihuae*]; WP\_004293186.1 carbohydrate-binding protein [*Bacteroides eggerthii*]; WP\_101690191.1 esterase [*Dysgonomonas* sp. Marseille-P4356]; WP\_007660993.1 hypothetical protein [*Bacteroides intestinalis*, Bi1033-CE1]; WP\_082717889.1 esterase [*Bacteroides cellulosilyticus*]; WP\_081743618.1 esterase [*Bacteroides timonensis*]; WP\_123396861.1 esterase [Muribaculaceae bacterium Isolate-102 (HZI)]; WP\_022104001.1 carbohydrate-binding protein [*Bacteroides stercoris*]; and WP\_026314837.1 esterase [*Bacteroides gallinarum*]. The accession number of the *B. intestinalis* enzyme with the crystal structure solved in this study is bolded. The amino acids that identical (shaded black) or similar (grey) in >50% of the positions are indicated. Amino acids with similar properties are grouped as LIMV, AG, YWF, DEQN, KRH and ST. The catalytic triad in each enzyme family are highlighted in red. The N-terminal regions are not shaded, as these regions represent signal peptides for secretion and are unlikely to be involved in the folding of the functional proteins.

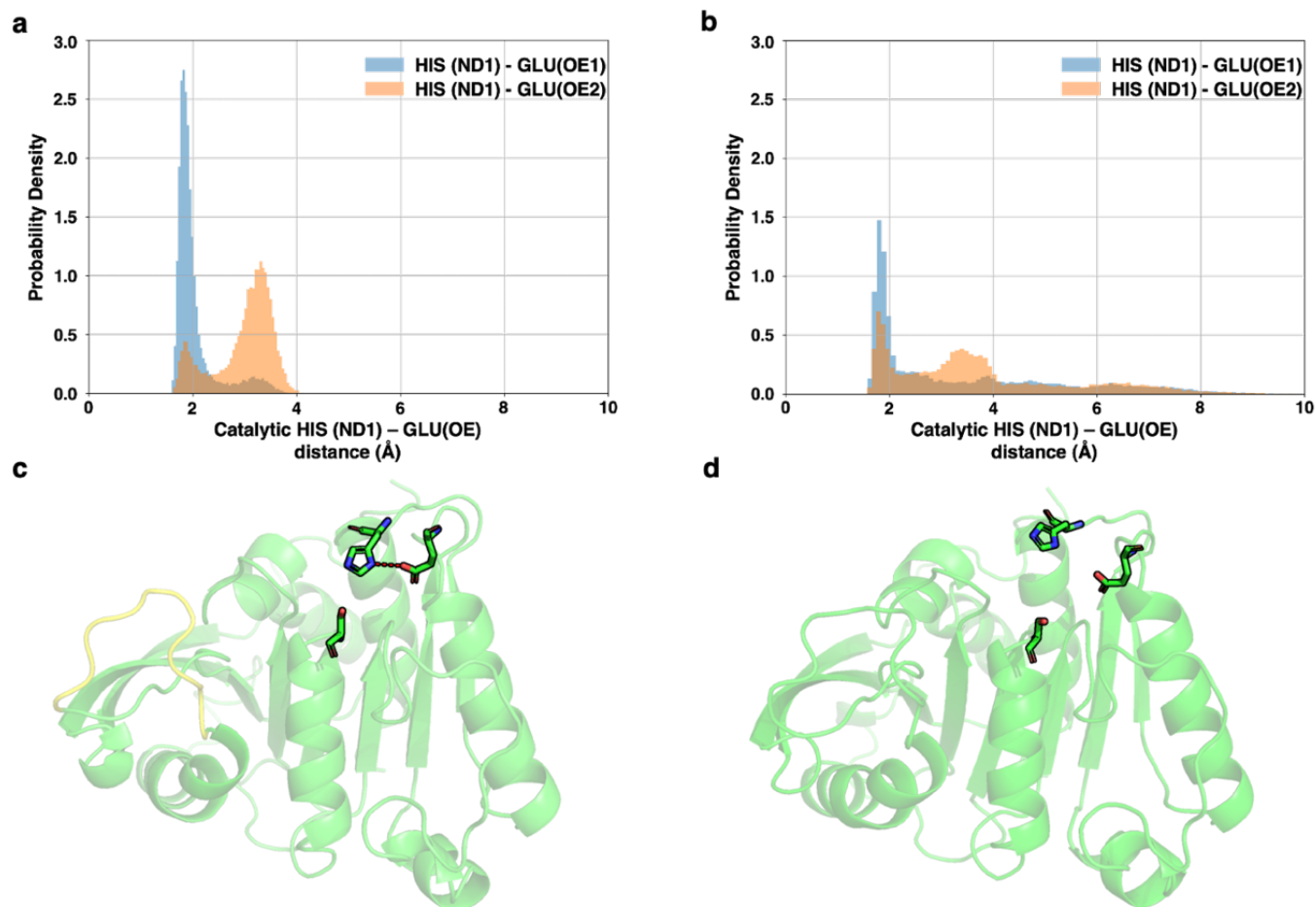


**Supplementary Fig. 14. Mutational analysis of the putative catalytic triads.** The residues (Ser-His-Glu/Asp) that constitute the proposed catalytic triad in BeGH43-Fae (S634, H774, D742), Bi1033-CE1 (S273, H365, E336) and Bi1039-CE1 (S266, H364, E332) were substituted with alanine by site-directed mutagenesis. Circular dichroism scans of mutant recombinant proteins of **a** BeGH43/FAE **b** Bi1033-CE1 and **c** Bi1039-CE1 compared to their wild-type counterparts. Proteins (0.1 mg/ml) were dialyzed in 10 mM potassium phosphate buffer pH 7.5 and CD spectra were collected at 25°C from 190 nm to 260 nm at a speed of 50 nm/min with a 0.1-nm wavelength pitch, with five accumulations. Relative feruloyl esterase activity of the mutant **d** BeGH43/FAE **e** Bi1033-CE1 and **f** Bi1039-CE1 proteins presented as a percentage of the activity of the corresponding wild-type proteins on insoluble wheat arabinoxylan (InWAX), the feruloylated trisaccharide (FAXX), and the synthetic substrate, para-nitrophenyl acetate (*p*NP Acetate), respectively. The bars are means  $\pm$  standard deviations. In **d** and **f**,  $n = 3$  independent reactions. In **e**,  $n = 2$  independent reactions. The source data underlying Supplementary Figs. 14d-f are provided in the Source Data file.

