BOVINE RUMEN AND FAECAL MICROBIOME ADAPTATION TO SEAWEEDS AND EXOTIC FORAGES

JEFFREY P. TINGLEY
Master of Science, University of Lethbridge, 2019

A Thesis submitted
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

BOVINE RUMEN AND FAECAL MICROBIOME ADAPTATION TO SEAWEEDS AND EXOTIC FORAGES

JEFFREY PIERS TINGLEY

Date of Defense: August 18th 2025

Dr. D. Wade Abbott	Research Scientist, Adjunct Professor	Ph.D.
Dr. R. Trushar Patel	Associate Professor	Ph.D.
Thesis Co-Supervisors		
Dr. W. Trevor Alexander	Research Scientist, Adjunct Professor	Ph.D.
Thesis Examination Committee Member		
Dr. Athanasios Zovoilis	Associate Professor	Ph.D.
Thesis Examination Committee Member		
Dr. Laura Parfrey	Associate Professor	Ph.D.
External Examiner		
University of British Columbia		
Vancouver, British Columbia		
Dr. Dylan Girodat	Assistant Professor	Ph.D.
Chair, Thesis Examination Committee		

Dedication

This dissertation is dedicated to my family and friends. I know my father is finally breathing a sigh of relief.

Hopefully now they will stop teasing me for still being in school.

Abstract

Seaweeds are an underdeveloped but valuable agriculture resource. Valorizing their value as CO_2 sinks, sustainable biomasses, and their promise as inhibitors of enteric methane emissions within ruminant livestock remains an important line of research. However, the dietary fate of seaweeds and other marine feedstocks within the ruminant digestive system is unknown. In this thesis, I explore the adaptation of ruminant microbiomes to three distinct seaweeds. The first study explores the adaptation of artificial ruminant systems to the brown seaweed *Saccharina latissima*; the second study explores the faecal and rumen microbiota of cattle supplemented red seaweed *Mazzaella japonica*; in the final study, I investigated the faecal microbiota of a rewilded herd of cattle on Meares Island to identify the adaptation towards unique shoreline foraging and the consumption of green seaweed. In all three studies, shotgun metagenomic datasets and functional prediction of seaweed saccharification pathways revealed the adaptation of the cattle microbiota to their respective seaweed supplementation. Further, biochemical characterization of carbohydrate-active enzymes from metagenomic sequences and isolated microorganisms confirmed that all bovine microbiomes had the capabilities to saccharify their respective seaweed cell wall polysaccharides. Local and international collaborations further corroborated these studies using metaproteomic, glycomic, and enzyme crystallographic analysis. Together, these studies represent the first fine-scale analysis of the adaptation of the ruminant gastrointestinal tract microbiome to seaweed polysaccharides. The adaptability of the ruminant gastrointestinal tract microbiota to distinct seaweed polysaccharides suggests that agriculture solutions can be tailored to locally grown seaweed supplementation.

Acknowledgements

These projects were funded by numerous grants including Agriculture and Agri-Food Canada Science and Technology Branch (Project no: J-002262 & J-003413), Indigenous Agriculture Science Partnerships (IASPP) proposals, and the Prize for Outstanding Achievement in Science (POAS) award (Project no: J-003135).

Firstly, thank you to all of the incredible lab members I have had throughout my degree. I have been here longer than I care to admit, but it has allowed me to form a second family in the lab. I wouldn't be near as successful if it wasn't for the insight and aid provided by that family. To current members Leeann Klassen, Dr. Kristin Low, "Bobby" Barinder Bajwa, Dr. "Mike" Xiaohui Xing, Anna Shearer, Naomi Hanna, Nadia Hand, and Drs. Clair and Cedric Dumon (yes you count as lab members), thank you all for indulging my coffee and baked goods addictions. It will be tough, but I know you will all find a way to move on without me. To the past members: Marrissa McCormack, Nic Jujihara, Dr. Jolene Garber, Dr. Greta Reintjes, Stephanie Monteith, Marshall Smith. It was an honor working with you all and growing together. I would like to 1) apologize for the all of the football talk, and 2) suggest someone else take up football as a hobby so Bobby doesn't go crazy. Leeann you're the last of the O.G. crew...good luck.

