Carbohydrate-active enzymes from *Akkermansia muciniphila* breakdown mucin O-glycans to completion

Cassie R. Bakshani[1], Taiwo O. Ojuri[1], Bo Pilgaard[2], Jesper Holck[2], Ross McInnes[1], Maria Zakhour[3], Manon Kerouedan[1], Emily Newton[1], David N. Bolam[3], and Lucy I. Crouch[1*]

1 Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT.
2 Protein Chemistry and Enzyme Technology Section, DTU Bioengineering, Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Kgs, Lyngby, Denmark.
3 Biosciences Institute, Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

*Address correspondence to l.i.crouch@bham.ac.uk

Abstract

Akkermansia muciniphila is a human microbial symbiont residing in the mucosal layer of the large intestine. Its main carbon source is the highly heterogeneous mucin glycoprotein that constitutes the majority of the mucus dry weight. *A. muciniphila* uses an array of Carbohydrate-Active enZymes (CAZymes) and sulfatases to access this complex energy source. Here we describe the biochemical characterisation of fifty-four glycoside hydrolases (from twenty-four families), twelve sulfatases, and one polysaccharide lyase to provide a holistic understanding of the O-glycan degrading activities *A. muciniphila* carries out in the colon. The results provide an extensive insight into the sequence of O-glycan degradation and the localisation of the different activities. One of the most outstanding elements of this work was the demonstration that these enzymes can act synergistically to degrade the O-glycans on the mucin polypeptide to completion, down to the core GalNAc.

Introduction

It is the 20-year anniversary of the discovery of *Akkermansia muciniphila* from the relatively understudied Planctomycetota-Verrucomicrobiota-Chlamydiota (PVC) superphylum[1]. This fastidious microbe is renowned for its focus on mucin as a nutrient source in the large intestine of humans and multiple different studies have shown its prevalence is inversely proportional to disease and inflammation markers[2-5]. *A. muciniphila* is detectable in the majority of people, colonises early on in life, and typically constitutes 1-3 % of the total microbiota[6-8]. *Akkermansia*-like sequences have also been detected in a wide variety of mammals, fish, and reptiles which suggests a long evolutionary history between the mucosal surface of the gastrointestinal tract in vertebrates and *Akkermansia* species[6].

The mucosal surface of the human large intestine is predominantly comprised of gel-forming secreted mucins and approximately 80 % of this glycoprotein is O-glycan. Mucin O-glycans comprise only five different monosaccharides, thus their considerable structural heterogeneity is attributed to linkage diversity between these monosaccharides and different sulfation patterns (Figure 1). The structure of an O-glycan chain is generally accepted to be categorised into three sections: 1) the core, consisting of an α-linked GalNAc attached to a serine or threonine, 2) the polyLacNAc extensions linked to the core GalNAc, and 3) the terminal 'capping' epitopes. The polyLacNAc extensions are generally LacNAc disaccharides linked through β1,3-bonds, but variable sulfation, fucosylation, and branching adds to the complexity and heterogeneity along these chains. The capping structures at the non-reducing end of the polyLacNAc chains are different types of α-linked monosaccharides, for example, sialylation, blood group structures, and Lewis structures. The details of O-glycan structure varies along the digestive tract, between individuals, and between species[9]. For example, in terms of core structures, gel-forming MUC5AC dominates in the oesophagus and stomach and largely has Core 2, but gel-forming MUC2 dominant in the large intestine mainly has Core 3 and 4[10,11]. In

terms of capping structures, MUC5AC in the upper sections of the GI tract, for example, has a characteristic capping GlcNAcα1,4Gal structure (stomach epitope), and blood groups and fucosylation will vary according to an individual's genetics.

A biochemical characterisation of the sialidases, fucosidases, and GH16 endo-O-glycanases from *A. muciniphila* ATCC BAA-835 has been carried out in detail previously and provided insights into how it tackles this complicated structure[12-15]. However, a comprehensive mechanistic understanding of mucin degradation has not previously been undertaken for any bacterial species and a significant number of enzymes from *A. muciniphila* remain uncharacterised. Here, we provide a comprehensive picture of how *A. muciniphila* ATCC BAA-835 degrades mucin O-glycans, related structures and other host carbohydrates.

Results

The specificities of the recombinantly expressed CAZymes and sulfatases from *A. muciniphila* ATCC BAA-835 (Supplementary Table 1) were initially characterised against a broad range of defined substrates (Supplementary Table 2). We then assayed these enzymes against mucin and human milk oligosaccharide (HMO) substrates. To contextualise these results, growth studies were performed with *A. muciniphila* ATCC BAA-835, alongside assessing enzyme activities on the surface of mucin-grown cells (whole cell assays). A pangenome analysis was used to explore conservation of enzymes across *A. muciniphila* and other *Akkermansia* species, as one indicator of functional importance. The description of these results has been broadly split up into those enzymes acting on the α-linked monosaccharides, β-linked galactose, and β-linked HexNAc sugars.

Investigating CAZymes with exo-acting activities towards α-linked monosaccharides

<u>α-sialidases</u>

A. muciniphila ATCC BAA-835 has three sialidases and six fucosidases that can tackle a wide range of substrates (Supplementary Figures 1-3)[12,13]. The two GH33 family members (Amuc_0625^{GH33} and Amuc_1825^{GH33}) are broad-acting sialidases in terms of linkage, whereas the Amuc_1547^{GH181} is α2,3-specific. In this work, we explored whether these enzymes act upon ganglioside structures, which have structural similarities to O-glycans (Supplementary Figure 1). Gangliosides are important components of eukaryotic plasma membranes and are linked to several genetic diseases and pathogen interactions[16]. Gangliosides (GM3 and GD3; Supplementary Figure 4) are also found in the membranes of milk fat globules in breast milk, so it is likely that gut microbes in the infant gut would encounter these[17]. Therefore, gangliosides would be found in the large intestine at all life stages from both host and dietary sources. We found that the two **GH33 enzymes from *A. muciniphila* ATCC BAA-835 can act on the relatively complex GD1a and GT1b ganglioside structures**, with one or two sialic acids being removed. The inability to remove the last sialic acid is likely due to this branching sialic acid being difficult to access. However, if we removed the first galactose with β-galactosidases from *A. muciniphila* ATCC BAA-835, then it was possible for Amuc_1825^{GH33} to remove all sialic acids (Supplementary Figure 1). Importantly, the removal of the galactose produces Sda antigen[18], which is present in 91 % of humans and is expressed in mucosal surfaces and other tissues[19]. Therefore, is it valuable to understand that sialidases from *A. muciniphila* can act on Sda antigen and gangliosides. Amuc_1825^{GH33} was recently highlighted as important for growth on mucin using a Tn-seq mutant library[20], and Amuc_0625^{GH33} and Amuc_1547^{GH181} were highly conserved (>90 % sequence identity) across all *Akkermansia* species assessed (Supplementary Figure 5).

We also tested release of sialic acids from bovine submaxillary mucin (BSM), which has sialylated Core 1 decorations, and found similar release by Amuc_0625^{GH33}, Amuc_1547^{GH181}, and Amuc_1825^{GH33} (Supplementary Figure 6). With this experiment we saw the release of both Neu5Ac and Neu5Gc and this was quantified using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD; Supplementary Figure 5). Interestingly, there was some indication that Amuc_0625^{GH33} could release Neu5Gc more efficiently than the other enzymes, but this was not explored further. Release of sialic acid also carried out on porcine gastric mucin III (PGM III) assessed by thin-layer chromatography (TLC) indicates that Amuc_0625^{GH33} and Amuc_1825^{GH33} are releasing the highest amounts, but this was not quantified (Supplementary Figure 1). Two types of commercially available HMOs were tested, but no release of sialic acid was observed (Supplementary Figure 7). Whole cell assays were carried out using PGM III-grown cells, tested against fresh PGM III and BSM, and sialic acid release was seen from both substrates (Supplementary Figure 8-10). We could not find any activities for Amuc_0623^{GHnc} or Amuc_1547^{GH177} during this work despite their associations with sialic acid hydrolysis (Supplementary Figure 1).

α-fucosidases

Between them, the GH29 and GH95 enzymes from *A. muciniphila* ATCC BAA-835 can degrade a wide variety of fucose-rich substrates (Supplementary Figure 2-3)[13,21]. These enzymes cannot access the fucose in blood group A and B (BGA and B) structures, so these glycan epitopes are first partially broken down by other CAZymes encoded in the *A. muciniphila* ATCC BAA-835 genome to produce blood group H structures first. In this work, we were able to characterise the specificity of Amuc_0010^{GH29} to α1,2-linked fucose on lactose/LacNAc (β1,4), but not Lacto-N-biose (β1,3). This means that **Amuc_0010^{GH29} exhibits specificity towards type II blood group structures (not type I) and the lacto-N-*neo*tetraose HMO series (not the lacto-N-tetraose)**. This specificity facilitates the probing of mucin glycan structures in more detail. For instance, only Amuc_1120^{GH95} can hydrolyse fucose from PGM III, indicating that this substrate has type I structures immediately available to the fucosidases at the non-reducing ends on the O-glycans (Supplementary Figure 11). Discrepancies are also seen between the activities of Amuc_0010^{GH29} and Amuc_1120^{GH95} against HMOs, with the latter releasing more fucose and, therefore, these particular HMOs have predominantly lacto-N-tetraose structures (Supplementary Figure 7). These two observations correspond to the understanding of type II structures being present on red blood cells and type I structures being associated with glycan decorations on secretions[16]. Notably, Amuc_0010^{GH29} clusters on a phylogenetic tree with 'Mfuc5' from a soil metagenome and has comparable activities[22].

Amuc_0146^{GH29} and Amuc_1120^{GH95} were recently highlighted as important for growth on mucin using a TnSeq mutant library[20] and Amuc_0392^{GH29} is highly conserved across different *Akkermansia* species (Supplementary Figure 5). Furthermore, five out of nine of the sialidases and fucosidases were highlighted by proteomics when *A. muciniphila* ATCC BAA-835 was grown on human milk (Supplementary Table 1)[23].

As observed previously, *A. muciniphila* ATCC BAA-835 did not grow directly on sialic acid and only minimally on fucose, which we could reproduce in this report (Supplementary figure 12), and these monosaccharides are cross-fed to other species[13]. Regarding monosaccharides, *A. muciniphila* ATCC BAA-835 could grow directly on GlcNAc and GalNAc, but not mannose or galactose.

α-galactosidases

There are two putative GH110 enzymes (Amuc_0480^{GH110} and Amuc_1463^{GH110}) encoded by the *A. muciniphila* ATCC BAA-875 genome, which have 28 % identity between them, and are predicted to be in the cytoplasm. These displayed **activity towards blood group B (BGB) Types I and II and Galα1,3-Gal/GalNAc** (Figure 1; Supplementary Figures 13-16). BGB is slightly less prevalent than BGA in the human population, ranging between 0-30 %, depending on geographical location[16]. In addition, Amuc_1463^{GH110} was also able to hydrolyse the α1,4-linked galactose on globotriose fully and P1 antigen partially under the conditions tested. **This makes Amuc_1463^{GH110} the broadest-acting GH110 characterised to date and the broadest α-galactosidase in *A. muciniphila* ATCC BAA-875[24-26]**. These are also the only enzymes that are active against blood group B from *A. muciniphila* ATCC BAA-875, which predicate the removal of fucose by the fucosidases. Whole cell assays using Galα1,3-Gal/GalNAc showed no degradation over a 9-hour period, supporting the proposal that these epitopes are deconstructed inside the cell (Supplementary Figure 8). The whole cell assay results from the blood group hexasaccharides are less clear, but we previously characterised GH16 activity against these substrates and this activity is prominent on the outside of the cell ([14]; Supplementary Figure 8).

When these GH110 enzymes were added to PGM III both alone and pre-treated with sialidase and fucosidase (Amuc_1835^{GH33} and Amuc_1120^{GH95}) there was release of galactose (Figure 1 and Supplementary Figure 11). Amuc_0480^{GH110} was able to release more galactose than Amuc_1463^{GH110}, however this was not quantified or the reasons for this difference explored

further in this study. No obvious release of galactose was observed for the purified HMOs (Supplementary Figure 7). There is a single structure available for the GH110 family from *Pseudoalteromonas distincta* and it has a Galα1,3-Gal crystallised into the active site[26]. We overlaid models of both Amuc_0480^{GH110} and Amuc_1463^{GH110} with this enzyme-substrate structure and were able to highlight where the fucose and +2 GlcNAc would most likely be accommodated (Supplementary Figure 17).

One putative GH27 (Amuc_1187^{GH27}) is encoded in the *A. muciniphila* ATCC BAA-875 genome and was active against α-linked galactose from the non-reducing end of most of the defined substrates tested here, but not blood group structures (Figure 1; Supplementary Figure 13-16). Amuc_1187^{GH27} can hydrolyse Galα1,3Gal/GalNAc and globotriose, but less activity was seen against P1 antigen. Galα1,3Gal (Galili antigen) is expressed by most mammals and some other animals, but not humans, apes, and Old World monkeys[27]. Therefore, gut microbes would only have access to this structure through animal products in the diet. Conversely, the similar Galα1,4Gal epitopes of P1 and P^k antigens are generally present in human tissues, so they would be present in the mucosal surface of the human large intestine. P1 and P^k structures are receptors for a variety of pathogens and their toxins[28]. When Amuc_1187^{GH27} was added to PGM III both alone and pre-treated with sialidase and fucosidase (Amuc_1835^{GH33} and Amuc_1120^{GH95}) there was release of galactose (Figure 1 and Supplementary Figure 11). No obvious release of galactose was observed for the purified HMOs (Supplementary Figure 7).

The Amuc_1187^{GH27} signal sequence indicates this enzyme is localised to the periplasm (Supplementary Table 1) and whole cell assays of mucin-grown *A. muciniphila* ATCC BAA-875 do not show degradation of Galα1,3Gal/GalNAc (Supplementary Figure 8 and 9). The characterised activities from GH27 family members so far include α-galactosidases, α-GalNAc'ases, β-L-arabinopyranosidases, and isomaltodextrinases. Bacterial GH27 family members with activity against α-Gal have been characterised in the context of plant polysaccharides pectic galactan and galactomannan, for example, and there are no structures of these enzymes[29,30]. **Therefore, Amuc_1187^{GH27} is the first example of an α-galactosidase GH27 family member from prokaryotic origin with specificity for animal-type substrates rather than those of plant.** An AlphaFold model of Amuc_1187^{GH27} is analysed in the context of relevant available crystal structures (Supplementary Discussion and Supplementary Figure 18).

A. muciniphila ATCC BAA-835 has one putative GH97 enzyme (Amuc_1420^{GH97}), which displayed a preference for α-linked galactose in the **α1,6-Gal raffinose-type substrates** within the defined substrate screen with only partial activity against host-type glycans Galα1,3-Gal/GalNAc substrates (Figure 1). However, Amuc_1420^{GH97} was able to hydrolyse relatively large amounts of galactose from PGM III from samples with or without sialidase and fucosidase pre-treatment (Figure 1 and Supplementary Figure 11). Intrigued by this contrast in activities against disaccharides compared to mucin we analysed an AlphaFold model of Amuc_1420^{GH97} alongside available GH97 structures. The only GH97 α-galactosidase crystal structure with a ligand is that of BT1871 from *Bacteroides thetaiotaomicron*, which has been crystallised with a lactose that also had an α-linked galactose to the Glc through a 1,1-linkage[31]. This ligand was overlaid into the AlphaFold model for Amuc_1420^{GH97} and suggests there would be significant interactions with the sugar in the +2 subsite (Supplementary Figure 19). This likely explains why poor activity is detected against disaccharide host-type substrates. This enzyme is predicted to either be localised to the outside of the cell or periplasm (Supplementary Table 1), but whole cell assays with cells grown on mucin did not show breakdown of raffinose, suggesting this enzyme is solely in the periplasm (Supplementary Figure 8).

Finally, there are three putative GH36 enzymes encoded in the *A. muciniphila* ATCC BAA-875 genome, and these have relatively low sequence identity between them (supplementary Table

3). The family members characterised so far display exo-acting activities towards α-Gal or α-GalNAc, with attention predominantly on the former in terms of raffinose-type oligosaccharides and galactomannan. One of these enzymes, **Amuc_0855^{GH36}, had specificity for Galili antigen Galα1,3Gal** in the defined substrate screen, with only trace activity against other substrates with α-Gal at the non-reducing end (Figure 1). Amuc_0855^{GH36} was able to liberate galactose from PGM III with or without sialidase and fucosidase treatment (Figure 1 and Supplementary Figure 11). On the phylogenetic tree, this enzyme clusters with α-galactosidases characterised to be active against both galactomannan and raffinose-type oligosaccharides (Supplementary Figure 20). Amuc_0855^{GH36} also had trace activity against these substrates, but no growth was observed on these substrates for *A. muciniphila* ATCC BAA-875, so this is unlikely to be physiologically relevant (Supplementary Figure 12). There is also no predicted signal sequence for Amuc_0855^{GH36}, so this enzyme likely resides in the cytoplasm (Supplementary Table 1).

<u>α-GalNAc'ases</u>

Two of the GH36 enzymes encoded in the *A. muciniphila* ATCC BAA-875 genome were found to have specificity towards α-GalNAc substrates. **Amuc_0216^{GH36} was able to act on all the substrates with α-GalNAc decorations we could test, including Forssman antigen and blood group A (BGA) hexasaccharide** (Figure 1). Forssman antigen is normally present in only 0.01 % of the human population but in that relatively small population this antigen would be present in the GI tract[32]. It can change susceptibility to some diseases and is also expressed in some cancers, but its true prevalence and impact on different diseases is understudied[33]. BGA is much more prevalent within the human population with this epitope being expressed in 10-40 % of the population depending on where you sample globally[34]. We were only able to test BGA type II in this study, but given the broad activity of Amuc_0216^{GH36}, it is likely that it will act on all BGA structures.

Amuc_0216^{GH36} could not remove as much core α-GalNAc from BSM peptide as other enzymes explored here (Supplementary Figure 6), indicating that these core sugars are not the true target for Amuc_0216^{GH36} and its specificity is for the non-reducing end of O-glycans or possibly core substrates on smaller peptides. **Notably, this was the only enzyme we found in *A. muciniphila* ATCC BAA-875 to degrade Tn antigen.** A phylogenetic tree of all the characterised bacterial GH36 enzymes with these from the *A. muciniphila* ATCC BAA-875 genome showed that Amuc_0216^{GH36} clustered with the other two examples of α-GalNAc'ases in the family (Supplementary Figure 20). One of these previously characterised enzymes also has activity on BGA and has been shown to convert BGA whole blood to universal type blood for transfusion applications[35]. The other α-GalNAc'ase example is from *Clostridium perfringens* and was found to be active against pNP-α-GalNAc[36]. The signal sequence of Amuc_0216^{GH36} is 74 % SPII indicating that it may be localised to the outside of the cell and would be the only BGA-degrading enzyme predicted to be outside the cell. Whole cell assays of mucin-grown *A. muciniphila* ATCC BAA-875 against Tn-antigen show no degradation, so this enzyme is likely not localised to the outside of the cell (Supplementary Figure 9).

Amuc_0517^{GH36} was much narrower in its **strict specificity for BGA hexasaccharide only** and predicted to be in the periplasm (Figure 1). This enzyme does not cluster on the phylogenetic tree with the other characterised α-GalNAc'ases, but most closely with the α-galactosidase from *Bifidobacterium breve* 203 (Supplementary Figure 20) that was shown to have activity against melibiose and raffinose[37]. Amuc_0517^{GH36} did display some activity towards raffinose-type oligosaccharides, but as the reactions did not go to completion overnight these are likely not real substrates. Finally, there was only trace activity seen against the core α-GalNAc from BSM for Amuc_0517^{GH36} (Supplementary Figure 6). **Amuc_0216^{GH36} and Amuc_0517^{GH36} were the only enzymes in *A. muciniphila* ATCC BAA-875 found to be active against BGA structures** and they must remove this sugar before the fucosidases can act. When tested against PGM III both enzymes could release GalNAc, with or without sialidase and fucosidase pre-treatment (Figure 1 and Supplementary Figure 11).

287 Amuc_0216^{GH36} was able to remove more than Amuc_0517^{GH36}, indicating both BGA and
288 other α-GalNAc capping structures are present in this substrate.
289
290 There are two putative GH109 enzymes (Amuc_0017^{GH109} and Amuc_0920^{GH109}) encoded by
291 the *A. muciniphila* ATCC BAA-875 genome, which have 66 % identity between them and
292 display **specificity towards GalNAcα1,3-Gal**. Whilst neither could act on α-GalNAc in the
293 context of blood group A (Figure 1), both could remove the core α-GalNAc from BSM
294 (Supplementary Figure 6). Amuc_0017^{GH109} could also remove the α-linked GalNAc from
295 Forssman antigen. No release of GalNAc was seen against PGM III, which is intriguing due to
296 their overlap in activity with the GH36 enzymes against defined substrates. **These**
297 **specificities represent the first examples of GH109 family members being unable to act**
298 **on blood group A structures** and rather, the ability of *A. muciniphila* ATCC BAA-875 to
299 access these substrates is facilitated by the two GH36 family members described above. Both
300 enzymes are predicted to have TAT signal peptides, so would be localised to the outside of
301 the cell (Supplementary Table 1). The whole cell assays showed no activity against
302 GalNAcα1,3-Gal but did for Forssman antigen (Supplementary Figure 8). This suggests that
303 these enzymes are localised to the outside, but why one of these substrates was degraded
304 but not the other may be due to preferences that were not resolved using the substrate
305 screening. Structural models are discussed in the context of the observed specificities
306 (Supplementary Discussion and Supplementary Figure 21).
307
308 <u>Activity against stomach antigen GlcNAcα1,4-Gal</u>
309 There are two putative GH89 enzymes (Amuc_0060^{GH89} and Amuc_1220^{GH89}) encoded for by
310 the *A. muciniphila* ATCC BAA-875 genome and both only show **activity against GlcNAcα1,4-**
311 **Gal disaccharide (stomach epitope)**, however Amuc_0060^{GH89} could not hydrolyse all the
312 GlcNAcα1,4-Gal overnight, which suggests this is not the preferred substrate (Figure 1 and
313 Supplementary Figure 13-16). Amuc_1220^{GH89} was also able to remove GlcNAc from PGM III
314 with or without sialidase and fucosidase pre-treatment, but Amuc_0060^{GH89} could not (Figure
315 1 and Supplementary Figure 11). Amuc_1220^{GH89} is predicted to be localised to the periplasm
316 (Supplementary Table 1) and whole cell assays using GlcNAcα1,4-Gal suggest that this
317 epitope is not broken down on the outside of the cell (Supplementary Figure 8 and 9).
318
319 There are currently five characterised GH89 family members, all with specificity for this α-
320 linked GlcNAc[38]. The first enzyme identified to have specificity for stomach epitope was one
321 from *C. perfringens*[39] and was subsequently crystallised with this disaccharide in the active
322 site (4A4A; Supplementary Figure 22)[40]. Other members of the GH89 family are active against
323 heparan sulfate and there is a crystal structure of this from *H. sapien* (4XWH)[41]. Mutations of
324 this GH89 gene leads to Sanfilippo syndrome or mucopolysaccharoidosis III, where recycling
325 of heparan sulfate is impaired, and there is currently no cure for this or treatments to slow
326 disease progression. Models of the two GH89 enzymes from *A. muciniphila* ATCC BAA-875
327 and these two crystal structures were superimposed and compared (Supplementary Figure
328 22). The substrate pocket of Amuc_1220^{GH89} has a similar conformation to the pocket from *C.*
329 *perfringens*, whereas the substrate pocket of Amuc_0060^{GH89} has a similar conformation to
330 the pocket from *H. sapien* (Supplementary Figure 22).
331
332 ***Investigating β-galactosidase activity***
333
334 Understanding the hydrolysis of the α-capping monosaccharides from the non-reducing ends
335 of O-glycans now leads us to the question of the deconstruction of the polyLacNAc chains.
336 Our previous work characterised the three GH16 family members from *A. muciniphila* BAA-
337 835 that hydrolyse within the O-glycan chains to produce a variety of fragments between 2
338 and 7 monosaccharides, including intrachain fucosylation and sulfation[14]. These GH16
339 enzymes can remove fragments with some α-capping epitopes still present, for example, we
340 demonstrated that the products included BGA structures. Furthermore, the GH16 enzymes
341 are also able to breakdown HMO structures. Subsequent, deconstruction of these O-glycan

342 and HMO fragments requires alternating β-galactosidase and β-HexNAc'ase activities with
343 concurrent defucosylation and desulfation.
344
345 There are nine putative CAZymes encoded in the *A. muciniphila* BAA-835 genome that were
346 highlighted as possible β-galactosidases from families GH2, GH35, and GH43 subfamily 24.
347 The five GH2 enzymes show low sequence identity, with the highest being 39 % between
348 Amuc_0824 and Amuc_1666 (Supplementary Table 4). However, the phylogenetic tree built
349 with these GH2 sequences, and other characterised GH2 family members, revealed clustering
350 of these enzymes with previously characterised β-galactosidases (Supplementary Figure 23).
351 The two GH35 enzymes share 37 % identity and the two GH43_24 enzymes have 87 %
352 identity.
353
354 Recombinant enzymes were screened against a variety of substrates to determine their
355 specificity (Figure 2; Supplementary Figure 24-27). Amuc_0539^{GH2} was found to be specific to
356 β1,4-linked Gal with either a Gal or GalNAc in the +1 position, with some activity against
357 lactose. We were unable to test longer defined substrates, but this enzyme did release
358 galactose from the HMO substrates. Conversely, Amuc_0771^{GH35} and Amuc_1666^{GH2} could
359 not hydrolyse these substrates and instead required a Glc/GlcNAc in the +1 position for β1,4-
360 linked Gal substrates. Amuc_1666^{GH2} was also able to act on TriLacNAc and Lacto-N-
361 neotetraose, but only Galβ1,3Glc out of those substrates with β1,3-linked Gal. Amuc_0771^{GH35}
362 was much broader with the capacity to also hydrolyse β1,3-linked Gal with Glc, GalNAc, and
363 Gal in the +1 position. This is also the only enzyme that could remove both Gal sugars from
364 the branching lacto-N-neohexaose. These branching structures, formed through a
365 GlcNAcβ1,6Gal linkage, are important due to their prevalence in humans (99 % of population)
366 and are referred to as "I antigens" (no branching is referred to as "i antigen"). Complementary
367 to the screening carried out in this report, Amuc_1666^{GH2} has previously been shown to have
368 a preferences for LacNAc[42] and Amuc_0771^{GH35} has also previously been shown to have
369 activity against mucin core 1 and core 2 structures, both β1,3 linkages[43].
370
371 Amuc_0824^{GH2} is highly specific towards β1,3-linked Gal with no preference observed for what
372 occupies the +1 positions and no activity against β1,4-linked Gal substrates. In line with our
373 findings, Amuc_0824^{GH2} has previously been shown to have preferences for Galacto-N-
374 biose[42]. Amuc_1686^{GH35} showed very similar specificity to Amuc_0824^{GH2} except that it could
375 also accommodate β1,6-linked Gal with Gal/GalNAc in the +1 (no β1,6-linked substrates with
376 a Glc-type sugar were available to test) and has previously been shown to have a preference
377 for Galβ1,3GalNAc[44]. Amuc_0290^{GH2} can accommodate β1,4 and β1,6-linked Gal substrates
378 with less preference for β1,3-linked Gal substrates. Finally, Amuc_1667^{GH2} was inactive
379 against most substrates, with only some activity towards the ganglioside core structure GA1.
380 Concerning biantennary complex N-glycans, all three enzymes that can accommodate other
381 Galβ1,4GlcNAc structures, could remove Gal from N-glycan that had been released from the
382 protein and de-sialylated. This is unlikely to be physiologically relevant, however, as *A.*
383 *muciniphila* BAA-835 will not grow on this glycoprotein substrate despite similarities to O-
384 glycans (Supplementary Figure 12). None of these enzymes showed activity towards 3-FL,
385 Lewis A, or Lewis X based structures, which confirms that there is a strict order of degradation,
386 with fucosidases acting on these substrates first, followed by galactosidases.
387
388 Regarding larger more complex substrates, Amuc_0771^{GH35} was able to release the largest
389 concentration of galactose from BSM (core 1; Supplementary Figure 6). Amuc_0824^{GH2} was
390 also able to release galactose, which complements the understanding that this substrate
391 constitutes core 1 structures, but Amuc_0771^{GH35} can release a higher concentration of
392 galactose indicating that it can accommodate the mucin polypeptide more easily than
393 Amuc_0824^{GH2} (Supplementary Figure 6). Galactose was released from both types of HMOs
394 and PGM III for most of the GH2 and GH35 family members (Figure 2 and Supplementary
395 Figure 29). The GH2 and GH35 enzymes also showed activity against Lactosylceramide

396 (Ganglioside and Globoside core structure), with Amuc_0290^{GH2}, Amuc_0771^{GH35}, and
397 Amuc_1686^{GH35} releasing the most (Figure 2).
398
399 Of these enzymes, Amuc_0824^{GH2} is predicted to localise to the cell surface, while the rest
400 are predicted to reside in the periplasm (Supplementary Table 1). Notably, the mucin-grown
401 whole cells showed rapid degradation of Lacto-N-biose, one of the substrates for this enzyme,
402 and this was reproducible over two different experiments (Supplementary Figure 8 and 9).
403 This was the fastest degradation of a substrate, with complete hydrolysis recorded in 30
404 minutes. Lactose and Galβ1,4GalNAc are not substrates for this enzymes and the whole cell
405 assays did not show degradation of these substrates (Supplementary Figure 9).
406 Amuc_0824^{GH2} cannot be directly linked to this activity, but techniques that have been
407 developed for the genetic manipulation of *A. muciniphila* may lend themselves to these types
408 of experiments in the future.
409
410 Of the fourteen enzymes identified as having >90 % sequence identity conservation across
411 the different *Akkermansia* species we assessed, all the GH2 and GH35 family members, apart
412 from Amuc_0771^{GH35} and Amuc_1667^{GH2}, were in this group (Supplementary Figure 5) and
413 Amuc_0290^{GH2}, Amuc_0539^{GH2}, Amuc_1666^{GH2} and Amuc_1667^{GH2} were highlighted as
414 important for growth on mucin[20]. This emphasises the important role these β-galactosidases
415 play in nutrient acquisition for *A. muciniphila* BAA-835 (Supplementary Tabel 1). No activities
416 for the GH43_24 enzymes were observed against the substrates tested here, but
417 Amuc_0698^{GH43} was active against pNP-Gal (Supplementary Figure 28). Interestingly, whilst
418 Amuc_0697^{GH43} was highlighted as important for growth on mucin[20], it was not well conserved
419 between *A. muciniphila* strains (Supplementary Figure 5). See Supplementary Discussion and
420 Supplementary Figure 29-30 for analysis of models of these enzymes.
421
422 ***Investigating β-HexNAc activities***
423
424 The GH20 family is the largest represented family in the *A. muciniphila* BAA-835 genome, with
425 11 putative enzymes encoded. The activity predominantly observed in this family is exo-acting
426 β-GlcNAc'ases with some able to act on GalNAc. The identity between the enzyme sequences
427 from *A. muciniphila* BAA-835 is generally low ~20-30 % with the exceptions of 72 % between
428 Amuc_2018^{GH20} and Amuc_2019^{GH20}. We also included Amuc_0052^{GH84}, Amuc_0803^{GH123},
429 and Amuc_2109^{GH3} in this screening process as the GH84 family is associated with activities
430 against β-linked GlcNAc, the GH123 family is associated with activities against β-linked
431 GalNAc, and Amuc_2109^{GH3} has been shown to have relevant activities
432 (pNPGlcNAc>GalNAc)[45]. These enzyme sequences were put into phylogenetic trees
433 alongside other characterised members of their families (Supplementary Figures 31-33).
434 These showed that the enzyme sequences from the *A. muciniphila* BAA-835 genome
435 clustered with previously characterised enzymes with β-HexNAc'ase activities.
436
437 The first step in this analysis was to screen a panel of defined oligosaccharides
438 (Supplementary Figures 34-36). The enzyme with the broadest activity was Amuc_0868^{GH20}
439 that could degrade most of the substrates provided, with the exception of GlcNAcβ1,3Man and
440 β1,2-linked GlcNAc on biantennary N-glycans, to which it was only partially active. This
441 enzyme was previously shown to have a preference for GlcNAc over GalNAc[46].
442 Amuc_2019^{GH20} showed a similar pattern but exhibited only partial activity against the P
443 antigen. Again, Amuc_2136^{GH20} showed a similar pattern to Amuc_0868^{GH20}, but with no
444 activity against the β1,4-linked GalNAc on GA2. Amuc_2136^{GH20} has also previously been
445 shown to prefer GlcNAc over GalNAc[47]. Amuc_0369^{GH20} was also relatively broad-acting, but
446 with a different pattern; it was able to access most of the host-type glycans tested and
447 demonstrated relatively less activity against chitooligosaccharides. Amuc_2018^{GH20} showed a
448 similar pattern to these three broad-acting GH20 enzymes, but with negligible activity against
449 chitooligosacchairdes and no activity against β1,4-linked GalNAc and P antigen.
450 Amuc_2018^{GH20} has previously been shown to prefer GlcNAc over GalNAc and have some

activity against N-glycan structures[46]. Amuc_0052^{GH84} showed specificity towards GlcNAc over GalNAc substrates and prefers smaller substrates (disaccharides). From the substrate screen, Amuc_1669^{GH20} is selective for host glycans with a preference for GlcNAcβ1,3Gal substrates. It can also hydrolyse GalNAc in the context of globoside P antigen, but not ganglioside GA2. The predicted shapes of these glycans are quite different and may play a role in these specificity differences (Supplementary Figure 4). Amuc_2148^{GH20} shows very similar pattern of activity to Amuc_1669^{GH20} but is even more specific as it will not act on P antigen. Amuc_1924^{GH20} prefers GalNAc in the -1 subsite but was unable to accommodate P antigen and this also may be due to overall glycan conformation. Amuc_1032^{GH20}, acted only upon GlcNAcβ1,3Gal in the context of TriLacNAc and lacto-N-triose, but not GlcNAcβ1,3Gal alone, indicating that this enzyme prefers substrates that are at least three sugars long (i.e. a sugar occupying the +2 subsite). Finally, Amuc_0803^{GH123} had strict specificity towards the GalNAc substrates, including GA2. Exact specificities could not be determined for Amuc_0397^{GH20} and Amuc_1815^{GH20}. Both only had trace activity against GalNAcβ1,3Gal, but it may be that they are active on longer versions of this substrate. Specificities could also not be found for Amuc_2109^{GH3}, with only trace activity against chitooligosaccharides and GaNcAcβ1,3Gal.

The β-HexNAc'ases were also tested against HMO substrates pre-treated with α-fucosidases and β-galactosidases (Figure 3). The specificities observed for the defined substrate screen facilitates the probing of more complicated structures in more detail and exemplifies how these enzymes may be useful in other contexts. For example, Amuc_0803^{GH123} showed no obvious activity towards the HMO substrates, implying that no β-linked GalNAc is present. Furthermore, Amuc_1032^{GH20} provides the same pattern as the broad-acting β-HexNAc'ases, so the glycans in this assay were at least 3 sugars long. Finally, Amuc_0052^{GH84} did not breakdown the glycans as extensively as the broad-acting β-HexNAc'ases, emphasizing its preference for smaller substrates.

There are currently three crystal structures available for Amuc_0868, Amuc_2018, and Amuc_2136, but a broad understanding of specificities for *A. muciniphila* GH20 enzymes has not been explored before[47-49]. It was generally difficult to make any observations about how the different structures were related to the observed specificities. However, there was one feature where we felt we could speculate, but frustratingly it was untestable in this work. One aspect we were not able to explore biochemically in this study is the removal of 6S-GlcNAc by any of the GH20 enzymes. However, there are three crystal structures of two GH20 6S-GlcNAc'ases available from *Bifidobacterium bifidum* and *Bacteroides thetaiotaomicron* with 6S-GlcNAc product in the active site. To hypothesise which GH20 enzymes from *A. muciniphila* may accommodate this sulfated monosaccharide in the -1 position, we superimposed the three crystal structures for Amuc_0868^{GH20}, Amuc_2018^{GH20}, and Amuc_2136^{GH20} and AlphaFold models for the other eight GH20 enzymes on to the 6S-GlcNAc'ases. From this superimposition, we were able to observe space for the sulfate group in the models for Amuc_1032^{GH20} and Amuc_1669^{GH20}, but not for any of the other structures or models (Supplementary Figure 37). Intriguingly, these two enzymes cluster closely with the characterised 6S-GlcNAc'ases on the phylogenetic tree (Supplementary Figure 33).

In summary, a range of different β-HexNAc'ase specificities, encoded for by the *A. muciniphila* BAA-835 genome, were detected and evidently the specificities of these enzymes are generally towards the glycan structures found in mucins rather than complex N-glycans or chitin. Between them, these β-HexNAc'ases could access all the substrates provided.