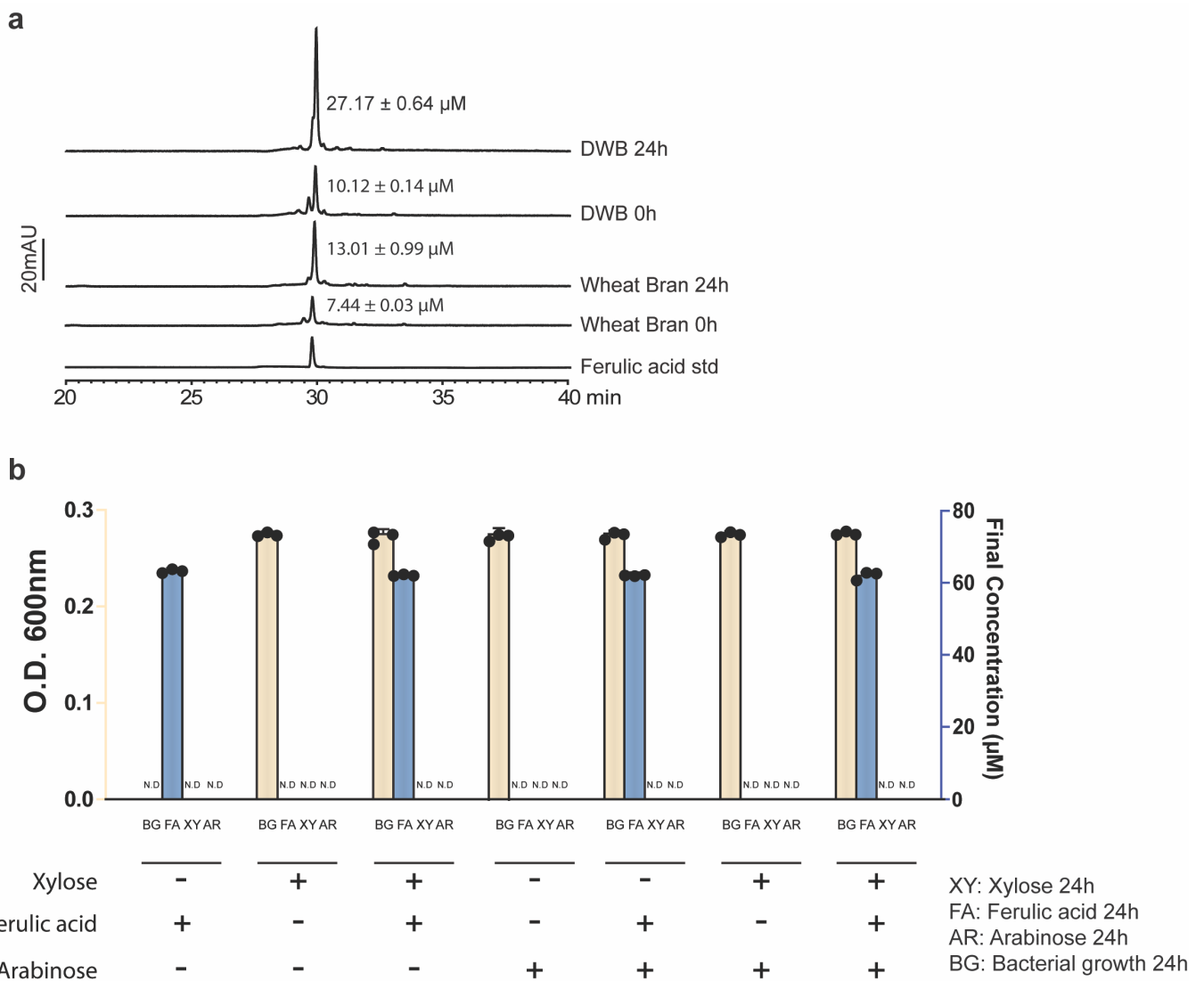




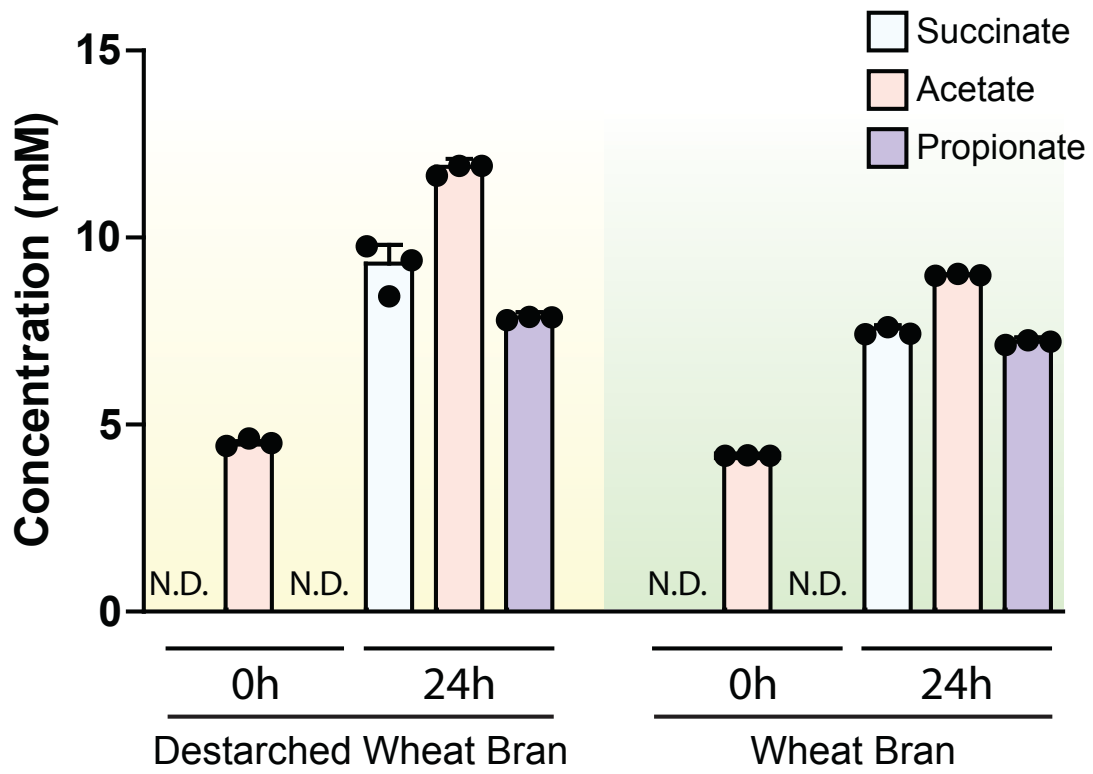
**Supplementary Fig 15. RMSD and RMSF analysis of esterase domains of three esterase enzymes** (Bi1033-CE1, Bi1039-CE1, and BeGH43/FAE in **a**, **b**, and **c**, respectively). The three figures in the left panel show root mean square deviations of all three proteins. The three figures in the right panel represent root mean square fluctuations of all three proteins. Approximate positions of catalytic triad residues are indicated by an arrow. The analysis was done on 2  $\mu$ s of simulation data on each protein. For Bi1033-CE1 and Bi1039-CE1, first 15 residues were omitted from calculations because of the large fluctuation of these residues.



