Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature

environmental microbiology reports

Environmental Microbiology Reports (2021) **00**(00), 00–00 doi:10.1111/1758-2229.12980

Minireview

Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature

Lauren S. McKee,[1,2*] Sabina Leanti La Rosa,[3] Bjørge Westereng,[4] Vincent G. Eijsink,[4] Phillip B. Pope[3,4] and Johan Larsbrink[2,5*]

[1]*Division of Glycoscience, Department of Chemistry, KTH Royal Institute of Technology, AlbaNova University Centre, Stockholm, 106 91, Sweden.*

[2]*Wallenberg Wood Science Center, Stockholm, 100 44, Sweden.*

[3]*Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway.*

[4]*Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway.*

[5]*Division of Industrial Biotechnology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, 412 96, Sweden.*

Summary

The Bacteroidetes phylum is renowned for its ability to degrade a wide range of complex carbohydrates, a trait that has enabled its dominance in many diverse environments. The best studied species inhabit the human gut microbiome and use polysaccharide utilization loci (PULs), discrete genetic structures that encode proteins involved in the sensing, binding, deconstruction, and import of target glycans. In many environmental species, polysaccharide degradation is tightly coupled to the phylum-exclusive type IX secretion system (T9SS), which is used for the secretion of certain enzymes and is linked to gliding motility. In addition, within specific *species* these two adaptive systems (PULs and T9SS) are intertwined, with PUL-encoded enzymes being secreted by the T9SS. Here, we discuss the most noteworthy PUL and non-PUL mechanisms that confer specific and rapid polysaccharide degradation capabilities to the Bacteroidetes in a range of environments. We also acknowledge that the literature showcasing examples of PULs is rapidly expanding and developing a set of assumptions that can be hard to track back to original findings. Therefore, we present a simple universal description of conserved PUL functions and how they are determined, while proposing a common nomenclature describing PULs and their components, to simplify discussion and understanding of PUL systems.

Received 17 December, 2020; accepted 23 May, 2021. *For correspondence. E-mail mckee@kth.se; Tel: +46 (0) 8 790 99 39. E-mail johan.larsbrink@chalmers.se; Tel. +46 (0) 31 772 38 39.

Introduction

The Bacteroidetes phylum dominates in glycan-rich environments including the gastrointestinal tract of bilaterians (humans and herbivores), freshwater and marine aquatic environments, and terrestrial ecosystems such as soil (Newton *et al*., 2011; Thomas *et al*., 2011; Fernández-Gómez *et al*., 2013; Krüger *et al*., 2019; Larsbrink and McKee, 2020). In each of these habitats, there is a constant supply of biomass rich in proteins and carbohydrates of plant, animal, and microbial origin. Bacteria rely mainly on glycoside hydrolases (GHs) and polysaccharide lyases (PLs) to deconstruct this diet of complex glycan polymers, which includes branched polysaccharides comprising multiple different monosaccharides connected via a range of different linkages (Fig. 1). In the marine environment, there is a strong reliance on sulfatase enzymes to metabolize the sulfate decorations of marine plant-derived polysaccharides (Fig. 1). Sulfate groups can also be found decorating polysaccharides in animal tissues (Fig. 1), and the variability of sulfation patterns makes metabolism of the polysaccharides a complex process. In addition, many plant-derived complex polysaccharides contain other non-carbohydrate decorations that can hinder GH action, and these may be enzymatically cleaved by enzymes like carbohydrate esterases (CEs). Due to this challenging complexity, a consortium of enzymes with complementary specificities is required for full conversion of a polysaccharide into simple sugars for further metabolism. To facilitate studies of such carbohydrate-active enzymes (CAZymes), these proteins have been grouped into classes and families based on sequence similarity in the Carbohydrate-Active Enzymes database (CAZy, www.cazy.org; Lombard *et al*., 2014).

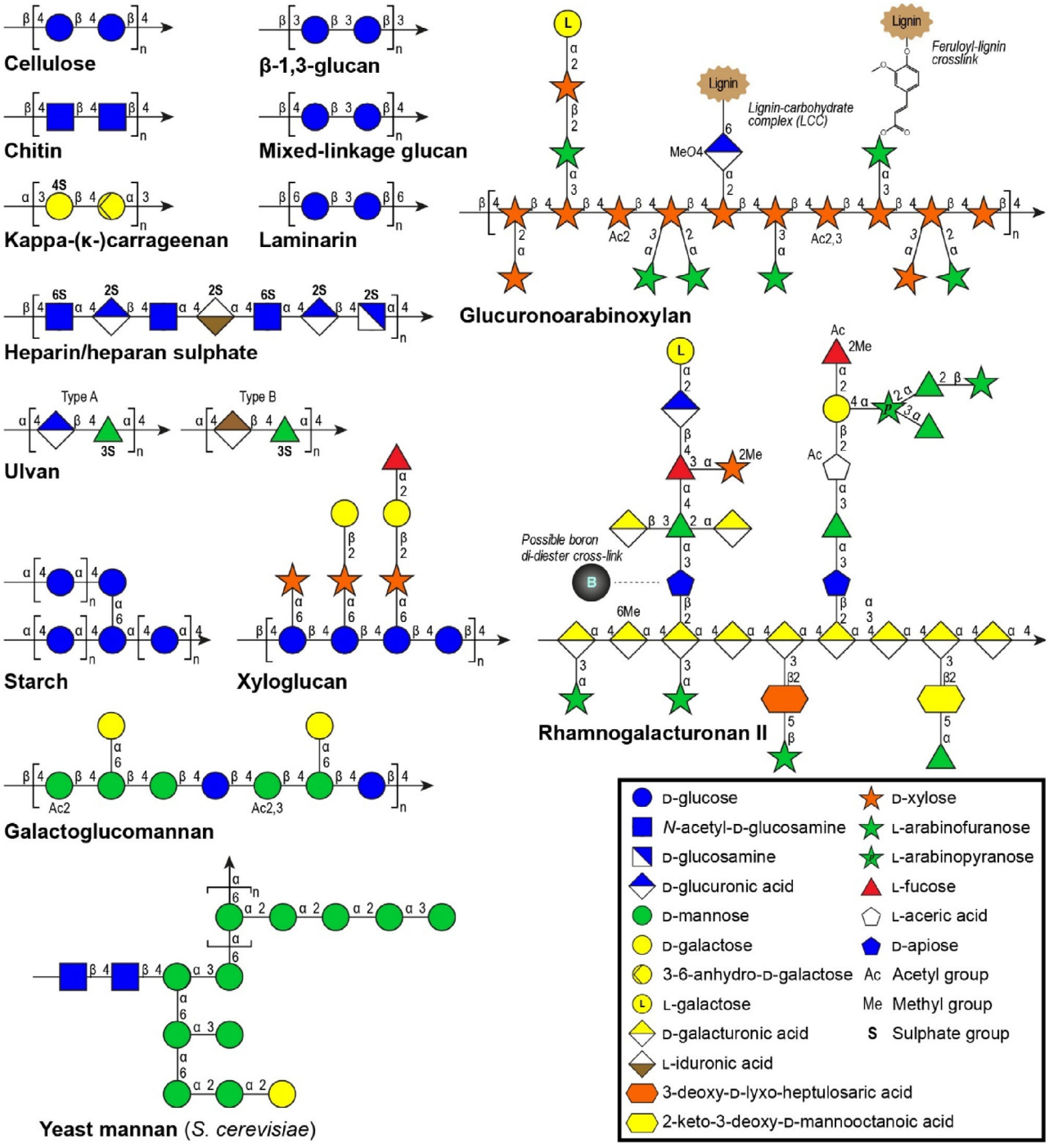

Fig. 1. Exemplary structures of some of the classes of biomass-derived complex carbohydrate structures that are abundant in nature and known to support Bacteroidetes growth. Conventional monosaccharide symbols are shown, glycosidic bond linkages are indicated, and square brackets highlight repeating elements. The figure illustrates only a fraction of the diversity and variability of carbohydrate and non-carbohydrate decorations and linkages found in Nature. Polysaccharide abbreviations used in the main text: cellulose – Cel; chitin – Chi; xylan – Xyl; arabinoxylan – AX; glucuronoxylan – GX; glucuronoarabinoxylan – GAX; xyloglucan – XyG; mannan – Man; glucomannan – GM or GluM; galactomannan – GM or GalM; galactoglucomannan – GGM; laminarin – Lam; mixed-linkage glucan (β-1,3/1,4-glucan) – MLG; β-1,3-glucan – β1-3G; arabinogalactan (protein) – AG(P); rhamnogalacturonan II – RGII; agarose – Aga; alginate – Alg; carrageenan – Car.

As will be discussed below, one of the most significant adaptations of the Bacteroidetes phylum, and perhaps the single adaptation that confers the most important advantage in hyper-competitive environments, is the ability to coordinate production of synergistic enzyme consortia in response to the availability of a relevant substrate.

While the Bacteroidetes are more or less universally recognized for their ability to produce consortia of complementary CAZymes, how they use these enzymes to interact with their target substrate can vary substantially. For example, certain species secrete large (high-molecular weight) multi-modular proteins that themselves contain several catalytic and non-catalytic domains. Such multi-modular proteins may or may not be produced as part of the CAZyme consortia encoded by the so-called polysaccharide utilization loci (PULs, discussed in detail below and first introduced in the study by Xu *et al*., 2003), which are clusters of genes encoding proteins with related functions used for the sensing, binding, deconstruction, and import of a particular polysaccharide. The secretion of large proteins and/or large numbers of enzymes is energetically demanding (Wallenstein and Burns, 2011; Traving *et al*., 2015), and so the Bacteroidetes have evolved systems to ensure a strong return on this energy investment. The ability to sense precisely which polysaccharides are nearby and to tailor CAZyme gene expression and protein secretion accordingly is vital, as is the ability to move rapidly over solid surfaces, perhaps towards an area more enriched with a particular carbon source. The type IX secretion system (T9SS), which is unique to the Bacteroidetes, is important in many species for secretion of the aforementioned CAZymes either to the cell surface or freely into the environment, and is intertwined with PULs in certain habitats. The different types of PULs that have thus far been described, and their mechanistic differences are discussed in this review in the context of the fitness advantages they confer to members of the Bacteroidetes phylum.

PULs are specialized saccharolytic systems with functional homology to the paradigmatic starch utilization system

The proteins necessary for capture and deconstruction of complex carbohydrates by Bacteroidetes species are typically encoded by discrete cassettes of contiguous genes called PULs (Grondin *et al*., 2017). All PULs studied to date confer the ability to digest one particular glycan. The first PUL to be identified was the starch utilization system (SUS) of *Bacteroides thetaiotaomicron*, a dominant member of the human gut microbiota (HGM) and a model species for studying polysaccharide digestion in the gut (Tancula *et al*., 1992; Shipman *et al*., 1999; Xu *et al*., 2003; Martens *et al*., 2009). The SUS remains the archetypal example of a PUL, and many of the methods still used to investigate PULs were established using the SUS. Several excellent reviews are available which describe the functioning of the canonical SUS, its discovery by Dr. Abigail Salyers (Whitaker, 2018), and the ground-breaking research involved in the dissection of the system (e.g., see Martens *et al*., 2009; Koropatkin and Smith, 2010; Cameron *et al*., 2012).

The eight genes comprising the SUS, *susRABCDEFG*, each encode a protein involved in the sensing, capture, import, or hydrolysis of starch (Table 1). The SUS still defines the PUL paradigm, but the literature describing examples of PULs from other species has expanded greatly since the biochemistry of the SUS was first established. With the breadth of PUL research reported and ongoing, a reader who is not well-versed in the history of the SUS may find some terminology unclear. The nomenclature used to refer to the SUS has become the convention when discussing other loci, and many components of newly discovered PULs are still referred to as SUS-equivalent proteins. The so-called SusC-like and SusD-like proteins found in all Bacteroidetes PULs have sequence, structure, and functional homology with the original SusC and SusD proteins. These proteins have sometimes been referred to as $SusC_H$ and $SusD_H$ to indicate that they are homologues of SusC and SusD (Ndeh *et al*., 2017; Luis *et al*., 2018). However, for the other SUS components, including the transcriptional regulator SusR and the cell-surface starch-binding proteins SusE and SusF, there are no sequence homologues in most PULs, and even the functional equivalence is not always clear. Likewise, the activities of the CAZymes encoded by PULs vary widely. In Table 1, we summarize the functions of the components of the original SUS, and their equivalents found in most other PULs characterized to date. Now that an abundance of PULs have been fully characterized, it is worth considering which canonical SUS features hold true for most or all cases, and which are specific adaptations that may be useful in starch metabolism but seem less relevant for other glycan substrates.

Lipid-anchoring to the outer membrane, which holds the protein on the external cell surface, is a common feature of PUL proteins (e.g., see Shipman *et al*., 1999; Larsbrink *et al*., 2014b; Cuskin *et al*., 2015; Tamura *et al*., 2017; Cartmell *et al*., 2018; Pereira *et al*., 2021), and indeed, four of the SUS proteins are anchored in this way. This includes the *endo*-acting amylase SusG which cleaves starch into maltodextrins, and functional analogues of this enzyme have been found in PULs targeting other glycans. All PULs characterized to date enable the import of resulting oligosaccharides of relatively high molecular weight into the periplasm by the combined action of SusC-like and SusD-like proteins acting in a complex (Glenwright *et al*., 2017; Bolam and van den Berg, 2018; Gray *et al*., 2021). The SusC/D complex functions with a 'pedal bin'-like mechanism, where SusD

Table 1. The components of the archetypal SUS.