Complete breakdown of porcine O-glycans by CAZymes from A. muciniphila ATCC BAA-875

Two putative GH31 enzymes are encoded in the *A. muciniphila* ATCC BAA-875 genome. Amuc_1008^{GH31} is from subfamily 18 and other members of this subfamily from *Bacteroides*

506 *caccae*, *Phocaeicola plebius, Enterococcus faecalis, Clostridium perfringens, and Bombix*
507 *mori* (domestic silk moth) have specificity for removing the α-GalNAc from peptide[50-52].
508 *Bc*GH31, *Pp*GH31, *Ef*GH31, *Cp*GH31 and *Bm*GH31 have 70, 65, 43, 43, and 41 % sequence
509 identity to Amuc_1008^{GH31}, respectively. A phylogenetic tree of characterised bacterial GH31
510 enzymes shows clustering of activities and Amuc_1008^{GH31} clusters with the other
511 characterised enzymes from GH31_18 (Supplementary Figure 38). **We also found that**
512 **Amuc_1008^{GH31} had specificity for the core α-GalNAc linked to peptide (BSM) with no**
513 **activity against any other substrates** (Figure 1). Upon removal of the sialic acid and β-Gal
514 from BSM, Amuc_1008^{GH31} was able to remove α-GalNAc. Quantification of this GalNAc
515 release using HPAEC-PAD showed that Amuc_1008^{GH31} removed the highest concentration
516 of all the enzymes tested here (Supplementary Figure 6). Furthermore, Amuc_1008^{GH31} could
517 not hydrolyse Tn antigen, so therefore requires more than one amino acid for activity (Figure
518 1). For the more complex substrate PGM III, Amuc_1008^{GH31} could not liberate any GalNAc
519 when incubated on its own, but when a cocktail of *A. muciniphila* ATCC BAA-875 enzymes
520 was used, GalNAc could then be released (Figure 1). **This observation is one of the most**
521 **notable in this report as it demonstrates that the enzymes characterised from *A.***
522 ***muciniphila* ATCC BAA-875 are capable of complete O-glycan degradation down to the**
523 **polypeptide, which to the best of our knowledge has not been demonstrated before.**
524 This enzyme is predicted to be in the periplasm and there was no obvious removal of GalNAc
525 from BSM during whole cells assays, supporting this prediction (Supplementary Figure 10).
526
527 A model of Amuc_1008^{GH31} was compared to the structure of the *E. faecalis* enzyme, which is
528 the only structure from this subfamily. An overlay indicates that the structure of
529 Amuc_1008^{GH31} has extra C-terminal domains compared to *Ef*GH31 (Supplementary Figure
530 39). Tn antigen that was crystallised in the active site of *Ef*GH31 overlays well into the active
531 site of Amuc_1008^{GH31} (Supplementary Figure 39). The residues interacting with the Tn-
532 antigen in the *Ef*GH31 crystal structure are all conserved in Amuc_1008^{GH31}.
533
534 ***Interaction with GAGs***
535
536 *A. muciniphila* ATCC BAA-875 did not grow directly on any of the glycosaminoglycans (GAGs)
537 tested: hyaluronic acid, chondroitin sulfate C, or heparin (Supplementary Figure 12). However,
538 its genome does encode Amuc_0778^{PL38} and Amuc_0863^{GH105}. The PL38 enzymes
539 characterised to date have specificities for glucuronan and alginate[53,54], but screens of these
540 substrates did not reveal any activity (Supplementary Figure 40). Characterised GH105
541 enzymes have been found active against the unsaturated products of PLs[55]. We found no
542 activities for GH105 alone against any of the other substrates used in this work. Therefore, we
543 tested host GAGs that are known substrates for PLs and are commonly available to the
544 microbiota in the large intestine. **Amuc_0778^{PL38} was found to be active against hyaluronic**
545 **acid (HA), chondroitin sulphate A (CSA), chondroitin sulphate C (CSC) and dermatan**
546 **sulfate (DS)** (Figure 4). **Amuc_0863^{GH105} had little or no activity when it was incubated**
547 **alone with these substrates but was able to act on the products of the PL38 from all**
548 **four substrates (Figure 4).**
549
550 The Michaelis-Menten kinetic analysis of Amuc_0778^{PL38} against the different GAGs revealed
551 the highest catalytic efficiency (k_{cat}/K_M) towards HA, followed closely by CSA and four- and
552 eight-fold lower towards CSC and DS, respectively (Supplementary Table 7, Supplementary
553 Figure 40). An attempt to resolve the kinetic parameters of Amuc_0863^{GH105} was performed
554 by running the reactions of Amuc_0778^{PL38} to completion (overnight), then adding
555 Amuc_0863^{GH105} and following the loss of the double bond of the unsaturated uronic acid at
556 235nm. While linear decreasing initial rates were observed for the lower substrate
557 concentrations, the higher concentration reactions were not reliable because of the
558 absorbance ceiling of the spectrophotometer. Attempts to downscale or dilute the reactions
559 did not provide data for which a reliable kinetic model could be fitted (Supplementary Figure
560 40). However, by isolating the highest observed initial rate for each substrate, a clear

561 selectivity of Amuc_0863^{GH105} was observed towards CSA, followed by the three other GAGs
562 at 36-40% relative activity (Figure 4, Supplementary Table 6). These results indicates that this
563 enzyme prefers an unsaturated GlcA in the -1 subsite and a 4SGalNAc in the +1 subsite.
564
565 The products of Amuc_0778^{PL38} were then characterised in more detail using LC-MS/MS with
566 UV detection (Figure 4 and Supplementary Figure 41). Time-course reactions were taken to
567 assess the product formations. For HA, Amuc_0778^{PL38} produced dimers, tetramers and
568 hexamers. For CSA, the sulfated disaccharide and double sulfated tetrasaccharide were
569 produced at equal rates initially, but as the disaccharide continud to increase in concentration,
570 the tetrasaccharide decreased, indicating degradationto disaccharides also. A different pattern
571 was observed for CSC, where the activity was slower and the tetrasaccharide was the
572 dominant product initially, indicating that the differing sulfation patter in this substrate reduces
573 activity. The activity was even slower for DS and the double sulfated tetrasaccharide was the
574 dominant product until much later in the time-course. The only difference between CSA and
575 DS is the presence of GlcA and IdoA, respectively, therefore based on this data,
576 Amuc_0778^{PL38} appears to have a preference for GlcA over IdoA containing substrates.
577
578 The two GH89 enzymes were also tested against GAGs both with and without Amuc_0778^{PL38}
579 and Amuc_0863^{GH105}, but no activity was observed (Supplementary Figure 40).
580
581 ***Starch-degrading activities***
582
583 Encoded in the *A. muciniphila* ATCC BAA-875 genome are two GH families that are currently
584 characterised to solely have specificity towards starch-type polysaccharides – one GH77 and
585 three GH13 enzymes. These enzymes were tested against several different substrates
586 composed of α-linked glucose, however activity was only found for Amuc_1621^{GH77}, primarily
587 against amylose and produced oligosaccharides of one to five glucose units (Supplementary
588 Figure 42). There was no activity against dextran or amylopectin, but potentially trace activity
589 against pullulan. There was no growth on these substrates (Supplementary Figure 12).
590 Interestingly, two of these enzymes were highlighted as important for growth on mucin using
591 a TnSeq library[20].
592
593 Concerning GAG and starch utilisation, why the *A. muciniphila* ATCC BAA-875 genome would
594 encode enzymes specific for these substrates when it does not use them as a nutrient source
595 directly, remains unresolved at present. One possibility, however, is that the large mucin
596 structures being taken up by *A. muciniphila* into the mucinosomes also commonly contain
597 GAG and starch contaminants. The PL38, GH105, sulfatases, GH13 and GH77 enzymes
598 would be present internally to break these down either for use as a nutrient source or to expel
599 them. We attempted to test this idea using different proportions of GAG-to-mucin, but the
600 decrease in growth with increasing GAG concentration suggests that this theory is incorrect
601 (Supplementary Figure 12).
602
603 ***Tackling sulfated substrates***
604
605 We highlighted twelve putative sulfatases from the *A. muciniphila* ATCC BAA-875 genome
606 using previously published upregulation data[20,56-58] and SulfAtlas database[59]. These enzymes
607 could be assigned to seven different families and one unknown (Supplementary Table 1). The
608 different subfamilies cluster on a phylogenetic tree and the activities (mainly explored in
609 *Bacteroides thetaiotaomicron*) within a subfamily are similar (Supplementary Figure 43)[60,61].
610 The sequence identity between the sulfatases is low, with the highest being ~60 % for pairs in
611 the same subfamily (Supplementary Table 8). The defined sulfated substrates tested were
612 sulphated monosaccharides and Lewis structures (Supplementary Figure 44-46). We first
613 screened the sulfatases against the sulfated monosaccharides, and found Amuc_1655^{Sulf16}
614 and Amuc_1755^{Sulf16} displayed activity against 4S-Gal, Amuc_1755^{Sulf16} also displayed activity
615 against 4S-GalNAc, and Amuc_0451^{Sulf20} displayed activity against 6S-GlcNAc, which aligns

with activities previously observed for these subfamilies[60]. We then used combinations of the sulfatases, α-fucosidases and β-galactosidase encoded in the *A. muciniphila* ATCC BAA-875 genome to determine how the different sulfated Lewis structures are broken down. For the 3'S and 6S-Lewis A and X glycans, Amuc_0392[GH29] removes the fucose to produce 3'S and 6S-LacNAc and LNB. From here, sulfatases from subfamily 20 act upon the 3'S disaccharides, whilst, in contrast, β-galactosidases act upon the 6S disaccharides first. In this study, we could not find a way to breakdown the 6'S-Lewis structures. Previous activities against these substrates have been observed for BT1624 from subfamily 15[60], but these reactions did not progress to completion. Further assays with O-glycan fragments may yield additional answers. Finally, we included the sulfatases with the Amuc_0778[PL38] (with and without Amuc_0863[GH105]) in assays with CSC and HS and no obvious activity could be seen by TLC (Supplementary Figure 47).

Growth of *A. muciniphila* ATCC BAA-875 on high-mannose N-glycoproteins

When testing the growth of *A. muciniphila* ATCC BAA-875 against a variety of substrates to support the biochemical work, we included glycoproteins that are decorated with mammalian complex N-glycans, as the sialidases, galactosidases, and GlcNAc'ases were observed to breakdown these substrates. The detected enzyme activities were unsurprising as mammalian complex N-glycans have very similar glycan structures to O-glycans. For completeness, a glycoprotein with high-mannose N-glycans was also included. Unexpectedly, ***A. muciniphila* ATCC BAA-875 was observed to grow on the high-mannose N-glycoprotein and not the two glycoproteins with complex N-glycans** (Supplementary Figure 12). Furthermore, we also observed growth on *Saccharomyces cerevisiae* mannan (*Sc*mannan), which is an N-glycan core structure with large antennae composed of α-linked mannose and a recognised nutrient source for other gut microbes[62].

Whole cell assays using cells that had been grown on high-mannose N-glycoprotein revealed a smear on the TLC that appeared overnight (Supplementary Figure 48). This smear did not migrate at the same rate as the free high-mannose N-glycan control (suggesting a higher molecular weight), but when the sample was treated with PNGaseL it ran the same by TLC. This implies that *A. muciniphila* ATCC BAA-875 is using the protein as a nutrient source and the by-product of this is high-mannose N-glycan still attached to a small peptide. The molecular mechanism underpinning why *A. muciniphila* ATCC BAA-875 can degrade this substrate (and not other proteins or glycoproteins included in this work) is beyond the scope of this study, but it is likely due to a particular glycopeptidase or peptidase. However, this observation indicates a possible important cross feeding relationship *A. muciniphila* may have with other human gut microbes. N-glycans are prevalent decorations on secreted proteins and high-mannose N-glycans are universal to eukaryotes, so microbes in the human gut would see this substrate from the host (secreted proteins and sloughed off dead cells) and any type of diet (animal- or plant-derived).

The whole cell assays for *Sc*mannan did not show any detectable glycobiology by TLC. All the CAZymes associated with the hydrolysis of α-linkages were tested against a panel of α-mannose glycans in trying to decipher this growth, but there were no positive results (Supplementary Figure 16). SDS-PAGE with Coomassie staining of *Sc*mannan also showed no obvious protein in the sample (Supplementary Figure 48). To rule out very small free peptides being used as a nutrient source, the *Sc*mannan was dialysed to try and remove this, but growth was comparable to the original substrate (Supplementary Figure 12). It may be that there is a small amount of peptide attached to the *Sc*mannan that *A. muciniphila* ATCC BAA-875 can access.

Discussion

In our previous work, we characterised the GH16 endo-O-glycanases from *A. muciniphila* ATCC BAA-875 (Amuc_0724^{GH16}, Amuc_0875^{GH16} Amuc_2136^{GH20}) that hydrolyse within the polyLacNAc chains of mucin to produce fragments of O-glycans[14]. Here, we expand on this work by taking a more holistic approach to assessing the enzymology of this organism (Figure 5). One of the most significant aspects of this report was showing that we could use a cocktail of CAZymes from *A. muciniphila* ATCC BAA-875 to reach the core GalNAcs of the O-glycan, thus demonstrating complete degradation of the constituent glycan of this complex substrate (Figure 1). By characterising substrate specificity, important insights were generated regarding the order of enzyme activity on different substrates. For example, 1) the removal of the BGA GalNAc or BGB galactose must precede fucose removal and 2) deconstruction of ganglioside glycans requires alternating sialidase and β-galactosidase activities (Figure 5).

One of the major impacts of this work will be the ability to characterise different mucins in much more detailed and in a higher throughput manner. We hope that these enzymes provide new glycotools for the community and have far-reaching applications, such as monitoring the change in mucin O-glycans between healthy and diseased tissue, or the detection of disease associated epitopes. This applies particularly to those enzymes active against the α-capping monosaccharides at the non-reducing ends of the O-glycans. These are the epitopes that can strongly influence the susceptibility to pathogens, for example.

An intriguing physiological phenomenon of *A. muciniphila* ATCC BAA-875 is its unusual substrate import mechanisms. In a paper describing the first genetic manipulation of *A. muciniphila*, Davey *et al.* highlighted a possible import system that implicated pili in the import of mucin[20]. They found that mucin was accumulating in intracellular compartments, dubbed "mucinosomes". Nine CAZymes were associated with this likely import system (Mul1A) (Supplementary Table 1). Interestingly, these CAZymes were also either very highly conserved across *Akkermansia* species, predicted to be localised to the outside of the cell, and/or highlighted as important for growth on mucin (Supplementary Figure 49). It is likely that some of these different observations are connected. Complementing the observation of mucinosomes, was the types of enzyme activity detected on the surface of *A. muciniphila* ATCC BAA-875 in this report. Prominent activities included GH16 and sialidase activities and fucosidase activities were also observed by Shuoker *et al.*[13], but other glycobiology was generally absent. This suggests that most of the degradation of mucin is taking place inside the cell. Finally, growth of *A. muciniphila* ATCC BAA-875 on human milk has also been observed [23], so there must also be a way to import human milk oligosaccharides and the very structurally similar GH16-derived O-glycan fragments.

In this report we have acknowledged several puzzles where activity could not be elucidated (Supplementary Table 1). It is possible that these enzymes are involved in another function besides mucin breakdown. *A. muciniphila* ATCC BAA-875 has 50 putative glycosyltransferase enzymes and it is probable that at least some of these will contribute to synthesising capsular and/or exopolysaccharides. The composition of these carbohydrate structures from microorganisms is a relatively understudied area of glycobiology and the types of structures produced by different microbes is likely highly variable. Therefore, some of the 'puzzle' enzymes discussed in this report may target these structures, possibly to recycle the component sugars and allow the cells to grow and divide. For example, there was no activity observed for Amuc_1868^{GH57} but it is highly conserved across *Akkermansia* species and was important for growth on mucin[20].

Materials and Methods

Substrates

A list of substrates and their sources are listed in Supplementary Table 2. Porcine gastric mucin III (Sigma) was prepared by dissolving overnight in sterile DI water. This is too cloudy to grow bacteria on, so the precipitate is removed by centrifugation at 1,500 x g for 5 minutes in sterile 15 mL falcon tubes. Polyglucuronic acid was prepared as described[53].

Recombinant protein expression and purification

Recombinant plasmids were transformed into Tuner cells (Novagen) in Luria-Bertani (LB) broth containing with 50 or 100 µg/mL kanamycin or ampicillin, respectively. The plates were used to inoculate 1 L flasks (also including relevant antibiotics) and grown at 37 °C with shaking at 180 rpm to mid-exponential, this was then cooled to 16 °C, and isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.2 mM. Recombinant His-tagged protein was purified from cell-free extracts using immobilized metal affinity chromatography (IMAC using Talon resin; Takara Bio). The buffer used during the purification process was 20 mM Tris-HCl, 100 mM NaCl, pH 8. Cell-free extracts (~30 mL) were passed through a 5 mL bed volume of resin. This was then washed with 50 mL of buffer and the protein eluted with 10 mM and then 100 mM imidazole. The purification is assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE; precast 4 to 20 % gradient from Bio-Rad stained with Coomassie Brilliant Blue), but proteins typically elute in the 100 mM fraction. The proteins were then concentrated using centrifugal concentrators (Vivaspin®, 10 kDa MWCO), the absorbance determined at 280 nm (NanoDrop™ One, Thermo Scientific), and concentrations calculated using extinction coefficients taken from ProtParam.

Recombinant enzyme assays

The concentrations of defined oligosaccharide substrates used in the assays are presented in Supplementary Table 2. The activities of the recombinant enzymes were typically assessed in 20 mM 4-morpholinepropanesulfonic acid (MOPS, pH 7) at 37 °C, and a final enzyme concentration of 1 µM. The volume varies according to how much was required for subsequent analysis.

Thin layer chromatography

For defined oligosaccharides, 3 µL of an assay containing 1 mM substrate was spotted onto silica plates. For assays against glycoproteins, GAGs and starch substrates, this was increased to 9, 6, and 9 µL, respectivley. The plates were resolved in running buffer containing butanol/acetic acid/water. Different ratios were used for different substrates - 2:1:1 (typically resolved once) was used for most of the defined oligosaccharides and 1:1:1 (typically resolved twice) was typically used for substrates where we expected large products (e,g, PGM III). All TLCs were stained using a diphenylamine-aniline-phosphoric acid (DPA) stain[63].

Growth of Akkermansia muciniphila on a variety of substrate

A. muciniphila was grown on chopped meat broth (CMB) as described in[14]. Overnight starter cultures were inoculated from a glycerol stock into 5 mL of chopped meat broth (CMB) and grown in an anaerobic cabinet (Whitley A35 Workstation, Don Whitley Scientific). Growth curves were collected in 96-well plates using Cerillo Stratus plate readers. 300 µL of 2x CMB (minus the monosaccharides) was mixed with 300 µL substrate and 200 µL was pipetted into three wells to provide replicates. Growth curves were also repeated with different starter cultures and on different days.

Whole cell assays

Overnight starter cultures were inoculated from a glycerol stock into 5 mL of CMB and grown under anaerobic conditions. This was then used to inoculate a 5 mL culture of *A. muciniphila* in CMB (minus the monosaccharides) with substrate and cells were pooled, harvested at mid-exponential, washed twice with PBS, and 250 µL of 2x PBS was used per 5 ml culture to resuspended. The assays were then mixed 50:50 cells to substrate, where substrate

concentrations were the same as used in the recombinant enzyme assays. PGM III, BSM, hyaluronic acid, chondroitin sulfate, RNaseB, and *Sc*mannan assays were 500 μL total and 50 μL samples were taken for each time point. Assays with defined oligosaccharides and HMOs were 200 μL and 20 μL samples were taken for each time point. Assays were carried out at 37 °C and to stop the reactions the samples were boiled.

High-performance anion exchange chromatography with pulsed amperometric detection

To analyse monosaccharide release, sugars were separated using a CARBOPAC PA-1 anion exchange column with a PA-1 guard using a Dionex ICS-6000 (ThermoFisher) and detected using pulsed amperometric detection (PAD). Flow was 1 mL/min and elution conditions were 0-25 min 5 mM NaOH and then 25-40 min, 5-100 mM NaOH. The software was Chromeleon Chromatography Data System. Monosaccharide standards were 0.1 mM, all data was obtained by diluting assays 1/10 before injection, apart from the ceramide samples, which were diluted 15 μL in 100 μL.

Steady-state kinetics of AmPL38

Initial velocities of purified recombinant Amuc_0778^{PL38} were quantified on HA, CSA, CSC and DS substrates, with concentrations ranging from 0.25 to 12 $g \cdot l^{-1}$ at 37 °C, 100 mM NaCl, and UB4 buffer at pH 7[64]. The average initial velocities, quantified in milli-absorbance units (mAU) at A235nm per second, were converted to μM of product generated by measuring the amount of Δ4,5 bonds formed per second, using the experimentally confirmed extinction coefficient for unsaturated glucuronic acid of 6150 M-1cm-1[53]. Kinetic parameters were determined by plotting initial velocities against substrate concentrations and fitting the Michaelis-Menten model using GraphPad Prism.

Initial rates of AmGH105

The steady-state reactions of Amuc_0778^{PL38} were sealed to prevent evaporation and incubated overnight at 37°C. Subsequently, Amuc_0863^{GH105} was added to a final concentration of 1 μM, and the decrease in A235nm was monitored for a minimum of two hours at 37°C. The substrate concentration was calculated by converting the initial absorbance of the Amuc_0778^{PL38} steady-state reactions, minus absorbance backgrounds, to μM double bonds. The initial rates of Amuc_0863^{GH105} were then calculated as the loss of double bonds in μM per minute in absolute values.

Liquid Chromatography-Mass Spectrometry (LC-MS) of GAG substrates

Duplicate time-course reactions for Amuc_0778^{PL38} were prepared under the same conditions as for the kinetics at 2 $g \cdot l^{-1}$ substrate concentrations. Reactions were terminated by heating the samples at 95°C for 5 min. Amuc_0863^{GH105} was added to the 20-hour reaction to a final concentration of 1 μM and left to run for two hours before heat inactivation. The final sample preparation and LC-MS analysis were carried out using an Iontrap coupled with GlycanPac chromatography as previously described[65]. Compounds were observed as single or double charge m/z, primarily as deprotonated adducts. The compounds were identified by MS and MS^2 fragmentation, if possible. Extracted ion chromatograms of identified compounds were prepared, and the areas of the peaks were used to quantify the products (Supplementary Figure 41). Fragments follow the nomenclature of Domon and Costello[66].

Bioinformatics

Putative signal sequences were identified using SignalP 6.0[67]. Identities between different sequences were determined using Clustal Omega using full sequences[68]. The CAZy database (http://www.cazy.org) was used as the main reference for CAZymes[38]. Phylogenetic trees were completed in SeaView[69] and final trees were produced in the Interactive Tree of Life (iTOL)[70].

Pangenome analysis: To analyse the CAZymes across different species and strains, CAZyme sequences were downloaded from UNIPROT using the get.seq function from the Bio3d package in RStudio[71]. Sequence selection was based on the accession numbers of the

CAZymes from the most up to date version of the CAZy repository at the time. Sequence txt files were converted to a FASTA format using the TabulartoFasta function from the following GitHub repository: https://github.com/lrjoshi/FastaTabular . All multiple sequence alignment and phylogenetic analysis was performed using Clustal Omega[68]. Percentage identity matrices were downloaded directly from the programme and merged with the original file from CAZY according to shared accession number. The presence or absence of different enzymes and how well conserved they were across strains and species was curated by hand.

Structural comparisons: Where possible, protein models were retrieved from the AlphaFold Protein Structure Database[72,73]. Where the models had not been built, this was completed using AlphaFold2 Colab[74]. Previously published protein structures were retrieved from the RCSB Protein Data Bank. Different structures or models were superimposed using COOT[75] and PyMOL was used to look at structures/models and generate figures[76]. Glycan models were built using the Carbohydrate Builder on http://glycam.org.

Acknowledgements

Many thanks to Professor William Willat's group (Newcastle University, UK) for providing some of the substrates for us to test. Thank you to Carl Morland (Newcastle University, UK) for technical support. Many thanks to Professor Anne S. Meyer and Professor Willem van Schaik for their insightful comments about the manuscript. The work was funded by The Academy of Medical Sciences (SBF0061175), the Wellcome Trust/Royal Society (224240/Z/21/Z), and T.O.Ojuri is funded by the BBSRC Midlands Integrative Biosciences Training Partnership (MIBTP) with his studentship in collaboration with industrial partners Ludger (Oxford, UK). Finally, thank you to fantastic parents/grandparents, Clare and Nick Crouch, who facilitated some of our experiments by babysitting late into the night.

Supplementary Information

Pangenome analysis

One way to assess the importance of these enzymes to the species for mucin degradation is to review how prevalent and conserved the enzymes are over multiple *A. muciniphila* isolates (Supplementary Figure 5). In this analysis we also included *Akkermansia biwaensis* WON2089, twelve *Akkermansia massiliensis* strains, and five *Candidatus Akkermansia timonensis* strains. The latter two species have recently been reclassified as their own species but were previously cluster II and III of *A. muciniphila*[77]. The isolates included in this analysis were from a range of human faecal sources, including different collection efforts, geographical location, adults, and children.

The striking aspect of this analysis was how highly most of these enzymes were conserved between *A. muciniphila* strains (>90 %). There were a few examples of enzymes not being present in many strains (e.g. Amuc_0875^{GH16} and Amuc_0146^{GH29}), but generally enzymes were present. As another generalisation, the other species assessed did have close homologues of the CAZymes in *A. muciniphila* BAA-835, with *A. massiliensis* having higher identity than *A. biwaensis* and *Candidatus A. timonensis*.

Additional enzymes were also identified during this analysis:

- A GH2 in some *A. muciniphila* strains
- A GH2 in *A. biwaensis* and *Candidatus A. timonensis*.
- A GH16 in *A. biwaensis* and *Candidatus A. timonensis*.
- A GH27 in *A. massiliensis*
- A GH27 in *A. biwaensis* and *Candidatus A. timonensis*.
- A GH29 in many of the *A. muciniphila* strains
- A GH29 in most *A. massiliensis*
- A GH29 in *A. biwaensis* and *Candidatus A. timonensis*.
- A GH36 in *A. biwaensis* and *Candidatus A. timonensis*.
- A GH36 in *A. massiliensis, A. biwaensis* and *Candidatus A. timonensis*.
- A GH84 in most of all strains and species assessed.
- A GH95 in *A. massiliensis*
- A GH95 in *A. biwaensis* and *Candidatus A. timonensis*.

CAZymes where no activities could be identified

GH31: The second GH31, Amuc_1870^{GH31} is from subfamily 1, which has 84 characterised members and all bacterial members exhibit α-glucosidase activity. This enzyme did have activity against pNP-α-Glc (Supplementary Figure 20), but no activity against substrates tested could be found including those composed of α-glucose: maltose, dextran, starch, and laminarin (Figure 1 and Supplementary Figure 16 & 19). A model of Amuc_1870^{GH31} was superimposed on to crystal structures of GH31 enzyme-substrate complexes with α-linked glucose disaccharides: isomaltose, nigerose, and kojibiose (Supplementary Figure 19). Only the isomaltose overlay indicated that this substrate could be accommodated, so we then tried isomaltose and pullulan (maltotriose linked via α1,6-linkages), but also found no activity (Supplementary Figure 16 & 19). Interestingly, however, the closest characterised homologues did prefer isomaltose over maltose (refs). Finally, we also tested this enzyme against glucosylceramide (core of gangliosides and globosides) and trehalose, which is an unusual disaccharide produced by some organisms and is two glucose monosaccharides linked via α1,1 bond. There was no activity against either of these substrates and for the glucosylceramide this was confirmed by HPAEC (Supplementary Figure 19). *A. muciniphila* ATCC BAA-875 did not grow on either maltose, starch, amylose, amylopectin, or pullulan (Supplementary Figure 12). This enzyme is one of the unsolved puzzles (Supplementary Table 1).

904 GH57: Amuc_1868^{GH57} is one of the unsolved puzzles (Supplementary Table 1). We found
905 activity on pNP-α-Gal (Supplementary Figure 20), but none against the wide variety of
906 carbohydrate substrates tested. Some galactose was released when tested against PGM III,
907 pre-treated with sialidase and fucosidase, but the context of this activity remains unresolved
908 (Figure 1). Interestingly, the protein sequence for Amuc_1868^{GH57} is one of the most highly
909 conserved across all the *Akkermansia* species considered here (>90 % identity over five
910 species) and is always present in different strains. Adding further intrigue is that the TnSeq
911 data previously reported by Davey *et al.* highlighted this enzyme as important for growth on
912 mucin (Supplementary Table 1)[20].
913
914 GH63: Amuc_1260^{GH63} is also an unsolved puzzle (Supplementary Table 1) with activity not
915 being identified on any substrate, including pNP-sugars.
916
917 GH105: Finally, Amuc_0863^{GH105} was included in the screening process to be thorough, but
918 no activities were observed apart from a relatively small release of GlcNAc from PGM III
919 (Figure 1).
920
921 **Structural comparisons and additional discussion points**
922 We have used available structures and models of the CAZymes from *A. muciniphila* ATCC
923 BAA-875 and compared them to relevant structures in the literature where possible. A
924 reasonable amount of caution was applied when trying to reach any conclusions based on
925 models - sometimes as we did not have good structural data to use in comparisons, in some
926 cases it is likely the active site is not as it would be with substrate bound, and sometimes the
927 contribution of accessory modules is particularly difficult to assess. We have included the ones
928 that we felt contributed to the report.
929
930 **GH27**
931 The structural and mutational studies of two GH27 enzymes from humans has previously been
932 insightful in terms of their differing specificities for α-galactose or α-GalNAc[78]. In humans, two
933 GH27 enzymes, one α-galactosidase and one α-GalNAc'ase (48 % sequence identity), are
934 involved in the recycling of glycolipids and glycoproteins. Deficiencies in these enzymes in
935 humans is linked to several lysosomal storage diseases due to the accumulation of
936 glycoconjugates (ref). In contrast to Amuc_1187^{GH27}, the human enzymes can accommodate
937 the fucose in the blood group structures to hydrolyse either α-Gal or α-GalNAc[79]. Between
938 these two human enzymes, eleven out of thirteen amino acids coordinating the sugar in the -
939 1 subsite were identical. Those that differed were key to the accommodation of either an α-
940 Gal or α-GalNAc. In the α-GalNAc'ase, there were relatively small residues (serine and
941 alanine) coordinating at the N-acetyl group of the GalNAc, but in the α-galactosidase these
942 residues were larger (glutamate and leucine) thus prohibiting the binding of GalNAc
943 (Supplementary Figure 17). Mutation of the larger residues to smaller ones and vice versa
944 allowed swapping in the specificities[78].
945 Amuc_1187^{GH27} has approximately 30 % sequence identity to the human enzymes, but
946 comparison of an Amuc_1187^{GH27} model with these structures reveals that the overall folds of
947 the enzymes are very similar. Amuc_1187^{GH27} does have an extra small β-sheet N-terminal
948 domain and one of the loops is much larger, but generally very similar at this level. In terms of
949 the active site, the residues coordinating the -1 sugar in the two human enzymes are generally
950 conserved in Amuc_1187^{GH27} and is has relatively large residues at the critical locations (S320
951 and N321), aligning with the observed activity of α-galactosidase (Supplementary Figure 17).
952
953 **GH109**
954 Two structures are available for the GH109 family – one of Amuc_0920^{GH109} and one from
955 *Elizabethkingia meningoseptica*, both have the NAD+ cofactor bound and a GalNAc. We
956 discuss these structures with the Alphafold2 models of Amuc_0017^{GH109} and the other
957 characterised family members of the GH109 family (Supplementary Figure 24). There has
958 been not BGA structure crystallised with a GH109 family member, so we used the BGA

959 tetrasaccharide from a GH98 structure 4D6H and overlaid the -1 GalNAc with the GalNAcs
960 from the two GH109 enzymes. This was to provide better understanding about the areas that
961 the +2 GlcNAc and +1' fucose may occupy. For all the enzymes active against BGA there was
962 a clear area for the fucose to be accommodated. The Amuc_0017^{GH109} model indicated that
963 the fucose of the blood group structure would not be accommodated, whereas in the
964 Amuc_0920^{GH109} structure it is not obvious why this enzyme cannot act on BGA. It may be that
965 this enzyme has conformational changes upon ligand binding, but also likely highlights the
966 limitations of this type of analysis. The ligand is also likely not to be in the exact conformation
967 it would be, which is exemplified by the GlcNAc clashing with the structure in most cases.
968
969 **β-galactosidases**
970 One of the striking aspects of CAZymes from *A. muciniphila* BAA-835 is the length of the
971 putative enzymes, which structural modelling predicts to be a high number of accessory
972 modules accompanying the catalytic module. This is exemplified by the β-galactosidases
973 (Supplementary Figure 29). How these accessory modules contribute to function is still yet to
974 be determined. The active sites of the different GH2 and GH35 enzyme models were
975 compared to relevant existing structures to provide insight into the different specificities being
976 observed.
977
978 Amuc_0539^{GH2} has specificity to β1,4-linked Gal with either a Gal or GalNAc in the +1 position,
979 not Glc/GlcNAc. Galactose and glucose type sugars have the hydroxyl in the axial and
980 equatorial positions at C4, respectively. The difference means that β1,4-linked sugars have
981 very different overall conformations, and this is reflected in some of the specificities observed
982 for the β-galactosidases. Models of different disaccharides is included to put these possible
983 conformational differences into context (Supplementary Figure 30). Current structures do not
984 include any examples with substrates with Gal/GalNAc in the +1 subsite, but LacNAc and
985 lactose examples are available. When we overlay LacNAc from *Streptococcus pneumoniae*
986 GH2 (4CUC) into the active site of Amuc_0539^{GH2} a tryptophan is highlighted as being very
987 close to the +1 GlcNAc. This suggests that this amino acid may contribute significantly to
988 excluding substrates with Glc-type sugars in the +1 subsite. Similarly, Amuc_0290^{GH2}, which
989 has a much broader activity, has W686 in the same region, but rotated approximately 180 °
990 along its plane. This may act to contribute to binding glycans rather than having a selectivity
991 role. Apart from these observations the active sites of these two enzymes are predicted to be
992 generally quite open.
993
994 In contrast to the GH2 and GH35 enzymes, the GH43 proteins only constitute a catalytic
995 module and they are very similar in sequence. The only difference is a short section of
996 sequence that corresponds to an extended loop in Amuc_0697^{GH43} that protrudes towards the
997 active site (Supplementary Figure 27).

1 Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* **54**, 1469-1476, doi:10.1099/ijs.0.02873-0 (2004).

2 Everard, A. *et al.* Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A* **110**, 9066-9071, doi:10.1073/pnas.1219451110 (2013).

3 Pittayanon, R. *et al.* Differences in Gut Microbiota in Patients With vs Without Inflammatory Bowel Diseases: A Systematic Review. *Gastroenterology* **158**, 930-946 e931, doi:10.1053/j.gastro.2019.11.294 (2020).

4 Png, C. W. *et al.* Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol* **105**, 2420-2428, doi:10.1038/ajg.2010.281 (2010).

5 Rajilic-Stojanovic, M., Shanahan, F., Guarner, F. & de Vos, W. M. Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm Bowel Dis* **19**, 481-488, doi:10.1097/MIB.0b013e31827fec6d (2013).

6 Belzer, C. & de Vos, W. M. Microbes inside-from diversity to function: the case of Akkermansia. *Isme J* **6**, 1449-1458, doi:10.1038/ismej.2012.6 (2012).

7 Collado, M. C., Derrien, M., Isolauri, E., de Vos, W. M. & Salminen, S. Intestinal integrity and Akkermansia muciniphila, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl Environ Microbiol* **73**, 7767-7770, doi:10.1128/AEM.01477-07 (2007).

8 Derrien, M., Collado, M. C., Ben-Amor, K., Salminen, S. & de Vos, W. M. The Mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. *Appl Environ Microbiol* **74**, 1646-1648, doi:10.1128/AEM.01226-07 (2008).

9 Hansson, G. C. Mucins and the Microbiome. *Annu Rev Biochem* **89**, 769-793, doi:10.1146/annurev-biochem-011520-105053 (2020).

10 Holmen Larsson, J. M., Thomsson, K. A., Rodriguez-Pineiro, A. M., Karlsson, H. & Hansson, G. C. Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *Am J Physiol Gastrointest Liver Physiol* **305**, G357-363, doi:10.1152/ajpgi.00048.2013 (2013).

11 Robbe, C., Capon, C., Coddeville, B. & Michalski, J. C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem J* **384**, 307-316, doi:10.1042/BJ20040605 (2004).

12 Anso, I. *et al.* Turning universal O into rare Bombay type blood. *Nat Commun* **14**, 1765, doi:10.1038/s41467-023-37324-z (2023).

13 Shuoker, B. *et al.* Sialidases and fucosidases of Akkermansia muciniphila are crucial for growth on mucin and nutrient sharing with mucus-associated gut bacteria. *Nat Commun* **14**, 1833, doi:10.1038/s41467-023-37533-6 (2023).

14 Crouch, L. I. *et al.* Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. *Nat Commun* **11**, 4017, doi:10.1038/s41467-020-17847-5 (2020).

15 Huang, K. *et al.* Biochemical characterisation of the neuraminidase pool of the human gut symbiont Akkermansia muciniphila. *Carbohydr Res* **415**, 60-65, doi:10.1016/j.carres.2015.08.001 (2015).

16 Jajosky, R. P. *et al.* ABO blood group antigens and differential glycan expression: Perspective on the evolution of common human enzyme deficiencies. *iScience* **26**, 105798, doi:10.1016/j.isci.2022.105798 (2023).

17 Rueda, R., Maldonado, J., Narbona, E. & Gil, A. Neonatal dietary gangliosides. *Early Hum Dev* **53 Suppl**, S135-147, doi:10.1016/s0378-3782(98)00071-1 (1998).

18 Stenfelt, L. *et al.* Missense mutations in the C-terminal portion of the B4GALNT2-encoded glycosyltransferase underlying the Sd(a-) phenotype. *Biochem Biophys Rep* **19**, 100659, doi:10.1016/j.bbrep.2019.100659 (2019).

19 Groux-Degroote, S., Vicogne, D., Cogez, V., Schulz, C. & Harduin-Lepers, A. B4GALNT2 Controls Sd(a) and SLe(x) Antigen Biosynthesis in Healthy and Cancer Human Colon. *Chembiochem* **22**, 3381-3390, doi:10.1002/cbic.202100363 (2021).

20 Davey, L. E. *et al.* A genetic system for Akkermansia muciniphila reveals a role for mucin foraging in gut colonization and host sterol biosynthesis gene expression. *Nat Microbiol* **8**, 1450-1467, doi:10.1038/s41564-023-01407-w (2023).

21 Padilla, L. *et al.* Mechanism of 2'-fucosyllactose degradation by human-associated Akkermansia. *J Bacteriol* **206**, e0033423, doi:10.1128/jb.00334-23 (2024).

22 Perna, V. N., Barrett, K., Meyer, A. S. & Zeuner, B. Substrate specificity and transglycosylation capacity of alpha-L-fucosidases across GH29 assessed by bioinformatics-assisted selection of functional diversity. *Glycobiology* **33**, 396-410, doi:10.1093/glycob/cwad029 (2023).

23 Kostopoulos, I. *et al.* Akkermansia muciniphila uses human milk oligosaccharides to thrive in the early life conditions in vitro. *Sci Rep* **10**, 14330, doi:10.1038/s41598-020-71113-8 (2020).

24 Liu, Q. P. *et al.* Identification of a GH110 subfamily of alpha 1,3-galactosidases: novel enzymes for removal of the alpha 3Gal xenotransplantation antigen. *J Biol Chem* **283**, 8545-8554, doi:10.1074/jbc.M709020200 (2008).

25 Wakinaka, T. *et al.* Bifidobacterial alpha-galactosidase with unique carbohydrate-binding module specifically acts on blood group B antigen. *Glycobiology* **23**, 232-240, doi:10.1093/glycob/cws142 (2013).

26 McGuire, B. E. *et al.* The structure of a family 110 glycoside hydrolase provides insight into the hydrolysis of alpha-1,3-galactosidic linkages in lambda-carrageenan and blood group antigens. *J Biol Chem* **295**, 18426-18435, doi:10.1074/jbc.RA120.015776 (2020).

27 Galili, U., Clark, M. R., Shohet, S. B., Buehler, J. & Macher, B. A. Evolutionary relationship between the natural anti-Gal antibody and the Gal alpha 1----3Gal epitope in primates. *Proc Natl Acad Sci U S A* **84**, 1369-1373, doi:10.1073/pnas.84.5.1369 (1987).

28 Hellberg, A., Westman, J. S., Thuresson, B. & Olsson, M. L. P1PK: the blood group system that changed its name and expanded. *Immunohematology* **29**, 25-33 (2013).

29 Luis, A. S. *et al.* Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat Microbiol* **3**, 210-219, doi:10.1038/s41564-017-0079-1 (2018).

30 Jindou, S. *et al.* alpha-Galactosidase Aga27A, an enzymatic component of the Clostridium josui cellulosome. *J Bacteriol* **184**, 600-604, doi:10.1128/JB.184.2.600-604.2002 (2002).

31 Okuyama, M. *et al.* Efficient synthesis of alpha-galactosyl oligosaccharides using a mutant Bacteroides thetaiotaomicron retaining alpha-galactosidase (BtGH97b). *FEBS J* **284**, 766-783, doi:10.1111/febs.14018 (2017).

32 Hult, A. K. & Olsson, M. L. The FORS awakens: review of a blood group system reborn. *Immunohematology* **33**, 64-72 (2017).

33 Yamamoto, M., Tarasco, M. C., Cid, E., Kobayashi, H. & Yamamoto, F. ABO blood group A transferase and its codon 69 substitution enzymes synthesize FORS1 antigen of FORS blood group system. *Sci Rep* **9**, 9717, doi:10.1038/s41598-019-46029-7 (2019).

34 Goel, R. *et al.* ABO blood group and COVID-19: a review on behalf of the ISBT COVID-19 Working Group. *Vox Sang* **116**, 849-861, doi:10.1111/vox.13076 (2021).

35 Rahfeld, P. *et al.* An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat Microbiol* **4**, 1475-1485, doi:10.1038/s41564-019-0469-7 (2019).

36 Calcutt, M. J., Hsieh, H. Y., Chapman, L. F. & Smith, D. S. Identification, molecular cloning and expression of an alpha-N-acetylgalactosaminidase gene from Clostridium perfringens. *FEMS Microbiol Lett* **214**, 77-80, doi:10.1111/j.1574-6968.2002.tb11327.x (2002).

1099 37 Lu, Y., Zhao, H., Wang, Q. P., Liu, W. F. & Xiao, M. [High expression of alpha-D-galactosidase
1100 gene (aga1) of Bifidobacterium breve 203 in Escherichia coli]. *Wei Sheng Wu Xue Bao* **45**, 241-
1101 246 (2005).
1102 38 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The
1103 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495,
1104 doi:10.1093/nar/gkt1178 (2014).
1105 39 Fujita, M. *et al.* Glycoside hydrolase family 89 alpha-N-acetylglucosaminidase from
1106 Clostridium perfringens specifically acts on GlcNAc alpha1,4Gal beta1R at the non-reducing
1107 terminus of O-glycans in gastric mucin. *J Biol Chem* **286**, 6479-6489,
1108 doi:10.1074/jbc.M110.206722 (2011).
1109 40 Ficko-Blean, E. & Boraston, A. B. Structural analysis of a bacterial exo-alpha-D-N-
1110 acetylglucosaminidase in complex with an unusual disaccharide found in class III mucin.
1111 *Glycobiology* **22**, 590-595, doi:10.1093/glycob/cwr165 (2012).
1112 41 Birrane, G. *et al.* Structural characterization of the alpha-N-acetylglucosaminidase, a key
1113 enzyme in the pathogenesis of Sanfilippo syndrome B. *J Struct Biol* **205**, 65-71,
1114 doi:10.1016/j.jsb.2019.02.005 (2019).
1115 42 Kosciow, K. & Deppenmeier, U. Characterization of three novel beta-galactosidases from
1116 Akkermansia muciniphila involved in mucin degradation. *Int J Biol Macromol* **149**, 331-340,
1117 doi:10.1016/j.ijbiomac.2020.01.246 (2020).
1118 43 Guo, B. S. *et al.* Cloning, purification and biochemical characterisation of a GH35 beta-
1119 1,3/beta-1,6-galactosidase from the mucin-degrading gut bacterium Akkermansia
1120 muciniphila. *Glycoconj J* **35**, 255-263, doi:10.1007/s10719-018-9824-9 (2018).
1121 44 Kosciow, K. & Deppenmeier, U. Characterization of a phospholipid-regulated beta-
1122 galactosidase from Akkermansia muciniphila involved in mucin degradation.
1123 *Microbiologyopen* **8**, e00796, doi:10.1002/mbo3.796 (2019).
1124 45 Qian, K. *et al.* Functional and structural characterization of a GH3 beta-N-
1125 acetylhexosaminidase from Akkermansia muciniphila involved in mucin degradation. *Biochem*
1126 *Biophys Res Commun* **589**, 186-191, doi:10.1016/j.bbrc.2021.12.022 (2022).
1127 46 Wang, M. *et al.* Cloning, purification and biochemical characterization of two beta-N-
1128 acetylhexosaminidases from the mucin-degrading gut bacterium Akkermansia muciniphila.
1129 *Carbohydr Res* **457**, 1-7, doi:10.1016/j.carres.2017.12.007 (2018).
1130 47 Chen, X., Li, M., Wang, Y., Tang, R. & Zhang, M. Biochemical characteristics and
1131 crystallographic evidence for substrate-assisted catalysis of a beta-N-acetylhexosaminidase in
1132 Akkermansia muciniphila. *Biochem Biophys Res Commun* **517**, 29-35,
1133 doi:10.1016/j.bbrc.2019.06.150 (2019).
1134 48 Xu, W., Yang, W., Wang, Y., Wang, M. & Zhang, M. Structural and biochemical analyses of
1135 beta-N-acetylhexosaminidase Am0868 from Akkermansia muciniphila involved in mucin
1136 degradation. *Biochem Biophys Res Commun* **529**, 876-881, doi:10.1016/j.bbrc.2020.06.116
1137 (2020).
1138 49 Chen, X. *et al.* Crystallographic evidence for substrate-assisted catalysis of beta-N-
1139 acetylhexosaminidas from Akkermansia muciniphila. *Biochem Biophys Res Commun* **511**, 833-
1140 839, doi:10.1016/j.bbrc.2019.02.074 (2019).
1141 50 Rahfeld, P. *et al.* Prospecting for microbial alpha-N-acetylgalactosaminidases yields a new
1142 class of GH31 O-glycanase. *J Biol Chem* **294**, 16400-16415, doi:10.1074/jbc.RA119.010628
1143 (2019).
1144 51 Miyazaki, T., Ikegaya, M. & Alonso-Gil, S. Structural and mechanistic insights into the substrate
1145 specificity and hydrolysis of GH31 alpha-N-acetylgalactosaminidase. *Biochimie* **195**, 90-99,
1146 doi:10.1016/j.biochi.2021.11.007 (2022).
1147 52 Ikegaya, M., Miyazaki, T. & Park, E. Y. Biochemical characterization of Bombyx mori alpha-N-
1148 acetylgalactosaminidase belonging to the glycoside hydrolase family 31. *Insect Mol Biol* **30**,
1149 367-378, doi:10.1111/imb.12701 (2021).