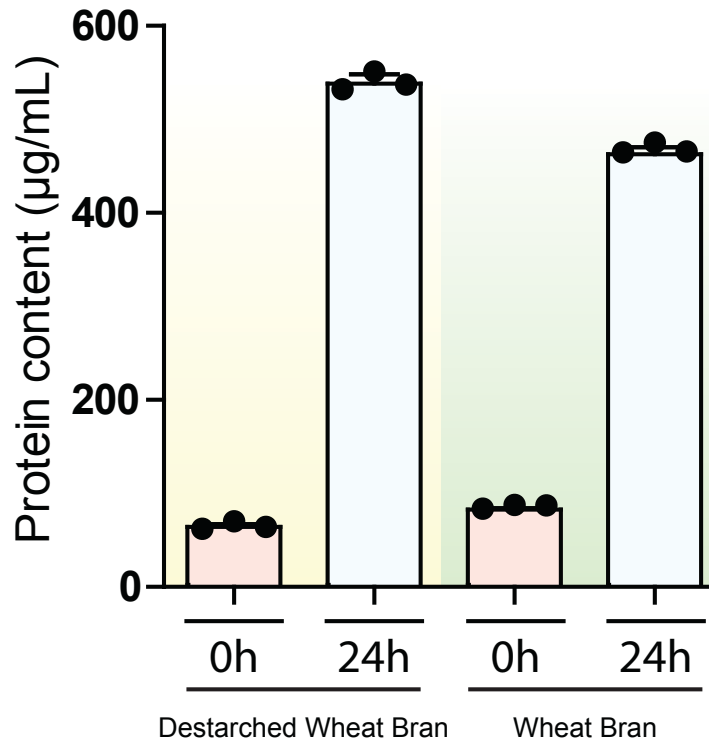
**Supplementary Fig. 16. Higher stability in acid-base hydrogen bonding may lead to higher catalytic activity.** **a, b** Distance between HIS (ND1) and GLU (OE1, color: Blue; OE2, color: Orange) atoms are represented as a 1-D histogram with increasing mean distance from left (Bi1033-CE1) to right (Bi1039-CE1). **c, d** Cartoon representation of esterase domain of Bi1039-CE1. Catalytic triad residues are shown as stick. Hydrogen Bond between HIS-GLU residue are shown as red dot. In **c**, hydrogen bond between base (HIS) and acid exists but in **d**, base (HIS) moves away from the acid molecule.



**Supplementary Fig. 17. Accumulation of ferulic acid in the spent medium of *B. intestinalis* during growth on arabinoxylan.** **a** HPLC-DAD chromatograms of *B. intestinalis* spent medium during growth on wheat bran or de-starched wheat bran (DWB) as sole carbon source, showing accumulation of ferulic acid in the medium after bacterial culture. The y-axis is absorbance in milli-absorbance units (mAU). **b** Growth of *B. intestinalis* in minimal medium containing ferulic acid as sole carbon source or supplemented with different monosaccharides from arabinoxylan, suggesting the lack of utilization of the phenolic compound by the bacterium. BG: bacterial growth; FA: ferulic acid; XY: xylose; AR: arabinose. In **a**, the results are presented as the mean  $\pm$  the standard deviation of three independent reactions ( $n=3$ ), and the traces are representatives of the three different BG reactions. In **b**, the bars indicate means  $\pm$  standard deviations of three independent reactions ( $n=3$ ). ND: not detected. The source data underlying Supplementary Fig. 17**a-b** are provided in the Source Data file.