To all of the great collaborators throughout the years, whether it be small contributions involving SACCHARIS work, or larger projects apart of this thesis, it was an honor to work with you all. Thank you to Dr. Phil Pope, Dr. Live Hagen, Dr. Thea Anderson, and Dr. Alessandra Ferrillo for the early morning/late night Canada-Norway-Australia meetings. Thank you to Dr. Alisdair Boraston, Liam Mihalynuk, Brendon Medley, and the rest of the Boraston lab for dealing with me as I fumbled through crystallography in your lab. Thank you to everyone at AAFC for all the fun events and treats provided. I may not have left the lab often, but coffee and treat breaks were always appreciated.

Thank you to my committee members, Drs. Athanasios Zovoilis and Trevor Alexander. You both always provided valuable insight and support on my projects and academia. Dr. Alexander thank you for your valuable insights on cattle production, microbiome, and the technicalities and viability of cattle systems – who would have

thought it would cost money to transport seaweed across Canada. To Dr. Zovoilis, thank you for your bioinformatic insights, from the Bioinformatics 1 course, to encouraging me to look into nanopore. To my co-supervisor Dr. Trushar Patel, thank you for being as supportive as you – I didn't deserve it. Every time I have reached out for insight on my projects and writing you have promptly and thoroughly given feedback. Your supportive nature has been a comfort when navigating my research, or something even tougher, like the politics and minutia of academia.

To Dr. Wade Abbott, we have come a long way since I first tagged along with my chemistry friend to his master's interview. It's funny to think about how if I never went to that interview, I likely never would have found my passion for research. You never blinked an eye when I came up with dumb ideas and let me indulge my scientific curiosity by letting me take on projects and travel to learn new methods. I don't know what will be more memorable, all the tense interactions like finishing my MSc thesis on deadlines or the first Meares Island collection, or the countless good times like actually finishing papers and conference beers. I have learned countless lessons from your guidance on research, academia, life and everything in between. Thank you and your family for always providing a place to party when things got tough and the lab needed to come together. The good news is, I know from the relationships that you build that it isn't the end, and there are still more lessons, beers, and laughs to go. See you at the next conference.

To my family and friends thank you for your unending support. Whether it be game nights with the Ryans and Mark, D&D sessions, or the endless support from my family through stressful times, I could not have done this without you. I often forget how privileged I am that I get to mess around in a lab all day, and I could not have done that without the support of my family. Finally, to my late father, how you put up with all the times I bent the truth about when I will finish my degree I will never know. Thank you for your continuing support. on to the next chapter.

Table of Contents

List of Figures

List of Tables

List of Abbreviation

#-Glc	Linage position
AA	Auxiliary activities
AIR	Alcohol insoluble residue
AMR	Antimicrobial resistance
ASV	Amplicon sequence variance
AUC	Alginate utilization cluster
AUL	Alginate utilization loci
BNR	Bacterial Neuraminidase Repeat
*Bo*12C04$_{HOM}$	*Bacteroides ovatus* CL02T12C04
*Bt*3731	*Bacteroides thetaiotaomicron* 3731
*Bx*Car$_{MOS}$	*Bacteroides xylanisolvens* strain from musk deer
*Bz*Car$_{GIR}$	*Bacteroides zhangwenhongii* strain from giraffe
CarPULs	Carrageenan-active polysaccharide utilization loci
CAZyme	Carbohydrate-active enzyme
CBM	Carbohydrate-binding module
CE	Carbohydrate esterase
CH_4	Methane
DAPI	4’6-diamidino-2-phenylindole
DCM	Dichloromethane
DM	Dry matter
ESI-MS	Electrospray ionization mass spectrometry
EtOH	Ethanol
FA	Formaldehyde
FDR	False discovery rate
FLA-AP	Fluorescently labelled amylopectin
FLA-*Mj*Ex	Fluorescently labelled *Mj*Ex