1150 53 Pilgaard, B. *et al.* Discovery of a Novel Glucuronan Lyase System in Trichoderma parareesei.
1151 *Appl Environ Microbiol* **88**, e0181921, doi:10.1128/AEM.01819-21 (2022).
1152 54 Ronne, M. E. *et al.* Three alginate lyases provide a new gut Bacteroides ovatus isolate with the
1153 ability to grow on alginate. *Appl Environ Microbiol* **89**, e0118523, doi:10.1128/aem.01185-23
1154 (2023).
1155 55 Itoh, T., Ochiai, A., Mikami, B., Hashimoto, W. & Murata, K. A novel glycoside hydrolase family
1156 105: the structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with
1157 a disaccharide in comparison with family 88 enzyme complexed with the disaccharide. *J Mol*
1158 *Biol* **360**, 573-585, doi:10.1016/j.jmb.2006.04.047 (2006).
1159 56 Shin, J. *et al.* Elucidation of Akkermansia muciniphila Probiotic Traits Driven by Mucin
1160 Depletion. *Front Microbiol* **10**, 1137, doi:10.3389/fmicb.2019.01137 (2019).
1161 57 Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier
1162 and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353.e1321,
1163 doi:10.1016/j.cell.2016.10.043 (2016).
1164 58 Ottman, N. *et al.* Genome-Scale Model and Omics Analysis of Metabolic Capacities of
1165 Akkermansia muciniphila Reveal a Preferential Mucin-Degrading Lifestyle. *Appl Environ*
1166 *Microbiol* **83**, doi:10.1128/aem.01014-17 (2017).
1167 59 Stam, M. *et al.* SulfAtlas, the sulfatase database: state of the art and new developments.
1168 *Nucleic Acids Res* **51**, D647-D653, doi:10.1093/nar/gkac977 (2023).
1169 60 Luis, A. S. *et al.* A single sulfatase is required to access colonic mucin by a gut bacterium.
1170 *Nature* **598**, 332-337, doi:10.1038/s41586-021-03967-5 (2021).
1171 61 Ulmer, J. E. *et al.* Characterization of glycosaminoglycan (GAG) sulfatases from the human gut
1172 symbiont Bacteroides thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase.
1173 *J Biol Chem* **289**, 24289-24303, doi:10.1074/jbc.M114.573303 (2014).
1174 62 Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish
1175 mechanism. *Nature* **517**, 165-169, doi:10.1038/nature13995 (2015).
1176 63 Zhang, Z., Xie, J., Zhang, F. & Linhardt, R. J. Thin-layer chromatography for the analysis of
1177 glycosaminoglycan oligosaccharides. *Anal Biochem* **371**, 118-120,
1178 doi:10.1016/j.ab.2007.07.003 (2007).
1179 64 Brooke, D., Movahed, N. & Bothner, B. Universal buffers for use in biochemistry and
1180 biophysical experiments. *AIMS Biophys* **2**, 336-342, doi:10.3934/biophy.2015.3.336 (2015).
1181 65 Pilgaard, B., Vuillemin, M., Holck, J., Wilkens, C. & Meyer, A. S. Specificities and Synergistic
1182 Actions of Novel PL8 and PL7 Alginate Lyases from the Marine Fungus Paradendryphiella
1183 salina. *J Fungi (Basel)* **7**, doi:10.3390/jof7020080 (2021).
1184 66 Domon, B. & Costello, C. E. A Systematic Nomenclature for Carbohydrate Fragmentations in
1185 Fab-Ms Ms Spectra of Glycoconjugates. *Glycoconjugate J* **5**, 397-409, doi:Doi
1186 10.1007/Bf01049915 (1988).
1187 67 Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language
1188 models. *Nat Biotechnol* **40**, 1023-1025, doi:10.1038/s41587-021-01156-3 (2022).
1189 68 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments
1190 using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).
1191 69 Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user
1192 interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**, 221-224,
1193 doi:10.1093/molbev/msp259 (2010).
1194 70 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
1195 display and annotation. *Nucleic Acids Res* **49**, W293-W296, doi:10.1093/nar/gkab301 (2021).
1196 71 Grant, B. J., Skjaerven, L. & Yao, X. Q. The Bio3D packages for structural bioinformatics. *Protein*
1197 *Sci* **30**, 20-30, doi:10.1002/pro.3923 (2021).
1198 72 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-
1199 589, doi:10.1038/s41586-021-03819-2 (2021).

73 Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444, doi:10.1093/nar/gkab1061 (2022).

74 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682, doi:10.1038/s41592-022-01488-1 (2022).

75 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/s0907444910007493 (2010).

76 The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. .

77 Ndongo, S., Armstrong, N., Raoult, D. & Fournier, P. E. Reclassification of eight Akkermansia muciniphila strains and description of Akkermansia massiliensis sp. nov. and Candidatus Akkermansia timonensis, isolated from human feces. *Sci Rep* **12**, 21747, doi:10.1038/s41598-022-25873-0 (2022).

78 Tomasic, I. B., Metcalf, M. C., Guce, A. I., Clark, N. E. & Garman, S. C. Interconversion of the specificities of human lysosomal enzymes associated with Fabry and Schindler diseases. *J Biol Chem* **285**, 21560-21566, doi:10.1074/jbc.M110.118588 (2010).

79 Olsson, M. L. & Clausen, H. Modifying the red cell surface: towards an ABO-universal blood supply. *Br J Haematol* **140**, 3-12, doi:10.1111/j.1365-2141.2007.06839.x (2008).

Supplementary Table 1. Summary of enzymes from *A. muciniphila* BAA-835 included in this paper. This is a summary of the enzymes explored in this report, how they were divided into different groups, their overall result, and where they are predicted to be localised in the cell according to Signal P 6.0 (grey shading highlights those possibly on the outside of the cell). The enzymes highlighted in pink are the ones that were highlighted by proteomics when *A. muciniphila* BAA-835 was grown on human breast milk (Kostopoulos *et al.* 2020).

	Enzyme	Headline	Predicted localisation	Importance judged by TnSeq (LogFC -1.5 and lower)	Interaction with Mul1A	Highly conserved across different *Akkermansia* species	Cloning
Sialidases	Amuc_0625^{GH33}	Broad-acting	90 % SPI, 7 % SPII			✓	pET28b Cl Kan
	Amuc_1547^{GH181}	α1,3 only	51 % SPII, 48 % SPI			✓	pET28b Cl Kan
	Amuc_1835^{GH33}	Broad-acting	99 % SPI	✓			pET28b Cl Kan
Fucosidases	Amuc_0010^{GH29}	Specific for BGH type II structures	99 % SPI				pET28b GS Kan
	Amuc_0146 ^{GH29}	★Puzzle★	71 % SPI, 27 % other	✓			pET28b Cl Kan
	Amuc_0186 ^{GH95}	2FL only	99 % SPI				pET21b Cl Amp
	Amuc_0392 ^{GH29}	Very broad for α1,3/4	99 % SPI			✓	pET28b Cl Kan
	Amuc_0846 ^{GH29}	Very broad for α1,3/4 (apart from SFLNT)	94 % SPI, 6 % SPII				pET28b Cl Kan
	Amuc_1120 ^{GH95}	Broad α 1,2 – only one that can handle BGH I	33 % SPI, 65 % other	✓	✓		pET28b Cl Kan
α-linked exo activities	Amuc_1187^{GH27}	Broad α-galactosidase	99 % SPI				pET28b Cl Kan
	Amuc_1008^{GH31}	Specificity for the core α-GalNAc	99 % SPI				pET21b Cl Amp
	Amuc_1870^{GH31}	★Puzzle★ pNP-α-Glc	Other				pET21b Cl Amp
	Amuc_0216^{GH36}	Broad α-GalNAc'ase. Only enzyme to degrade Tn and Forssman antigens.	74 % SPII, 26 % SPI				pET28b Cl Kan
	Amuc_0517^{GH36}	Specificity for BGA	99 % SPI				pET28b Cl Kan
	Amuc_0855^{GH36}	Specificity for Galα1,3Gal	Other				pET21b Cl Amp
	Amuc_1868^{GH57}	★Puzzle★ pNP-α-Gal	Other	✓		✓	pET28b Kan
	Amuc_1260^{GH63}	★Puzzle★	99 % SPI				pET28b Cl Kan
	Amuc_0060^{GH89}	α-GlcNAc'ase but true substrate unclear	99 % SPI	✓			pET28b Kan
	Amuc_1220^{GH89}	GlcNAcα1,4-Gal disaccharide	63 % SPI, 37 % SPII	✓	✓		pET28b Cl Kan
	Amuc_1420^{GH97}	α-galactosidase, likely requires +2 subsite	60 % SPI, 39 % SPII				pET28b Cl Kan
	Amuc_0017^{GH109}	α-GalNAc'ase (not BGA or Tn)	TAT			✓	pET28b Cl Kan
	Amuc_0920^{GH109}	α-GalNAc'ase (not BGA or Tn)	TAT				pET28b Cl Kan
	Amuc_0480^{GH110}	α-galactosidase (not globotriose or P1 antigen)	23 % SPI, 75 % other				pET28b Cl Kan
	Amuc_1463^{GH110}	Broad α-galactosidase	99 % SPI				pET28b GS Kan
β-galactosidas	Amuc_0290^{GH2}	Broad β-galactosidase	99 % SPI	✓	✓	✓	pET21b Cl Amp
	Amuc_0539^{GH2}	Galβ1,3Gal/GalNAc	99 % SPI	✓	✓	✓	pET28b Cl Kan
	Amuc_0771^{GH35}	Requires a Glc/GlcNAc in the +1 position for	98 % SPI				pET21b Cl Amp

		β1,4-linked. Not β1,6 substrates.					
	Amuc_0824^{GH2}	β1,3-galactosidase	55 % SPII, 44 % SPI		✓	✓	pET21b Cl Amp
	Amuc_1666^{GH2}	β1,4-galactosidase with a Glc sugar in the +1 subsite	99 % SPI	✓		✓	pET28b Kan
	Amuc_1667^{GH2}	True substrate not identified, partial against GA2	99 % SPI	✓	✓		pET28b Kan
	Amuc_1686^{GH35}	β1,3/6-galactosidase	99 % SPI			✓	pET21b Cl Amp
	Amuc_0497^{GH43}	★Puzzle★	99 % SPI	✓			pET28b Kan
	Amuc_0498^{GH43}	★Puzzle★ pNP-β-Gal	94 % SPII, 6 % SPI				pET29b Kan
β-HexNAc' ases	Amuc_0369^{GH20}	Broad β-HexNAc'ase	99 % SPI				pET28b Kan
	Amuc_0397^{GH20}	β1,3-GalNAc'ase, true substrate not identified	99 % SPI	✓			pET28b Kan
	Amuc_0868^{GH20}	Broad β-HexNAc'ase	55 % SPI, 43 % SPII		✓		pET28b Kan
	Amuc_1032^{GH20}	GlcNAcβ1,3Gal specificity, requires +2 sugar	99 % SPI			✓	pET21b Cl Amp
	Amuc_1669^{GH20}	β1,3-HexNAc preference (host substrates only)	90 % SPI, 10 % SPII	✓	✓		pET28b Kan
	Amuc_1815^{GH20}	β1,3-GalNAc'ase, true substrate not identified	61 % SPI, 39 % SPII				pET28b Kan
	Amuc_1924^{GH20}	β-GalNAc'ase, not P antigen	99 % SPI				pET21b Cl Amp
	Amuc_2018^{GH20}	β1,3-HexNAc preference, not P antigen	99 % SPI				pET28b Kan
	Amuc_2019^{GH20}	Broad β-HexNAc'ase	69 % SPI, 29 % SPII				pET21b Cl Amp
	Amuc_2136^{GH20}	Broad β-HexNAc'ase, not GA2	99 % SPI	✓			pET21b Cl Amp
	Amuc_2148^{GH20}	β1,3-GlcNAc'ase preference	99 % SPI	✓			pET28b Kan
	Amuc_0052^{GH84}	Broad β-GlcNAc'ase (host substrates only)	99 % SPI	✓			pET21b Cl Amp
	Amuc_0803^{GH123}	Broad β-GalNAc'ase	99 % SPI		✓	✓	pET21b Amp
	Amuc_2109^{GH3}	β-GlcNAc'ase, true substrate not identified	Other				pET28b Kan
GAGs	Amuc_0778^{PL38}	HA, CSA, CSC, and DS	99 % SPI				pET28b Kan
	Amuc_0863^{GH105}	Products of PL38	93 % SPI				pET28b Kan
Sulfatases	Amuc_0451^{Sulf20}	6S-GlcNAc 3'S-LNB and LacNAc	51 % SPII, 48 % SPI		✓	Not tested	pET28b GS Kan
	Amuc_0491^{Sulf20}	3'S-LNB (LacNac partial)	99 % SPI	✓	✓	Not tested	pET29b GS Kan
	Amuc_0565^{Sulf4}		99 % SPI			Not tested	pET28b GS Kan
	Amuc_0864^{Sulf26}		Other			Not tested	pET28b GS Kan
	Amuc_0953^{Sulf20}	3'S-LNB and LacNAc	99 % SPI			Not tested	pET28b GS Kan
	Amuc_1033^{Sulf11}	6S-GlcNAc	87 % SPI, 12 % SPII			Not tested	pET29b GS Kan
	Amuc_1182^{Sulf19}		99 % SPI	✓		Not tested	pET28b GS Kan
	Amuc_1290^{Sulf}		Other	✓		Not tested	pET28b GS Kan
	Amuc_1655^{Sulf16}	4S-Gal/GalNAc	99 % SPI			Not tested	pET29b GS Kan
	Amuc_1755^{Sulf16}	4S-Gal/GalNAc	70 % SPII, 30 % SPI			Not tested	pET29b GS Kan
	Amuc_1074^{Sulf11}		99 % SPI			Not tested	pET29b GS Kan

	Amuc_0121^{Sulf15}	Partial 6S-GlcNAc	99 % SPI			Not tested	pET29b GS Kan
?	Amuc_1216^{GH177}	★Puzzle★	TAT			✓	pET28b Cl Kan
	Amuc_0623^{GHnc}	★Puzzle★	99 % SPI				pET28b Cl Kan
Starch	Amuc_1621^{GH77}	Ladder of products from amylose	Other				pET29b GS Kan
	Amuc_1637^{GH13}	★Puzzle★	Other				pET29b GS Kan
	Amuc_1751^{GH13}	★Puzzle★	Other	✓			pET29b GS Kan
	Amuc_1812^{GH13}	★Puzzle★	Other	✓			pET28b Cl Kan

GS – gene synthesis. Cl – cloned.

Supplementary Table 2. Substrates. Details about the substrates used in this work and the concentration they were used at.

Substrate	Source	Further info	Concentration used at
2’ fucosyllactose	Biosynth	OF067391402	1 mM
2'-Fucosyllactose synthetic	Biosynth	OF067391501	1 mM
3 Fucosyllactose	Biosynth	OF056731501	1 mM
3' Sulfated lewis A	Biosynth	OS05428	1 mM
3' Sulfated Lewis X	Biosynth	OS04132	1 mM
3'-O-sulfated Lewis A	Biosynth	OS054281601	1 mM
3'-O-sulfated Lewis X	Biosynth	OS041321801	1 mM
3'-Sialyl Lewis A	Biosynth	OS007451301	1 mM
3'-Sialyl Lewis X	Biosynth	OS040581801	1 mM
3’-sialylactose	Biosynth	OS043971701	1 mM
4-O-sulfated GalNAc	Dextra	G1054	1 mM
6-O-sulfated GalNAc	Dextra	G1052	1 mM
6-O-sulfated GlcNAc	Dextra	G1004	1 mM
6'-O-sulfated Lewis A	Biosynth	OS58478	1 mM
6-O-sulfated Lewis A	Biosynth	OS58480	1 mM
6'-O-sulfated Lewis X	Biosynth	OS58479	1 mM
6-O-sulfated Lewis X	Biosynth	OS58481	1 mM
6SL	Biosynth	OS043981301	1 mM
Agarose	Sigma	A-941450g	10 mg/mL
Amylopectin	Sigma	10120-250g	10 mg/mL
Amylose	Sigma	A0512-250mg	10 mg/mL
Blood Group A Type II	Biosynth	OA11696	1 mM
Blood Group B Type I	Biosynth	OB159055	1 mM
Blood Group B Type I hexasaccharide	Biosynth	OB159055	1 mM
Blood Group B Type II	Biosynth	OB09577	1 mM
Blood Group B Type II hexasaccharide	Biosynth	OB09577	1 mM
Blood Group H Disaccharide	Biosynth	OB059071501	1 mM
Blood Group H Type I	Biosynth	OB64806	1 mM
Blood Group H Type II	Biosynth	OB083581401	1 mM
Bovine Ribonuclease B	Sigma	R7884-100mg	10 mg/ml
Bovine submaxillary mucin	Sigma	M3895-500mg	10 mg/mL
Cellulase from Trichoderma viride	VWR	390744E	10 mg/mL
Chitobiose	Biosynth	OD007691201	1 mM
Chitohexose	Biosynth	OH07433	1 mM
Chitopentaose	Biosynth	OP04834	1 mM
Chitotetraose	Biosynth	OT04211	1 mM
Chitotriose	Biosynth	OT06497	1 mM
Chondroitin sulphate	Biosynth	YC314581701	20 mg/mL
Colominic acid			25 mg/ml
Dextran from Leuconostoc mesenteroides	Sigma	D1037-50g CAS 9004-0 (C6H1005)n	10 mg/mL
D-galactose-4-O-sulfate	Dextra	G0012	1 mM
D-galactose-6-O-sulfate	Dextra	G0011	1 mM
Difucosyllacto-N-hexaose	Biosynth	OD06532	1 mM
Forssman antigen	Biosynth	OF456291801	1 mM
Fucan	Danisco	W6-16	10 mg/mL
Fucoidan (Fucus vesiculosus)	Sigma	F5631	10 mg/mL
Galactomannan	Megazyme	10501b	10 mg/mL
GalNAc-α1,3-gal	Biosynth	OA080851601	1 mM
GalNAc-β1,3-Gal	Biosynth	OA080841401	1 mM
Gal-α1,3-Gal	Biosynth	OG064911501	1 mM
Gal-α1,3-GalNAc	Biosynth	OA06694	1 mM
Gal-β1,3-GalNAc	Biosynth	OA01686	1 mM
Gal-β1,3-GalNac-β1,4-Gal-β1,4-Glc	Biosynth	OG448341901	1 mM
Gal-β1,3-Glc	Biosynth	OG101421201	1 mM
Gal-β1,4-Gal	Biosynth	OG150711701	1 mM
Gal-β1,4-GalNAc	Biosynth	OA09634	1 mM
Gal-β1,6-Gal			1 mM
Gal-β1,6-GalNAc	Biosynth	OA04443	1 mM
GD1a	Biosynth	OO35336	1 mM
GlcNAc-α1,4-gal	Dextra	GN104	1 mM
GlcNAc-β1,2-Man	Biosynth	OA044511001	1 mM

GlcNAc-β1,3-Gal	Elicityl	GLY155-5mg	1 mM
GlcNAc-β1,3-Man	Biosynth	OA067381201	1 mM
Globotriose	Biosynth	OG016261801	1 mM
Glucosyl ceramide	Biosynth	MG16326	10 mM
Glucuronan			1 mM
GT1b	Biosynth	OO16323	1 mM
Heparin	Biosynth	YH09354	10 mg/mL
Hyaluronic acid	Biosynth	YH72470	10 mg/mL
Isomaltose			1 mM
LacNAc	Biosynth	OA0822441801	1 mM
Lacto-N-biose	Biosynth	OL017071301	1 mM
Lacto-N-difucohexaose I	Biosynth	OL01664	1 mM
Lacto-N-difucohexaose II	Biosynth	OL06826	1 mM
Lacto-N-fucopentaose II	Biosynth	OL03876	1 mM
Lacto-N-fucopentaose II	Biosynth	OL03876	1 mM
Lacto-N-hexose	Biosynth	OL02436	1 mM
Lacto-N-neotetraose	Biosynth	OL05765	1 mM
Lacto-N-tetraose	Dextra or Biosynth	CCDX233-132 or OL1626651802	1 mM
Lacto-N-tetraose	Biosynth	OL1626651802	1 mM
Lacto-N-triose	Made from Lacto-N-tetraose		1 mM
Lactose	Merck		1 mM
Lactosylceramide	Biosynth	OL16325	1 mM
Laminarin	Sigma	I-9634	1 mM
Lewis A	Biosynth		1 mM
Lewis A Trisaccharide	Biosynth	OL045411102	1 mM
Lewis B	Biosynth	OL02434	1 mM
Lewis B tetrasaccharide	Biosynth	OL024341601	1 mM
Lewis X	Biosynth	OL064901301	1 mM
Lewis Y	Biosynth	OL065211701	1 mM
Lewis Y tetrasaccharide	Biosynth	OL005210901	1 mM
Lichenan	Biosynth	BS-3263c-1g	10 mg/mL
Maltose	Sigma		1 mM
Man-α1,2-man	Biosynth		1 mM
Man-α1,3-man	Biosynth		1 mM
Man-α1,4-man	Biosynth		1 mM
Man-α1,6-man	Biosynth		1 mM
Mannan fom Saccharomyces cerevisiae	Sigma	M7504-5g	10 mg/mL
Melibiose	Sigma	O47H1430	1 mM
Mucin from porcine stomach, type III	Sigma	M1778-100g	10 mg/mL
N-acetyl-D-glucosamine-3,6-di-O-sulfate	Dextra	G1033	1 mM
Neutral HMOs	Biosynth	OH165969	6 mM
P1 antigen			1 mM
paralacto-N-hexaose	Biosynth	OL068391403	1 mM
Paralacto-N-neohexose	Biosynth	OL068391403	1 mM
paralacto-N-neohexose			1 mM
Porphyran	Extracted from Porphyra sp.		10 mg/mL
Pullulan	Megazyme	90801a	10 mg/mL
Raffinose	Sigma	RO250-25g	1 mM
Sialylated HMOs	Biosynth	OH165970	2 mM
Sialylfucosyllacto-N-tetraose	Biosynth	OS06576	1 mM
Siaylfucosyllacto-N-tetraose	Biosynth	OS06576	1 mM
Sodium alginate food grade	Sigma		10 mg/mL
Stachyose	Sigma	S4001-1g	1 mM
Starch	Sigma	S-9765	10 mg/mL
T antigen	Biosynth	OG075251401	1 mM
TetraLacNAc	Elicityl	GLY007-3	1 mM
Tn antigen	Biosynth	MA073781302	1 mM
Trehalose dihydrate	EMD Millipore	625625-50gm	1 mM
TriLacNAc	Elicityl	GLY007-2	1 mM
Ulvan (Enteromorpha spp.)	Elicityl Oligotech	Ulv101	10 mg/mL
Ulvan (Ulva spp.)	Elicityl Oligotech	Ulv100	10 mg/mL
α1 AGP	Sigma	G3643-100mg	10 mg/ml
ι-carrageenan	Extracted from Eucheuma denticulatum		10 mg/mL
κ-carrageenan	Extracted from Kappaphycus alvarezii		10 mg/mL

λ-carrageenan	Sigma	Rec lot MO68102-002/ Man lot: BCBP8978V/ Exp Date: NOV 21, 2021	10 mg/mL
Polygalacturonic acid	Sigma		10 mg/mL

Supplementary Table 3 Prevalence of GH36 enzymes in *Akkermansia* strains. For this analysis, we took the GH36 sequences Extra 1 and 2 from *A. biwaensis.*

	Amuc_0517	Amuc_0855	Amuc_0216	Extra 1	Extra 2
Amuc_0517	100	17	17	14	17
Amuc_0855		100	19	17	20
Amuc_0216			100	32	33
Extra 1				100	35
Extra 2					100

Scale colour coding	100-91	90-81	80-71	70-61	60-51	50-41	40-31	30-21	20-11	10-0

Supplementary Table 4 Percentage identity between GH2 family members from *A. muciniphila* BAA-835

	Amuc_0824	Amuc_1666	Amuc_0290	Amuc_0539	Amuc_1667
Amuc_0824	100	39	22	22	20
Amuc_1666		100	20	21	24
Amuc_0290			100	26	27
Amuc_0539				100	28
Amuc_1667					100

Scale colour coding	100-91	90-81	80-71	70-61	60-51	50-41	40-31	30-21	20-11	10-0

Supplementary Table 5. Percentage identity between GH20 family members from *A. muciniphila* BAA-835

	Amuc_1924	Amuc_0367	Amuc_1815	Amuc_2136	Amuc_1032	Amuc_2018	Amuc_2019	Amuc_2148	Amuc_0868	Amuc_0369	Amuc_1669
Amuc_1924	100	27	29	21	23	23	23	26	27	26	27
Amuc_0367		100	30	20	25	27	28	30	29	27	24
Amuc_1815			100	20	23	24	23	21	23	23	25
Amuc_2136				100	21	23	23	24	22	23	21
Amuc_1032					100	41	42	26	31	28	26
Amuc_2018						100	72	29	32	30	27
Amuc_2019							100	26	32	30	26
Amuc_2148								100	31	32	30
Amuc_0868									100	37	35
Amuc_0369										100	54
Amuc_1669											100

Scale colour coding	100-91	90-81	80-71	70-61	60-51	50-41	40-31	30-21	20-11	10-0

Supplementary Table 6. Specific activities calculated from the highest rates observed in the kinetic analysis experiments of Amuc_0778^{PL38} and Amuc_0863^{GH105}. One Unit was defined as µmol double bond produced/converted per minute. Initial substrate concentrations (µM double bonds) for Amuc_0863^{GH105} was calculated based on the absorbance (A235) of the samples after terminated overnight reactions of Amuc_0778^{PL38}.

	Amuc_0778^{PL38}			**Amuc_0863^{GH105}**		
Substrates	**Specific activity (U · mg^{-1})**	**Substrate concentration (g · l^{-1})**	**Relative activity (%)**	**Specific activity (U · mg^{-1})**	**Substrate concentration (uM)**	**Relative activity (%)**
Hyaluronic acid	29.373 ± 0.710	12	100	0.029 ± 0.002	1036	42
Chondroitin sulfate A	29.393 ± 0.401	12	100	0.069 ± 0.012	1016	100
Chondroitin sulfate C	4.426 ± 0.080	12	15	0.025 ± 0.002	1172	36
Dermatan sulfate	2.274 ± 0.258	8	8	0.026 ± 0.001	540	38
Heparin	nd.	-	-	nd.	-	-
Polymannuronic acid	nd.	-	-	nd.	-	-
Polyguluronic acid	nd.	-	-	nd.	-	-
Polyglucuronic acid	nd.	-	-	nd.	-	-
Polygalacturonic acid	nd.	-	-	nd.	-	-

Supplementary Table 7. Kinetic parameters for Amuc_0778^{PL38} on four different substrates. Parameters were obtained by fitting the Michaelis-Menten model to the initial rates for increasing substrate concentrations. ± represents standard deviations from triplicate experiments.

	k_{cat} (s^{-1})	K_M ($g \cdot l^{-1}$)	k_{cat} / K_M (s^{-1} $g \cdot l^{-1}$)
Hyaluronic acid	23.13 ± 0.81	2.98 ± 0.27	7.79 ± 0.45
Chondroitin sulfate A	29.24 ± 2.00	6.38 ± 0.93	4.62 ± 0.38
Chondroitin sulfate C	3.06 ± 0.11	3.06 ± 0.41	1.01 ± 0.11
Dermatan sulfate	0.77 ± 0.12	1.51 ± 0.50	0.53 ± 0.10

Supplementary Table 8. Percentage identity between putative sulfatases from *A. muciniphila* BAA-835

	Amuc_1033	Amuc_1074	Amuc_0121	Amuc_0491	Amuc_0565	Amuc_1182	Amuc_1655	Amuc_1755	Amuc_0451	Amuc_0953	Amuc_0864	Amuc_1290
Amuc_1033	100	61	28	26	26	24	26	27	22	22	13	14
Amuc_1074		100	27	24	27	25	25	26	21	20	15	15
Amuc_0121			100	29	28	30	27	25	24	24	12	11
Amuc_0491				100	32	29	26	24	29	29	15	15
Amuc_0565					100	37	27	26	25	26	11	14
Amuc_1182						100	28	26	25	24	14	15
Amuc_1655							100	63	28	30	12	15
Amuc_1755								100	27	27	10	15
Amuc_0451									100	58	11	16
Amuc_0953										100	12	19
Amuc_0864											100	20
Amuc_1290												100

Scale colour coding	100-91	90-81	80-71	70-61	60-51	50-41	40-31	30-21	20-11	10-0

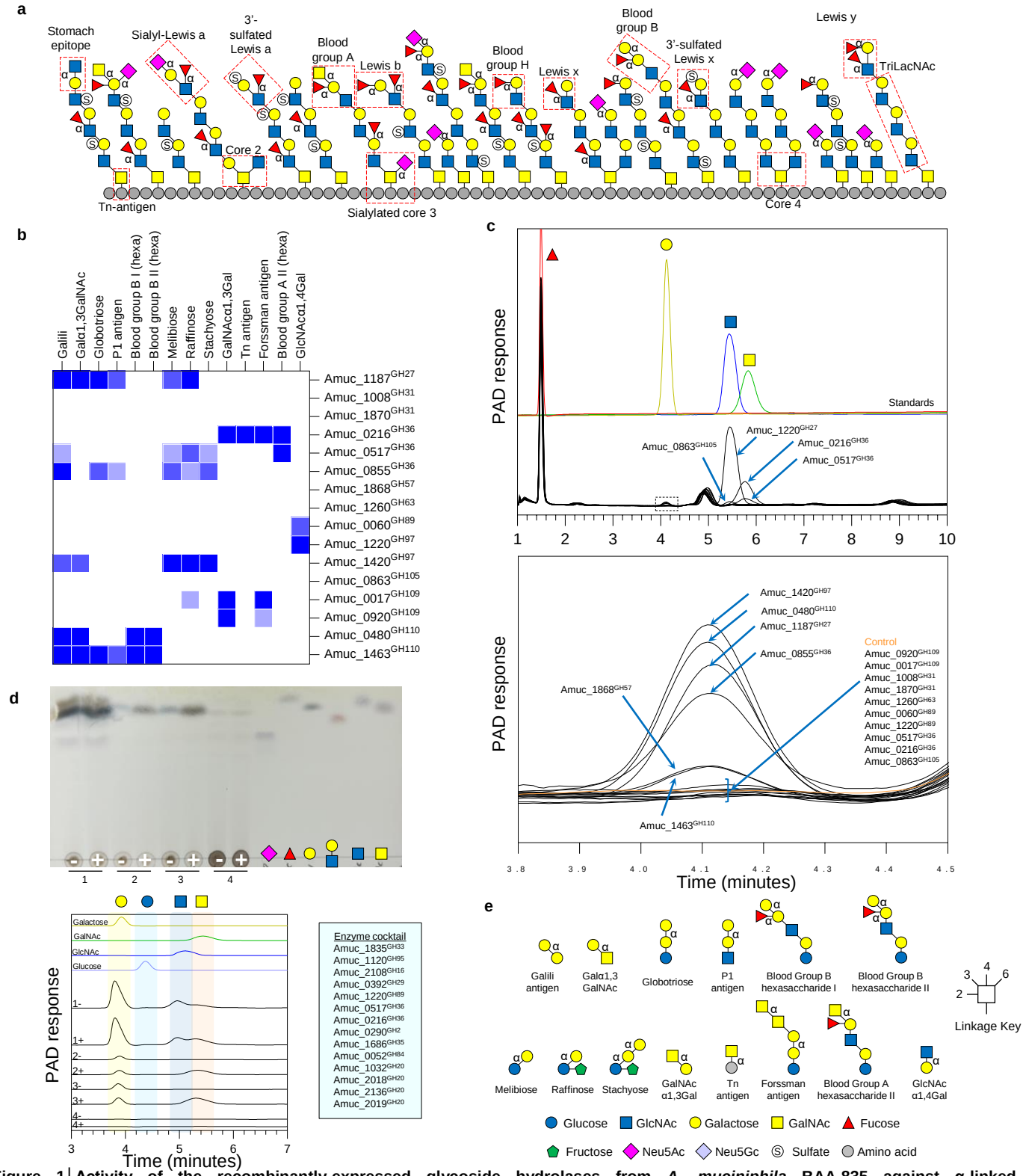


Figure 1 | Activity of the recombinantly-expressed glycoside hydrolases from *A. muciniphila* BAA-835 against α-linked monosaccharides. a, Structural features that are expected in natural mucin glycoproteins with epitopes highlighted. Only alpha linkages are labelled apart from the core GalNAc monosaccharides which are also alpha linkages. **b,** Heat map of recombinant enzyme activities against defined oligosaccharides. The dark blue and white indicate full and no activity, respectively, and partial activities are represented by the lighter blues. **c** Activity of enzymes against PGMIII (pre-treated with Amuc_1835^{GH33} and Amuc_1120^{GH95}) and analysed using HPAEC-PAD. Standards were also run to identify monosaccharide products. Top panel: relevant area of chromatograms. Bottom panel: zoom in of where galactose elutes to show detail (dashed box in top panel). **d** A cocktail of enzymes from *A. muciniphila* BAA-835 can completely degrade the O-glycans from PGMIII down to Tn antigen. Top and bottom panel are a TLC and HPAEC-PAD of the results, respectively. PGMIII was incubated with a cocktail of enzymes (in pale blue box) and the reaction stopped by boiling (1), this reaction was then dialysed to remove all free monosaccharides and glycans (2) and concentrated (3). Untreated PGMIII was also included as a control (4). The application of Amuc_1008^{GH31} is indicated by plus signs. Standards have also been included. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 μM enzymes. **e**, Structures of the defined substrates used to characterise the enzymes active against α-linked monosaccharides.

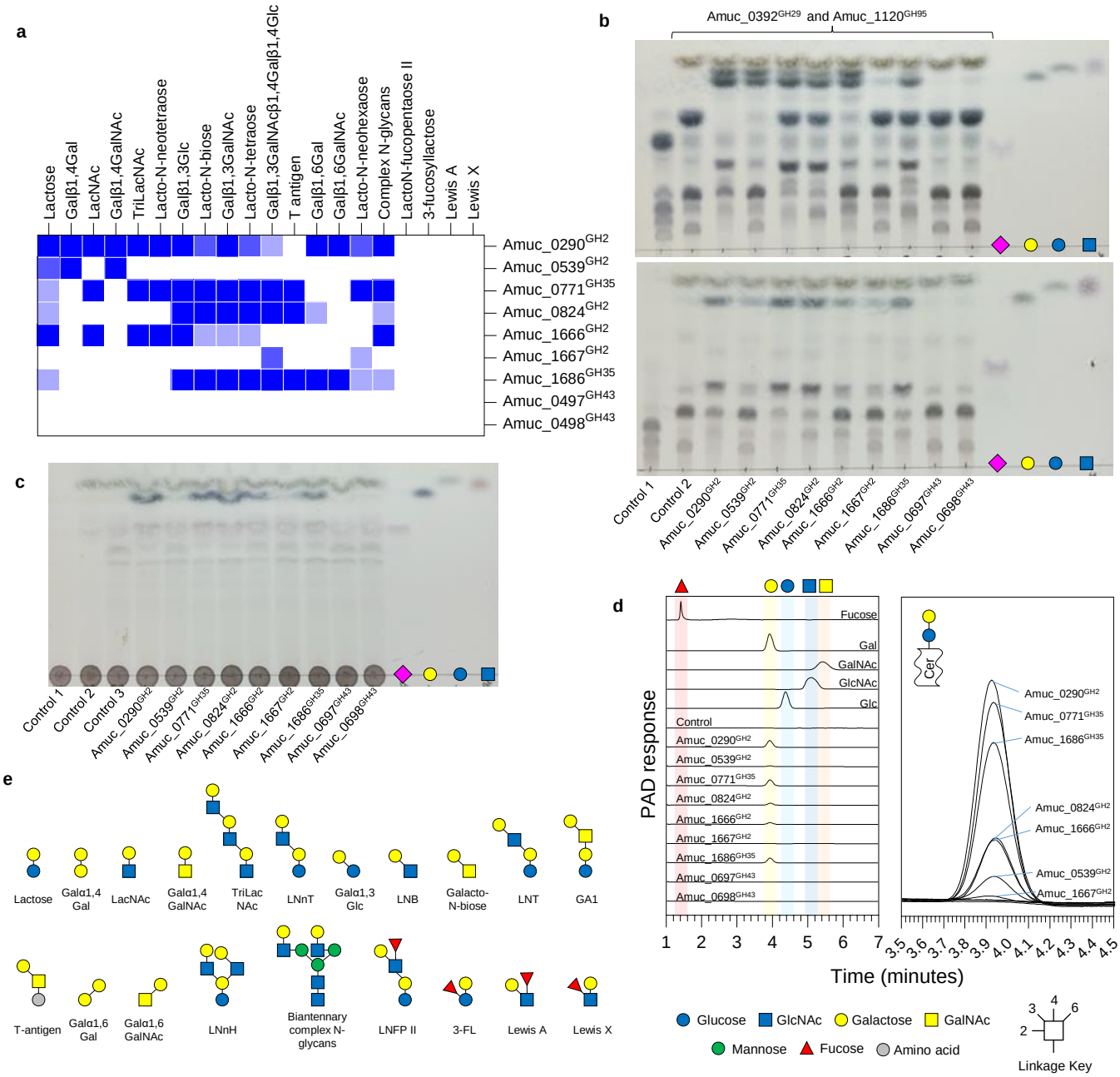


Figure 2 | Activity of the glycoside hydrolases from familys 2, 35, and 43 from *A. muciniphila* BAA-835 against β-linked galactose. **a,** Heat map of recombinant enzyme activities against defined oligosaccharides. The dark blue and white indicate full and no activity, respectively, and partial activities are represented by the lighter blues. **b**, Thin layer chromatography results of GH2/35/43 enzymes against HMOs. Top and bottom panels are sialylated and neutral HMO preparations (Biosynth). Control 1 – no enzymes added and Control 2 - Amuc_0392^{GH29} and Amuc_1120^{GH95}. The rest of the lanes are the individual enzymes with the fucosidases are also included in these reactions. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 μM enzymes. Monosaccharide standards are shown in the right lanes. **c**, This layer chromatography results of GH2/35/43 enzymes against PGM III. Control 1 – no enzymes added, Control 2 - Amuc_1835^{GH33} and Amuc_1120^{GH95}, Control 3 – Control 2 + Amuc_2108^{GH16}. The rest of the lanes are the individual enzymes added in the background of 'Control 3'. These assays were performed in stages, so the reactions were boiled after the 'Control 2' incubation and then again after the 'Control 3' incubation, after which the β-galactosidases were added. This makes the results much easier to interpret. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 μM enzymes. Monosaccharide standards are shown in the right lanes. **d**, Activity of the GH2/35/43 enzymes against Lactosylceramide. The assays were analysed using HPAEC-PAD and the different controls confirm the release of galactose. Left panel – the different controls and samples are stacked. Right panel – The assay chromatograms are overlaid for comparison so the galactose peaks can been observed in more detail. The substrate could not be resolved by this method, so whether the reaction had gone to completion could not be determined and thin layer chromatography of the samples was also inconclusive. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 μM enzymes. **e**, Structures of the defined substrates used to characterise the enzymes active against β-linked galactose.

