THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Carbohydrate utilization and metabolite production by diverse members of the phylum Bacteroidota

YI-HSUAN LEE

Department of Life Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2026

Carbohydrate utilization and metabolite production by diverse members of the phylum Bacteroidota
YI-HSUAN LEE
ISBN 978-91-8103-488-2

© YI-HSUAN LEE, 2026.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5945
ISSN 0346-718X
https://doi.org/10.63959/chalmers.dt/5945

Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:

Model kits for glycoscience. Artist: Yi-Hsuan Lee

Printed by Chalmers Digitaltryck,

Gothenburg, Sweden 2026

—「上善若水。」

— shàng shàn ruò shuǐ.

Preface

This dissertation serves as partial fulfilment of the requirements for obtaining the degree of Doctor of Philosophy at the Department of Life Sciences at Chalmers University of Technology. The work was primarily funded by Swedish Research Council Formas (Dnr 2021-01079). The PhD studies were carried out between May 2022 and November 2026 under the supervision of Professor Johan Larsbrink, co-supervision of Professor Verena Siewers, and with Professor Carl Johan Franzén as examiner. The work in this thesis was carried out at the Division of Industrial Biotechnology, Department of Life Sciences, at Chalmers University of Technology.

Abstract

The human microbiota consists of a huge number of microorganisms. Countless metabolic pathways within the microbiota help sustain the microbes themselves, but also contribute positive effects to the host, in a highly dynamic and complex interplay. The gut microbiota contains the highest number of microorganisms in mammals, and complex carbohydrates from dietary fiber represent a major nutrient source in this environment. The degradation of carbohydrates in mammalian guts thus relies on species that possess efficient carbohydrate-active enzymes (CAZymes). Some species, particularly from the phylum Bacteroidota, cluster CAZyme encoding genes in so-called polysaccharide utilization loci (PULs), which are large gene clusters that encode the necessary proteins to bind, transport, and break down specific polysaccharides. As a result of carbohydrate metabolism, many gut species are known to produce various profiles of metabolites such as organic acids (OAs) that in turn can be utilized by the host, and these are linked to multiple health benefits. However, whether metabolism of different types of carbohydrates results in production of static or variable OA output remains poorly understood.

In this thesis, I have investigated bacterial consumption of various carbohydrates and the resulting production of extracellular metabolite profiles. Biochemical characterization of individual CAZymes involved in breakdown of complex glycans was additionally done to assess unexpected poor degradation abilities in select species. Lastly, a system to transfer entire PULs between species was established, which enables modification of which carbohydrates a species can grow on. I show through metabolite profiling of species from the *Bacteroides* genus and the *Prevotellaceae* family that different carbohydrates indeed lead to different metabolite profiles both within and across species. Biochemical characterization of α-glucuronidases encoded by the poorly studied species *Paraprevotella clara* showed that while it does not readily grow on glucuronoxylan, it encodes functional xylan-active enzymes. Finally, the developed PUL transfer method enabled hemicellulose conversion in the model species *Bacteroides thetaiotaomicron* as well as deeper studies of a PUL sourced from a metagenome. In summary, this thesis contributes to fundamental knowledge regarding carbohydrate utilization by gut microorganisms, which has implications for improved human gut health.

Keywords: PUL, Bacteroides*,* Prevotellaceae*, organic acid, CAZyme*

List of publications

I. **Lee, Y.H.**, Siewers, V. and Larsbrink, J., 2026. Extracellular metabolite production by *Bacteroides* strains varies depending on carbohydrates used as carbon sources. *Applied and Environmental Microbiology*, 92(8):e0054226. DOI: 10.1128/aem.00542-26.

II. **Lee, Y.H.**, Calzetti, D., Siewers, V. and Larsbrink, J. Carbohydrate consumption and production of extracellular metabolites in select *Prevotellaceae* species. *Manuscript*.

III. **Lee, Y.H.***, Mazurkewich, S.*, Siewers, V. and Larsbrink, J. Biochemical characterization of the α-glucuronidases from the gut bacterium *Paraprevotella clara*. *Manuscript*.

IV. Porter, N.T., Kmezik, C., **Lee, Y.H.**, Siewers, V., Pope, P.B., Koropatkin, N., Martens, E. and Larsbrink, J., 2026. A system for transferring large genetic loci in *Bacteroides* enables hemicellulose utilization in *Bacteroides thetaiotaomicron* and characterization of a locus from an uncultivated strain. *Applied and Environmental Microbiology*, *92*(5):e00176-26. DOI: 10.1128/aem.00176-26.

Contribution summary

I. I performed the experimental work, analyzed the data, interpreted and visualized the results, wrote the original draft, and finalized the manuscript together with my co-authors.

II. I performed the majority of the experimental work, analyzed the data, interpreted and visualized the results, wrote the original draft, and finalized the manuscript together with my co-authors.

III. *Shared first author. I performed the experimental work together with my co-author, participated in data analysis, interpretation, validation, and visualization, wrote the original draft, and finalized the manuscript together with my co-authors.

IV. I performed parts of the experimental work, participated in partial data analysis, interpretation, validation, and visualization, and wrote parts of the manuscript, and finalized the manuscript together with my co-authors.

Abbreviations

AA	Auxiliary activities
ABC-transporter	ATP-binding cassette transporter
ATP	Adenosine triphosphate
att site	Integrase-recognized attachment site (in *Bacteroides*)
BLASTP	Basic local alignment search tool — protein
BGX	Beech wood glucuronoxylan
CAZy	Carbohydrate-Active enZymes database
CAZyme	Carbohydrate-active enzyme
CBM	Carbohydrate-binding module
CE	Carbohydrate esterase
CelUL	Cellulose utilization locus
ChiUL	Chitin utilization locus
ECF-σ/anti-σ pair	Extracytoplasmic function σ/anti- σ pair
EMP pathway	Embden-Meyerhof-Parnas pathway
FUdR	Floxuridine (5-fluoro-2'-deoxyuridine)
GH	Glycoside hydrolase
GI tract	Gastrointestinal tract
GlcA	Glucuronic acid
GalNAc	*N*-acetylgalactosamine
GlcNAc	*N*-acetylglucosamine
GMUL	Galactomannan utilization locus
GX	Glucuronoxylan
gpPUL	Gram-positive polysaccharide utilization locus
HMM	Hidden Markov model
HPLC	High-performance liquid chromatography
HPAEC-PAD	High-performance anion-exchange chromatography with pulsed amperometric detection
HTCS	Hybrid two-component system
LPMO	Lytic polysaccharide monooxygenase
MAG	Metagenome-assembled genome
MFS	Major facilitator superfamily
MLG	Mixed-linkage glucan
MLGUL	Mixed-linkage glucan utilization locus
NADPH	Reduced nicotinamide adenosine dinucleotide phosphate
NBU2	Nonreplicating *Bacteroides* Unit 2
NMR	Nuclear magnetic resonance
OA	Organic acid

PDB	Protein Data Bank
PEP	Phosphoenolpyruvate
PL	Polysaccharide lyase
PPi	Inorganic pyrophosphate
PPP	Pentose phosphate pathway
PUL	Polysaccharide utilization locus
PULDB	Polysaccharide Utilization Loci DataBase
qPCR	Quantitative polymerase chain reaction
RG-I	Rhamnogalacturonan type I
RG-II	Rhamnogalacturonan type II
RNA	Ribonucleic acid
rt-PCR	Real-time-polymerase chain reaction
SCFA	Short-chain fatty acid
SGBP	Surface glycan-binding protein
Sus	Starch utilization system
tdk	Thymidine kinase gene
XyGUL	Xyloglucan utilization locus
(4*O*Me)GlcA	4-*O*-methyl-glucuronic acid

Table of Contents

Preface i
Abstract ii
List of publications iii
Contribution summary iii
Abbreviations iv
Table of Contents vi
Chapter 1: Introduction 1
1.1 Questions and specific aims of the thesis 1
Chapter 2: Microbiota 5
2.1 Natural and anthropogenic microbiota 5
2.2 Human microbiota 6
2.2.1 Human gut microbiota 8
2.2.2 Microbial metabolite formation 9
2.2.3 Carbohydrate catabolism in Bacteroidota species 10
2.2.4 “Good” or “bad” microbes? 11
Chapter 3: Carbohydrates representing a major nutrient source for microbiota 13
3.1 Carbohydrates: sources and structures 13
3.1.1 Cellulose 14
3.1.2 Hemicelluloses 14
3.1.3 Pectins 15
3.1.4 Plant storage polysaccharides: starch and fructans 15
3.1.5 Non-plant-derived polysaccharides 16
3.2 Carbohydrate-Active enZymes 17
3.2.1 Glycoside hydrolases 19
3.3 Polysaccharide utilization loci 21
3.3.1 Polysaccharide utilization loci in the Bacteroidota phylum 24
3.4 Substrate preferences in carbohydrate degradation 26
Chapter 4: Original work 29

4.1 Substrate-directed extracellular metabolite production in Bacteroidota species 29
4.1.1 Metabolites produced by *Bacteroides* species from carbohydrate utilization 31
4.1.2 Substrate probing for metabolic composition of the understudied *Prevotellaceae* species 33
4.1.3 Future directions for metabolite output linked to carbohydrate utilization 37
4.2 Characterization of α-glucuronidases from *Paraprevotella clara* 38
4.2.1 Biochemical investigation of GH67 and GH115 α-glucuronidases from *Paraprevotella clara* 40
4.2.2 Biological roles of α-glucuronidases encoded by *Paraprevotella clara* 42
4.3 Development of a genetic toolkit for interspecies PUL transfers 45
4.3.1 Probing the limits of PUL transfer 46
4.3.2 Optimization of PUL transfer efficiency 49
Chapter 5: Conclusions and outlook 51
Acknowledgements 55
References 59

Chapter 1: Introduction

The gut microbiota is the community of trillions of mostly anaerobic microorganisms living predominantly in the distal/lower gastrointestinal tract of animals. This community feeds largely on host mucin and dietary fibers that the host cannot digest, and often the gut microbiota is dominated by species from certain phyla, such as Bacteroidota and Bacillota (formerly Firmicutes) (1). Bacteroidota species, for instance, can utilize a broad machinery for deconstruction of complex carbohydrates in dietary fibers and then ferment the released monosaccharides. Many of these species dedicate a significant portion of their genomes to carbohydrate degradation. Especially in Bacteroidota species, genes related to carbohydrate degradation are often organized in dedicated, co-regulated gene clusters known as poly-saccharide utilization loci (PULs), that encode the proteins needed for glycan-binding, transport, and degradation of a specific substrate (2, 3). The subsequent fermentation of released monosaccharides produces organic acids (OAs) such as acetate, propionate, and succinate, which matter both for the host, as energy sources and is affecting inflammatory parameters linked to health (4-6), and for the microbial community, where one species' products can feed others and shape competition and cooperation (7). The fiber structure can steer which acids are made (8), but such studies are done on complex consortia which limits molecular understanding of the process. Knowing the finer details of carbohydrate utilization within the gut, from the species and enzymes involved in breakdown to the resulting metabolite output, could help explain how diet affects health and potentially contribute to developing so-called precision nutrition (9-11).

1.1 Questions and specific aims of the thesis

This thesis investigated how different types of carbohydrates are metabolized into OAs by Bacteroidota species from both the human gut and other environments, spanning from the commonly studied genus *Bacteroides* to the less studied family *Prevotellaceae* (represented by the genera *Prevotella*, *Paraprevotella*, and *Alloprevotella*). The overarching aim was to generate new knowledge on carbohydrate utilization in these species. This was pursued along three complementary lines of work: (i) phenotypic profiling of carbohydrate utilization and metabolite output across Bacteroidota strains, using a diverse set of defined substrates, from mono-, to oligo- and polysaccharides.; (ii) assessing the annotation of putative PUL-encoded enzymes in a poorly studied bacterium, which did not conform to its predicted carbohydrate degradation potential; (iii) engineering of interspecies PUL transfer to test whether genetic loci can confer new degradative or metabolic capabilities in recipient strains. Linked to these studies, three main questions guided the work:

Question 1: Do different carbohydrates lead to different metabolic outputs in Bacteroidota species?

As mentioned previously, some bacteria possess dedicated gene clusters for carbohydrate utilization, known as PULs. Detailed studies of how PULs and their encoded enzymes can confer different degradative abilities have been reported for many species, especially in the species *Bacteroides thetaiotaomicron* (3, 12). Furthermore, many species thought to play important roles in polysaccharide breakdown in human guts remain understudied, partly due to difficulties in culturing.

Bacteroidota species are known to produce different extracellular metabolites, especially OAs. However, whether this output is a fixed, species-specific trait or instead varies depending on the specific carbohydrate consumed — even among species considered "generalist" degraders — has remained largely untested beyond single-substrate characterization, which typically has relied on glucose (13-15). To address this question, I grew several Bacteroidota species, including species both from the well-studied genus *Bacteroides* and the less studied family *Prevotellaceae*, on a panel of carbon sources ranging from monosaccharides to complex polysaccharides, and the resulting extracellular OAs were quantified using chromatography techniques to compare the metabolite profiles across species and substrates. The results are reported in **Paper I** and **Paper II**, and collectively show substrate-dependent trends of extracellular OA production.

Question 2: Can biochemical enzyme characterization explain unexpected lack of carbohydrate degradation in a single species?

In **Paper II**, I observed that the carbohydrate utilization of several *Prevotellaceae* species was restricted to simple sugars, despite them encoding many putative carbohydrate-active enzymes (CAZymes) and multiple PULs. This led to the question whether these enzymes were actually functional, or if the lack of growth could be explained by other reasons. I chose to focus on the species *Paraprevotella clara*, which encodes many enzymes and PULs expected to be involved in the degradation of the polysaccharide xylan, which this species could not grow on in my experiments. Xylan is a major hemicellulose in plant cell walls, and consequently also in dietary fiber, for instance in cereals (16, 17). Its β-1,4-linked xylose backbone is often decorated with side chains such as arabinose or 4-*O*-methyl-glucuronic acid ((4*O*Me)GlcA), which restrict access for backbone-cleaving xylanases. Efficient xylan degradation therefore needs accessory enzymes to remove backbone decorations, such as α-glucuronidases (18, 19). *P. clara* encodes α-glucuronidases from both glycoside hydrolase family 67 (GH67) and GH115, and these were produced and biochemically characterized, to assess whether they are active enzymes that can liberate (4*O*Me)GlcA from glucuronoxylan, and should play a role in xylan metabolism in this species. The results are presented in **Paper III**, where we show that the enzymes are functional and have different activities and

specificities, and thus should endow *P. clara* with part of the enzymes needed for efficient xylan conversion.

Question 3: Can PULs be functionally transplanted across *Bacteroides* species?

In Bacteroidota species, PULs confer the ability to degrade specific carbohydrates, and often highly similar PULs can be found in many species, both closely and more distantly related. These observations indicate that PULs can somehow be shared between species, and there is support for horizontal gene transfer enabling such sharing (20). To be able to mimic PUL sharing in controlled settings would enable modulation of carbohydrate degradation abilities of chosen recipient species, but while such examples exist in literature, they are scarce (21). We therefore developed a new system to facilitate PUL transfers between species, and used the well-studied species *B. thetaiotaomicron* as the model system. Our new plasmid system, named pICKUP, was successfully shown to enable site-specific genomic integration of two PULs into *B. thetaiotaomicron*: the mixed-linkage glucan utilization locus (MLGUL) from the related species *Bacteroides ovatus*, and a putative cellulose-targeting utilization locus (CelUL) derived from an uncultivated rumen bacterium. The results are covered in **Paper IV**, where we, in addition to showing that the PUL transfers also resulted in altered phenotypes in the recipient strains, also conducted a series of genetic modifications to better understand PUL function and regulation in the new host.

Chapter 2: Microbiota

Having the prefix “micro-” originating from the Greek term “μικρο (mikro)” as "small" and the root “-biota” from “βιοτή (biotē)” and related to “βίωμα (bíōma)”, representing “life", “way of life”, or "lived experience", the two parts together could give “the small things alive" or “the small (short) experience lived". The “small living things”, known as “micro-organisms” to scientists, can be found everywhere, from broad open environments to narrow closed spaces (22). The term “microbiota”, however, scientifically refers to the collection of living microorganisms – bacteria, archaea, fungi, and protists – inhabiting a defined environment, whereas “microbiome” more broadly encompasses these organisms together with their genomes, structural elements, metabolites, and the surrounding environmental conditions that shape them (23, 24). The two terms are often used interchangeably in the literature but strictly describe “the community” versus “the community plus its genetic and functional context”.

2.1 Natural and anthropogenic microbiota

Microbial communities in any environment, whether soil, water, or the mammalian gut, are shaped fundamentally by energy acquisition. Among them, the need of an electron transfer chain to gain adenosine triphosphate (ATP) distinguishes microbial energy metabolism into respiration and fermentation. Respiration requires molecules as an external terminal electron acceptor to generate ATP, which is oxygen in aerobic respiration, or other molecules such as nitrate, sulfate, or fumarate in anaerobic respiration. In fermentation ATP is instead produced by substrate-level phosphorylation, which does not require an external electron acceptor. Microorganisms also response differently to oxygen. Facultative anaerobes can switch between respiratory and fermentative metabolism depending on the conditions. Aerotolerant anaerobes grow fermentatively and tolerate some oxygen. Obligate anaerobes are strictly sensitive to oxygen (25).

Another important classification, specifically for bacteria, is Gram staining (Gram-positive versus Gram-negative) which reflects differences in cell envelope structure. Gram-positive bacteria have a thick peptidoglycan wall, whereas Gram-negative bacteria have a thin one enclosed between the inner membrane and a lipopolysaccharide (LPS)-containing outer membrane, with consequences for antibiotic susceptibility and immune recognition (26, 27).

Since carbohydrates are energy-rich and abundant molecules, polysaccharide-degrading microorganisms can be found throughout the environment (2). While fungi are often regarded as the main decomposers of complex biomass in nature, the focus of this thesis is instead on different types of bacteria, which also play important roles in the global carbon cycle. For instance, the soil microbiota is typically dominated by aerobic and facultative bacterial taxa including Actinomycetota, Pseudomonadota, and soil-dwelling Bacteroidota, and these

species can be major decomposers of plant litter, which requires the ability to break down cellulose, hemicellulose, and/or lignin-associated polysaccharides (28-31). Marine Bacteroidota species similarly specialize in degrading algal polysaccharides such as laminarin, alginate, and fucoidan during phytoplankton blooms, using PULs or PUL-like gene clusters closely related in architecture to those found in gut residents (32, 33). More discussion on PULs and their architecture will be covered in Chapter 3. Extremophiles that utilize terrestrial carbohydrates can also be found in environments of extreme conditions, such as *Sulfolobus solfataricus*, a thermoacidophilic archaeon found in solfataric fields (34). Airborne microbiota, by contrast, are typically transient and much lower in biomass than soil or aquatic communities, largely comprised of mixtures of cells and spores dispersed from soil, water, plant, and animal surfaces instead of a resident, actively metabolizing community. Rather than a site of substantial *in situ* polysaccharide degradation, air is generally understood as a dispersal medium for microorganisms, though molecular evidence for potential metabolic activities from airborne bacteria was reported (35).

In industrial settings such as anaerobic digesters, mixed anaerobic consortia are deliberately cultivated. These often include species that are able to break down lignocellulosic biomass or food waste into fermentable carbohydrates and biofuel precursors from Bacillota with contributing proportion from Bacteroidota and Fibrobacterota (formerly Fibrobacteres or informally Fibrobacteraeota) (36, 37).

Functionally comparable communities are found in the rumen and hindgut of mammalian farm animals, which host dense polysaccharide-degrading anaerobic microbiota that resemble engineered anaerobic digesters in their substrate specificity and fermentation product profile (38). This analogy can be stretched to human gut microbiology, as the distal gut is a densely populated, oxygen-poor niche dominated by obligate and facultative anaerobes that rely on anaerobic fermentation rather than aerobic respiration to extract energy from nutrients; anaerobic respiration activities can be found in this part of the colon as well, but is not discussed in the scope of this thesis (39-41).

2.2 Human microbiota

The gastrointestinal tract (GI tract) of animals hosts diverse and often highly dense microbial communities, from insects to mammals (42, 43). Depending on the diet of the host, these communities face different nutritional exposures. In this thesis, the main focus is the microorganisms associated with humans. These species come from different taxonomic ranks, from phylum down to individual species and strains, with relevant examples such as the genera *Bacteroides* and *Prevotella* in Bacteroidota phylum and the genera *Lactobacillus* and *Staphylococcus* in Bacillota phylum. The different evolutionary origins and cellular structures of these organisms are also reflected in which parts of the body they colonize; for

example, the skin, oral cavity, vagina, and GI tract each harbor distinct and adapted communities (44, 45).

The skin is the largest organ for humans, but it is a nutrient-poor and acidic habitat that is rather dry. Yet it hosts a diverse microbiota comprised by commensals that interact with each other and with the host immune system, such as *Staphylococcus epidermidis* and *Cutibacterium acnes* (46). An organ that presents a structurally distinct, yet comparably constrained environment to the human GI tract is the vaginal microbiome. The vaginal microbiota is characterized by a *Lactobacillus*-dominated, strongly acidic milieu (pH ~4 - 5) with conserved lactic acid microbial functions, in clear contrast to the compositionally diverse, near-neutral colonic community (45, 47, 48). Debatably, while early studies proposed low-biomass bacterial colonization of placental tissue (49), a subsequent contamination-controlled analysis of 537 placentas have failed to substantiate the existence of placenta microbiome (50), which lends greater support to the “sterile womb" model, in which meaningful fetal microbial exposure begins at or around birth rather than *in utero*.

The human GI tract is an example of how the microbiome composition varies substantially not only between anatomical sites, but also within a single organ system; see Figure 1 (51, 52). The oral cavity, where chewing mechanically breaks down the ingested food, chewed into a so-called bolus, hosts a diverse microbiota of over 600 common species-level taxa, with different subsets dominating distinct sites like the teeth, tongue, and tonsils (53). The esophagus then leads to the stomach, which is described as harboring a bacterial community despite its acidity with phyla spanning largely across Pseudomonadota (formerly Proteobacteria), Bacillota, Actinomycetota (formerly Actinobacteria), Bacteroidota, and Fusobacteria (54). The food transit time to transform from the state of bolus to chyme in the upper GI tract is rather short, especially compared to the long residence time in the lower counterpart, where fermentation takes place and nurtures a dense population of micro-organisms, which will be discussed in Chapter 2.2.1 (1, 45, 48).

The luminal pH changes markedly along the GI tract's length – from the oral cavity to the highly acidic stomach (changing from pH ~6.7 to ~0.3), through the rapidly neutralizing duodenum (changing from pH ~4.8 to ~7.0), to the near-alkaline terminal ileum (changing from pH ~6.4 to ~8.2) before dropping again in the fermentation-active proximal colon and rising once more towards the rectum (48). This physicochemical gradient, combined with the residual undigested dietary substrate that reaches each segment, and the residence time available for microbial growth (only a few hours in the small intestine to up to a day in the colon), collectively conditions which and how much microbial populations can be established and maintained. The microbial communities then in turn determine which fermentation metabolites like short-chain fatty acids (SCFAs), OAs, and gases predominate locally.