Protein	Function in the SUS	Equivalent proteins in other PULs	Identification in other PULs
SusR	Sensor protein spanning the inner membrane and recognizing the disaccharide maltose and larger malto-oligosaccharides (MaltOs) in the periplasm, then signalling to induce SUS transcription	Classical or hybrid two-component sensor systems, or other types of sensor proteins, such as AraC-type regulators. Typically recognizes an oligosaccharide fragment of the polysaccharide targeted by the PUL's enzymes and induces upregulation of PUL transcription	Genomic proximity to SusC/D pair, annotation as transcriptional regulator protein. The glycan recognized can be identified by binding studies or inferred by transcriptomic investigation of the other PUL-encoded genes after growing the bacterium on pure carbon sources
SusA	Periplasmic neopullulanase, synergistically converting MaltOs to glucose together with SusB	One or more periplasmic CAZymes with specificity for imported oligosaccharides. Often *exo*-acting	CAZy family annotation, proximity to SusC/D pair. Activity determined by recombinant production and characterization
SusB	Periplasmic α-glucosidase, synergistically converting MaltOs to glucose together with SusA	One or more periplasmic CAZymes with specificity for imported oligosaccharides. Often *exo*-acting	CAZy family annotation, proximity to SusC/D pair. Activity determined by recombinant production and characterization
SusC	Integral outer membrane protein, TonB-dependent sugar transporter, importing oligosaccharides delivered by SusD that are derived from the polysaccharide targeted by the PUL	Commonly referred to as SusC-like protein, TonB-dependent sugar transporter, specific for fragments of the target polysaccharide	SusC-like proteins are identified by sequence identity. Always adjacent to a SusD-like protein
SusD	Lipid-anchored starch/MaltO-binding protein, forms a complex with SusC, and shuttles starch fragments into the SusC pore	Commonly referred to as SusD-like protein (sometimes as SGBP-A), binds to fragments of the target polysaccharide, shuttles them into the cognate SusC-like protein pore	SusD-like proteins are identified by sequence identity. Always adjacent to a SusC-like protein. Binding specificity can be determined by recombinant production and characterization
SusE	Cell surface-anchored starch binding protein, aiding SUS complex formation, and assisting starch capture (Cameron *et al*., 2012)	Sometimes referred to as cell surface glycan-binding proteins (SGBPs). Some SGBPs promote glycan capture, some contribute to SusC/D complex architecture	Proximity to SusC/D pair. Recombinant production and characterization as glycan binders. Generally, no detectable sequence similarity with known proteins
SusF	Cell surface-anchored starch binding protein, aiding SUS complex formation, and assisting starch capture (Cameron *et al*., 2012)	SGBP. Some SGBPs promote glycan capture, some contribute to SusC/D complex architecture	Proximity to SusC/D pair. Recombinant production and characterization as glycan binders. Generally, no detectable sequence similarity with known proteins
SusG	Amylase lipid-anchored to the cell surface, converting starch into importable MaltOs. Additional non-catalytic MaltO binding site and appended CBM further promote starch sequestration (Koropatkin and Smith, 2010)	One or more extracellular CAZyme(s) (typically *endo*-acting) with specificity for the target polysaccharide. Can be freely secreted and/or outer-membrane tethered.	CAZy family annotation, proximity to SusC/D pair. Activity determined by recombinant production and characterization.

The table describes functional equivalents of SUS components found in other PULs, and how they may be identified via the gene sequence or by functional description. A complete PUL may be identified as a discrete cassette of contiguous genes that are upregulated during growth on a particular glycan. Knock-outs of an entire PUL, or one or more components thereof, can be used to explore the importance of PUL proteins. Once a PUL is identified in a genome, functional characterization of the CAZymes and carbohydrate-binding proteins is required for an accurate description of PUL specificity and function, as CAZyme family annotations are indicative but not always precisely predictive of enzyme specificity.

acts as a lid with carbohydrate-binding properties, facilitating shuttling of oligosaccharides into the SusC pore which is closed to the periplasm by a plug. As the SusD lid is closed, the plug is pulled out of the SusC pore through interaction with TonB, a protein complex spanning the periplasm, and the oligosaccharide cargo is then released into the periplasm for further depolymerization (Bolam and van den Berg, 2018).

Another key feature of the SUS is the use of outer membrane-tethered non-catalytic glycan-binding proteins, which is a common but not ubiquitous feature in other PULs. The so-called cell surface glycan-binding proteins (SGBPs) have functional analogy to the original SusE and SusF and have now been discovered in PULs targeting multiple different glycans (Rogowski *et al*., 2015; Ndeh *et al*., 2017; Cartmell *et al*., 2018; Tamura *et al*., 2021). In some publications, the SusD-like proteins have also been referred to as SGBPs (see examples in the studies by Tauzin *et al*., 2016; Tamura *et al*., 2019; Déjean *et al*., 2020; Tamura and Brumer, 2021). SGBPs are typically thought to facilitate substrate acquisition by the PUL. In the SUS, SusE and

SusF appear to be crucial for formation of the greater SUS protein complex formed on the surface of the cells; super-resolution imaging and single-molecule tracking studies suggest that SusE and SusF are immobile on the cell surface, whereas the SusC/D pair and SusG enzyme are highly mobile across the cell surface, and are recruited into the larger protein complex only when needed. This indicates that the entire SUS protein complex may form around SusE and SusF (Tuson *et al*., 2018). To the best of our knowledge, the same has not been demonstrated for SGBPs in other PULs, with the exception of a β-glucan PUL for which it was shown that an SGBP supports function of the SusC-like protein (Déjean *et al*., 2020). Functionally equivalent complex-recruiter proteins are not included in current PUL discovery algorithms, which are discussed in detail below.

In addition to carbohydrate binding, breakdown, and transport, an important conserved feature of the SUS and all other PULs is system activation and specific sensing of imported oligosaccharides. In the SUS, MaltOs imported through SusC activate SusR, an inner-membrane sensor/regulator with glycan specificity for both linkage and monosaccharide type that exclusively upregulates expression of the other genes in the locus (D'Elia and Salyers, 1996; Cho *et al*., 2001). This preference for oligosaccharide ligands over simple monosaccharides gives more information about the structure of available polysaccharides, and likely contributes to fine distinctions in substrate preferences between PULs (Martens *et al*., 2011). Upon MaltO binding to SusR, the production of all SUS components is upregulated (except SusR itself; D'Elia and Salyers, 1996), and this is a common PUL feature. While the SUS does not need to encode an inner membrane transporter for importing the ubiquitous monosaccharide glucose into the cytosol, some other PULs do include inner membrane sugar transporters that are essential for growth on the PUL's target glycan (Larsbrink *et al*., 2016; Terrapon *et al*., 2018).

All PULs, including the SUS, are believed to be transcribed constitutively at low levels, permitting a minor 'background' capability to hydrolyse the target polysaccharide and import the resulting oligosaccharides (Pereira *et al*., 2021). As early degradation products enter the periplasm and activate the sensor system (SusR in the SUS), the PUL is upregulated. For the SUS, this leads to increased concentrations of the SusA, SusB, and SusG enzymes. Other PULs correspondingly encode enzymes suited to the deconstruction of their target glycans, some of which are secreted to the outside of the cell, and some of which remain in the periplasm to complete the deconstruction of imported glycan fragments. Thus, upon sensor activation, an array of specific CAZymes is deployed to rapidly hydrolyse the PUL-inducing glycan polymer into importable oligosaccharides, followed by periplasmic degradation to monosaccharides (Cameron *et al*., 2012; Tamura *et al*., 2017). A positive feedback loop is activated: enhanced enzyme production leads to a higher rate of polysaccharide depolymerization, higher concentrations of the activating oligosaccharide in the periplasm, and an ensured persistent activation of the PUL's transcriptional regulator until the target polysaccharide is depleted or a polysaccharide higher in the bacterium's 'preference list' is detected (discussed below). PUL regulators include classical two-component systems or hybrid two-component variants, extracytoplasmic function (ECF) sigma factors with corresponding anti-sigma factors, GntR-like transcription factors, and AraC-like regulators (Lowe *et al*., 2012; Terrapon *et al*., 2018).

New PULs can be identified by algorithmic comparison to the SUS or by de novo *bacteriology*

Since the characterization of the now-paradigmatic SUS, genes homologous to SusC and SusD have become recognized as a conserved signature motif that can be used to identify new PULs in Bacteroidetes genomes (Martens *et al*., 2011; Terrapon *et al*., 2015; Stewart *et al*., 2018; Terrapon *et al*., 2018). Indeed, there is consensus that a PUL is defined and identified by the observation of at least one tandem *susC*/*D*-like pair of genes closely flanked by at least one CAZyme gene, and some form of transcriptional regulator. While *susC* and *susD* homologues can be identified by sequence similarity, other PUL-encoded proteins are only functional analogues to SUS components, as discussed above and in Table 1. PULs not targeting starch do not encode enzymes similar to SusABG, but instead encode CAZymes acting on the PUL's target glycan. Similarly, while the great majority of PULs lack homologues of the non-catalytic SusE and SusF, some do include functionally related SGBPs (Table 1). There is now an automated system, PULDB (http://www.cazy.org/PULDB/) (Terrapon *et al*., 2015; Terrapon *et al*., 2018), which predicts PULs within Bacteroidetes genomes by identifying SusC/D homologues. CAZy-annotated proteins and transcriptional regulators in close proximity are then annotated as belonging to the putative PUL. In addition, dbCAN-PUL serves as a repository of experimentally validated PULs (http://bcb.unl.edu/dbCAN_PUL/) (Ausland *et al*., 2021). Table 2 shows the number of PULs predicted by the PULDB algorithm for the genomes of some species found in different environments. Some Bacteroidetes possess over 100 different PULs (Lapébie *et al*., 2019), and the number of PULs within a genome is strongly correlated with the number of

Table 2. A selection of species from various habitats and the numbers of PULs they possess (as automatically predicted by the PULDB tool at http://www.cazy.org/PULDB/; Terrapon *et al*., 2015; Terrapon *et al*., 2018).

Species	Number of PULs[a]	Habitat
Bacteroides thetaiotaomicron	90	Gut
Bacteroides ovatus	118	Gut
Bacteroides fragilis	60	Gut
Bacteroides xylanisolvens	100	Gut
Proteinphilum acetatigenes	77	Wastewater sludge
Chitinophaga pinensis	106	Soil
Chitinophaga niabensis	153	Soil
Cytophaga hutchinsonii	2	Soil
Flavobacterium johnsoniae	40	Soil/freshwater
Prevotella melaninogenica	25	Upper respiratory tract
Prevotella salivae	32	Mouth
Prevotella ruminicola	24	Rumen
Zobellia galactanivorans	61	Marine

Only a small proportion of these PULs have been experimentally verified. Note that these PULDB predictions show some discrepancies with previously published numbers of PULs in some genomes. The first publication describing PUL discovery in *B. thetaiotaomicron* manually identified 88 candidate PULs (Martens *et al*., 2008), and the equivalent publication for *B. ovatus* identified 112 candidate PULs (Martens *et al*., 2011). A 2009 description of the *F. johnsoniae* genome described 33 candidate PULs and PUL-like loci (McBride *et al*., 2009).

[a]If several strains have been analysed, the number corresponds to the highest number of PULs among different strains.

polysaccharides metabolized by a particular species (Martens *et al*., 2011). However, it should be stressed that predictions of the exact number of true PULs in a bacterial genome based solely on the number of SusC/D homologues is not without risk, as SusC/D pairs can be found without any neighbouring CAZymes or regulators and still be listed as predicted PULs (Terrapon *et al*., 2018). These pairs may be involved in the acquisition of non-carbohydrate nutrients, or be part of PUL-like cassettes such as the phosphate utilization systems, comprising SusC/D-like proteins and phosphatase enzymes, that were recently identified in *Flavobacterium* strains (Lidbury *et al*., 2020). Other metabolic applications of SusC/D-like protein pairs may yet be uncovered. Additionally, there may be cases where a SusC/D pair and distally located CAZyme-encoding genes are regulated by as-yet unidentified transcriptional regulators, forming non-canonical PULs that are not organized into the typical contiguous loci. Such 'polysaccharide utilization regulons' would not be identified by the current PUL-predicting algorithms but could be detected in transcriptomic or proteomic investigations.

Importantly, despite the abundance of PULs in many Bacteroidetes genomes, not all polysaccharide metabolism is necessarily directed by these systems. For example, enzymes metabolizing starch, glycogen and peptidoglycan are often found outside of PULs, likely because their target substrates are found within the bacterial cell or cell wall, making the sensing and import functions of PULs superfluous (Lapébie *et al*., 2019). Likewise, despite the large numbers of CAZymes with fine specificity encoded by many PULs, there are examples where the CAZymes of one PUL are not sufficient to fully metabolize one complex polysaccharide (Cuskin *et al*., 2015; Ndeh *et al*., 2017; Briliūt *et al*., 2019; Lapébie *et al*., 2019), meaning that PUL activities may be complemented by the action of non-PUL enzymes encoded elsewhere in the genome.

The naming of PULs is a difficult matter

Once a PUL has been identified and fully biochemically characterized using recombinant techniques, it is typically given a concise name describing its function. In many cases, the 'polysaccharide' in 'PUL' is replaced by the name of the specific glycan being metabolized, as in 'xyloglucan utilization locus, XyGUL' (Larsbrink *et al*., 2014b), or 'mixed-linkage β-glucan (MLG) utilization locus, MLGUL' (Tamura *et al*., 2017). In other cases, the name of the targeted glycan is appended to the existing PUL abbreviation, as in 'Mannan-degrading PUL, Man-PUL' (Reddy *et al*., 2016; Bågenholm *et al*., 2017), or 'PUL for degrading xylan, PUL-Xyl' (Rogowski *et al*., 2015). These short names given to characterized PULs are certainly more useful than referring to a PUL as a collection of contiguous locus tags in an annotated genome, but as the PUL literature expands, there is a need for one consistent naming convention for characterized PULs, as this would be of use for both readers and database listings.

One reason for the current inconsistency in PUL naming is the lack of a strong consensus on how to abbreviate the often complex names of polysaccharides themselves. For example, in chemical and biochemical literature, the plant heteroglycan xyloglucan is comm only abbreviated to XG (Benselfelt *et al*., 2016), but in some publications is referred to as XyG (Larsbrink *et al*., 2014b), or even TXG, with this latter form indicating that tamarind seeds (T) are the source of the polysaccharide (McKee and Brumer, 2015). Three-letter codes are typically preferred where they are natural and appropriate and are used wherever possible in the naming of CAZymes, following a standard nomenclature system proposed in the 1990's (Henrissat *et al*., 1998): examples include Chi (chitin), Xyl (xylan), Cel (cellulose), and Man (mannan). We believe that such indicatory codes are a

useful route to standardization of nomenclature, but that an effective PUL name should confer sufficient detail regarding the structure of the targeted polysaccharide to enable a reader to understand fine differences in substrate specificity where they exist. For example, the use of 'ManPUL' may be confusing now that there are examples in the literature of PULs targeting both α- and β-linked mannans from microbes and plants, respectively (Cuskin *et al*., 2015; Bågenholm *et al*., 2017). Furthermore, regarding plant mannans, 'ManPUL' as a general term would not distinguish between PULs specifically targeting glucomannan (GM or GluM), galactomannan (GM or GalM), or galactoglucomannan (GGM). Similarly, 'XylPUL' would be insufficient to distinguish between PULs preferentially targeting arabinoxylan (AX), glucuronoxylan (GX), glucuronoarabinoxylan (GAX or AGX), or xylo-oligosaccharides (XOs or XylOs). We must also consider the extent to which subtle variations in polysaccharide structure (such as the degree of arabinosylation of an AX) are even relevant to the activation of a PUL.

