Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*

Ana S. Luis[1¶], Jonathon Briggs[1¶], Xiaoyang Zhang[1¶], Benjamin Farnell[2,3], Didier Ndeh[1], Aurore Labourel[1], Arnaud Baslé[1], Alan Cartmell[1], Nicolas Terrapon[4], Katherine Stott[5], Elisabeth C. Lowe[1], Richard McLean[2], Kaitlyn Shearer[2], Julia Schückel[6], Immacolata Venditto[1], Marie-Christine Ralet[7], Bernard Henrissat[4,8,9], Eric C. Martens[10], Steven C. Mosimann[3], D. Wade Abbott[2,3]* and Harry J. Gilbert[1,*]

[1]*Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, U.K.*

[2]*Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Alberta, Canada, T1J 4B1*

[3]*Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada, T1K 6T5*

[4]*Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique (CNRS), Aix-Marseille University, F-13288 Marseille, France*

[5]*Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom*

[6]*Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Copenhagen, Denmark*

[7]*INRA, UR1268 Biopolymères Interactions Assemblages, 44300 Nantes, France*

[8] *INRA, USC 1408 AFMB, F-13288 Marseille, France*

[9]*Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia*

[10]*Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA*

¶These authors contributed equally.*To whom correspondence should be addressed: D. Wade Abbott (wade.abbott@agr.gc.ca); Harry J. Gilbert (harry.gilbert@ncl.ac.uk),

The major nutrients available to human colonic *Bacteroides* species are glycans exemplified by pectins, a network of covalently linked plant cell wall polysaccharides containing galacturonic acid (GalA). Metabolism of complex carbohydrates by the *Bacteroides* genus is orchestrated by polysaccharide utilisation loci or PULs. In *Bacteroides thetaiotaomicron*, a human colonic bacterium, the PULs activated by the different pectin domains have been identified, however, the mechanism by which these loci contribute to the degradation of these GalA-containing polysaccharides is poorly understood. Here we show that each PUL orchestrates the metabolism of specific pectin molecules, recruiting enzymes from two previously unknown glycoside hydrolase (GH) families. The apparatus that depolymerizes the backbone of rhamnogalacturonan-I (RGI) is particularly complex. This system contains several GHs that trim the remnants of other pectin domains attached to RGI, while nine enzymes contribute to the degradation of the backbone compring of a rhamnose-GalA repeating unit. The catalytic properties of the pectin degrading enzymes are optimized to protect the glycan cues that activate the specific PULs ensuring a continuous supply of inducing molecules throughout growth. The contribution of *Bacteroides* spp. to the metabolism of the pectic network is illustrated by cross-feeding between organisms.

The human gut microbiota (HGM) impacts on host physiology and health[1,2]. Understanding the mechanisms of nutrient acquisition by the HGM, exemplified by glycan metabolism[3-5], underpins the development of probiotic and prebiotic strategies that maximize human health. While glycan acquisition by human colonic *Bacteroides* species is well established[6-9], it should be emphasised that Firmicutes are more abundant in the HGM of Western populations, however, the mechanism by which they metabolise complex carbohydrates is less well understood[3]. Indeed, it is likely that Firimicutes make a substantial contribution to the degradation of dietary and host glycans in the HGM. The glycan degrading systems of *Bacteroidetes* are encoded by polysaccharide utilization loci (PULs) that are activated by the target carbohydrate[3]. These systems comprise surface glycan binding proteins (SGBPs), outer membrane oligosaccharide transporters; SusC and SusD homologues ($SusC_H$ and $SusD_H$, respectively), and surface and periplasmic carbohydrate active enzymes (CAZymes) that are grouped into sequence based families in the CAZy database[10]. Relevant to this work are glycoside hydrolase (GH) and polysaccharide lyase (PL) families[11].

Pectins are D-galacturonic acid (D-GalA) rich plant cell wall polysaccharides that are abundant in fruits and vegetables. The two major pectins (see[12] for review) are homogalacturonan (HG) and rhamnogalacturonan-I (RGI) (**Fig. 1a**). HG comprises

α-1,4-linked D-GalA and the backbone of RGI is a repeating unit of the disaccharide α-1,2-L-rhamnose (Rha)-α-1,4-D-GalA. Depending on the plant species, the RGI backbone is decorated with galactans (β-1,4-D-galactose (D-Gal) units) and/or arabinans (α-1,5-linked L-arabinofuranose (L-Ara*f*) units with additional L-Ara*f* side-chains)[13]. The backbones of HG and RGI are covalently linked[14]. Although individual microbial pectin degrading enzymes have been described[15], the mechanism by which these biocatalysts participate in the concerted degradation of intact pectin remains opaque. *Bacteroides thetaiotaomicron*, a member of the HGM, utilizes all known pectins structures and discreet PULs activated by these glycans have been identified[16]. Here we have characterized the function of PULs associated with pectin metabolism, and explored how they contribute to interactions within the HGM foodweb. The data show how these loci coordinate the complex degradative interactions between the backbone and oligosaccharide decorations of these acidic polysaccharides.

Results

The PULs that orchestrate pectin degradation. Transcriptomic data[16] revealed the PULs upregulated in response to arabinan (*bt0348-bt0369*, Ara-PUL), galactan (*bt4667-bt4673*, Gal-PUL), RGI backbone (*bt4145-bt4183*, RGI-PUL) and HG (*bt4108-bt4124*, HG-PUL) (**Fig. 1b**). To determine the mechanisms by which these loci mediate pectin degradation, the biochemical functions of recombinant proteins encoded by the PULs were determined (**Supplementary Tables 1-5**).

Cell surface degradation, substrate binding and import. Initial degradation of the pectins by *B. thetaiotaomicron* is mediated by endo-acting CAZymes on the surface of the bacterium (**Fig. 2c** and **Supplementary Fig. 1**). These enzymes are essential for pectin utilization as they generate glycans with an appropriate degree of polymerization (DP) for transport into the periplasm[17]. Consistent with this premise, deletion of the genes encoding the single outer membrane endo-acting enzymes encoded by RGI-PUL (BT4170) and Gal-PUL (BT4668), **Supplementary Fig. 2** and **Fig. 2c**, prevented growth on the respective pectin (**Fig. 2a**). The surface location of these enzymes was consistent with whole cell assays of *B. thetaiotaomicron* under aerobic conditions (**Fig. 2b**), which report only the activity of surface proteins. To

explore the function of the rhamnogalacturonan lyase BT4170, a key component of the RGI degrading apparatus, the crystal structure of the enzyme was determined in complex with ligands. The data (**Supplementary Fig. 3**) showed that the catalytic apparatus of BT4170 and a HG lyase (Pel9A, 1RU4) both located in family PL9, comprising a Brønstead base (Lys285 in BT4170) and a calcium, was conserved. Specificity determinants were identified in subsites distal to the active site, explaining why Pel9A and BT4170 target distinct substrates (see **Supplementary Discussion** and **Supplementary Fig. 3**).

Ara-PUL and HG-PUL each encode two surface enzymes. The enzymes derived from Ara-PUL, BT0360 and BT0367, are α1,5-arabinanases that display endo- and endo-processive activity, respectively (**Supplementary Fig. 4** and **Supplementary Discussion**). Only *Δbt0367* led to the loss of the arabinan utilization phenotype (**Fig. 2a**). Gene deletion studies showed that of the two surface PLs (BT4116 and BT4119, **Supplementary Fig. 4** and **Supplementary Table 2**) encoded by HG-PUL, only BT4116 was essential for growth on HG (**Fig. 2a**). The functional significance of BT0360 and BT4119 are unclear, but may reflect the targeting of substrates not evaluated here.

Gene deletion studies explored the functional significance of the $SusC_H/SusD_H$ pairs encoded by each pectin PUL (**Supplementary Fig. 5**). In HG-PUL, which contains two $SusC_H/SusD_H$ pairs, only the *Δbt4114* mutant displayed no growth on HG, indicating that BT4114 plays a key role in the import of this pectic glycan. Similarly, the Ara-PUL encodes two pairs of $SusC_H/SusD_H$ transporters (BT0361/BT0362 and BT0363/BT0364)[16]. Deletion of *bt0364*, but not *bt0362*, prevented growth on arabinan. This indicates that only the BT0363/BT0364 complex is capable of transporting arabinooligosaccharide products. The rationale for the presence of two $SusC_H/SusD_H$ pairs in the HG-PUL and Ara-PUL remains unknown, but likely increases access to additional pectins.

SGBPs contribute to glycan degradation by bringing substrates into proximity of membrane bound enzymes[7]. Here a single SGBP encoded by each PUL was identified and shown to be specific for the target polysaccharide (**Supplementary**

Table 5). Only $SusD_H$s encoded by Gal-PUL and HG-PUL displayed affinity for their cognate glycans (**Supplementary Table 5**). The lack of glycan recognition by $SusD_H$s associated with the arabinan and RGI degrading systems suggests that the corresponding $SusC_H$ partner is required for ligand recognition. Recent structural data demonstrate that the tight association of $SusC_H$-D_H pairs[18], supporting the concept that initial ligand recognition can require participation of both protein partners.

Periplasmic degradation of pectins. The oligosaccharides imported into the periplasm were degraded by GHs and PLs. The oligosaccharides generated from galactan and arabinan were depolymerized exclusively by exo-GHs, with HG and the RGI backbone by endo-PLs and exo-GHs (**Supplementary Fig. 2** and **6**).

With respect to galactan degradation only a single GH2 β1,4-galactosidase (BT4667) depolymerized galactooligosaccharides generated by the surface endo-galactanase (**Supplementary Fig. 6** and **Supplementary Table 3**). Surprisingly the *Δbt4667* mutant displayed no growth defect on galactan (**Supplementary Fig. 5**). This may reflect an element of redundancy within the large number of predicted *B. thetaiotaomicron* β-galactosidases[19].