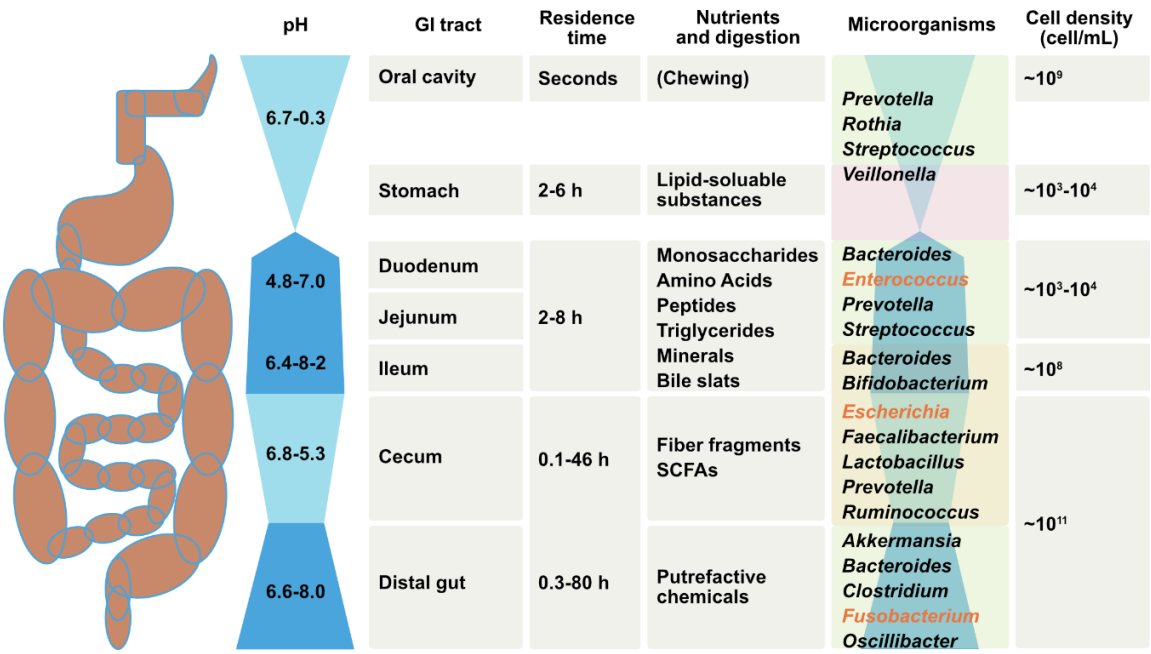


Figure 1. A scheme for digestive processes and micro-ecology along the GI tract. Entering from the oral cavity and exiting from the anus at the end of distal gut, the ingested bolus is gradually processed and broken down to utilizable molecules that facilitate human nutrient absorption. Short-chained oligomers like oligosaccharides from amylase degradation in saliva, peptides and triglycerides by a combination of protease, pepsins, HCl, and bile salts are mostly absorbed in the beginning of small intestines, duodenum, and jejunum. The recalcitrant fibers remain intact and serve as the colonic luminal contents, based on which microorganisms co-develop into a unique microbiome. The long residence time of food enables longer processing time of the microorganisms, where average transit times have been reported to be around a full day (24 h). While the cecum is not used in humans, it plays an important digestive role in many animals. The common bacteria in microbiomes of each part along GI tract are listed in a transient fashion from site to site, where the potential pathobionts are color-coded in orange. (44, 45, 48)

2.2.1 Human gut microbiota

The lower human GI tract, particularly the colon, hosts the most dense and diverse microbial community in the body, dominated by two bacterial phyla: the Gram-negative Bacteroidota and the Gram-positive Bacillota. Bacteroidota includes genera such as *Bacteroides*, *Parabacteroides,* and *Prevotella*, while Bacillota includes species from genera *Faecalibacterium* and *Roseburia*. Additionally, species from other phyla such as Actinomycetota, Pseudomonadota, and Verrucomicrobiota (formerly Verrucomicrobia) are typically found in significant proportions (1, 45, 55-60). The relative composition of these groups is highly individual and shaped by diet, geography, host genetics, and early-life colonization events, and this composition can shift (61-63).

In many studies, the impact of diet on microbiota composition is strongly emphasized (64, 65). In a report from de Filippo et al. (2010), microbiota in children in rural regions on a high-fiber diet had significantly more Bacteroidota and fewer Bacillota species than European children (66). Within the same phylum, dietary habits can define microbial composition as well. For instance, a relative enrichment of *Bacteroides* species is often found in individuals consuming the so-called Western style diet – fiber-poor and fat/protein-rich – which contrasts an enrichment of *Prevotella* species in populations consuming non-Western, i.e., fiber-rich and plant-based diets. This has become a widely cited example of diet-microbiota co-adaptation (67-69). Other gut taxa outside Bacteroidota are still recognized as efficient carbohydrate degraders, like *Ruminococcus bromii* from Bacillota, which is a keystone degrader of resistant starch found in the human colon. In co-culture, it has been shown to boost resistant starch usage by other starch-degrading species, such as *B. thetaiotaomicron* (70). *Bacteroides* species are known as broad glycan-degrading generalists that can turn to host mucin-*O*-glycan foraging when plant-derived dietary fibers are scarce, allowing them to persist on fiber-poor diets whereas species from the *Prevotellaceae* family are more tightly linked to specific plant-polysaccharide degradation. The apparent lack of a fallback capacity for non-plant glycan degradation, makes their abundance more strictly coupled with continuous high-fiber dietary habits (71, 72).

2.2.2 Microbial metabolite formation

Gut anaerobes derive energy largely through fermentation of dietary components remaining undigested by the host-encoded enzymes, where a rich portion of the digesta in gut is composed of complex carbohydrates. Gut bacteria can convert imported carbohydrates through glycolysis. For hexoses, one of the catabolism routes is entering via the Embden-Meyerhof-Parnas (EMP) pathway. For pentoses like xylose, via the pentose phosphate pathway (PPP) it generates pyruvate and reducing equivalents, where both lead to downstream catabolism after pyruvate formation (73-75). Because oxidative phosphorylation is limited or absent under the anaerobic conditions of the colon, pyruvate and its derivatives are instead routed through various downstream fermentative branches, to illuminate, via mixed-acid fermentation (76), the succinate-propionate pathway (77), or reductive pathways to butyrate production (78, 79). The collection of fermentation processes maintains the redox balance of the cell and can be coupled to excretion of the resulting OAs as end products.

The metabolic end products from fermentation, most notably SCFAs and OAs, can further be absorbed by the host or neighboring microbial cells and serve a myriad of biological functions. For instance, the OAs from fermentation have been reported to be beneficial to hosts by serving as energy substrates for colonocytes, signaling molecules that influence host metabolism and the immune system, and also serve as substrates for cross-feeding among microbial community members (5, 80, 81). The specific profile and ratio of OAs produced

in the gut are orchestrated by which microorganisms are present, what substrates they have access to, and how their respective central carbon metabolism is wired and interconnected within the consortium, making metabolite output a property of the whole microbial community rather than a fixed trait of any single species *in vivo*, one of the themes this thesis centers via its original work.

2.2.3 Carbohydrate catabolism in Bacteroidota species

In *B. thetaiotaomicron*, this entry route of EMP pathway has been inferred from genome annotation and homology to canonical glycolytic enzymes, including glucokinase, phosphoglucose isomerase, phosphofructokinase, and fructose-bisphosphate aldolase that process hexoses through fructose-6-phosphate and fructose-1,6-bisphosphate to phospho-enolpyruvate (PEP) and further pyruvate (82). Unlike the downstream OA-producing branches, the individual steps between glucose and PEP have not been confirmed by targeted gene deletion (13). Mannose is thought to converge on the same route at fructose-6-phosphate following conversion by a mannose-6-phosphate isomerase, though this step also remains annotation-based rather than genetically validated. The transcriptional regulators controlling the transporters and catabolic operons responsible for hexose uptake and utilization have been mapped through comparative genomic reconstruction across *Bacteroides* species (73).

The profile of products resulting from carbohydrate fermentation by Bacteroidota species commonly includes acetate, propionate, butyrate, succinate, lactate, and formate (13). These compounds primarily serve to maintain the cellular redox balance, and they are excreted from the cells (7, 15, 83). However, the secretion of these end products is not merely a passive overflow of metabolism but is itself subject to regulation. A previous nuclear magnetic resonance (NMR)-based metabolomic study showed that *B. thetaiotaomicron* reduced its secretion of OAs and amino acids when environmental acetate concentration was elevated, suggesting an internal feedback mechanism that leans to energy conservation rather than continuous export of fermentation products (84).

Beyond their role as terminal metabolic products, catabolic intermediates can also confer benefits to the bacterial cell that extend past energy storing and redox balance. Most metabolic studies in Bacteroidetes have typically used only glucose as carbon source. The previously mentioned examples of NMR-based exometabolomic study on *B. thetaiotaomicron*, monitoring the secreted OAs and amino acids (84) as well as OA synthesis pathways investigated by gene deletion (13) have likewise been performed in defined minimal medium supplemented with glucose. Studies that compare several carbohydrates are fewer, where one instance can be found in a *Phocaeicola vulgatus* (formerly *Bacteroides vulgatus*) study, which characterized its growth and metabolite production using a range of mono-, oligo-, and polysaccharides as carbon source (85). In this study, the different carbohydrates did result in differences such as lag phase length or final culture density

measured by optical density, while no major differences were observed in the extracellular metabolite output profiles among various single-substrate cultures. The OA profile mainly consisted of acetate, succinate, and lactate. Overall, previous studies are in many aspects limited, since they have covered only a small part of carbohydrate-induced metabolite variation of Bacteroidota, which will be addressed more in the original work of this thesis in Chapter 4.

2.2.4 "Good" or "bad" microbes?

As binary as many scientific questions seem to be, whether a given microbe is pathogenic or not is not a fixed biological property, but a context-dependent outcome (86, 87). Many microorganisms behave as harmless or even beneficial commensals in their native niche but can cause disease if they translocate to a sterile site, or in cases where the host is immuno-compromised, or if community balance is disrupted, which is referred to as dysbiosis (88, 89). Thus, pathogenicity is best understood along a spectrum ranging from strict pathogens to opportunistic pathogens to mutualists, rather than as a binary label (90-94).

The proportions of which members constitute the microbiota depends on many factors, such as which nutrients are present. The study from de Filippo et al. (2010) is a great example, where species from the *Prevotellaceae* family were sustained through high-fiber diet in the studied group, accompanied by a higher SCFA production compared to the low-fiber diet individuals (66). The activities of the community then result in diverse mixtures of metabolites, and many metabolites from human-dwelling microorganisms have been linked in observational and mechanistic studies to specific health outcomes. Lactic acid in different body parts, for instance, can be both positive and negative. In the vaginal micro-biome, the vaginal lactobacilli acidify the vagina, which is associated with protection against infection. With the loss of lactate-producing strains, the mixed SCFAs increase, accompanied by bacterial vaginosis (95). Lactic acid produced by oral streptococci, on the other hand, contributes to formation of dental caries (96). In the gut, butyrate is widely recognized as protective for colonic epithelial integrity and is reduced in some inflammatory bowel disease patients (59, 97, 98); propionate has been associated with beneficial effects on host lipid and glucose metabolism (99), while excess microbial succinate and certain bile acids or trimethylamine-derived metabolites have instead been implicated in metabolic dysfunction and cardiovascular disease risk (100, 101). The level of lactic acid producing bacteria that provide probiotic properties in the human gut was also correlated with sulfate-reducing bacteria, which generate toxic hydrogen sulfide (H_2S) that can be harmful for the gut cell lining (102). This illustrates how the identity and balance of microbial metabolites, not just the microbial community composition itself, is directly relevant to host health.

Chapter 3: Carbohydrates representing a major nutrient source for microbiota

Dietary carbohydrates, particularly those human enzymes cannot digest, such as non-starch polysaccharides, represent the principal carbon currency and, by extension, the energy sources for the colonic microbiota (2, 3). The efficiency and specificity with which different bacterial taxa can capture, degrade, and convert these substrates are major determinants of the microbial community composition, nutrient pools, and the resulting spectrum of fermentation products provided to the host (7, 103).

In general, bacterial polysaccharide digestion requires a coordinated set of components. Extracellular or surface-associated enzymes initiate the breakdown of intact polymers by recognition of them as substrates, followed by cleavage. Transport systems then import the resulting shorter fragments across the cell membrane(s), and, subsequently, the intracellular enzymes complete depolymerization of substrate fragments into monomers that can enter central metabolism; see Chapter 2.2.2 (2, 104). Since the fermentation then leads to production of various OAs or other molecules that the host can benefit from, harboring diverse sets of carbohydrate-degrading gut microorganisms indirectly enables the host to utilize dietary fiber nutrients to a non-negligible extent.

The degradation capacities in microorganisms differ widely for catabolizing various polysaccharides, and hierarchical preferences of substrates can exist, which will be discussed in Chapter 3.4 (105). This capacity is greatly determined by CAZymes encoded in their genomes, but also factors such as the cellular location of these proteins (104). Some species are equipped to attack complex and recalcitrant polysaccharides such as resistant starch, xylans, or pectins (105), whereas others feed on simpler forms of carbohydrates or on oligosaccharides released by primary degraders in the same community (70). The microbiota can therefore comprise species occupying different nutritional niches, with primary degraders, secondary degraders, and cross-feeders sharing and shaping the nutrient pools (83). In essence, it is the constituent carbohydrates that provide the conditions for a nutrient niche to build upon. In the following sections, carbohydrate properties and what enables microbial carbohydrate catabolism are discussed.

3.1 Carbohydrates: sources and structures

In humans and other mammals, complex carbohydrates mainly enter the gut through dietary fibers, a heterogeneous group of polysaccharides and oligosaccharides largely derived from plant cell walls. These include cellulose and hemicelluloses, plant storage polysaccharides like starch and fructans, and other sources including fungal and algal glycans. While starch is mostly digested by host enzymes in the small intestine, other polysaccharides resist human digestion and hence reach the colon essentially intact. Even certain forms of starch, labelled

resistant starch, are recalcitrant to degradation by human enzymes (83, 106, 107). Altogether, the undigested polysaccharides pass to the large intestine as dietary fiber containing luminal content.

Complex carbohydrates vary enormously in structure, as a result of them comprising various monosaccharide building blocks, glycosidic linkages, degree of branching, and chemical decorations, such as acetylation, methylation, and feruloylation (108, 109). This structural diversity is why a diverse and abundant gut microbiome, possessing a variety of enzymatic functions, is necessary for the human host to indirectly utilize dietary fiber (2, 11, 110, 111). Gut Bacteroidota species can meet this challenge with PULs, which they use to degrade the fibers for subsequent fermentation into SCFAs and other OAs that the host can absorb and use (3, 111). The following sections describe the structure, sources, and properties of the main polysaccharides found within dietary fiber, as well as the enzymatic machineries used to process them, which will be covered in Chapter 3.2 and Chapter 3.3.

3.1.1 Cellulose

Cellulose is the main structural component of the plant cell walls of all land plants and is also produced by many algae and certain bacteria. Structurally, it is a simple linear homo-polysaccharide of D-glucose units, linked by β-1,4-glycosidic bonds; see Figure 2. Despite this structural simplicity, individual cellulose chains associate through extensive intra- and inter-chain hydrogen bonding and van der Waals interactions, packing into highly ordered crystalline microfibrils with less-ordered amorphous regions. The dense crystalline packing renders native cellulose insoluble in water and highly resistant to enzymatic and chemical hydrolysis. Its efficient degradation requires the specialized, often complementary sets of cellulolytic enzymes like endoglucanases, cellobiohydrolases, lytic polysaccharide mono-oxygenases (LPMOs), and β-glucosidases in organisms capable of using it as a carbon source (112, 113).

The relevance of cellulose as a nutrient source in the human gut is rather limited, though cellulose is considered a major component of dietary fiber. Humans lack native enzymes to degrade cellulose, and only some gut microorganisms, like cellulosome-producing *Ruminococcus* species (111), are able to utilize them. In the human GI tract lumen, cellulose in practice passes through the gut mostly intact. The dietary polysaccharides described in the coming subsections are rather the ones that gut microorganisms can actually utilize.

3.1.2 Hemicelluloses

Hemicelluloses are a structurally diverse group of plant cell wall polysaccharides that, unlike cellulose, are not crystalline, but instead they form hydrogen bonds with the surface of cellulose microfibrils, effectively coating and cross-linking them within the wall matrix. They share a common β-1,4-linked backbone, but differ substantially in backbone sugar

composition and decorations, giving rise to distinct classes including xylan, xyloglucan, mannan, galactomannan, and mixed-linkage β-glucans that have a β-1,3/1,4-linked backbone (Figure 2) (16).

The dietary sources of these polysaccharides vary by type. Xylans are enriched in cereal grains and bran; xyloglucan in fruits and vegetables; mannans in seeds; mixed-linkage β-glucans are found specifically in grasses and cereals such as barley and oat (114-116). Their structural variability directly translates into them having variable physicochemical properties, where the hemicelluloses range from highly water-soluble and viscous – such as arabinoxylan, xyloglucan, and galactomannan, to poorly soluble and tightly cell wall associated, like glucuronoxylan in hardwood (16). Correspondingly, the diversity in structure is reflected by the range of CAZymes and PULs that gut Bacteroidota species have evolved to target hemicelluloses (117).

3.1.3 Pectins

Pectins are a family of structurally complex, acidic polysaccharides that form a major component of the primary cell wall and middle lamella in dicotyledonous plants. They are particularly abundant in fruits such as citrus and apple as well as root vegetables like sugar beet. Unlike cellulose, pectins are not a single defined polymer but a composite of several related and often connected polysaccharides. The main types are: homogalacturonan, a linear backbone of α-1,4-linked D-galacturonic acid that can be variably methyl-esterified and acetylated; rhamnogalacturonan I (RG-I), featuring a repeating galacturonic acid-rhamnose backbone decorated with arabinan and galactan side chains; rhamnogalacturonan II (RG-II), a highly complex branched structure containing several rare sugars, which is often referred to as the most structurally complex carbohydrate known. Pectins are typically viscous and gel-forming. For gut microbes, as for hemicelluloses, the structural complexity of pectins demands correspondingly elaborate multi-enzyme systems comprising diverse pectin/pectate lyases and glycoside hydrolases for complete depolymerization (109, 118-120).

3.1.4 Plant storage polysaccharides: starch and fructans

Starch is the main storage polysaccharide of plants and the most abundant digestible carbohydrate in the human diet. It is a mixture of two α-glucan types: amylose, linear chains of α-1,4-linked glucose (Figure 2), and amylopectin, structures in which α-1,6 linkages create branch points for the α-1,4-linked amylose chains every 20-30 residues. Starch is packed into semi-crystalline granules whose size, crystallinity and amylose-to-amylopectin ratio depend on the botanical source (121). Host α-amylases made by salivary glands, ptyalin, initiates bolus digestion in the oral cavity; α-amylases made by the pancreas continue the digestible starch digestion in the small intestine. Depending on granule structure, crystallinity, and food processing, the resistant fraction escapes the previous digestion and reaches the colon, where it is fermented by the microbiota (106).

Fructans, on the contrary, are fructose-based polymers that both plants and some bacteria can use as storage and protective carbohydrates, and for which humans innately produce no enzymes to digest (107). The two main types are inulin, with β-2,1 linkages, and levan, with β-2,6 linkages (Figure 2) (122). Inulin can be found in chicory, Jerusalem artichoke, and onion in nature, while levan is mostly synthesized by bacteria, with a few plants like timothy grass is also capable of producing it.

3.1.5 Non-plant-derived polysaccharides

Beyond the plant cell wall, the human gut microbiota also encounters a range of structurally distinct polysaccharides of animal, algal, fungal, and host origin, many of which represent important nutrient niches (12). Several starch-like α-glucans share the same glucose building blocks but are from other origins and differ in linkage and/or branching. Glycogen has the same general structure as starch, but is much more branched and serves as an animal and bacterial storage polymer (123). Pullulan, a fungal polysaccharide, consists of maltotriose units joined by α-1,6 linkages (124); and dextrans, which are produced by bacterial transglycosylases (glucansucrases), are built up mainly from α-1,6 linkages (125).

Animal-derived glycosaminoglycans such as chondroitin sulfate (Figure 2) and heparin/heparan sulfate consist of sulfated, uronic-acid-containing disaccharide repeats and occur in connective tissue and cartilage; desulfation is thus a major activity for organisms targeting these polysaccharides as a nutrient source (126, 127). Marine red algae also contain sulfated polysaccharides, in the form of galactans, carrageenan and porphyran, which have backbones of alternating galactose units (128). Ingested fungal biomass furthermore provides fungal cell wall components, such as chitin, β-glucans, and mannans (129). The host itself contributes mucin-*O*-glycans, which are heavily glycosylated Ser/Thr-rich proteins with attached chains of galactose, *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), fucose, *N* -acetylneuraminic acid, which are continuously produced by goblet cells at the mucosal surface independent of diet (130, 131). Collectively, these non-plant glycans broaden the nutritional niches that different gut species may occupy, beyond typical plant-derived dietary fiber. Whether an organism can utilize a particular glycan is, however, dependent on the enzymes and related proteins it can produce, as it governs the types of substrates and linkages that it can efficiently cleave.

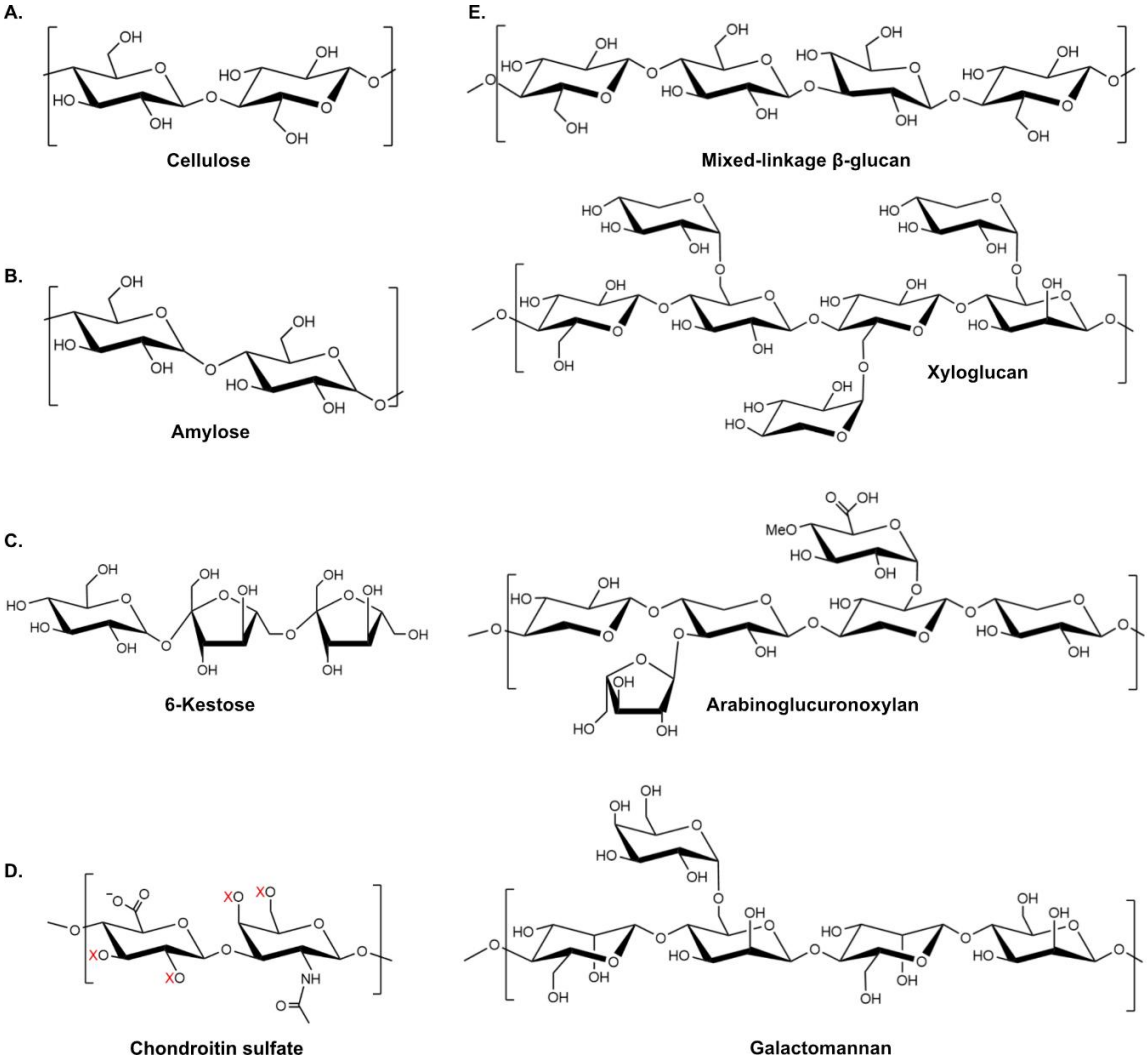


Figure 2. Chemical structures of glycans that are commonly present in human diets. A. Cellulose (β-1,4-glucan). **B.** Amylose (α-1,4-glucan). **C.** 6-Kestose, the shortest levan-type fructo-oligosaccharide (β-2,6-fructan). **D.** Chondroitin sulfate (repeating β-1,4-GlcA-β-1,3-GalNAc). X – sulfation; labelled in grey. **E.** Xylans: Mixed-linkage β-1,3/1,4-glucan, xyloglucan (β-1,4-glucan backbone with α-1,6-linked xylose branches), arabinoglucuronoxylan (β-1,4-xylan backbone with α-1,2-linked (4*O*Me)GlcA and α-1,2/1,3-liked arabinofuranose branches), and galactomannan (β-1,4-mannan backbone with α-1,6-linked galactose branches).