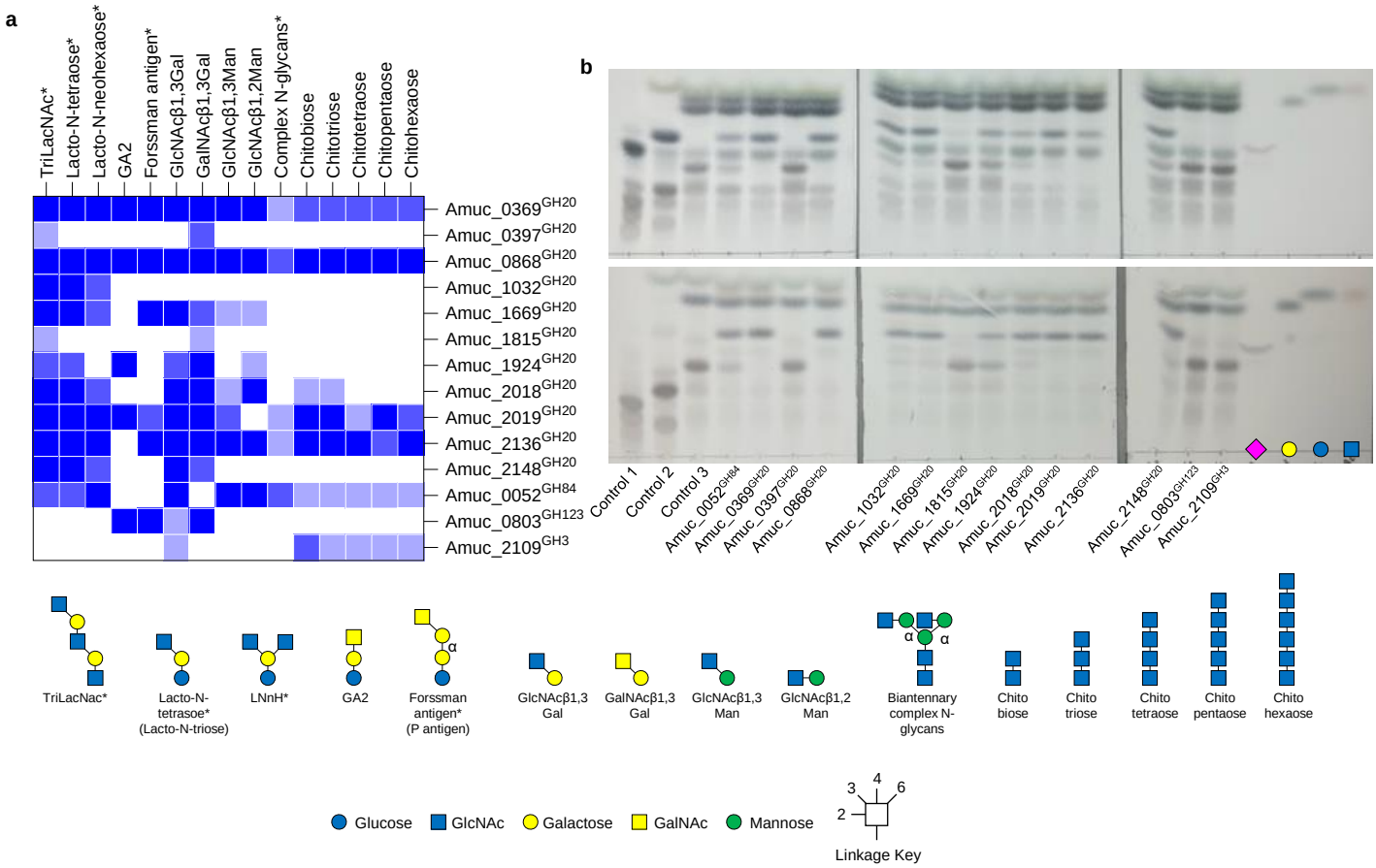


Figure 3 | Activity of β-HexNAc'ases from glycoside hydrolases families 20, 84,, 123 and 4 from *A. muciniphila* BAA-835. a, Heat map of recombinant enzyme activities against defined oligosaccharides. The dark blue and white indicate full and no activity, respectively, and partial activities are represented by the lighter blues. Asterisk represents those substrates that were generated by pre-treating with other CAZymes. **b**, Thin layer chromatography results of enzymes against HMOs. Top and bottom panels are sialylated and neutral HMO preparations (Biosynth). Control 1 – no enzymes added and Control 2 - Amuc_0392^{GH29} and Amuc_1120^{GH95}. Control 3 – Top panel: Amuc_0517^{GH36} and Amuc_0290^{GH2}. Bottom panel: Amuc_0216^{GH36} and Amuc_0771^{GH35}. The controls were boiled in between all stages to produce a sequential degradation. The rest of the lanes are the individual β-HexNAc'ases. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 μM enzymes. Monosaccharide standards are shown in the right lanes.

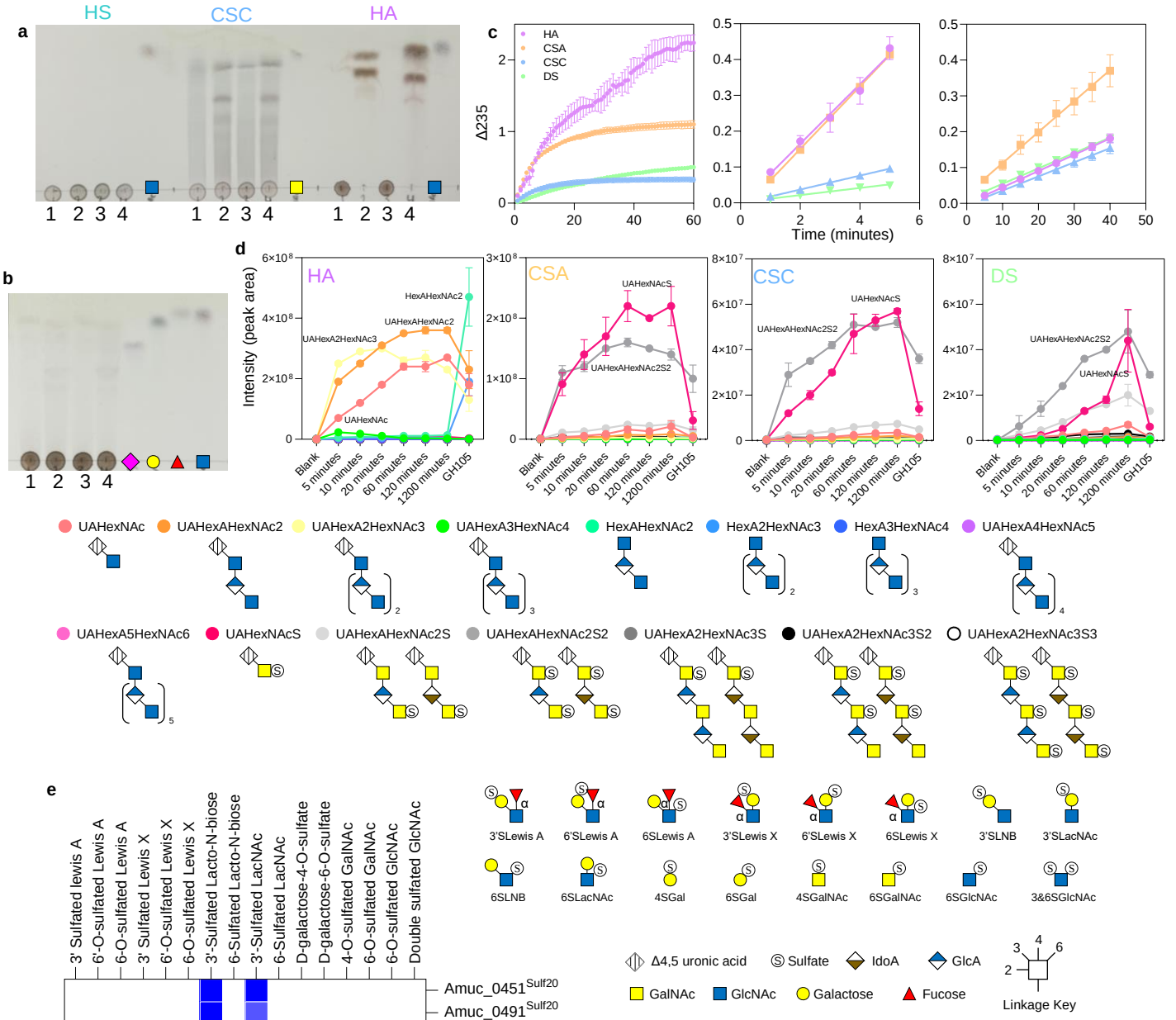


Figure 4 | Activity of the GAG-active enzymes and breakdown of sulfated host glycans by *A. muciniphila* BAA-835. a, Activities of the Amuc_0771^{PL38} and Amuc_0863^{GH105} against different GAGs. 1) control, 2) PL38, 3) GH105, and 4) PL38+GH105. **b**, Testing Amuc_0771^{PL38} and Amuc_0863^{GH105} against PGM III. 1) control, 2) PL38, 3) GH105, and 4) PL38+GH105. **c**, Left: substrate depletion of different GAGs (2 g/L) by Amuc_0771^{PL38}. This wavelength allows the monitoring of the generation of unsaturated products (double bonds) of the PL38. Middle: Highest initial rates of Amuc_0771^{PL38}. Right: Initial rates of the Amuc_0863^{GH105}. The change in absorbance monitors the GH105 activity against the Amuc_0771^{PL38} product. **d**, Quantification of the Amuc_0771^{PL38} activity over time from four different GAG substrates by LC-MS. The area under the peaks from the chromatograms are shown and each glycan has been given a different colour, apart from xxx For sulfated glycans, the position of the sulfation is unknown. CSA and DS typically has 4S on GalNAc, whereas CSC has relatively high levels of 6S on GalNAc, but also low levels of 4S. The GlcA in CSC can also have 2S sulfation. HA has no sulfation. **e**, Heat map of the activity of enzymes against sulfated monosaccharides and glycans.

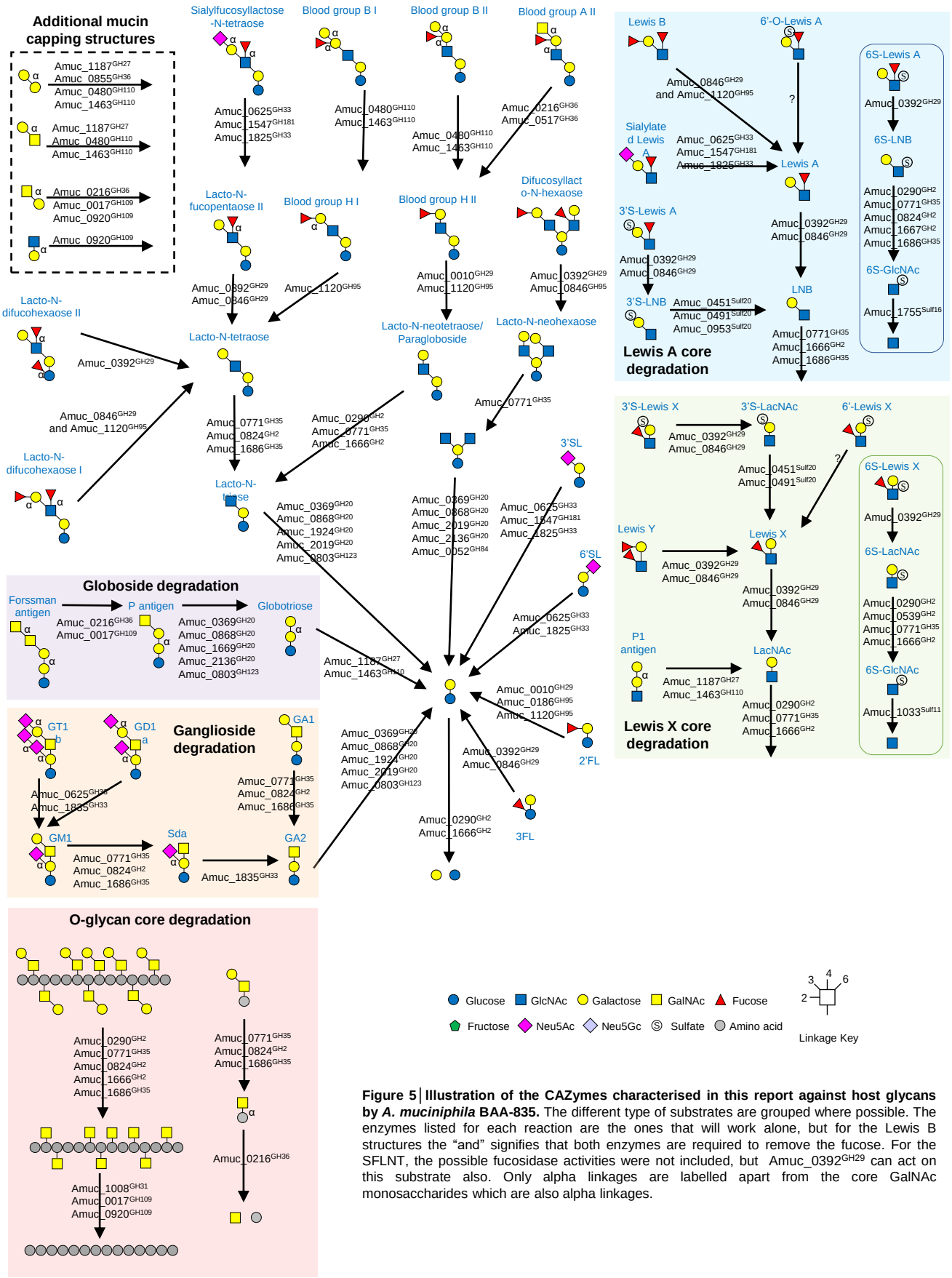
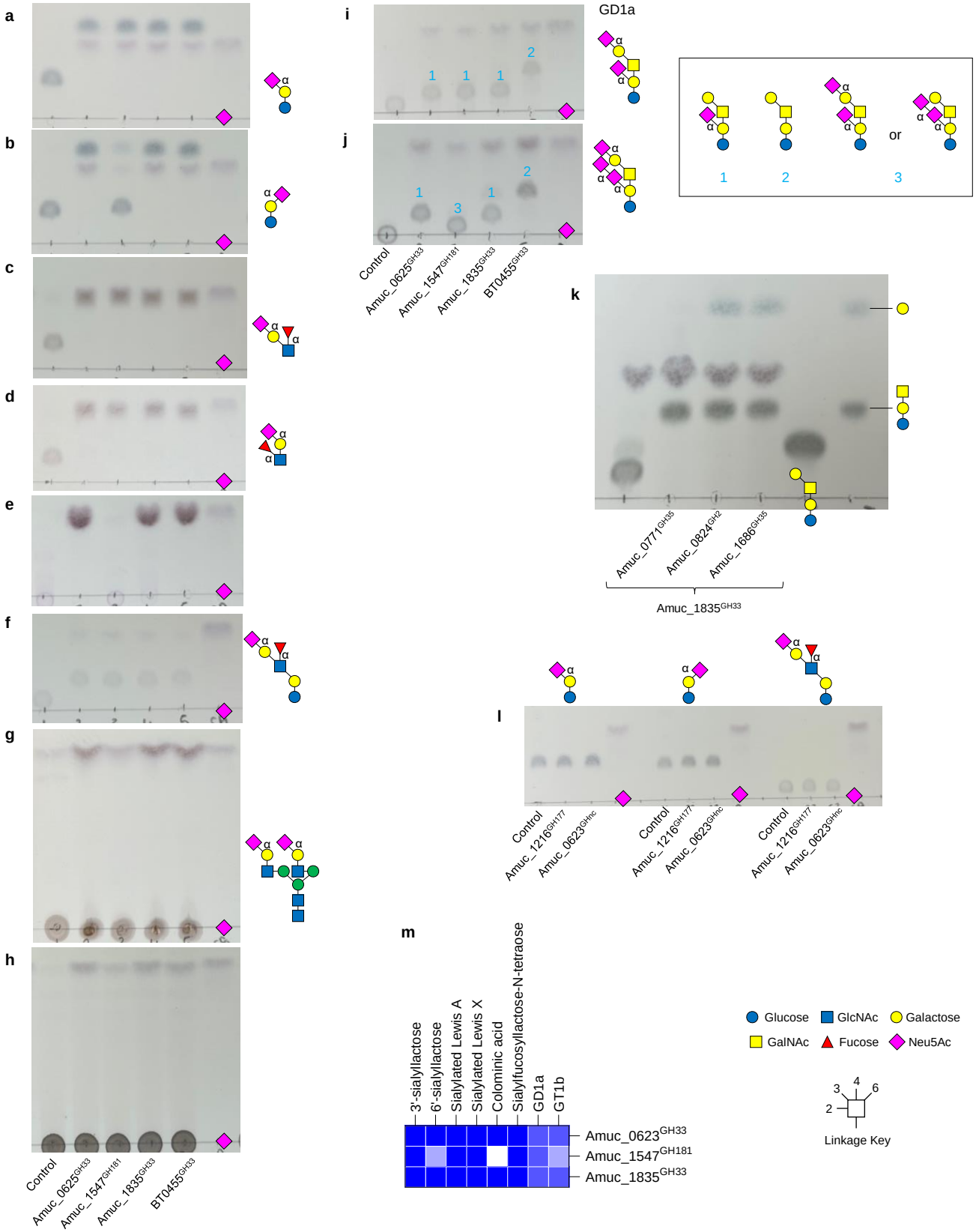
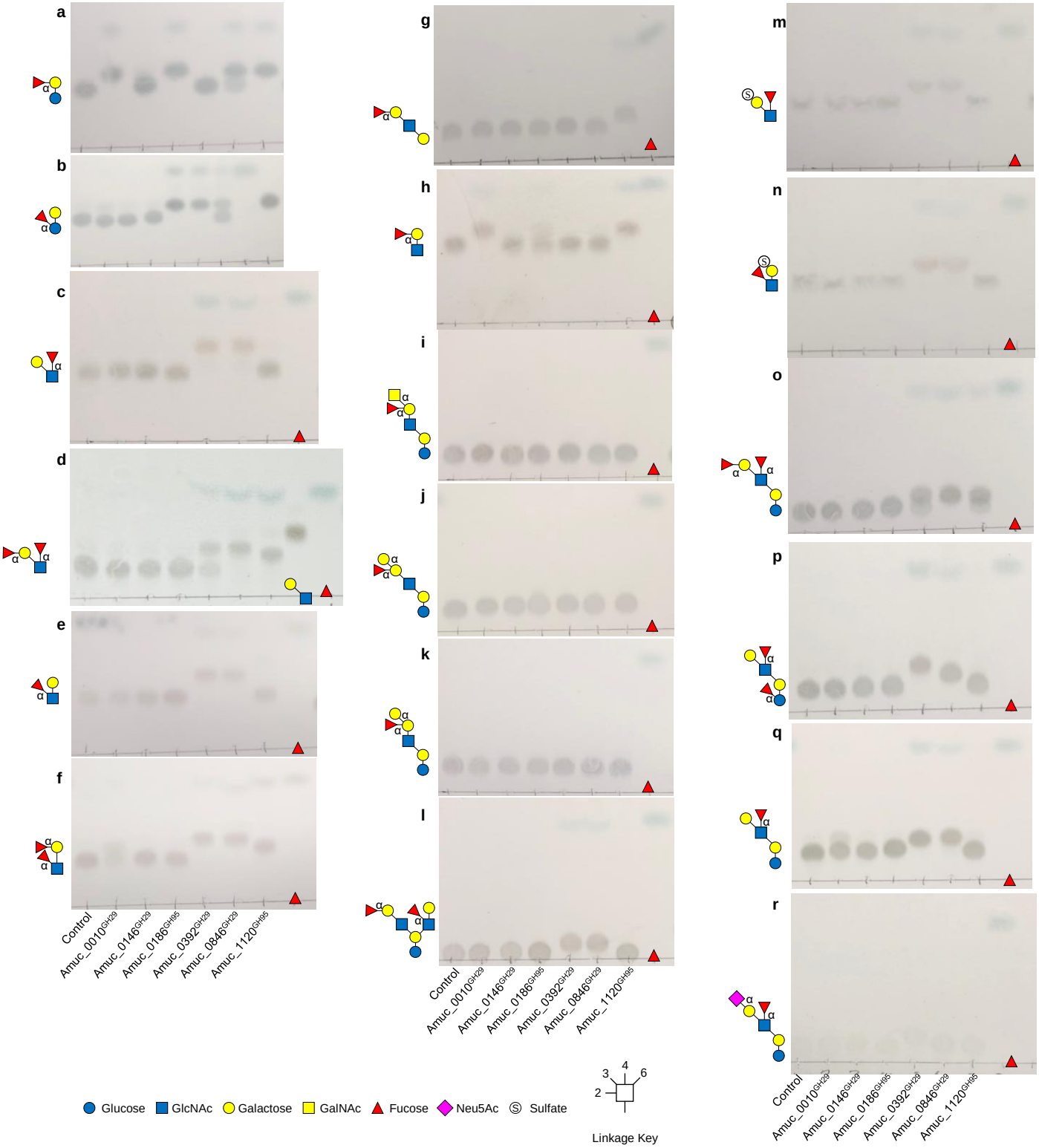


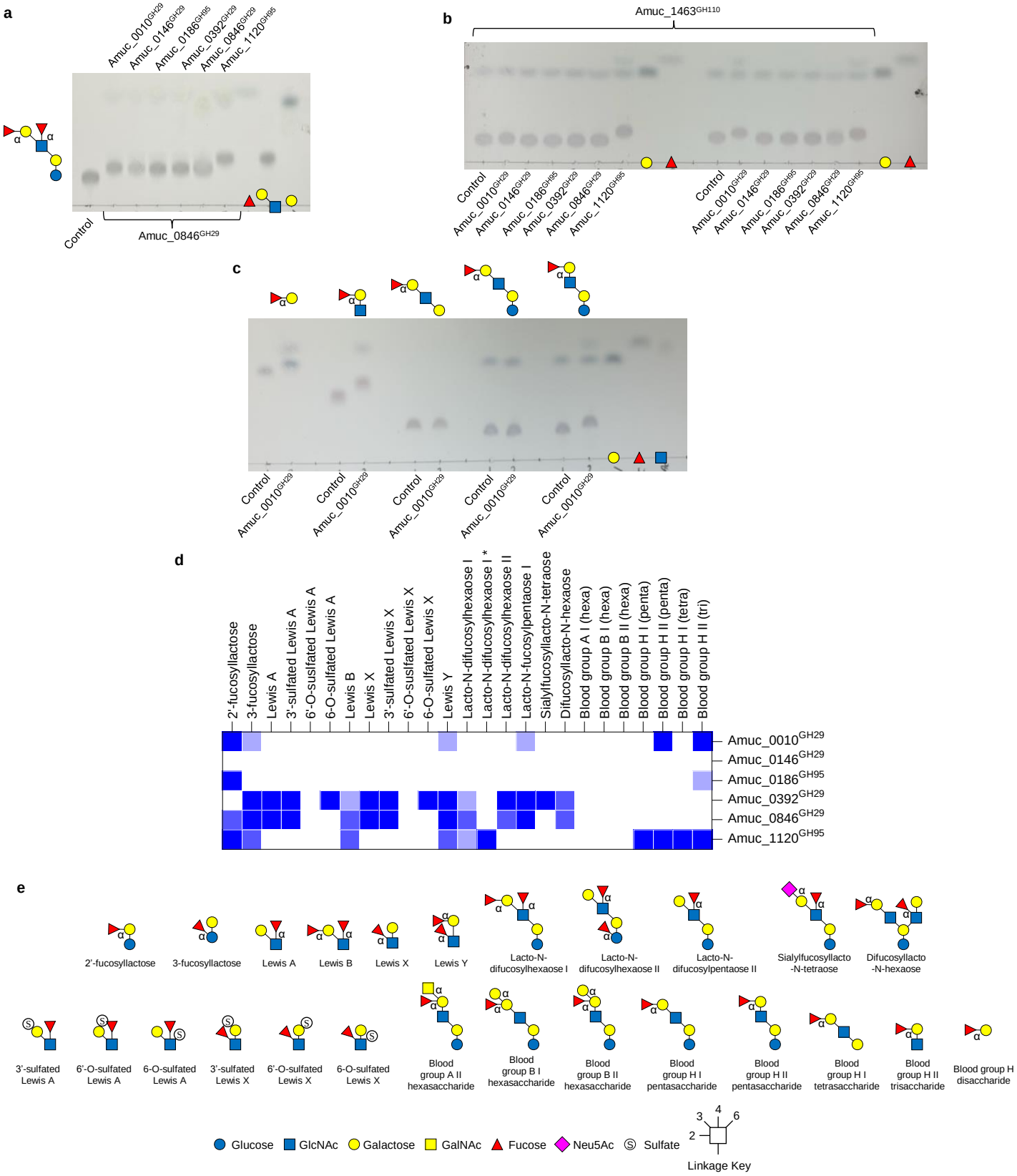
Figure 5 | Illustration of the CAZymes characterised in this report against host glycans by *A. muciniphila* BAA-835. The different type of substrates are grouped where possible. The enzymes listed for each reaction are the ones that will work alone, but for the Lewis B structures the "and" signifies that both enzymes are required to remove the fucose. For the SFLNT, the possible fucosidase activities were not included, but Amuc_0392GH29 can act on this substrate also. Only alpha linkages are labelled apart from the core GalNAc monosaccharides which are also alpha linkages.



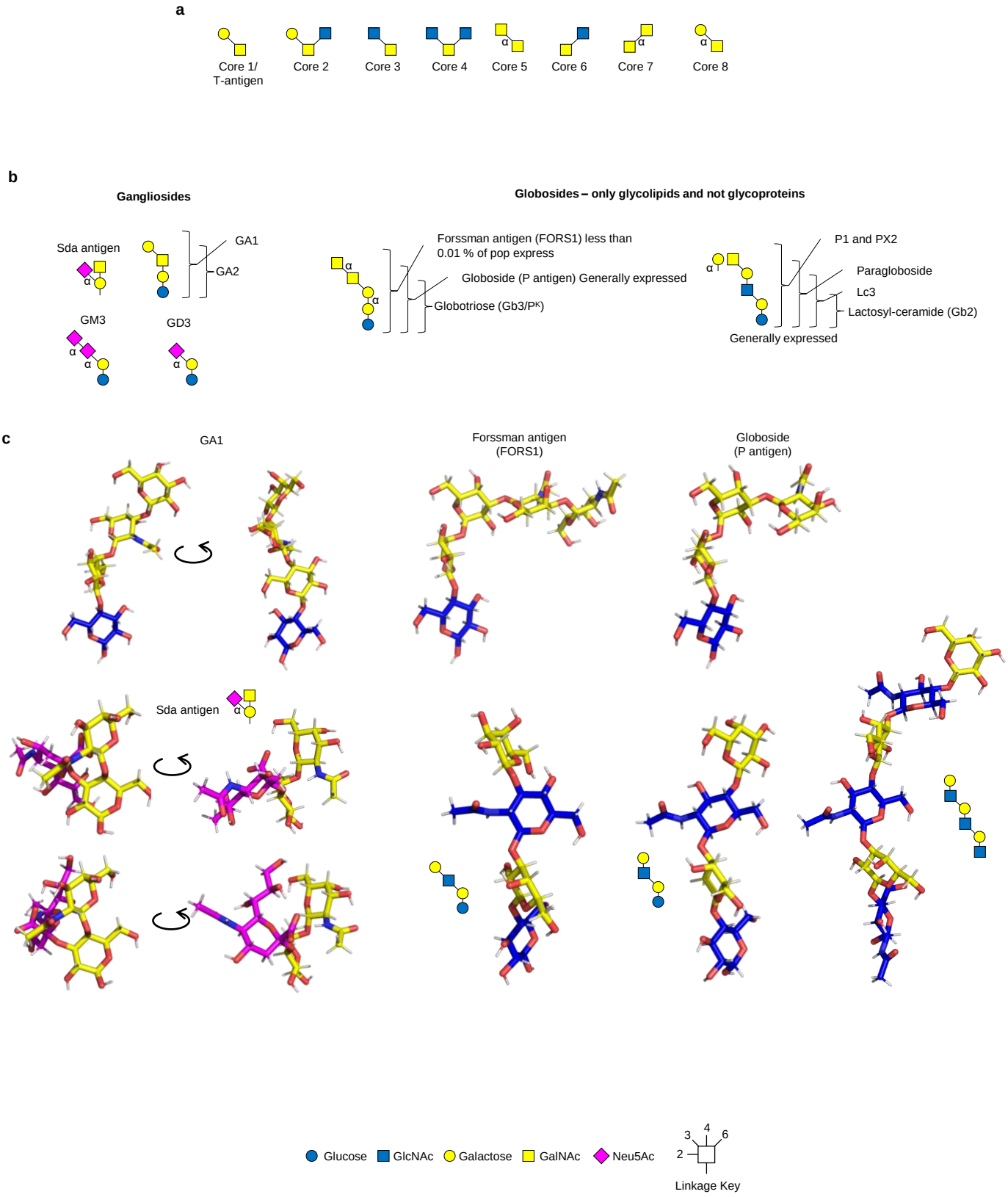
Supplementary Figure 1 | Activity of the GH33 and GH181 family members from *A. muciniphila* BAA-835 against defined oligosacchairdes, polysaccharides, and mucin substrates. a, 3'-sialyllactose. **b**, 6'-sialyllactose. **c**, Sialylated Lewis A. **d**, Sialylated Lewis X. **e**, colominic acid. **f**, sialylfucosyllacto-N-tetraose. **g**, biantennary complex N-glycans ($\alpha_1$acid glycoprotein). **h**, PGM III. **i,** GD1a **j,** and GT1b. **k,** Exploring if removing the galactose also allows the release of the final sialic acid. **l,** Testing Amuc_1216^{GH177} and Amuc_0623^{GHnc} against three different sialylated substrates. **m,** Heat map of recombinant enzyme activities against defined oligosaccharides. The dark blue and white indicate full and no activity, respectively, and partial activities are represented by the lighter blues. Standards have also been included on the TLCs and substrates have been included on the right where possible. For the gangliosides (I and j) the structures of the different products are in the key to the right. Enzyme assays were carried out at a final substrate concentration of 1 mM (except N-glycans at 10 mg/ml and colominic acid at 25 mg/ml), pH 7, 37 °C, overnight, and with 1 μM enzymes. BT0455 from Bacteroides thetaiotaomicron has been used as a positive control a-i.



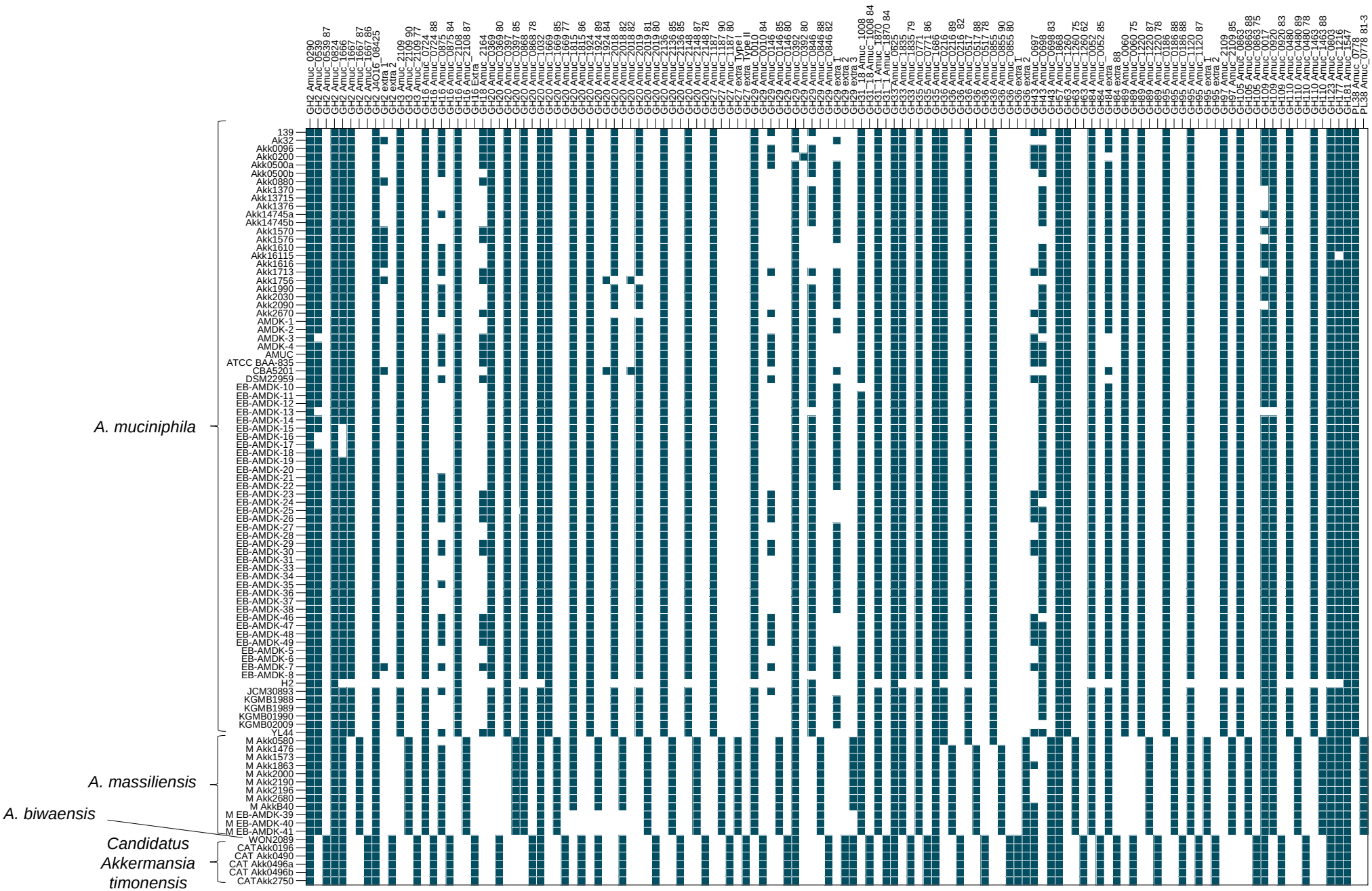
Supplementary Figure 2 | Activity of the GH29 and GH95 family members from *A. muciniphila* BAA-835 against defined oligosacchairdes, Part 1. a, 2'-fucosyllactose. **b**, 3-fucosyllactose. **c**, Lewis A. **d**, Lewis B. **e**, Lewis X. **f**, Lewis Y. **g**, Blood group H I. **h**, Blood group H II. **i**, Blood group A II. **j**, Blood group B I. **k**, Blood group B II. **l**, difucosyllacto-N-hexaose. **m**, 3'-sulphated Lewis A. **n**, 3'-sulphated Lewis X. **o**, Lacto-N-difucohexaose I. **p**, Lacto-N-difucohexaose I. **q**, Lacto-N-fucopentaose II. **r**, Sialylfucosyllacto-N-tetraose. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



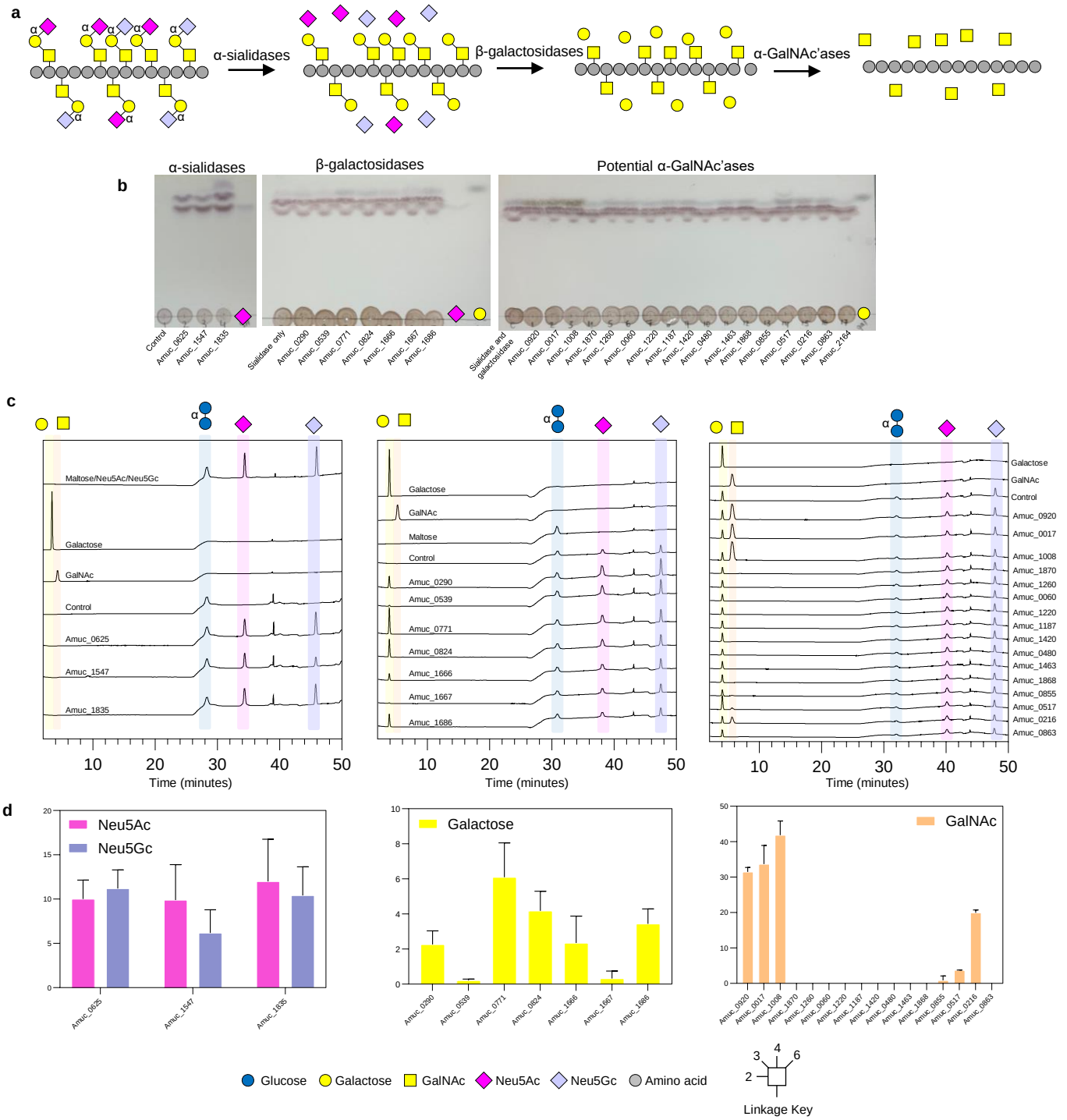
Supplementary Figure 3 | Activity of the GH29 and GH95 family members from *A. muciniphila* BAA-835 against defined oligosacchairdes, Part 2. a, Exploring the full degradation of LNDFH I by pre-treating with Amuc_0846^{GH29} and then adding the other fucosidases. **b**, Fucosidase activity against different types of blood group B structures where the α-linked galactose has been removed. **c**, Amuc_0010^{GH29} activity against a range of blood group H structures. For all assays, standards have been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme. **d**, A. heat map of recombinant enzyme activities against defined oligosaccharides. The dark blue and white indicate full and no activity, respectively, and partial activities are represented by the lighter blues. The asterisk indicates that one fucose has been remove by Amuc_0846^{GH29}. **e**, Structures of the defined substrates used to characterise the fucosidases.



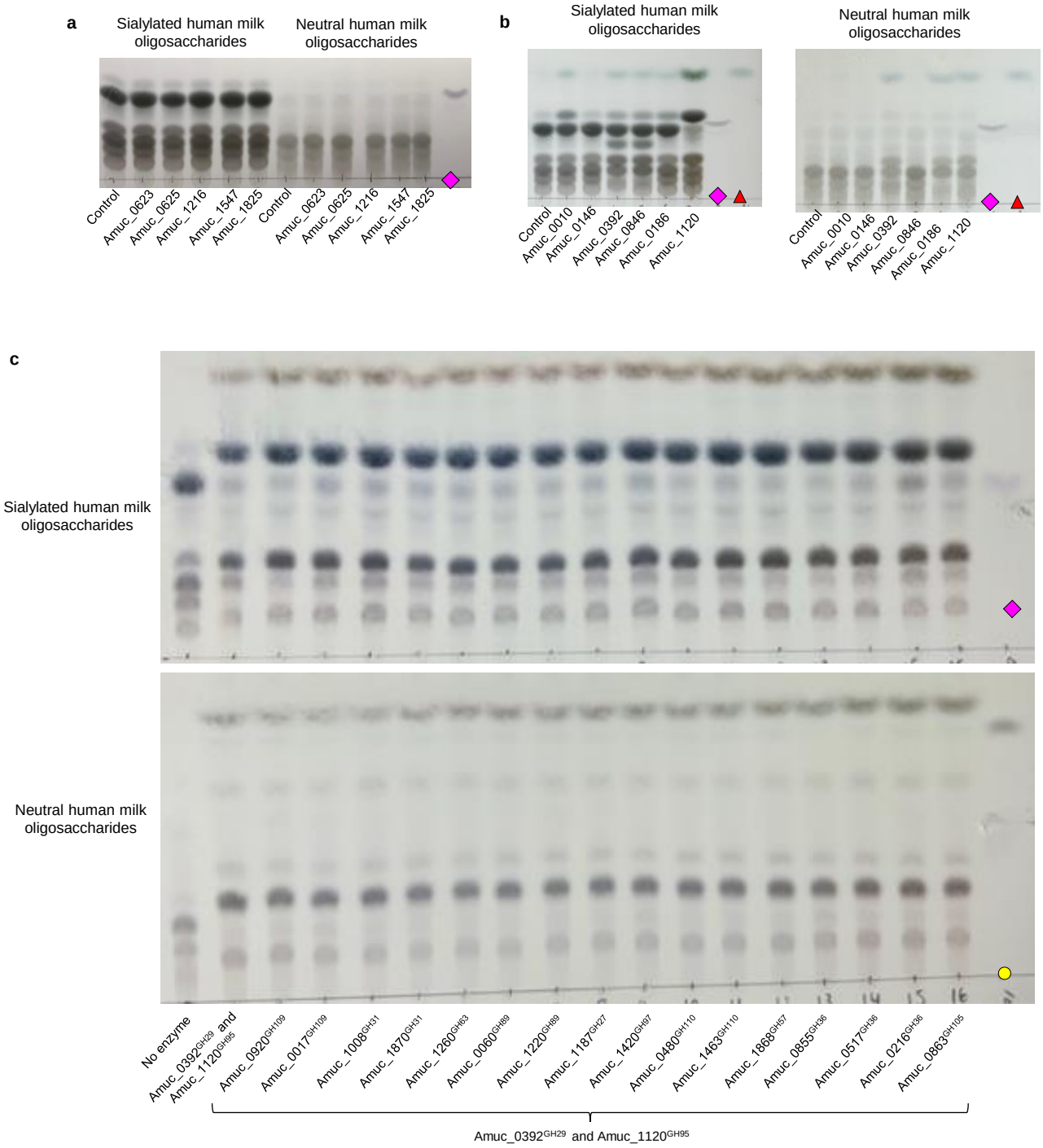
Supplementary Figure 4 | Further structural information about human-associated glycans. a, O-glycan core structures. Only alpha linkages are labelled. **b**, Details of gangliosides and globosides, with some specific structures where the paper requires. **c**, Models of some glycans to help visualise and analyse the specificities of the different CAZymes. These were made in GYLCAM using the Carbohydrate Builder.