Periplasmic degradation of arabinan-derived oligosaccharides was mediated by three exo-α-L-arabinofuranosidases (**Supplementary Fig. 6**). BT0369 removed α-1,2-L-arabinofuranose side chains[20]. Here we demonstrate the GH51 enzymes BT0348 and BT0368 target arabinan side chains, likely α-1,3-arabinofuranosyl linkages, and the backbone α-1,5-arabinofuranosyl linkages, respectively (**Supplementary Table 4** and **Supplementary Fig. 6**). BT0349 released β-L-arabinose from an arabinan derived oligosaccharide (**Supplementary Fig. 7** and **Supplementary Table 7**). The enzyme reveals a previously unknown GH family now designated GH146 (**Supplementary Fig. 8** and **Supplementary Discussion**).

RGI released from the pectin network contains remnants of arabinan, galactan and HG. Prior to RGI backbone depolymerisation these accessory structures must be removed, explaining why RGI-PUL is so complex. To characterise these accessory

enzymes we used RGI from potato galactan (RGI-P), which contains many of these remnants. Galactan substitutions were cleaved from the RGI-P backbone by the synergistic action of three exo-β-1,4-galactosidases, BT4151, BT4156 and BT4160 (**Supplementary Table 1**). BT4160 targeted galactooligosaccharides, while the other two enzymes released galactose only from RGI-P. The lack of functional arabinofuranosidase genes in RGI-PUL likely reflects the role of single β1,4-D-Gal units in linking arabinan chains to the RGI backbone[13]. Enzyme cocktail data indicate that BT4151 and BT4156 play a pivotal role in exposing the backbone of RGI to enzymatic attack (**Supplementary Fig. 9**). RGI-PUL also encodes the esterase BT4158 (**Supplementary Table 1**), which releases acetyl groups from D-GalA in the RGI backbone, was also shown to be important for the depolymerisation of the glycan (**Supplementary Fig. 9**). A GH28 α-D-galacturonidase (BT4155), which targets HG (**Supplementary Table 1**), removed D-GalA from RGI-P but not from the glycan in Arabidopsis mucilage (RGI-AM), which contains no HG decorations. The crystal structure of BT4155 (**Supplementary Fig. 10**) revealed the expected β-helix for a GH28 enzyme[21]. In the center of the helix is a pocket that houses three carboxylate residues that comprise the predicted catalytic apparatus based on conservation with other GH28 enzymes[22] and mutagenesis data (**Supplementary Table 6**). The pocket extends into a channel-like structure that likely accommodates the conformation adopted by HG but not the RGI backbone.

In addition to enzymes classically associated with pectin degradation, RGI-PUL encodes BT4157, which is located in the apparent "non-pectinase family" GH27. The enzyme (**Supplementary Table 1**) was shown to be a α-galactosidase, which likely targets single α-galactose units that decorate the RGI backbone from Okra plants[23]. Another example of enzyme diversity is the β-D-glucuronidase activity displayed by BT4181 against sugar beet arabinan in which the RGI backbone is known to contain GlcA[24] (**Supplementary Fig. 6**). It is evident, therefore, that the pectin degrading systems are able to accommodate diversity in the fine-chemistry of RGI structures from a variety of plants.

In contrast to exo-cleavage of arabinan and galactan, the backbone of RGI and HG were initially cleaved by endo-PLs and the products depolymerized by exo-GHs (**Fig.**

1 and **Supplementary Fig. 2** and **6**). The different degradative strategy likely reflects the high DP of the imported RGI- and HG-derived oligosaccharides compared to the neutral glycans. Thus, the initial concentration of available substrate for exo-GHs is low, but is increased by the endo-PLs. RGI-PUL encodes three periplasmic PLs, respectively. Of particular note is BT4175, which was shown to accommodate glycans appended to backbone rhamnose units (**Supplementary Fig. 2**), ensuring that cleavage of the $(Rha\text{-}GalA)_n$ polymer occurred in concert with, and not subsequent to, side-chain removal.

The β-elimination of the RGI and HG backbone by the PLs generated Δ4,5-GalA. The unsaturated residues were removed from RGI oligosaccharides with a DP of 2 or ≥4 by BT4176 or BT4174, respectively (**Supplementary Table 1**); and HG oligosaccharides by BT4108 (**Supplementary Table 2**), which expands the activity for the GH105 family. BT4108 products were then depolymerized to GalA by the exo-α-galacturonidase BT4123 (**Supplementary Table 2** and **Supplementary Fig. 6**). The RGI-AM oligosaccharides were degraded through the successive action of a RGI-specific GH106 α-L-rhamnosidase (BT4145) and one of three GH28 rhamnogalacturonidases (BT4146, BT4153 and BT4149) that target $[D\text{-}GalA\textit{p}\text{-}\alpha\text{-}1,2\text{-}L\text{-}Rha\textit{p}]_n$ with a DP of 2, ≥2 or ≥4, respectively (**Supplementary Table 1**). BT4145 cleaved rhamnosidic linkages through an inverting mechanism (**Supplementary Fig. 11**). The biological rationale for galacturonidases that target substrates with different DPs is unclear. Surprisingly deletion of BT4145 only extended lag phase (**Supplementary Fig. 5**), likely reflecting the slow but complete degradation of RGI-AM by the PLs and Δ4,5-unsaturated-α-rhamnogalacturonidases.

The ligands that activate the pectin degradative system. Previously arabinooligosaccharides with DP ≥6 were shown to activate Ara-PUL[16]. Here we determined ligands that bound and activated the hybrid two component system (HTCS) of the pectin PULs. The data (**Supplementary Table 5** and **Supplementary Fig. 12**) demonstrate that the HTCS of Ara-PUL bound linear but not decorated arabinan, and the sensor of the Gal-PUL HTCS (BT4673) recognised small galactooligosaccharides. Only the oligosaccharide Δ4,5GalA-α-1,2-Rha-α-1,4-GalA-α-1,2-Rha, a major limit product of the rhamnogalacturonan lyases, bound the HTCS

(BT4178) that regulates RGI-PUL. Saturated RGI oligosaccharides failed to bind the sensor, indicating that unsaturation of the non-reducing terminal sugar is a recognition determinant. The HTCS that unregulates HG-PUL recognised only saturated HG-derived oligosaccharides. The mRNA levels of the *susC*$_H$ genes of the pectin PULs showed that activation of RGI-PUL resulted in a small up-regulation of HG-PUL and RGII-PUL1 (**Supplementary Fig. 13**). This may reflect the need to extract RGI from pectins networks through cleavage of adjacent HG segments, as a prelude to its degradation.

Exo-acting enzymes that target the remnants of RGI side chains and α-1,3-L-Ara*f* units that decorate arabinan were substantially less active than GHs that depolymerised the backbone of the respective glycans (**Supplementary Tables 1 and 4**). Additionally, BT4108, which removed 4,5ΔGalA from HG-oligosaccharides generating the HTCS activating ligand was slow compared to the other enzymes that act on these pectins (**Supplementary Tables 1 and 2**). The biological rationale for this difference in catalytic competence, may reflect the need to protect the inducing ligand (**Fig. 3)**, as proposed for the chondroitin sulfate utilization system[25]. Slow release of the side chain stubs or unsaturated uronic acids will block the rapid degradation of the backbone ensuring that there is continuous production of the activating molecules throughout growth on the respective glycan.

Pectin utilization within the HGM *Bacteroidetes* and the extent of cross-feeding.

To explore pectin utilization by HGM *Bacteroidetes* growth of the different species on these GalA-rich polysaccharides was determined. The data showed that only *B. ovatus, B. thetaiotaomicron* and *B. finegoldii* utilised all the pectins, although the majority of other organisms could grow on at least some of these glycans (**Supplementary Fig. 14** and **Supplementary Table 8**). Around 70% of the organisms grew on HG and galactan, while only four *Bacteroides* strains utilise the RGI backbone (RGI-AM). However, 56% and 100% of the strains unable to utilise RGI-AM and potato galactan, respectively, grew on the respective oligosaccharides, demonstrating that these organisms utilise pectin degradation products. A key question is the source of oligosaccharides available to these organisms. Evidence of