3.2 Carbohydrate-Active enZymes

While dietary proteins are built from amino acids that are joined by a uniform type of peptide bond, degraded by a set of proteases, e.g., pepsin and trypsin (132); ester bonds in dietary fats are hydrolyzed mainly by lipases that release fatty acids from a glycerol backbone (133), dietary fibers that are not naturally digestible for humans are comprised of many different

monosaccharides with a large variety of linkages, branches, anomeric configuration (α/β), and substitutions. To reach their complete breakdown thus requires a far broader repertoire of specific bond-cleaving enzymes to orchestrate. These various enzymes that modify carbohydrates are referred to as CAZymes, and these are classified into different functional classes and families in the Carbohydrate-Active enZymes database (CAZy; https://www.cazy.org/) (134)). As classes, CAZy includes glycoside hydrolases (GHs), which cleave glycosidic bonds by hydrolysis; polysaccharide lyases (PLs), which cleave glycans via β-elimination; carbohydrate esterases (CEs), which hydrolyze ester-linked decorations such as acetyl or feruloyl groups; glycosyltransferases (GTs), which synthesize glycosidic bonds, and auxiliary activities (AA), which are redox enzymes that act alongside other CAZymes. Additionally, non-catalytic proteins with carbohydrate-binding abilities, carbohydrate-binding modules (CBMs), are listed as an additional associated class.

CAZy grew out of a classification of GHs into families based on amino acid sequence similarity, which was introduced by Henrissat in 1991 after it was noted that sequence-based families can group enzymes of different substrate specificities (135, 136). The scheme was extended with new families starting from 1993, and by 1995 it covered more than 45 GH families (137, 138). The classification was integrated into an online resource, the first iteration of CAZy, between 1998 and 1999, and the same family-based approach was later applied to include GTs, PLs, CEs, and CBMs (135, 139), whereas AAs were added in 2013 (140). The database is continuously curated and now links sequence families to function, structure, and the literature (134).

This classification method for CAZymes is purely based on amino acid sequence similarity, which does not necessarily correlate with substrate specificity or reaction mechanism, though within the vast majority of families the mechanism is shared for all members. A family number is assigned once a set of sequences is shown to share enough sequence identity to be grouped together, and the numbering itself simply follows the chronological order in which each family was established (134, 136). For instance, GH family 5 (GH5), comprising many cellulases (141), and GH6, comprising mainly cellobiohydrolases (142), are not evolutionarily closely related and not functionally related just because they are numerically adjacent.

As sequence-based grouping does not guarantee a single substrate specificity, several large GH families have since been subdivided into subfamilies that can provide insights for into their biological roles as a group. For example, GH43 contains dozens of subfamilies spanning β-xylosidase, α-L-arabinofuranosidase, and endo-xylanase activities, despite sharing the same five-bladed β-propeller fold (143); GH5 likewise has around 50 sub-families, with several β-glycan-cleaving activities (141); GH2, though best known for β-galactosidase/β-mannosidase activity, also contains subfamily members active on xylo-oligosaccharides

(144). These cases illustrate why family assignment is often insufficient to predict an enzyme's exact substrate, and why subfamily division is often needed, but of all, the biochemical evidence of specificity provides the highest degree of confidence in an activity assignment.

Bacteroidota gut species often encode an unusually large and diverse CAZyme repertoire, with often around 3-4% of the genome devoted to such enzymes (145). GHs as the main enzymes responsible for hydrolytically cleaving glycosidic bonds that are stable and abundant in various glycans, currently there are almost 200 GH families listed in CAZy (134). Without GH-based catalysis, a typical *O*-glycosidic bond would have a half-life of millions of years (146). GHs help overcome this barrier, accelerating bond cleavage by many orders of magnitude, and hence play huge roles in underpinning carbohydrate catabolism, remodeling, and signaling across all kingdoms of life.

3.2.1 Glycoside hydrolases

Most GHs follow one of two classical Koshland mechanisms, both involving two catalytic carboxylate residues (Figure 3) (147). Inverting enzymes act in a single displacement step, with the carboxylates about 10 Å apart to accommodate a water molecule. Retaining enzymes act by double displacement through a covalent glycosyl-enzyme intermediate, with the carboxylates only about 5.5 Å apart. The two steps in the catalysis results in the anomeric configuration of the product being retained relative to the substrate (148). Beyond the mechanism, substrate specificity and mode of action are governed by details of the active-site architecture more than by the global fold (148). The substrate binding site decides whether an enzyme cleaves within a chain (endo-) or from its end (exo-), and the sugar-binding subsites are numbered relative to the cleaved bond (149). A few bacterial enzymes cleave glycosidic bonds by non-Koshland routes involving oxidation/reduction and elimination/hydration steps (147).

Xylans, being a major hemicellulose in plant cell walls, provide humans with miscellaneous sources of dietary fibers of different monosaccharide composition, linkages, and decorations (117). Xylanase activities are thus found wherever organisms degrade it, and these are widespread among fungi and bacteria (150-153). By strict definition, xylanases are endo-1,4-β-xylanases (EC 3.2.1.8) that cleave the internal β-1,4-xylosidic bonds of the xylan backbone, producing xylo-oligosaccharides of varying length. To exemplify, GH10 enzymes under this Xylans, being a major hemicellulose in plant cell walls, provide humans with miscellaneous sources of dietary fibers of different monosaccharide composition, linkages, and decorations (117). Xylanase activities are thus found wherever organisms degrade it, and these are widespread among fungi and bacteria (150-153). By strict definition, xylanases are endo-1,4-β-xylanases (EC 3.2.1.8) that cleave the internal β-1,4-xylosidic bonds of the xylan backbone,

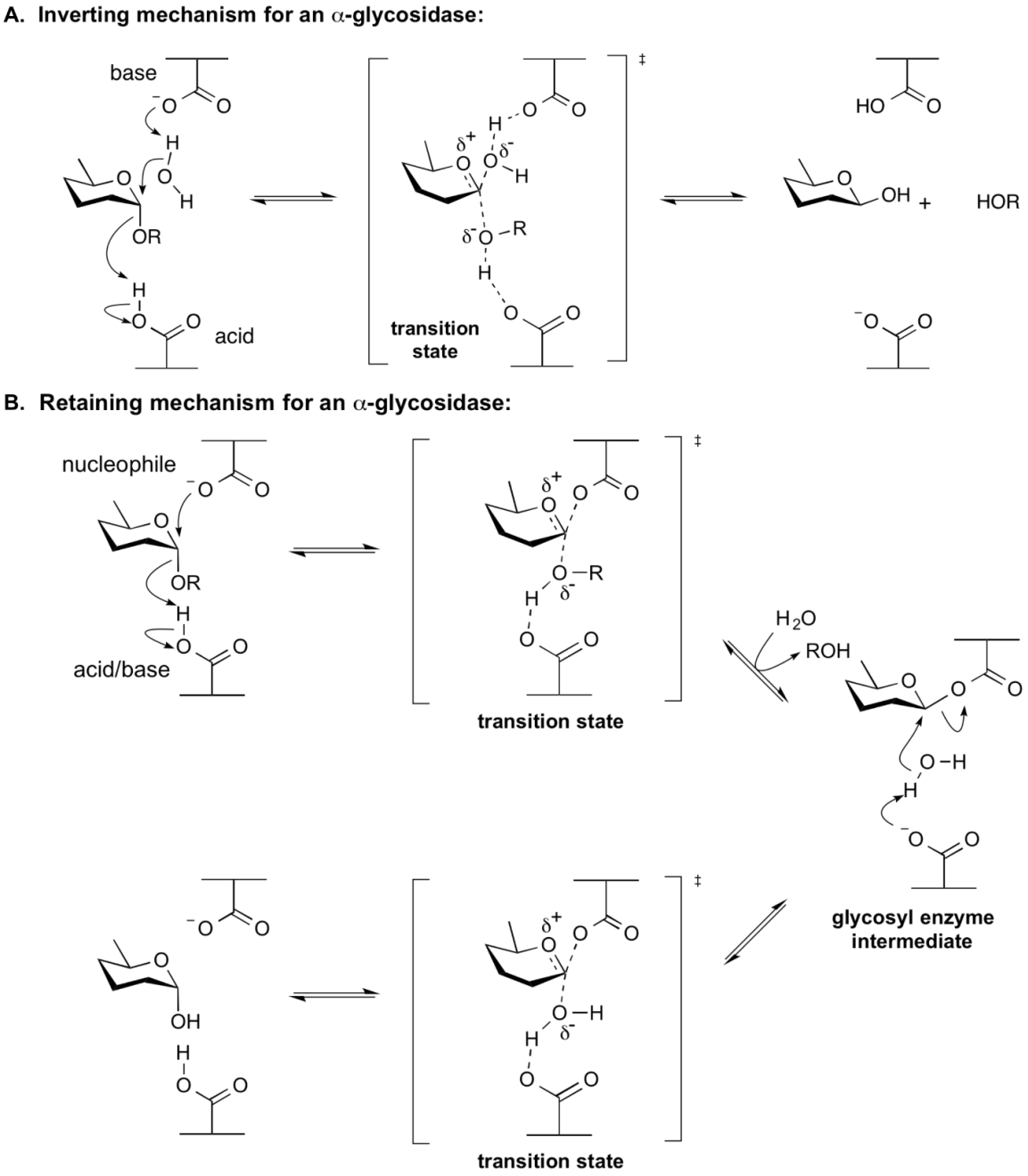


Figure 3. Mechanisms of glycoside hydrolases. A. Mechanism of an inverting α-glycosidase. The general base makes a nucleophilic attack on the water molecule, which in turn hydrolyzes the glycosidic bond, inverting the anomeric configuration. **B.** Mechanism of a retaining α-glycosidase. The nucleophile attacks the anomeric center, creating a covalent adduct glycosyl enzyme intermediate and a deprotonated acid/base. In the second step, the deprotonated carboxyl acts as a base and assists a water molecule in the hydrolysis of the intermediate. This creates the hydrolyzed product with a net retention of the anomeric configuration. Figure adapted from CAZypedia under an open access license (154). Accessed 2026-10-21.

producing xylo-oligosaccharides of varying length. To exemplify, GH10 enzymes under this category have a $(\beta/\alpha)_8$-barrel fold and are found in plants, fungi, and bacteria, while the other main endo-xylanase family, GH11, contains smaller enzymes with a β-jelly-roll fold, and these occur only in fungi and bacteria (152, 155). Xylanase activities can meanwhile also be found in other families, including GH5, GH7, GH8, GH30, and GH43, as well as a group of surface xylanases from *B. ovatus* include GH10, GH30 and GH98 members (117, 152).

As in nature, homoxylan rarely occurs in the cell walls of land plants (156). The xylan backbone from beech wood glucuronoxylan (BGX), to illustrate, carries decorations such as glucuronic acid (GlcA), (4*O*Me)GlcA, arabinose, and acetyl groups. Prior to xylan backbone cleaving, for some endo-xylanases such as those from GH10, the decorations pose as steric hindrance in the active sites, thus restricting binding of the main chain and limiting cleavage. Removing the decoration enable xylanase activities, and as a whole, xylan degradation requires a cohort of side-chain-cleaving enzyme activities (18), such as α-arabinofuranosidases from GH43, GH5, and GH146, α-glucuronidases from GH67 and GH115, and acetyl and feruloyl esterases in CE families 1 and 6 (108).

3.3 Polysaccharide utilization loci

Together with proteins of roles such as substrate sensing, capturing, and importing, the genes encoding CAZymes that share proximal/relevant substrates are often co-regulated within their genome. Their enzyme activities are co-organized genetically rather than scattered individual genes, forming their intricate multi-gene systems. This co-regulated locus allows efficient cellular carbohydrate utilization and is the essence of PULs.

The first PUL described in the literature was the starch utilization locus (Sus) of *B. thetaiotaomicron* from the Salyers lab in the late 1980s (157). The Sus locus contains eight genes, from *susA* to *susG* with a regulator gene, *susR.* The encoded proteins together bind, cleave, and import starch and its oligosaccharides (malto-oligosaccharides) in a coordinated fashion. The process starts with SusG, which is a surface-exposed, outer-membrane-anchored GH13 α-amylase that both binds and cleaves starch (104), and it is assisted by the starch/maltooligosaccharide binding abilities of the other outer membrane lipoproteins, SusE and SusF. Next, SusD, an outer membrane lipoprotein, binds and directs malto-oligosaccharides into SusC, an outer membrane TonB-dependent transporter. The captured substrate is then imported into the periplasm. SusC and D work in sync, and were reported to be a transporter pair anchoring the outer membrane protein complex in a “pedal bin” mechanism, where the SusD lid opens to bind the glycan and then closes it into the SusC barrel, and the inner-membrane TonB system powers the import via the proton motive force (158-160). More GHs await in the periplasm, where the neopullulanase SusA (GH13) and α-glucosidases SusB (GH97) fully cleave oligosaccharides to glucose. The Sus genes are thought to be expressed at a low basal level, but once maltose or malto-oligosaccharides are

present in the periplasm, they are sensed by the periplasmic portion of the regulator SusR, which leads to upregulation of the whole locus and hence rapid starch degradation (160).

When the *B. thetaiotaomicron* genome was sequenced, it was shown to contain dozens of *susC*/*susD*-like pairs, pointing to existence of many Sus-like systems (161), which were later named PULs (82). The *susC/susD*-like pair thus became the standard marker for finding PULs in Bacteroidota species because they are encoded as an adjacent pair in virtually every PUL. Most PULs encode their own regulator, making each PUL a self-contained unit with its own sensor, enzymes, and transporter, switched on by the glycan it targets (2). In Sus, SusR binds maltose or malto-oligosaccharides and activates transcription of *susA-G* (158). Many other PULs also use periplasmic-sensing hybrid two-component system (HTCS), an extracytoplasmic function σ/anti-σ pair (ECF-σ/anti-σ pair), or a SusR-like one-component regulator (162, 163). In each case, low basal expression only produces a small amount of oligosaccharides. When they bind the regulator, the regulator switches on then transcription of the whole locus. This typical setting of having a homologous SusC/SusD pair, regulators, sensor/transcription factor systems, CAZymes and sometimes sulfatases is also what defined a canonical PUL structure (Figure 4) (2, 104).

Essentially, PULs are a carbohydrate uptake solution confined to Gram-negative phyla by architecture, since the defining transport-and-binding pair encoded by PULs operates in the outer bacterial membrane, a structural component Gram-positive bacterium simply do not possess (104). Gram-positive gut bacteria, by contrast, lack SusC/SusD-type systems and instead rely on extracellular free enzymes, cell-surface-anchored enzymes, or multi-enzyme cellulosome-like complexes to degrade polysaccharides outside the cell before importing smaller fragments inwards (73, 163-165).

Whether using freely detached or surface-anchored enzymes, glycan degradation strategies can vary considerably among bacterial phyla. Yet all strategies exist to balance degradation efficiency against the risk of "leaking" valuable sugars to competing microbes in the same environment, which at the same time shapes the community ecosystem (166). The polysaccharide fragments released by one strain's degradation can in turn be utilized by other strains, creating cross-feeding networks within the community. *Ruminococcus bromii*, to illustrate, is a great example of a keystone degrader of resistant starch in the human colon, releasing oligosaccharides that are subsequently used by other species (70). Meanwhile, the unique substrate accessibilities of CAZymes that different microorganisms harbor to deploy leads to the development of their nutritional niche in gut microenvironment, which is directly shaped by individual’s dietary habits.

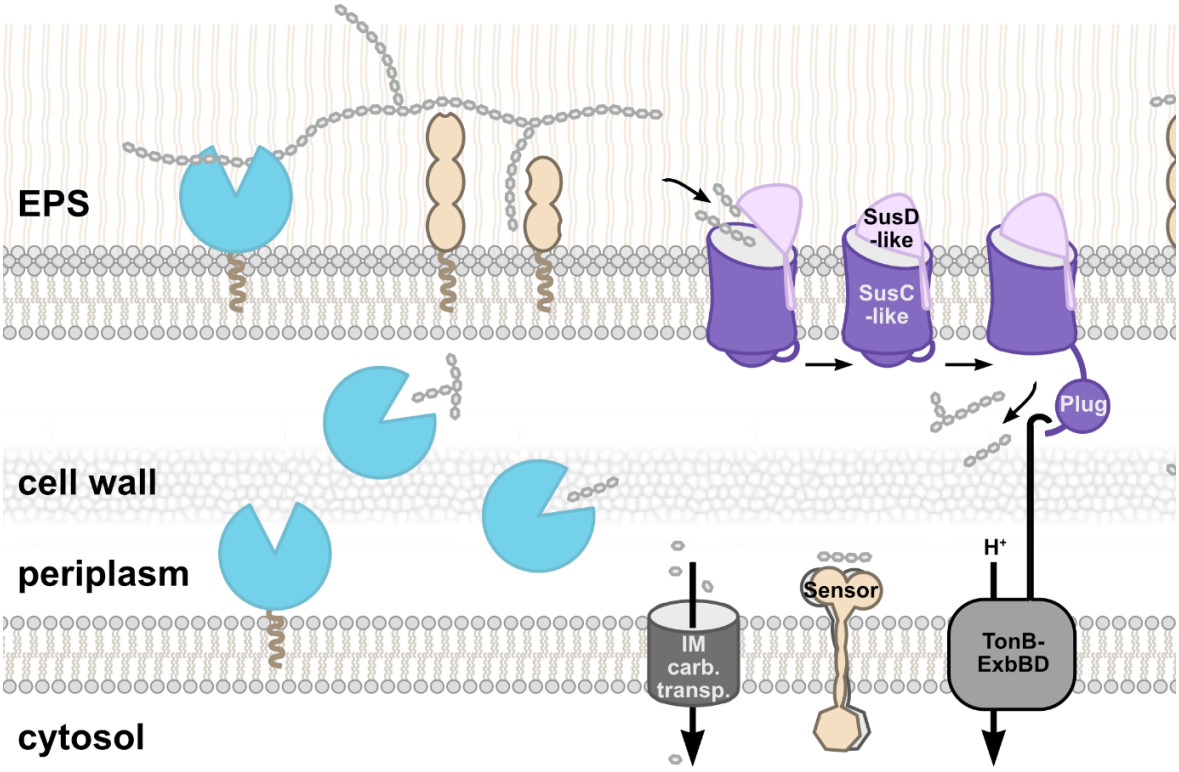


Figure 4. Overview of a canonical PUL system. A 'classical' PUL system typically has SusC/SusD-like pairs and regulator protein(s) (in purple), where the SusD-like protein captures carbohydrates to be imported through the SusC-like pore, and the sensor at the inner membrane (in light yellow) regulates the upregulation of the PUL upon binding of signature oligosaccharides with the help of TonB dependent transporter. In the 'classical' PUL setup, the surface-attached CAZymes (in blue) cleave polysaccharides into oligosaccharides to be imported and fully degraded to monosaccharides within the periplasm. The system may also include surface-tethered non-catalytic glycan-binding proteins (in light yellow), as well as inner-membrane transporters (in grey). Figure adapted from (104) under the CC BY 4.0 license.

Based on several PUL characterization studies, the Polysaccharide Utilization Loci DataBase (PULDB) was launched in 2015, a database of predicted and experimentally characterized PULs from sequenced genomes (167). Initially, the PULDB included PUL predictions for about 70 Bacteroidota species from the human gut together with PULs based on literature (163). Predictions are anchored on loci having *susC*/*susD*-like gene pairs and CAZymes in close proximity. Since the 2018 updated version, PULDB has expanded from 820 genomes to having 84k PUL predictions in around 2600 Bacteroidetes species, ranging from human, rumen, soil and marine environments, with more detailed categories like "Literature-derived PULs" (previously "Experimentally validated PULs") and "CAZyme clusters", which are predictions for regions rich in CAZymes but lacking a *susC*/*susD*-like gene pair (167).

PULDB has served as the data source for large-scale surveys, such as the analysis of thousands of PULs by their CAZyme composition (168). The annotated PULs from this database of the species relevant to this work are exemplified in Table 1, where it is evident that for species from the *Bacteroidaceae* family, the number of annotated PULs, especially the literature-derived ones, is much higher than in *Prevotellaceae*.

3.3.1 Polysaccharide utilization loci in the Bacteroidota phylum

B. thetaiotaomicron was the first species from which a PUL was described – the starch utilization locus (Sus), describe above. Owing to its genetic tractability, well-annotated genome, and the historical discovery of the founding Sus system in this species, it has served as the principal model organism for PUL research (104, 169). Much of the mechanistic understanding of how PULs sense, bind, and degrade glycans derives from work in this single organism. Notable examples of characterized PULs after the Sus are: the xyloglucan utilization locus (XyGUL) of *B. ovatus*, which was the first holistic description of a PUL in a single study (115); the yeast mannan-targeting PUL from *B. thetaiotaomicron*, which revealed how the PUL functions in a selfish manner where products are not released to neighboring cells (30); the chitin utilization locus (ChiUL) of *Flavobacterium johnsoniae*, which was the first description of a PUL targeting a crystalline polysaccharide (170); the xylan targeting PULs, PUL-XylL and PUL-XylS, from *B. ovatus*, which highlighted how this organism uses different PULs to utilize different types of this hemicellulose (117). Many other examples exist in a range of different plant-, microbial-, and algal polysaccharides (104). In *Zobellia galactanivorans*, an unconventional PUL with a SusC/D-like pair outside it was found to be degrading carrageenan, where the association of horizontal gene transfer with *Bacteroides plebeius* of its porphyran utilization locus was established (20, 128). PULs for one of the most structurally complex glycans known, pectins, was found in *B. thetaiotaomicron* with new enzyme families discovered, where to break down the complex pectin RG-II, a coordinated set of enzymes were deployed to fully depolymerize the various pectic domains (120).