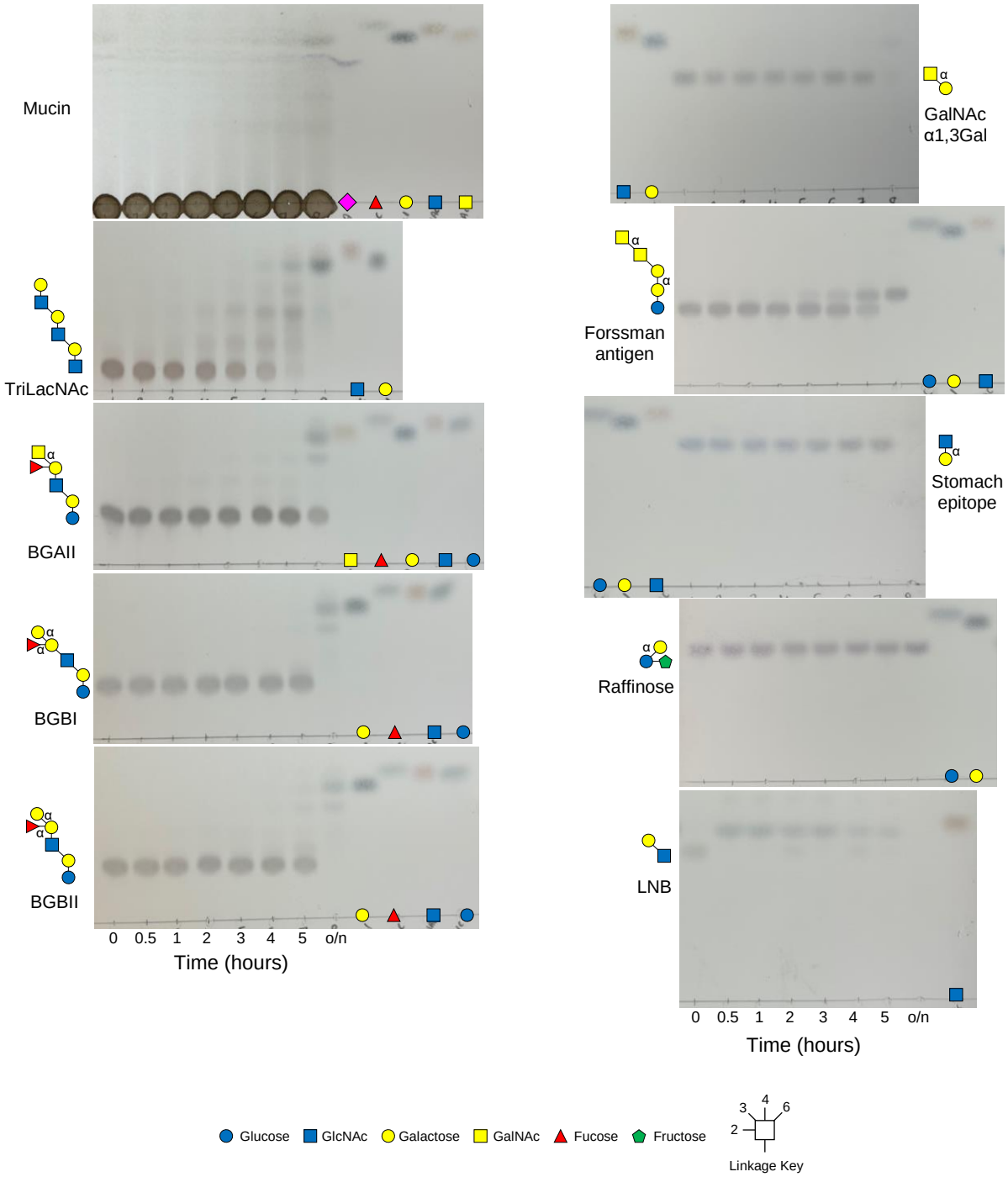
Supplementary Figure 5 | Presence of homologues of the putative CAZymes from ATCC BAA-835 across other *A. muciniphila* strains and *Akkermansia* species. The heat map shows the presence of homologues to the CAZymes in ATCC BAA-835 (dark teal). When the homologue is <90 % identical, <80 % identical, and <70 % identical then these are placed in a separate columns. For instance, the homologues for Amuc_0290^{GH2} are all >90 % across strains and species, whereas the homologues for Amuc_1667^{GH2} all have lower percentage identity across the species but >90 % conservation within strains. The y axis is split up according to species and the x axis is in numerical GH order.



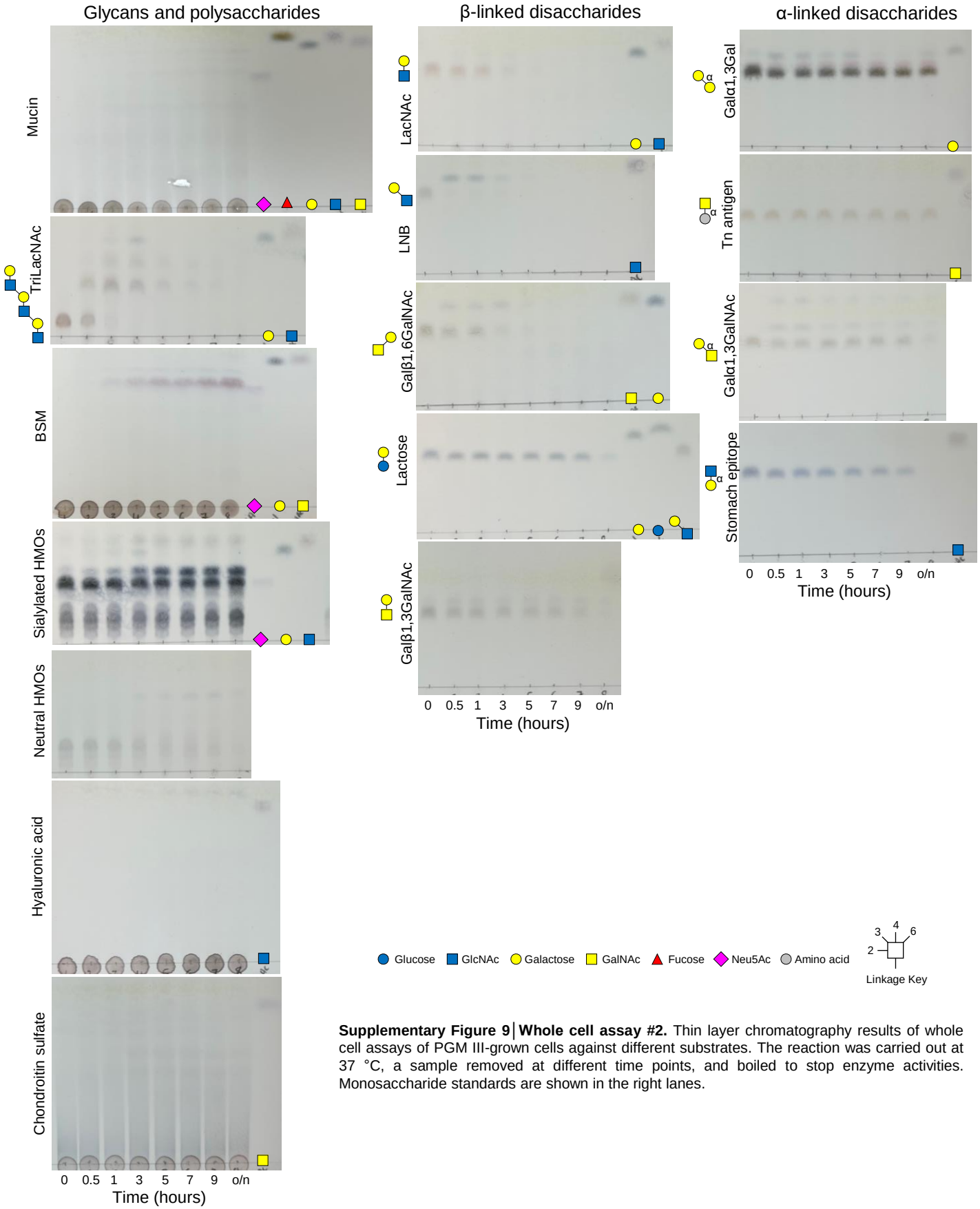
Supplementary Figure 6 | Sequential degradation of O-glycans on BSM with CAZymes from *A. muciniphila* BAA-835. The sialidases, β-galactosidases, and CAZymes associated with α-linked monosaccharide hydrolysis were added sequentially to understand specificity of the CAZymes. **a,** Structural features that are expected in bovine submaxilliary mucin (left) and how it is broken down in this experiment. Only alpha linkages are labelled apart from the core GalNAc monosaccharides which are also alpha linkages. **b**, Thin layer chromatography results of the sequential degradation. Standards have also been included on the left of the TLCs. **c**, The chromatograms of the assays analysed by HPAEC-PAD. Maltose was used as the internal standards and other standards were run separately. **d**, The areas of the peaks were quantified. The data is from three distinct reactions.



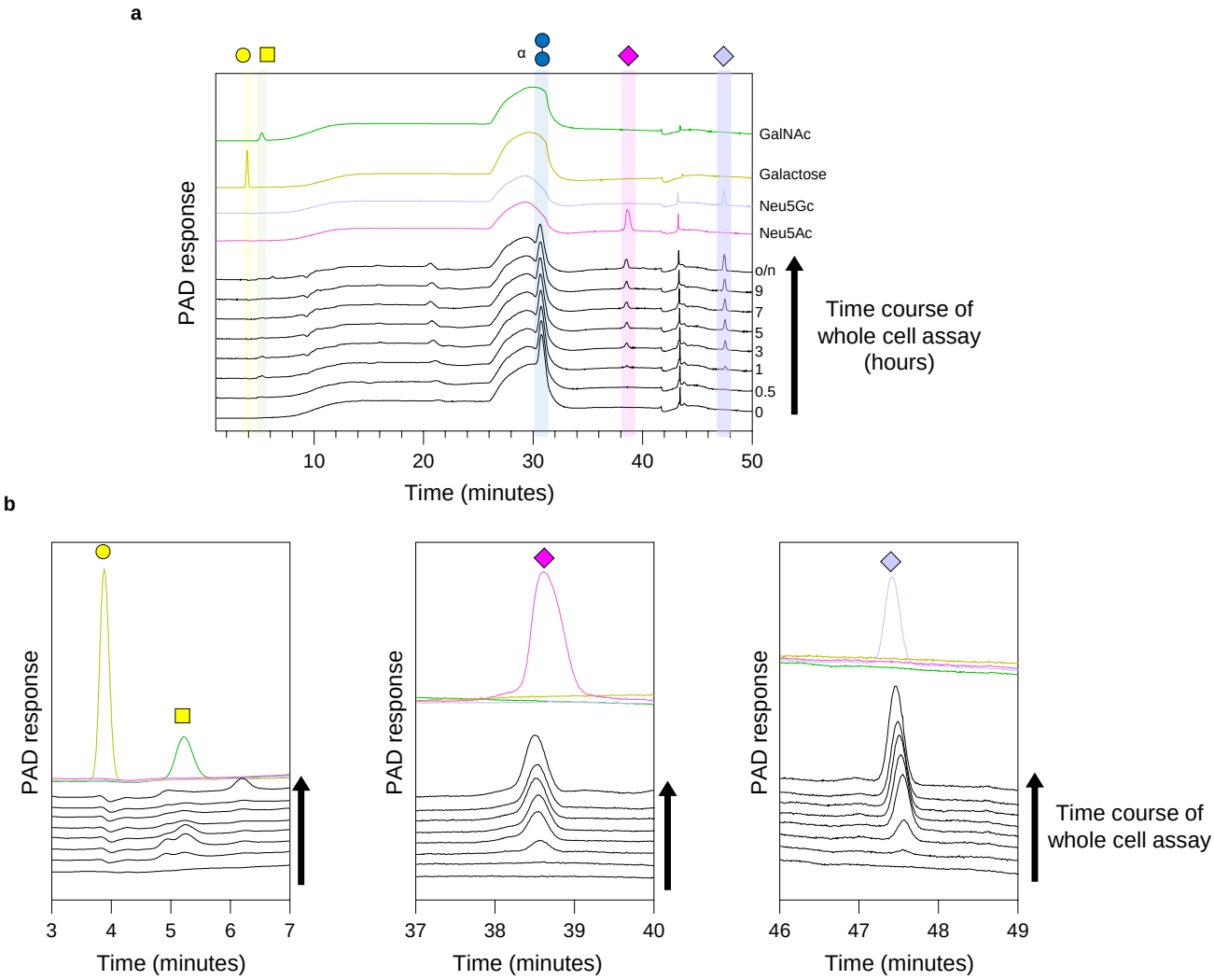
Supplementary Figure 7│Activity of sialidases, fucosidases and other α-linkage acting GH enzymes from *A. muciniphila* BAA-835 against human milk oligosaccharides. a, Sialidases and potential sialidases. **b**, Fucosidases. **c**, CAZymes to predicted to act on α-linked monosaccharides. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



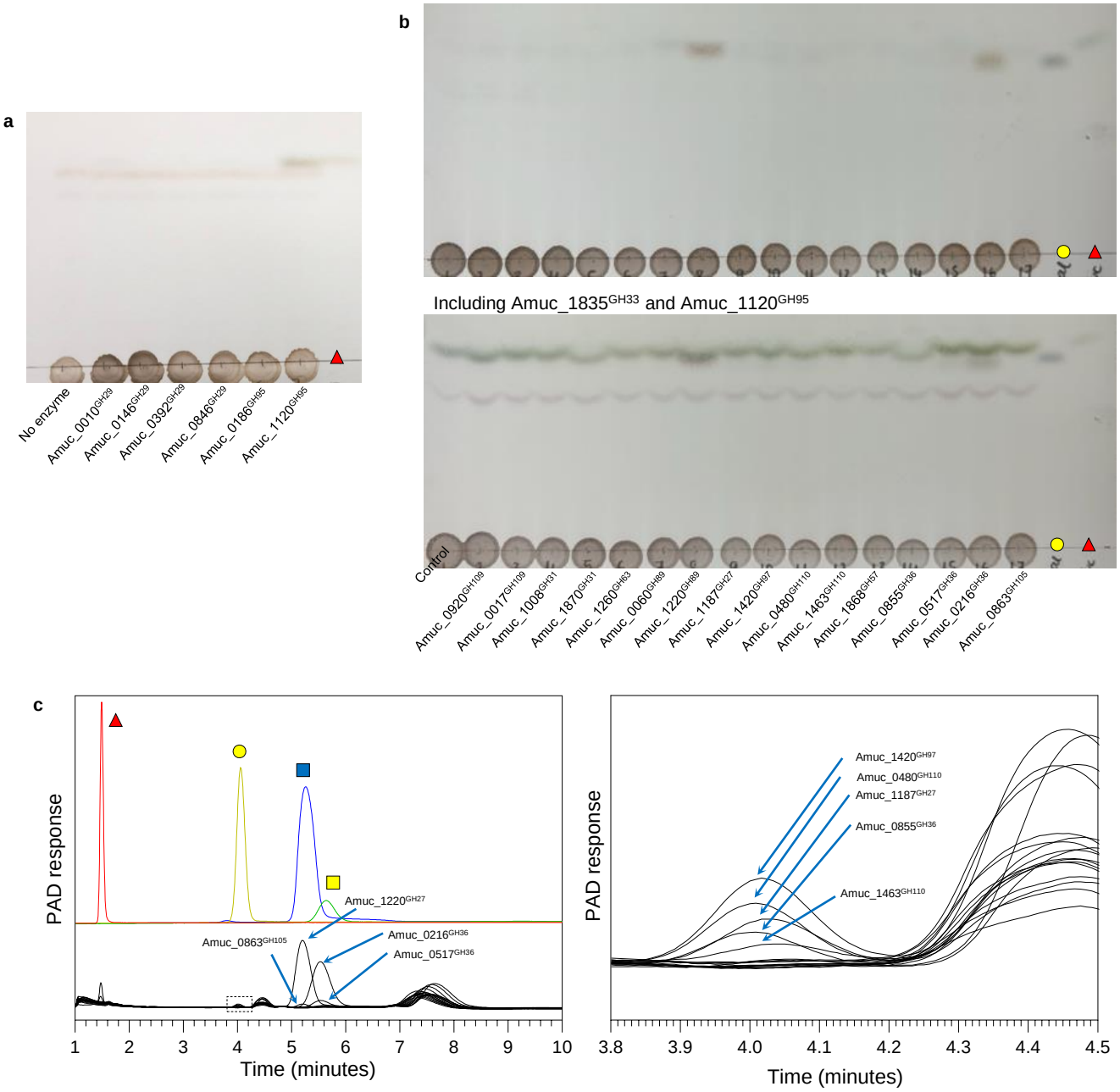
Supplementary Figure 8 | Whole cell assay #1. Thin layer chromatography results of whole cell assays of PGM III-grown cells against different substrates. The reaction was carried out at 37 °C, a sample removed at different time points, and boiled to stop enzyme activities. Monosaccharide standards are shown in the right lanes.



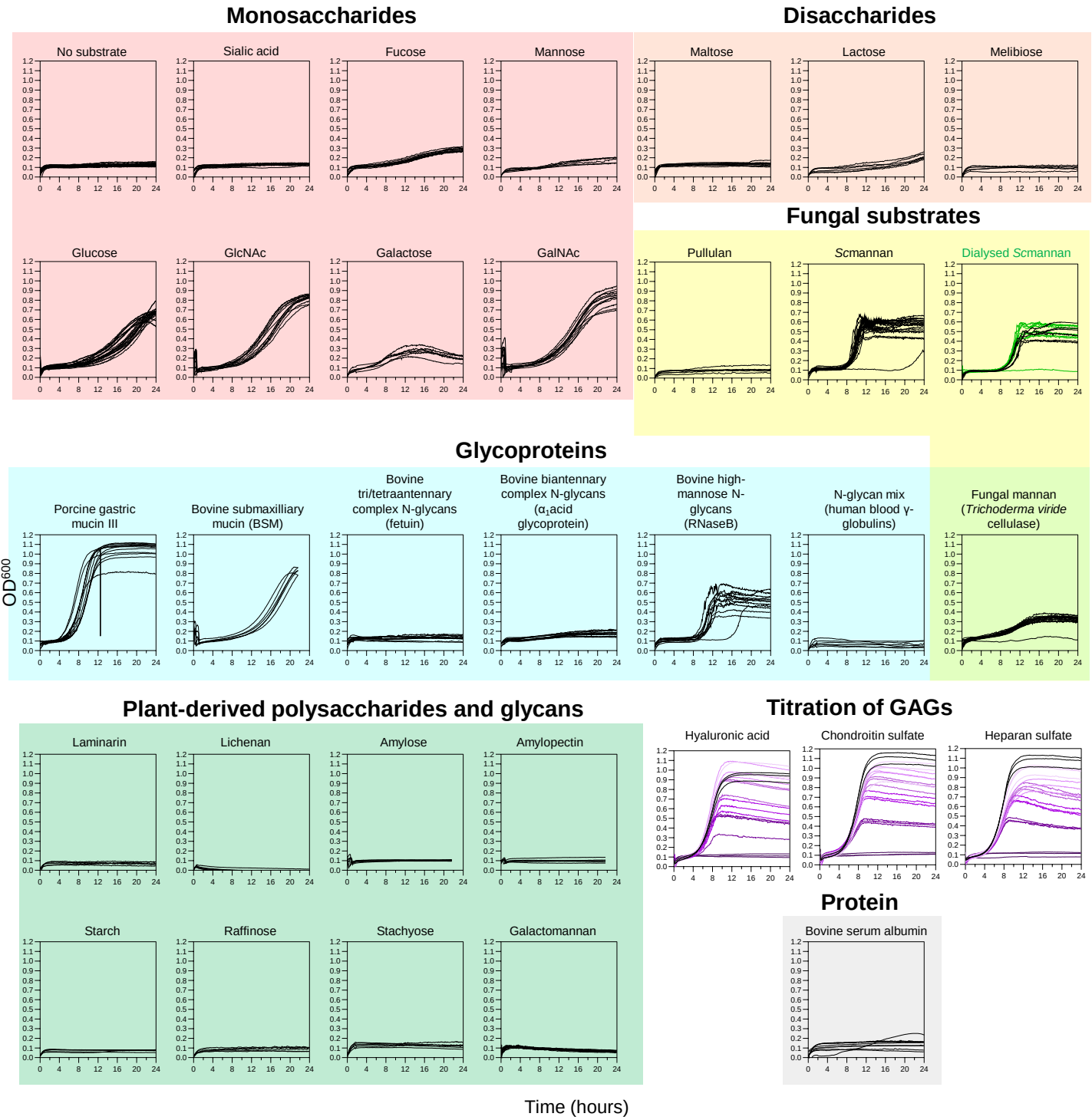
Supplementary Figure 9 | Whole cell assay #2. Thin layer chromatography results of whole cell assays of PGM III-grown cells against different substrates. The reaction was carried out at 37 °C, a sample removed at different time points, and boiled to stop enzyme activities. Monosaccharide standards are shown in the right lanes.



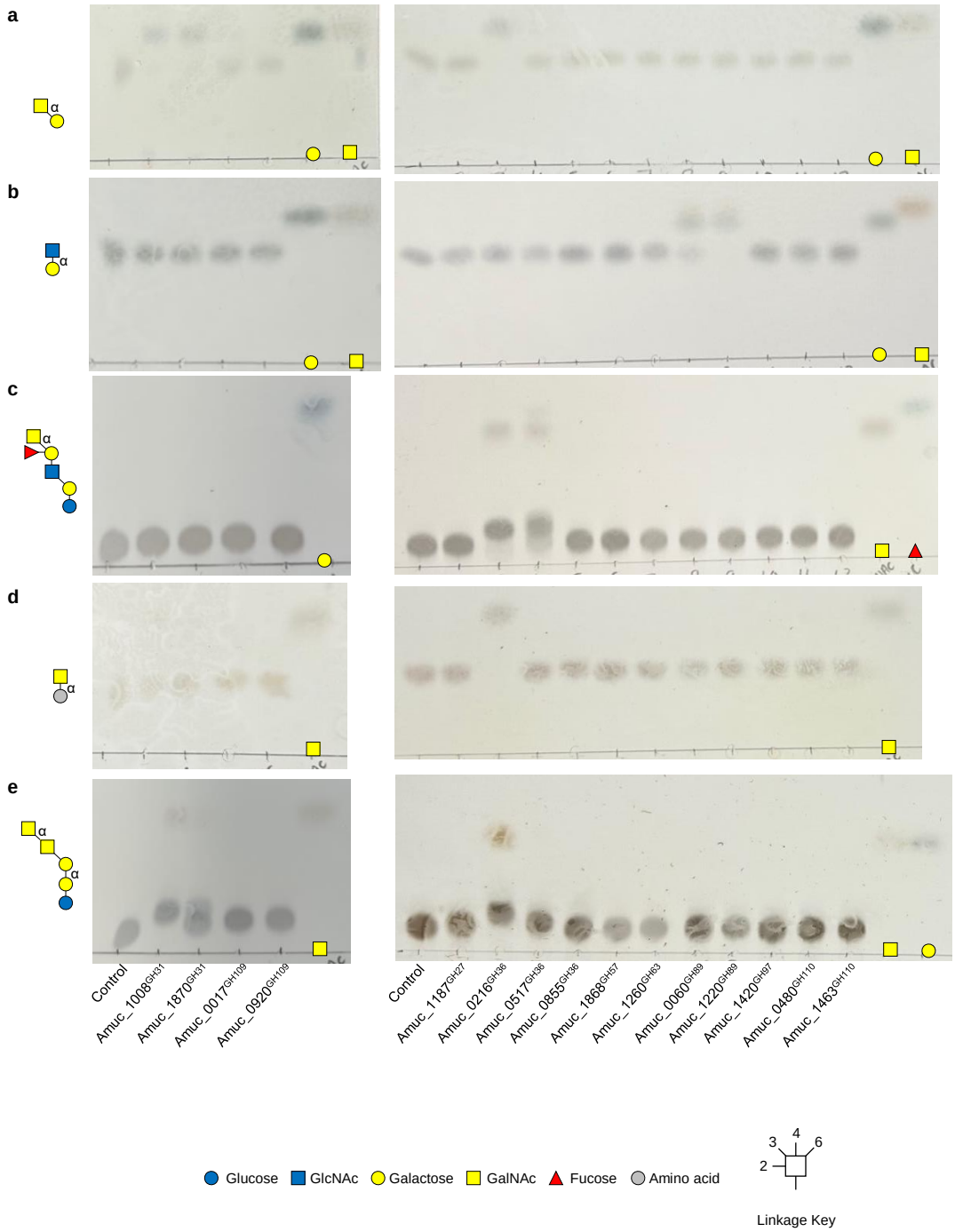
Supplementary Figure 10 | HPAEC-PAD of BSM whole cell assays time points. Samples from Supplementary Figure 9 of PGM III-grown whole cell assays against BSM were analysed by HPAEC-PAD. **a,** The full chromatograms are shown alongside standards of the different expected monosaccharides (colourful lines). **b,** Certain sections of the chromatograms are showed in more detail and stacked to show the increase in sialic acids produced over time.



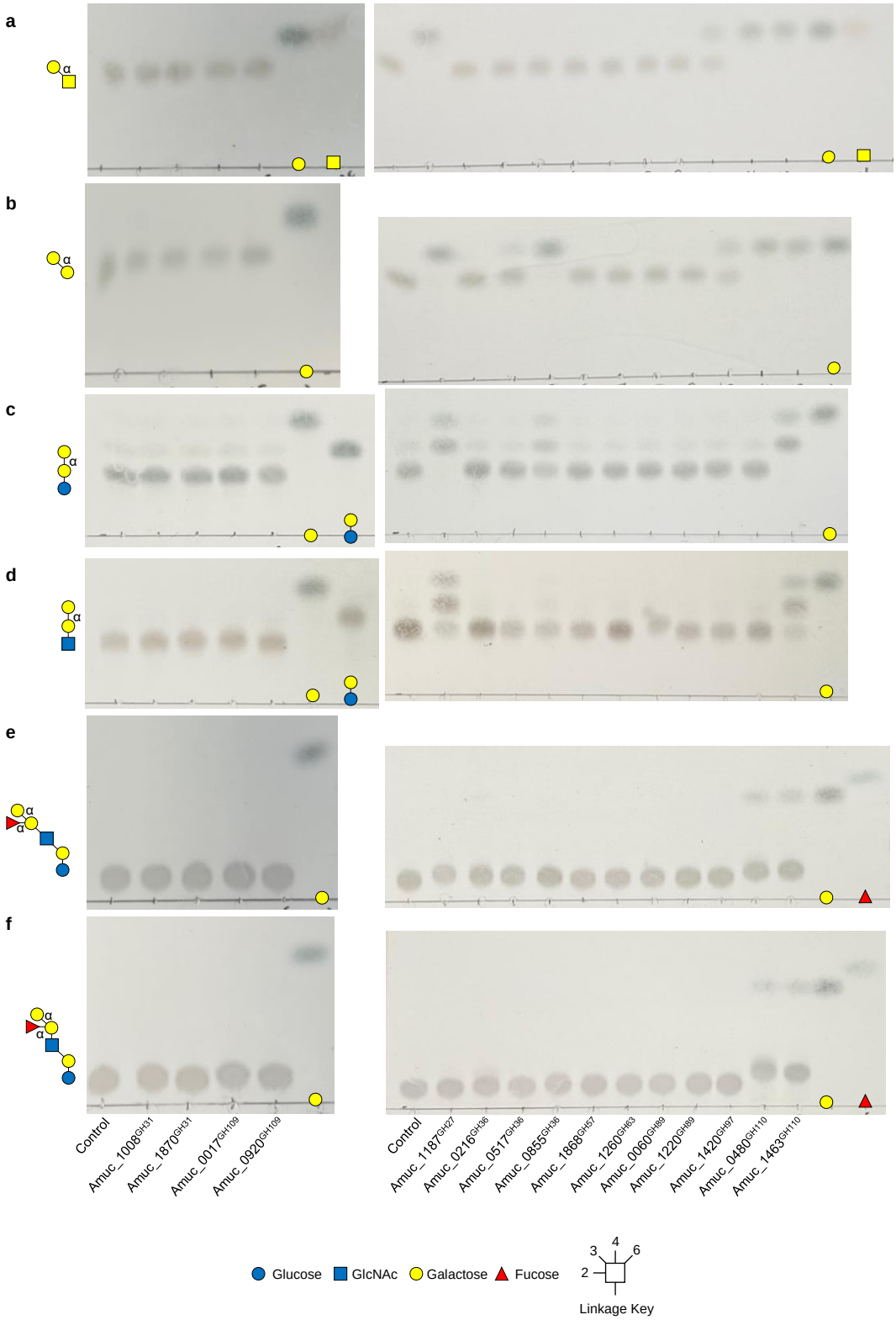
Supplementary Figure 11 | Degradation of PGM III with fucosidases and alpha-linkage CAZymes from *A. muciniphila* BAA-835 and analysis by thin layer chromatography. Enzyme assays were carried out pH 7, 37 °C, overnight, and with 1 µM enzymes. **a,** The activity of GH29 and GH95 family members against PGM III. **b**, The activity of GH enzymes associated with the hydrolysis of alpha-linkage removal. Top panel is without enzyme pre-treatment included and the HPAEC-PAD for these samples is presented in 'c'. The bottom panel are the same assays with Amuc_1835^{GH33} and Amuc_1120^{GH95} included and the HPAEC-PAD for these samples is presented in Figure 1.



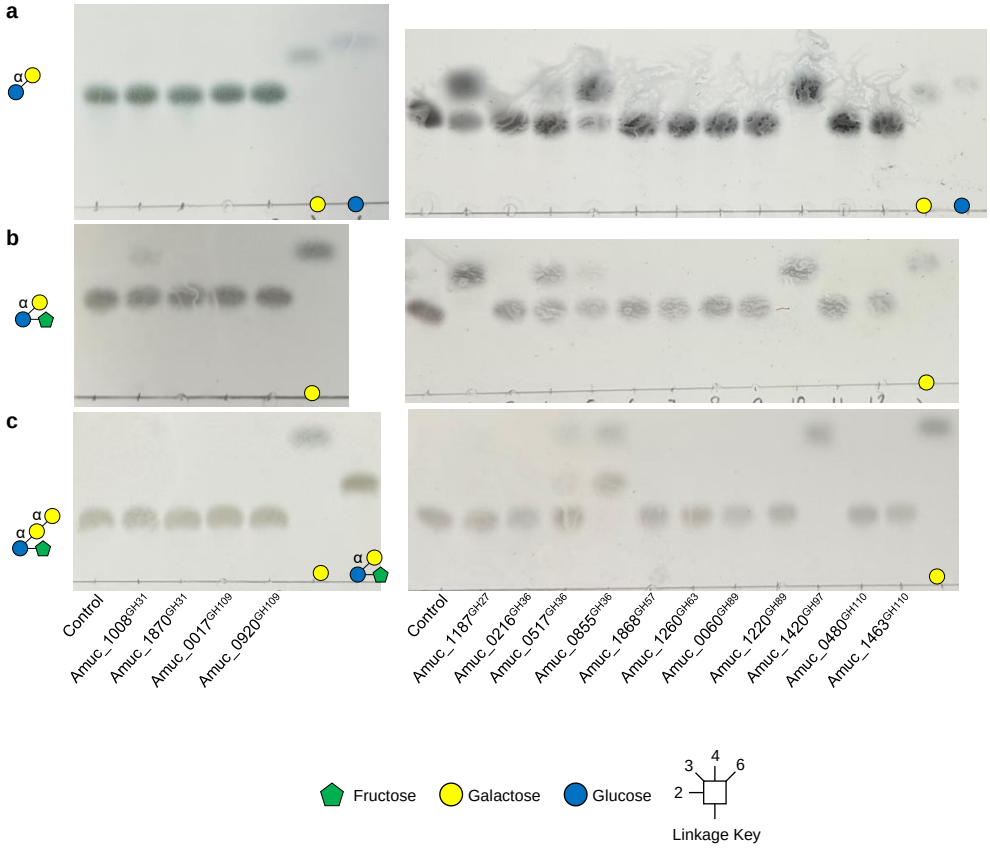
Supplementary Figure 12 | Growth of *A. muciniphila* ATCC BAA-835 on a variety of glycans and polysaccharides. *A. muciniphila* was grown anaerobically in minimal media containing a different potential nutrient sources. Growth was monitored continuously at OD600 using a 96-well plate in a plate reader. Concentrations of substrates are listed in Supplementary Table 2. For the dialysed Scmannan, the black curves are original sample and the green it dialysed. For the GAG substrates, this was completed with varying concentrations of mucin:GAG. The black lines are the mucin growth curves and the darkest purple are the GAGs alone. The lighter the purple gets the more mucin in those growths.



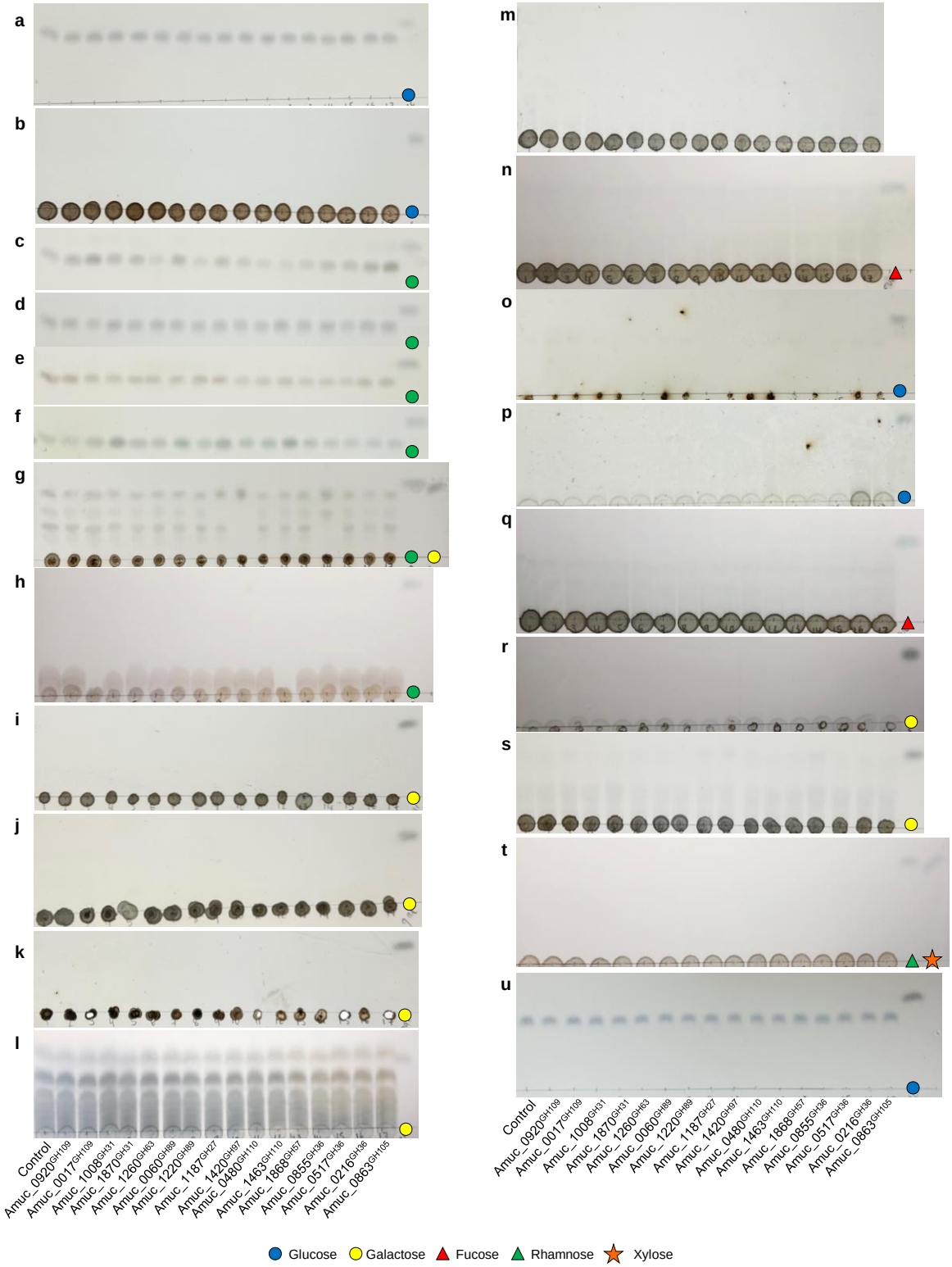
Supplementary Figure 13 | Activity of GH enzymes from *A. muciniphila* BAA-835 against defined oligosacchairdes capped with α-GalNAc or α-GlcNAc. a, GalNAcα1,3Gal. **b**, GlcNAcα1,4Gal. **c**, Blood group A II. **d**, Tn antigen. **e**, Forssman antigen. Standards have also been included on the TLCs on the right. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



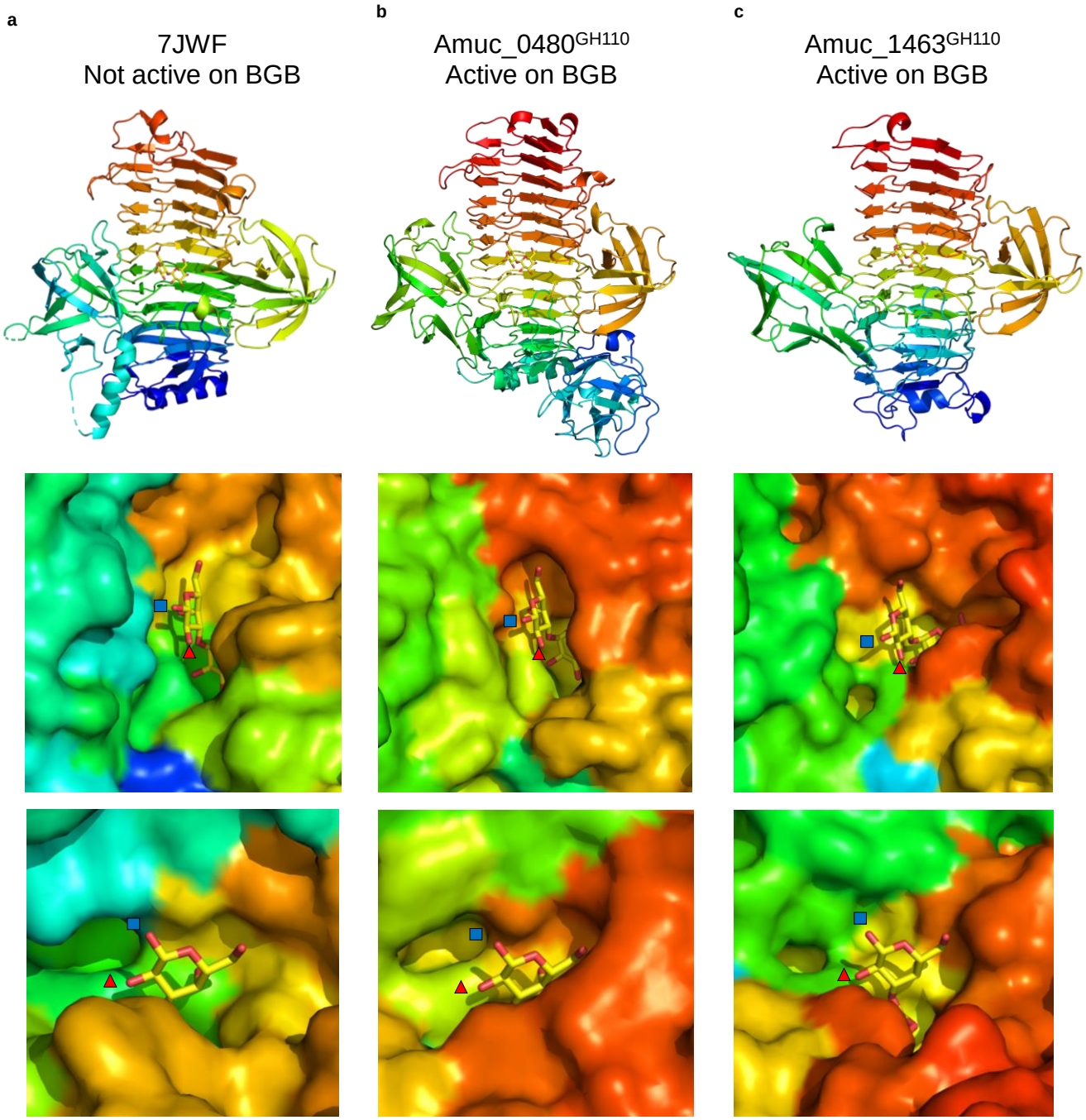
Supplementary Figure 14 | Activity of GH enzymes from *A. muciniphila* BAA-835 against defined oligosacchairdes capped with α-GalNAc or α-GlcNAc. a, Galα1,3GalNAc. **b**, Galα1,3Gal. **c**, globotriose. **d**, P1 antigen. **e**, Blood group B I. **f**, Blood group B II. Standards have also been included on the TLCs on the right. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



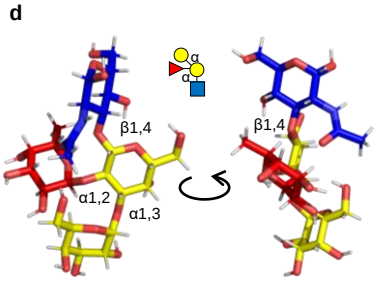
Supplementary Figure 15│Activity of GH enzymes from *A. muciniphila* BAA-835 against raffinose-type defined oligosacchairdes. a, melibiose. **b**, raffinose. **c**, stachyose. Standards have also been included on the TLCs on the right. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.