FLA-SLAT	Fluorescent *S. latissima* extract
Fuc*p*	Fucose pyranose
Gal*p*	Galactose pyranose
GH	Glycoside hydrolase
GH16_17	Glycoside hydrolase family 16 subfamily 17
GIT	Gastrointestinal tract
Glc*p*	Glucose pyranose
GT	Glycosyl transferase
Gul*p*A	Guluronic acid pyranose
HCD	Higher-energy collisional dissociation
HGT	Horizontal gene transfer
HILIC	UPLC BEH Amide
HMM	Hidden markov model
HPAEC-PAD	High-performance anion-exchange chromatography with pulsed Amperometric detection
LC-ESI-MS	Liquid chromatography with electrospray ionization mass spectrometry
LFQ	Label-free quantification
MAG	Metagenomic assembled genome
Man*p*	Mannose pyranose
Man*p*A	Mannuronic acid pyranose
MBR	Match-between-runs
*Mj*Ex	*M. japonica* water-soluble extract
*Mj*SM	*M. japonica* surface-associated microbiome
MM	Minimalized medium
OTU	Operational taxonomic units
PBS	Phosphate saline buffer
*Pf*GH16B	*Pseudoalteromonas fuliginea* PS47 GH16
PL	Polysaccharide lyase

PMAA	Permethylated alditol acetates
PUL	Polysaccharide utilization loci
RDA	Redundancy analysis
RUSITEC	Rumen simulation technique
S#	Sulfatase subfamily
SCFA	Short chain fatty acid
SRA	Sequence read archive
sugar kelp	*Sarccharina latissima*
SUS	Starch utilization system
TEAH	Triethylamine hydrochloride
TFA	Trifluoroacetic acid
t-Glc	Terminal sugar
TLC	Thin layer chromatography
UHPLC	Ultra-high performance liquid chromatography
UlvPUL	Ulvan-active PUL
Xyl*p*	Xylose pyranose

Chapter 1 Literature Review and Introduction

1.1 Preface

This chapter contains sections which were written for submission to the Journal Biotechnology for Biofuels [1]. As first author, I conceptualized, wrote, and revised these sections, which have been further modified for this thesis. These sections were revised by Drs. D. Wade Abbott, Kristin E. Low, and Xiaohui Xing. The citation style of these sections has been updated from the original publication to maintain continuity in this thesis.

1.2 The ruminant gastrointestinal tract

The Ruminantia suborder have adapted to consume a diverse range of vegetation and distinct foraging habits. Harvesting energy from fiber-rich vegetation is facilitated through a four-chamber stomach (foregut) consisting of the rumen, reticulum, omasum, and the abomasum, followed by the traditional mammalian hindgut of the small intestine, the cecum, the colon and the rectum [2]. The rumen and reticulum function as primary sites of carbohydrate fermentation, with larger particles being regurgitated and chewed further (rumination), and smaller particles passing through the reticulum to the omasum, the abomasum (true stomach) and lower intestinal tract [3].

1.2.1 The ruminant GIT microbiome

Throughout the gastrointestinal tract (GIT), carbohydrate and protein fermentation is primarily facilitated by complex communities of anaerobic microorganisms. In mammals, these communities are critical in the production of short chain fatty acids (SCFA), immune development and function [4], and colonization resistance of pathogens [5]. The symbiotic relationship between mammals and the GIT microbiome are essential in providing energy for the host. This is exemplified within ruminants, which contain segregated and distinct microbial communities for the fermentation of feedstocks both pre- and post-stomach digestion. The most well-studied of these communities is housed within the rumen.