Table 1. PULDB annotations for species studied in this thesis. The numbers of literature-derived PULs, predicted PULs, and CAZyme clusters annotated are listed below. The family *Bacteroidaceae* in general harbors a huge number of predicted PULs compared to *Prevotellaceae*. *Bacteroides cellulosilyticus*, *Bacteroides ovatus*, and *Bacteroides thetaiotaomicron* have more characterized PULs, illustrating the intensive research focus on these three prominent glycan-degrading species. The data also reveals how species from *Prevotellaceae* have received relatively low attention for CAZyme studies. Data collected from PULDB (167). Accessed 2026-10-21.

Strains	Literature-derived PULs	Predicted PULs	CAZyme clusters
***Bacteroidaceae* family**			
Bacteroides cellulosilyticus WH2	113	93	10
Bacteroides eggerthii DSM 20697	0	40	2
Bacteroides ovatus ATCC 8483	115	107	10
Bacteroides thetaiotaomicron VPI-5482	96	86	10
Dysgonomonas mossii DSM22836	0	29	0
Phocaeicola vulgatus ATCC 8482	0	67	4
***Prevotellaceae* family**			
Alloprevotella rava F0323*	0	7	0
Paraprevotella clara YIT11840	0	17	8
Paraprevotella xylaniphila YIT11841	0	14	8
Prevotella amnii DSM 23384^T	0	10	1
Prevotella dentalis DSM 3688 ^{T}	0	30	6
Prevotella herbatica WR041	0	31	6
Prevotella pectinovora DSM 29996 ^{T}	No data	No data	No data
Prevotella scopos JCM 17725 W2052	0	22	2

*Different to the strain tested in this work, which was *Alloprevotella rava* DSM 22548^T

As mentioned earlier, PULs are large loci in the genomes of certain species that encode all necessary proteins for the degradation of a target polysaccharide. Within the Bacteroidota phylum, the *Bacteroides* genus has been the most extensively studied regarding its PUL content and function (171). Species such as *B. thetaiotaomicron* and *B. ovatus* can illustrate development of different types of glycan specialization. *B. thetaiotaomicron* encodes around a hundred distinct PULs and can utilize a broad range of plant- and host-derived glycans, some of which, such as complex *N*-glycans, are degraded not by one single enzyme but by multiple enzymes drawn from several, acting together (172). This species is however unable

to use plant hemicelluloses as nutrients, which its relative *B. ovatus* is instead adept at, including xylans, mannan, and xyloglucan (115, 117, 173). In a study of *Bacteroides* species screened with fructose-based glycans by Sonnenburg et al. (2010), among all Bacteroidetes present in the experimental set-up, only *B. thetaiotaomicron* grew on levan, owing to the absence of an endo-levanase homologue in the fructan utilization loci of the others (103).

Species in the *Prevotellaceae* family, on the other hand, comprising genera such as *Prevotella*, *Paraprevotella*, and *Alloprevotella* that have been investigated in this thesis, are far less studied despite their ecological dominance in gut microbiotas of individuals with non-Western diets. *Prevotellaceae* species generally encode plant-targeting PULs, based on their CAZyme content and SusC/SusD-like proteins (104, 174). The gene content and arrangement, however, can differ from those of *Bacteroides*. The xylan PUL of *Segatella copri* DSM 18205 (previously classified as *Prevotella copri*), for example, lacks synteny with the previously characterized *Bacteroides* PULs, yet its xylan PUL is exactly syntenic with that of the rumen isolate *Prevotella bryantii* B14 (175), which suggests conserved gene arrangements within the family.

Complex as dietary fibers are structurally, PULs are highly substrate-specific, each tuned to a single glycan or a closely related family of glycans, whereas a single bacteroidete genome can contain over 100 predicted PULs, reflecting an extensive and specialized capacity to exploit diverse dietary and host-derived carbohydrates (167). The pattern, content, and abundance of PULs a species carries therefore constrain and usually predict which dietary fibers the species can utilize as substrates for growth. For example, surface-binding proteins are essential for glycan capture at the cell surface before import can occur. Tamura, *et al*., demonstrated that SusD-like surface glycan-binding proteins are indispensable for *B. ovatus* to utilize cereal β-glucan, showing that loss of these binding partners abolishes growth on the polysaccharide even when the corresponding catabolic enzymes remain intact (116). This also underscores that PUL function depends not only on catalytic CAZymes, but equally on the non-catalytic binding proteins that anchor and concentrate substrate at the cell surface prior to hydrolysis.

3.4 Substrate preferences in carbohydrate degradation

The PUL repertoire of a species defines which glycans it is able to use. PUL content can theoretically predict what a species can grow on. However, the PUL content does not describe what a species prefers when several carbohydrates are present in the same environment. Different *Bacteroides* species have different sets of PULs, which govern their overall abilities to utilize complex glycans as nutrients, but they also have different preferences for these. Their preferences are shown as differences in the order of substrates used, when they are presented with several types. Several human gut Bacteroidota species utilize glycans hierarchically, in a way that resembles catabolite repression, and in *B. ovatus* and *B.*

thetaiotaomicron, the hierarchical orders generally remain the same in co-culture, which is its "hard-wired" preference list. In a paper from Rogers et al. (2013), for *B. thetaiotaomicron* in the co-culture setting, this priority of glycan consumption was experimentally validated and the strict consumption order was not affected by prior exposure to lower-priority glycans (105). Thus, the presence of a PUL shows that a glycan can be used, but not when it will be used relative to other available glycans. Such hard-wired differences in capability and priority are thought to promote stable coexistence by partitioning shared glycan pools rather than direct competition for identical resources (3, 176).

The fact that many homologous or syntenic PULs are found across the phylum suggests that they have spread by horizontal gene transfer, though the mechanism is unclear (20). By extension, polysaccharide degrading capabilities carried by individual microbial community members are not necessarily confined to one niche. This also makes industrial biomass valorization a possible application that can refer to the same PUL- and CAZyme-driven degradation strategies that have been studied in gut bacteria (36). Understanding whether the species-level specialization in different carbohydrate utilization would also reflect differences at the metabolite level, e.g., changes in extracellular OA production, could further elucidate if substrate preference in Bacteroidetes shapes not just their growth capability but also the downstream fermentation products each species contributes to the gut microbiome.

Chapter 4: Original work

Although PULs have been extensively researched in Bacteroidota species, studies of which metabolites they produce are limited and have mainly relied on one single substrate, usually glucose. Metabolic output profiles resulting from carbohydrates with different degrees of complexity, that better simulate the substrate variability in a real diet, are currently missing. PUL studies have mostly focused on *Bacteroides* species, while studies of both substrate utilization and metabolites produced by *Prevotellaceae* species are lacking. Even though multiple putative PULs have been identified in genomes of *Prevotellaceae*, these predictions require experimental evidence.

This chapter revisits the aims of the thesis and summarizes the four lines of original research: extracellular metabolite profiling across Bacteroidota species (**Paper I** and **Paper II**), which inspired the following biochemical characterization of GH67 and GH115 α-glucuronidase enzymes from putatively annotated gene clusters in the genome of *P. clara* (**Paper III**), and lastly, development of methods for interspecies PUL transfer (**Paper IV**).

4.1 Substrate-directed extracellular metabolite production in Bacteroidota species

As mentioned in the previous sections, PULs are abundant in many Bacteroidota species. This enables the tools needed to deconstruct carbohydrates and thereby allows bacteroidetes to collectively grow on a wide variety of polysaccharides. However, the metabolites that result from this growth have received less attention. So far, most data on species-level metabolite output of gut Bacteroidota species has been performed using a defined growth phase with a single, standard carbon source — most often glucose — and quantifying the resulting SCFAs and OAs by chromatography (13, 14). Generally, gut Bacteroidota species are associated with propionate- and succinate-dominated fermentation, in contrast to the more butyrogenic output typical of many Bacillota species (177, 178). These data are most extensively documented in *B. thetaiotaomicron*, where gene-level studies have proposed several biosynthetic routes for each product (Figure 5) (13). A glucose-centric metabolic profile therefore leaves open how these microorganisms behave on other simple carbohydrates or on the more ecologically relevant complex dietary polysaccharides. Whether a species' metabolite profile is a fixed, intrinsic property or instead depends on the substrate being fermented is a question the earlier studies were not designed to answer (13, 14).

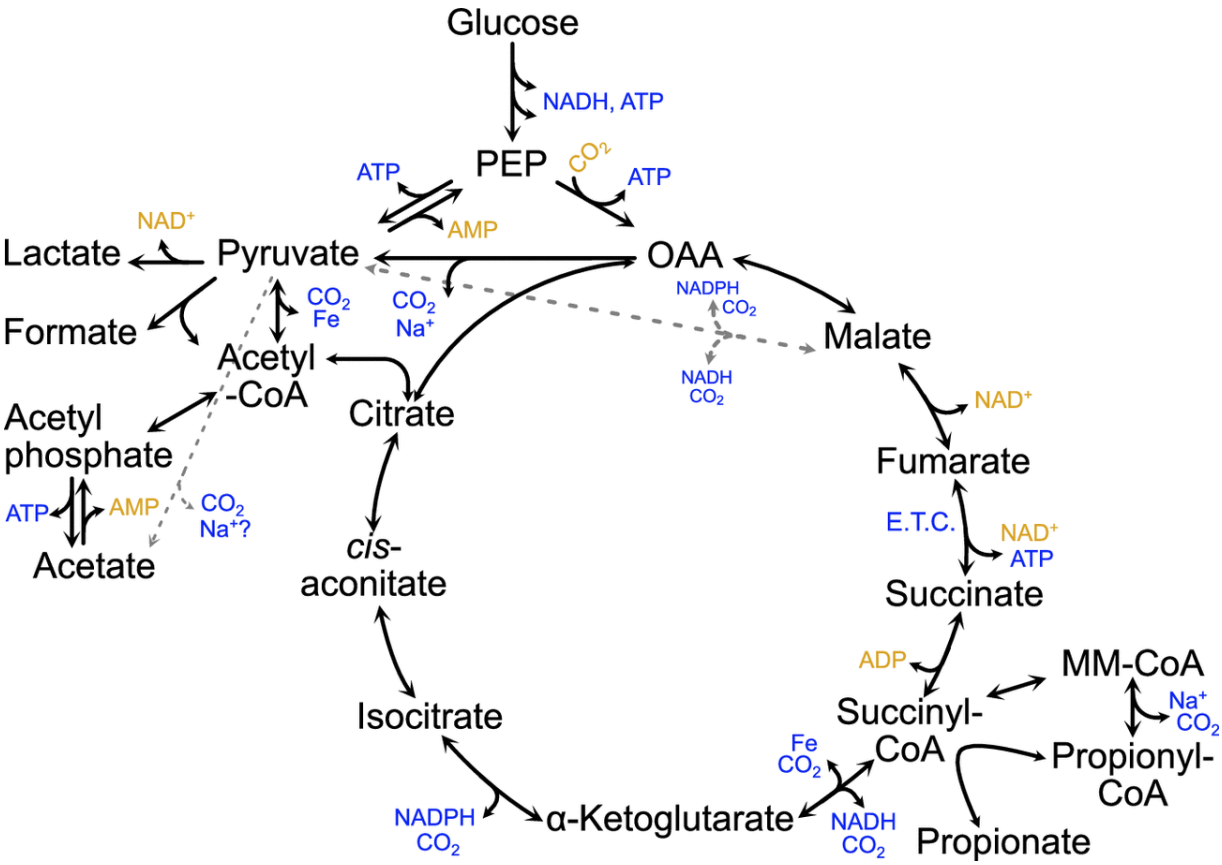


Figure 5. Model of *Bacteroides thetaiotaomicron* VPI-5482 organic acid production using glucose as cycle entry point. Glycolysis is inferred between glucose and PEP. Production of molecules is shown in blue, whereas consumption of the same compounds is shown in gold. PEP - Phosphoenolpyruvate. OAA - Oxaloacetate. CoA - Coenzyme A. E.T.C. - Electron transport chain. MM-CoA - Methylmalonyl-CoA. Fe - Ferredoxin/flavodoxin. Figure adapted from (179) under the CC BY 4.0 license.

Given the stated knowledge gap, I aimed in the studies, leading to **Paper I** and **Paper II,** to profile the metabolite outputs of individual *Bacteroides* species (**Paper I**) and several members of the *Prevotellaceae* family (**Paper II**). A gut community is more complex than these single-species- and substrate studies, where the gut microbiota would instead encounter a dynamically changing mixture of fibers as well as neighboring species. However, to reach a community-level understanding of gut microbiome dynamics, undertaking metabolic analysis in mono-cultures can provide fundamental understanding of how single species behave, which can then be expanded to co-cultures and larger communities.

4.1.1 Metabolites produced by *Bacteroides* species from carbohydrate utilization

In order to investigate the metabolic modulation of *Bacteroides* species in **Paper I,** the work began with a first-stage growth assay screen of six species against an array of carbohydrates in 96-well plates. These species were chosen more broadly than the *Bacteroides* genus, either based on having been well-studied regarding carbohydrate utilization (*B. thetaiotaomicron*, *B. ovatus*, and *Bacteroides cellulosilyticus*), or the fact that they had been less studied yet have high potential capability for growth on a range of carbohydrates (*Bacteroides eggerthii*, *Phocaeicola vulgatus*, and *Dysgonomonas mossii*). From the initial growth screen, species with weak growth were however excluded from the subsequent analyses (*B. eggerthii*, *P. vulgatus*, and *D. mossii*), limiting those studies to species from the *Bacteroides* genus.

Unlike *B. eggerthii* and *D. mossii*, among the three excluded species, *P. vulgatus* had measurable growth on several mono- and oligosaccharides, but it grew poorly on starch and barely on polysaccharides like polygalacturonic acid. Considering its genomic content with 67 predicted PULs, the finite number of substrates selected in **Paper I** could be a limiting factor, and possibly it has a wider-ranged carbohydrate utilization than observed (85). The three species with robust growth, *B. thetaiotaomicron*, *B. ovatus*, and *B. cellulosilyticus*, were further cultivated on defined carbon sources in larger scale (2-mL tubes), from mono-saccharides to complex polysaccharides, followed by metabolite collection and analysis. Extracellular OAs were quantified by high performance liquid chromatography (HPLC) to assess whether growth on each substrate produced distinct and predictable metabolite profiles.

In the prior literature using glucose as the reference carbohydrate, *Bacteroides* typically secrete a core set of OAs, mainly acetate, succinate, formate, and propionate, and also some lactate, which are produced following glycolysis, mixed-acid fermentation and the succinate-propionate pathway (13, 178). The acid production comes from central metabolism (Figure 5), mostly the non-oxidative branch initiating from downstream pyruvate formation (179). Comparing the metabolic profiles of the three *Bacteroides* species when grown on glucose, the extracellular metabolic profiles from monosaccharides did not show significant differences for other neutral hexoses or pentoses. However, for some monosaccharides there were clear differences compared to glucose, such as for uronic acids, deoxy-sugars, and amino-sugars, where overall most OAs decreased in abundance except for acetate which instead increased substantially. Thus, already for monosaccharides, it was observed how the metabolite output can be quite variable.

The studies concept was were further extended to oligosaccharides. For these reasons, no large differences were seen compared to glucose or the other neutral hexoses for the overall profiles, but significant concentration differences could be observed in some cases, such as reduced malate from maltose in *B. thetaiotaomicron* cultures compared to glucose (Figure 6).

Possibly, this reduction reflects the requirement to synthesize additional enzymes to cleave the oligosaccharides to monosaccharides. To measure the extent of carbohydrate utilization in the cultures, the medium broth after each culture reached its stationary phase (the sampling timepoint T_3, as referred to in **Paper I**, was analyzed using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD), which indicated little to no residual carbohydrate remaining in the culture broth (13, 178).

The metabolic profile comparison was then extended to polysaccharides. For homopolysaccharide substrates, one could expect that the metabolic composition would be the same or proportional to that of the constituent monosaccharide. Namely, the metabolite output from starch or various glucose-based oligosaccharides would be the same as for glucose, as the same monosaccharide enters the central metabolic pathways. For heteropolysaccharides, this would then instead result in a metabolite mixture reflecting the proportions of the profiles observed from their constituent monosaccharides. However, in **Paper I**, unexpectedly, growth on a polysaccharide often yielded a different profile from growth on its own constituent monosaccharide(s). As exemplified in Figure 6, metabolic profiles from *B. thetaiotaomicron* cultures of amylopectin and starch were compared to both glucose and maltose. A similar pattern could be seen but with a gradually decreasing overall metabolite concentration measured by the order "glucose-maltose-starch (monosaccharide-oligosaccharide-polysaccharide)". Significant differences were found, for lactate and acetate levels from growth on both amylopectin and starch. For other polysaccharides, the OA output sometimes correlated with the main monosaccharide constituents, but not consistently, and it was observed that the extracellular metabolite profiles were often somewhat stochastic.

The same type of analyses was done for *B. ovatus* and *B. cellulosilyticus*, where similar to the results for *B. thetaiotaomicron*, the metabolite profiles were observed to vary, both within and among the species, depending on the carbon source provided. Furthermore, the assumption that a more complex carbohydrate would result in lower amounts of OAs due to the need to synthesize a corresponding set of degradative enzymes was not always the case. For instance, when *B. ovatus* was grown on certain hemicelluloses, it produced equally higher amounts of OAs compared to on monosaccharides.

To summarize, a species' metabolic output to the gut environment as demonstrated is therefore not necessarily a fixed nor predictable trait of any one microorganism, but can depend on which nutrients are available, even when the carbon source is always a carbohydrate. This adds an important layer of nuance to how a bacterium's ecological role in

the mixed-glycan environment, such as guts, can be dynamic. There is still no obvious explanation for the mechanism behind these observations, but future studies may shed light on the molecular and regulatory aspects governing the final OA output. The point is especially relevant for species such as *B. cellulosilyticus*, which carries one of the largest and most extensive glycobiomes described among human gut symbionts (180).

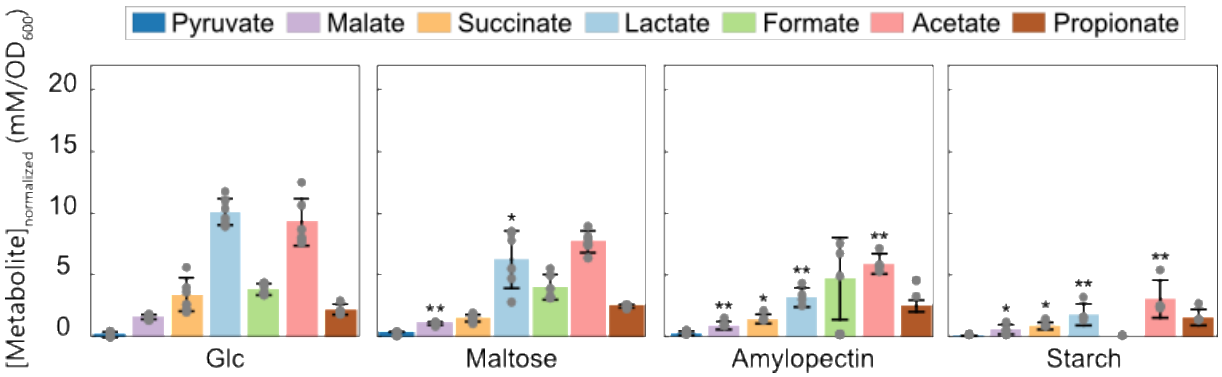


Figure 6. Metabolite profiles from *Bacteroides thetaiotaomicron* on polysaccharides and its constituent monosaccharides. These results demonstrated various substrate accesses did not translate into a single, fixed metabolic outcome. Species accessing substrates formed by the same monomers could still generate distinguishable OA signatures. Downstream metabolic routing is therefore dependent on substrate choice instead of the past glucose-centric uniformly-produced OA assumption, which rather needs checking substrate by substrate for metabolite data. Cultures were grown with 5 g/L carbohydrate as starting carbon source and samples collected at peak growth, and metabolites monitored using HPLC. The concentrations are normalized to the OD_{600} values, which were all in a similar range (around 1.5; see Supplemental data file 1 in **Paper I**). Individual data points are shown, where the bars show the mean value, and error bars represent standard deviation, from at least triplicate experiments. Asterisks indicate p-values from t-tests comparing the amount of each metabolite to what was produced from Glc, with * for $p \leq 0.01$ and ** for $p \leq 0.001$. Figure adapted from **Paper I** under the CC BY 4.0 license.

4.1.2 Substrate probing for metabolic composition of the understudied *Prevotellaceae* species

The metabolic profiling results from *Bacteroides* species inspired the decision to investigate a less studied family from Bacteroidota. Members of the *Prevotellaceae* family have diverse habitats, and carry varying sets of PULs that suggest extensive carbohydrate-degradation abilities (Table 1). *Prevotella melaninogenica*, for example, one of the earliest species that was originally isolated from patient samples and first named *Bacteroides melaninogenica*, is a commensal but opportunistically pathogenic bacterium commonly residing in human oral

cavity and the upper respiratory tract. It has been shown to carry a putative starch-targeting PUL (181). Like Bacteroidetes, genomes of many *Prevotellaceae* species can encode several CAZymes, and they are, therefore, expected to be proficient carbohydrate consumers. Broader genomic surveys across dozens of *Prevotellaceae* species from oral, gut, and other niches have identified nearly 1200 SusD-like proteins, with substantial variation in PUL number and repertoire per genome (174). PUL-based glycan foraging thus appears to be a general strategy rather than a gut-specific adaptation for *Prevotellaceae* species. Historically, though, research on the family has focused on clinical and disease-associated mechanisms, leaving their carbohydrate metabolism largely unstudied (182).

Existing PUL surveys of the family have disproportionately focused on a small number of gut-associated species, primarily *S. copri* (174, 183, 184). The PUL content and predicted substrate specificities on *Prevotellaceae* spreading out in other microbiomes therefore remain comparatively at genomic prediction level in relation to other gut-associated relatives within the phylum Bacteroidota. *Prevotella dentalis* (185), first described in infected dental root canals, and *Prevotella amnii* (186), originally isolated from amniotic fluid, are examples of such that reside likewise in the human oral cavity microbiome. Given the lack of their empirical data, most of their metabolic mechanisms have also not been thoroughly tested, including the recurring topic of carbohydrate modulation ability for their metabolite output.

Therefore, **Paper II** of this thesis addressed this gap following the same methodology as previously described in **Paper I**, by conducting growth assays using a panel of carbohydrates as sole carbon sources for growth with eight previously uncharacterized *Prevotellaceae* species with annotated genomes. A recurring assumption in PUL genomics is that the presence of a PUL predicted to target a given glycan is sufficient to infer growth capability on polysaccharides (103). The selection of species was thus based on the level of carbohydrate catabolizing capabilities each strain carries, inferred from their respective annotated PULs and gene clusters on PULDB. *P. amnii* DSM 23384^T, *P. dentalis* DSM 3688^T, *Prevotella herbatica* DSM 112534^T, *Prevotella pectinovora* DSM 29996^T, *Prevotella scopos* DSM 22613^T, *P. clara* DSM 19731^T, *Paraprevotella xylaniphila* DSM 19681^T, and *Alloprevotella rava* DSM 22548^T were thus selected from that they putatively contain multiple GH families, PULs, and CAZyme clusters in their genomes that are likely to target complex glycans; see Table 1 in Chapter 3.3.1 (167). Additionally, these species were originally isolated from a diversity of environments across different human body, from environmental or industrial settings like methanogenic bioreactor, and fecal samples from mammals (185-191).