A consistent system for polysaccharide naming should be agreed upon by researchers active within the CAZy community to permit consistent PUL nomenclature. Subsequent to this agreement, we encourage the use of 'GlycanUL' to refer to a PUL directing the metabolism of a given glycan, where a consistent abbreviation is used to denote a particular polysaccharide. The abbreviations should conform to those already used in polysaccharide and CAZyme research, where possible. This would give, for example: ChiUL (chitin utilization locus); XyGUL (xyloglucan utilization locus); GAXUL (glucuronoarabinoxylan utilization locus); GGMUL (galactoglucomannan utilization locus); and αManUL (α-mannan utilization locus, an example where a three-letter polysaccharide code must be expanded as it gives insufficient detail on substrate structure). With the ongoing rapid expansion of the PUL literature, there are now multiple examples of PULs targeting the same polysaccharide type, and so species indicators will become increasingly useful, as already used in publications describing the so-called *Bo*MANPUL (Reddy *et al*., 2016; Bågenholm *et al*., 2017) and *Bo*XyGUL (Larsbrink *et al*., 2014b) of *B. ovatus*, or the β-1,3-glucan targeting PULs of *B. fluxus* (*Bf*1,3GUL) and *B. uniformis* (*Bu*1,3GUL) (Déjean *et al*., 2020). Eventually, it may be necessary to additionally include information about the order of discovery of PULs found within an organism's genome (e.g. *Bo*XyGUL-A or *Bo*XyGUL1 for the first such characterized example, and *Bo*XyGUL-B or *Bo*XyGUL2 for the second, and so on). Using letters for this (A, B, C, etc) would echo the long-standing nomenclature used for naming characterized CAZymes (Henrissat *et al*., 1998). The decided names of newly characterized PULs could be submitted to the CAZy database using something akin to the online form that now allows researchers to directly input enzyme function data, for integration into the PULDB. Finally, we recognize that this is a dynamic discussion that will benefit from engagement and advertisement within the greater CAZyme community at a public forum such as the biennial CAZymes for Glycan Conversions meeting.

All classes of polysaccharides are metabolized by PULs across multiple investigated ecosystems

Due to a certain anthropocentric focus in the literature, and the interest in Bacteroidetes species as indicators of and contributors to human gut health (Fan and Pedersen, 2020), a substantial proportion of PUL investigations have focussed on HGM species and the metabolism of glycans found in the human intestinal ecosystem. While cellulose degradation is a major activity in the rumen of herbivores and in soils, cellulose is recalcitrant to digestion in the human gut (Slavin *et al*., 1981; Chassard *et al*., 2010), whereas other abundant plant glycans such as hemicelluloses, starch, and pectin, constitute major nutrient sources for the HGM (Scheller and Ulvskov, 2010) (Fig. 1). Of the hemicelluloses, xylans are particularly enriched in grasses, grains, and cereals (Vogel, 2008); β-mannans are found in certain nuts and seeds and are used as food-thickeners (Scheller and Ulvskov, 2010); xyloglucan is found in all terrestrial plants including vegetables and cereals (Vogel, 2008); and mixed linkage β-glucan (MLG) is mostly found in cereals. Plant-based diets are also rich in pectin and related glycans such as arabinan and arabinogalactan, which are particularly abundant in some fruits, berries, and processed foods like jams (Mohnen, 2008). Arabinogalactan is also found in the form of arabinogalactan proteins (AGPs) (Fincher *et al*., 1983), which are particularly enriched in red wine (Vidal *et al*., 2003), instant coffee (Capek *et al*., 2010), and natural gums used in food processing (Phillips, 1998; Atgié *et al*., 2019). Fungal biomass represents an additional nutrient source for diverse microbiota and is arguably even more abundant and important in soils than gut systems. Fungi offer a buffet of complex glycan-based cell walls built of chitin, α-glucans, β-glucans, α-mannans, galactans, and glycoproteins (Fig. 1) (Gow *et al*., 2017). In marine environments, the cell walls of aquatic plants present linkages, monosaccharides, and sulfated groups that are not typically found in terrestrial plants, including polysaccharides such as carrageenans, agarose, porphyran, ulvan, alginate, and laminarin (Popper *et al*., 2014; Synytsya *et al*., 2015) (Fig. 1). In all of these environments, complex heteroglycans require multiple synergistic CAZymes for complete deconstruction,

addressing the multitude of monosaccharide types, linkages, and non-carbohydrate decorations (including sulfate, acetyl, and feruloyl groups, among others).

Bacteroidetes species are abundant in gut, soil, and aquatic environments, and PULs have been discovered that target every major glycan class in plant, algal,

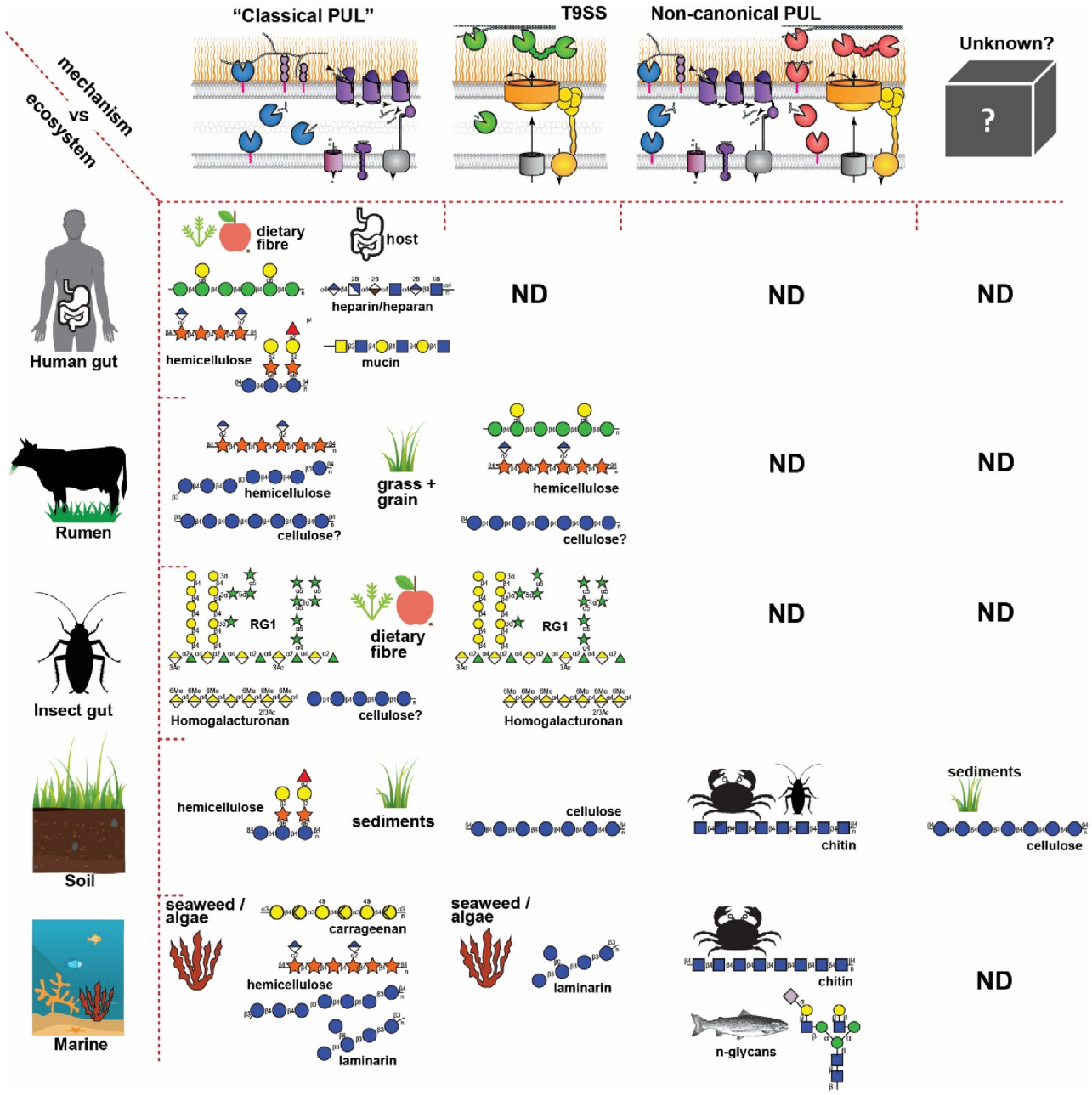

Fig. 2. Examples of saccharolytic mechanisms discovered in various environments colonized by the Bacteroidetes. Detailed structural depictions of 'classical PULs', the T9SS, and non-canonical 'hybrid' PULs are illustrated below in Fig. 3. 'Unknown' represents examples of saccharolytic Bacteroidetes that have been isolated in pure culture, but whose approach to polysaccharide metabolism has yet to be fully elucidated, despite there being clear phenotypic evidence of polysaccharide degradation (e.g. *Cytophaga hutchinsonii*) (Taillefer *et al*., 2018). Example substrates are shown in the row corresponding to their source ecosystem and specific source organisms are indicated alongside substrate structure depictions. Columns indicate the type of mechanism used in each example shown. For specific studies, please refer to Grondin *et al*. (2017) for human 'classical PULs'; (Barbeyron *et al*., 2016) for marine ecosystems; (Pérez-Pascual *et al*., 2017) for salmon gut microbiome; (Larsbrink *et al*., 2016) for chitin degrading in soil and marine environments; (Vera-Ponce de León *et al*., 2020) for cockroach gut microbiome; (Grondin *et al*., 2017; McKee *et al*., 2019) for soil; and (Rosewarne *et al*., 2014) for rumen. ND denotes 'not discovered', which means such mechanisms may yet still exist in these exemplar environments. The shown glycan structures are representative and do not fully cover the larger variety of building blocks and structures that are present in Nature. For example, the mucin structure only shows a core 3 type, which is one of the more abundant glycan structures in MUC2.

animal, and microbial biomass (Fig. 2), highlighting the enormous substrate diversity and adaptive flexibility that the PUL system provides to the phylum (Glowacki and Martens, 2020). However, environmental studies outside of the HGM also indicate that the PUL paradigm and its mechanisms are not necessarily conserved or of equal significance across these environments. Examples of how adaptations of PULs have arisen in different environments are highlighted in Fig. 2 and discussed further below.

PUL-mediated metabolism of complex glycans in the human gut microbiota. Several PULs from HGM species have been characterized in detail, and these typically conform to the standard, or 'classical', view of how PULs operate, with surface-bound enzymes cleaving target glycans into oligosaccharides, which are imported through the SusC/D-like protein complex for final degradation within the periplasmic space (Table 1) (Fig. 3). Notable examples of some of the key polysaccharides metabolized by the HGM and the corresponding PULs are described below, without providing fine details of individual enzyme specificities, which are not

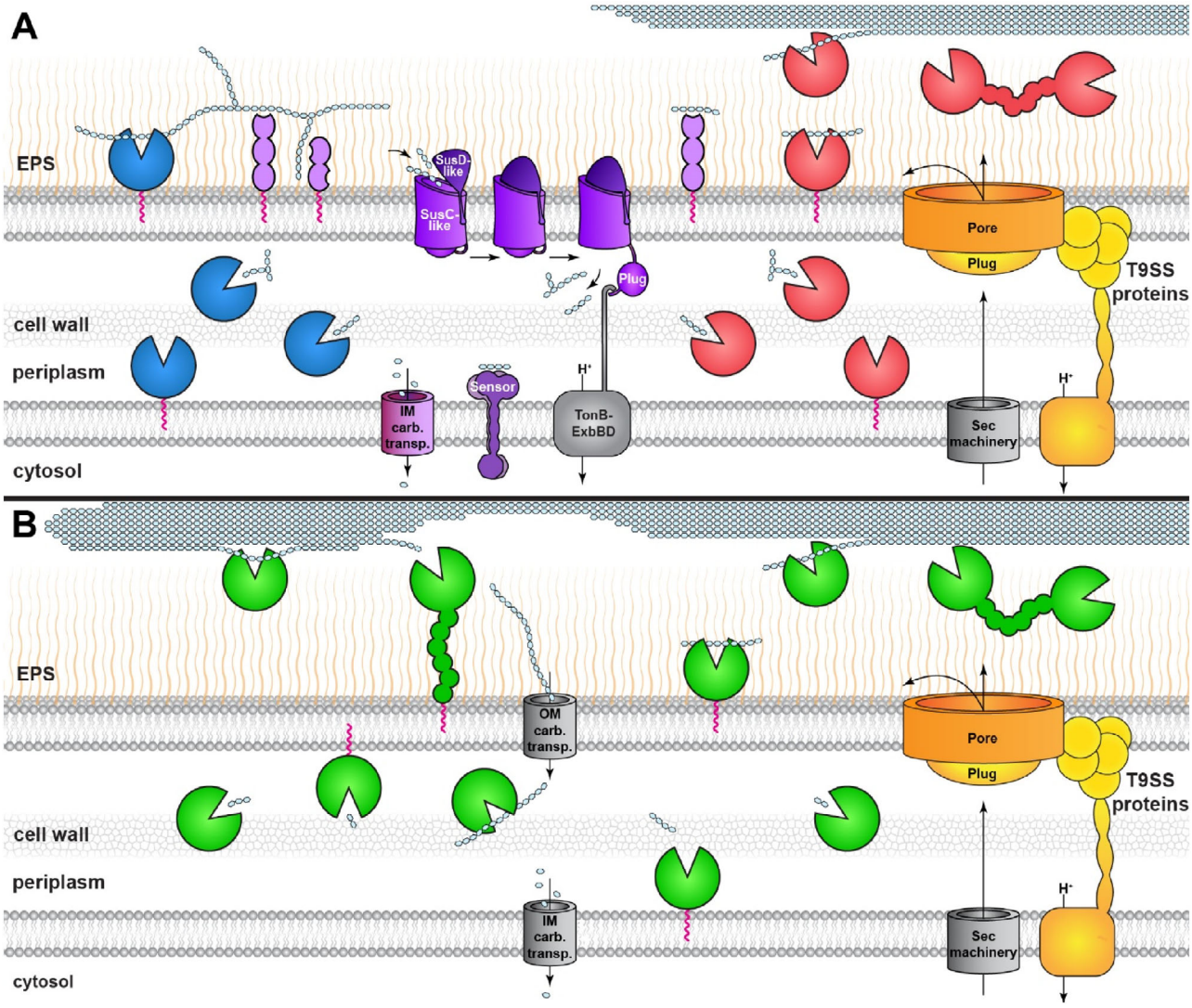

Fig. 3. Overview of PUL systems and their connection to the T9SS in some species.A. The 'classical' PUL setup, where blue-coloured surface-attached CAZymes cleave polysaccharides into oligosaccharides to be imported and fully degraded to monosaccharides within the periplasm. In red, the enzymes of a non-canonical 'hybrid' PUL are shown; these are released into the environment following T9SS-mediated secretion, and sometimes include large enzymes comprised of multiple catalytic domains. Both 'classical' and 'hybrid' PULs have SusC/D-like pairs and regulator protein(s) in common, coloured in purple, where the SusD-like protein captures carbohydrates to be imported through the SusC-like pore, and the sensor regulates the upregulation of the PUL upon binding of signature oligosaccharides. Both systems may also include surface-tethered non-catalytic glycan-binding proteins (in purple), as well as inner-membrane transporters (in purple).B. Non-PUL systems employed by Bacteroidetes species, with enzymes in green. Here, the core PUL apparatus is absent and genes encoding CAZymes and related proteins can be spread throughout the genome. These systems have been shown to generally rely on T9SS secretion, large multi-catalytic enzymes, and apparently redundant systems of extracellular soluble, membrane-bound, and periplasmic proteins (e.g. the cellulolytic *C. hutchinsonii*; Taillefer *et al*., 2018).

within the scope of this review and may be found in the cited literature.