**a**Fermentation end products *B. intestinalis***b**

## Protein content in medium



**Supplementary Fig. 18. The end products during growth of *B. intestinalis* in a minimal medium containing wheat bran or de-starched wheat bran as a sole carbon source. a** The major end products of fermentation of *B. intestinalis* cells grown on either substrate for 24 hours. **b** Protein accumulation in the medium containing either substrate after culturing with *B. intestinalis* cells for 24 hours. In **a** and **b**, the bars indicate means  $\pm$  standard deviations of three independent reactions (n=3). The source data underlying Supplementary Figs. 18**a-b** are provided in the Source Data file.

**Supplementary Table 1: X-ray data collection and refinement statistics.**

Protein	Bi1033 (dimer/ASU)	Bi1033 (monomer/ASU)	BeGH43-FAE	Bi1039
PDB ID	6MOU	6MOT	6MLY	6NE9
Resolution (Å)	76.37 - 2.24 (2.32 - 2.24)	86.57 - 1.71 (1.77 - 1.71)	48.16 - 2.7 (2.8 - 2.7)	28.6 - 1.74 (1.8 - 1.74)
Space group	P 3 <sub>1</sub> 2 1	P 6 <sub>4</sub> 2 2	C 1 2 1	C 2 2 2 <sub>1</sub>
Unit cell (Å)	a=b=95.2, c=202.6; $\gamma$ =120	a=b=99.9, c=164.204; $\gamma$ =120	a=254.7,b=93.3, c=214.0, $\beta$ =123.38	a=116.7, b=126.6, c=110.3
Total reflections	302305 (30106)	2143630 (125664)	383527 (32441)	145030 (2851)
Unique reflections	52004 (5143)	52998 (5201)	113690 (10279)	76952 (2380)
Multiplicity	5.8 (5.9)	40.4 (24.2)	3.4 (3.2)	1.9 (1.2)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	98.8 (90.1)	91.7 (28.6)
Mean I/sigma(I)	7.5 (1.9)	19.3 (2.3)	8.0 (1.2)	15.3 (2.0)
Wilson B-factor	30.9	24.6	49.0	18.2
$R_{\text{merge}}$	0.18 (0.78)	0.14 (1.02)	0.16 (1.03)	0.06 (0.23)
$R_{\text{meas}}$	0.20	0.15	0.19	0.09
CC1/2	0.99 (0.57)	0.99 (0.89)	0.98 (0.52)	0.99 (0.21)
CC*	0.99 (0.85)	1 (0.97)	0.99 (0.83)	0.99 (0.59)
$R_{\text{work}}$	0.19 (0.30)	0.17 (0.34)	0.22 (0.29)	0.18 (0.28)
$R_{\text{free}}$	0.24 (0.38)	0.19 (0.33)	0.27 (0.36)	0.22 (0.35)
Number non-hydrogen atoms	6163	3080	24144	6165
macromolecules	5702	2758	23692	5568
ligands	16	16	56	110
water	445	306	396	487
Protein residues	713	345	3029	705
RMS(bonds) Å	0.01	0.01	0.01	0.004
RMS(angles) °	1.2	1.2	1.3	0.9
Ramachandran favored (%)	97	97	93	99
Ramachandran outliers (%)	0	0	0.27	0
Clashscore	6.13	2.38	4.32	2.59
Average B-factor	32.7	32	64.8	24.1
macromolecules	32.4	30.6	64.9	23.1
ligands	39.8	40.8	57.9	35.7
solvent	36	44.3	62.2	33

**Supplementary Table 2.** Comparison of the secondary structure the wild-type and mutant BeGH43/FAE, Bi1033-CE1 and Bi1039-CE1 proteins.