The experimental design that was different from the setting in **Paper I** involved the use of a more nutrient-rich medium, including beef extract, to the 96-well growth screening, as the more limited medium for *Bacteroides* did not sustain basal growth. The carbon sources that

enabled bacterial growth to at least OD_{600} values around 0.25-0.3 were then as before scaled up to 2-mL cultures for metabolite analysis.

In contrast to the *Bacteroides* species studied in **Paper I**, the overall growth of the *Prevotellaceae* species was poor. It was largely restricted to mono- and disaccharides, and the final OD_{600} values were well below those of previously observed in *Bacteroides* in **Paper I** (often below 0.8 compared to a general >1 value for *Bacteroides* species), despite the richer medium used. Based on cultures using glucose, their metabolite composition was appreciably different from those of *Bacteroides* in terms of types and proportions of metabolites. Fewer types of OAs were produced throughout the fermentation, where the dominating types were lactate, acetate, and at times formate and slight propionate production in some species. Lactate was frequently the most abundant OA, while succinate, malate, and pyruvate, which are commonly reported for *Bacteroides* (13), were not detected in our experimental conditions.

Surprisingly, the metabolite profiles from other neutral monosaccharides did not resemble that on glucose, in contrast to the results from the *Bacteroides* species. Of the selected *Prevotellaceae* species, monosaccharides like glucose, fructose, galactose, and mannose were the preferred substrates, where overall, *P. herbatica*, *P. xylaniphila*, *P. clara*, and *P. pectinovora* exhibited more stable growth responses on the supplemented carbohydrates and slightly higher metabolite concentrations overall than the other four species (Figure 7).

Once again, differences in metabolite profiles observed as per substrate suggested that carbon flux through central metabolism is not static. Possibly, rather it depends on how each substrate enters metabolism and on the redox demand it creates, respectively. Substrates entering at different points of the metabolic cycle, such as hexoses via glycolysis or pentoses via the PPP, yield different amounts of ATP and reducing equivalents, where the cell compensates by shifting the balance between redox-consuming products, like succinate and lactate, and redox-releasing products, acetate, formate, and potentially H_2 (84, 192-194). Some products can also arise directly as a result of the substrate choice, as with propane-1,2-diol observed from cultures with fucose and rhamnose in *P. vulgatus* (76). As complex polysaccharides are degraded gradually, contrasting free carbohydrates that can cause overflow metabolism and faster acidification, where OA secretion thus can itself be feedback-inhibited as an impact from the speed of substrate release (84).

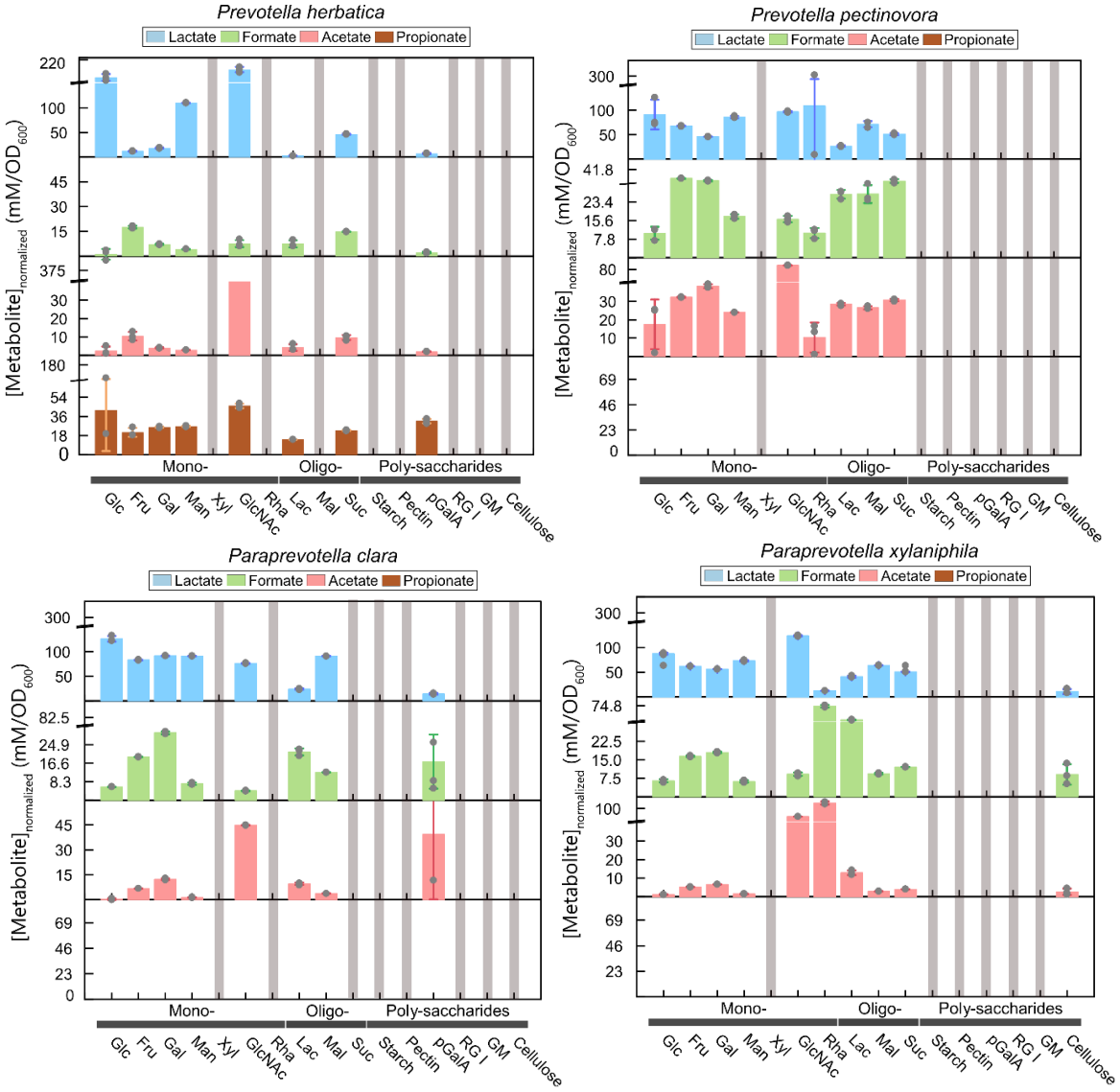


Figure 7. Metabolite profiles from the four *Prevotellaceae* that exhibited better growth. Samples collected from the four species, *Prevotella herbatica*, *Prevotella pectinovora*, *Paraprevotella clara*, and *Paraprevotella xylaniphila*, were processed in HPLC for OA quantification; run in triplicate. The secretion profiles from these cultures were different with pattern changes seen among different carbohydrates both inner-species and interspecies (refer also to Figure 6 for comparisons with *Bacteroides thetaiotaomicron*). The supplemented carbohydrate carbon source was 5 g/L and samples for chromatography analysis were collected at stabilized maximum growth normalized to OD_{600}. Carbohydrates that did not sustain the cultures were marked dark gray. Taken from **Paper II**.

Unexpectedly, the *Prevotellaceae* contrasted species exhibited poor or no growth on polysaccharides, contrasting the PULs number of predicted PULs and CAZymes overall predicted (Table 1). For instance, *P. clara* is predicted to have 17 PULs and 8 CAZyme clusters, including PUL with seemingly clear targets, such as one encoding a GH10 gene that is very likely a xylanase, given the family's characterized enzymes to date (134), and other likely xylan-active enzymes. Such predictions strongly indicated that *P. clara* would grow on xylan, yet no growth on xylan was seen from the growth assay.

Species in the GI microbiome, especially in the guts, encounter a complex mixture of digesta composition *in vivo*, including oligosaccharides and metabolites released by other primary degraders (70). Because PULs are induced by their cognate substrates, the activation of PULs in a microbiome may depend not only on the species' own genetic potential, but also on the free metabolites as donated substrates supplied by other members (7, 166). This possibility is what a monoculture on defined substrates cannot test. The lack of growth on xylan by *P. clara*, despite seemingly obvious capacity based on its genome, led to the study reported in **Paper III**, where individual enzymes were functionally investigated.

4.1.3 Future directions for metabolite output linked to carbohydrate utilization

Taken together, the results presented in **Paper I** and **Paper II** suggest that *Prevotellaceae* may occupy niches different from *Bacteroides* in fiber degradation and metabolite production, a possibility that will require testing in co-culture and community settings. The two metabolite profiling studies showed that the metabolic outputs of Bacteroidota species are not a fixed trait, as in both the *Bacteroides* genus and the *Prevotellaceae* family, the OAs secreted varied with the carbon source supplied. Since all measurements were made using monocultures on defined substrates, the metabolic profiles here described only what each species produces in isolation, which is not necessarily what a communal microbe does in interplay with others.

Making co-culture and defined-community experiments using the same substrates that resulted in growth could be one way to extend the monoculture profiling to see whether outputs add up or lead to different responses that suggest interspecies interactions, and these may then begin to mimic the dynamics of the microbiota. Co-culture competition assays, in which two species are grown together on a defined substrate (mixture) with their relative abundance tracked over time by strain-specific quantitative polymerase chain reaction (qPCR), alongside OA or carbohydrate measurements by HPLC/HPAEC-PAD, would help illustrate if interactions such as cross-feeding or competition occur.

Transcriptomic or proteomic profiling under matched conditions can further show which genes and enzymes drive substrate-dependent output, with an extension of the substrate panel

to more “natural” mixtures and to more species as an expansion of experiment. The known OA pathways could be a good starting point for the mechanistic part, for that the core routes have been mapped genetically in reported *B. thetaiotaomicron* study by deletion mutants and HPLC quantification (13). Genome surveys also showed that the core pathways for acetate, formate, lactate, and succinate-propionate formation are widely present among gut Bacteroidota genomes (178), yet the metabolite profiles produced differ between species with the carbon source supplied according to a study from Louis et al. (2017) (177). Knockout studies can be applied to identify fermentation pathways and to investigate in which way blocking a specific route can influence the metabolite profiles on different substrates. The same approach is, however, limited in *Prevotellaceae* due to the scarcity of established genetic tools and by less complete genome annotation (183).Other than omics studies, intracellular NMR metabolomics or metabolic flux analysis with ^{13}C-labelled substrates could help resolve the underlying flow of carbon within cells (84).

4.2 Characterization of α-glucuronidases from *Paraprevotella clara*

The unexpected lack of growth on polysaccharides by *Prevotellaceae* species, reported in **Paper II,** raised the question of whether their predicted PULs and CAZymes were functional. As xylan is a ubiquitous polysaccharide in plants, and as such found in most diets, it was chosen as a model substrate for deeper studies. Of the investigated *Prevotellaceae* species, only *P. amnii* lacks apparent xylan-degrading enzymes, where the others have multiple copies of predicted relevant enzymes: xylanases, arabinofuranosidases, and α-glucuronidases (Table 2). To assess whether such enzymes were functional, I focused on *P. clara*, as it encodes at least nine putative xylan-acting enzymes, from GH10, GH30, GH67, and GH115, where several are found within putative xylan-targeting PULs (167, 189).

As *P. clara* is an understudied species, it was unclear if its lack of growth on xylan was a result of the cells entirely lacking activities for xylan degradation and metabolism. In **Paper III**, the aim was to investigate the activities, or lack thereof, of the select GHs encoded by *P. clara*. Enzymes from the α-glucuronidase families GH67 and GH115 which should remove (4*O*Me)GlcA substituents from xylan were chosen, as these enzymes are important to facilitate complete degradation of the polysaccharide (18, 151, 195).

Table 2. Xylan-degrading-relevant GHs predicted in the studied *Prevotellaceae* species. The numbers of predicted GHs that are linked to xylan-degrading activities are listed below. Species like *Paraprevotella clara*, *Paraprevotella xylaniphila*, and *Prevotella herbatica* have an abundant number of putative xylosidases found in genomes, where xylanases can be found occasionally in the same predicted PULs. This hints on the high likelihood of the predicted PULs being a xylan-degrading PULs, which was not reflected by the otherwise poor growth of cells on xylan observed from the results of **Paper II**. Data collected from PULDB (167). Accessed 2026-10-21.

Strains	Putative xylanases	Putative β-xylosidases/α-L-arabinofuranosidases	Putative arabino-furanosidases	Putative α-glucuronidases	Total[*1]
Alloprevotella rava F0323[*2]	0	0	0	0	0
Paraprevotella clara YIT11840	3	21	0	2	26
Paraprevotella xylaniphila YIT11841	3	24	1	3	31
Prevotella amnii DSM 23384[T]	0	0	0	0	0
Prevotella dentalis DSM 3688 [T]	2	9	1	3	15
Prevotella herbatica WR041	6	13	1	2	22
Prevotella pectinovora DSM 29996 [T]	ND	ND	ND	ND	ND
Prevotella scopos JCM 17725 W2052	0	0	0	0	0

Putative xylanases: GH10, GH11, and GH30; β-xylosidases: GH43 and GH51; arabinofuranosidases: GH5 and GH8

Putative α-glucuronidases: GH67 and GH115

[*1] Total – Total no. of the putative xylan-acting GHs found in predicted PULs

[*2] Different to the strain tested in this work, which was *Alloprevotella rava* DSM 22548[T]

[*3] ND – No data

4.2.1 Biochemical investigation of GH67 and GH115 α-glucuronidases from *Paraprevotella clara*

Xylan is an abundant plant polysaccharide material aside from cellulose (196). As a major dietary fiber, it reaches the colon intact, where gut Bacteroidota species can degrade it with dedicated PULs. In the human gut, the enzymes that carry xylan-degrading activities can be found in *B. ovatus*, where the surface-exposed endo-acting enzymes cut the xylan backbone before the products are imported (18), whereas the host itself cannot depolymerize dietary fiber. Complete degradation of xylan requires several activities acting together. Of these, the xylanases, having the primary biological role of main chain-cleaving enzymes, fall mainly in GH10 and GH11, which CAZy lists as the main xylanase families (134). The xylan α-glucuronidases that remove the α-glucuronic decorations on xylan backbone belong to GH67 and GH115.

P. clara encodes four putative α-glucuronidases, one from GH67 (*Pc*GH67) and three from GH115 (*Pc*GH115A, B, and C). GH67 and GH115 α-glucuronidases have been biochemically characterized from a range of bacterial and fungal sources, with prior work establishing a consistent mechanistic distinction between them — GH67 enzymes act only on the non-reducing terminal (4*O*Me)GlcA substituents of xylo-oligosaccharides, whereas GH115 enzymes can also release internal (4*O*Me)GlcA from polymeric GX. This makes the two families functionally complementary rather than redundant in organisms that encode both (18, 19, 197, 198). Both families remain currently less studied. The number of characterized GH115 enzymes is low, and the ones studied differ in domain composition, oligomeric state, and the position of a mobile loop, which affects activity on polymeric substrates. A reported structural comparison of *Bo*Agu115A from *B. ovatus* with the Protein Data Bank (PDB) database identified the *Geobacillus stearothermophilus* GH67 α-glucuronidase as the closest structural homolog, yet with weak overall profile of sequence identity of 11% (18). Moreover, GH115 is reported to be poly-specific, as shown by a member active on decorated arabinogalactans (199), showing their activity is hard to infer from sequence alone. with no prior biochemical characterization available for enzymes from both GH67 and GH115 encoded within the putative xylan-targeting CAZyme clusters of *P. clara*. Although xylan utilization is a key trait of *Prevotellaceae* (200), the putative xylan PULs in *P. clara* lack synteny with the characterized *Bacteroides* ones (117). Whether the GH67 and GH115 enzymes in the genome of *P. clara* resemble the characterized ones in sequence, in domain architecture, or in genomic context is not known, which motivated the characterization of GH67 and GH115 enzymes of *P. clara* in this work.

Bioinformatic analysis also placed *Pc*GH67, *Pc*GH115A, and *Pc*GH115C in xylan-related gene contexts. *Pc*GH67 sits in a likely xylan PUL, and *Pc*GH115C lies in a nearby cluster with one putative GH30 xylanase present. The genomic context of *Pc*GH115B is however

not as apparently linked to xylan depolymerization, as the possible targets for the cluster where *Pc*GH115B is located could be either xylan or complex arabinogalactan, such as acacia gum which comprises terminal (4*O*Me)GlcA residues. Phylogenetically, *Pc*GH115A and *Pc*GH115C are grouped with known glucuronoxylan-active enzymes, whereas *Pc*GH115B is closest to *Bt*GH115A *thetaiotaomicron*, an enzyme found active on acacia gum arabinogalactan and not on GX (199). AlphaFold2 models showed conserved catalytic residues in all four enzymes, with loop insertions in *Pc*GH115A and *Pc*GH115B that may affect substrate binding (Figure 8).

Considering the α-glucuronidases encoded by *P. clara* have predicted preferences for both terminal (GH67 and GH115) as well internal (GH115) glucuronidation, the biochemical characterization was carried out in combination with xylanases from four main xylanase families, GH8, GH10, GH11, and GH30, on GX from beech wood (BGX). Since these xylanases respectively have different substrate binding sites and structures, they generate different GX fragments, allowing observation of substrate preferences for the different α-glucuronidases.

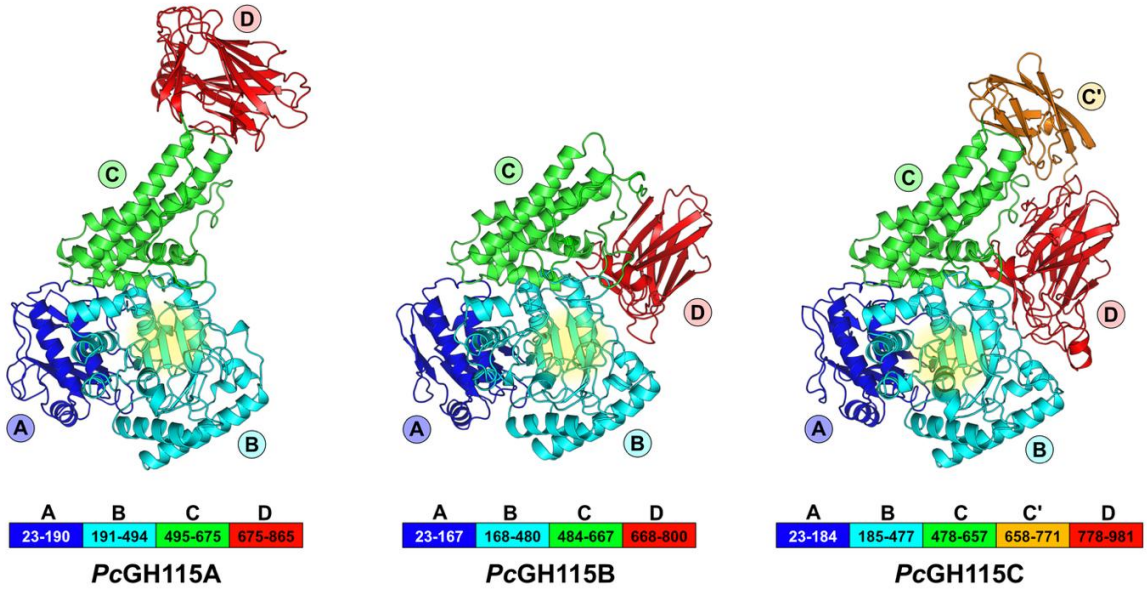


Figure 8. AlphaFold models of the overall structures of *Pc*GH115 enzymes. The enzymes comprise at least four distinct domains (A-D) while *Pc*GH115C comprises an additional domain (C') between domains C and D. The domains are distinctly colored with the residue ranges for each domain provided below the structures. The putative active site pocket, from homology to characterized members, is highlighted in yellow. The different relative position of domain D to the others between *Pc*GH115B and *Pc*GH115A/C. Taken from **Paper III**.

From product profiles using HPAEC-PAD, it was observed that *Pc*GH115A and *Pc*GH115C released (4*O*Me)GlcA from intact BGX, both alone and when used together with GH8, GH10, GH11, or GH30 xylanases. *Pc*GH67 released barely measurable (4*O*Me)GlcA only from BGX fragments generated by a GH10 xylanase, as expected for enzymes acting on the terminal non-reducing residues of oligosaccharides. However, addition of xylanase to pre-digest BGX before adding most of α-glucuronidases did not improve activities of the GH115 enzymes. The fragments from GH8, GH11, or GH30 left the activity of *Pc*GH115C largely unchanged. GH10 fragments rose *Pc*GH115B activity by 61-fold, however, relative to using intact BGX. More substrate verification is needed before validating *Pc*GH115B as functional α-glucuronidases, as the biological role of *Pc*GH115B remains speculative, though its specificity with GH10 digested products did suggest a role as a α-glucuronidase for possibly small linear fragments. As *Pc*GH115B had an overall low activity to BGX and BGX fragments, and also did not show any apparent clear activity on acacia gum (data not shown), it suggests that we have not been able to identify a relevant biological substrate for it. However, *Pc*GH115A, *Pc*GH115C, and, in combination with GH10, *Pc*GH67, could therefore be validated as functional α-glucuronidases.

4.2.2 Biological roles of α-glucuronidases encoded by *Paraprevotella clara*

Although the recombinant *Pc*GH67 and *Pc*GH115 enzymes were catalytically active *in vitro*, *P. clara* was still unable to grow on α-glucuronidase/xylanase-pre-digested BGX supplemented culture broth. There was also no growth on very simple substrates such as xylobiose and xylose, suggesting that the most likely explanation for the lack of growth on xylan by *P. clara* is not directly related to lack of enzyme activity, but rather related to malfunctioning or missing substrate uptake or downstream xylose/xylo-oligo metabolism.

Looking into the gene content in the *P. clara* genome as indicated in Table 3, two putative xylose metabolism genes, xylose isomerase (*xylA*) and xylulokinase (*xylB*), are found co-located. This neighboring gene arrangement is similar to the homologous genes in *B. ovatus* ATCC 8483, which can metabolize xylose; the two-enzyme route used by *Bacteroides* (196) is likely to be the xylose-utilizing pathway in *P. clara*, although this has not yet been validated experimentally. However, the equivalent genes to ribose-5-phosphate isomerase B (*rpi*) and transketolase (*tkt*) are not directly neighboring the xylose metabolism putative genes, and there is a lack of high identity to the *B. ovatus* regulator (*xylR*) in the *P. clara* genome. Altogether, this might suggest the xylose pathway in *P. clara* could be either constitutive or is controlled by regulators of other monosaccharides or a global regulator, rather than being co-activated regulated. These findings imply the inability of *P. clara* to utilize xylan as the

Table 3. Putative xylose metabolic genes involved in xylose metabolism in *Paraprevotella clara.* Putative proteins for xylose metabolism *Paraprevotella clara* YIT 11840 were blasted using Basic Local Alignment Search Tool – Protein (BLASTP) against *Bacteroides ovatus* ATCC 8483 to investigate the potential analogues. Data retrieved from BLASTP (201). Accessed 2026-10-21.

Genes in *Bacteroides ovatus* ATCC 8483 and its function description		Locus tags		Per. Ident (%)
		Paraprevotella clara YIT11840	*Bacteroides ovatus* ATCC 8483	
xylR	NrtR-family regulator	HMPREF9441_03585	BACOVA_02532	28.12
xylB	Xylulokinase	HMPREF9441_00786	BACOVA_02533	68.71
xylA	Xylose isomerase	HMPREF9441_00785	BACOVA_02534	86.76
xylE	MFS transporter	HMPREF9441_01443	BACOVA_02535	85.51
rpi	Ribose-5-phosphate isomerase B	HMPREF9441_02892	BACOVA_01705	72.22
tkt	Transketolase	HMPREF9441_02891	BACOVA_01706	80.18

main carbon source in our growth assays might likely not relate to missing critical metabolic pathways, or catalytically inactive xylan-active enzymes, but possibly rather in sensing/cellular signaling, or substrate transport.