277 cross-feeding was provided by mutants of *B. thetaiotaomicron* engineered to utilize
278 only pectic oligosaccharides [lacking the surface endo-galactanase (*Δbt4668*) or RGI
279 lyase (*Δbt4170*)], which grew on the cognate polysaccharide when co-cultured with
280 wild type *B. thetaiotaomicron* (**Fig. 4a**). These data show that wild type *B.*
281 *thetaiotaomicron* released polysaccharide breakdown products (PBP; **Fig. 2f**) into
282 culture media, which were available to other organisms. This is consistent with *B.*
283 *uniformis* (grows on galactooligosaccharides but not galactan in mono-culture)
284 utilisation of the polysaccharide when co-cultured with wild type *B. thetaiotaomicron*
285 (**Fig. 4b**). This pectin cross-feeding between *B. thetaiotaomicron* and other
286 *Bacteroides* species, however, is variable. Although *B. massiliensis* utilize HG or RGI
287 oligosaccharides, the bacterium failed to grow on the cognate polysaccharides when
288 co-cultured with *B. thetaiotaomicron* (**Fig. 4b**). This likely reflects the large PBPs
289 generated by *B. thetaiotaomicron* from these pectins, while *B. massilensis* appears
290 to import only RGI and HG oligosaccharides with a low DP. The lack of cross-feeding
291 of some pectin-derived PBPs is evident in arabinan utilization. *B. ovatus* and
292 engineered *B. thetaiotaomicron* (*Δbt0360/Δbt0367*, lacking the two surface endo-
293 arabinanases) both grew on arabinooligosaccharides but not arabinan. The two
294 organisms, however, failed to utilise arabinan when co-cultured with wild type *B.*
295 *thetaiotaomicron* (**Fig 4b**). Although *B. thetaiotaomicron* released
296 arabinooligosaccharides (**Fig. 2f**), the high DP of these molecules (reflects slow
297 activity of the surface endo-arabinanases[20]) may have prevented transport into the
298 periplasm of these organisms.
299
300 **Genetic basis of pectin utilization among the *Bacteroides*.** Loci corresponding to
301 *B. thetaiotaomicron* Gal-PUL in other *Bacteroides* species (**Supplementary Fig. 15**)
302 contained an additional ORF, which, in the *B. ovatus* Gal-PUL, encodes a β-
303 galactosidase (BACOVA_05493) with a retaining mechanism (**Supplementary Fig.**
304 **16**) that belongs to a previously unknown CAZy family (assigned GH147). The
305 enzyme was particularly active against galactohexaose and galactan
306 (**Supplementary Fig. 6** and **Supplementary Table 3**). The importance of this
307 enzyme is illustrated by the severe growth defect displayed by *Δbacova_05493* on
308 galactan (**Fig. 2)**. Whole cell assays with galactan revealed the accumulation of
309 galactose and not galactooligosaccharides, as occurs in *B. thetaiotaomicron*,

suggesting that BACOVA_05493 is located on the bacterial surface. This was confirmed by whole cell assays of *Δbacova_05493*, which revealed no products were generated from galactan. This not only demonstrates that BACOVA_05493 is a surface enzyme but shows that the endo-galactanase, BACOVA_05488, consistent with its very low activity (**Supplementary Fig. 2**), did not contribute to galactan degradation.

We examined whether the *B. thetaiotaomicron* pectin PULs provide a genetic model for *Bacteroides* utilisation of these glycans. A 16S-based phylogenetic tree of the Bacteroides species was constructed and the organisms labelled for PUL conservation and growth on the respective oligo- and polysaccharides (**Supplementary Fig. 14**). There is ≥80% agreement between the presence of a PUL and growth of the bacterium on the corresponding poly- or oligosaccharide. Growth, however, was apparent in some organisms without an equivalent PUL, showing that bacteria can deploy alternative pathways to utilise a particular glycan. There were also examples of the presence of the cognate PUL without growth of the corresponding pectin. PUL conservation was also not always congruent with the 16S phylogeny (**Supplementary Fig. 14** and **Supplementary Fig. 15**). Thus, *B. ovatus*, *B. xylanisolvens* and *B. caccae* form a monophyletic group yet only *B. xylanisolvens* has a galactan PUL, while only *B. caccae* has lost the RGI PUL (**Supplementary Fig. 14**). The arabinan PUL of *B. ovatus* is fragmented and that of *B. caccae* absent (**Supplementary Fig. 15a**). *B. egghertii*, *B. stercoris* and *B. clarus*, also form a monophyletic group, but only *B. stercoris* has conserved the galactan PUL, while *B. eggerthii* has an arabinan PUL but has lost its RGI PUL (**Supplementary Fig. 15d**). A final example involves *B. thetaiotaomicron* and *B. xylanisolvens*, which are closely related species, yet have different arabinan PULs. The arabinan PUL of *B. thetaiotaomicron* is identical to *B. cellulosilyticus/oleiciplenus/intestinalis*, while the arabinan PUL of *B. xylanisolvens* is similar to *B. egghertii*. Our observations suggest that in the course of evolution *Bacteroides* rapidly gain and lose PULs that target different pectin structures.

Discussion

Combining biochemical properties and cellular location of the enzymes that target pectins, with growth profiles of mutants containing gene deletions in the appropriate PULs, enabled models for the metabolism of each pectic substructure, showing how the individual pathways are coordinated by *B. thetaiotaomicron* (**Fig 5**). The data revealed that 30 GHs and PLs are required to degrade the major pectin domains. Given that large numbers of enzymes are also required to degrade starch and the hemicelluloses, it is evident that plant glycan metabolism explains the extremely large repertoire of CAZyme gene clusters in colonic *Bacteroides* species.

In contrast to several *Bacteroides* glycan degrading systems where the surface GHs act slowly and target infrequent linkages[7,26,27], the equivalent enzymes of *B. thetaiotaomicron* that cleave galactan and the backbone of HG and RGI rapidly degrade their target polysaccharide. This likely reflects substrate accessibility to enzyme attack, and thus organisms with efficient surface enzymes that target accessible carbohydrates would be more competitive than bacteria in which the corresponding GHs or PLs were inefficient. This model (**Fig. 5**), however, does not apply to arabinan degradation where low activity of the surface enzymes was evident. This observation underpins the distinct mechanisms, distributive or selfish, by which glycans are metabolized by *Bacteroides spp.*

The RGI-PUL, in addition to orchestrating RGI backbone depolymerisation, removes remnants of linked polysaccharides and single sugar sidechains (**Fig. 5**). In contrast, PULs that mediate degradation of other branched glycans[7,8,26,27] depolymerize both the respective side chains and backbone structures. We propose that *B. thetaiotaomicron* does not necessarily target intact pectin structures but are able to utilise pectin domains generated by other organisms in the HGM. The RGI backbone exposed through symbiotic relationships with other intestinal microorganisms, or upstream processing by other PULs of *B. thetaiotaomicron*, is likely to contain additional pectin remnants explaining the complexity of enzymes encoded by the RGI-PUL.

The cross-feeding experiments demonstrate that galactooligosaccharides released by *B. thetaiotaomicron* are used by other organisms. The utilization of other pectin-

derived PBPs, however, is more restricted. These data illustrate how glycans are made available to the general community by primary degraders. Such cross-feeding has been observed between strains of *Bacteroides* cultured on fructans and soluble starch[28], with the recipient organism providing a benefit to the glycan degrading bacterium[29]. Possible non-*Bacteroides* beneficiaries of pectin-derived cross-feeding within the HGM are *Bifidobacterium* species, which generally utilize PBPs rather than the polysaccharide[27]. Contrasting oligosaccharide utilisation profiles observed among *Bacteroides spp.* may allow for co-existence of species within the same niche targeting different components of the same glycans without competition.

The critical role played by a surface exo-β-galactosidase in galactan metabolism in some *Bacteroides* species is intriguing. This contrasts with all other *Bacteroides* glycan degrading systems described to date, which deploy endo-acting CAZymes[2,7,8,27,30]. These organisms may target galactooligosaccharides, albeit with a high DP, released by other organisms within the HGM, obviating that need for endo-cleavage. This indicates that different *Bacteroides* target galactans in distinct nutritional niches within the gut. The data also illustrate the risk associated with generating models for glycan degradation based solely on prediction of enzyme function through CAZy family assignment. To fully understand glycan metabolism a molecular genetics approach informed by biochemical and transcriptional data in harness with bioinformatics predictions is required.

This report provides a model for how the pectic network is metabolized by a *Bacteroides* species in the HGM. Surprising variations in selective glycan metabolism and the constitution of individual pathways were apparent. This contrasts with the extensive conservation of other PULs[7-9,27]. This suggests that organisms have adopted a variety of strategies to metabolise dietary pectins. A salient feature of pectin utilization is the elaboration of enzymes in the RGI-PUL, reflecting the requirement to remove remnants from other pectic glycans and the extraordinary number of enzymes deployed in depolymerizing the disaccharide backbone. Dissecting the mechanism of pectin degradation contributes to our understanding of the foodweb within the HGM.

Methods

Producing recombinant proteins

DNA fragments encoding predicted CAZymes and binding proteins were amplified without signal sequence by PCR using appropriate primers. The resultant DNA was then cloned into pET21a or pET28a/b linearized using appropriate restriction enzymes. The expressed protein included a His6-tag fusion at the N-terminus. *Escherichia coli* strains BL21(DE3) or TUNER were transformed with the plasmids and grown to mid-exponential phase before induction with 1 mM (BL21(DE3)) or 0.2 mM (TUNER) isopropyl β-D-galactopyranoside (IPTG), and the culture was grown for a further 5 h at 37 °C or 16 h at 16 °C, respectively. The recombinant proteins were purified to >90% electrophoretic purity by immobilized metal ion affinity chromatography (IMAC) using Talon, a cobalt-based matrix, with bound proteins eluted with 100 mM imidazole, describe previously [9]. To generate seleno-methionine (Se-Met) proteins for structure resolution, *E. coli* cells were cultured as described previously [9], and the proteins were purified using IMAC as described above. For crystallization, the Se-Met proteins were further purified by size exclusion chromatography. After IMAC, fractions containing the purified proteins were buffer-exchanged, using PD-10 Sephadex G-25M gel-filtration columns (GE Healthcare), into 10 mM Na-HEPES buffer, pH 7.5, containing 150 mM NaCl and were then subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml min^{-1}. For crystallization trials, purified proteins were concentrated using an Amicon 10-kDa molecular mass centrifugal concentrator and washed three times with 5 mM DTT (for the Se-Met proteins) or water (for native proteins).

Site-directed mutagenesis

Site-directed mutagenesis was carried out employing a PCR-based NZY-Mutagenesis kit (NZYTech Ltd) using the plasmids encoding the appropriate enzymes as the template. The mutated DNA clones were sequenced to ensure that only the appropriate DNA change was introduced after the PCR.