Bacteroides ovatus grows on the majority of plant glycans, and has together with the closely related *B. thetaiotaomicron* become a key species for studying PUL diversity within the HGM. Xylan metabolism in *B. ovatus* is mediated by loci referred to in the original publication as PUL-XylL and PUL-XylS (Fig. 2) (Rogowski *et al*., 2015), which are conserved in Bacteroidetes derived from other mammals (e.g. rumen-isolated *P. bryantii*; Dodd *et al*., 2010). The larger locus, PUL-XylL, enables SGBP-mediated binding to and degradation of complex GAX, while the smaller PUL-XylS is responsible for binding and hydrolysing simpler GXs and undecorated linear xylan (Rogowski *et al*., 2015). Following the nomenclature framework outlined above, these loci would be named *Bo*GAXUL and *Bo*GXUL (or *Bo*XylUL), respectively. Similar to *Bo*GXUL, growth on MLG is enabled by the small *Bo*MLGUL, where the size reflects the complexity of the target glycan (Tamura *et al*., 2017). Xyloglucan deconstruction by *B. ovatus* is mediated via the *Bo*XyGUL (Fig. 2) (Larsbrink *et al*., 2014b), which can fully degrade the arabinofuranosylated XyG found in solanaceous plants but lacks enzymes targeting fucosyl decorations. A PUL with somewhat looser specificity is the so-called *Bo*ManPUL (Fig. 2; suggested name *Bo*GalMUL), which targets both glucomannan and galactomannan (Reddy *et al*., 2016; Bågenholm *et al*., 2017). These examples highlight how, even within one species, the highly adaptable PUL system is permitting both specific and more general polysaccharide metabolism, by varying enzyme repertoires. These PULs of *B. ovatus* all have in common that in addition to an expected SusD-like protein, sometimes referred to as SGBP-A, they also encode functional analogues to SusE/F ('SGBP-B proteins') which help sequester the target glycan to the cell surface (Larsbrink *et al*., 2014b; Rogowski *et al*., 2015; Tauzin *et al*., 2016; Bågenholm *et al*., 2017; Tamura *et al*., 2017). Similarly organized (syntenic) PULs to those functionally characterized have been observed in other members of the phylum, including some outside of the *Bacteroides* genus (Larsbrink *et al*., 2014b; Terrapon *et al*., 2015; Tamura *et al*., 2017; Terrapon *et al*., 2018), and this has community-level ecosystem implications if certain species can internalize and hoard large oligosaccharides from a broad range of structurally related glycans.

Bacteroides uniformis encodes a PUL conferring both MLG and β-1,3-glucan metabolism (Déjean *et al*., 2020), and similar PULs from *B. thetaiotaomicron* and *Bacteroides fluxus* have also been studied (Tamura *et al*., 2021), each encoding an SGBP in addition to the PUL's SusD-like protein. The current naming of these β1,3GULs perfectly exemplifies the nomenclature problem: fine details of substrate specificity are obscured, if they exist, because there is no standard name for this type of substrate, unlike the xylans, mannans, and xyloglucans, which show wide structural variability depending on the source organism, but which can always be referred to by their generic polysaccharide name. Depending on the origin, β-1,3-glucans can have several different names, including callose (plants), curdlan (bacteria), pachyman (fungi), or laminarin (algae), and these may in some cases show variability in terms of degree of polymerization, acetylation, or glycosyl substitution. Polysaccharides enriched with the Glc-β1,3-Glc linkage also include yeast β-glucan and fungal scleroglucan/schizophyllan, which additionally contain the Glc-β1,6-Glc linkages that are also found in laminarin (Manners *et al*., 1973; Kadam *et al*., 2015). Of note, the *Bu*β1,3GUL was shown to deconstruct yeast β-glucans and laminarin (Déjean *et al*., 2020), as well as MLG, indicating that this PUL may simply target any polysaccharide containing Glc-β1,3-Glc linkages regardless of finer polysaccharide differences, indicating that this name is sufficient as there is not a preference for one particular named glucan. But another PUL that may hypothetically show preference for branched β-1,3-glucans or linear β-1,3-glucans within a particular range of molecular weight or degree of acetylation may need a more specific name. Differences in purity, molecular weight, and structural features like acetylation are common between β-glucans extracted in different ways, but these data are not commonly reported for commercial substrate preparations; while this information is not always accessible to enzymologists, these differences may in some cases be relevant when dissecting the precise functions of PULs, as they can influence substrate solubility and hence influence cell adhesion capabilities, as well as the efficiency of enzymes and binding proteins.

Bacteroides thetaiotaomicron encodes two PULs (currently named PUL_{AGPL} and PUL_{AGPS}) that target highly complex and variable AGP (Martens *et al*., 2011; Cartmell *et al*., 2018). In addition to β-1,3-galactan cleavage, each PUL encodes different abilities to remove the variable AGP side chains (Cartmell *et al*., 2018), such as β-1,6-linked galactose side groups. The latter are only addressed by PUL_{AGPL}, in another example of the very fine distinctions that sometimes occur between PULs acting within the same substrate group. Sufficiently descriptive short names for these PULs are not obvious, but *Bt*AGPUL-A and *Bt*AGP UL-B seem reasonable. The ability of *B. thetaiotaomicron* to target one of the most complex polysaccharides known, RGII, has also been characterized in detail and is attributed to three distinct PULs in the genome (RG-II PULs 1–3; suggested re-naming to *Bt*RGIIUL-A-C), thus correlating PUL complexity with the structural complexity of the target glycan (Ndeh *et al*., 2017). In addition to dietary plant glycans, polysaccharides deriving from dietary fungi such as baker's yeast and fungal HGM members, the so-called mycobiome (Huseyin *et al*., 2017; Sam *et al*., 2017), are also important drivers of Bacteroidetes metabolism and specialization. Three loci, currently named MAN-PULs 1–3, have been identified in *B. thetaiotaomicron* as being activated by α-mannan deriving from *Saccharomyces cerevisiae* (Fig. 1), *Schizosaccharomyces pombe*, or the pathogen *Candida albicans*. These loci might with a new nomenclature be referred to as *Bt*αManUL-A-C. In contrast to *B. thetaiotaomicron*, *Bacteroides xylanisolvens* is able to metabolize α-1,6-mannan, but not intact complex mannan from *S. cerevisiae*, suggesting a higher selectivity for α-mannan degradation (Cuskin *et al*., 2015). Such differences in encoded PUL repertoires may also be a driver of microdiversity in substrate niche colonization (Hehemann *et al*., 2016). As for *B. ovatus*, the studied PULs from *B. thetaiotaomicron* rely on surface-attachment of key enzymes, analogous to the archetypal SUS.

Marine plant-derived polysaccharides are commonplace in the diets of only a few restricted human populations, and so PULs targeting these glycans have been mostly discovered in aquatic microbiomes (discussed below). In 2010, Hehemann *et al*. revealed that a porphyran- and agar-degrading PUL identified in the genome of the marine species *Zobellia galactanivorans* is present in the HGM of Japanese individuals (Hehemann *et al*., 2010). This PUL was acquired by *Bacteroides plebeius* via lateral gene transfer, leading to an adaptation within a specific human population with a traditionally high consumption of seaweed. Phylogenomic analyses further uncovered horizontally acquired alginolytic PULs originating from an ancestral *Z. galactanivorans* in other HGM *Bacteroides* in the guts of Japanese individuals (Thomas *et al*., 2012). In addition, a *B. uniformis* strain was shown to have acquired an agarose-targeting 'Ag-PUL' (Pluvinage *et al*., 2018). Laminarin can be degraded by the aforementioned *Bu*β1,3GUL (Déjean *et al*., 2020), and PUL-mediated carrageenan metabolism has recently been described (Pudlo *et al*., 2020). Following our suggested nomenclature, the PULs targeting carrageenan, laminarin, alginate, and agarose could be re-named CarUL (if needed adding Greek letters, e.g. κ for κ-carrageenan), LamUL (or β1,3GUL), AlgUL, and AgaUL. In common with the plant polysaccharide-targeting PULs from the HGM, these systems appear to act in a classical manner, relying on surface-bound *endo*-acting enzymes and periplasmic degradation of oligosaccharides.

Finally, in addition to all of the dietary and microbial glycans available in the human intestine, several HGM members can forage directly on the mucosal layer lining the human large intestine, the site of bacterial colonization (Glowacki and Martens, 2020). Cell-surface glycans on intestinal epithelial cells can serve as a carbon source for HGM residents either as a major nutrient source or during dietary fibre deprivation in infants before weaning or in individuals consuming a low-fibre diet, depending on the degree of species specialization (Marcobal *et al*., 2011; Desai *et al*., 2016). Among the mucus-eroding micro biota, *B. thetaiotaomicron* encodes several PULs targeting host *N*- and *O*-linked glycans found in mucin (Martens *et al*., 2008; Martens *et al*., 2011), and uses combinations of CAZymes (Martens *et al*., 2008; Crouch *et al*., 2020) and sulfatases (Cartmell *et al*., 2017; Luis *et al*., 2020) to metabolize a range of host-derived glycans, including mucin, heparin, and keratan- and heparan sulfates (Fig. 2). Underlining the importance of this endogenous source of microbiota-accessible carbohydrates is the recent observation that mucin-derived *O*-glycans are effective prebiotics that can mitigate dysbiosis and suppress the pathogen *Clostridium difficile* (Pruss *et al*., 2021).

More than just classical PULs and CAZymes are important in non-HGM environments. Outside of the HGM, we also observe broad representation of classical PUL mechanisms being employed by Bacteroidetes, for example, in marine, soil, and host-associated ecosystems found in herbivores (Fig. 2). Ocean-dwelling Bacteroidetes are considered central degraders of the algal glycans which predominate in aquatic environments (Arnosti *et al*., 2021), and corresponding PULs targeting these polysaccharides have, as mentioned previously, on occasion been acquired by HGM residents (Hehemann *et al*., 2010). In contrast to PULs targeting plant cell wall glycans, these PULs need to make extensive use of polysaccharide lyases and sulfatases as well as GHs (Arnosti *et al*., 2021) (Fig. 1). Two PULs targeting marine glycans from the marine Bacteroidete *Gramella forsetii* KT0803 have been studied, one LamUL and one AlgUL (Kabisch *et al*., 2014). The response of *G. forsetii* when growing on laminarin and alginate, compared to glucose, was studied by cellular fractionation and proteomics. *Gf*LamUL is similar to the previously mentioned *Bu*LamUL/β1,3GUL from *B. uniformis* (Déjean *et al*., 2020), and similar genes and gene organizations were also found in putative PULs from related marine *Flavobacteriaceae*, again suggesting a conserved strategy for utilization of major glycans found in brown algae. These syntenic PULs all appear to operate according to the classical PUL mechanistic paradigm (Fig. 3). PULs rich in polysaccharide lyase-encoding genes permit metabolism of both mannuronate and guluronate components of alginate in *Maribacter dokdonensis* 62–1, which co-habits a similar metabolic niche as *Z. galactanivorans* (Wolter *et al*., 2021).

Z. galactanivorans is highly proficient in degrading marine glycans, and the large CarUL it utilizes for degradation of carrageenan is an example of a PUL not operating in the classical manner of the SUS archetype (Ficko-Blean *et al*., 2017). The *Zg*CarUL contains enzymes, a regulator, and an inner-membrane sugar transporter, but the expected SusC/D-like proteins are encoded elsewhere in the genome, as are other key carrageenolytic enzymes (Ficko-Blean *et al*., 2017). The CarUL (Fig. 2) is highly conserved within marine Bacteroidetes but varies in other phyla of marine bacteria in ways that indicate an evolutionary history of gene losses, duplications, and horizontal acquisitions around a conserved 3,6-anhydro-D-galactose core metabolism. Indeed, horizontal gene transfer between microbes in the ocean is a primary driver of micro-diversification in substrate acquisition capacity, as species target increasingly narrow niches of specific glycan structure (Hehemann *et al*., 2016). A similar finding is the AlgUL from *Z. galactanivorans*, which, as previously mentioned, is found in the genomes of both marine and gut bacteria (Thomas *et al*., 2012).

The system of *Z. galactanivorans* conferring agarose and porphyran metabolism represents another example of a non-canonical PUL situation (Hehemann *et al*., 2012b). The majority of genes encoding this complex system are found within two distally located loci, encoding the signature SusC/D-like proteins and a sensor, in addition to several enzymes. However, the bacterium also relies on enzymes located elsewhere in the genome, activated to different degrees by agarose and porphyran, without neighbouring genes with related function. While most of the β-porphyran-degrading enzymes of the system are predicted to reside in the periplasm, one is located in the outer membrane, as are several agarases, but additionally several key enzymes are found secreted as free enzymes using the T9SS. The PULs found in marine species thus contain the expected classical PULs, similar to those found in HGM species, but also non-canonical PULs complemented by distally located genes acting on the same polysaccharide. Such 'hybrid' PULs rely on secretion of proteins by the T9SS, including extracellular soluble enzymes,

and thus represent a departure from the reliance on surface-tethered enzymes (Fig. 3).

On land, the most abundant carbohydrate is plant biomass-derived cellulose, a rich source of glucose in soil and herbivorous habitats. Due to the recalcitrant crystalline structure of cellulose, only specialized species possess the consortium of enzymes required to fully break it down, including oxygen-dependent lytic polysaccharide monooxygenases (LPMOs), cellobiohydrolases (CBHs), and other GH types (Vaaje-Kolstad *et al*., 2010; Horn *et al*., 2012; Østby *et al*., 2020). Although Bacteroidetes are abundant in cellulose-rich environments, and cellulolytic species are known, no PUL from an isolated species has yet been conclusively shown to target cellulose. In a study by Naas *et al*., a putative CelUL containing enzymes with experimentally verified cellulose specificity was identified from a rumen metagenome assembled genome (AC2a) (Naas *et al*., 2014). But the most strongly cellulolytic Bacteroidetes species that have been characterized to date appear to use a completely 'PUL-free' mechanism for cellulose metabolism. The aerobic soil bacteria *Cytophaga hutchinsonii* and *Sporocytophaga myxococcoides* are proficient cellulose degraders, though the enzymatic systems they use are still enigmatic (Zhu and McBride, 2017; Taillefer *et al*., 2018) (Fig. 2). They lack the LPMOs (which are in fact absent from the phylum as a whole), CBHs, and multi-enzyme cellulosomes (Artzi *et al*., 2017) that are typically expected for efficient cellulose depolymerization, and instead appear to rely on T9SS-mediated secretion of large multi-domain enzymes and redundant repertoires of extracellular soluble, membrane-tethered, and periplasmic enzymes (Zhu and McBride, 2017; Taillefer *et al*., 2018) (Fig. 3). Within the anaerobic habitat of the HGM, oxygen-dependent LPMOs are not expected, but it is striking that no functionally similar enzyme activities have yet been uncovered in the few known aerobic cellulolytic soil-dwelling Bacteroidetes. While neither *C. hutchinsonii* or *S. myxococcoides* encode any obvious PULs, *C. hutchinsonii* does possess two SusC/D-like pairs, although the encoding genes are not found in proximity to any CAZymes and their deletion does not impair growth on cellulose (Zhu *et al*., 2015).