To test whether basal xylose metabolism genes are transcribed, ideally the genes for xylose isomerase (*xylA*), xylulokinase (*xylB*), and any xylose transporter or PPP genes could be quantified by real-time quantitative polymerase chain reaction (RT-qPCR) or RNA sequencing (RNAseq) in cells grown on xylose, xylan, and glucose. However, since the cells do not grow on the xylose-based substrates, this is not feasible. Potentially, using or developing another basal medium could help the cells grow on these carbon sources. As for the substrate recognition at cellular level, characterization of the surface glycan-binding proteins (SGBPs) can be done, including the SusD-like proteins, by producing them heterologously. The binding affinity could be tested against xylan, xylo-oligosaccharides, and its substituted forms by affinity gel electrophoresis, pull-down with insoluble xylan, or isothermal titration calorimetry (115, 116).

Though *P. clara* carries genes for both xylose metabolism and xylan degradation, using α-glucuronidases and xylanases from other species may not be enough to trigger its expected sensing of xylan and related PULs, and possibly other enzymes are needed to provide key

Table 4. Identity comparison of the genes of carbohydrate-active enzymes from a putative xylan-targeting polysaccharide utilization locus in the *Paraprevotella clara* genome. CAZymes encoded by the most complete PUL identified from *Paraprevotella clara* genome (locus tag HMPREF9441_02099 to 02142) were compared to CAZymes found in the xylan-targeting PUL of *Bacteroides ovatus* ATCC 8483 genome by Rogowski et al. (2015) by BLASTP (117). Multiple GHs, CEs, PLs, and CBMs in *P. clara* loci are not located within the studied PULs in this *B. ovatus* study. Even in cases where GHs from the same families were found, the enzymes presented in both loci have low Per. Identity (%) in protein sequences, questioning the biological and functional purposes of CAZymes encoded by *P. clara*. Data retrieved from BLASTP (201). Accessed 2026-10-21.

***Paraprevotella clara* YIT11840**		***Bacteroides ovatus* ATCC 8483**		**Per. Ident (%)**
CAZymes	**Locus tags**	**Locus tags**	**Function description(s)**	
MFS	HMPREF9441_02099	BACOVA_04388	Xylo-oligosaccharide transporter	26.28
GH10	HMPREF9441_02100	BACOVA_04387 /04390	Periplasmic endo-xylanase	51.24
GH43_1	HMPREF9441_02101	BACOVA_04386	β-xylosidase	76.09
GH67	HMPREF9441_02102	BACOVA_04385	α-glucuronidase	57.33
SusC/SusD-SusC/SusD	HMPREF9441_02105-02106/02109-02110	BACOVA_03427 -03430/04392-04393	(Putative) SusC/SusD-like	26.19
PL6	HMPREF9441_02112	NA	Pectate/alginate/chondroitin-type lyases	
HTCS	HMPREF9441_02118	BACOVA_04394	HTCS sensor-regulator	35.84
GH115C	HMPREF9441_02122	BACOVA_03449	α-glucuronidase	42.90
CBM48-CE1	HMPREF9441_02123	NA	Glycogen/starch binding domain and feruloyl/acetyl esterase	
GH43_29	HMPREF9441_02124-02126	BACOVA_03424	α-L-arabinofuranosidase	41.33
GH43_10	HMPREF9441_02127	BACOVA_03417	α-1,3-specific arabinofuranosidase	59.05
CE6	HMPREF9441_02129	BACOVA_03435	CE6–CE1 bimodular esterase (acetyl and feruloyl esterase)	63.37
CE1	HMPREF9441_02130			22.22
GH95	HMPREF9441_02131	BACOVA_03438	α-1,2-fucosidase	68.87
GH2_9	HMPREF9441_02133	NA	β-galactosidase	
GH146	HMPREF9441_02134	NA	β-L-arabinofuranosidase	

CE20	HMPREF9441_02136	NA	(Putative) arabinoxylan acetylesterase	
GH98-CBM35	HMPREF9441_02137	BACOVA_03433	Endo-xylanase and xylan binding domain	42.80
GH43_35	HMPREF9441_02138	NA	β-endoxylosidase	
GH30-GH43_16-CBM6	HMPREF9441_02139	BACOVA_03432, BACOVA_03424 /03426	Glucuronoxylanas/arabinosidase/xylosidase type and xylan binding domain	28.78
GH2_1	HMPREF9441_02140	NA	β-galactosidase	
GH2_20	HMPREF9441_02142	NA	β-galactosidase	

* NA - Not applicable

xylan fragments. Comparing the putative xylan-active enzymes in *P. clara* to those in *B. ovatus* indicates a range of sequence identities, and also several proteins in the large putative xylan-active PUL of *P. clara* lack homologs in the corresponding locus in *B. ovatus*. Further studies of the *P. clara* enzymes may reveal activities leading to unique, or rare, products that could explain what types of factors could trigger xylan degradation in *P. clara*, if any.

In summary, this work characterized *P. clara* enzymes from GH67 and GH115 and validated them as active α-glucuronidases. *Pc*GH115A and *Pc*GH115C were shown to be functional on GX, releasing 4*O*MeGlcA, both individually and in combination with xylanases. *Pc*GH67 acted as an α-glucuronidase only on GH10-generated fragments, whereas the 'true' substrate of *Pc*GH115B likely remains unidentified, based on its weak activity. Despite these active enzymes, *P. clara* did not grow on any of the prepared BGX-derived, enzymatically predigested substrates. This suggested that the utilization limitation might lie in substrate recognition, uptake, transport, or xylose metabolism instead.

4.3 Development of a genetic toolkit for interspecies PUL transfers

Characterizing individual enzymes such as the GH67 and GH115 α-glucuronidases from *P. clara* in **Paper III** shows what single proteins can do, but not whether a whole PUL, with its transporters, regulators, and accessory enzymes, will work as a unit in a living cell. Because PULs act as coordinated systems, understanding them requires expanding focus from single genes to complete loci and asking which components are necessary and sufficient for growth on a given glycan. This is where genetic tools for transferring the whole PULs become handy.

Functional characterization of PULs has historically relied almost entirely on reverse genetics within a single, native host — most commonly *B. thetaiotaomicron* — via deletion of whole PULs or individual PUL genes via allelic exchange systems such as those based on

the thymidine kinase gene (*tdk*)-based counterselection vector, pExchange-*tdk*, to test whether a locus is necessary for growth on its predicted substrate (202, 203). Experiments as such establish that a PUL is necessary in its native genomic and regulatory context, but they cannot show whether it is sufficient to confer a new degradative capacity in a different genomic background. Transferring the locus from a foreign donor into a tractable recipient such as *B. thetaiotaomicron* would however make such studies possible.

Paper IV attempted to expand the toolkit for PUL studies by developing the pICKUP-based plasmid system, which enables site-specific genomic integration of an entire PUL into a recipient strain, and with repeated transfers possible through recycling of antibiotic markers. The system was validated with two test cases, one PUL targeting mixed-linkage glucan (MLG) and one assumed to target cellulose.

4.3.1 Probing the limits of PUL transfer

Before this work, moving a PUL between strains would have relied on tools built for other purposes. Allelic-exchange plasmids such as pExchange-*tdk* have been used extensively for single-gene insertions or deletions (204). The system is versatile as it can be targeted to virtually any site in the genome. pNBU2-type insertion vectors instead insert their cargo at one of only two native attachment sites (*att)* in *B. thetaiotaomicron*, so each construct uses up a site and an antibiotic marker. The *att* sites are short, specific DNA sequences at which mobile genetic elements such as phages and integrative plasmids insert into a chromosome by site-specific recombination (205, 206). The element carries an *att* site and encodes an integrase, which recognizes the matching sequence on the chromosome (*attB*) and on the element (*attP*), cuts both within the shared sequence, and rejoins them so that the element becomes part of the chromosome. In *Bacteroides*, this is the basis of pNBU2-type vectors, derived from the mobilizable element NBU2, Nonreplicating *Bacteroides* Unit 2, which integrate themselves at *att* sites that lie in tRNA genes (203, 207). *B. thetaiotaomicron* has two such sites, so at most two pNBU2 constructs, each with a different marker, can be inserted in one strain, where the integration positions are fixed. For large inserts such as a whole PUL, the NBU2 integrase was too weakly expressed as demonstrated in Shepherd et al. (2018): in this study, the 20-40 kbp fragments of interest were successfully integrated only after a first pNBU2 construct had been used to boost its integrase expression (21). This restricts both the number of constructs and, for large cargo such as the whole PUL fragments, the efficiency of integration, which motivated tools that would insert at chosen sites by homologous recombination.

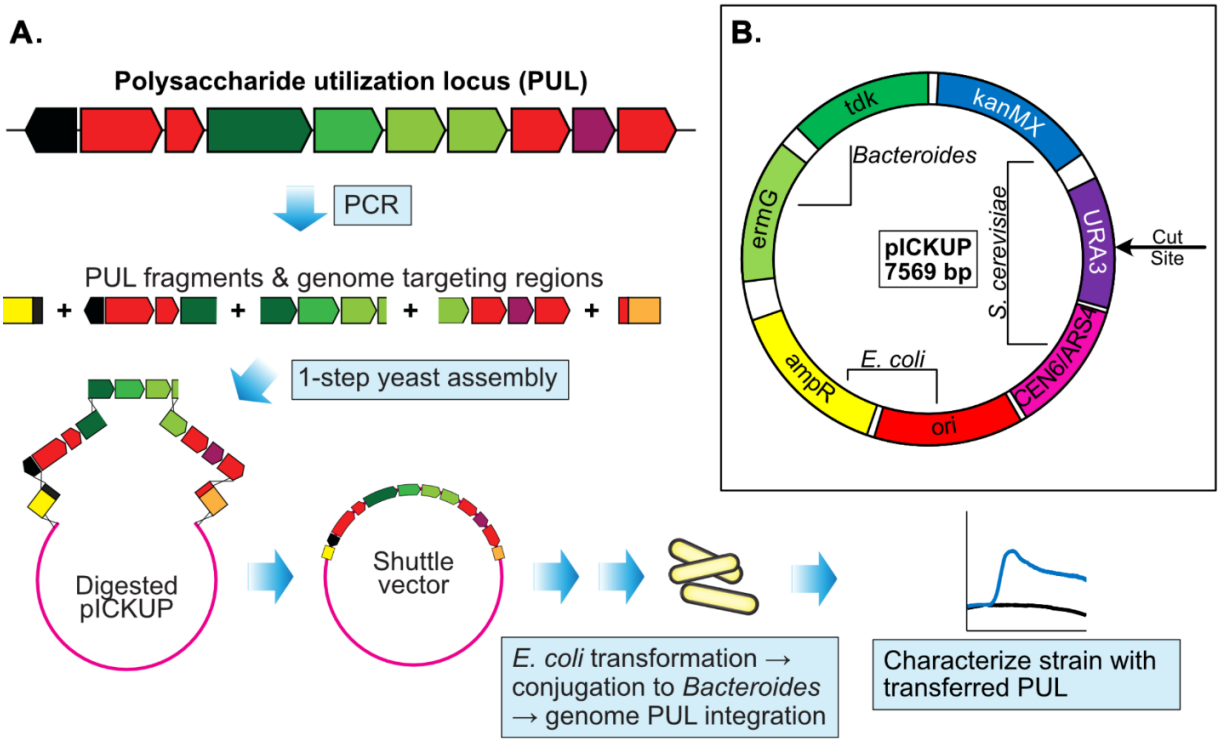


Figure 9. Overview of the pICKUP-based workflow for insertion of foreign PULs into *Bacteroides*. A. Overlapping fragments of the PUL of transfer interest are amplified by PCR. DNA flanking the chosen insertion site (targeting regions) is also amplified with engineered overlaps to the PUL and to the pICKUP backbone vector. The various parts are assembled into a single vector by homologous recombination in *Saccharomyces cerevisiae* prior to transferring the plasmid to *E. coli* and, at last, to the bacterial strain of interest. **B.** Vector map of pICKUP. pICKUP contains factors for replication and selection in both *S. cerevisiae* and *E. coli*, as well as genes for selection and counterselection in *Bacteroides*. Figure adapted from **Paper IV** under the CC BY 4.0 license.

To better enable transfer of large whole loci into chosen positions with the option of repeating the procedure, **Paper IV** developed a pICKUP-based system, a plasmid system for integrating a PUL into a defined chromosomal site by homologous recombination with removal of the plasmid backbone and recycling of the selection marker, so the same markers can be reused in later rounds (Figure 9). In practice, the PUL is amplified in pieces, e.g., for the successful transfer case of MLGUL, three pieces of 5-6 kbp in total were assembled on the plasmid by homologous recombination in *Saccharomyces cerevisiae* CEN.PK113-5D, which was then transformed into *Escherichia coli* S17-1 *λpir* and then conjugated into *B. thetaiotaomicron*. After the first crossover, *tdk*-based counterselection on floxuridine (5-fluoro-2'-deoxyuridine, FUdR) selects the second crossover, which leaves only the insert in the chromosome. The system was first tested with a previously characterized MLGUL from *B. ovatus* (116, 208), which was transferred into *B. thetaiotaomicron*, which cannot grow on

MLG natively. After the PUL transfer, *B. thetaiotaomicron* was able to utilize MLG almost as efficiently as *B. ovatus*, and thus the system successfully conferred a new phenotype into the recipient as hypothesized.

The application of pICKUP was however further tested by extending the transfer to a PUL of uncertain origin, in this case a more distantly related PUL extracted from a metagenome-assembled genome (MAG) from an uncultivated bovine rumen bacterium (209). This putative cellulose utilization locus (CelUL), around 23.7 kbp, was successfully transferred into *B. thetaiotaomicron*. This locus had no prior functional characterization as a whole system, but encodes a SusC/SusD-like transporter as binding-protein pair (CelC/CelD), an additional SGBP CelE, a predicted regulator CelR, a predicted inner-membrane transporter CelT, with three glycoside hydrolases, Cel5, Cel9, and Cel94.

Among them, Cel5, Cel9, CelD, and CelE had been studied biochemically (210-212) as isolated proteins. Despite the successful pICKUP-based transfer into *B. thetaiotaomicron*, the resulting strain was not found grow on cellulose. Instead, it was however able to utilize cello-oligosaccharides like cellobiose and cellotriose. The PUL transfer enabled further studies, such as determination of its regulation, and the CelUL was shown to be induced merely about 30- to 70-fold higher by cellobiose compared to its growth in minimal medium with glucose. Previously studied native PULs are instead expected to be upregulated by hundreds to thousands of folds (3, 105, 115, 213).

A series of genetic dissections of the transferred locus was done, using the same allelic-exchange and integration workflow as the transfer itself to delete, complement, or overexpress the proteins within the locus. Deleting *cel9* produced a lag phase during growth on cellotriose, while overexpressing it from a constitutive *B. thetaiotaomicron* promoter shortened the lag and improved growth. The predicted MFS transporter of the CelUL, CelT, has no counterpart in the canonical Sus system, but it is proposed to move a phosphorylated product of Cel94, glucose-1-phosphate, across the inner membrane. Overexpressing *celT* using the strong *B. thetaiotaomicron* promoter P_BfP2E5 improved both growth rate and the final cell density grown on cellobiose and cellotriose (214). A double mutant with both upregulated *cel9* and *celT* however did not further improve the phenotype.

The two examples of MLGUL and CelUL transfers together hinted the potential importance of the phylogenetic relatedness of donor and recipient to a successful and interpretable PUL transfer, a variable that single-host reverse genetics cannot reveal. Showing that a locus from an uncultivated rumen bacterium could function, at least partially, in a human gut *Bacteroides,* illustrated that polysaccharide-degradation strategies are conserved to a degree in nature.

4.3.2 Optimization of PUL transfer efficiency

The cause of the GMUL transfer failure is unclear. That the PUL to be transferred was too large of a cargo is an unlikely explanation, as the CelUL (~24,000 bp) is only about 5.5 kb shorter than the GMUL (~29,000 bp), and each half of the GMUL (about 14.6 kb) was smaller than the CelUL and of similar size to the full MLGUL (about 16 kbp). The second GMUL transfer step might have not recombined precisely for a seamless joint in the genome, making the complete GMUL partially disrupted. As the two-step recombination of split GMUL was designed taking advantage of the overlapping targeting regions carried by the two pICKUP vectors in two transfers, the insertion-site efficiency based on base-pair overlapping in the pICKUP system could be improved possibly by lengthening the homologous genome targeting arms. Different ways of splitting the GMUL could further be assessed to see if the two-step transfer would be successful.

By moving complete loci into a genetically tractable host, for instance, using the pICKUP system described here in **Paper IV**, many new research lines should become available. The transferred PUL can, in a genetically tractable host, be further modified, by for instance deleting, swapping, or adding individual genes, and assess the resulting phenotypes. Potentially, the system can be used also for more applied studies to create designer strains with chosen abilities to grow on different types of polysaccharides, either for gut systems or for production of metabolites.

Chapter 5: Conclusions and outlook

This thesis aimed to explore carbohydrate metabolism in various Bacteroidota species, from determining which carbohydrates they can grow on and which OAs they produce as a result, and furthermore how their carbohydrate-degradation abilities can be explained or engineered.

The thesis started with profiling of the extracellularly secreted metabolites produced by both well-studied *Bacteroides* species but also less-studied organisms from the *Prevotellaceae* family. In **Paper I** and **Paper II**, I could demonstrate how different carbohydrates used as carbon sources led to different extracellular metabolite output, both intra- and interspecies. Interestingly, the original descriptions of many *Prevotellaceae* species report succinate and acetate as the family's main fermentation products, which was not observed in **Paper II**. As described earlier, most physiological and metabolic profiling of Bacteroidota species has been done with glucose as the carbon source, which is probably due to practical reasons. It is a defined, soluble, cheap, and simple hexose that essentially all strains can consume, and allows quick comparison across species, while complex poly-saccharides instead are typically heterogeneous and sometimes even insoluble, leading to issues with standard measurements of optical density. The main take-home message from both **Paper I** and **Paper II** is that the metabolic output of carbohydrate-consuming cells is not a static trait of a species. Whether similar results with varied OA production should be expected from all gut Bacteroidota species, or furthermore when extending the studies to other types of bacteria, remains to be tested.

It is also important to stress that the results from these two metabolite profiling studies only addressed how single species behave on single purified substrates. The gut environment, as well as the vast majority of other environments, can be highly variable and dynamic. The nutrient flow and metabolic exchanges do not only happen for isolated bacteria, but also between the different types of bacteria in the community and with the host. In single-species studies, metabolites may accumulate over time, which likely does not happen in the gut, and this might also have affected the strains used in these studies. For instance, a previous study from Catlett et al. (2020) (84), reported that acetate accumulation would suppress the growth of *B. thetaiotaomicron*.

Future studies building on **Paper I** and **Paper II** should include defined multi-species communities firstly on purified substrates, then mixed fibers, and eventually in *in vivo* models, to show whether the substrate-dependent OA production patterns seen in mono-culture are sustained during cross-feeding and/or competition, as such interactions among *Bacteroides* have previously been reported (166).

A surprise in **Paper II** was the lack of growth of the studied species on the majority of polysaccharide substrates, despite apparently abundant CAZymes encoded in their genomes.

The characterization of α-glucuronidases from *P. clara* in **Paper III**, including determination of their activities as single enzymes and combined activity with xylanases, showed the enzymes to be fully functional *in vitro*. Despite testing a range of xylan- and xylan-derived potential carbon sources, as well as simple xylose, I was still unable to demonstrate growth of *P. clara*. The reasons for this remain obscure, and further analysis of the putative xylose metabolism genes in the species suggests an ability to metabolize this sugar. Further studies of these putative genes/enzymes would be needed to verify that they function as expected, or of other related proteins such as putative xylose/xylo-oligosaccharide transporters. If some of these targets do not function as predicted, it could possibly explain the lack of growth of the species. It is also possible that the right inducers for xylan uptake are yet to be discovered, and that the source type of xylan used in these studies did not include the right motifs.

Finally, in **Paper IV**, a new system was introduced that enabled whole-PUL transfer into *B. thetaiotaomicron*. The transfer of the MLGUL from *B. ovatus* resulted in a strain with hemicellulose-degrading ability, which is absent in the wildtype organism, and transfer of the metagenomic CelUL enabled growth on cello-oligosaccharides. While the latter was not as exciting as cellulose utilization would have been, which was the predicted role of the PUL, it nonetheless showcases how transfer of a PUL of unknown function enables testing *in vivo* and also further genetic manipulations of the PUL. The study highlights how carbohydrate degradation ability can be transferred between species, which could be way to create designer strains in the future.

The whole-PUL transfer method developed here, however, still has some apparent limitations yet to be investigated. While the MLGUL and the CelUL were successfully transferred into *B. thetaiotaomicron*, the attempts to similarly transfer the *B. ovatus* GMUL was unsuccessful. It remains to be evaluated if the limitation was its size, a faulty design of the intended two-step transfer, or some other reason. Whole-genome sequencing after each transfer step could possibly reveal if unwanted PUL- or genome rearrangements happened in the process. A straightforward way to test the possibilities to conduct double transfers into one recipient strain could be done by splitting the MLGUL, since we know it should result in a clear phenotype when transferred intact into the recipient strain. It would also be highly interesting to investigate the previously mentioned aspect of genetic relatedness for the donor-recipient choice. By testing PUL transfer from a range of genetically close to distant donors into the recipient could further illuminate the impact of other cellular factors, like codon usage or promoter functionality, for successful PUL transfer, as such factors likely become less efficiently recognized the less similar the donor- and recipient strains are. Possibly, the system could also be used to assess whether the putative xylan-active PUL from *P. clara* is functional, thus tying together the studies from **Paper II**, **Paper III**, and **Paper IV**.

Tying together the metabolite profiling studies with the ability to do PUL transfers would additionally be of high interest. How transfer of a PUL, and thus a phenotype, into a recipient strain affects its OA output is untested, but potentially new strains with both desired degradation abilities and OA production could be a future possibility as a result from these studies. Many OAs are important both for gut health, but are also regarded as key platform chemicals, and thus such designer strains could also have a role in more applied settings.

Overall, this thesis has contributed to the knowledge on how various bacteria, especially those commonly found in the human gut, respond to different nutrients. The results illustrate a previously somewhat overlooked complexity in these species, showing that their physiological responses are not static during growth on glycans. The results presented in this thesis could be highly important to improve future prediction of gut microbiota interactions, and potentially also other environments, as well as creation of novel strains relevant for both health and industry.

Acknowledgements

Four years and a bit more in a flash – words cannot fully unravel the emotion I feel at this moment and could barely describe how enriched my doktorand life has been.

Before anything, a heartfelt thank you to everyone who helped me draft the thesis. It would have been impossible without any of you.

First and foremost, I would like to express my greatest gratitude towards my supervisor Johan and my co-supervisor Verena. The stretch from Monday evening until Tuesday morning, our routine meeting time, was like an emotional-rollercoaster to me. I (almost always) entered the office dreading it with anxieties, while always left feeling mentally rebooted. Your full presence at all stages of the project has helped me believe I could see this Ph.D. journey through, even if most of the time I felt drowned in doubts and worries. Both scientifically and for my personal development, thank you deeply for your firm support, inspiring guidance, and patience throughout the supervision. Our talks have always given me a lot of strength to carry on, and I am truly thankful for you giving me the opportunity to work with you.