Purification of oligosaccharides

Galactooligosaccharides were generated by incubation of 3 g of galactan with 100 mM HCl incubated for 3 h at 100 ˚C and neutralised by NaOH titration. The oligosaccharide mixture was freeze dried and resuspended in water before being applied to a P2-BioGel (BioRad) column with a 0.22 ml/min flow rate. Fractions were evaluated for oligosaccharide content and purity by TLC. Pure fractions of defined oligosaccharides were pooled and concentrated. Oligosaccharide size was confirmed by Mass Spectrometry and HPAEC. Crude oligosaccharide mixtures were generated by partial digestion with appropriate enzymes; BT0360 and BT0367 (arabinan), BT4668 (galactan), BT4170 (P-RGI/AM-RGI) and BT4116 (HG). Reactions were boiled and filter sterilised to remove precipitate before being evaluated by TLC.

Preparation of RGI-AM

Arabisopsis thaliana seeds were resuspended in distilled water (1 g/ml) and incubated at 4 ˚C for 16 h while stirring. The solution was centrifuged and supernatant filtered through G1 glass filter (15-40 μm pore size). This was then dialysed against 2 x 40 volumes of water before freeze drying. Typical yield was 1 g from 80 g seeds.

CAZyme Assays

Spectrophotometric quantitative assays for the □-L-rhamnosidase BT4145, L-arabinofuranosidases (BT0349, BT0348 and BT0368), □-D-galactosidases (BT4667, BT4151, BT4156, BT4160 and BACOVA_05493) and carbohydrate esterase (BT4158) were monitored by the formation of NADH, at $A_{340\,nm}$ using an extinction coefficient of 6,230 M^{-1} cm^{-1}, with an appropriately linked enzyme assay system. The assays were adapted from purchased Megazyme International assay kits. These kits were as follows: the L-rhamnose assay kit (K-RHAMNOSE); L-arabinose/D-galactose assay kit (K-ARGA); acetic acid detection kit (K-ACET). Activity of pectic lyases (BT4170, BT4175, BT4115, BT4116) were measured at A_{235nm}. Activity on 4-nitophenyl-glycosides was monitored at A_{400nm}. The activity of BT4668 to hydrolyse galactan was determined in 20 mM sodium phosphate buffer, pH 7.5 at 37 °C containing an appropriate concentration of the polysaccharide and 1 mg ml^{-1} BSA. Reactions were incubated at 37 °C and at regular time intervals

500 µl aliquots were removed and the amount of reducing sugar was quantified using the dinitrosalicylic acid reagent [31] and a standard curve of xylose in the reaction conditions used. Substrate depletion assays were performed as described previously [7] to determine BT4668 activity on galactooligosaccharides while production of D-galactose was used to assay BT4160 activity on galactooligosaccharides. The mode of action of enzymes were determined using PAEC or TLC, as appropriate. In brief, aliquots of the enzyme reactions were removed at regular intervals and, after boiling for 10 min to inactivate the enzyme and centrifugation at 13,000*g*, the amount of substrate remaining or product produced was quantified by HPAEC using standard methodology. The reaction substrates and products were bound to a Dionex CarboPac PA100 (Galactooligosaccharides/Arabinooligosaccharides), PA1 (Monosaccharides) or PA20 (Polygalacturonic acid oligosaccharides) column and glycans eluted with an initial isocratic flow of 100 mM NaOH then a 0–200 mM sodium acetate gradient in 100 mM NaOH at a flow rate of 1.0 ml min^{-1}, using pulsed amperometric detection. Linked assays were checked to make sure that the relevant enzyme being analysed was rate limiting by increasing its concentration and ensuring a corresponding increase in rate was observed. A single substrate concentration was used to calculate catalytic efficiency (k_{cat}/K_M), and was checked to be markedly less than K_M by halving and doubling the substrate concentration and observing an appropriate increase or decrease in rate. The equation $V_0 = (k_{cat}/K_M)[S][E]$ was used to calculate k_{cat}/K_M unless substrate depletion was used then the calculation was as follows $\ln(k_{cat}/K_M) = (S_0/S_t)/[E]$, in which $[E]$ and $[S]$ are enzyme and substrate concentration, respectively. All reactions were carried out in 20 mM sodium phosphate buffer, pH 7.0, with 150 mM NaCl (defined as standard conditions) and performed in at least technical triplicates.

Isothermal Titration Calorimetry

The binding of proteins to their glycan ligands was quantified by isothermal titration calorimetry (ITC), as described previously[27]. Titrations were carried out in 50 mM Na-HEPES buffer, pH 7.5 at 25 °C. The reaction cell contained protein at 50–100 µM, while the syringe contained either the oligosaccharide at 1–10 mM or the polysaccharide at 3–10 mg ml^{-1}. Integrated heats were fitted to a single-site model using Microcal Origin v7.0 to derive n, K_a, and ΔH values. ΔG and ΔS were

calculated from the equation $-RT\ln K_a = -\Delta G = \Delta H - T\Delta S$ where R is the gas constant and T temperature in Kelvins.

Electrospray ionisation mass spectrometry (ESI-MS)

The molecular mass of purified oligosaccharides (in 10 mM ammonium acetate, pH 7.0) were analysed via negative ion mode infusion/offline ESI-MS following dilution (typically 1:1 (v/v)) with 5% trimethylamine in acetonitrile.

Electrospray data was acquired using an LTQ-FT mass spectrometer (Thermo) with a FT-MS resolution setting of 100,000 at $m/z = 400$ and an injection target value of 1,000,000. Infusion spray analyses were performed on 5–10 µl of samples using medium 'nanoES' spray capillaries (Thermo) for offline nanospray mass spectrometry in negative ion mode at 1 kV.

^{1}H-NMR determination of catalytic mechanism

Enzymes BT4145 and BACOVA_05493 were freeze dried in 20 mM Tris-HCl, 500 mM NaCl, pH 7.5 as were substrates α-L-Rha-α1,4-D-GalA and $(\beta1,4\text{-Gal}p\text{-})_3$, respectively and resuspended in deuterium oxide. Prior to addition of enzyme an initial 1H-NMR spectra was obtained. Enzyme was added and spectra recorded at appropriate time intervals. The ratio of α- and β- monosaccharide products was determined to deduce catalytic mechanism.

2D NMR of arabinotetraose before and after treatment with BT0349

NMR spectra were recorded at 298 K in D_2O with a Bruker AVANCE III spectrometer operating at 600 MHz equipped with a TCI CryoProbe. Two-dimensional ^{1}H-^{1}H TOCSY, ROESY, DQFCOSY, ^{13}C HSQC and HSQC-TOCSY experiments were performed, using established methods[32]; the mixing times were 70 ms and 200 ms for the TOCSY and ROESY experiments, respectively. Chemical shifts were measured relative to internal acetone (δ_H =2.225, δ_C=31.07 ppm). Data were processed using the Azara suite of programs (v. 2.8, copyright 1993-2017, Wayne Boucher and Department of Biochemistry, University of Cambridge, unpublished) and chemical-shift assignment was performed using Analysis v2.4[33].

Growth of Bacteorides and generation of mutants

Bacteroides mutants were generated by deletion or replacement of the target gene with an inactive version by counter selectable allelic exchange using the pExchange-tdk plasmid. The full method is described in[34]. Mutants generated in this study are distinguished by the locus tag of the gene deleted/inactivated (*Δbtxxx* or *ΔBACOVAxxxxx*).

Bacteroides spp. were routinely cultured under anaerobic conditions at 37 °C using an anaerobic cabinet (Whitley A35 Workstation; Don Whitley) in culture volumes of 0.2, 2 or 5 ml) of TYG (tryptone-yeast extract-glucose medium) or minimal medium (MM) containing 0.5-1% of an appropriate carbon source and 1.2 mg ml^{-1} porcine haematin (Sigma-Aldrich) as previously described[8]. The growth of the cultures were routinely monitored at $OD_{600\,nm}$ using a Biochrom WPA cell density meter for the 5 ml cultures or a Gen5 v2.0 Microplate Reader (Biotek) for the 0.2 and 2 ml cultures.

Protein cellular localisation

Cellular localization of proteins was carried out as described previously[7]. In brief, *B. thetaiotaomicron* cultures were grown overnight ($OD_{600\,nm}$ value of 2.0) in 5 ml MM containing 0.5 % potato rhamnogalacturonan I (P-RGI) or homogalacturonan. The next day, cells were collected by centrifugation at 5,000*g* for 10 min and resuspended in 2 ml PBS. Proteinase K (0.5 mg ml^{-1} final concentration) was added to 1 ml of the suspension and the other half left untreated (control). Both samples were incubated at 37 °C overnight followed by centrifugation (5,000*g* for 10 min) to collect cells. To eliminate residual proteinase K activity, cell pellets were resuspended in 1 ml of 1.5 M trichloroacetic acid and incubated on ice for 30 min. Precipitated mixtures were then centrifuged (5,000*g*, 10 min) and washed twice in 1 ml ice-cold acetone (99.8%). The resulting pellets were allowed to dry in a 40 °C heat block for 5 min and dissolved in 250 µl Laemmli buffer. Samples were heated for 5 min at 98 °C and mixed by pipetting several times before resolving by SDS–PAGE using 12% gels. Electrophoresed proteins were transferred to nitrocellulose membranes by Western blotting followed by immunochemical detection using primary rabbit polyclonal antibodies (Eurogentec) generated against various proteins and secondary goat anti-rabbit antibodies (Santa Cruz Biotechnology). For BT4119 the anti-sera failed to produce the desired reactivity. Thus, a C-terminal Flag peptide

(DYKDDDDK) was incorporated at the C-terminals of the native proteins expressed by *B. thetaiotaomicron* through counter-selectable allelic exchange[34]. This allowed for their detection using rabbit anti-Flag antibodies (Sigma) as primary antibodies. In the case of BT4668, BT0360 and BT0367 mutations (that lead to the inactivation of the encoded enzymes) were made in each gene within the *B. thetaiotaomicron* genome to generate the mutants *Δbt4668, Δbt0360, Δbt0367* and *Δbt0360/Δbt0367*. Cells were grown in MM containing 0.5% arabino- or galacto-oligosaccharides to activate the target PULs. The cells were harvested from mid-log phase 5 ml cultures and concentrated in 0.5 ml PBS. The resuspended cells were incubated with the target glycans in an aerobic environment, conditions in which only the activity of only the surface enzymes can be monitored. The appropriate time intervals samples were taken, subjected to HPAEC analysis. The data were compared to that of wild type *B. thetaiotaomicron* to explore whether the loss in enzyme activity occurred at the bacterial surface.