Like cellulose, chitin is a highly recalcitrant and abundant crystalline polysaccharide. Instead of being produced by plants, it is abundant in fungal cell walls and arthropod exoskeletons. *Flavobacterium johnsoniae* encodes a PUL (ChiUL) enabling rapid metabolism of chitin (Larsbrink *et al*., 2016). The main chitinase, ChiA, is an unusually large (~160 kDa) multi-modular CAZyme that is secreted from the cells by the T9SS and comprises two catalytic domains with complementary *endo*- and *exo*-activities, separated by an extended chitin-binding domain (Mazurkewich *et al*., 2020). ChiA is the only T9SS-secreted enzyme in this PUL, and the presence of similar multi-catalytic chitinase-encoding genes in syntenic ChiULs from fresh-water and marine species was found to correlate with the ability to grow on crystalline chitin (Larsbrink *et al*., 2016), reflecting the importance of such multi-modular proteins in chitin conversion. Thus, this ChiUL represents a 'hybrid' PUL (Fig. 3), similar to some of the PULs that use T9SS secretion to target algal polysaccharides.

In a similar vein to both the *Fj*ChiUL findings and the described cellulolytic soil bacteria, the recently studied rumen bacterium '*Candidatus Paraporphyromonas polyenzymogenes*' encodes no apparent PULs, but instead relies heavily on large multicatalytic cellulases, several of which are secreted using the T9SS (Naas *et al*., 2018) (Figs. 2 and 3). Furthermore, Naas *et al*. used meta-omics studies to show that such T9SS-dependent 'PUL-free' systems could be important for ruminal deconstruction of cellulose and hemicelluloses.

The type 9 secretion system: driving cellular motility and enzyme secretion

As mentioned above, there are several examples of 'hybrid' PULs that, in addition to the PUL-encoded proteins, also rely on the phylum-exclusive T9SS (Figs. 2 and 3). Additionally, the T9SS is important for the gliding motility system in motile Bacteroidetes species, which relies on the T9SS for secretion of components in a mechanism that uses surface-tethered adhesins linked to intracellular helical tracks and motors (Nakane *et al*., 2013; Kharade and McBride, 2014; McBride, 2019). As an example, disruption of genes coding for proteins involved in gliding motility in *C. hutchinsonii*, which does not rely on PULs, abolished both motility and the ability to grow on cellulose (Zhu and McBride, 2014). The T9SS spans across the entire Bacteroidetes phylum, with the notable exception of the *Bacteroides* genus that dominates the HGM, which lacks the T9SS and the ability to glide (Bacic and Smith, 2008). An exception within the *Bacteroides* genus is *B. salyersiae*, a species that does not glide but where genome analysis indicates the presence of T9SS components (Coyne *et al*., 2014).

Several excellent reviews and articles have recently described the current knowledge of this complex system (McBride, 2019; Gorasia *et al*., 2020a, 2020b). In short, secretion via the T9SS is a two-step process: firstly, an N-terminal signal peptide directs the protein for translocation by the Sec system through the inner membrane into the periplasm, where it folds. Next, a conserved ~70–100 amino acid residue C-terminal domain (CTD) directs the protein for transport through the outer membrane via the T9SS protein complex (Fig. 3) (Gorasia *et al*., 2020a), facilitated by a large pore, typically with concomitant removal of the CTD by a specific peptidase. Cryo-EM studies have shown that SprA, the T9SS pore protein in *F. johnsoniae*, forms a channel with an inner diameter/cavity as large as ~70 Å (Lauber *et al*., 2018), which explains how even very large folded proteins can be translocated to the cell's exterior. Following translocation, the protein may be released from the cell in a freely soluble form, or tethered to the cell surface through a sortase-like mechanism, where the newly formed C-terminal carboxylate is fused to an anionic

lipopolysaccharide that inserts into the membrane (Gorasia *et al*., 2015; McBride, 2019).

Two types of T9SS CTD (A and B; TIGRFAM family annotation TIGR04183 and TIGR04183, respectively) are known, and conceivably they are used for different subsets of proteins (de Diego *et al*., 2016; Lasica *et al*., 2016; Kulkarni *et al*., 2017; Kulkarni *et al*., 2019). Further research is however needed to fully clarify this. The CTDs were originally identified in the Bacteroidetes human pathogen *Porphyromonas gingivalis* as being involved in cell-surface tethering of secreted proteins, a function in line with the typical outer membrane attachment of *endo*-acting PUL enzymes. Figure 3 shows a schematic overview of the T9SS and how it is used to secrete PUL-encoded CAZymes in non-canonical 'hybrid' PULs.

The T9SS complements PULs by permitting the secretion of large enzymes

As more species outside the HGM are being investigated, it is becoming more evident that T9SS-mediated secretion of modular CAZymes is wide-spread in the phylum, but it is not universally utilized for CAZyme secretion in any species. While there are several examples of PUL-encoded CAZymes in soil-dwelling Bacteroidetes that are secreted through this pathway, not all CAZymes (including PUL and non-PUL proteins) from such species are secreted in this way. In addition to the characterized 'hybrid' PUL examples described above, an example of the heavy reliance on the T9SS is a PUL predicted to target chitin and fungal β-glucans that was identified in proteomic analysis of the *Chitinophaga pinensis* secretome; all CAZymes encoded by this PUL possess CTDs for secretion via the T9SS (Larsbrink *et al*., 2017). Other 'classical' PULs of *C. pinensis* have no CTD-tagged enzymes, while others have a mixture of tagged and untagged enzymes (McKee *et al*., 2019). These are key examples of how the PUL system is complemented by the T9SS in many species, with the secretion system bringing additional adaptive flexibility where it is needed. The nature and functional implications of (the connections between) the varying CAZyme secretion mechanisms, polysaccharide-degrading abilities, and gliding motility remain largely unknown. Additional work on non-HGM species is needed to fill these knowledge gaps.

The current view is that for many non-HGM species, the T9SS functions alongside the PUL system, giving additional secretion routes for large proteins, both PUL and non-PUL enzymes. Especially in Bacteroidetes species that lack PULs entirely, the T9SS appears to be crucial for polysaccharide degradation, as was found in proteomic studies of the cellulolytic *C. hutchinsonii* and *S. myxococcoides*, where the majority of *endo*-acting CAZymes detected in the outer membrane or as extracellular proteins were secreted by the T9SS (Taillefer *et al*., 2018). Indeed, *C. hutchinsonii* provides the key example of PUL-free polysaccharide hydrolysis in motile soil-dwelling Bacteroidetes, as it has been shown to use the T9SS to both secrete cellulolytic enzymes and to enable gliding over physical surfaces such as cellulose itself (Zhu *et al*., 2016). Recently, genes coding for all necessary components of the T9SS complex were identified in the genomes of several Bacteroidetes isolated from the digestive tract of the omnivorous cockroach *Periplaneta americana* (Vera-Ponce de León *et al*., 2020), showing another under-explored environment where this secretion system is likely playing a major role in glycan nutrient acquisition. In addition, a recent investigation into soil-derived *Flavobacteria* grown on pectin and pectin components found that certain defined carbon sources stimulated not only CAZyme secretion and SusCD production but also colony spreading on agar plates, showing yet another way that the PUL system and T9SS are intertwined (Kraut-Cohen *et al*., 2021).

Current knowledge on the regulation of PULs

Whether a particular Bacteroidetes species is a general biomass scavenger or is more specialized at deconstructing a particular class of glycans, it is common for their genomes to encode large numbers of discrete PULs (Table 2) (Lapébie *et al*., 2019). In these cases, a 'preference list' for the different polysaccharides ava ilable may come into play and determine which glycans are targeted first. This would be controlled via non-concurrent activation of specific PULs. Such a hierarchical list of substrate preferences has indeed been demonstrated in several cases (Rogers *et al*., 2013; Pudlo *et al*., 2015; Tuncil *et al*., 2017). The sensing of degradation products from glycans that are highly prioritized can even repress transcription of PULs of lower preference. For some PULs in some species, there is likely a balance between activation by early degradation products of the target glycan and repression by the early degradation products from a distal PUL that targets a glycan of higher priority (Pudlo *et al*., 2015). The ranking of different glycans seems to be hard-wired in the genomes of studied species, regardless of whether they are cultured alone or together with other species. High-priority substrates will trigger upregulation of the corresponding PUL even if the cells have been exposed to and are growing on abundant but lower-priority glycans (Rogers *et al*., 2013; Tuncil *et al*., 2017).

These polysaccharide preferences have great implications for the composition of microbial communities and

the temporal variance in competition for various glycans between species. It is fascinating to consider the implications of PUL preferences in a real natural context, where glycans are almost never available in the pure forms in which we study them, but instead are found enmeshed within complex food and biomass material. Early work on the model HGM symbiont *B. thetaiotaomicron* explored the response of this species in mono-colonized gnotobiotic mice to a complex chow diet comprising multiple plant polysaccharides and found that hexose sugars were preferentially liberated and consumed before other glycan moieties, and that host mucus glycans were a 'last reserve' carbon source when dietary glycans were lacking (Bäckhed *et al*., 2005; Sonnenburg *et al*., 2005). These pioneering studies showcase polysaccharide preferences that dictate conversion steps of complex intact biomass and also show that the induction of enzyme production can be achieved even when polysaccharides are bound within a cell wall matrix. Within the marine ecosystem, taxonomically distinct groups of Bacteroidetes and related phyla are enriched as the profile of available phytoplankton-derived organic matter shifts, as reflected in observed changes in the expression profile of genes encoding sulfatases, CAZymes, and PUL-like transporter proteins (Teeling *et al*., 2012). In a clearer example, it has been shown that marine bacteria within a mixed-species particle showed a preference for alginate metabolism even when directly scaffolded onto pectin (Bunse *et al*., 2021). Similar phenomena have been observed in bacterial soil communities actively degrading fungal necromass, where degradative changes to substrate composition drive temporal changes in community composition and CAZyme gene expression profiles (Brabcová *et al*., 2016). In the marine environment, the situation may be further complicated by the tendency of microbes to form physical aggregates with complex interregulation phenomena that are difficult to parse from metabolic investigation (Cordero and Datta, 2016).

In a few cases, PUL-mediated upregulation of genes located outside the PUL in question has been observed, in what we refer to here as 'non-canonical' PULs (Hehemann *et al*., 2012a, b; Ficko-Blean *et al*., 2017). This reliance on distally located accessory genes/proteins for carbohydrate turnover shows how not all PULs are perfectly independent loci. Possibly, the common notion that PULs are discrete loci encoding all necessary functions to deconstruct a specific glycan is a reflection on the strong focus on HGM species thus far. Further complicating the matter, some Bacteroidetes have been shown to use outer membrane vesicles to facilitate glycan depolymerization and cross-feeding between species (Elhenawy *et al*., 2014; Valguarnera *et al*., 2018). For instance, SusG has been shown to be packed into secreted vesicles, which could enable better access for the enzyme to act on starch particles than when the enzyme is locked to the cell surface (Valguarnera *et al*., 2018).

As discussed above, some PULs make very fine distinctions between related polysaccharides with subtle variations in structure, while others show less discernment and appear able to metabolize a relatively broad group of glycan structures. To a great extent, this specificity is regulated via the SusC/D complex and the transcriptional regulator, such that characterization of the ligand-binding specificity of the SusD-like protein produced recombinantly is often taken as an indicator of the PUL target glycan. It has been shown that a *B. thetaiotaomicron* fructan-targeting PUL permits metabolism of inulin instead of levan in certain strains with a variant *susC/D gene* pair (*Joglekar et al*., 2018). In some cases, regulation of PUL activation is instead directed by the elegant orchestration of enzymes with low efficiencies, which prolongs PUL activation. This was demonstrated in an investigation of the metabolism of complex pectin by *B. thetaiotaomicron* and other members of the same genus that use multiple PULs to target different specific pectin components. Specifically, Luis *et al*. showed these species were able to access their target glycan structures within a complex pectin matrix substrate, and that their PULs are functionally regulated by means of carefully controlled enzyme efficiency differences, which ensure that the glycans activating other pectin PULs are not depleted too quickly (Luis *et al*., 2018).

Finally, as we increase exploration into complex and dynamic microbiomes in their native habitat, we are beginning to improve our understanding of PUL regulation at a community level. By combining different -omic technologies, one can link expression of multiple PULs from multiple Bacteroidetes populations simultaneously, which when linked to substrate availabilities, can be used to reconstruct 'food-webs' that depict polysaccharide degradation at a system-wide level. Examples of where multi-omic approaches have been used to monitor PUL expression include the rumen of moose (Solden *et al*., 2018) and the colon of pigs (Michalak *et al*., 2020), which both highlighted specific niche specializations for different hemicellulose fibres. As the resolution of technologies rapidly improve, so will our appreciation of how Bacteroidetes populations deploy their saccharolytic strategies in synergistic and/or competitive contexts.

Are these systems selfish or sharing?

The archetypal SUS employs carbohydrate-binding proteins on the cell surface, allowing *B. thetaiotaomicron* to effectively sequester starch by use of enzyme-appended CBMs, the SusE/F starch-binding proteins, and non-

catalytic substrate binding sites on SusG. Because relatively high molecular weight oligosaccharides are imported through SusC, and subsequent glycan deconstruction into monosaccharides occurs in the periplasm, very low amounts of oligosaccharides escape the cell to feed competitors. Such nutrient-hoarding is referred to as a 'selfish' adaptation as it limits cross-feeding in dense microbial communities and several different examples of this phenomenon exist (Hehemann *et al*., 2010; Cuskin *et al*., 2015; Rogowski *et al*., 2015; Pluvinage *et al*., 2018). Indeed, it is almost considered paradigmatic that SUS-like systems are selfish in this manner, thus boosting the competitiveness of PUL-encoding organisms by helping a species to sequester a bulky polysaccharide like starch, or even more complex carbohydrate-rich particles, close to the bacterial cell surface (Cameron *et al*., 2012). In the landmark study of the *Bt*αManULs described by Cuskin *et al*., it was shown how the weakly acting surface-tethered enzymes of the *Bt*αManULs minimize extracellular polysaccharide cleavage, so that the large fragments generated are rapidly imported, and breakdown to metabolizable monosaccharides occurs almost entirely in the periplasm. Further, the large and complex oligosaccharides generated outside the cell would be inaccessible to several other HGM species in contrast to small ManOs (Cuskin *et al*., 2015). Indeed, in co-culturing experiments using yeast α-mannan as sole carbon source, *B. thetaiotaomicron* did not support the growth of other *Bacteroides* species that can metabolize mannose and ManOs (Cuskin *et al*., 2015). Single-cell microscopy has corroborated these results for *B. thetaiotaomicron* growing on both α-mannan and RGII (Hehemann *et al*., 2019), and similar hoarding of oligosaccharides in the periplasm has been observed among marine Bacteroidetes such as *G. forsetii* metabolizing phytoplankton-derived polysaccharides (Reintjes *et al*., 2017). This adaptation in the marine environment would prevent diffusive loss of oligosaccharides to the environment and provide a competitive advantage for well-equipped species, with potential impact on our understanding of carbon flow in the ocean, which so far has not typically accounted for bacterial sequestration of higher molecular weight glycans (Reintjes *et al*., 2017).