I would also like to thank my master thesis supervisor Lauren, who inspired me to pursue a Ph.D. years back by showing me how fun and fulfilling it can be to do science with heart. And, certainly, to my people in Stockholm: you are the reason that made me decide to stay in Sweden, and I look forwards to more memories with you, wherever and whenever that is.

My sincere thanks also go to my examiner Calle, who has been present since the beginning of the adventure, offering support for the project but also genuine care. Your encouragement helps me push through difficulties and build confidence little by little.

Many thanks also to my opponent and the grading committee members, for spending time on evaluating the thesis and taking part in the defense.

It is without a doubt my pleasure to be working at Chalmers, and more particularly here at IndBio among such amazing co-workers. I have always felt lucky to be able to start my first official employment at IndBio. From everyone I have had the pleasure of crossing paths with, I have learned something, professionally or from the perhaps shortest but kindest of encounters. I am very grateful to be surrounded by people with good hearts, and these are the memories I carry with me wherever the future leads.

It is certainly my pleasure to be part of *the Maplers*, to whom I am very thankful, especially for the rounds and rounds of proof-reading my messy thesis drafts over the past weeks. Being part of the team brings me so much joy to routine works, thanks to the great team atmosphere we share – one less non-barker barking occasionally in the next group meeting soon! Special thanks are extended to Scott, with whom I have the honor of collaborating and learning (have to learn from the best!). Thank you for your supportive encouragement that guided me

through panicking moments, and for opening up the fascinating world of enzymes to me – maximizing the discovery! Special thanks to the ex-non-barker Tove, who was there at the very first second I stepped into IndBio for the interview and, hopefully not but maybe, has been traumatized by my awkwardness ever since. Thank you for your contagious calmness, which never fails to bring me calm happiness.

I would also like to thank all the administrative staff of IndBio and in LIFE, above all, the one and only, the best Gunillas, who have helped me with administrative documentation and kindly rescued me numerous times when I forgot my door keys (typical). Huge thanks to our research engineers, Vijay and Pun, who rescued me not only when I forgot my key (it is not pathological), but also when I was in despair in-and-out the lab, although at times the issue turned out to be just me being eye-crossed, for which I am incredibly sorry. Your efforts make the easily messy lab work less unmanageable. Thank you for all your assistance throughout the years.

Approaching the finale of a journey often comes with greater challenges. To me, the start of a time is no easier either. My sincere gratitude to those who were there at the beginning of my Ph.D. life. Professionally or/and privately, you witnessed, perhaps tolerated, and helped me evolve to the (hopefully) better version of me now. Thank you greatly for simply being there when I felt clueless and helpless.

Nathan, thank you for initiating an interesting project that later introduced me to the *magical* Bacteroidota et al. There were countless times you reached out at the exact moment I needed help, and with patience you cleared up my confusion again and again. Thank you for making time for a newbie in genetic engineering like me back then.

Kamesh, one of the top ten (or perhaps three) most frequent conversations we have had was you saying you would not speak to me and me carrying on anyway. Thank you sincerely nonetheless, for being present and being my lab buddy when I first arrived at IndBio. My Gothenburg experience in general would have been very different without you - thank you!

Bohyun, being punctuated by you was an underestimated fun task at work which I did not know existed. The office hours spent discussing miscellaneous topics were recharging, and I could always turn to you for professional advice. Thank you for putting up with my randomness and just letting it be, and for being an amazing office-mate.

There are two Aquariuses of all that could really make me snap in seconds, yet I am still deeply thankful to —

Teddy was actually where the story began, though he was never truly in the story. Immediately after your — now to think back — absentminded encouragement: “It will be fun to work at the same city again.”, I applied for the only position in a foreign city on my list years ago (and it turns out to be a wonderful decision:)). All this, only for you to casually

remark around a year into my time in Gothenburg, “By the way, in the end I didn’t apply for the position there.” Thank you truly for always being there for me.

Andrés always finds ways to tease me relentlessly, and it would be as if all the existing or non-existing rage in me were ignited at once. However, at my lowest you were also the one asking around back home for me just because I missed your calls with my phone on sleep mode. Thank you for being present, and I am delighted that we are in the same time zone.

Also, Wayne — not an Aquarius finally, who took a detour of hours just to check on how I was doing in life. Thank you for caring for me all this time; you always turn care into action.

In those low moments, or when all purposes seem vague, Alex, Andrea, Branda, Chloe, Dany, Elin, Facu, Julie, Linus, Luisa, Marta, Marisol, Nathálie, Ric, Roger, Sara, Tove, you bring me comfort by staying a little delusional with me, yet not forgetting to gently slap me in the face at last and pick me back up (maybe more of Julie and Nathálie, also Chloe). Thank you wholeheartedly for all the good times and memories we share, and for being my warmest source of emotional support.

Facu and Andrés el Hombre, you always find ways to mock and tease me wherever, whenever, and for whatever unbelievable reason, but you were also the first to notice when I subconsciously withdrew into self-isolation and actually tried to cheer me up, for which I am truly grateful. Besides, you don’t know how many times I have metaphorically sacrificed you in my dark magic for the good of all humankind; you also don’t know if your autoclaved water is actually sterile — who could know?

VerBon, Punchi, Tommy, Chloaki, Mr. Leman, and Doro, who are my go-to people for whatever (sane or not) hang-out ideas I bring up out of nowhere. Thank you for adding joy to my days. The only thing we have not (yet) agreed on so far is spray tanning (have you really ever been sprayed with a wall-nozzle?). We will come back to this later.

Andrea, you are the afternoon sunlight, like, upon breathing in the brisk air, you know the day is such a good day. And Julie, with you I always feel so loved, secure, and understood. And then there is Arne, who might have contributed to my caffeine addiction at work. And, of course, Maata C., who would never have stopped bickering about this forever if I had chosen not to mention her — thank you for the laughter, tears, ridiculous/hangry/cozy moments we share. I treasure them greatly.

A few months prior to this last mile of my Ph.D. time, I repeatedly tried warning Max about my potential mood swings. At times I could not silence my work anxieties, even just temporarily. Even so, every morning he still makes us perfect coffee, listens patiently, and calms me down whenever I start murmuring again. Thank you dearly for being there for me. You remind me of how beautiful life has always been, and I am glad we get to appreciate it together. The other day he told me he had a nightmare about carbohydrate structures.

The kindness I receive from my family is what supports me wherever and whenever. Thank you for your unwavering love and trust. I am a little less afraid on the way to where I would like to be, knowing you are with me. Of all, Echo somehow makes the time-zone difference nonexistent, as we can easily be on goodnight calls at my bedtime and pick the conversation up again knowing you are already up. You give me tremendous strength throughout all these years abroad, even if we are physically far apart. I love you dearly, my dearest sister. I miss you!

預設的字體真的不是最好看。我早上在猶豫要不要在句子裡隨機加「早餐店小籠包」。

對我來說，想家是生活裡難以承擔的念頭；不敢想起也不敢想——愛的重量。謝謝家人們的愛。你們一直以來給予的溫暖支持，讓我說真的異想天開的腦袋得以實現那些某天突然想做的事。大部分的時候我也不曉得這些想法的來由，但因為有你們的愛，我很幸運能隨心所往。我最不實際的幻想也許是盼望任意門的到來，我們就可以常一起吃飯；很多時候真的只是想吃頓飯就好。

想所有家人們。想下次做什麼呢？想阿公、阿婆，想爸爸、媽媽，想欸口；愛你們。

Lastly, thank you to whoever has read it this far. You make my heart full.

References

1. Qin, J., Li, R., Raes, J., Arumugam, M., and Burgdorf, K. S. (2010) A human gut microbial gene catalogue established by metagenomic sequencing *Nature* 464, 59-65
2. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide Utilization Loci: Fueling Microbial Communities *Journal of Bacteriology* 199, e00860-00816
3. 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 Biology* 9, e1001221
4. Koh, A., De Vadder, F., Kovatcheva-Datchary, P., and Bäckhed, F. (2016) From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites *Cell* 165, 1332-1345
5. Hirmas, B., Gasaly, N., Orellana, G., Vega-Sagardía, M., Saa, P., Gotteland, M. *et al.* (2022) Metabolic modeling and bidirectional culturing of two gut microbes reveal cross-feeding interactions and protective effects on intestinal cells *mSystems* 7, e0064622
6. Louis, P., Duncan, S. H., Sheridan, P. O., Walker, A. W., and Flint, H. J. (2022) Microbial lactate utilisation and the stability of the gut microbiome *Gut Microbiome* 3, e3
7. Fischbach, M. A., and Sonnenburg, J. L. (2011) Eating for two: how metabolism establishes interspecies interactions in the gut *Cell Host & Microbe* 10, 336-347
8. Deehan, E. C., Yang, C., Perez-Muñoz, M. E., Nguyen, N. K., Cheng, C. C., Triador, L. *et al.* (2020) Precision microbiome modulation with discrete dietary fiber structures directs short-chain fatty acid production *Cell Host & Microbe* 27, 389-404.e386
9. Makki, K., Deehan, E. C., Walter, J., and Bäckhed, F. (2018) The impact of dietary fiber on gut microbiota in host health and disease *Cell Host & Microbe* 23, 705-715
10. Kolodziejczyk, A. A., Zheng, D., and Elinav, E. (2019) Diet-microbiota interactions and personalized nutrition *Nature Reviews Microbiology* 17, 742-753
11. Hamaker, B. R., and Tuncil, Y. E. (2014) A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota *Journal of Molecular Biology* 426, 3838-3850
12. Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012) How glycan metabolism shapes the human gut microbiota *Nature Reviews Microbiology* 10, 323-335
13. Porter, N. T., and Larsbrink, J. (2022) Investigation and alteration of organic acid synthesis pathways in the mammalian gut symbiont *Bacteroides thetaiotaomicron Microbiology Spectrum* 10, e0231221
14. Mayhew, J. W., Onderdonk, A. B., and Gorbach, S. L. (1975) Effects of time and growth media on short-chain fatty acid production by *Bacteroides fragilis Applied Microbiology* 29, 472-475
15. Macfarlane, S., and Macfarlane, G. T. (2003) Regulation of short-chain fatty acid production *Proceedings of the Nutrition Society* 62, 67-72
16. Cosgrove, D. J. (2024) Structure and growth of plant cell walls *Nature Reviews Molecular Cell Biology* 25, 340-358

17. Broekaert, W. F., Courtin, C. M., Verbeke, K., Van de Wiele, T., Verstraete, W., and Delcour, J. A. (2011) Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides *Critical Reviews in Food Science and Nutrition* 51, 178-194
18. Rogowski, A., Baslé, A., Farinas, C. S., Solovyova, A., Mortimer, J. C., Dupree, P. *et al.* (2014) Evidence that GH115 α-glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility *Journal of Biological Chemistry* 289, 53-64
19. Nurizzo, D., Nagy, T., Gilbert, H. J., and Davies, G. J. (2002) The structural basis for catalysis and specificity of the *Pseudomonas cellulosa* alpha-glucuronidase, GlcA67A *Structure* 10, 547-556
20. 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
21. Shepherd, E. S., DeLoache, W. C., Pruss, K. M., Whitaker, W. R., and Sonnenburg, J. L. (2018) An exclusive metabolic niche enables strain engraftment in the gut microbiota *Nature* 557, 434-438
22. Prescott, S. L. (2017) History of medicine: Origin of the term microbiome and why it matters *Human Microbiome Journal* 4, 24-25
23. Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T. *et al.* (2020) Microbiome definition re-visited: old concepts and new challenges *Microbiome* 8, 103
24. Marchesi, J. R., and Ravel, J. (2015) The vocabulary of microbiome research: a proposal *Microbiome* 3, 31
25. Lu, Z., and Imlay, J. A. (2021) When anaerobes encounter oxygen: mechanisms of oxygen toxicity, tolerance and defence *Nature reviews microbiology* 19, 774-785
26. Takeuchi, O., Hoshino, K., Kawai, T., Sanjo, H., Takada, H., Ogawa, T. *et al.* (1999) Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components *Immunity* 11, 443-451
27. Silhavy, T. J., Kahne, D., and Walker, S. (2010) The bacterial cell envelope *Cold Spring Harbor perspectives in biology* 2, a000414
28. Book, A. J., Lewin, G. R., McDonald, B. R., Takasuka, T. E., Doering, D. T., Adams, A. S. *et al.* (2014) Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose *Applied and Environmental Microbiology* 80, 4692-4701
29. De Boer, W., Folman, L. B., Summerbell, R. C., and Boddy, L. (2005) Living in a fungal world: impact of fungi on soil bacterial niche development *FEMS Microbiology Reviews* 29, 795-811
30. Larsbrink, J., Tuveng, T. R., Pope, P. B., Bulone, V., Eijsink, V. G. H., Brumer, H. *et al.* (2017) Proteomic insights into mannan degradation and protein secretion by the forest floor bacterium *Chitinophaga pinensis Journal of Proteomics* 156, 63-74
31. McKee, L. S., Martinez-Abad, A., Ruthes, A. C., Vilaplana, F., and Brumer, H. (2019) Focused metabolism of beta-glucans by the soil Bacteroidetes species *Chitinophaga pinensis Applied and Environmental Microbiology* 85, e02231-02218
32. 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 *The ISME Journal* 13, 2800-2816
33. Bennke, C. M., Krüger, K., Kappelmann, L., Huang, S., Gobet, A., Schüler, M. *et al.* (2016) Polysaccharide utilisation loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic *Environmental Microbiology* 18, 4456-4470
34. Quehenberger, J., Shen, L., Albers, S.-V., Siebers, B., and Spadiut, O. (2017) *Sulfolobus* - a potential key organism in future biotechnology *Frontiers in Microbiology* 8, 2474
35. Klein, A. M., Bohannan, B. J. M., Jaffe, D. A., Levin, D. A., and Green, J. L. (2016) Molecular evidence for metabolically active bacteria in the atmosphere *Frontiers in Microbiology* 7, 772
36. Güllert, S., Fischer, M. A., Turaev, D., Noebauer, B., Ilmberger, N., Wemheuer, B. *et al.* (2016) Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies *Biotechnology for biofuels* 9, 121
37. Campanaro, S., Treu, L., Rodriguez-R, L. M., Kovalovszki, A., Ziels, R. M., Maus, I. *et al.* (2020) New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters *Biotechnology for biofuels* 13, 25
38. Ozbayram, E. G., Ince, O., Ince, B., Harms, H., and Kleinsteuber, S. (2018) Comparison of rumen and manure microbiomes and implications for the inoculation of anaerobic digesters *Microorganisms* 6, 15
39. Rigottier-Gois, L. (2013) Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis *The ISME Journal* 7, 1256-1261
40. Rivera-Chavez, F., Lopez, C. A., and Baumler, A. J. (2017) Oxygen as a driver of gut dysbiosis *Free Radical Biology and Medicine* 105, 93-101
41. Winter, S. E., Lopez, C. A., and Baumler, A. J. (2013) The dynamics of gut-associated microbial communities during inflammation *EMBO Reports* 14, 319-327
42. Engel, P., and Moran, N. A. (2013) The gut microbiota of insects - diversity in structure and function *FEMS Microbiology Reviews* 37, 699-735
43. Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S. *et al.* (2008) Evolution of mammals and their gut microbes *Science* 320, 1647-1651
44. Sender, R., Fuchs, S., and Milo, R. (2016) Revised estimates for the number of human and bacteria cells in the body *PLoS Biology* 14, e1002533
45. Human Microbiome Project, C. (2012) Structure, function and diversity of the healthy human microbiome *Nature* 486, 207-214
46. Byrd, A. L., Belkaid, Y., and Segre, J. A. (2018) The human skin microbiome *Nature Reviews Microbiology* 16, 143-155
47. Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S., McCulle, S. L. *et al.* (2011) Vaginal microbiome of reproductive-age women *Proceedings of the National Academy of Sciences* 108, 4680-4687
48. Procházková, N., Falony, G., Dragsted, L. O., Licht, T. R., Raes, J., and Roager, H. M. (2023) Advancing human gut microbiota research by considering gut transit time *Gut* 72, 180-191

49. Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., and Versalovic, J. (2014) The placenta harbors a unique microbiome *Science Translational Medicine* 6, 237ra265
50. De Goffau, M. C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S. J. *et al.* (2019) Human placenta has no microbiome but can contain potential pathogens *Nature* 572, 329-334
51. Ding, T., and Schloss, P. D. (2014) Dynamics and associations of microbial community types across the human body *Nature* 509, 357-360
52. Vuik, F., Dicksved, J., Lam, S., Fuhler, G., van der Laan, L., van de Winkel, A. *et al.* (2019) Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals *United European Gastroenterology Journal* 7, 897-907
53. Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H. *et al.* (2010) The human oral microbiome *Journal of Bacteriology* 192, 5002-5017
54. Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F. *et al.* (2006) Molecular analysis of the bacterial microbiota in the human stomach *Proceedings of the National Academy of Sciences of the United States of America* 103, 732-737
55. Kyrochristou, I., Kyrochristou, G., Fousekis, F., Katsanos, K., Schizas, D., Vlachos, K. *et al.* (2026) Determinants of the healthy gut microbiome: core features, modifying factors and normal functions *Annals of Gastroenterology* 39, 191
56. Johnson, E. L., Heaver, S. L., Walters, W. A., and Ley, R. E. (2017) Microbiome and metabolic disease: revisiting the bacterial phylum Bacteroidetes *Journal of Molecular Medicine* 95, 1-8
57. Shin, N. R., Whon, T. W., and Bae, J. W. (2015) Proteobacteria: Microbial signature of dysbiosis in gut microbiota *Trends in Biotechnology*
58. Hughes, E. R., Panzetta, M. E., Sharma, A., and Valdivia, R. H. (2025) The biology of *Akkermansia Annual Review of Microbiology* 79
59. Snodgrass, J. L., and Velayudhan, B. T. (2026) Butyrate-producing bacteria as a keystone species of the gut microbiome: a systemic review *International Journal of Molecular Sciences* 27(3), 1289
60. Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M. *et al.* (2005) Diversity of the human intestinal microbial flora *Science* 308, 1635-1638
61. Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R. *et al.* (2014) Human genetics shape the gut microbiome *Cell* 159, 789-799
62. Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D. *et al.* (2018) Environment dominates over host genetics in shaping human gut microbiota *Nature* 555, 210-215
63. Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R. *et al.* (2011) Enterotypes of the human gut microbiome *Nature* 473, 174-180
64. Carter, M. M., Olm, M. R., Merrill, B. D., Dahan, D., Tripathi, S., Spencer, S. P. *et al.* (2023) Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes *Cell* 186, 3111-3124. e3113
65. Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S., and Sonnenburg, J. L. (2016) Diet-induced extinctions in the gut microbiota compound over generations *Nature* 529, 212-215

66. De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S. *et al.* (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa *Proceedings of the National Academy of Sciences of the United States of America* 107, 14691-14696
67. Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F. D., & Lewis, J. D. (2011) Linking long-term dietary patterns with gut microbial enterotypes *science* 334, 105-108
68. Gorvitovskaia, A., Holmes, S. P., and Huse, S. M. (2016) Interpreting *Prevotella* and *Bacteroides* as biomarkers of diet and lifestyle *Microbiome* 4, 15
69. Senghor, B., Sokhna, C., Ruimy, R., and Lagier, J.-C. (2018) Gut microbiota diversity according to dietary habits and geographical provenance *Human Microbiome Journal* 7, 1-9
70. Ze, X., Duncan, S. H., Louis, P., and Flint, H. J. (2012) Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon *The ISME Journal* 6, 1535-1543
71. 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
72. Gellman, R. H., Olm, M. R., Terrapon, N., Enam, F., Higginbottom, S. K., Sonnenburg, J. L. *et al.* (2023) Hadza *Prevotella* require diet-derived microbiota-accessible carbohydrates to persist in mice *Cell reports* 42
73. Ravcheev, D. A., Godzik, A., Osterman, A. L., and Rodionov, D. A. (2013) Polysaccharides utilization in human gut bacterium Bacteroides thetaiotaomicron: comparative genomics reconstruction of metabolic and regulatory networks *BMC Genomics* 14, 873
74. Gough, A. M., Anita C. Parker, Patricia J. O'Bryan, Terence R. Whitehead, Sourav Roy, Brandon L. Garcia, Paul S. Hoffman, C. Jeffrey Smith, and Edson R. Rocha. (2024) Novel roles for pirin Proteins and 2-ketoglutarate: ferredoxin oxidoreductase in *Bacteroides fragilis* central metabolism *bioRxiv*
75. Basen, M., and Kurrer, S. E. (2021) A close look at pentose metabolism of gut bacteria *The FEBS journal* 288, 1804-1808
76. Clausen, U., Vital, S.-T., Lambertus, P., Gehler, M., Scheve, S., Wöhlbrand, L. *et al.* (2024) Catabolic network of the fermentative gut bacterium *Phocaeicola vulgatus* (phylum Bacteroidota) from a physiologic-proteomic perspective *Microbial Physiology* 34, 88-107
77. Gindt, M. E., Lück, R., and Deppenmeier, U. (2024) Genetic optimization of the human gut bacterium *Phocaeicola vulgatus* for enhanced succinate production *Applied Microbiology and Biotechnology* 108, 465
78. Franke, T., and Deppenmeier, U. (2018) Physiology and central carbon metabolism of the gut bacterium *Prevotella copri Molecular Microbiology*
79. Lück, R., and Deppenmeier, U. (2022) Genetic tools for the redirection of the central carbon flow towards the production of lactate in the human gut bacterium *Phocaeicola* (*Bacteroides*) *vulgatus Applied microbiology and biotechnology* 106, 1211-1225