	Helix (%)	Strand (%)	Turn (%)	Unordered (%)
<b>BeGH43/FAE</b>				
wt	31.0 ± 0.00	25.5 ± 0.70	19.0 ± 2.82	24.5 ± 2.12
S634A	31.5 ± 0.70	25.0 ± 2.82	18.5 ± 2.12	25.0 ± 4.24
H744A	30.5 ± 0.70	27.0 ± 0.00	21.0 ± 1.41	21.5 ± 2.12
D742A	31.5 ± 2.12	27.0 ± 1.41	20.5 ± 0.70	21.0 ± 2.82
<b>Bi1033-CE1</b>				
wt <sup>α</sup>	27.3 ± 1.52	30.0 ± 2.64	15.3 ± 1.52	27.3 ± 5.03
S273A <sup>α</sup>	31.0 ± 2.64	31.3 ± 0.57	14.0 ± 1.00	23.6 ± 3.51
H365A <sup>α</sup>	24.0 ± 2.64	35.6 ± 6.50	16.6 ± 3.51	23.6 ± 6.65
E336A <sup>α</sup>	28.0 ± 3.00	34.0 ± 3.00	12.3 ± 2.51	25.6 ± 3.51
<b>Bi1039-CE1</b>				
wt <sup>α</sup>	30.6 ± 1.52	30.6 ± 1.15	11.3 ± 1.15	27.3 ± 0.57
S266A <sup>α</sup>	30.6 ± 2.08	30.0 ± 1.73	10.6 ± 2.08	28.6 ± 1.52
H364A <sup>α</sup>	31.0 ± 1.00	29.6 ± 1.15	12.0 ± 1.00	27.3 ± 0.57
E332A <sup>α</sup>	30.3 ± 3.51	31.6 ± 3.21	10.0 ± 1.00	28.0 ± 0.99

<sup>α</sup>: No statistical significant difference between each mutant and the wild-type (wt). Statistical analyses were not carried out for the BeGH43/FAE data because they come from only two readings. The source data underlying Supplementary Table 2 are provided in the Source Data file.

**Supplementary Table 3:** Primers used to amplify the genes encoding putative esterase and putative glycosyl hydrolases.

Primer (5'-3')	Gene	Sequence <sup>a</sup>
Bi1033For	BACINT_01033	catatgCAACAGCAAGATTTTCCGGCAGGAAC
Bi1033Rev		ctcgagTTATTTTCGTTTTAAATAATAGGGGAGCAAATTCCTTTCAGGC
Bi1035For	BACINT_01035	catatgCAAATCGGCACTCCATACATCCACGATC
Bi1035Rev		ctcgagCTAATGGTCGCGGAAATTCATTTGGAATTG
Bi1038For	BACINT_01038	catatgTTGAATAGAATGAAAAAGCTGTTATTATTTATCGCATGCTTG
Bi1038Rev		ctcgagTTACTTCTTAAATAAATTCGGTAAAAATTCATTACAGACATCTGC
Bi1039For	BACINT_01039	catatgCAGACAGTGGAGGATTTCAAACCATCG
Bi1039Rev		ctcgagTTATTTAAAAAGAAGCGGAGCAAACCTCATTCAATGC
Bi1040For	BACINT_01040	catatgCAGATTACGCAATGGACTGATATCAACTATGC
Bi1040Rev		ctcgagTTAATAAAGGGTATAAAACTGTTTTTGAGGAGGATTTTTTC
Bi1041For	BACINT_01041	catATGAAGATACTGTTTCATTTACAATAACTCTGTTCCG
Bi1041Rev		ctcgagCTAAGGATGATATTTCCCAGTATTCTAATTTTGTCTC
Bi1042For	BACINT_01042	catatgCAAACATTGCCGTATCAGAATCCTGAACTAAG
Bi1042Rev		ctcgagTTATTGTAAAGTGACTTTGACAGATTGCAGGTC
Bi1043For	BACINT_01043	catatgCAGAATCCCATTATTACGGATCAGTTCACTG
Bi1043Rev		ctcgagTTATTGAAAAGTATCCAGTCGATTTCAACTTTACC
BeGH43/FaeFor	HMPREF1016_RS0111555	ccgggatccCAAAGCCTGCAACTAATCCTGTGA
BeGH43/FaeRev		ccgctcgagTCATTTAATGTCATCGCATTTTATCGGCC



**Supplementary Table 4.** Primers used for Reverse Transcriptase quantitative PCR in *B. intestinalis*, *B. cellulosilyticus*, and *B. oleiciplenus*.