Although the SUS and the *Bt*αManULs are the best-known examples of how PULs can facilitate nutrient hoarding, there are now a number of demonstrations of PUL-mediated cross-feeding among the Bacteroidetes (Rakoff-Nahoum *et al*., 2014; Porter and Martens, 2016; Grondin *et al*., 2017). Whether a PUL confers cross-feeding or not might reflect the balance between the catalytic rates of the key extracellular *endo*-acting enzyme(s) and the rate of sugar transport into the periplasm (Briggs *et al*., 2020). Extracellular XylOs resulting from xylan degradation by *B. ovatus* can cross-feed other gut commensals such as *Bifidobacterium adolescentis* (Rogowski *et al*., 2015), and this sharing of nutrients may be further mediated through secreted enzyme-packed vesicles produced by certain Bacteroidetes species (Valguarnera *et al*., 2018). *B. thetaiotaomicron* is able to metabolize AGP-derived oligosaccharides released by other members of the HGM such as *Bacteroides cellulosilyticus*, which produces AGP-degrading enzymes that complement the activities encoded by *Bt*AGPUL-A and *Bt*AGPUL-B (Cartmell *et al*., 2018). A follow-up study showed that it is specifically the release of β-1,3-linked di- and tri-saccharides of galactose by *B. cellulosilyticus* that permits cross-feeding of other *Bacteroides* and even of *Bifidobacteria* species (Munoz *et al*., 2020).

This phenomenon of PUL-encoded enzymes creating oligosaccharides that become public goods for the greater microbiome has also been observed in the terrestrial soil environment. In these cases, the release of oligosaccharides is often due to the free secretion of depolymerizing enzymes into the extracellular environment. For example, in contrast to the lipid-anchored *endo*-acting enzymes used in many PULs, *F. johnsoniae* secretes ChiA freely into the environment via the T9SS, (Kharade and McBride, 2014). Studies of the related soil species *C. pinensis* showed secretion of a high level of PUL-derived chitinases and β-glucanases in several growth conditions (Larsbrink *et al*., 2017), and that a considerable fraction of the generated oligosaccharides is released to the surroundings (McKee *et al*., 2019). The lower microbial density in soils compared to gut environments may explain the use of freely secreted enzymes in the soil, as opposed to the nutrient-hoarding cell surface-tethered enzymes that are often used in the hyper-competitive HGM ecosystem.

Both selfish and sharing behaviours have been observed in the marine environment. A study of polysaccharide degradation by communities sampled from the Atlantic Ocean found selfish, sharing, and scavenging species (Reintjes *et al*., 2019). The 'selfish' members of the community included many Bacteroidetes, which transported higher molecular weight glycans into the periplasm, likely mediated by PUL systems specific for the substrates tested (laminarin, xylan, chondroitin sulfate, arabinogalactan, pullulan, and fucoidan) (Reintjes *et al*., 2019). The 'sharing' species tended to use cell-surface associated or freely released enzymes for polysaccharide hydrolysis, which released sufficient levels of low-molecular weight oligosaccharides into the environment to support the growth of 'scavenging' species that do not secrete hydrolytic enzymes and cannot import high-molecular weight glycans. Highlighting the consistency of the Bacteroidetes approach to polysaccharide digestion, this three-way interplay between selfish,

sharing, and scavenging species has also been proposed to occur within the HGM where *B. ovatus* displays both selfish and sharing approaches to xylan depolymerization, leading to oligosaccharides that are subsequently assimilated by members of the Actinobacteria, Firmi cutes, and Proteobacteria phyla, in some cases with additional deconstruction of XylOs (Ndeh and Gilbert, 2018).

Another aspect that challenges the view of PULs as species-specific adaptations for selfish nutrient hoarding is the frequency with which genes encoding CAZymes or entire PULs are swapped between species via horizontal gene transfer. In-depth studies have revealed that syntenic PULs are often found in both closely and more distantly related species (Martens *et al*., 2011; Coyne *et al*., 2014; Larsbrink *et al*., 2014b; Reddy *et al*., 2016; Ficko-Blean *et al*., 2017; Déjean *et al*., 2020), suggesting that sharing of PULs through horizontal gene transfer is common in the phylum as a whole. The exact manner in which intact PULs are shared between species, and the apparent fine-tuning of PULs through deletion or incorporation of genes encoding CAZymes and glycan binding proteins, is still not well understood (Martens *et al*., 2014), but there are several notable examples of PUL transfers occurring in different environments. The aforementioned acquisition of a porphyran- and agar-degrading PUL from *Z. galactanivorans* by the Japan-specific human symbiont *B. plebeius* is perhaps the best known example (Hehemann *et al*., 2010). Marine microbes themselves are also able to acquire novel PULs, including from terrestrial species. For example, it was shown that a PUL-like system transferred from terrestrial microbes, comprising a sugar transporter and several glycoside hydrolases, displays pectin-responsive expression and confers pectin degradation abilities to the marine species *Pseudoalteromonas haloplanktis* (Hehe mann *et al*., 2017).

Bacteroidetes PULs and genomes as tools facilitating scientific inquiry

The clustering of CAZyme genes into PULs has not only permitted advances in our understanding of carbohydrate metabolism in the Bacteroidetes phylum and its role in complex ecosystems. The proficiency of the Bacteroidetes in polysaccharide degradation, and the view of PULs as synergistic gene clusters whose CAZyme functions essentially describe the structure of the target polysaccharide, has also enabled estimations of the global diversity of glycans in nature in bioinformatic meta-studies (Lapébie *et al*., 2019). Specifically, the inventory of PULs from marine bacteria is helping to generate an increasingly nuanced view of the structure and abundance of marine polysaccharides (Becker *et al*., 2017; Becker *et al*., 2020; Arnosti *et al*., 2021).

Furthermore, the convenient PUL clustering of synergistic CAZyme genes targeting a particular polysaccharide has facilitated the discovery of new CAZyme functions, which has great implications for both fundamental understanding of microorganisms and enzymes, but also for industry aiming to develop more sustainable methods for conversion of renewable biomass. Luis and Martens recently described a general 'functional microbiology' scheme for identifying novel catalytic functions within the HGM that is based on identifying PULs that are upregulated by specific glycans of interest (Luis and Martens, 2018). As an example, the aforementioned PUL mediating hydrolytic deconstruction of xyloglucan (*Bo*XyGUL) was first identified in a global transcriptomic study as being upregulated during growth on galactoxyloglucan and it was found in the genome of the human gut symbiont *B. ovatus*, while being absent in the genome of *B. thetaiotaomicron* (Martens *et al*., 2011). Pursuant to this observation, a series of knock-out variants of *B. ovatus* were created, to probe the role of the *Bo*XyGUL and some of its components. This work verified that the *Bo*XyGUL was necessary for growth of *B. ovatus* on xyloglucan, strongly suggesting a role in enzymatic digestion of xyloglucan polysaccharides. Through gene cloning, recombinant protein production, and activity determination of all eight enzymes encoded by the XyGUL, a degradative pathway of xyloglucan into monosaccharides could be assembled (Larsbrink *et al*., 2014b).

Following this general approach to PUL discovery and characterization can often lead to the discovery of novel functions embedded in the PUL itself, including novel specificities for known polysaccharide structures, or entirely new activities targeting previously unknown substrate features. As genes co-expressed within a PUL are expected to have complementary activities on one polysaccharide, the PUL's predicted activities can give clues to the activities of other PUL proteins of unknown function (Luis and Martens, 2018). The characterization of two *B. ovatus* PULs upregulated during growth on xylan uncovered novel xylanolytic enzymes from existing GH families not previously known to possess such functionality (Martens *et al*., 2011; Rogowski *et al*., 2015). In another study, a novel acetyl xylan esterase, FjoAcXE, that can deacetylate MeGlcA-substituted xylose residues, was discovered using a PUL-guided search in *F. johnsoniae* (Razeq *et al*., 2018). The other CE in this xylan-targeting PUL was shown to be a multi-catalytic protein with combined acetyl-feruloyl esterase activity (Kmezik *et al*., 2020), contributing to a XylUL that provides the full capacity to deconstruct esterified heteroxylans. Building on the frequent discoveries of novel

CAZyme activities, many new CAZyme structures have been identified from HGM-derived PULs, as recently reviewed by Tamura and Brumer (2021).

Several PUL characterization studies have led to the establishment of entirely new CAZy families (Lombard *et al*., 2014). In the study of pectin-targeting PULs from *B. thetaiotaomicron*, biochemical characterization of PUL-encoded hypothetical genes led to seven new CAZy families being founded, and the observed (novel) enzymatic activities were further used to improve structural descriptions of the RGII polysaccharide (Ndeh *et al*., 2017). For instance, the activity of a newly characterized α-L-rhamnosidase was used to revise the previously accepted identification of a β-linked rhamnose sugar in two of the four established and highly complex side chains common to RGII from all land plants (O'Neill *et al*., 2004). In addition, the characterization of this PUL led to the discovery of a novel side chain of RGII, comprising an arabinosyl moiety appended to a backbone galacturonic acid (Ndeh *et al*., 2017). As another example, a new polysaccharide lyase family was established following the characterization of a number of PUL-derived hypothetical proteins produced when *B. thetaiotaomicron* is grown on AGPs (Munoz-Munoz *et al*., 2017).

Concluding remarks

Investigations into polysaccharide degradation by the Bacteroidetes have so far been dominated by studies of human gut symbiont *Bacteroides* species, and in this context, there has been a great focus on PULs, both in fundamental bacteriology and in directed enzyme discovery efforts. There are now increasing numbers of projects focussing on species from non-gut environments, and a more nuanced vision of polysaccharide degradation by bacteria from this phylum is emerging. Investigations of soil, ruminant, insect, and marine communities have shown that PUL systems are often complemented by a connection to the T9SS, which permits the secretion of large multi-modular enzymes that are released freely into the environment rather than being tethered to the cell surface as seems common in the human intestine. In such 'hybrid' PULs, some, or even all, of the encoded CAZymes may be secreted via the T9SS. Additionally, the genomes of most Bacteroidetes encode several CAZymes that are not part of PULs, and often these represent important activities needed to complement PULs to reach full depolymerization of target glycans. Importantly, while PUL-based mechanisms are common, certain Bacteroidetes appear to rely heavily on a 'PUL-free' approach, deploying T9SS-secreted CAZymes with multiple linked catalytic domains to efficiently depolymerize crystalline structures such as cellulose and chitin. In all studied environments, genes encoding individual CAZymes as well as entire PULs have been traded between species within the habitat, but also between different habitats (e.g. aerobic vs anaerobic), showcasing global and dynamic sharing of metabolic capacity.

In all ecosystems, the energy expended in the production and free release of secreted enzymes must be recouped by metabolism of the target polysaccharide, but there will be a loss of released sugars to cross-feeding species if low-molecular weight hydrolysis products are released. The 'selfish' strategy of cellular adhesion to a bulk substrate and import of high-molecular weight oligosaccharides is abundantly found in the HGM, where it likely provides a strong advantage in a dense and hypercompetitive environment. Selfish systems are also encountered in the ocean, where they will limit the diffusive loss of substrate to the broader environment. Although there are now a plethora of observations of cross-feeding among HGM *Bacteroides*, there is still a general agreement in the literature that polysaccharide deconstruction in the highly competitive HGM is mostly performed in a 'selfish' manner, achieved via the use of multiple cell-surface carbohydrate binding/sequestering proteins, careful control and balancing of enzyme reaction rates to prevent glycan depletion and PUL de-activation, and the import of large oligosaccharides via importer proteins specific to one PUL. Observations to date indicate that a 'sharing' approach, with plentiful release of accessible oligosaccharides to the surrounding environment, may be more common in the terrestrial soil ecosystem, where free release of multi-catalytic CAZymes is common, and may happen in some cases without any obvious PUL-like association that would permit rapid uptake of oligosaccharide products. A 'sharing' approach may however reflect a lower microbial density in this environment, and the impact of such leaky systems on community dynamics in the soil is a hot topic for microbial ecologists.

Metagenomic data are also expanding our understanding of polysaccharide degradation by the Bacteroidetes, such as by uncovering secreted multi-modular CAZymes in ruminant ecosystems, suggesting that many fibre-degrading capabilities remain to be discovered in this environment. Further ecological insights regarding community cross-feeding surely also remain. The data accrued thus far show that the Bacteroidetes are fit to dominate diverse glycan-rich environments, thanks to their specialized PULs, the T9SS, gliding motility, and sophisticated enzyme architectures. As the knowledge of HGM, terrestrial as well as marine species increases, more insight into non-Bacteroidete PULs, either seemingly complete systems or smaller cassettes comprising TonB-dependent transporters (analogous to SusC-like proteins) and a few CAZymes, is obtained. Examples of such PUL-like systems include: laminarin-, alginate-, and

pectin-targeting PULs of *Alteromonas macleodii* (Koch *et al*., 2019); small loci encoding enzymes and TonB-dependent transporters for xyloglucooligosaccharides in *Cellvibrio japonicus* (Larsbrink *et al*., 2014a) and human milk oligosaccharide metabolism in Clostridiales (Pichler *et al*., 2020); and the β-mannan-targeting loci of the Firmicute *Roseburia intestinalis* (La Rosa *et al*., 2019). The prevalence of this broadly beneficial adaptation that ties enzyme production to substrate availability shows that it has relevance for a diversity of species and habitats, and that there may be a multitude of PUL-like systems that have not yet been defined.

References

Arnosti, C., Wietz, M., Brinkhoff, T., Hehemann, J.H., Probandt, D., Zeugner, L., and Amann, R. (2021) The biogeochemistry of marine polysaccharides: sources, inventories, and bacterial drivers of the carbohydrate cycle. *Ann Rev Mar Sci* **13**: 81–108.

Artzi, L., Bayer, E.A., and Moraïs, S. (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat Rev Microbiol* **15**: 83–95.

Atgié, M., Chennevière, A., Masbernat, O., and Roger, K. (2019) Emulsions stabilized by gum arabic: how diversity and interfacial networking lead to metastability. *Langmuir* **35**: 14553–14565.

Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., *et al*. (2021) dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Res* **49**: D523–D528.

Bacic, M.K., and Smith, C.J. (2008) Laboratory maintenance and cultivation of bacteroides species. *Curr Protoc Microbiol*. May; Chapter 13: Unit 13C.1. https://doi.org/10.1002/9780471729259.mc13c01s9.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005) Host-bacterial mutualism in the human intestine. *Science* **307**: 1915.

Bågenholm, V., Reddy, S.K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C.M., *et al*. (2017) Galactomannan catabolism conferred by a polysaccharide utilization locus of *Bacteroides ovatus*: enzyme synergy and crystal structure of a β-mannanase. *J Biol Chem* **292**: 229–243.

Barbeyron, T., Thomas, F., Barbe, V., Teeling, H., Schenowitz, C., Dossat, C., *et al*. (2016) Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium *Zobellia galactanivorans* Dsij T. *Environ Microbiol* **18**: 4610–4627.

Becker, S., Scheffel, A., Polz, M.F., and Hehemann, J.H. (2017) Accurate quantification of laminarin in marine organic matter with enzymes from marine microbes. *Appl Environ Microbiol* **83**. https://doi.org/10.1128/AEM.03389-16.