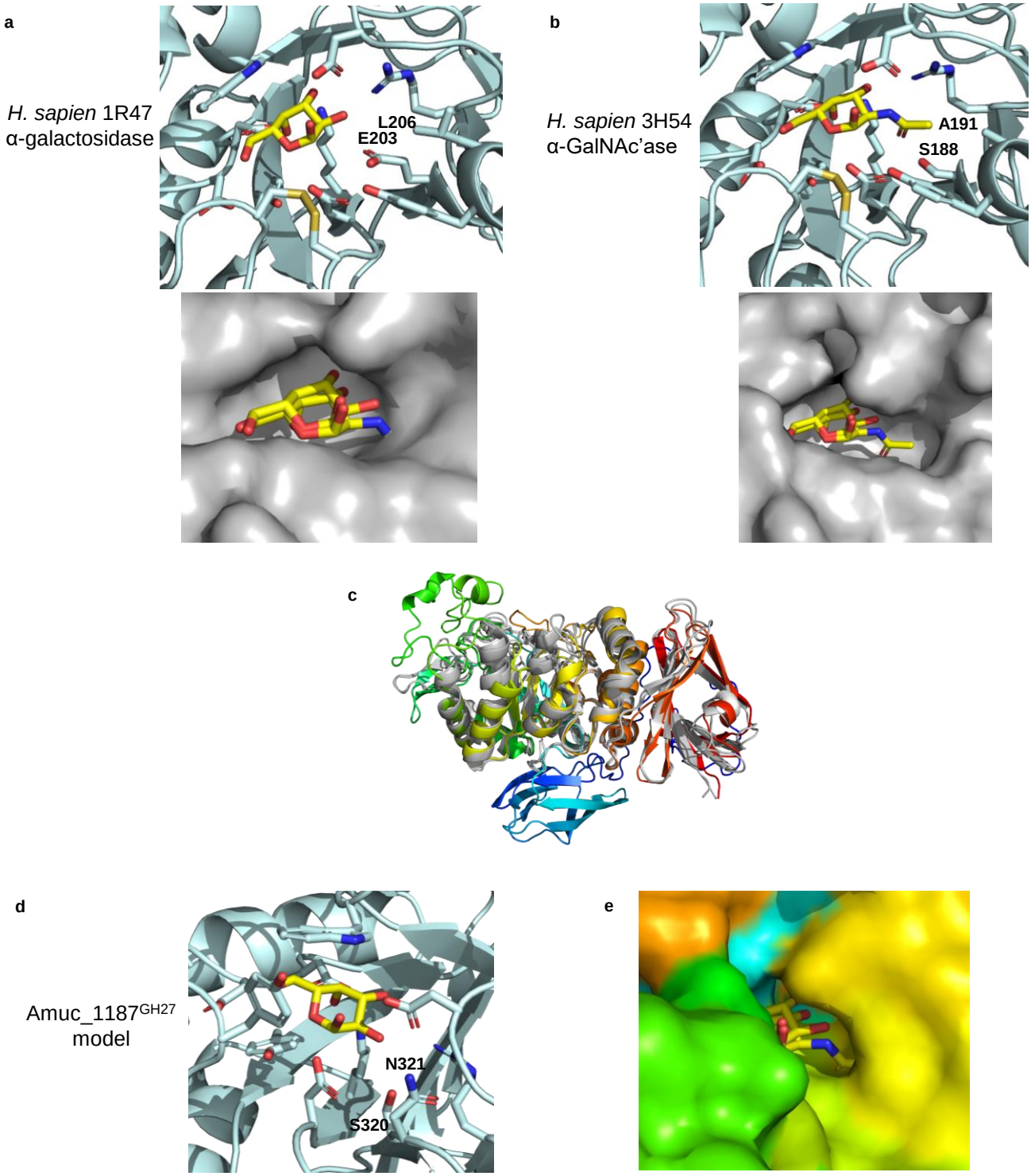


Supplementary Figure 16 | Activity of GH enzymes from *A. muciniphila* BAA-835 against a wide variety of substrates. a, maltose. **b**, dextran. **c**, α1,2-mannobiose. **d**, α1,3-mannobiose **e**, α1,4-mannobiose **f**, α1,6-mannobiose. **g,** galactomannan. **h,** RNaseB (high-mannose N-glycans). **i,** L-carrageenan. **j,** K-carrageenan. **k,** I-carrageenan. **l,** porphyran (L). **m,** ulvan. **n,** fucan. o**,** starch. **p,** laminarin. **q,** fucoidan. **r,** agarose. **s,** porphyrin (W). **t,** ulvan (entero). **u,** isomaltose. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.

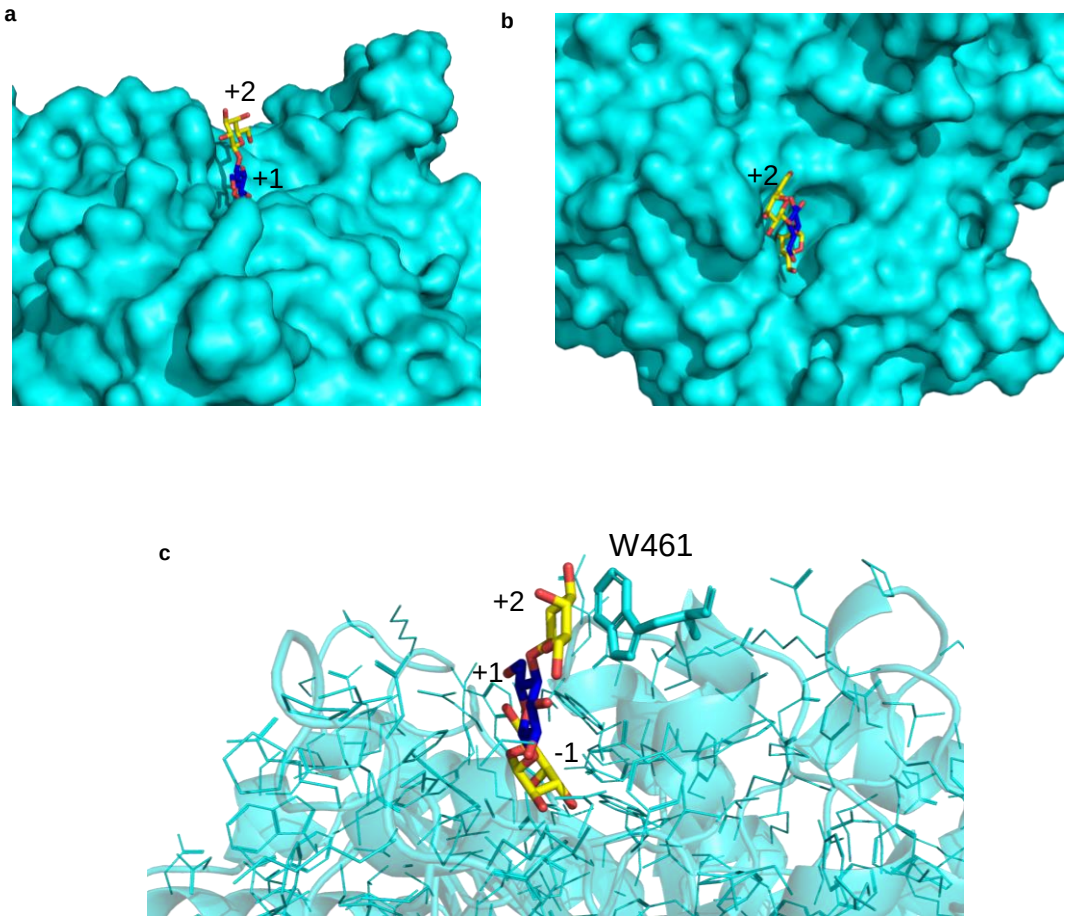


Supplementary Figure 17 | Structural analysis of the *Pseudoalteromonase distincta* GH110 structure with models of Amuc_0480^{GH110} and Amuc_1463^{GH110}. a. The structure of the GH110 from *Pseudoalteromonase distincta* is a right-handed parallel β-helix of 11 complete turns with two small β-barrel domains. All three domains contribute to the active site and this was crystallised with Galα1,3Gal Yellow sticks) in the active site. **b.** Model of Amuc_0480^{GH110} **c.** Model of Amuc_1463^{GH110}. All three are coloured rainbow blue-red from the N- to C-terminus, respectively. Top panels: cartoons to show overall structures. Middle and bottom panels show a surface representation of structure and models with which positions the fucose and GlcNAc would be in for a full BGB structure. The two enzymes from *A. muciniphila* do appear to have much more open active sites to accommodate the BGB glycan, but any observations cannot be extrapolated very far as these are only models. **d.** Structural models of BGB II tetrasaccharide made with GLYCAM carbohydrate builder to provide an insight into dimensions of the full BG structure.

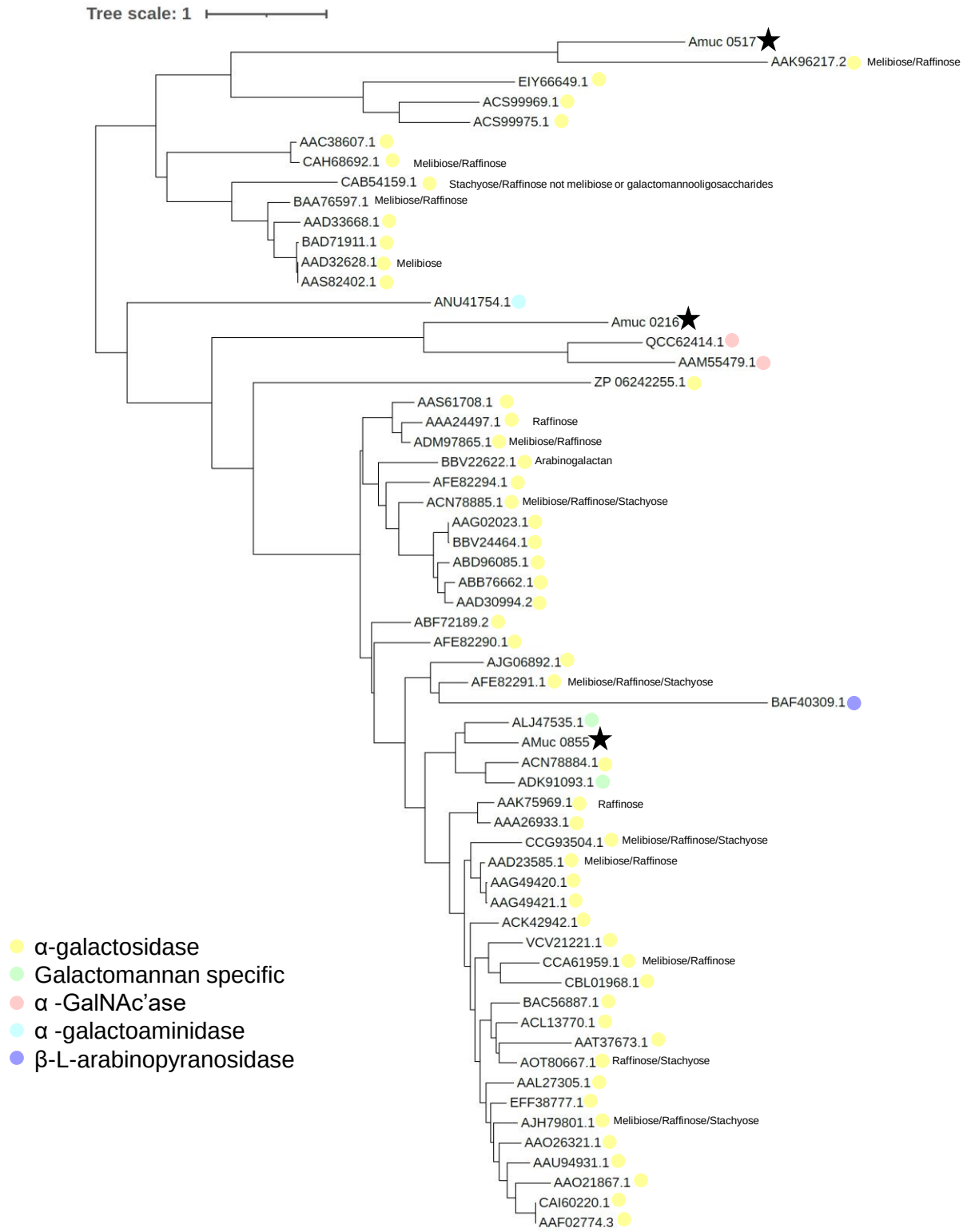




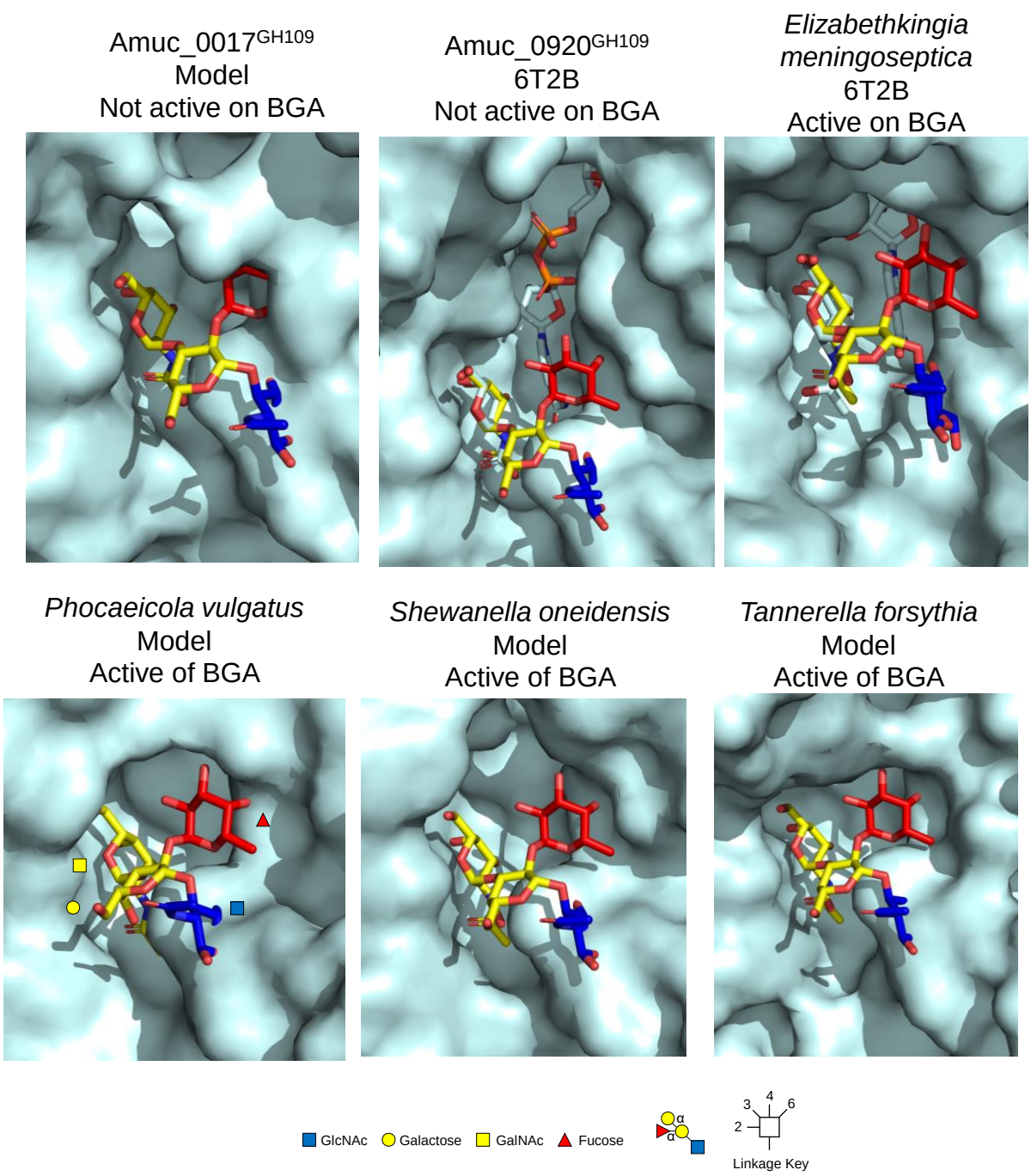
Supplementary Figure 18│Models of the GH27 family member from *A. muciniphila* ATCC BAA-835. a. Human GH27 α-galactosidase with galactose in the -1 subsite and **b.** Human GH27 α-GalNAc'ase with GalNAc in the -1 subsite. The residues coordinating these -1 sugars differ only in the residues that have been labelled. The larger residues coordinate the galactose in 1R47 and the smaller residues allow the accommodation of GalNAc in 3H54. **c.** An overlay of Amuc_1187^{GH27} (rainbow, N- to C-termini are blue-red, respectively) with the two *H. sapien* enzymes (grey) show that the overall fold is largely similar. **d.** A model of Amuc_1187^{GH27} with the ligand from 1R47 and the two equivalent residues. These are relatively large residues, so the enzyme has specificity for galactose rather than GalNAc. **e.** A surface representation of the model of Amuc_1187^{GH27} to show the pocket for galactose more clearly.



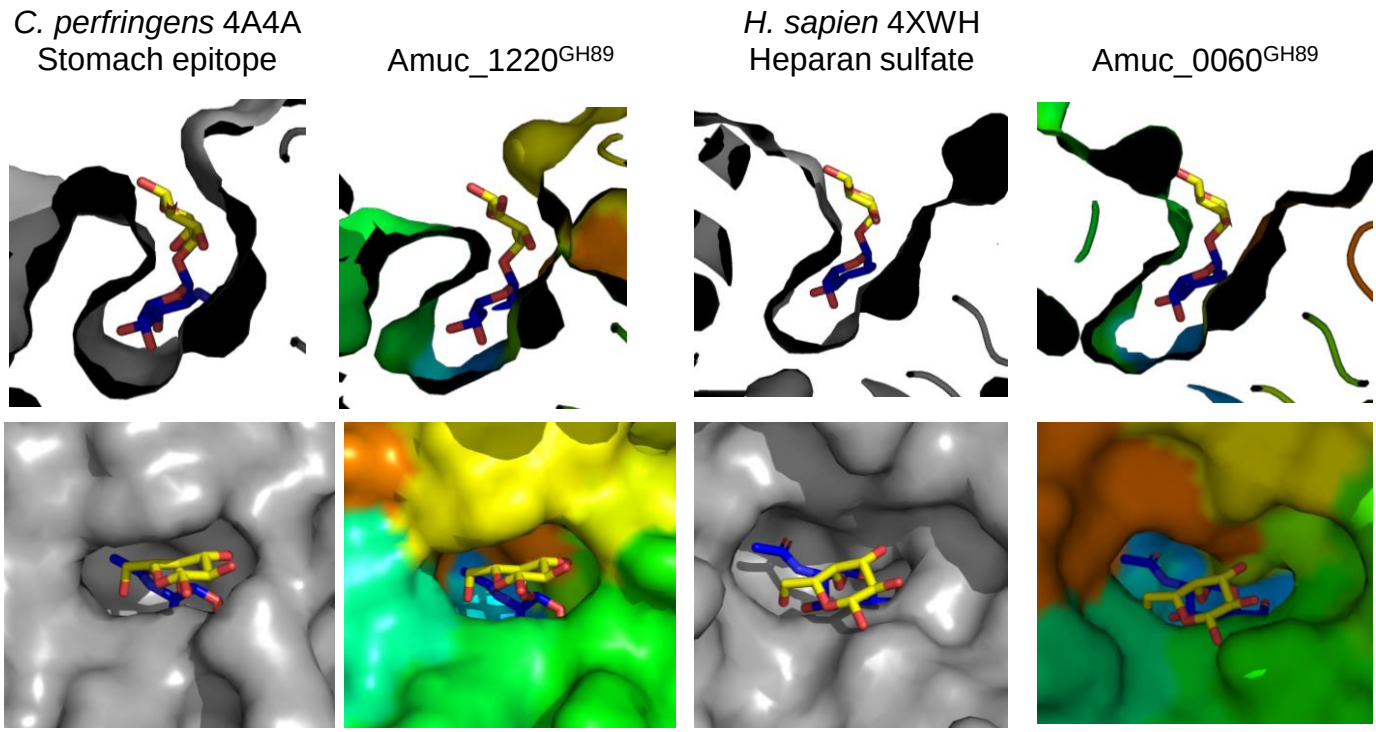
Supplementary Figure 19 | The structural model of the Amuc_1420^{GH97}. a, The active site of the model is viewed in surface mode from the side. Overlaid into the active site is a trisaccharide (lactose that also has an α-linked galactose to the Glc through a 1,1-linkage) from the *Bacteroide thetaiotaomicron* BT1871 structure. b, same structure looking down into the active site. c, a close up of the active site, showing the secondary structure as cartoon, the amino acid side chains as lines, apart from one tryptophan that may interact with the sugar in the +2 subsite.



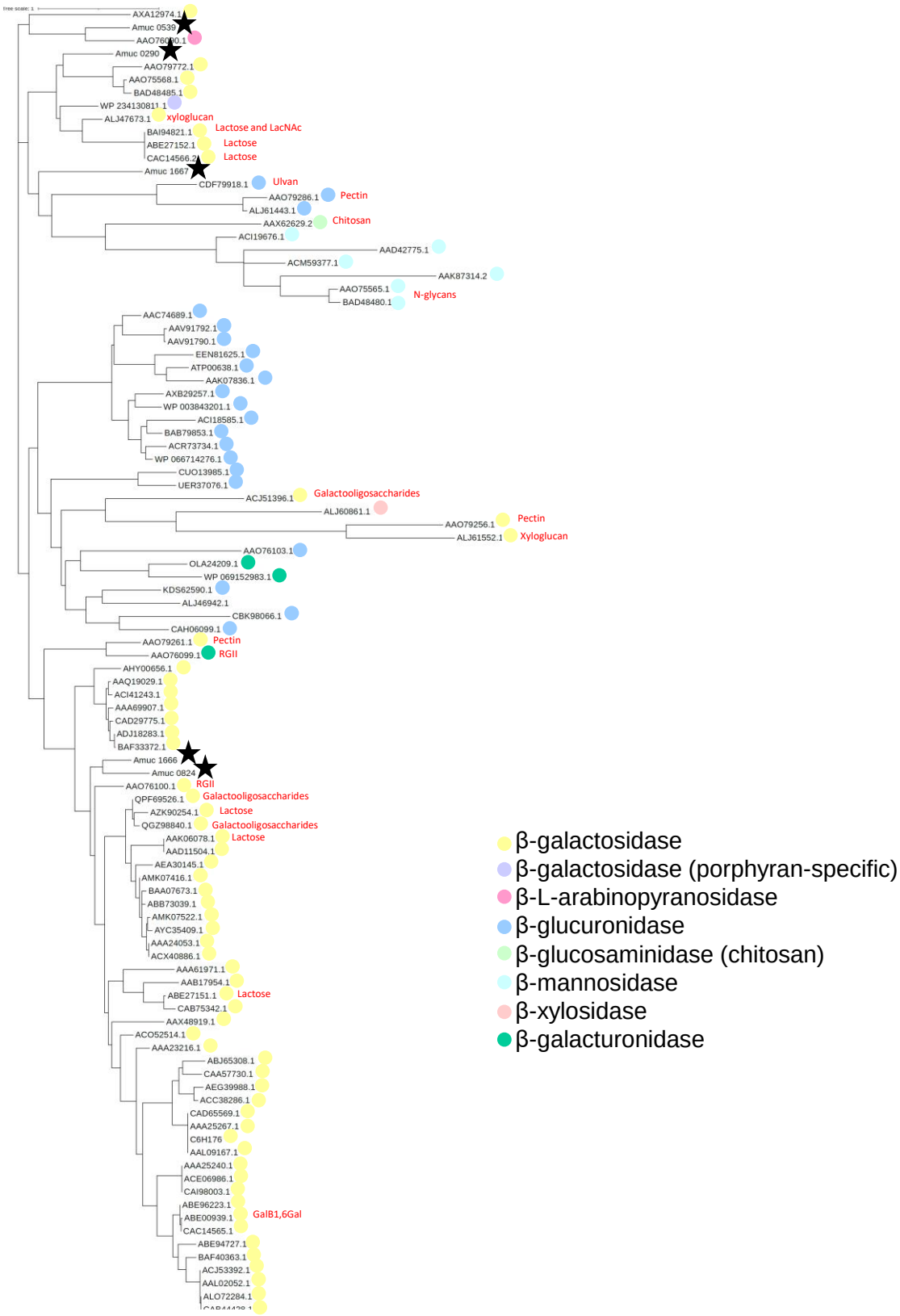
Supplementary Figure 20 | Phylogenetic tree of characterised GH36 family members with those from *A. muciniphila* ATCC BAA-835. The sequences of the GH36 family members with reported activities (CAZy database) and the ones from *A. muciniphila* ATCC BAA-835 were compared as described in the methods. The different specificities are indicated by different colours and the *A. muciniphila* ATCC BAA-835 enzymes are highlighted by black stars. The enzymes are represented by their accession numbers of locus tags.



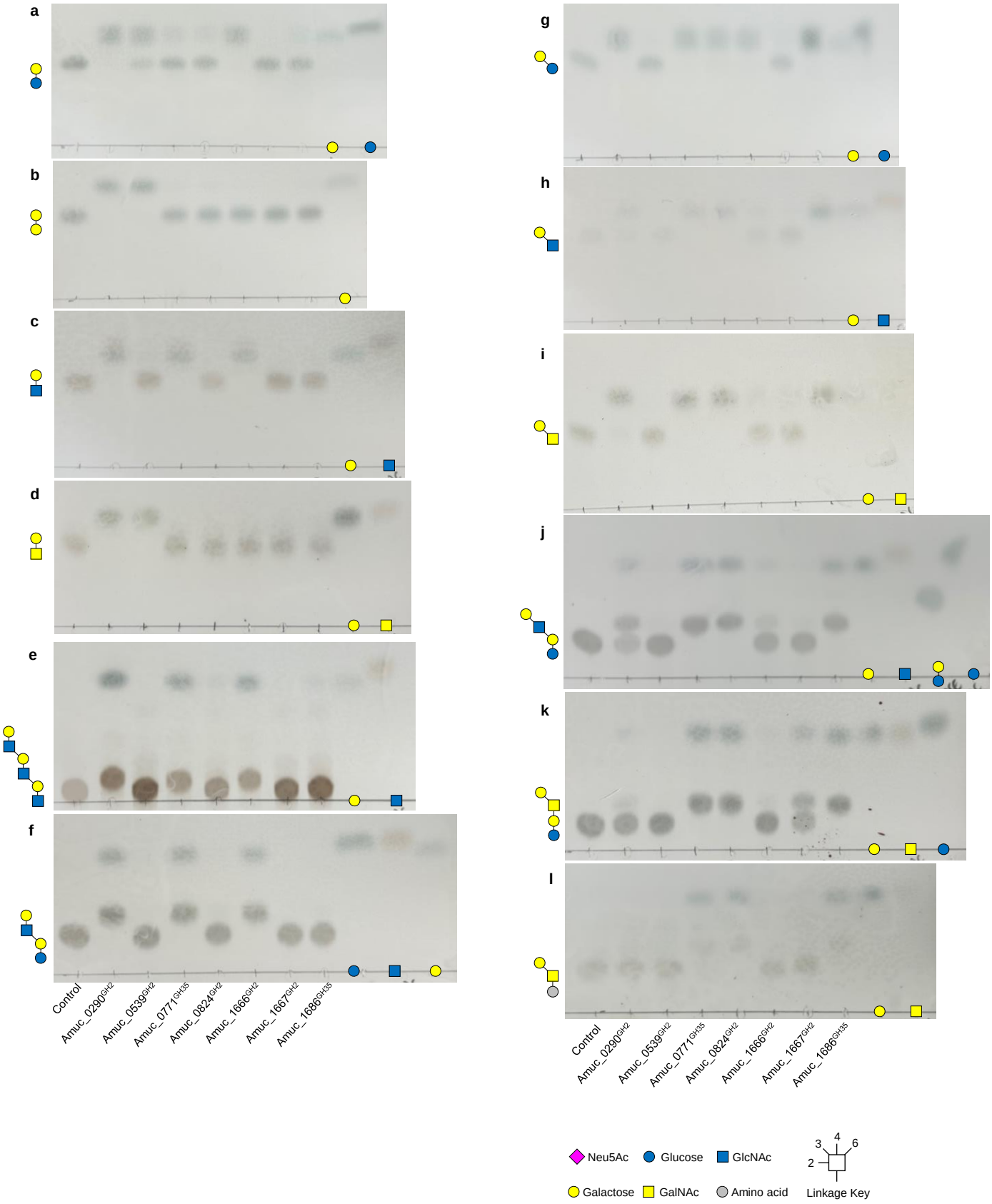
Supplementary Figure 21| Structural analysis of available GH109 structures with models of Amuc_0017^{GH109} and previously characterised GH109 enzymes. For the two structures, the GalNAc and NAD+ are in cyan and orange. BGA tetrasaccharide from GH98 4D6H was overlaid into the active sites, with the monosaccharides coloured according to the key.



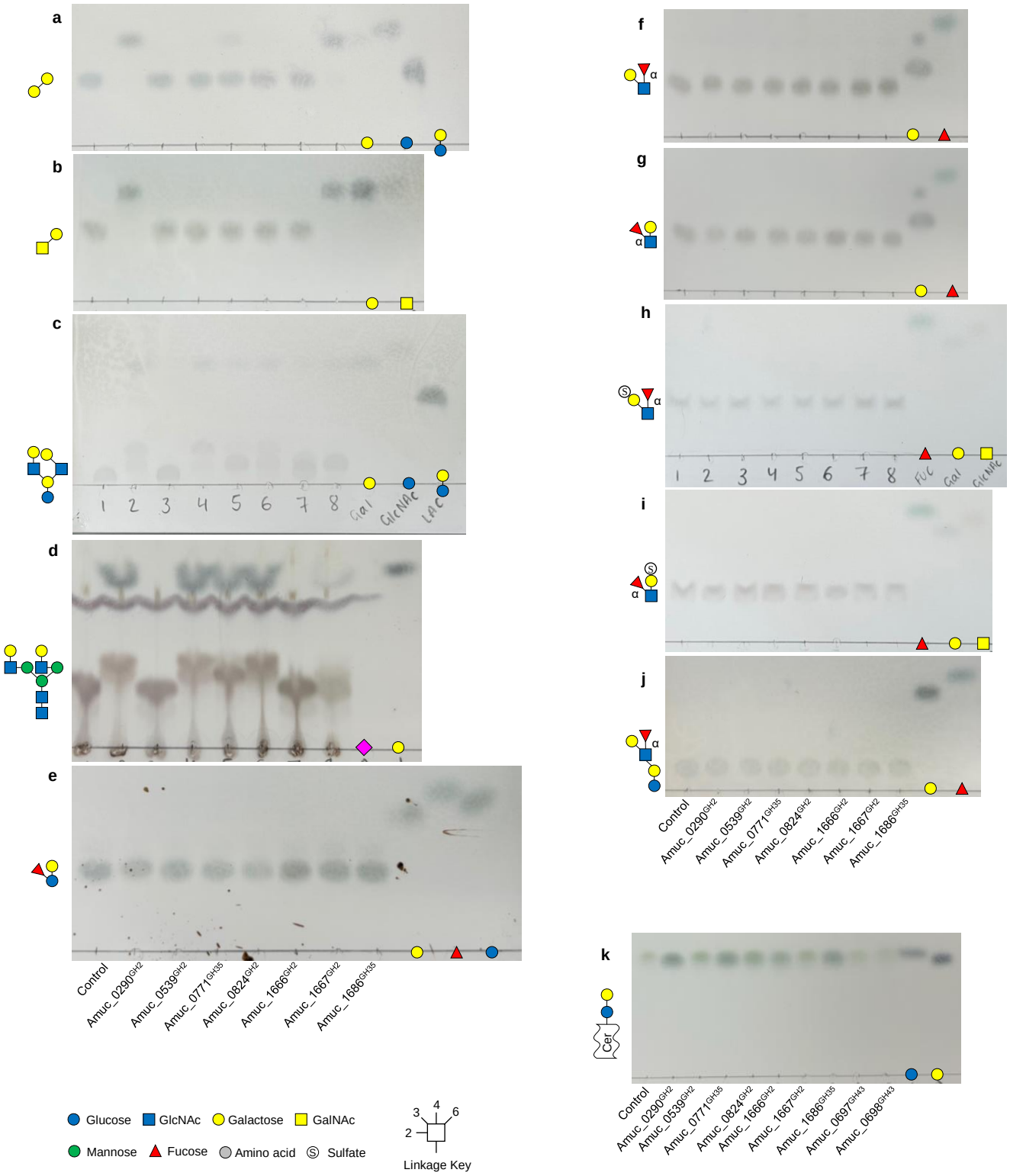
Supplementary Figure 22 | Structural insights into GH89 enzymes from *A. muciniphila* ATCC BAA-835. Two crystal structures (grey) with different substrate preferences from the GH89 family were compared to models of the two GH89 enzymes analysed in this work (rainbow). The *C. perfringens* 4A4A structure was crystallised with GlcNAcα1,4Gal in its active site and this was overlaid into the structure from *H. sapien* 4XWH and the two models. GlcNAc and Gal are blue and yellow, respectively. The top panels are slices through the active site to show how this disaccharide sits in there and the bottom panels are looking down at the entrance of the active site. The active site of the *C. perfringens* 4A4A structure packs tightly around the sugars in the -1 and +1 subsites, which has a considerable curve in its structure. The conformation of the Amuc_1220^{GH89} active site pocket is predicted to be very similar to this and accommodate this disaccharide well. The *H. sapien* 4XWH and Amuc_0060^{GH89}model do not look like they can accommodate the angle of this substrate very well.



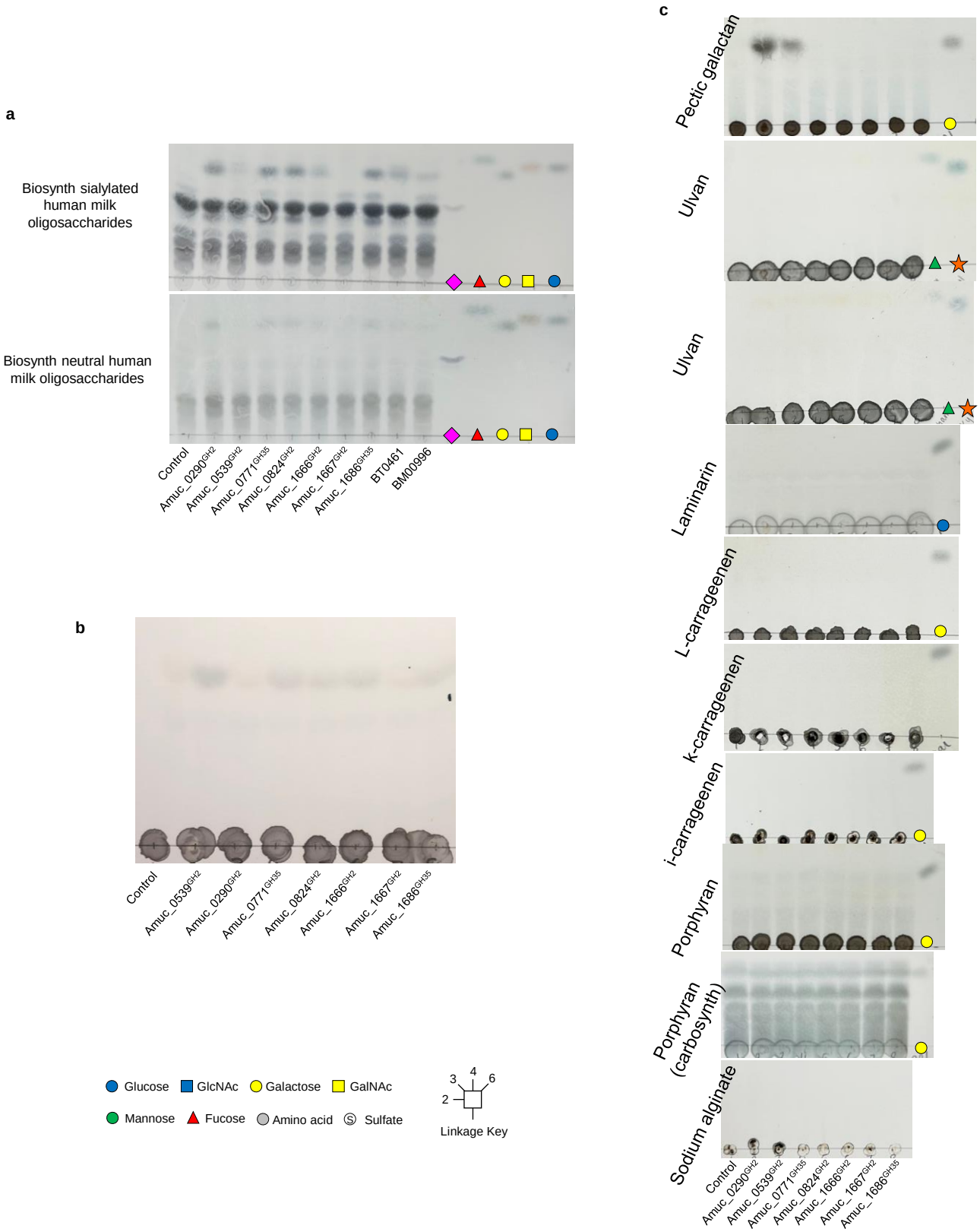
Supplementary Figure 23 | Phylogenetic tree of characterised GH2 family members with those from *A. muciniphila* ATCC BAA-835. The sequences of the GH2 family members with reported activities (CAZy database) and the ones from *A. muciniphila* ATCC BAA-835 were compared as described in the methods. The different specificities are indicated by different colours and the *A. muciniphila* ATCC BAA-835 enzymes are highlighted by black stars. Some more specific information about activity is supplied were possible. The enzymes are represented by their accession numbers of locus tags. Different specificities do cluster in this analysis and the observed activity of the *A. muciniphila* ATCC BAA-835 enzymes correlate to where they cluster on the tree.



Supplementary Figure 24 | Activity of GH2 and GH35 enzymes from *A. muciniphila* BAA-835 against a wide variety of substrates. a, lactose. **b**, Galβ1,4Gal. **c**, LacNAc. **d**, Galβ1,4GalNAc. **e**, TriLacNAc. **f,** Lacto-N-*neo*tetraose. **g,** Galβ1,3Glc. **h,** LNB. **i,** Galβ1,3GalNAc. **j,** Lacto-N-tetrasoe. **k,** GA1. **l,** Tn antigen. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



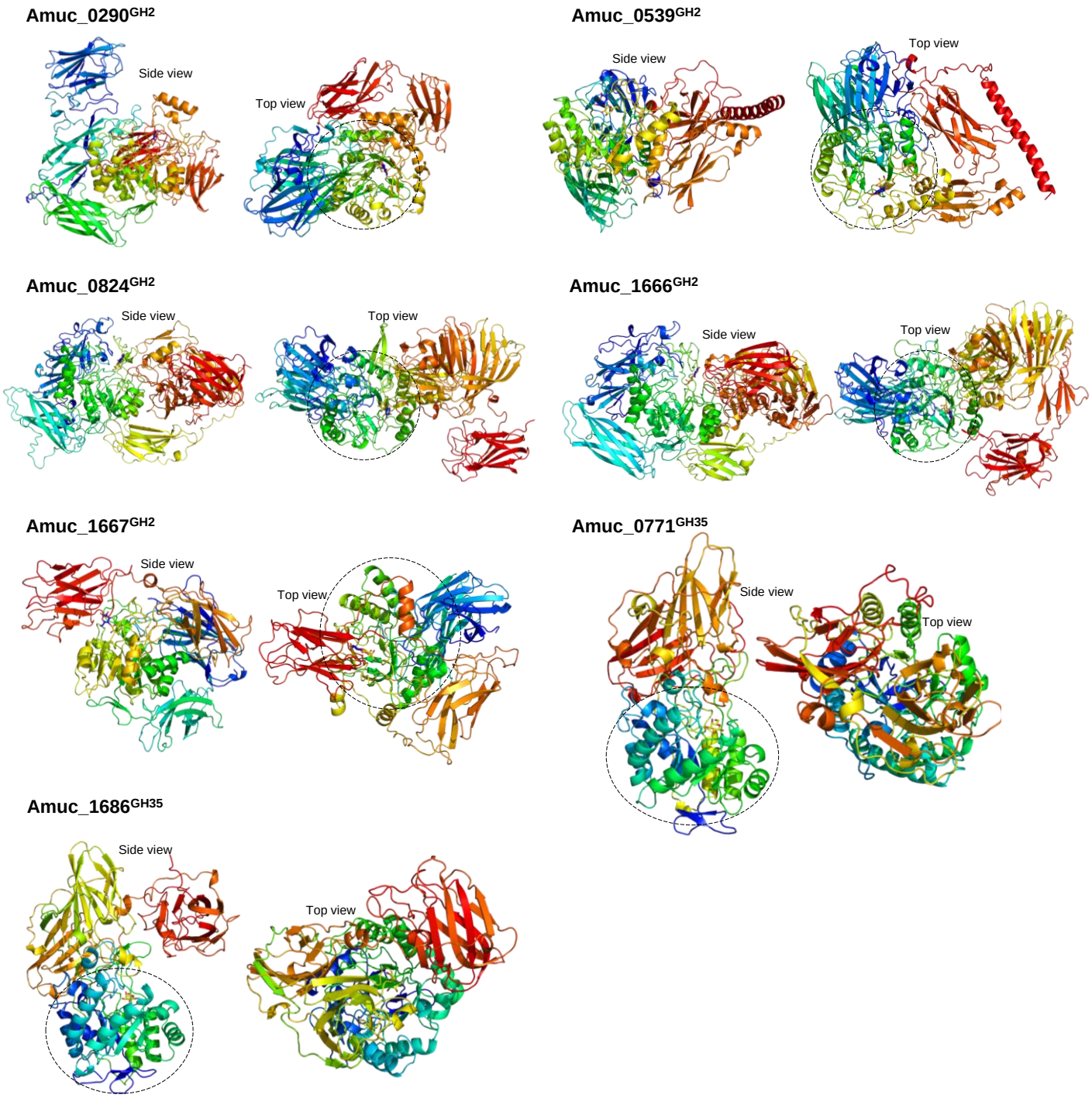
Supplementary Figure 25 | Activity of GH2 and GH35 enzymes from *A. muciniphila* BAA-835 against a wide variety of substrates. a, Galβ1,6Gal. **b**, Galβ1,6GalNAc. **c**, Lacto-N-*neo*hexaose. **d**, biantennary complex N-glycans. **e**, 2'FL. **f**, Lewis A. **g,** Lewis X. **h,** 3S Lewis A. **i,** 3S Lewis X. **j,** LNFP II. **k,** Lactosylceramide. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM (or 10mg/ml for glycoproteins), pH 7, 37 °C, overnight, and with 1 μM enzyme.