Primer (5'-3')	Gene	Sequence
Bi1036For	BACINT_01036	TGATTCTAACTACACTCTCTTTGGTG
Bi1036Rev		TTGACAGGAGTAACGTTTCATGTAAGC
Bi1037For	BACINT_01037	TATCAATCATGAGTTTGCTCACAAAATG
Bi1037Rev		TTCGTCAAATTCTTTATCTGCGCG
Bi16sFor	Bi16s rDNA	GGAGCGTAGGCGGATTATTAAG
Bi16sRev		GGAGCGTAGGCGGATTATTAAG
Bc2149For	BACCELL_02149	GAATGGGCAGGTAACAGAAA
Bc2149Rev		GGATGGCAACCGTACAGATAG
Bc2148For	BACCELL_02148	CACTCCAGACGAGTTCTTATAC
Bc2148Rev		GGCCTTCTGTACTCTTGATACC
Bc16sFor	Bc16s rDNA	AGCAACACAATGCTATG
Bc16sRev		CACGTAAACCACTTTCTT
Bo2534For	HMPREF9447_02534	CGTATCCGGCTTCACCTATTC
Bo2534Rev		GTCGAAACCTACACCCATATCC
Bo2533For	HMPREF9447_02533	GATGCGCTGGCATAACAATAC
Bo2533Rev		CTCTACTCGGAATCGGAAGAATG
Bo16sFor	Bo16s rDNA	CCATTCATTGGGCATAG
Bo16sRev		CGTACTTTCTTACCGATAC

**Supplementary Table 5.** Primers used for site directed mutagenesis in BeGH43/FAE, Bi1033-CE1, and Bi1039-CE1.

Primer (5'-3')	Gene	Sequence <sup>a</sup>
BeGH43/Fae S634A For	BeGH43/Fae	TACCGGCTTT <b>gca</b> TCGGGTGGAC
BeGH43/Fae S634A Rev		ATTCCTATAAAAGAAGTATCCAACCTATACTTGGC
BeGH43/Fae D742A For	BeGH43/Fae	TGGTGAAGCG <b>gcc</b> ACTGTGGTTC
BeGH43/Fae D742A Rev		TGAATAACAATAAATTTAGGATCGTTTTTATCTATATAAGTAATG
BeGH43/Fae H774A For	BeGH43/Fae	CGGCGGACAA <b>gct</b> GGTCCTGTCCAC
BeGH43/Fae H774A Rev		GGAACAGAAATAAATTCTTCCAAAC
Bi1033-CE1 S273A For	Bi1033-CE1	GGCCGGACTT <b>gcc</b> TGGGGTGGAC
Bi1033-CE1 S273A Rev		ATGGCGCGATTATCCCTATCGGTCAG
Bi1033-CE1 E336A For	Bi1033-CE1	CGGTTCCGAA <b>gca</b> CATCCGGAAA
Bi1033-CE1 E336A Rev		ATGCCCAGGAAGAAGACG
Bi1033-CE1 H365A For	Bi1033-CE1	GGGCACTGCC <b>gcc</b> GAGTTCCTCA
Bi1033-CE1 H365A Rev		GGCGATTGTAATAGATCG
Bi1039-CE1 S266A For	Bi1039-CE1	TGCAGGATT <b>Ggcg</b> ATGGGGGCTA
Bi1039-CE1 S266A Rev		ATGGCGCGGTCTTCACGAC
Bi1039-CE1 E332A For	Bi1039-CE1	TGGTACGGCC <b>gca</b> CCGCATCCTT
Bi1039-CE1 E332A Rev		AGTCCCAGAAAGAAAACCTTAAACTGAACGTTTAC
Bi1039-CE1 H364A For	Bi1039-CE1	CGATACGGCC <b>gct</b> GAATGGCTCAC
Bi1039-CE1 H364A Rev		GGAGATTCATAATACACATATTTAATTC

a: Nucleotides mutated are in lowercase and bolded

## References:

- 1 Colquhoun, I. J., Ralet, M. C., Thibault, J. F., Faulds, C. B. & Williamson, G. Structure identification of feruloylated oligosaccharides from sugar-beet pulp by NMR spectroscopy. *Carbohydr. Res.* **263**, 243-256 (1994).