Becker, S., Tebben, J., Coffinet, S., Wiltshire, K., Iversen, M. H., Harder, T., *et al*. (2020) Laminarin is a major molecule in the marine carbon cycle. *Proc Natl Acad Sci* **117**: 6599.

Benselfelt, T., Cranston, E.D., Ondaral, S., Johansson, E., Brumer, H., Rutland, M.W., and Wågberg, L. (2016) Adsorption of xyloglucan onto cellulose surfaces of different morphologies: an entropy-driven process. *Biomacromolecules* **17**: 2801–2811.

Bolam, D.N., and van den Berg, B. (2018) TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Curr Opin Struct Biol* **51**: 35–43.

Brabcová, V., Nováková, M., Davidová, A., and Baldrian, P. (2016) Dead fungal mycelium in forest soil represents a decomposition hotspot and a habitat for a specific microbial community. *New Phytol* **210**: 1369–1381.

Briggs, J.A., Grondin, J.M., and Brumer, H. (2020) Communal living: glycan utilization by the human gut microbiota. *Environ Microbiol* **23**: 15–35.

Briliūtė, J., Urbanowicz, P.A., Luis, A.S., Baslé, A., Paterson, N., Rebello, O., *et al*. (2019) Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat Microbiol* **4**: 1571–1581.

Bunse, C., Koch, H., Breider, S., Simon, M., and Wietz, M. (2021) Sweet spheres: succession and CAZyme expression of marine bacterial communities colonizing a mix of alginate and pectin particles. *Environ Microbiol*. https://doi.org/10.1111/1462-2920.15536.

Cameron, E.A., Maynard, M.A., Smith, C.J., Smith, T.J., Koropatkin, N.M., and Martens, E.C. (2012) Multidomain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism. *J Biol Chem* **287**: 34614–34625.

Capek, P., Matulová, M., Navarini, L., and Suggi-Liverani, F. (2010) Structural features of an arabinogalactan-protein isolated from instant coffee powder of *Coffea arabica* beans. *Carbohydr Polym* **80**: 180–185.

Cartmell, A., Lowe, E.C., Baslé, A., Firbank, S.J., Ndeh, D. A., Murray, H., *et al*. (2017) How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc Natl Acad Sci* **114**: 7037.

Cartmell, A., Muñoz-Muñoz, J., Briggs, J.A., Ndeh, D.A., Lowe, E.C., Baslé, A., *et al*. (2018) A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation. *Nat Microbiol* **3**: 1314–1326.

Chassard, C., Delmas, E., Robert, C., and Bernalier-Donadille, A. (2010) The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol Ecol* **74**: 205–213.

Cho, K.H., Cho, D., Wang, G.R., and Salyers, A.A. (2001) New regulatory gene that contributes to control of *Bacteroides thetaiotaomicron* starch utilization genes. *J Bacteriol* **183**: 7198–7205.

Cordero, O.X., and Datta, M.S. (2016) Microbial interactions and community assembly at microscales. *Curr Opin Microbiol* **31**: 227–234.

Coyne, M.J., Zitomersky, N.L., McGuire, A.M., Earl, A.M., and Comstock, L.E. (2014) Evidence of extensive DNA transfer between Bacteroidales species within the human gut. *mBio* **5**: e01305–e01314.

Crouch, L.I., Liberato, M.V., Urbanowicz, P.A., Baslé, A., Lamb, C.A., Stewart, C.J., *et al*. (2020) Prominent members of the human gut microbiota express endo-acting O-

glycanases to initiate mucin breakdown. *Nat Commun* **11**: 4017.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E. A., Pudlo, N.A., *et al*. (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**: 165–169.

D'Elia, J.N., and Salyers, A.A. (1996) Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron*. *J Bacteriol* **178**: 7180–7186.

de Diego, I., Ksiazek, M., Mizgalska, D., Koneru, L., Golik, P., Szmigielski, B., *et al*. (2016) The outer-membrane export signal of *Porphyromonas gingivalis* type IX secretion system (T9SS) is a conserved C-terminal β-sandwich domain. *Sci Rep* **6**: 23123.

Déjean, G., Tamura, K., Cabrera, A., Jain, N., Pudlo, N.A., Pereira, G., *et al*. (2020) Synergy between cell surface glycosidases and glycan-binding proteins dictates the utilization of specific beta(1,3)-glucans by human *Bacteroides*. *mBio* **11**: e00095–e00020.

Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., *et al*. (2016) A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**: 1339–1353.e1321.

Dodd, D., Moon, Y.-H., Swaminathan, K., Mackie, R.I., and Cann, I.K.O. (2010) Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic bacteroidetes. *J Biol Chem* **285**: 30261–30273.

Elhenawy, W., Debelyy, M.O., and Feldman, M.F. (2014) Preferential packing of acidic glycosidases and proteases *Bacteroides* outer membrane vesicles. *mBio* **5**: e00909–e00914.

Fan, Y., and Pedersen, O. (2020) Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* **19**: 55–71.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., and Pedrós-Alió, C. (2013) Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* **7**: 1026–1037.

Ficko-Blean, E., Préchoux, A., Thomas, F., Rochat, T., Larocque, R., Zhu, Y., *et al*. (2017) Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat Commun* **8**: 1685.

Fincher, G.B., Stone, B.A., and Clarke, A.E. (1983) Arabinogalactan-proteins: structure, biosynthesis, and function. *Annu Rev Plant Physiol* **34**: 47–70.

Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorev, D.S., Baslé, A., Firbank, S.J., *et al*. (2017) Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**: 407–411.

Glowacki, R.W.P., and Martens, E.C. (2020) If you eat it, or secrete it, they will grow: the expanding list of nutrients utilized by human gut bacteria. *J Bacteriol* **203**. https://doi.org/10.1128/JB.00481-20.

Gorasia, D.G., Veith, P.D., and Reynolds, E.C. (2020a) The type IX secretion system: advances in structure, function and organisation. *Microorganisms* **8**: 1173.

Gorasia, D.G., Veith, P.D., Chen, D., Seers, C.A., Mitchell, H.A., Chen, Y.Y., *et al*. (2015) *Porphyromonas gingivalis* type IX secretion substrates are cleaved and modified by a sortase-like mechanism. *PLoS Pathog* **11**: e1005152.

Gorasia, D.G., Chreifi, G., Seers, C.A., Butler, C.A., Heath, J.E., Glew, M.D., *et al*. (2020b) *In situ* structure and organisation of the type IX secretion system. *bioRxiv*: https://doi.org/10.1101/2020.05.13.094771.

Gow, N., Latge, J., and Munro, C. (2017) The fungal cell wall: structure, biosynthesis, and function. In *Microbiol Spect*. Heitman, J., Howlett, b., Crous, P., Stukenbrock, E., James, T., and Gow, N. (eds.). The Fungal Kingdom. Washington, DC: ASM Press, pp. 267–292. https://doi.org/10.1128/microbiolspec.FUNK-0035-2016.

Gray, D.A., White, J.B.R., Oluwole, A.O., Rath, P., Glenwright, A.J., Mazur, A., *et al*. (2021) Insights into SusCD-mediated glycan import by a prominent gut symbiont. *Nat Commun* **12**: 44–44.

Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., and Brumer, H. (2017) Polysaccharide utilization loci: fueling microbial communities. *J Bacteriol* **199**: e00860–e00816.

Hehemann, J.-H., Kelly, A.G., Pudlo, N.A., Martens, E.C., and Boraston, A.B. (2012a) Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc Natl Acad Sci* **109**: 19786.

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**: 908–912.

Hehemann, J.-H., Reintjes, G., Klassen, L., Smith, A.D., Ndeh, D., Arnosti, C., *et al*. (2019) Single cell fluorescence imaging of glycan uptake by intestinal bacteria. *ISME J* **13**: 1883–1889.

Hehemann, J.-H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., *et al*. (2012b) Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *J Biol Chem* **287**: 30571–30584.

Hehemann, J.-H., Arevalo, P., Datta, M.S., Yu, X., Corzett, C.H., Henschel, A., *et al*. (2016) Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat Commun* **7**: 12860.

Hehemann, J.H., Truong, L.V., Unfried, F., Welsch, N., Kabisch, J., Heiden, S.E., *et al*. (2017) Aquatic adaptation of a laterally acquired pectin degradation pathway in marine gammaproteobacteria. *Environ Microbiol* **19**: 2320–2333.

Henrissat, B., Teeri, T.T., and Warren, R.A.J. (1998) A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Lett* **425**: 352–354.

Horn, S.J., Vaaje-Kolstad, G., Westereng, B., and Eijsink, V. (2012) Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels* **5**: 45.

Huseyin, C.E., O'Toole, P.W., Cotter, P.D., and Scanlan, P.D. (2017) Forgotten fungi - the gut mycobiome in human health and disease. *FEMS Microbiol Rev* **41**: 479–511.

Joglekar, P., Sonnenburg, E.D., Higginbottom, S.K., Earle, K.A., Morland, C., Shapiro-Ward, S., *et al*. (2018) Genetic variation of the SusC/SusD homologs from a polysaccharide utilization locus underlies divergent fructan

specificities and functional adaptation in *Bacteroides thetaiotaomicron* strains. *mSphere* **3**: e00185–e00118.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M., *et al*. (2014) Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes '*Gramella forsetii*' KT0803. *ISME J* **8**: 1492–1502.

Kadam, S.U., Tiwari, B.K., and O'Donnell, C.P. (2015) Extraction, structure and biofunctional activities of laminarin from brown algae. *Int J Food Sci Technol* **50**: 24–31.

Kharade, S.S., and McBride, M.J. (2014) *Flavobacterium johnsoniae* chitinase ChiA is required for chitin utilization and is secreted by the type IX secretion system. *J Bacteriol* **196**: 961.

Kmezik, C., Bonzom, C., Olsson, L., Mazurkewich, S., and Larsbrink, J. (2020) Multimodular fused acetyl–feruloyl esterases from soil and gut Bacteroidetes improve xylanase depolymerization of recalcitrant biomass. *Biotechnol Biofuels* **13**: 60.

Koch, H., Dürwald, A., Schweder, T., Noriega-Ortega, B., Vidal-Melgosa, S., Hehemann, J.-H., *et al*. (2019) Biphasic cellular adaptations and ecological implications of *Alteromonas macleodii* degrading a mixture of algal polysaccharides. *ISME J* **13**: 92–103.

Koropatkin, N.M., and Smith, T.J. (2010) SusG: a unique cell-membrane-associated α-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure* **18**: 200–215.

Kraut-Cohen, J., Shapiro, O.H., Dror, B., and Cytryn, E. (2021) Pectin induced colony expansion of soil-derived Flavobacterium strains. *Front Microbiol* **12**.

Krüger, K., Chafee, M., Ben Francis, T., Glavina del Rio, T., Becher, D., Schweder, T., *et al*. (2019) In marine Bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. *ISME J* **13**: 2800–2816.

Kulkarni, S.S., Zhu, Y., Brendel, C.J., and McBride, M.J. (2017) Diverse C-terminal sequences involved in *Flavobacterium johnsoniae* protein secretion. *J Bacteriol* **199**: e00884–e00816.

Kulkarni, S.S., Johnston, J.J., Zhu, Y., Hying, Z.T., and McBride, M.J. (2019) The carboxy-terminal region of *Flavobacterium johnsoniae* SprB facilitates its secretion by the type ix secretion system and propulsion by the gliding motility machinery. *J Bacteriol* **201**: e00218–e00219.

La Rosa, S.L., Leth, M.L., Michalak, L., Hansen, M.E., Pudlo, N.A., Glowacki, R., *et al*. (2019) The human gut Firmicute *Roseburia* intestinalis is a primary degrader of dietary β-mannans. *Nat Commun* **10**: 905–905.

Lapébie, P., Lombard, V., Drula, E., Terrapon, N., and Henrissat, B. (2019) Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat Commun* **10**: 2043.

Larsbrink, J., and McKee, L.S. (2020) Chapter two - Bacteroidetes bacteria in the soil: glycan acquisition, enzyme secretion, and gliding motility. In *Advances in Applied Microbiology*. Gadd, G.M., and Sariaslani, S. (eds). Cambridge, MA, USA: Academic Press, pp. 63–98.

Larsbrink, J., Thompson, A.J., Lundqvist, M., Gardner, J.G., Davies, G.J., and Brumer, H. (2014a) A complex gene locus enables xyloglucan utilization in the model saprophyte *Cellvibrio japonicus*. *Mol Microbiol* **94**: 418–433.

Larsbrink, J., Tuveng, T.R., Pope, P.B., Bulone, V., Eijsink, V.G.H., Brumer, H., and McKee, L.S. (2017) Proteomic insights into mannan degradation and protein secretion by the forest floor bacterium *Chitinophaga pinensis*. *J Proteomics* **156**: 63–74.

Larsbrink, J., Zhu, Y., Kharade, S.S., Kwiatkowski, K.J., Eijsink, V.G.H., Koropatkin, N.M., *et al*. (2016) A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnol Biofuels* **9**: 260.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., *et al*. (2014b) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**: 498–502.

Lasica, A.M., Goulas, T., Mizgalska, D., Zhou, X., de Diego, I., Ksiazek, M., *et al*. (2016) Structural and functional probing of PorZ, an essential bacterial surface component of the type-IX secretion system of human oral-microbiomic *Porphyromonas gingivalis*. *Sci Rep* **6**: 37708.

Lauber, F., Deme, J.C., Lea, S.M., and Berks, B.C. (2018) Type 9 secretion system structures reveal a new protein transport mechanism. *Nature* **564**: 77–82.

Lidbury, I.D.E.A., Borsetto, C., Murphy, A.R.J., Bottrill, A., Jones, A.M.E., Bending, G.D., *et al*. (2020) Niche-adaptation in plant-associated Bacteroidetes specialisation in organic phosphorus mineralisation. *ISME J* **15**: 1040–1055.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490–D495.

Lowe, E.C., Baslé, A., Czjzek, M., Firbank, S.J., and Bolam, D.N. (2012) A scissor blade-like closing mechanism implicated in transmembrane signaling in a *Bacteroides* hybrid two-component system. *Proc Natl Acad Sci* **109**: 7298.

Luis, A.S., and Martens, E.C. (2018) Interrogating gut bacterial genomes for discovery of novel carbohydrate degrading enzymes. *Curr Opin Chem Biol* **47**: 126–133.

Luis, A.S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., *et al*. (2018) Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. *Nat Microbiol* **3**: 210–219.

Luis, A.S., Jin, C., Pereira, G.V., Glowacki, R.W.P., Gugel, S., Singh, S., *et al*. (2020) A single bacterial sulfatase is required for metabolism of colonic mucin *O*-glycans and intestinal colonization by a symbiotic human gut bacterium. *bioRxiv*. https://doi.org/10.1101/2020.11.20.392076.

Manners, D.J., Masson, A.J., and Patterson, J.C. (1973) The structure of a beta-(1 leads to 3)-D-glucan from yeast cell walls. *Biochem J* **135**: 19–30.