Cross-feeding and competition assays

Prior to co-culture each *Bacteroides spp.* was grown in TYG and washed in PBS before being used to inoculate MM containing 0.5% glycan. Samples of 0.5 ml were taken at regular intervals during growth, which were serially diluted and plated onto Brain-Heart Infusion (BHI, Sigma-Aldrich) with agar and porcine hematin for determination of total CFU/ml of the culture. Genomic DNA was purified from the remainder of the sample (Bacterial genomic DNA purification kit, Sigma-Aldrich). Quantitate PCR was used to determine the ratio of different *Bacteroides spp.* or mutants in the sample using primers specific for unique regions in each *Bacteroides sp.* genome or tag introduced into one of two *att* sites. The Ratio of each species/mutant was used to calculate the CFU/ml of each organisms in the culture.

Quantitative RT-PCR (RT-qPCR)

Comparison of the levels of transcription of *susC* homologues (*susC*$_H$) from each of the pectin PULs was performed by RT-qPCR. Previous studies have shown *sus*C$_H$ genes are a good proxy for expression of their cognate PUL[35]. *B. ovatus* was cultured in 5 ml of MM containing 0.5% (w/v) carbon source, as described above. Triplicate bacterial cultures were harvested at mid-log phase (OD_{600} ~0.8) and placed in RNAprotect (Qiagen), then stored at −80 °C overnight, before purification

608 with RNeasy kit (Qiagen). RNA purity was assessed spectrophotometrically, and
609 1 µg of RNA was used immediately for reverse transcription (QuantiTect Reverse
610 Transcription kit, Qiagen). RT-qPCR was performed in a 96-well plate on a
611 LightCycler 480 System (Roche) with FastStart Essential DNA Green Master
612 (Roche) using the standard primer. Reactions were carried out in 10 µl, consisting of
613 5 µl SYBR Green mix, 20 ng of cDNA, and 1 µM (*sus*C_H genes) or 0.125 µM (16 S
614 rRNA) primer mix. Reaction conditions were 95 °C 600 s, followed by 45 cycles of
615 95 °C for 10 s, 55 °C for 10 s, 72 °C for 10 s. Cq values were calculated using
616 LightCycler 480 SW 1.5. Data were normalized to 16 S rRNA transcript levels, and
617 change in expression level calculated as fold-change compared with minimal media,
618 glucose cultures.
619
620 **Crystal structure determination**
621 ***Crystallisation:*** All proteins were concentrated to 10 mg/ml. BT4170 native
622 crystallised in 20 mM sodium/potassium phosphate 20% (w/v) polyethylene glycol
623 (PEG) 3350. BT4170 co-crystallised with 10 mM of oligosaccharide reaction
624 products generated by BT4170 (defined as ligand) in 100 mM succinic acid, sodium
625 phosphate glycine buffer at pH 6.0 and 25 % (w/v) PEG 1500. BT4170 inactive
626 mutant K285A was co-crystalized with 30 mM ligand in 200 mM potassium chloride
627 and 20% PEG 3350. Selenomethionine-containing BT4155 crystalized in 200 mM
628 sodium chloride, 100 mM Bis-Tris buffer pH 5.5 and 25% PEG 3350. BT0349 with
629 500 mM L-arabinose was crystallised in 20 % PEG 3350 and 200 mM ammonium
630 formate. All samples were cryo-protected by supplementing the mother liquor with
631 20% PEG 400.
632
633 ***Data collection and processing:*** BT0349, BT4170 and BT4170 K285A ligand data
634 were indexed and integrated with the automated pipeline Xia2 (3da protocol)[36].
635 BT4170 in complex with ligand and BT4155 were indexed and integrated with XDS[37].
636 The data were scaled with either XDS or Aimless[38]. Space group determination was
637 confirmed with Pointless[39]. The phase problem for BT0349 and BT4155 was solved
638 by SeMet-SAD using hkl2map[40] and the shelx pipeline[41]. BT4170 native apo data
639 were solved by molecular replacement with the pipeline Balbes[42] with the PDB
640 model 1RU4 as search model. Initial models of BT0349, BT4155 and BT4170 were
641 improved by successive runs of automated model building program arp_warp[43] and

buccaneer[44]. BT4170 TRI SCACCHARIDE and BT4170 inactive mutant K285A data were solved using the 4170 native apo model. All models were refined and improved using successive cycle of Refmac[45] and manual model building with Coot[46]. All models were validated using Coot[46] and molprobity[47]. Five percent of the observations were randomly selected for the Rfree set. The data processing, refinement statistics and protein database (PDB) codes are reported in **Supplementary Table 9**

Comparative genomics analysis

PULs similar to the RGI, galactan, arabinan and homogalacturonan PULs were searched in HGM Bacteroidetes genomes. The identification of similar PULs was based on PUL alignments. Gene composition and order of Bacteroidetes PULs were computed using the PUL predictor described in PULDB[48]. Then, in a manner similar to amino acid sequence alignments, the predicted PULs were aligned to the appropriate pectin PULs according to their modularity as proposed in the RADS/RAMPAGE method[49]. Modules taken into account include CAZy families, sensor-regulators and *suscd*-like genes. Finally, PUL boundaries and limit cases were refined by BLASTP-based analysis. The previously unknown glycoside hydrolase families discovered in this study are listed in the main text.

Data availability. The data that support the findings of this study are available from the corresponding author upon request. The authors declare that the data supporting the findings of this study are available within the paper and the Supplementary Information. Complete western blot images are provided in **Supplementary Fig. 1**. The crystal structure datasets generated (coordinate files and structure factors) have been deposited in the Protein Data Bank (PDB) and are listed in **Supplementary Table 9** together with the PDB accession codes.

REFERENCES

1 Gilbert, J. A. *et al.* Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94-103, (2016).

2 Sonnenburg, J. L. & Backhed, F. Diet-microbiota interactions as moderators of human metabolism. *Nature* **535**, 56-64, (2016).

3 Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* **10**, 323-335, (2012).

4 Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353 e1321, (2016).

5 Wu, M. *et al.* Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. *Science* **350**, aac5992, (2015).

6 Cartmell, A. *et al.* How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc Natl Acad Sci U S A* **114**, 7037-7042, (2017).

7 Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165-169, (2015).

8 Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498-502, (2014).

9 Ndeh, D. *et al.* Complex pectin metabolism by gut bacteria reveals novel catalytic functions *Nature* **544**, *65-70*, (2017).

10 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495, (2014).

11 Gilbert, H. J. The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol* **153**, 444-455, (2010).

12 Caffall, K. H. & Mohnen, D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res* **344**, 1879-1900, (2009).

13 Lau, J. M., Mcneil, M., Darvill, A. G. & Albersheim, P. Treatment of Rhamnogalacturonan-I with Lithium in Ethylenediamine. *Carbohyd Res* **168**, 245-274, (1987).

14 Coenen, G. J., Bakx, E. J., Verhoef, R. P., Schols, H. A. & Voragen, A. G. J. Identification of the connecting linkage between homo- or xylogalacturonan and rhamnogalacturonan type I. *Carbohyd Polym* **70**, 224-235, (2007).

15 Bonnin, E., Garnier, C. & Ralet, M. C. Pectin-modifying enzymes and pectin-derived materials: applications and impacts. *Appl Microbiol Biotechnol* **98**, 519-532, (2014).

16 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221, (2011).

17 Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, 24673-24677, (2009).

18 Glenwright, A. J. *et al.* Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407-411, (2017).

19 Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**, 2074-2076, (2003).

20 Cartmell, A. *et al.* The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *J Biol Chem* **286**, 15483-15495, (2011).

21 Pickersgill, R., Smith, D., Worboys, K. & Jenkins, J. Crystal structure of polygalacturonase from Erwinia carotovora ssp. carotovora. *J Biol Chem* **273**, 24660-24664, (1998).

22 van Santen, Y. *et al.* 1.68-angstrom crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis. *J Biol Chem* **274**, 30474-30480, (1999).

23 Sengkhamparn, N. *et al.* Okra pectin contains an unusual substitution of its rhamnosyl residues with acetyl and alpha-linked galactosyl groups. *Carbohydr Res* **344**, 1842-1851, (2009).