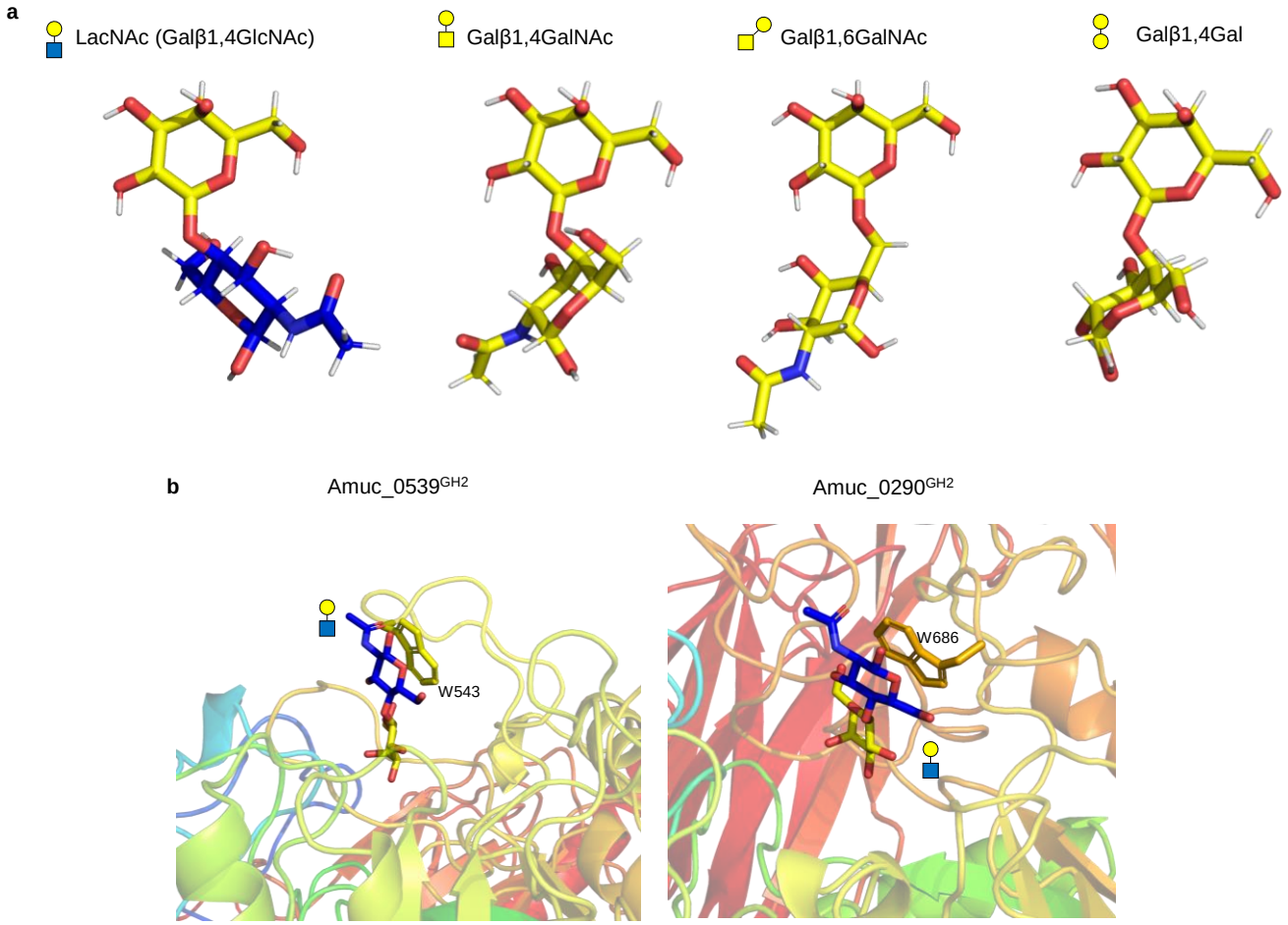
Supplementary Figure 26 | Activity of GH2 and GH35 enzymes from *A. muciniphila* BAA-835 against a wide variety of substrates. a, Human milk oligosaccharides. **b**, PGM III. **c**, Range of plant polysaccharides**,** Lactosylceramide. Standards have also been included on the TLCs.

Supplementary Figure 27 | The GH43 enzymes of *A. muciniphila* BAA-835. a, Testing activity against a variety of defined substrates. Standards have also been included on the TLCs on the right. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme. **b,** Sequence alignment (left) showing that the sequences are conserved apart from a section in the middle (purple dashed box). Right: Structural models of the two GH43 enzymes superimposed with a galactose in the active site (from 6EUI, which is BT3683 from *Bacteroides thetaiotaomicron*). The only difference between the models is one of the loops on the side of the active site and corresponds to the sequence difference seen with the alignment purple dashed box).

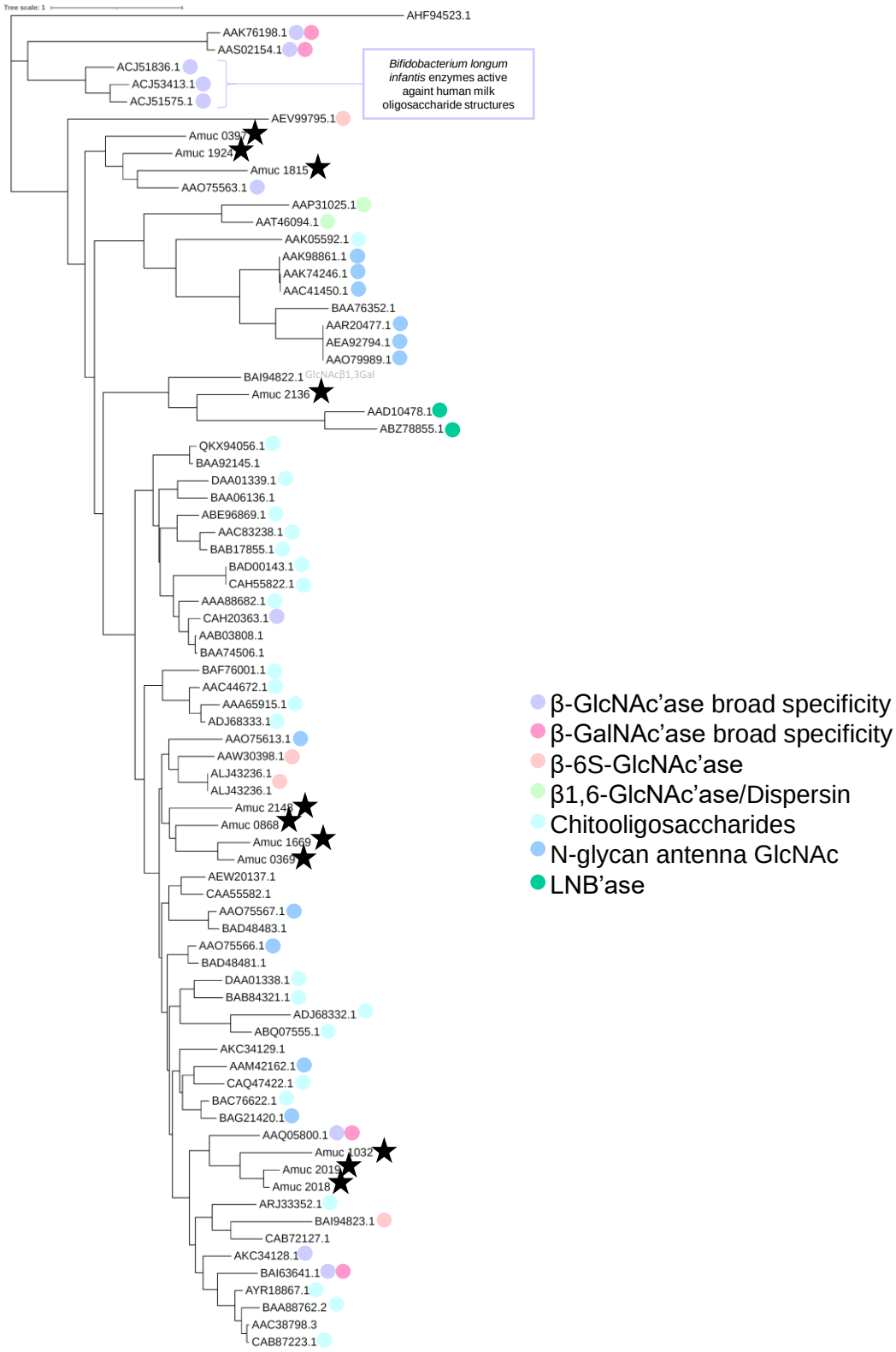
Supplementary Figure 28 | pNP assays with some of the GH enzymes from *A. muciniphila*. If the enzyme is active the reaction turns yellow.



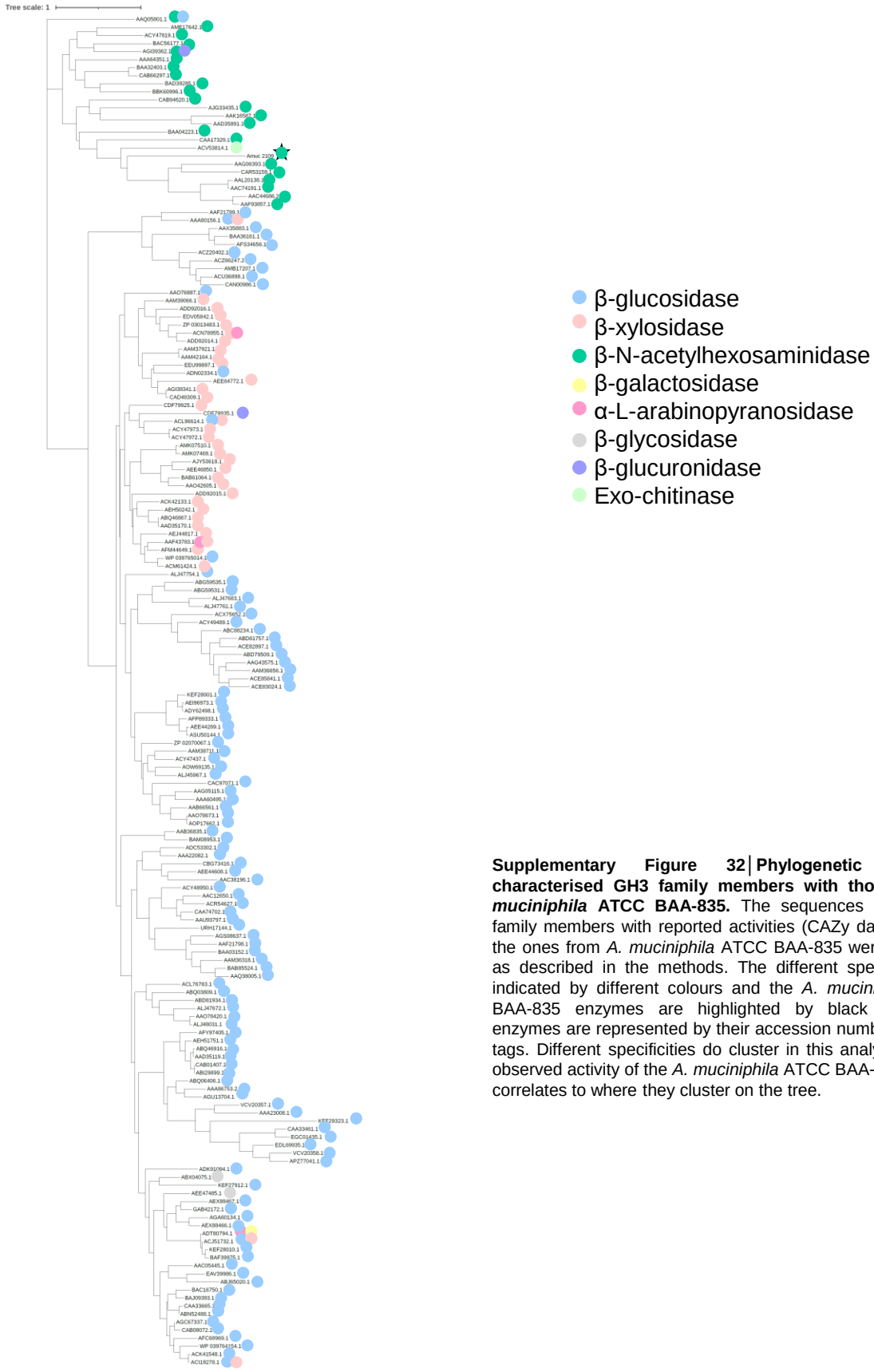
Supplementary Figure 29│Models of the GH2 and GH35 enzymes from *A. muciniphila* BAA-835. Models were built using AlphaFold2 Colab as described in the materials and methods. Lactose from 1JYN was overlaid into the active site of the GH2 enzymes and galactose from 4E8C was overlaid into the active sites of the GH35 enzymes and are shown as sticks. The enzymes are coloured blue to red from the N- to the C-termini, respectively. The catalytic modules are circled with a dotted line to highlight the variety of different accessory modules.



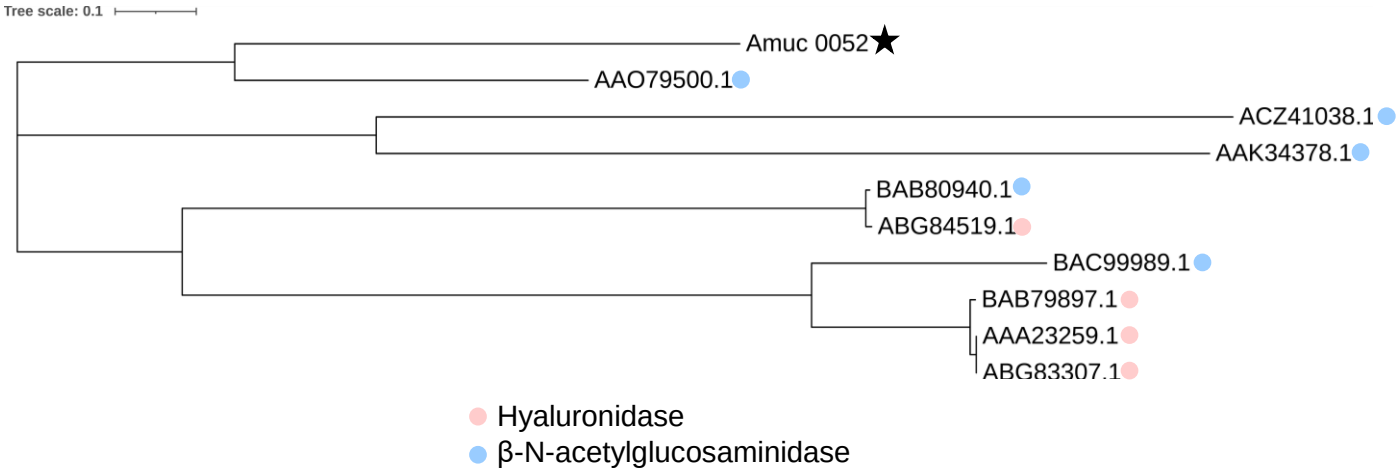
Supplementary Figure 30 | Analysis of models of the GH2 enzymes from *A. muciniphila* BAA-835. a, Models of different disaccharides. **b,** Models of two GH2 enzymes with LacNAc overlaid into the active site (from a crystal structure of a *Streptococcus pneumoniae* GH2, 4CUC). Aromatic residues contact the +1 sugar to likely play a role in selectivity and substrate binding.



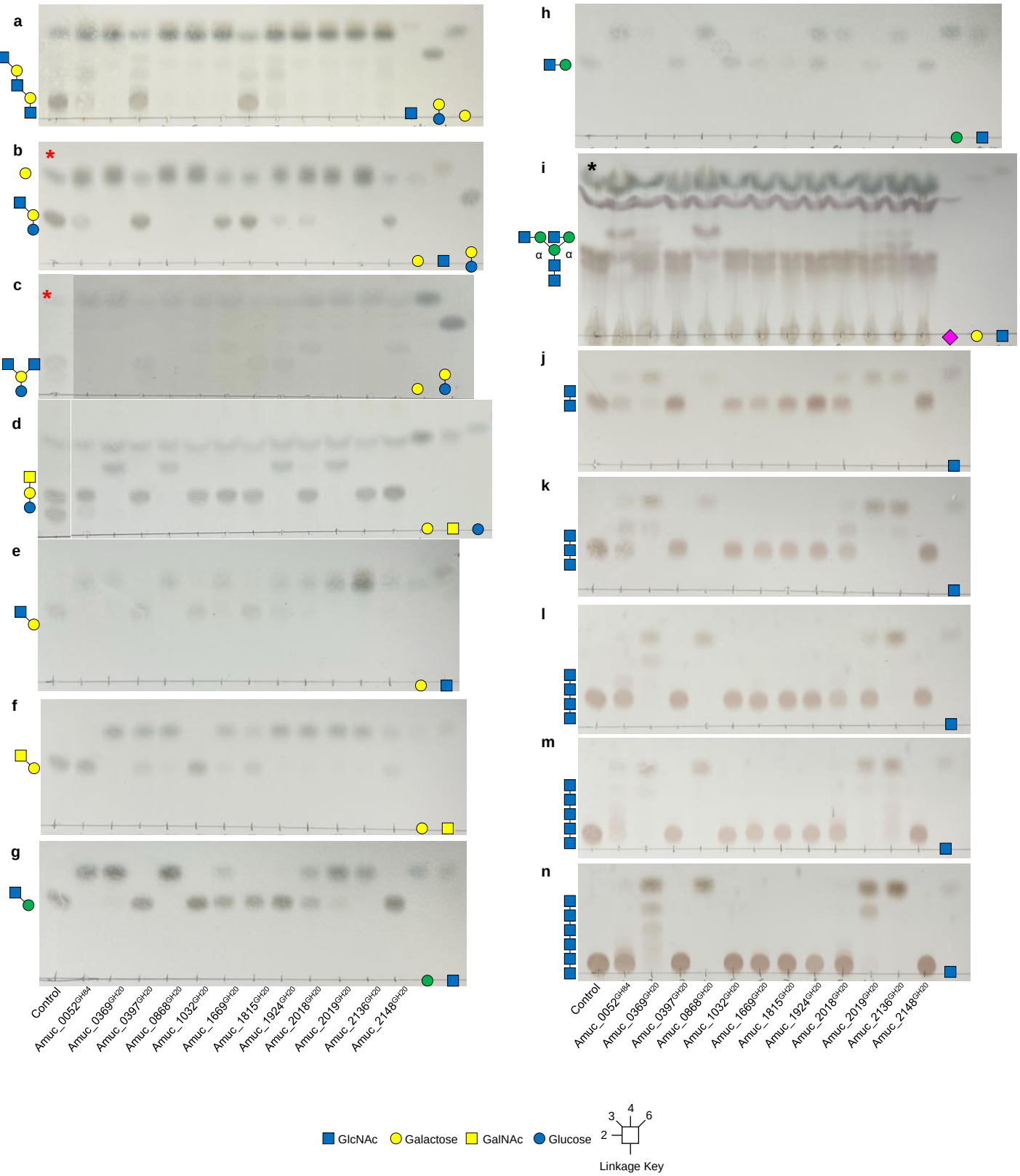
Supplementary Figure 31 | Phylogenetic tree of characterised GH20 family members with those from *A. muciniphila* ATCC BAA-835. The sequences of the GH20 family members with reported activities (CAZy database) and the ones from *A. muciniphila* ATCC BAA-835 were compared as described in the methods. The different specificities are indicated by different colours and the *A. muciniphila* ATCC BAA-835 enzymes are highlighted by black stars. Some more specific information about activity is supplied were possible. The enzymes are represented by their accession numbers of locus tags. Different specificities do cluster in this analysis and the observed activity of the three *A. muciniphila* ATCC BAA-835 enzymes correlates to where they cluster on the tree.



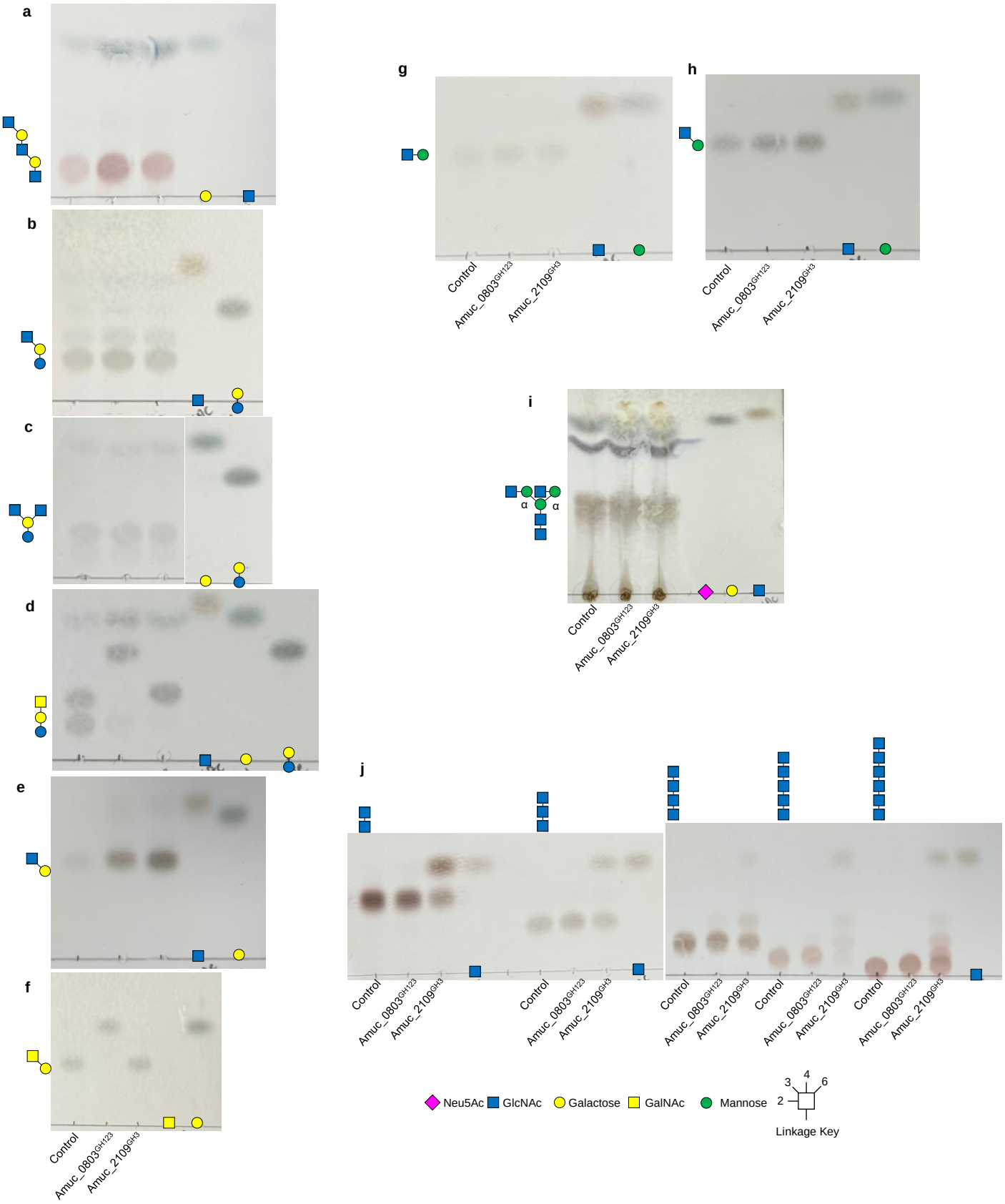
Supplementary Figure 32 | Phylogenetic tree of characterised GH3 family members with those from *A. muciniphila* ATCC BAA-835. The sequences of the GH3 family members with reported activities (CAZy database) and the ones from *A. muciniphila* ATCC BAA-835 were compared as described in the methods. The different specificities are indicated by different colours and the *A. muciniphila* ATCC BAA-835 enzymes are highlighted by black stars. The enzymes are represented by their accession numbers of locus tags. Different specificities do cluster in this analysis and the observed activity of the *A. muciniphila* ATCC BAA-835 enzyme correlates to where they cluster on the tree.



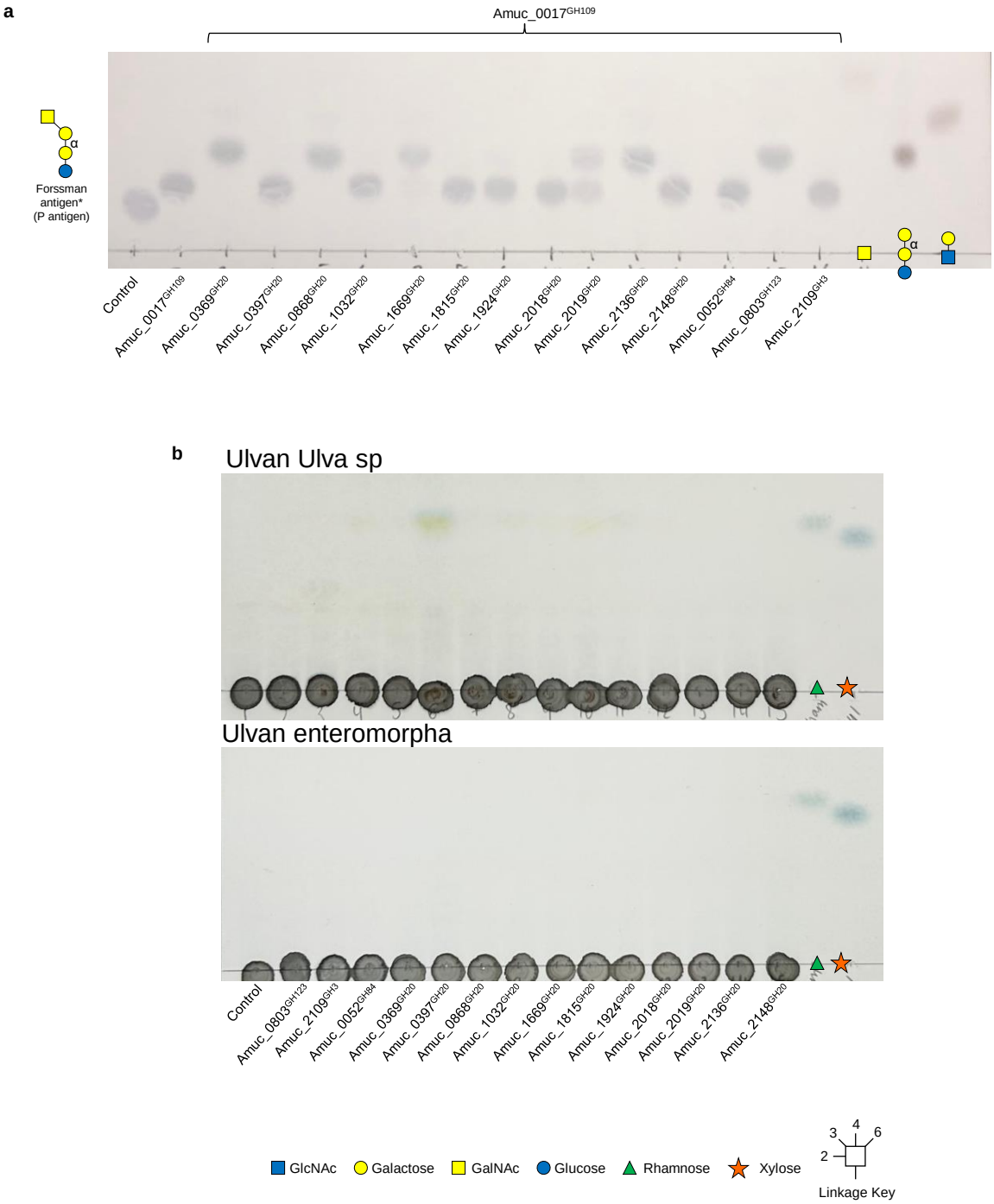
Supplementary Figure 33 | Phylogenetic tree of characterised GH84 family members with those from *A. muciniphila* ATCC BAA-835. The sequences of the GH84 family members with reported activities (CAZy database) and the ones from *A. muciniphila* ATCC BAA-835 were compared as described in the methods. The different specificities are indicated by different colours and the *A. muciniphila* ATCC BAA-835 enzymes are highlighted by black stars. The enzymes are represented by their accession numbers of locus tags. Different specificities do cluster in this analysis and the observed activity of the *A. muciniphila* ATCC BAA-835 enzyme correlates to where they cluster on the tree although there are not many enzymes characterised.



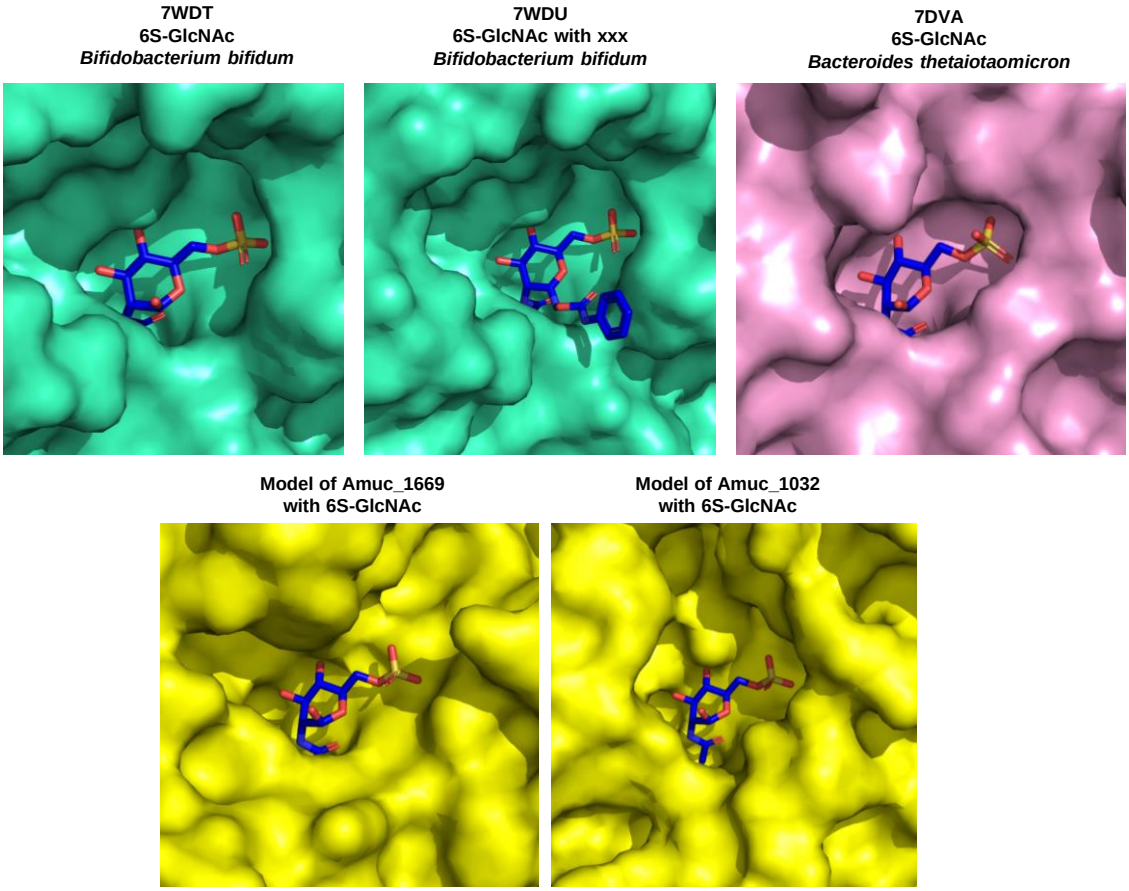
Supplementary Figure 34 | Activity of the GH20 and GH84 family members from *A. muciniphila* BAA-835 against defined oligosacchairdes. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzymes. **a,** TriLacNAc. **b**, Lacto-N-neotetraose. **c**, Lacto-N-neohexaose. **d**, Galβ1,3GalNAcβ1,4Galβ1,4Glc. **e**, GlcNAcβ1,3Gal. **f**, GalNAcβ1,3Gal. **g**, GlcNAcβ1,3Man. **h**, GlcNAcβ1,2Man. **i**, $\alpha_1$acidglycoprotein. **j**, Chitobiose. **k**, Chitotriose. **l**, Chitotetraose. **m**, Chitopentaose. **n**, Chitohexaose. Some pre-treatments were required to generate the substrates: Red asterisk - pre-treated with BT0461^{GH2} to remove the non-reducing end β1,4galactose, cyan asterisk - pre-treated with B035DRAFT_00996^{GH2} to remove the non-reducing end β1,3galactose, black asterisk – pre-treated with BT0455^{GH33}, BT0461^{GH2}, and PNGaseL, which are a broad-acting sialidase, β1,4-galactosidase, and remove N-glycans from the protein, respectively. The glycan that the GH20 enzymes are tested on are shown to the right of each TLC and all linkages are beta. Standards have also been included on the left of the TLCs.



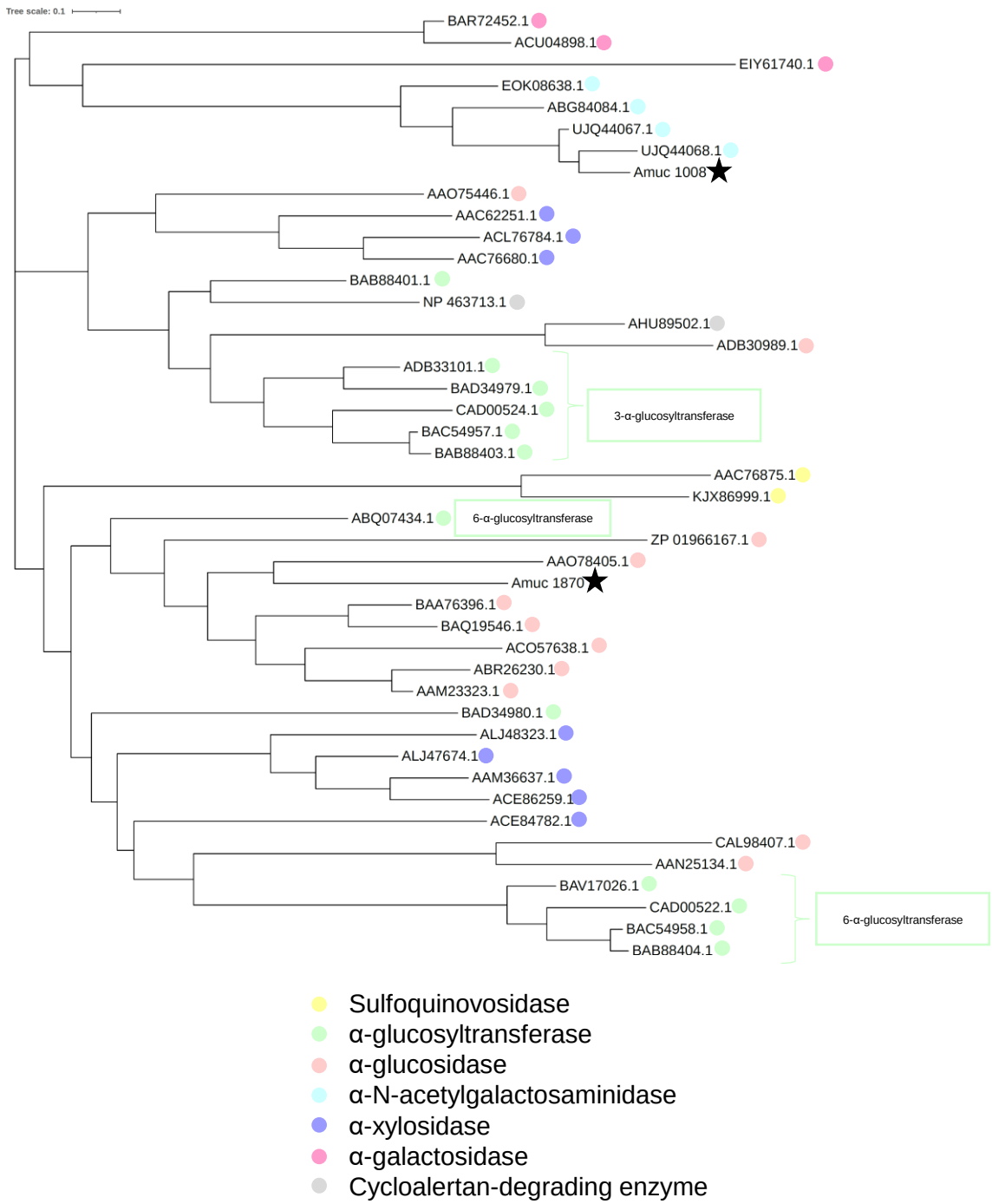
Supplementary Figure 35 | Activity of the GH123 and GH3 family members from *A. muciniphila* BAA-835 against defined oligosacchairdes. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzymes. **a,** TriLacNAc. **b**, Lacto-N-neotetraose. **c**, Lacto-N-neohexaose. **d**, Galβ1,3GalNAcβ1,4Galβ1,4Glc. **e**, GlcNAcβ1,3Gal. **f**, GalNAcβ1,3Gal. **g**, GlcNAcβ1,3Man. **h**, GlcNAcβ1,2Man. **i**, $\alpha_1$acidglycoprotein. **j**, Chitobiose-Chitohexaose, left to right. Some pre-treatments were required to generate the substrates: Red asterisk - pre-treated with BT0461^{GH2} to remove the non-reducing end β1,4galactose, cyan asterisk - pre-treated with B035DRAFT_00996^{GH2} to remove the non-reducing end β1,3galactose, black asterisk – pre-treated with BT0455^{GH33}, BT0461^{GH2}, and PNGaseL, which are a broad-acting sialidase, β1,4-galactosidase, and remove N-glycans from the protein, respectively. The glycan that the GH20 enzymes are tested on are shown to the right of each TLC and all linkages are beta. Standards have also been included on the left of the TLCs.