Marcobal, A., Barboza, M., Sonnenburg, E.D., Pudlo, N., Martens, E.C., Desai, P., *et al*. (2011) *Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10**: 507–514.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008) Mucosal glycan foraging enhances fitness and transmission of

a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**: 447–457.

Martens, E.C., Roth, R., Heuser, J.E., and Gordon, J.I. (2009) Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem* **284**: 18445–18457.

Martens, E.C., Kelly, A.G., Tauzin, A.S., and Brumer, H. (2014) The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *J Mol Biol* **426**: 3851–3865.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., *et al*. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**: e1001221.

Mazurkewich, S., Helland, R., Mackenzie, A., Eijsink, V.G. H., Pope, P.B., Brändén, G., and Larsbrink, J. (2020) Structural insights of the enzymes from the chitin utilization locus of *Flavobacterium johnsoniae*. *Sci Rep* **10**: 13775.

McBride, M.J. (2019) Bacteroidetes gliding motility and the type IX secretion system. In *Protein Secretion in Bacteria*. Washington, DC, USA: American Society of Microbiology.

McBride, M.J., Xie, G., Martens, E.C., Lapidus, A., Henrissat, B., Rhodes, R.G., *et al*. (2009) Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol* **75**: 6864.

McKee, L.S., and Brumer, H. (2015) Growth of *Chitinophaga pinensis* on plant cell wall glycans and characterisation of a glycoside hydrolase family 27 β-L-Arabinopyranosidase implicated in arabinogalactan utilisation. *PLOS One* **10**: e0139932.

McKee, L.S., Martínez-Abad, A., Ruthes, A.C., Vilaplana, F., and Brumer, H. (2019) Focused metabolism of β-glucans by the soil Bacteroidetes *Chitinophaga pinensis*. *Appl Environ Microbiol* **85**: e02231–e02218.

Michalak, L., Gaby, J.C., Lagos, L., La Rosa, S.L., Hvidsten, T.R., Tétard-Jones, C., *et al*. (2020) Microbiota-directed fibre activates both targeted and secondary metabolic shifts in the distal gut. *Nat Commun* **11**: 5773–5773.

Mohnen, D. (2008) Pectin structure and biosynthesis. *Curr Opin Plant Biol* **11**: 266–277.

Munoz-Munoz, J., Cartmell, A., Terrapon, N., Baslé, A., Henrissat, B., and Gilbert, H.J. (2017) An evolutionarily distinct family of polysaccharide lyases removes rhamnose capping of complex arabinogalactan proteins. *J Biol Chem* **292**: 13271–13283.

Munoz, J., James, K., Bottacini, F., and Van Sinderen, D. (2020) Biochemical analysis of cross-feeding behaviour between two common gut commensals when cultivated on plant-derived arabinogalactan. *J Microbial Biotechnol* **13**: 1733–1747.

Naas, A.E., Mackenzie, A.K., Mravec, J., Schückel, J., Willats, W.G.T., Eijsink, V.G.H., and Pope, P.B. (2014) Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *mBio* **5**: e01401–e01414.

Naas, A.E., Solden, L.M., Norbeck, A.D., Brewer, H., Hagen, L.H., Heggenes, I.M., *et al*. (2018) "Candidatus Paraporphyromonas polyenzymogenes" encodes multimodular cellulases linked to the type IX secretion system. *Microbiome* **6**: 44.

Nakane, D., Sato, K., Wada, H., McBride, M.J., and Nakayama, K. (2013) Helical flow of surface protein required for bacterial gliding motility. *Proc Natl Acad Sci* **110**: 11145.

Ndeh, D., and Gilbert, H.J. (2018) Biochemistry of complex glycan depolymerisation by the human gut microbiota. *FEMS Microbiol Rev* **42**: 146–164.

Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Baslé, A., Gray, J., *et al*. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**: 65–70.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., and Bertilsson, S. (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**: 14–49.

O'Neill, M.A., Ishii, T., Albersheim, P., and Darvill, A.G. (2004) Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol* **55**: 109–139.

Østby, H., Hansen, L.D., Horn, S.J., Eijsink, V.G.H., and Várnai, A. (2020) Enzymatic processing of lignocellulosic biomass: principles, recent advances and perspectives. *J Ind Microbiol Biotechnol* **47**: 623–657.

Pereira, G.V., Abdel-Hamid, A.M., Dutta, S., D'Alessandro-Gabazza, C.N., Wefers, D., Farris, J.A., *et al*. (2021) Degradation of complex arabinoxylans by human colonic Bacteroidetes. *Nat Commun* **12**: 459–459.

Pérez-Pascual, D., Lunazzi, A., Magdelenat, G., Rouy, Z., Roulet, A., Lopez-Roques, C., *et al*. (2017) The complete genome sequence of the fish pathogen *Tenacibaculum maritimum* provides insights into virulence mechanisms. *Front Microbiol* **8**: 1542.

Phillips, G.O. (1998) Acacia gum (Gum Arabic): a nutritional fibre; metabolism and calorific value. *Food Additiv Contam* **15**: 251–264.

Pichler, M.J., Yamada, C., Shuoker, B., Alvarez-Silva, C., Gotoh, A., Leth, M.L., *et al*. (2020) Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. *Nat Commun* **11**: 3285.

Pluvinage, B., Grondin, J.M., Amundsen, C., Klassen, L., Moote, P.E., Xiao, Y., *et al*. (2018) Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nat Commun* **9**: 1043.

Popper, Z.A., Ralet, M.-C., and Domozych, D.S. (2014) Plant and algal cell walls: diversity and functionality. *Ann Bot* **114**: 1043–1048.

Porter, N.T., and Martens, E.C. (2016) Love thy neighbor: sharing and cooperativity in the gut microbiota. *Cell Host Microbe* **19**: 745–746.

Pruss, K.M., Marcobal, A., Southwick, A.M., Dahan, D., Smits, S.A., Ferreyra, J.A., *et al*. (2021) Mucin-derived O-glycans supplemented to diet mitigate diverse microbiota perturbations. *ISME J* **15**: 577–591.

Pudlo, N.A., Urs, K., Kumar, S.S., German, J.B., Mills, D.A., and Martens, E.C. (2015) Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *mBio* **6**: e01282–e01215.

Pudlo, N.A., Pereira, G.V., Parnami, J., Cid, M., Markert, S., Tingley, J.P., *et al*. (2020) Extensive transfer of genes for

edible seaweed digestion from marine to human gut bacteria. *bioRxiv*. https://doi.org/10.1101/2020.06.09.142968.

Rakoff-Nahoum, S., Coyne, M.J., and Comstock, L.E. (2014) An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr Biol* **24**: 40–49.

Razeq, F.M., Jurak, E., Stogios, P.J., Yan, R., Tenkanen, M., Kabel, M.A., *et al*. (2018) A novel acetyl xylan esterase enabling complete deacetylation of substituted xylans. *Biotechnol Biofuels* **11**: 74.

Reddy, S.K., Bågenholm, V., Pudlo, N.A., Bouraoui, H., Koropatkin, N.M., Martens, E.C., and Stålbrand, H. (2016) A β-mannan utilization locus in *Bacteroides ovatus* involves a GH36 α-galactosidase active on galactomannans. *FEBS Lett* **590**: 2106–2118.

Reintjes, G., Arnosti, C., Fuchs, B.M., and Amann, R. (2017) An alternative polysaccharide uptake mechanism of marine bacteria. *ISME J* **11**: 1640–1650.

Reintjes, G., Arnosti, C., Fuchs, B., and Amann, R. (2019) Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. *ISME J* **13**: 1119–1132.

Rogers, T.E., Pudlo, N.A., Koropatkin, N.M., Bell, J.S.K., Moya Balasch, M., Jasker, K., and Martens, E.C. (2013) Dynamic responses of *Bacteroides thetaiotaomicron* during growth on glycan mixtures. *Mol Microbiol* **88**: 876–890.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., *et al*. (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* **6**: 7481.

Rosewarne, C.P., Pope, P.B., Cheung, J.L., and Morrison, M. (2014) Analysis of the bovine rumen microbiome reveals a diversity of sus-like polysaccharide utilization loci from the bacterial phylum Bacteroidetes. *J Ind Microbiol Biotechnol* **41**: 601–606.

Sam, Q.H., Chang, M.W., and Chai, L.Y. (2017) The fungal mycobiome and its interaction with gut bacteria in the host. *Int J Mol Sci* **18**: 330.

Scheller, H.V., and Ulvskov, P. (2010) Hemicelluloses. *Annu Rev Plant Biol* **61**: 263–289.

Shipman, J.A., Cho, K.H., Siegel, H.A., and Salyers, A.A. (1999) Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. *J Bacteriol* **181**: 7206–7211.

Slavin, J.L., Brauer, P.M., and Marlett, J.A. (1981) Neutral detergent fiber, hemicellulose and cellulose digestibility in human subjects. *J Nutr* **111**: 287–297.

Solden, L.M., Naas, A.E., Roux, S., Daly, R.A., Collins, W. B., Nicora, C.D., *et al*. (2018) Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat Microbiol* **3**: 1274–1284.

Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.-H., Westover, B.P., Weatherford, J., *et al*. (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**: 1955.

Stewart, R.D., Auffret, M.D., Roehe, R., and Watson, M. (2018) Open prediction of polysaccharide utilisation loci (PUL) in 5414 public Bacteroidetes genomes using PULpy. *bioRxiv*. https://doi.org/10.1101/421024.

Synytsya, A., Čopíková, J., Kim, W.J., and Park, Y.I. (2015) Cell wall polysaccharides of marine algae. In *Springer Handbook of Marine Biotechnology*. Kim, S.-K. (ed). Berlin, Heidelberg: Springer, pp. 543–590.

Taillefer, M., Arntzen, M.Ø., Henrissat, B., Pope, P.B., and Larsbrink, J. (2018) Proteomic dissection of the cellulolytic machineries used by soil-dwelling Bacteroidetes. *mSystems* **3**: e00240–e00218.

Tamura, K., and Brumer, H. (2021) Glycan utilization systems in the human gut microbiota: a gold mine for structural discoveries. *Curr Opin Struct Biol* **68**: 26–40.

Tamura, K., Dejean, G., Van Petegem, F., and Brumer, H. (2021) Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota. *J Biol Chem* **296**: 100415.

Tamura, K., Hemsworth, G.R., Déjean, G., Rogers, T.E., Pudlo, N.A., Urs, K., *et al*. (2017) Molecular mechanism by which prominent human gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides. *Cell Rep* **21**: 417–430.

Tamura, K., Foley, M.H., Gardill, B.R., Dejean, G., Schnizlein, M., Bahr, C.M.E., *et al*. (2019) Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cell Mol Life Sci* **76**: 4319–4340.

Tancula, E., Feldhaus, M.J., Bedzyk, L.A., and Salyers, A.A. (1992) Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron*. *J Bacteriol* **174**: 5609–5616.

Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A., *et al*. (2016) Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *mBio* **7**: e02134–e02115.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., *et al*. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608.

Terrapon, N., Lombard, V., Gilbert, H.J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**: 647–655.

Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H.J., and Henrissat, B. (2018) PULDB: the expanded database of polysaccharide utilization loci. *Nucleic Acids Res* **46**: D677–D683.

Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M., and Michel, G. (2011) Environmental and gut Bacteroidetes: the food connection. *Front Microbiol* **2**: 93.

Thomas, F., Barbeyron, T., Tonon, T., Génicot, S., Czjzek, M., and Michel, G. (2012) Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut *Bacteroides*. *Environ Microbiol* **14**: 2379–2394.

Traving, S.J., Thygesen, U.H., Riemann, L., and Stedmon, C.A. (2015) A model of extracellular enzymes in free-living microbes: which strategy pays off? *Appl Environ Microbiol* **81**: 7385.

Tuncil, Y.E., Xiao, Y., Porter, N.T., Reuhs, B.L., Martens, E. C., and Hamaker, B.R. (2017) Reciprocal prioritization to dietary glycans by gut bacteria in a competitive environment promotes stable coexistence. *mBio* **8**: e01068–e01017.

Tuson, H.H., Foley, M.H., Koropatkin, N.M., and Biteen, J.S. (2018) The starch utilization system assembles around stationary starch-binding proteins. *Biophys J* **115**: 242–250.

Vaaje-Kolstad, G., Westereng, B., Horn, S.J., Liu, Z., Zhai, H., Sørlie, M., and Eijsink, V.G.H. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**: 219.

Valguarnera, E., Scott, N.E., Azimzadeh, P., and Feldman, M.F. (2018) Surface exposure and packing of lipoproteins into outer membrane vesicles are coupled processes in *Bacteroides*. *mSphere* **3**: e00559–e00518.

Vera-Ponce de León, A., Jahnes, B.C., Duan, J., Camuy-Vélez, L.A., and Sabree, Z.L. (2020) Cultivable, host-specific Bacteroidetes symbionts exhibit diverse polysaccharolytic strategies. *Appl Environ Microbiol* **86**: e00091–e00020.

Vidal, S., Williams, P., Doco, T., Moutounet, M., and Pellerin, P. (2003) The polysaccharides of red wine: total fractionation and characterization. *Carbohydr Polym* **54**: 439–447.

Vogel, J. (2008) Unique aspects of the grass cell wall. *Curr Opin Plant Biol* **11**: 301–307.

Wallenstein, M.D., and Burns, R.G. (2011) Ecology of extracellular enzyme activities and organic matter degradation in soil: a complex community-driven process. *Methods Soil Enzymol* **9**: 35–55.

Whitaker, R.J. (2018) Abigail salyers: an almost unbeatable force. In *Women in Microbiology*. Whitaker, R.J., and Barton, H.A. (eds). Washington DC: American Society for Microbiology, pp. 243–251.

Wolter, L.A., Mitulla, M., Kalem, J., Daniel, R., Simon, M., and Wietz, M. (2021) CAZymes in Maribacter dokdonensis 62–1 from the Patagonian shelf: genomics and physiology compared to related Flavobacteria and a co-occurring Alteromonas strain. *Front Microbiol* **12**: 628055.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., *et al*. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**: 2074.

Zhu, Y., and McBride, M.J. (2014) Deletion of the *Cytophaga hutchinsonii* type IX secretion system gene sprP results in defects in gliding motility and cellulose utilization. *Appl Microbiol Biotechnol* **98**: 763–775.

Zhu, Y., and McBride, M.J. (2017) The unusual cellulose utilization system of the aerobic soil bacterium *Cytophaga hutchinsonii*. *Appl Microbiol Biotechnol* **101**: 7113–7127.

Zhu, Y., Han, L., Hefferon, K.L., Silvaggi, N.R., Wilson, D.B., and McBride, M.J. (2016) Periplasmic *Cytophaga hutchinsonii* endoglucanases are required for use of crystalline cellulose as the sole source of carbon and energy. *Appl Environ Microbiol* **82**: 4835.

Zhu, Y., Kwiatkowski, K.J., Yang, T., Kharade, S.S., Bahr, C. M., Koropatkin, N.M., *et al*. (2015) Outer membrane proteins related to SusC and SusD are not required for *Cytophaga hutchinsonii* cellulose utilization. *Appl Microbiol Biotechnol* **99**: 6339–6350.