24 Renard, C. M., Crepeau, M. J. & Thibault, J. F. Glucuronic acid directly linked to galacturonic acid in the rhamnogalacturonan backbone of beet pectins. *Eur J Biochem* **266**, 566-574, (1999).
25 Raghavan, V., Lowe, E. C., Townsend, G. E., 2nd, Bolam, D. N. & Groisman, E. A. Tuning transcription of nutrient utilization genes to catabolic rate promotes growth in a gut bacterium. *Mol Microbiol* **93**, 1010-1025, (2014).
26 Bagenholm, V. *et al.* Galactomannan catabolism conferred by a polysaccharide utilization locus of *Bacteroides ovatus*: enzyme synergy and crystal structure of a beta-mannanase. *J Biol Chem* **292**, 229-243, (2017).
27 Rogowski, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* **6**, 7481, (2015).
28 Rakoff-Nahoum, S., Coyne, M. J. & Comstock, L. E. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr Biol* **24**, 40-49, (2014).
29 Rakoff-Nahoum, S., Foster, K. R. & Comstock, L. E. The evolution of cooperation within the gut microbiota. *Nature* **533**, 255-259, (2016).
30 Foley, M. H., Cockburn, D. W. & Koropatkin, N. M. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell Mol Life Sci* **73**, 2603-2617, (2016).
31 Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426-428, (1959).
32 Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*. (Academic Press, 1996).
33 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696, (2005).
34 Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105-1115, (2008).
35 Despres, J. *et al.* Unraveling the pectinolytic function of Bacteroides xylanisolvens using a RNA-seq approach and mutagenesis. *BMC Genomics* **17**, 147, (2016).
36 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190, (2010).
37 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, (2010).
38 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214, (2013).
39 Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**, 282-292, (2011).
40 Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Crystallogr* **37**, 843-844, (2004).
41 Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479-485, (2010).
42 Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* **64**, 125-132, (2008).
43 Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-1179, (2008).

44 Cowtan, K. Fitting molecular fragments into electron density. *Acta Crystallogr D Biol Crystallogr* **64**, 83-89, (2008).
45 Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195, (2004).
46 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, (2004).
47 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, (2010).
48 Terrapon, N., Lombard, V., Gilbert, H. J. & Henrissat, B. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, 647-655, (2015).
49 Terrapon, N., Weiner, J., Grath, S., Moore, A. D. & Bornberg-Bauer, E. Rapid similarity search of proteins using alignments of domain arrangements. *Bioinformatics* **30**, 274-281, (2014).
50 Varki, A. *et al.* Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324, (2015).

Correspondence and requests for materials should be addressed to H.J.G.

Acknowledgements

This work was supported in part by an Advanced Grant from the European Research Council (Grant No. 322820) awarded to H.J.G. and B.H. supporting A.S.L., D.N., A.C. and N.T., a Wellcome Trust Senior Investigator Award to HJG (grant No. WT097907MA) that supported J.B. and E.C.L. a European Union Seventh Framework Initial Training Network Programme entitled the “WallTraC project” (Grant Agreement number 263916) awarded to M-C.R. and H.J.G, which supported X.Z. and J.S. The Biotechnology and Biological Research Council project “Ricefuel” (grant numbers BB/K020358/1) awarded to H.J.G. supported A.L. We thank Diamond Light Source for access to beamline I02, I04-1 and I24 (mx1960, mx7854 and mx9948) that contributed to the results presented here, and to Joes Gray at Newcastle University for assistance with the mass spectrometry.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article

Author contributions

Enzyme characterisation was carried out by A.S.L., J.B., X.Z., A.L., I.V. R.M., K.C., B.F., J.S. The generation of oligosaccharide products by M-C.R., X.Z., A.S.L., A.C. and D.N. Gene deletion strains were constructed by A.S.L., D.N., R.M., B.F., J.B. and D.W.A. Co-culturing experiments were carried out by J.B. and A.S.L. Phylogenetic reconstruction and metagenomic analysis was by N.T. and B.H. Bacterial growth and transcriptomic experiments: Y.X., E.C.L. and E.C.M. X-ray protein crystallography was by A.B., A.C., A.S.L. and J.B. N.M.R. experiments was by A.S.L. and K.S. Experiments were designed by D.W.A., H.J.G. and E.C.L., S.C.M. and H.J.G. The manuscript was written by H.J.G. with substantial contributions from D.W.A. E.C.L., N.T. and B.H. Figures were prepared by E.C.L. and A.S.L.

FIGURE LEGENDS

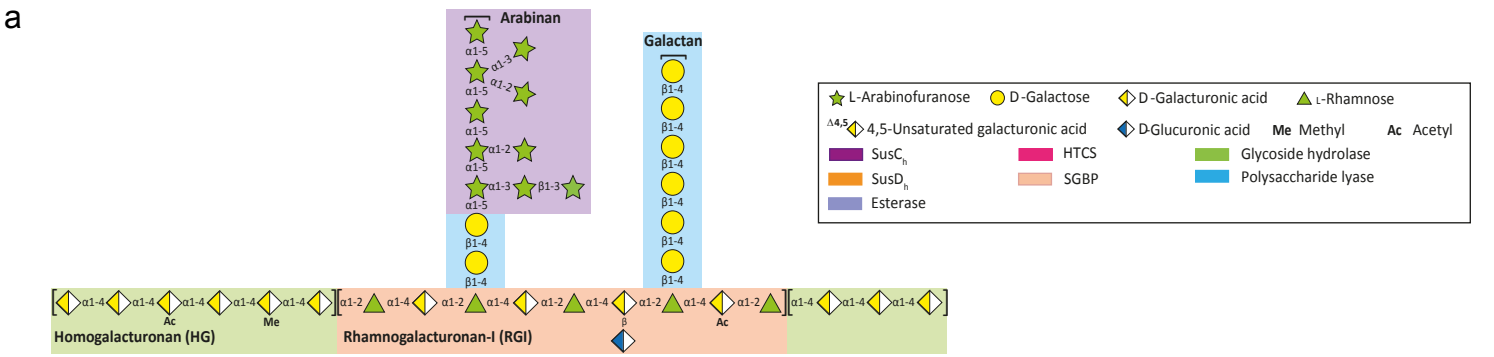
Figure 1. Genomic organization of pectin PULs. a, Schematic of pectin structure showing the different polysaccharides highlighted with different coloured backgrounds. The respective linkages and monosaccharide composition are represented according to the Symbol Nomenclature for Glycans system[50]. **b,** genes encoding proteins of known or predicted functionalities are colour coded. GHs, CEs and PLs located in a known CAZy family are indicated by GHXX, CEXX or PLXX where XX indicates the number of the family.

Figure 2. Depolymerization of pectins at the cell surface of *B. thetaiotaomicron* cell surface. a, Growth of wild-type and mutants of *B. thetaiotaomicron* (BtWT and ∆btxxxx) or *B. ovatus* (BoWT and ∆bacovaxxxx) in minimal media containing the indicated pectic polysaccharide; HG, homogalacturonan; SBA, sugar beet arabinan; RGI-AM, rhamnogalacturonan I backbone from Arabidopsis mucilage (biological replicates, n=3, error bars denote s.e.m). **b,** BtWT, BoWT and mutants lacking functional outer membrane enzymes were incubated with appropriate polysaccharides in aerobic conditions for the times indicated. Under these conditions substrate is only available to the surface enzymes. Products released from the glycans were monitored by High performance anion exchange chromatography (HPAEC) with pulsed amperometric detection (PAD) or UV detection at 235 nm (Abs_{235nm}). The degree of polymerisation of the peaks corresponding to the galactose (Gal) and arabinose (Ara) oligosaccharides are shown in subscript numbers. **c,** Western blot detection of selected *B. thetaiotaomicron* enzymes encoded by the HG-PUL and RGI-PUL after treatment with proteinase K (PK+) or untreated (-). BT4661 is a known surface glycan binding protein (control)[6]. The cellular localization is indicated as periplasmic (P) or cell surface (CS). The example is from biological replicates n=3. The full western blots are shown in **Supplementary Fig. 1**.

Figure 3. Signal molecule protection. Each panel shows the affinities of the signal molecules to respective sensors (top) and the catalytic efficiency of key enzymes implicated in signal molecule degradation (bottom) for **a,** galactan, **b,** arabinan and **c**, RGI. The data were from technical replicates, $n \geq 3$.

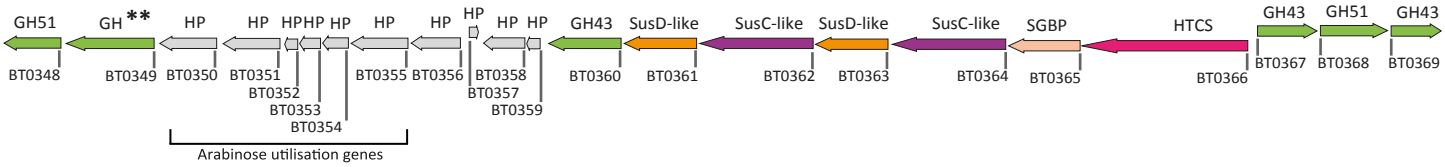
Figure 4. Cross-feeding of polysaccharide breakdown products between *Bacteroides* species. a, Wild type *B. thetaiotaomicron* (WT) and mutants of the bacterium lacking the key surface degrading enzymes for each polysaccharide were mono-cultured and co-cultured with the wild type bacterium as indicated. Samples were taken at different time points. The colony forming units of these samples were determined by plating onto rich media (top panels) and the ratio of each bacterium in the culture (bottom panel) was determine by qPCR with primers unique to each strain. Error bars represent the s.e.m of biological replicates (n=3). **b,** *B. ovatus* (Bo), *B. massiliensis* (Bm) and *B. uniformis* were mono-cultured or co-cultured with wild type *B. thetaiotaomicron* (Bt) using the same experimental approach described in **a**.