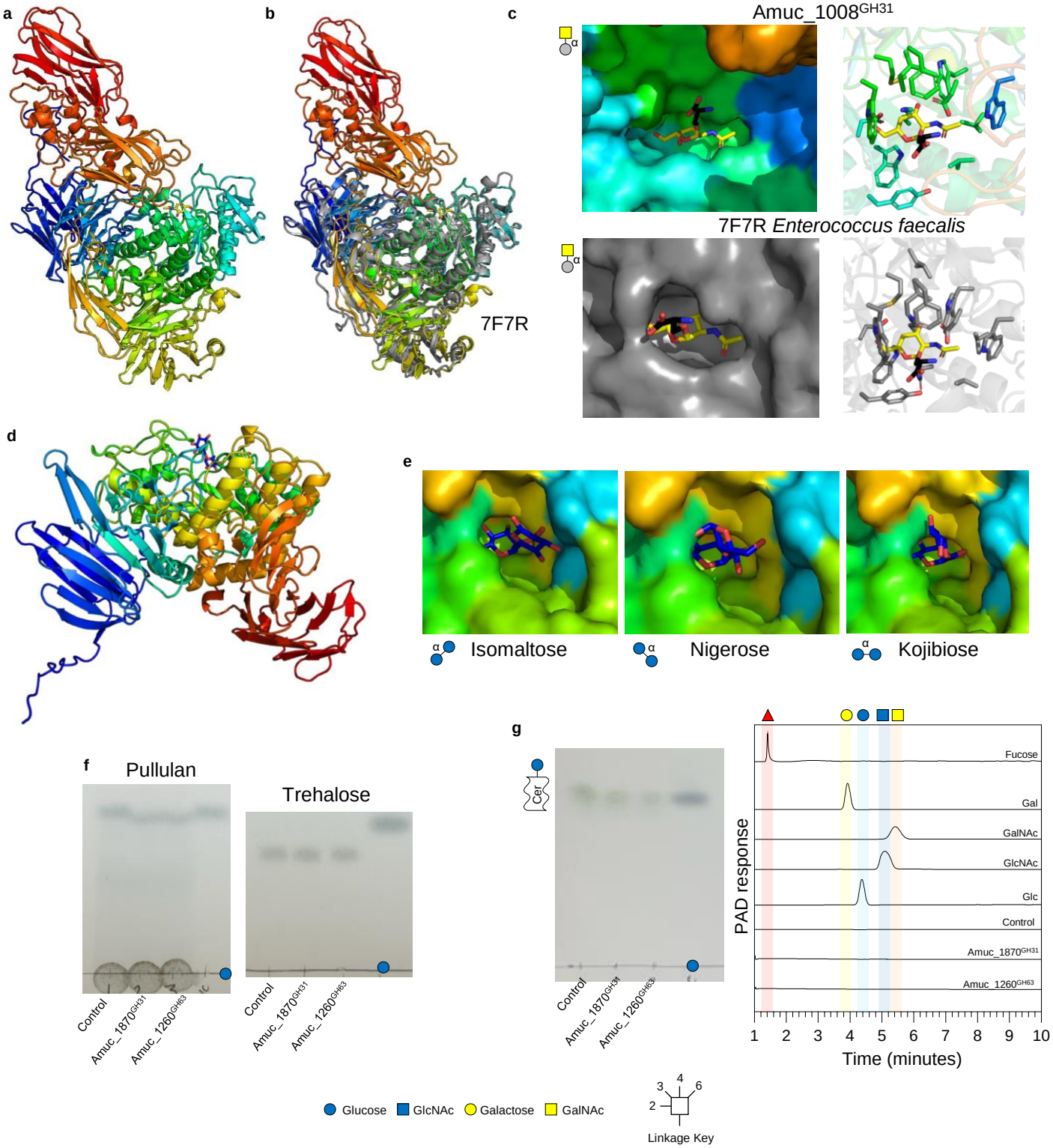
Supplementary Figure 36 | Activity of the β-HexNAc'ases from *A. muciniphila* BAA-835 against P antigen and ulvan substrates. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 μM enzymes. Standards have also been included on the left of the TLCs. The Forssman antigen was treated with Amuc_0017^{GH109}



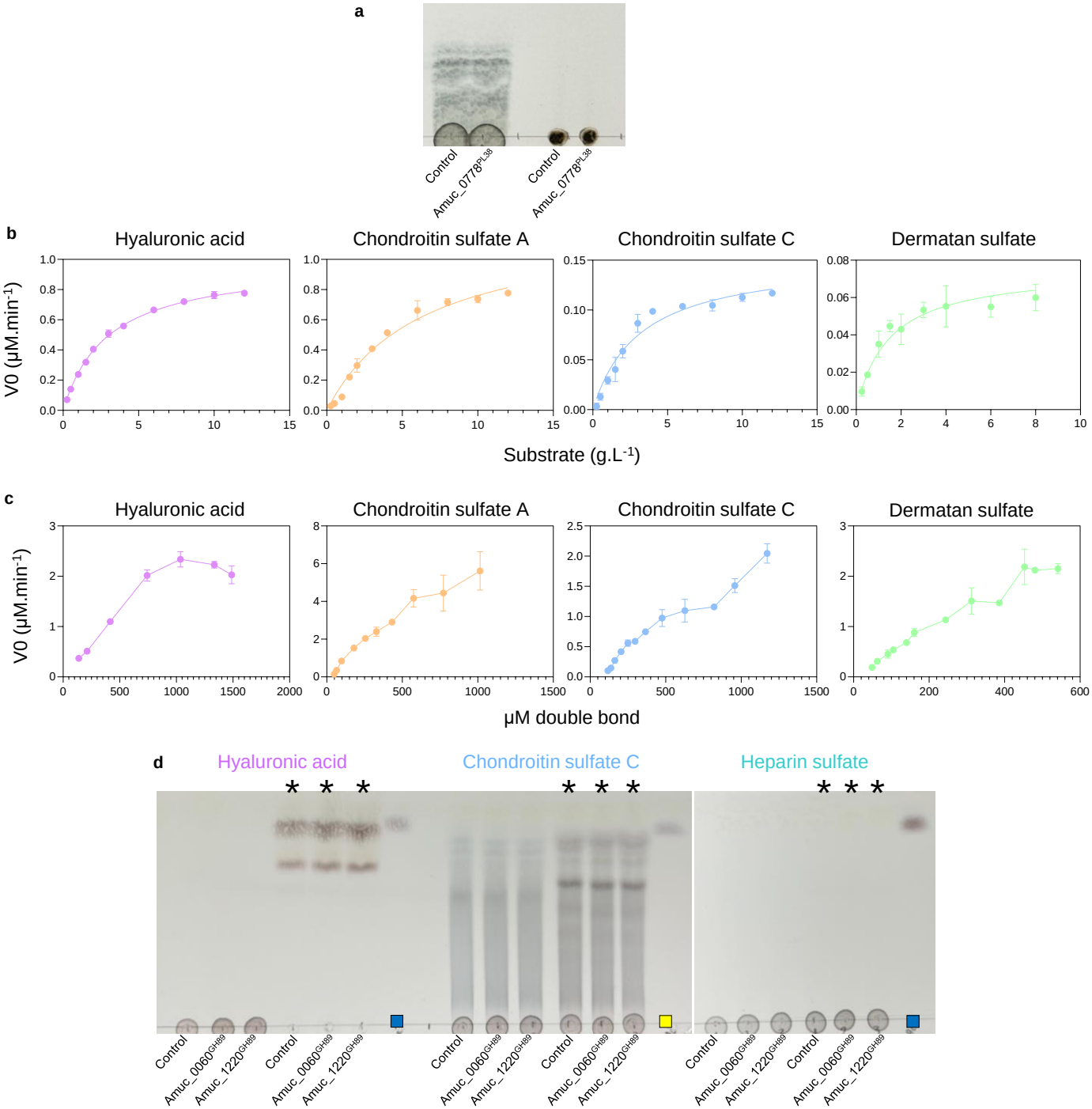
Supplementary Figure 37 | Structures of the active sites of GH20 active against 6S-GlcNAc. Three crystal structures of enzymes that can accommodate 6S-GlcNAc are shown along the top row. Two of the A. muciniphila GH20 Alphafold models were able to accommodate the sulfate when overlaid into the active site.



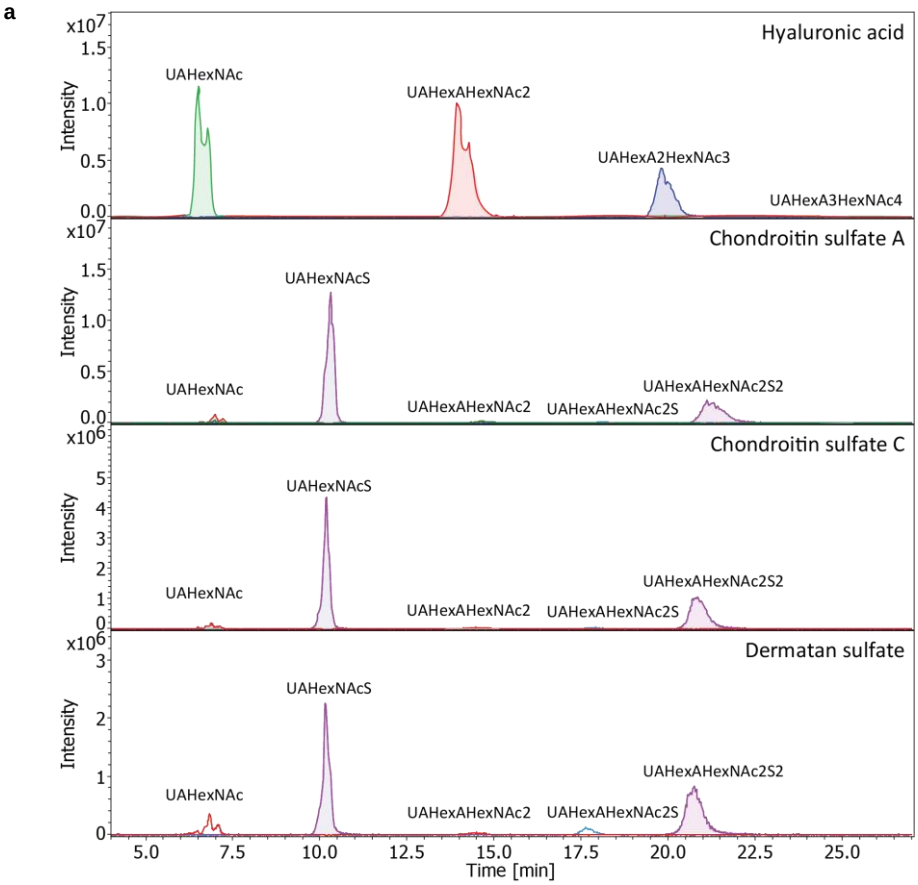
Supplementary Figure 38│Phylogenetic tree of characterised GH31 family members with those from *A. muciniphila* ATCC BAA-835. The sequences of the GH31 family members with reported activities (CAZy database) and the ones from *A. muciniphila* ATCC BAA-835 were compared as described in the methods. The different specificities are indicated by different colours and the *A. muciniphila* ATCC BAA-835 enzymes are highlighted by black stars. Some more specific information about activity is supplied were possible. The enzymes are represented by their accession numbers of locus tags. Different specificities do cluster in this analysis and the observed activity of the two *A. muciniphila* ATCC BAA-835 enzymes correlates to where they cluster on the tree.

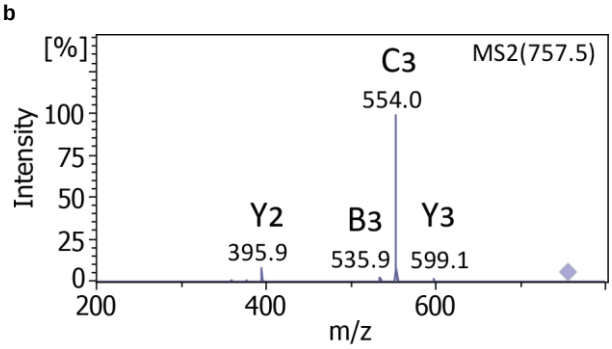


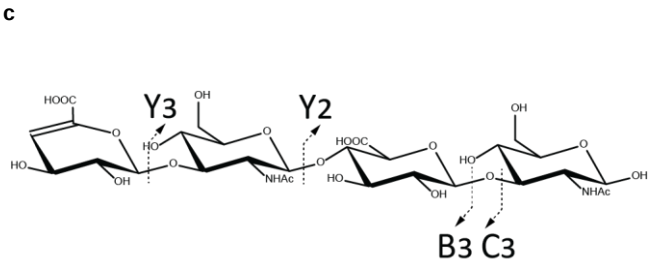
Supplementary Figure 39 | GH31 family members from *A. muciniphila* ATCC BAA-835. a. Amuc_1008^{GH31} cartoon model coloured rainbow blue-red from the N- to C-terminus, respectively. GalNAc (yellow) from 7F7Q was overlaid into the active site. **b.** The cartoon of Amuc_1008^{GH31} (rainbow) overlaid with the *E. faecalis* GH31 (grey) with the same specificity showing that they overlay well in terms of secondary structures, but the *A. muciniphila* enzyme has additional modules. **c.** Top: Amuc_1008^{GH31} active site surface overlaid with Tn-antigen from 7F7R *Enterococcus faecalis* GH31_18. Bottom: active site of 7F7R for comparison. Left and right panels show the surface of the enzymes and residues interacting with the sugar in the -1 subsite as sticks, respectively. **d.** Amuc_1187^{GH31} cartoon model coloured rainbow blue-red from the N- to C-terminus, respectively. Isomaltose (blue) from 3MKK was overlaid into the active site. **e.** Amuc_1187^{GH31} active site surface overlaid with α-linked glucose disaccharides from different crystal structures. From left to right: Isomaltose from 3MKK, Nigerose from 7WJC, and Kojibiose from 7WJF. Only the isomaltose was not making obvious steric clashes with the predicted protein structure. **f.** Activity against pullulan and trehalose. Isomaltose reactions are in Supplementary Figure 16. **f.** Activity against glucosylceramide. Left panel is the TLC and right panel are the assays run on the HPAEC alongside standards. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



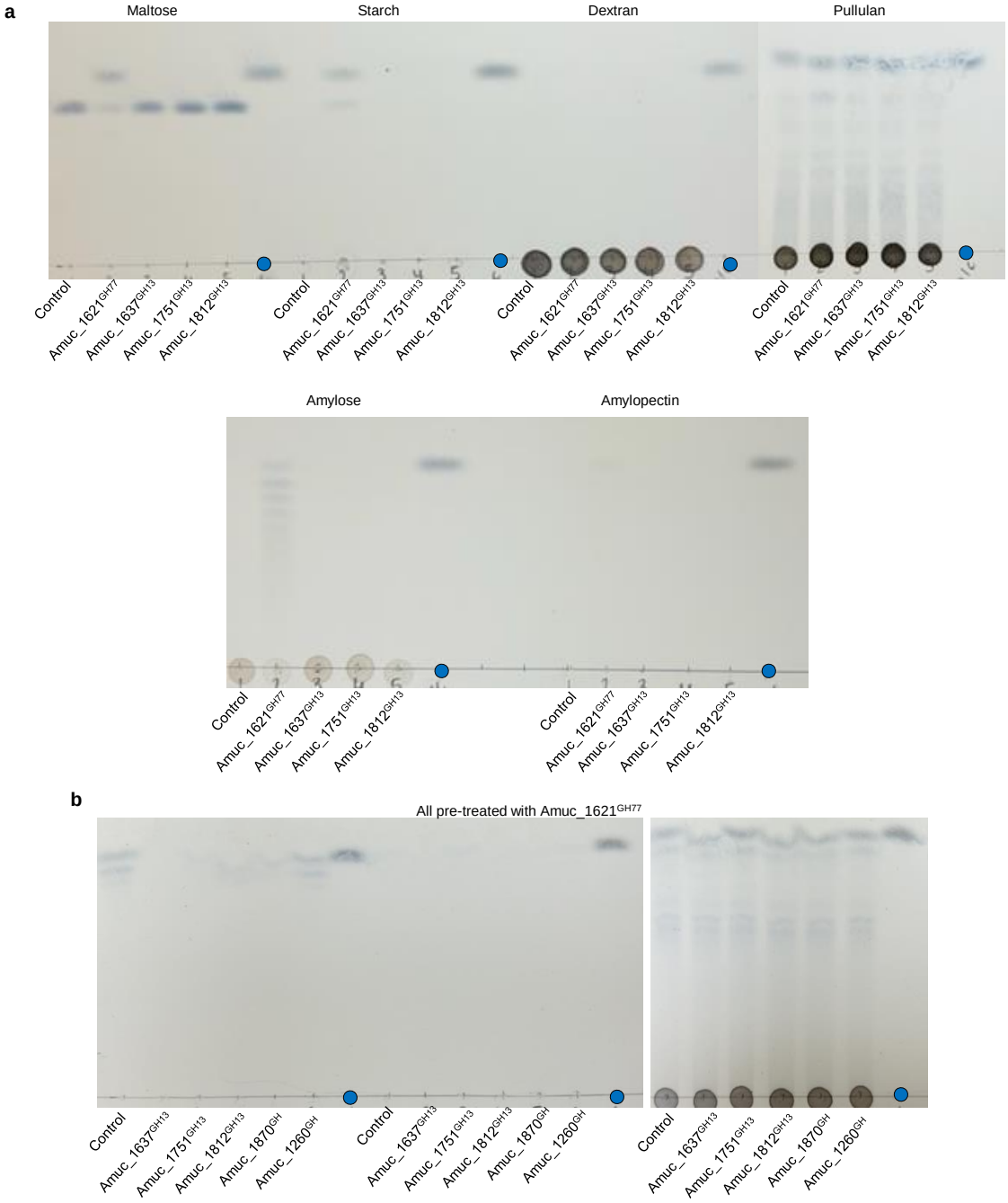
Supplementary Figure 40 | Enzyme kinetics for Amuc_0778^{PL38} and Amuc_0863^{GH105} against GAG substrates. a, Activity of Amuc_0778^{PL38} against known PL38 substrates glucuronan (left) and alginate (right). Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 µM enzyme. **b,** Initial rates of Amuc_0778^{PL38}. with increasing substrate concentrations. The Michaelis-Menten model was fitted in Graph-Prism. Error-bars represents standard deviations from triplicate experiments. **c,** Initial rates of Amuc_0863^{GH105} for increasing substrate concentrations. The substrate concentrations were obtained directly from the AmPL38 kinetic experiments, and the initial rate defined as loss/converted double-bond per minute. In cases were the initial absorbance of substrate exceeded 2.7 (A235) the data was omitted. No kinetic models could be fitted reliably to any of the datasets. Error-bars represents standard deviations from triplicate experiments. **d,** Testing if the GH89 enzymes have activity against GAGs. Asterisks indicate the presence of Amuc_0778^{PL38} and Amuc_0863^{GH105}. Standards have also been included on the TLCs. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 µM enzyme.



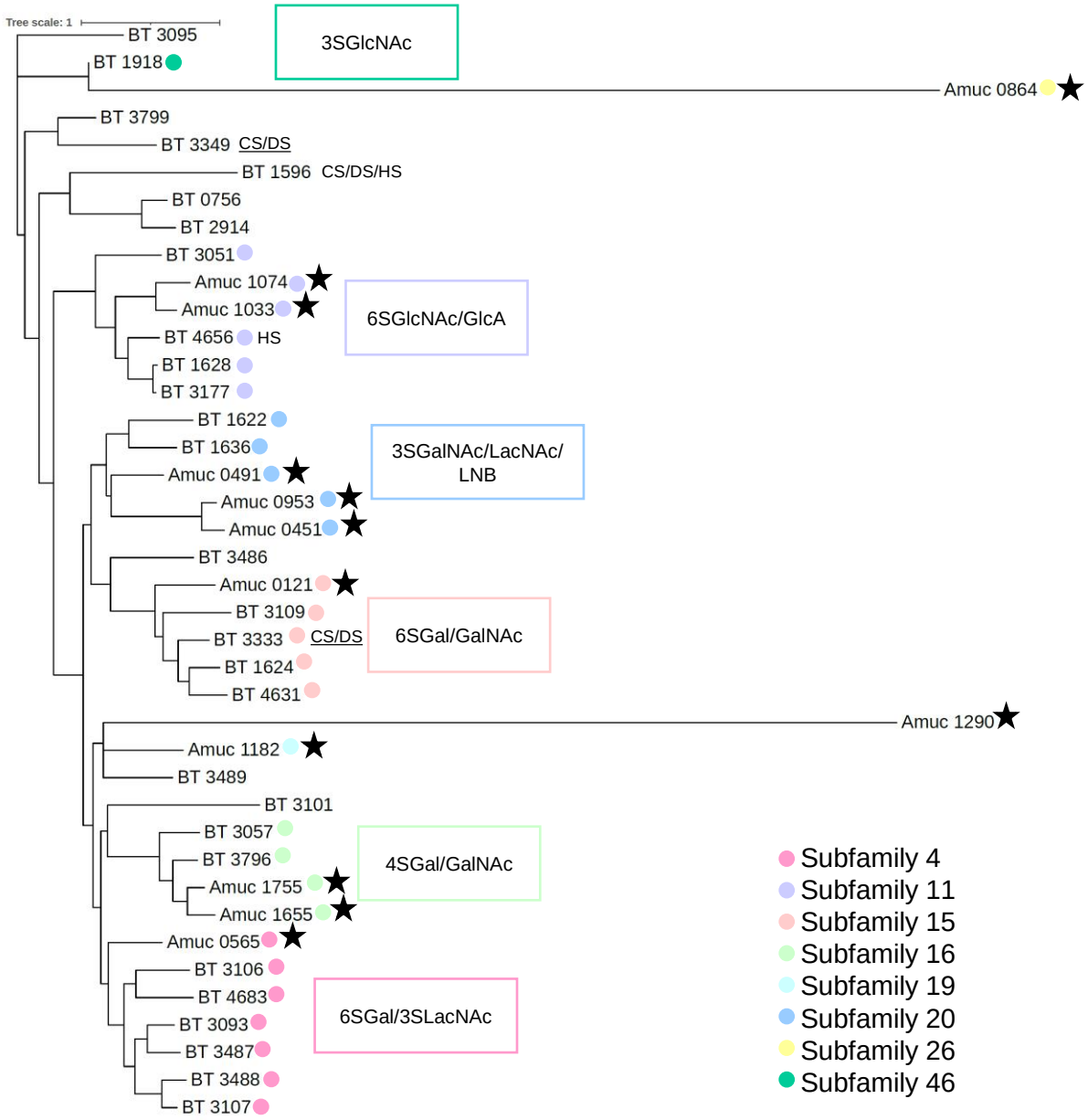




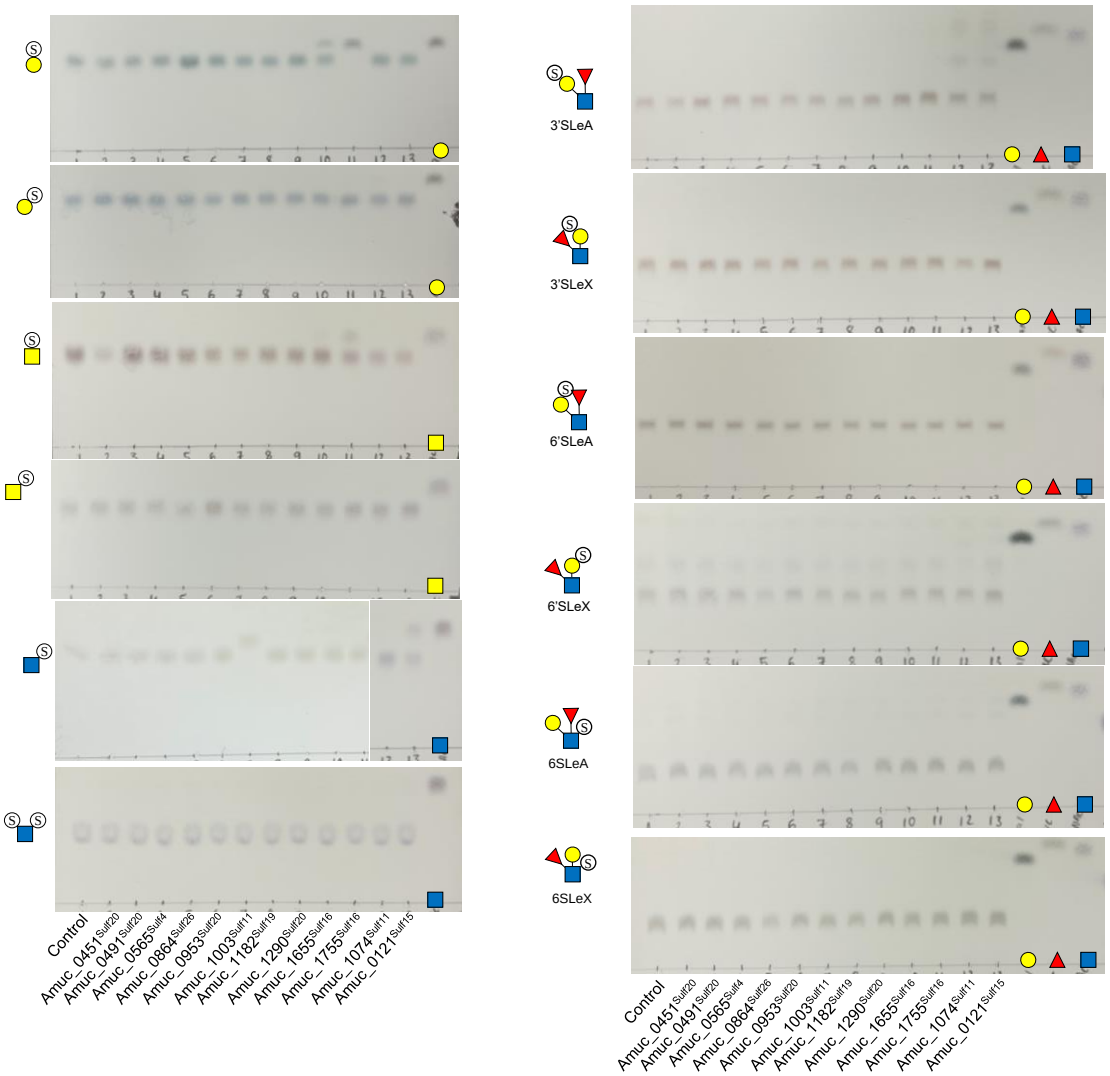
Supplementary Figure 41│Examples of the LC-MS data of of Amuc_0778^{PL38} against GAG substrates. a, Extracted ion chromatograms corresponding to the most abundant compounds during Amuc_0778^{PL38} degradation of hyaluronic acid, chondroitin sulfate A, chondroitin sulfate C, and dermatan sulfate. Extracted masses: UAHexNAc2: [M-H]- 378 Da, [2M-H]- 757 Da; UAHexAHexNac2: [M-H]2- 378 Da, [M-H]- 757 Da; UAHexA2HexNac3: [M-H]2- 567.5 Da, [M-H]- 1136.1 Da, UAHexA3HexNac4: [M-H]2- 757 Da, UAHexNAc2S: [M-H]- 458 Da; UAHexAHexNac2S: [M-H]2- 418 Da, [M-H]- 837 Da; UAHexAHexNac2S2: [M-H]2- 458 Da. **b,** Negative mode ESI MS^2 fragmentation of UAHexAHexNAc2, a product of hyaluronic acid degradation by Amuc_0778^{PL38}. Diamond indicates precursor ion at 757.5 Da. Fragments are following the nomenclature of Domon and Costello (ref) are indicated on both MS spectrum and structure. **c,** A structure of the glycan corresponding to the fragmentation pattern in 'b'.



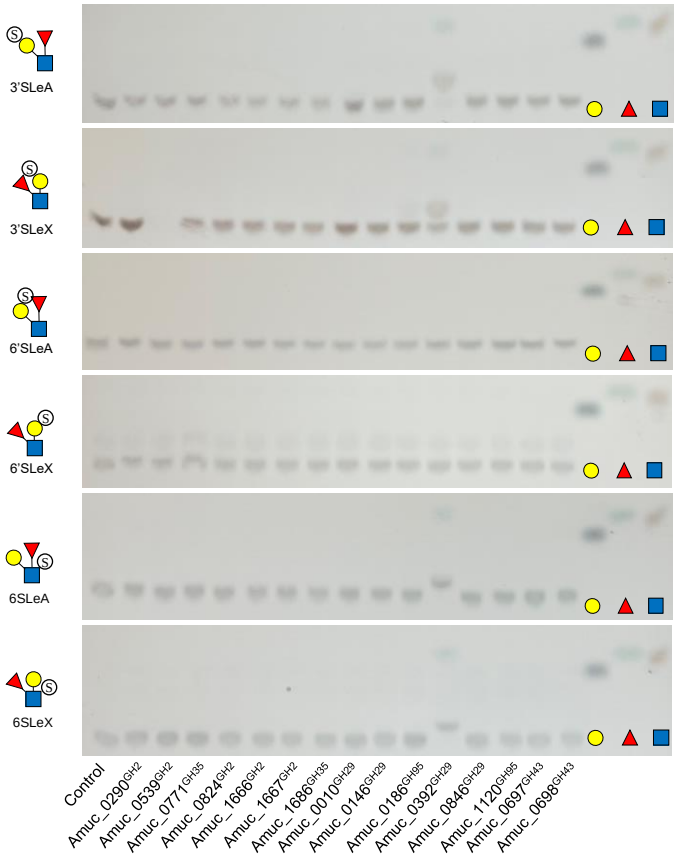
Supplementary Figure 42│Activities of the GH13 and GH77 family members from *A. muciniphila* ATCC BAA-835 against different α-linked glucose polysaccharides. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 μM enzymes. **a,** Assays of the individual enzymes against possible substrates. **b**, Assays where the substrates were treated with Amuc_1621^{GH77} as well as the other enzymes. Standards have also been included on the left of the TLCs.



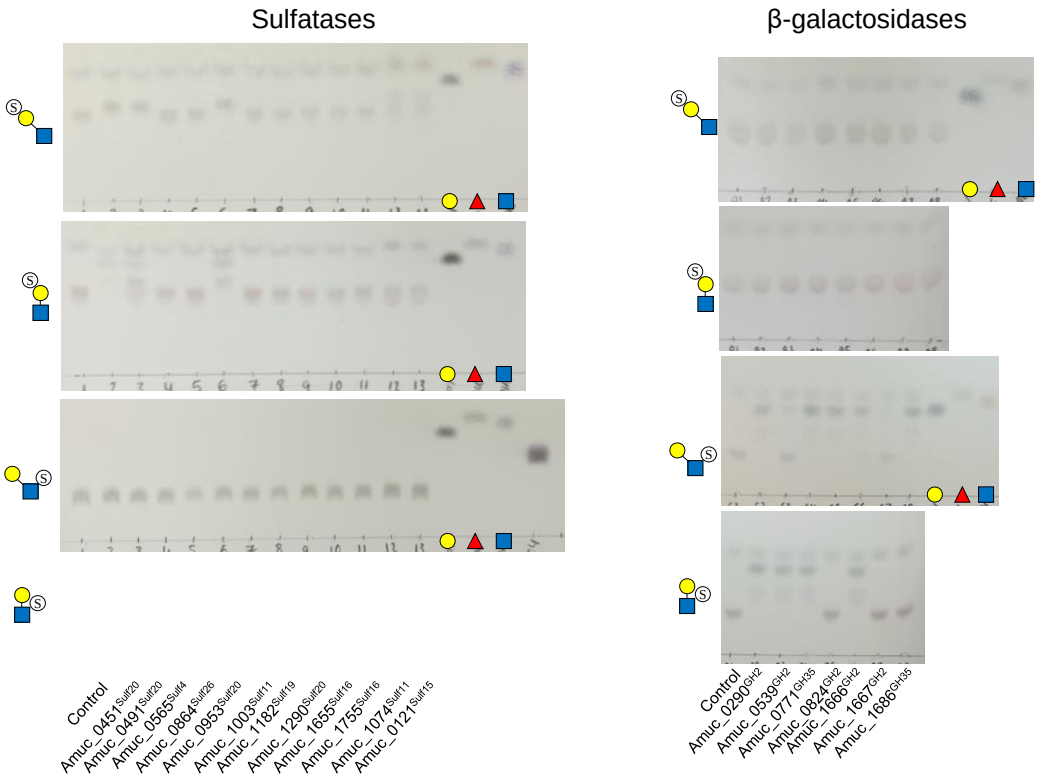
Supplementary Figure 43 | Phylogenetic tree of characterised sulfatase family members from *Bacteroides thetaiotaomicron* with those from *A. muciniphila* ATCC BAA-835. The sequences of the sulfatases with reported activities and the ones from *A. muciniphila* ATCC BAA-835 were compared as described in the methods. The different subfamilies are indicated by different colours and the *A. muciniphila* ATCC BAA-835 enzymes are highlighted by black stars. Some more specific information about activity is supplied were possible. When a sulfatase is known to act on a particular GAG, then this is noted and underlined. The enzymes are represented by their locus tags.



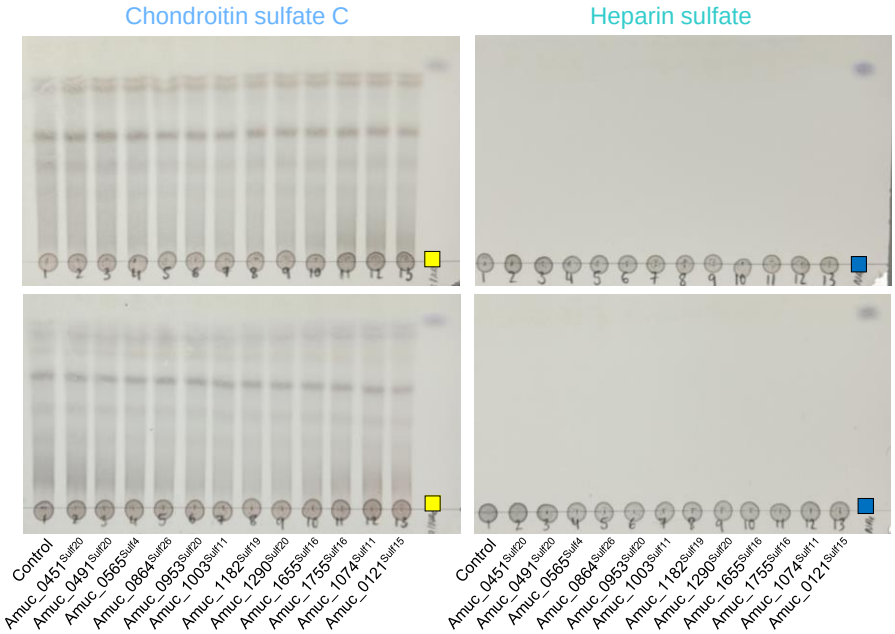
Supplementary Figure 44 | Activity of sulfatases from *A. muciniphila* BAA-835 against defined sulfated monosaccharides and Lewis Structures. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



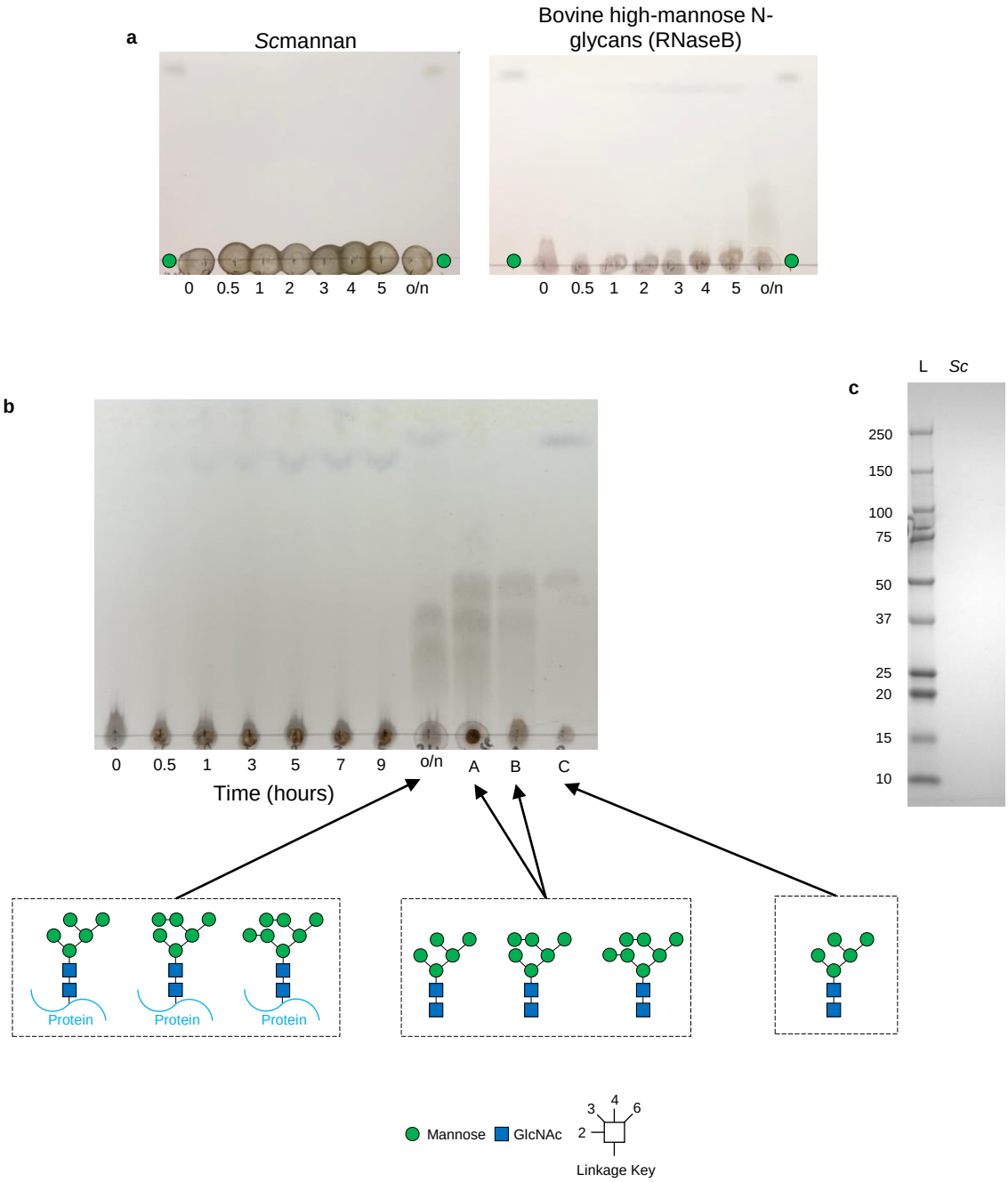
Supplementary Figure 45│Activity of β-galactosidases and α-fucosidases from *A. muciniphila* BAA-835 against sulfated Lewis Structures. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 μM enzyme.



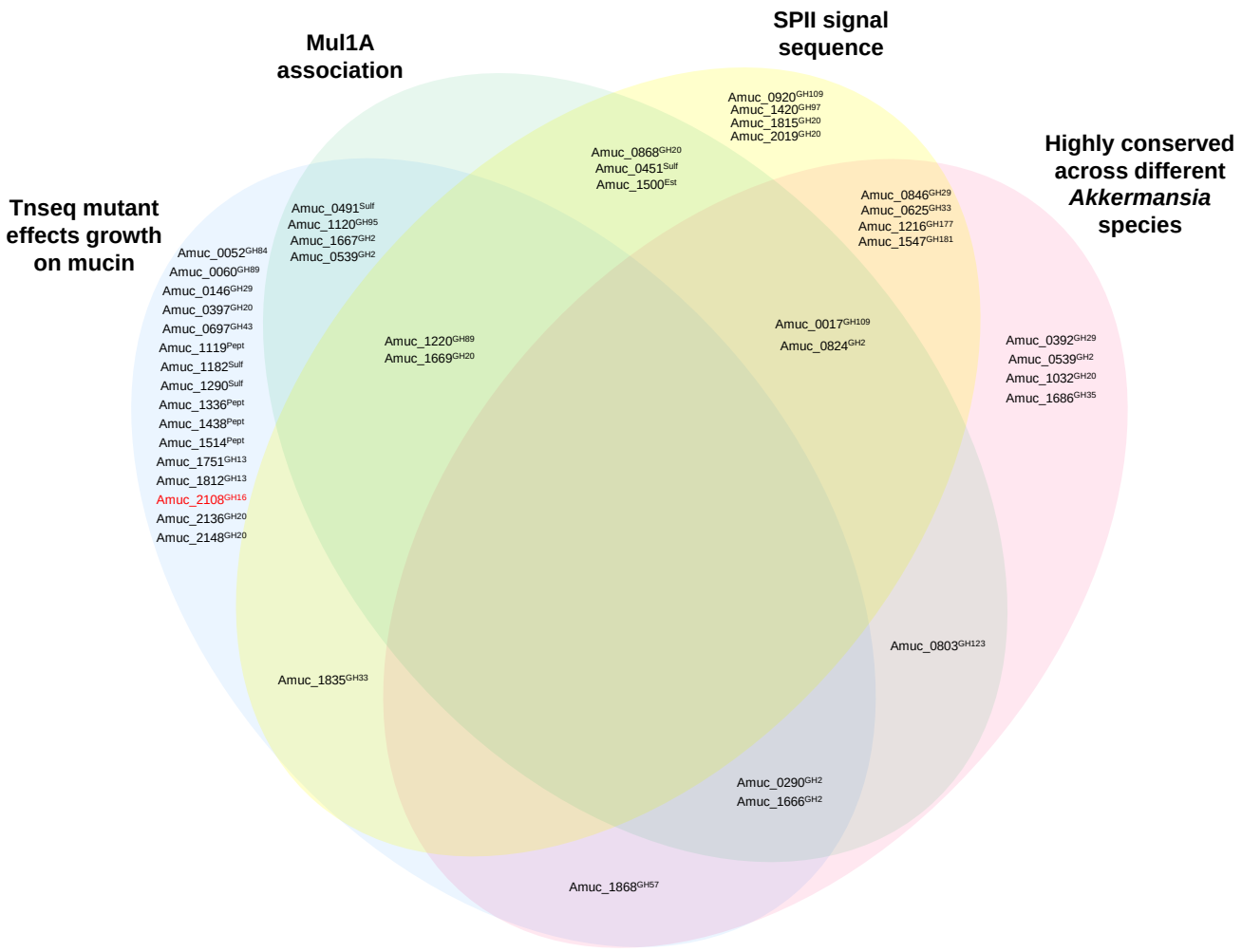
Supplementary Figure 46│Activity of fucosidases and β-galactosidases from *A. muciniphila* BAA-835 against sulfated Lewis Structures. Standards have also been included on the TLCs. Enzyme assays were carried out at a final substrate concentration of 1 mM, pH 7, 37 °C, overnight, and with 1 µM enzyme.



Supplementary Figure 47 | Activity of sulfatases from *A. muciniphila* BAA-835 against GAGs. The putative sulfatases were tested against two types of sulfated GAGs in the presence of Amuc_0778^{PL38} and Amuc_0863^{GH105}. Standards have also been included on the TLCs. Enzyme assays were carried out at pH 7, 37 °C, overnight, and with 1 µM enzyme.



Supplementary Figure 48 | Whole cell assay with high-mannose N-glycoprotein and *Sc*mannan. a, The results from the whole cell assay showing that no obvious degradation of the mannan is occurring (left) and a smear appears overnight for the high-mannose N-glycoprotein (right). **b,** TLC is repeated with controls. A: the overnight whole cell assay sample treated with PNGaseL, which is a broad-acting PNGase. B: RNaseB treated with PNGaseL. The smear seen here is a mix of high-mannose N-glycans of between 5 and 7 mannose sugars. C: RNaseB treated with PNGaseL and BT3990, which is an α1,2-mannosidase that trims all the high-mannose N-glycans down to a five-mannose structure. results of whole cell assays of PGM III-grown cells against different substrates. The reaction was carried out at 37 °C, a sample removed at different time points, and boiled to stop enzyme activities. Monosaccharide standards are shown in the right lanes. All mannose linkages are alpha. **c,** SDS-PAGE gel stained for protein using Coomassie Brilliant Blue. L – labber and *Sc* – *Sc*mannan.



Supplementary Figure 49│Venn diagram summary of some observations made about the different CAZymes and sulfatases. Four different features are included to assess the importance of different enzymes and to consider if some of these observations are associated. The first is if the signal sequence indicates that the enzyme may be localised to the outside of the cell and the second is if the enzyme was highly conserved (>90 %) across all species assessed in the pangenome analysis. The other two aspects come from Davey et al. and include the enzymes that were found to effect growth on mucin during a Tnseq library experiment. Finally, this publication also highlighted the putative mucin import system and these enzymes were found to associate with this complex.