80. Milani C, L. G., Duranti S, Turroni F, Mancabelli L, Ferrario C, Mangifesta M, Hevia A, Viappiani A, Scholz M, Arioli S. (2015) Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut *Scientific Reports* 28;5(1):15782
81. Shaheen, N., Khursheed, W., Gurung, B., & Wang, S. (2025) *Akkermansia muciniphila*: A key player in gut microbiota-based disease modulation *Microbiological research* 128317
82. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C. *et al.* (2003-3-28) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis *Science* 299
83. Flint, H. J., Scott, K. P., Louis, P., and Duncan, S. H. (2012) The role of the gut microbiota in nutrition and health *Nature Reviews Gastroenterology & Hepatology* 9, 577-589
84. Catlett Jennie, L., Catazaro, J., Cashman, M., Carr, S., Powers, R., Cohen Myra, B. *et al.* (2020) Metabolic feedback inhibition influences metabolite secretion by the human gut symbiont *Bacteroides thetaiotaomicron mSystems* 5
85. Keitel, L., Schick, B., Pohen, G., Yordanov, S., and Büchs, J. (2025) Online monitored characterization of *Phocaeicola vulgatus* for organic acid production using anaerobic microtiter plate cultivations *Biotechnology Progress* 41, e3526
86. Glazunova, E., Molodtsova, P., Grabarnik, I., Kurnosov, A., Bikaeva, I., Shipulin, G., & Zlobovskaya, O. (2025) Healthy human gut microbiome: Towards standardized research *AIMS Microbiology* 11(4), 786
87. Sola, L., Candeliere, F., Busi, E., Raimondi, S., Amaretti, A., & Rossi, M. (2026) A genomic atlas of gut clostridia: phylogeny, butyrate, and propionate production *Frontiers in Microbiology* 17, 1761627
88. Winter, S. E., and Bäumler, A. J. (2023) Gut dysbiosis: ecological causes and causative effects on human disease *Proceedings of the National Academy of Sciences* 10.1073/pnas.2316579120
89. Tingler, A. M., & Engevik, M. A. (2205) Breaking down barriers: is intestinal mucus degradation by *Akkermansia muciniphila* beneficial or harmful? *93(9), e00503-24*
90. Aja, E., Zeng, A., Gray, W., Connelley, K., Chaganti, A., & Jacobs, J. P. (2025) Health Effects and Therapeutic Potential of the Gut Microbe *Akkermansia muciniphila Nutrients* 17(3), 562
91. Wucher, B. R., Carlos J. Pardo-De la Hoz, Isaac Stamper, Sneh Sharma, Diego Kaune, Pranali Bendale, Jonathan Peled, and Joao B. Xavier. (2026) Metabiosis underlies a microbiota permissive to Pseudomonadota and increases the risk of gut-borne bloodstream infection *bioRxiv*
92. Binda, C., Lopetuso, L. R., Rizzatti, G., Gibiino, G., Cennamo, V., & Gasbarrini, A. (2018) Actinobacteria: a relevant minority for the maintenance of gut homeostasis *Digestive and Liver Disease* 50(5), 421-428
93. Precup, G., and Vodnar, D. C. (2019) Gut *Prevotella* as a possible biomarker of diet and its eubiotic versus dysbiotic roles *British Journal of Nutrition*
94. Abdelsalam, N. A., Hegazy, S. M., and Aziz, R. K. (2023) The curious case of *Prevotella copri Gut Microbes* 15(2), 2249152
95. Aldunate, M., Srbinovski, D., Hearps, A. C., Latham, C. F., Ramsland, P. A., Gugasyan, R. *et al.* (2015) Antimicrobial and immune modulatory effects of lactic

acid and short chain fatty acids produced by vaginal microbiota associated with eubiosis and bacterial vaginosis *Frontiers in Physiology* Volume 6 - 2015
96. Lemos, J. A., Palmer, S. R., Zeng, L., Wen, Z. T., Kajfasz, J. K., Freires, I. A. *et al.* (2019) The biology of *Streptococcus mutans Microbiology Spectrum* 7
97. Deleu, S., Machiels, K., Raes, J., Verbeke, K., and Vermeire, S. (2021) Short chain fatty acids and its producing organisms: An overlooked therapy for IBD? *EBioMedicine* 66
98. Stoeva, M. K., Jeewon Garcia-So, Nicholas Justice, Julia Myers, Surabhi Tyagi, Madeleine Nemchek, Paul J. McMurdie, Orville Kolterman, and John Eid (2021) Butyrate-producing human gut symbiont, *Clostridium butyricum*, and its role in health and disease *Gut microbes* 13, no. 1 (2021): 1907272
99. Facchin, S., Calgaro, M., and Savarino, E. V. (2025) Rethinking short-chain fatty acids: a closer look at propionate in inflammation, metabolism, and mucosal homeostasis *Cells* 14, 1130
100. Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., DuGar, B. *et al.* (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease *nature* 472, 57-63
101. Bhattacharjee, D., Millman, L. C., Seesengood, M. L., Martineau, L. M., & Seekatz, A. M. (2026) Genomic insights into the functional and metabolic versatility of gut microbiome *Anaerostipes* species *Microbial Genomics* 12(2), 001617
102. Dordević, D., Jančíková, S., Vítězová, M., & Kushkevych, I. (2021) Hydrogen sulfide toxicity in the gut environment: Meta-analysis of sulfate-reducing and lactic acid bacteria in inflammatory processes *Journal of advanced research* 27, 55-69
103. Sonnenburg, E. D., Zheng, H., Joglekar, P., Higginbottom, S. K., Firbank, S. J., Bolam, D. N. *et al.* (2010) Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations *Cell* 141, 1241-1252
104. McKee, L. S., La Rosa, S. L., Westereng, B., Eijsink, V. G., Pope, P. B., and Larsbrink, J. (2021) Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature *Environmental Microbiology Reports* 13, 559-581
105. Rogers, T. E., Pudlo, N. A., Koropatkin, N. M., Bell, J. S. K., Moya Balasch, M., Jasker, K. *et al.* (2013) Dynamic responses of *Bacteroides thetaiotaomicron* during growth on glycan mixtures *Molecular Microbiology* 88, 876-890
106. Birt, D. F., Boylston, T., Hendrich, S., Jane, J.-L., Hollis, J., Li, L. *et al.* (2013) Resistant starch: promise for improving human health *Advances in nutrition* 4, 587-601
107. El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D., and Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota *Nature Reviews Microbiology* 11, 497-504
108. Kmezik, C., Mazurkewich, S., Meents, T., McKee, L. S., Idström, A., Armeni, M. *et al.* (2021) A polysaccharide utilization locus from the gut bacterium *Dysgonomonas mossii* encodes functionally distinct carbohydrate esterases *J Biol Chem* 296, 100500
109. Mohnen, D. (2008) Pectin structure and biosynthesis *Current opinion in plant biology* 11, 266-277

110. 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 *Journal of Molecular Biology* 426, 3851-3865
111. Porter, N. T., and Martens, E. C. (2017) The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology *Annual Review of Microbiology* 71, 349-369
112. Lynd, L. R., Weimer, P. J., Van Zyl, W. H., and Pretorius, I. S. (2002) Microbial cellulose utilization: fundamentals and biotechnology *Microbiology and molecular biology reviews* 66, 506-577
113. Tõlgo, M., Hegnar Olav, A., Østby, H., Várnai, A., Vilaplana, F., Eijsink Vincent, G. H. *et al.* (2022) Comparison of six lytic polysaccharide monooxygenases from *Thermothielavioides terrestris* shows that functional variation underlies the multiplicity of LPMO genes in filamentous fungi *Applied and Environmental Microbiology* 88, e00096-00022
114. Moreira, L., and Filho, E. (2008) An overview of mannan structure and mannan-degrading enzyme systems *Applied microbiology and biotechnology* 79, 165-178
115. Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, O. *et al.* (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes *Nature* 506, 498-502
116. 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 *Cellular and Molecular Life Sciences* 76, 4319-4340
117. 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 *Nature communications* 6, 7481
118. Voragen, A. G. J., Coenen, G. J., Verhoef, R. P., and Schols, H. A. (2009) Pectin, a versatile polysaccharide present in plant cell walls *Structural Chemistry* 20, 263-275
119. 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 Nature microbiology* 3, 210-219
120. 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
121. Buléon, A., Colonna, P., Planchot, V., and Ball, S. (1998) Starch granules: structure and biosynthesis *International Journal of Biological Macromolecules* 23, 85-112
122. Vijn, I., and Smeekens, S. (1999) Fructan: more than a reserve carbohydrate? *Plant Physiology* 120, 351-360
123. Roach, P. J., Depaoli-Roach, A. A., Hurley, T. D., and Tagliabracci, V. S. (2012) Glycogen and its metabolism: some new developments and old themes *Biochemical Journal* 441, 763-787
124. Leathers, T. D. (2003) Biotechnological production and applications of pullulan *Applied Microbiology and Biotechnology* 62, 468-473
125. Leemhuis, H., Pijning, T., Dobruchowska, J. M., van Leeuwen, S. S., Kralj, S., Dijkstra, B. W. *et al.* (2013) Glucansucrases: three-dimensional structures, reactions, mechanism, alpha-glucan analysis and their implications in biotechnology and food applications *Journal of Biotechnology* 163, 250-272

126. 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 *Proceedings of the National Academy of Sciences* 114, 7037-7042
127. Ndeh, D., Baslé, A., Strahl, H., Yates, E. A., McClurgg, U. L., Henrissat, B. *et al.* (2020) Metabolism of multiple glycosaminoglycans by *Bacteroides thetaiotaomicron* is orchestrated by a versatile core genetic locus *Nature communications* 11, 646
128. Campo, V. L., Kawano, D. F., da Silva Jr, D. B., and Carvalho, I. (2009) Carrageenans: Biological properties, chemical modifications and structural analysis–A review *Carbohydrate polymers* 77, 167-180
129. Gow, N. A., Latge, J.-P., and Munro, C. A. (2017) The fungal cell wall: structure, biosynthesis, and function *Microbiology spectrum* 5
130. Tailford, L. E., Crost, E. H., Kavanaugh, D., and Juge, N. (2015) Mucin glycan foraging in the human gut microbiome *Frontiers in genetics* 6, 81
131. Luis, A. S., Jin, C., Pereira, G. V., Glowacki, R. W., Gugel, S. R., Singh, S. *et al.* (2021) A single sulfatase is required to access colonic mucin by a gut bacterium *Nature* 598, 332-337
132. Wu, G. (2016) Dietary protein intake and human health *Food & function* 7, 1251-1265
133. Mu, H., and Høy, C.-E. (2004) The digestion of dietary triacylglycerols *Progress in lipid research* 43, 105-133
134. Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022) The carbohydrate-active enzyme database: functions and literature *Nucleic Acids Research* 50, D571-D577
135. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013 *Nucleic Acids Research* 42, D490-D495
136. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino-acid sequence similarities *Biochemical Journal* 280, 309-316
137. Henrissat, B., and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases *Biochemical Journal* 316, 695-696
138. Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities *Biochemical Journal* 293, 781-788
139. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics *Nucleic Acids Research* 37, D233-D238
140. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes *Biotechnology for Biofuels* 6, 41
141. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer III, H., and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5) *BMC evolutionary biology* 12, 186
142. Sandgren, M., Wu, M., Karkehabadi, S., Mitchinson, C., Kelemen, B. R., Larenas, E. A. *et al.* (2013) The structure of a bacterial cellobiohydrolase: the catalytic core

of the *Thermobifida fusca* family GH6 cellobiohydrolase Cel6B *Journal of molecular biology* 425, 622-635
143. Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016) Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization *Applied and Environmental Microbiology* 82, 1686-1692
144. Lebreton, A., Garron, M. L., Vuillemin, M., Pilgaard, B., Hornung, B. V. H., Drula, E. *et al.* (2025) Division of the large and multifunctional glycoside hydrolase family 2: high functional specificity and biochemical assays in the uncharacterized subfamilies *Biotechnol Biofuels Bioprod* 18, 68
145. Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T. *et al.* (2018) dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation *Nucleic Acids Research* 46, D516-D521
146. Wolfenden, R., Lu, X., and Young, G. (1998) Spontaneous hydrolysis of glycosides *Journal of the American Chemical Society* 120, 6814-6815
147. Nasseri, S. A., Lazarski, A. C., Lemmer, I. L., Zhang, C. Y., Brencher, E., Chen, H.-M. *et al.* (2024) An alternative broad-specificity pathway for glycan breakdown in bacteria *Nature* 631, 199-206
148. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases *Structure* 3, 853-859
149. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases *Biochemical Journal* 321, 557-559
150. Chakdar, H., Kumar, M., Pandiyan, K., Singh, A., Nanjappan, K., Kashyap, P. L. *et al.* (2016) Bacterial xylanases: biology to biotechnology *3 Biotech* 6, 150
151. Biely, P., Singh, S., and Puchart, V. (2016) Towards enzymatic breakdown of complex plant xylan structures: state of the art *Biotechnology Advances* 34, 1260-1274
152. Collins, T., Gerday, C., and Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases *FEMS Microbiology Reviews* 29, 3-23
153. Beg, Q. K., Kapoor, M., Mahajan, L., and Hoondal, G. S. (2001) Microbial xylanases and their industrial applications: a review *Applied Microbiology and Biotechnology* 56, 326-338
154. Steve Withers, S. W. Glycoside hydrolases. CAZypedia, [online]
155. Hu, J., and Saddler, J. N. (2018) Why does GH10 xylanase have better performance than GH11 xylanase for the deconstruction of pretreated biomass? *Biomass and Bioenergy* 110, 13-16
156. Cordeiro, L. M. C., de Almeida, C. P., and Iacomini, M. (2015) Unusual linear polysaccharides: (1→5)-α-l-Arabinan, (1→3)-(1→4)-α-d-glucan and (1→4)-β-d-xylan from pulp of buriti (*Mauritia flexuosa*), an edible palm fruit from the Amazon region *Food Chemistry* 173, 141-146
157. Anderson, K. L., and Salyers, A. A. (1989) Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes *Journal of Bacteriology* 171, 3192-3198
158. 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 *Journal of Bacteriology* 183, 7198-7205

159. 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
160. Koropatkin, N. M., Martens, E. C., Gordon, J. I., and Smith, T. J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices *Structure* 16, 1105-1115
161. 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, 2074-2076
162. 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 *Proceedings of the National Academy of Sciences of the United States of America* 109, 7298-7303
163. Terrapon, N., Lombard, V., Gilbert, H. J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species *Bioinformatics* 31, 647-655
164. Artzi, L., Bayer, E. A., and Moraïs, S. (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides *Nature Reviews Microbiology* 15, 83-95
165. Sheridan, P. O., Martin, J. C., Lawley, T. D., Browne, H. P., Harris, H. M. B., Bernalier-Donadille, A. *et al.* (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes *Microbial Genomics* 2, e000043
166. Rakoff-Nahoum, S., Foster, K. R., and Comstock, L. E. (2016) The evolution of cooperation within the gut microbiota *Nature* 533, 255-259
167. Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H. J. *et al.* (2018) PULDB: the expanded database of Polysaccharide Utilization Loci *Nucleic Acids Research* 46, D677-D683
168. Lapébie, P., Lombard, V., Drula, E., Terrapon, N., and Henrissat, B. (2019) Bacteroidetes use thousands of enzyme combinations to break down glycans *Nature Communications* 10, 2043
169. Foley, M. H., Cockburn, D. W., and Koropatkin, N. M. (2016) The *Sus* operon: a model system for starch uptake by the human gut Bacteroidetes *Cellular and Molecular Life Sciences* 73, 2603-2617
170. Larsbrink, J., Zhu, Y., Kharade, S. S., Kwiatkowski, K. J., Eijsink, V. G., Koropatkin, N. M. *et al.* (2016) A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin *Biotechnol Biofuels* 9, 260
171. Hao, Z., Wang, X., Yang, H., Tu, T., Zhang, J., Luo, H., Huang, H., & Su, X. (2021) PUL-mediated plant cell wall polysaccharide utilization in the gut Bacteroidetes *International Journal of Molecular Sciences* 22, 3077
172. Briliute, J., Urbanowicz, P. A., Luis, A. S., Basle, 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 *Nature Microbiology* 4, 1571-1581
173. Reddy, S. K., Bågenholm, V., Pudlo, N. A., Bouraoui, H., Koropatkin, N. M., Martens, E. C. *et al.* (2016) A β-mannan utilization locus in *Bacteroides ovatus*

involves a GH36 α-galactosidase active on galactomannans *FEBS Lett* 590, 2106-2118
174. Accetto, T., and Avguštin, G. (2015) Polysaccharide utilization locus and CAZYme genome repertoires reveal diverse ecological adaptation of *Prevotella* species *Systematic and Applied Microbiology* 38, 453-461
175. Panwar, D., Briggs, J., Fraser, A. S. C., Stewart, W. A., and Brumer, H. (2025) Transcriptional delineation of polysaccharide utilization loci in the human gut commensal *Segatella copri* DSM18205 and co-culture with exemplar *Bacteroides* species on dietary plant glycans *Applied and Environmental Microbiology* 91, e0175924
176. 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, e01068-01017
177. Louis, P., and Flint, H. J. (2017) Formation of propionate and butyrate by the human colonic microbiota *Environmental Microbiology* 19, 29-41
178. Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K. P. *et al.* (2014) Phylogenetic distribution of three pathways for propionate production within the human gut microbiota *The ISME Journal* 8, 1323-1335
179. Porter, N. T., and Larsbrink, J. (2022) Investigation and Alteration of Organic Acid Synthesis Pathways in the Mammalian Gut Symbiont Bacteroides thetaiotaomicron *Microbiology Spectrum* 10
180. McNulty, N. P., Wu, M., Erickson, A. R., Pan, C., Erickson, B. K., Martens, E. C. *et al.* (2013) Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome *PLoS Biol* 11, e1001637
181. Albright, C. E., Spurgeon, K., Ntamubano, S., and Kozik, A. J. (2025) Development of a basal media for carbon utilization assay and analysis of the starch utilization locus in *Prevotella melaninogenica bioRxiv*
182. Ntamubano, S., Parker, D. J., and Kozik, A. J. (2026) The redefined identity of *Prevotella*: new implications for oral, respiratory, and vaginal health *Journal of Bacteriology* 208, e00133-00126
183. Li, J., Gálvez, E. J. C., Amend, L., Almási, É., Iljazovic, A., Lesker, T. R. *et al.* (2021) A versatile genetic toolbox for *Prevotella copri* enables studying polysaccharide utilization systems *The EMBO Journal* 40, e108287
184. Fehlner-Peach, H., Magnabosco, C., Raghavan, V., Scher, J. U., Tett, A., Cox, L. M. *et al.* (2019) Distinct polysaccharide utilization profiles of human intestinal *Prevotella copri* isolates *Cell Host & Microbe* 26, 680-690
185. Willems, A., and Collins, M. D. (1995) 16S rRNA gene similarities indicate that *Hallella seregens* (Moore and Moore) and *Mitsuokella dentalis* (Haapasalo et al.) are genealogically highly related and are members of the genus *Prevotella*: emended description of the genus *Prevotella* (Shah and Collins) and description of *Prevotella dentalis* comb. nov *International Journal of Systematic Bacteriology* 45, 832-836
186. Lawson, P. A., Moore, E., and Falsen, E. (2008) *Prevotella amnii* sp. nov., isolated from human amniotic fluid *International Journal of Systematic and Evolutionary Microbiology* 58, 89-92
187. Downes, J., Dewhirst, F. E., Tanner, A. C. R., and Wade, W. G. (2013) Description of *Alloprevotella rava* gen. nov., sp. nov., isolated from the human oral cavity, and

reclassification of *Prevotella tannerae* Moore et al. 1994 as *Alloprevotella tannerae* gen. nov., comb. nov *International Journal of Systematic and Evolutionary Microbiology* 63, 1214-1218
188. Downes, J., and Wade, W. G. (2011) *Prevotella fusca* sp. nov. and *Prevotella scopos* sp. nov., isolated from the human oral cavity *International Journal of Systematic and Evolutionary Microbiology* 61, 854-858
189. Morotomi, M., Nagai, F., Sakon, H., and Tanaka, R. (2009) *Paraprevotella clara* gen. nov., sp. nov. and *Paraprevotella xylaniphila* sp. nov., members of the family '*Prevotellaceae*' isolated from human faeces *International Journal of Systematic and Evolutionary Microbiology* 59, 1895-1900
190. Nograšek, B., Accetto, T., Fanedl, L., and Avguštin, G. (2015) Description of a novel pectin-degrading bacterial species *Prevotella pectinovora* sp. nov., based on its phenotypic and genomic traits *Journal of Microbiology* 53, 503-510
191. Ueki, A., Tonouchi, A., Kaku, N., and Ueki, K. (2022) *Prevotella herbatica* sp. nov., a plant polysaccharide-decomposing anaerobic bacterium isolated from a methanogenic reactor *International Journal of Systematic and Evolutionary Microbiology* 72, 005221
192. Garschagen, L. S., Franke, T., and Deppenmeier, U. (2021) An alternative pentose phosphate pathway in human gut bacteria for the degradation of C5 sugars in dietary fibers *FEBS Journal* 288, 1839-1858
193. Franke, T., and Deppenmeier, U. (2018) Physiology and central carbon metabolism of the gut bacterium Prevotella copri *Molecular microbiology* 109, 528-540
194. Keitel, L., Braun, K., and Finger, M. (2023) Carbon dioxide and trace oxygen concentrations impact growth and product formation of the gut bacterium *Phocaeicola vulgatus BMC Microbiology* 23, 391
195. Murciano Martínez, P., Appeldoorn, M. M., Gruppen, H., and Kabel, M. A. (2016) The two *Rasamsonia emersonii* α-glucuronidases, *Re*GH67 and *Re*GH115, show a different mode-of-action towards glucuronoxylan and glucuronoxylo-oligosaccharides *Biotechnology for Biofuels* 9, 105
196. Dodd, D., Mackie, R. I., and Cann, I. K. O. (2011) Xylan degradation, a metabolic property of rumen bacteria *Molecular Microbiology* 79, 292-304
197. Bronnenmeier, K., Meissner, H., Stocker, S., and Staudenbauer, W. L. (1995) α-D-Glucuronidases from the xylanolytic thermophiles *Clostridium stercorarium* and *Thermoanaerobacterium saccharolyticum Microbiology* 141, 2033-2040
198. Wilkens, C., Vuillemin, M., Pilgaard, B., Polikarpov, I., and Morth, J. P. (2022) A GH115 α-glucuronidase structure reveals dimerization-mediated substrate binding and a proton wire potentially important for catalysis *Acta Crystallographica Section D: Structural Biology* 78, 658-668
199. Aalbers, F., Turkenburg, J. P., Davies, G. J., Dijkhuizen, L., and Lammerts van Bueren, A. (2015) Structural and functional characterization of a novel family GH115 4-*O*-methyl-α-glucuronidase with specificity for decorated arabinogalactans *Journal of Molecular Biology* 427, 3935-3946
200. Gálvez, E. J. C., Iljazovic, A., Amend, L., Lesker, T. R., Renault, T., Thiemann, S. *et al.* (2020) Distinct polysaccharide utilization determines interspecies competition between intestinal *Prevotella* spp *Cell Host & Microbe* 28, 838-852

201. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs *Nucleic Acids Research* 25, 3389-3402
202. 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
203. Wang, J., Shoemaker, N. B., Wang, G.-R., and Salyers, A. A. (2000) Characterization of a *Bacteroides* mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene *Journal of Bacteriology* 182, 3559-3571
204. Alexeyev, M. (1999) The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of Gram-negative bacteria *BioTechniques* 26, 824-826, 828
205. Groth, A. C., and Calos, M. P. (2004) Phage integrases: biology and applications *Journal of Molecular Biology* 335, 667-678
206. Grindley, N. D. F., Whiteson, K. L., and Rice, P. A. (2006) Mechanisms of site-specific recombination *Annual Review of Biochemistry* 75, 567-605
207. Williams, K. P. (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies *Nucleic Acids Research* 30, 866-875
208. Tamura, K., Hemsworth, G. R., Dejean, 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 Reports* 21, 417-430
209. Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G. *et al.* (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen *Science* 331, 463-467
210. Naas, A. E., Mackenzie, A. K., Mravec, J., Schückel, J., Willats, W. G. T., Eijsink, V. G. H. *et al.* (2014) Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *mBio* 5, e01401-01414
211. Mackenzie, A. K., Naas, A. E., Kracun, S. K., Schückel, J., Fangel, J. U., Willats, W. G. T. *et al.* (2015) A polysaccharide utilization locus from an uncultured Bacteroidetes phylotype suggests ecological adaptation and substrate versatility *Applied and Environmental Microbiology* 81, 187-195
212. Naas, A. E., MacKenzie, A. K., Dalhus, B., Eijsink, V. G. H., and Pope, P. B. (2015) Structural features of a Bacteroidetes-affiliated cellulase linked with a polysaccharide utilization locus *Scientific Reports* 5, 11666
213. Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Yeoman, C. J. *et al.* (2016) Unraveling the pectinolytic function of *Bacteroides xylanisolvens* using a RNA-seq approach and mutagenesis *BMC Genomics* 17, 147
214. Whitaker, W. R., Shepherd, E. S., and Sonnenburg, J. L. (2017) Tunable expression tools enable single-cell strain distinction in the gut microbiome *Cell* 169, 538-546