876 **Figure 5. Model of pectin utilization by *B. thetaiotaomicron***. Chemical structures
877 of the sugars in the major pectins (**a**)Models for degradation of galactan (**b**, blue),
878 arabinan (**c**, purple), homogalacturonan (**d**, green) and rhamnogalacturonan I (**e**,
879 peach) are displayed. The black arrows indicate the linkage cleaved by the various
880 enzymes, while the grey arrows show the direction of the degradative pathway.
881

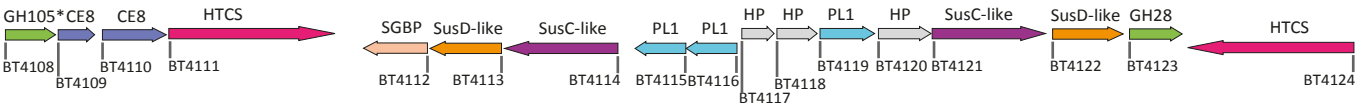


b

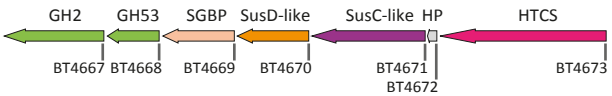
Ara-PUL



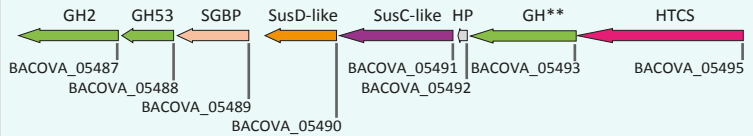
HG-PUL



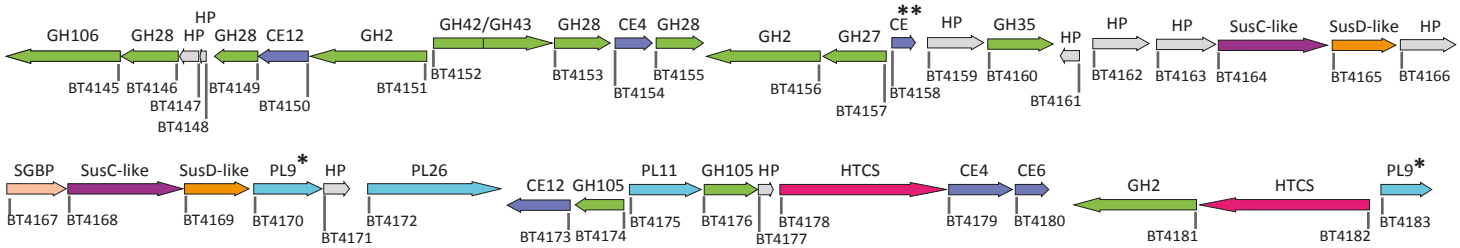
Gal-PUL



B. ovatus Gal-PUL

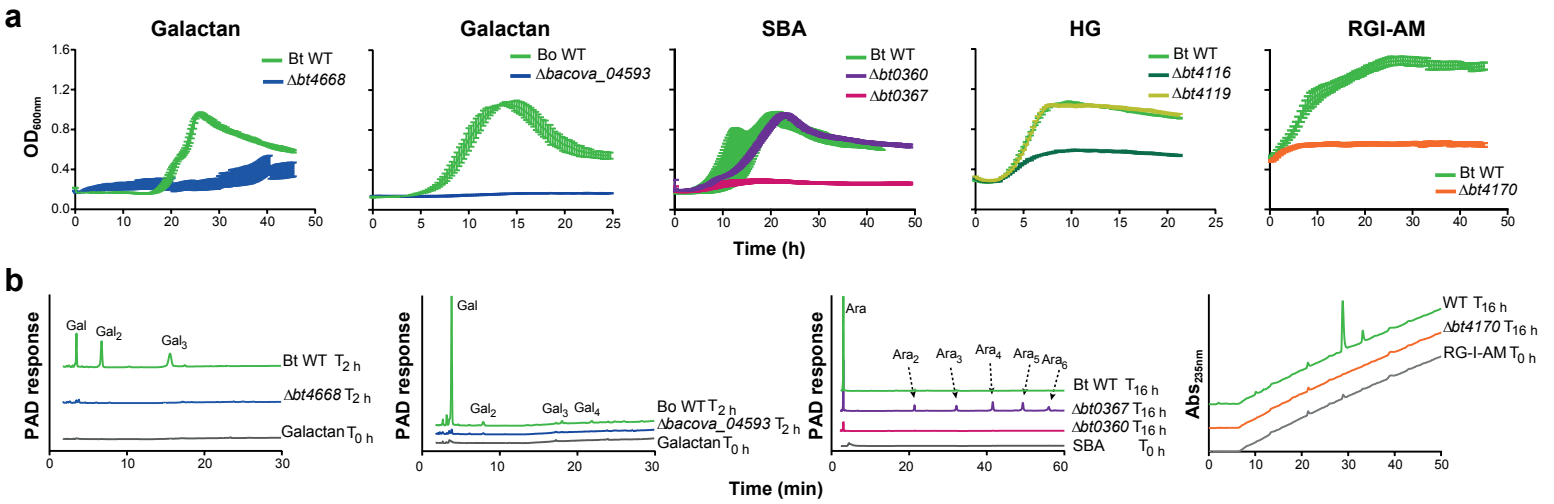


RGI-PUL

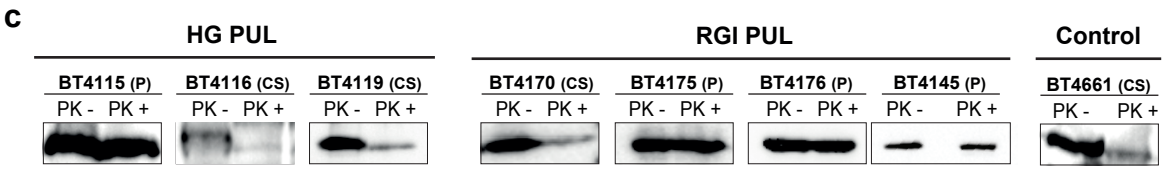


GH (Glycoside hydrolase) | CE (Carbohydrate Esterase) | SusC-like | SGBP (Surface Glycan Binding Protein) | * New activity for family

PL (Polysaccharide lyase) | HP (Hypothetical Protein) | SusD-Like | HTCS (Hybrid Two Component System) | ** New family


a
Galactan
Bt WT
Δbt4668
Galactan
Bo WT
Δbacova_04593
SBA
Bt WT
Δbt0360
Δbt0367
HG
Bt WT
Δbt4116
Δbt4119
RGI-AM
Bt WT
Δbt4170
OD600nm
Time (h)

b
PAD response
Gal
Gal2
Gal3
Bt WT T2 h
Δbt4668 T2 h
Galactan T0 h
Gal
Gal2
Gal3 Gal4
Bo WT T2 h
Δbacova_04593 T2 h
Galactan T0 h
Ara
Ara2
Ara3
Ara4
Ara5 Ara6
Bt WT T16 h
Δbt0367 T16 h
Δbt0360 T16 h
SBA T0 h
Time (min)
Abs235nm
WT T16 h
Δbt4170 T16 h
RG-I-AM T0 h


c
HG PUL
RGI PUL
Control
BT4115 (P)
BT4116 (CS)
BT4119 (CS)
BT4170 (CS)
BT4175 (P)
BT4176 (P)
BT4145 (P)
BT4661 (CS)
PK - PK +

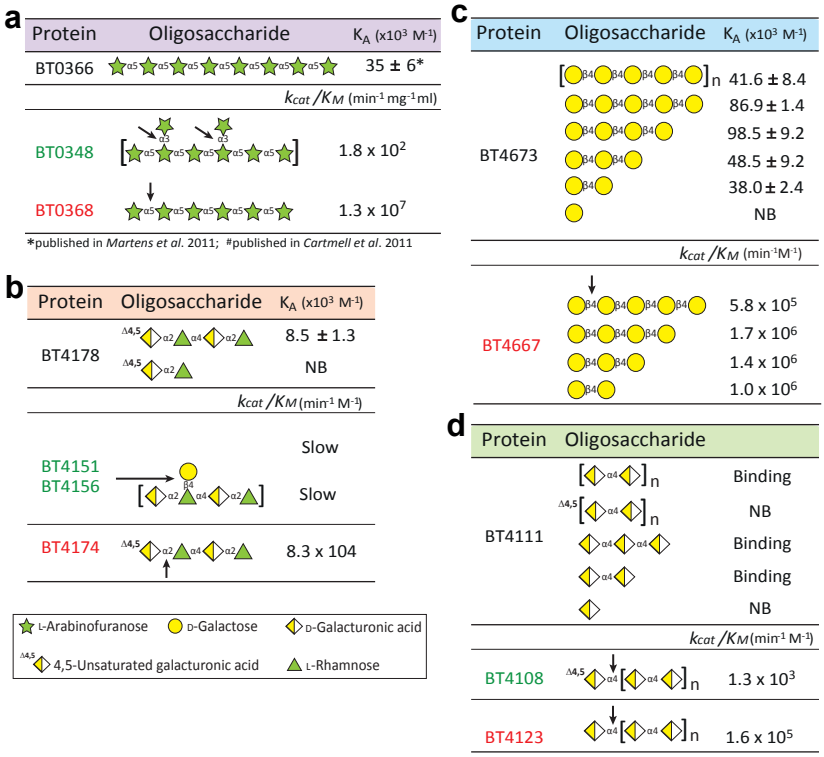
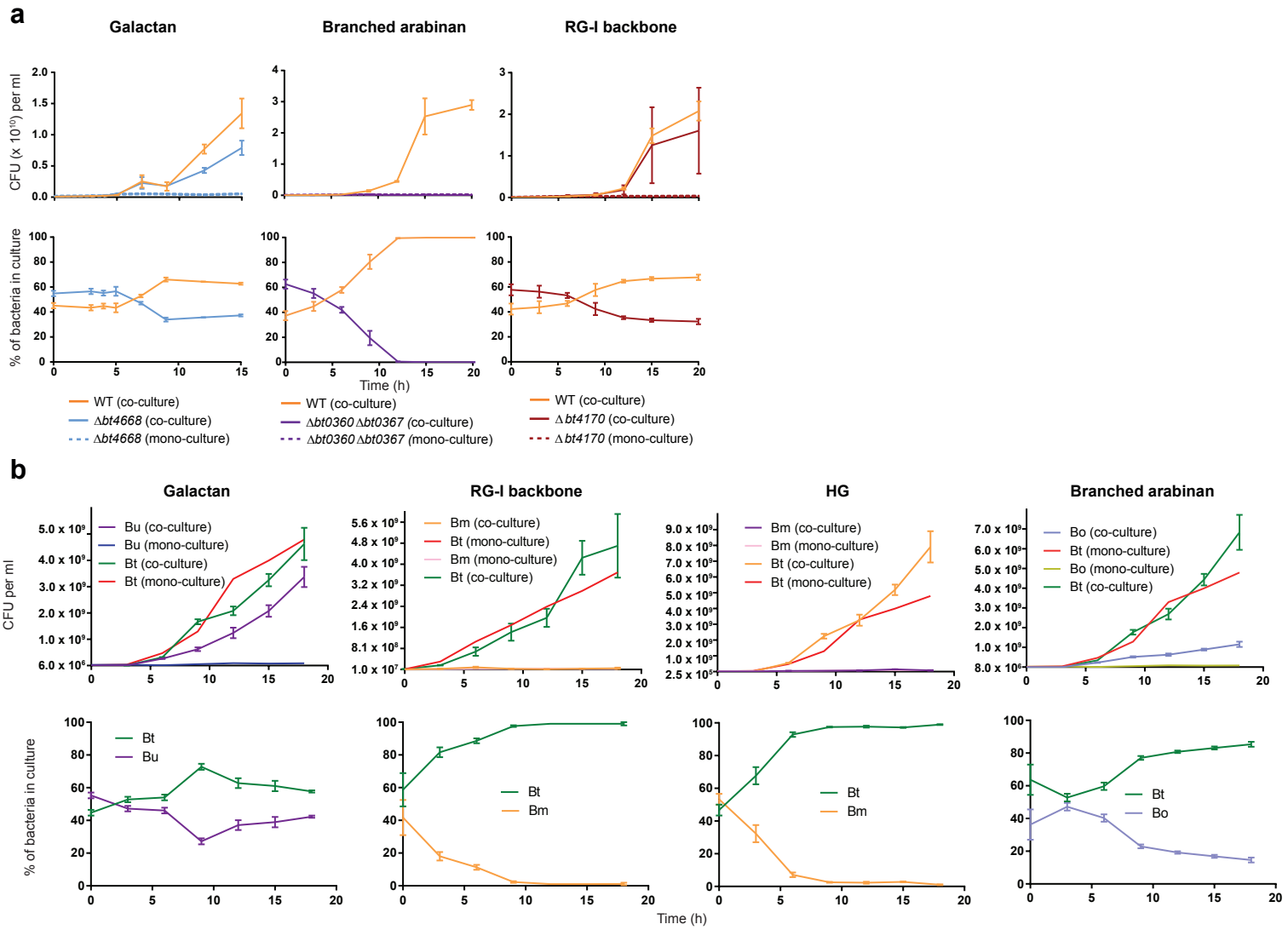
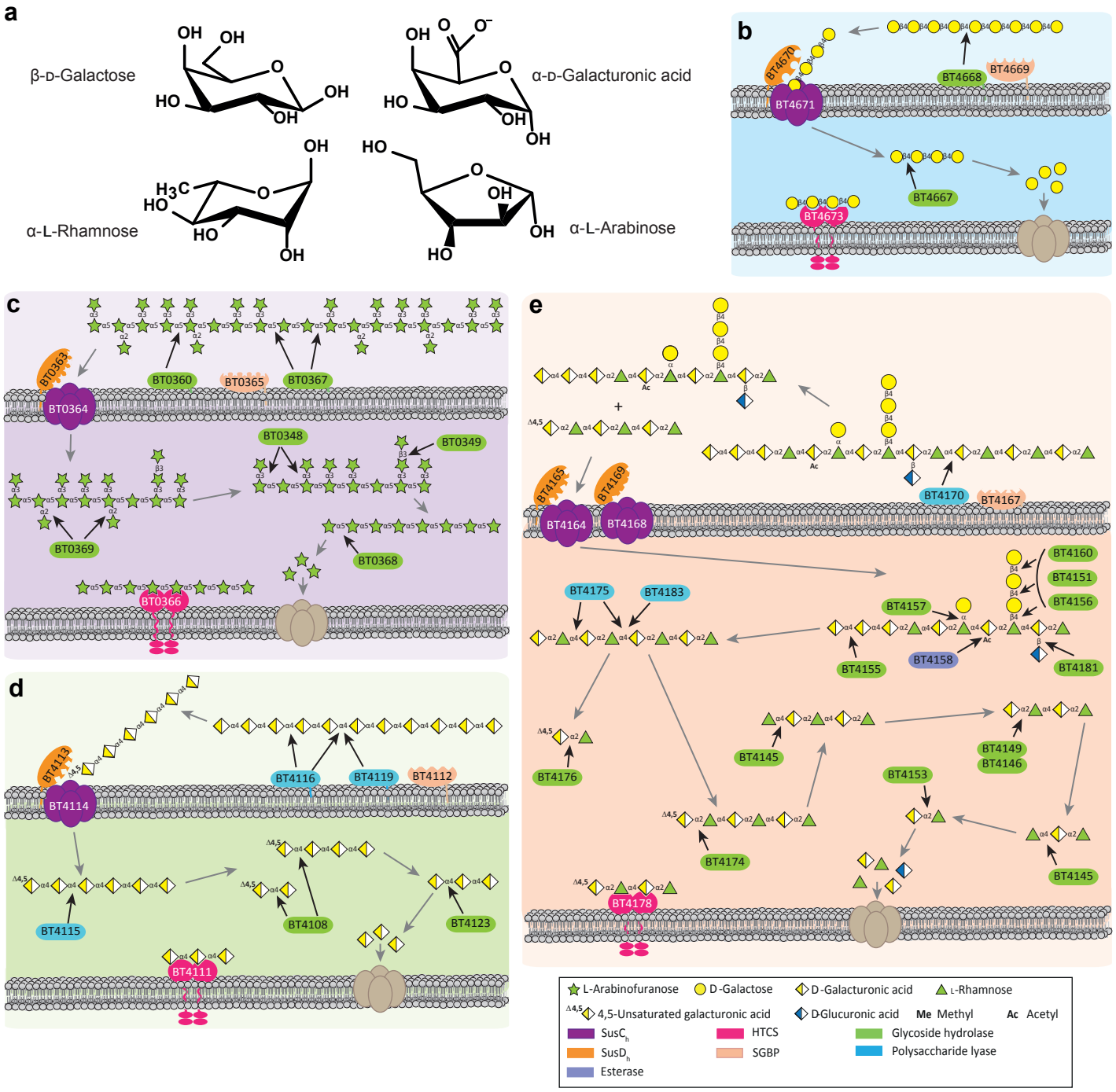


Figure 3. Signal molecule protection. Each panel shows the affinities of the signal molecules to respective sensors (top) and the catalytic efficiency of key enzymes implicated in signal molecule degradation (bottom) for **a**, galactan, **b**, arabinan, **c**, RGI and **d**, HG. The data were from technical replicates, n ≥ 3.

FIGURE 3


a
Galactan
Branched arabinan
RG-I backbone
CFU (x 10^10) per ml
% of bacteria in culture
Time (h)
WT (co-culture)
Δbt4668 (co-culture)
Δbt4668 (mono-culture)
WT (co-culture)
Δbt0360 Δbt0367 (co-culture)
Δbt0360 Δbt0367 (mono-culture)
WT (co-culture)
Δbt4170 (co-culture)
Δbt4170 (mono-culture)
b
Galactan
RG-I backbone
HG
Branched arabinan
CFU per ml
% of bacteria in culture
Time (h)
Bu (co-culture)
Bu (mono-culture)
Bt (co-culture)
Bt (mono-culture)
Bm (co-culture)
Bt (mono-culture)
Bm (mono-culture)
Bt (co-culture)
Bm (co-culture)
Bm (mono-culture)
Bt (co-culture)
Bt (mono-culture)
Bo (co-culture)
Bt (mono-culture)
Bo (mono-culture)
Bt (co-culture)
Bt
Bu
Bt
Bm
Bt
Bm
Bt
Bo


a
β-D-Galactose
α-D-Galacturonic acid
α-L-Rhamnose
α-L-Arabinose
b
BT4670
BT4671
BT4668
BT4669
BT4667
BT4673
c
BT0363
BT0364
BT0360
BT0365
BT0367
BT0348
BT0349
BT0369
BT0368
BT0366
d
BT4113
BT4114
BT4116
BT4119
BT4112
BT4115
BT4108
BT4123
BT4111
e
BT4165
BT4164
BT4169
BT4168
BT4170
BT4167
BT4160
BT4151
BT4156
BT4175
BT4183
BT4157
BT4158
BT4155
BT4181
BT4176
BT4145
BT4149
BT4146
BT4153
BT4174
BT4178
L-Arabinofuranose
D-Galactose
D-Galacturonic acid
L-Rhamnose
4,5-Unsaturated galacturonic acid
D-Glucuronic acid
Me Methyl
Ac Acetyl
SusC_h
SusD_h
Esterase
HTCS
SGBP
Glycoside hydrolase
Polysaccharide lyase