Neonatal calves have an undeveloped foregut, which is bypassed the esophageal groove during pre-weaning [6]. The pre-weaning period is thought to be the most malleable the rumen microbiome will be throughout the host's lifetime and represents a promising window for early-life interventions [6]. Interestingly, the neonatal rumen microbiome is

immediately colonized during the birthing process, regardless of the esophageal groove and lack of solid feed [7, 8], providing calves with the means to change dietary habits during weaning. Once matured, the bulk of fermentation occurs within the rumen [9]. The mature rumen microbiome is highly resistant towards community perturbations and shows high resilience in returning to its original community state [10, 11]. This phenomenon has been observed following drastic interventions (up to 95% of the community being replaced with donor rumen content) [11], and only overcome with intensive washing and complete replacement [12]. In individual animals the "core" microbiome of the rumen is thought to consist of roughly 30 bacterial groups which vary in abundance with dietary influences and host factors [13, 14]. Of these, the phyla *Bacillota* (formally *Firmicutes*) and *Bacteroidota* (formally *Bacteroidetes*) are generally the most dominant. Bacterial from the genera *Prevotella*, *Butyrivibro*, *Fibrobacter*, and *Ruminococcus* have been reported as dominant species of the rumen microbiome [14-16], all playing key roles in the digestion of plant cell wall polysaccharides. These communities can be further separated into communities within the lumen (rumen chamber) - either residing on solid particles or those suspended within the liquid fraction of the digesta - and those on the rumen epithelial surface [17, 18].

Solid-associated communities are often thought of as pioneers, breaking down complex cell wall polysaccharides; whereas liquid communities often feed off fermentation products and soluble carbohydrates [17]. The communities can be conceptualized as trophic levels, with the first level being polymer degradation, the second being soluble sugar fermentation, and the third being fermentation of byproducts from the first two trophic levels (i.e. SCFA and gasses) [16, 19].

The rumen epithelium layer hosts a unique microbiome composition, consisting of higher abundances of *Bacillota* which serve to metabolize host products including oxygen and urea [18]. In addition, *Butyrivbrio* members were significantly enriched near the epithelium layer, potentially providing the host with sufficient SCFA butyrate for host cell proliferation [20]. Bacteria are not the sole inhabitants of the rumen, however, as 5 archaeal groups all identified methanogens, are considered core members of the ruminant microbiome [14]. These organisms are critical in utilizing excess hydrogen within the rumen, in turn producing methane gas (CH_4) [21]. CH_4 production in ruminants is driven

by only a small portion (<4% based on total rRNA 16S/18S) of the rumen microbiota classified as methanogenic archaea [22, 23]. Methanogens convert ruminal H_2 and CO_2 into CH_4, resulting in 2-12% of gross energy intake converted to CH_4 and considered as energy loss in meat and dairy production [21]. In large part, the variation in CH_4 production results from dietary influences [21]. Eukaryotes including fungi and protists also reside within the rumen, with the latter representing the second most dominant group with the rumen [24]. Protozoa are not well characterized but are known to have roles in methane production and soluble carbohydrate fermentation [24].

The remainder of the GIT is less well-characterized; however, recent studies have shown that each section has a distinct microbiome [20, 25-28]. These studies confirm that the small intestine contains far less diversity than other GIT regions, and higher abundances of *Pseudomonadota* (formally *Proteobacteria*) and *Bacillota* [20, 26, 27]. The relative abundance of *Pseudomonadota* and *Bacteroidota* differed between studies, which aligned with studies demonstrating that the small intestine had the highest levels of β-diversity between animals and lowest levels of species richness (α-diversity) [25]. These findings are explained through the portfolio effect: high diversity of microorganisms can buffer the effect of microbial perturbation [10, 25] and are further corroborated through studies of human small intestinal communities [29]. The ruminant hindgut is responsible for under 10% of energy intake, but is important for the saccharification of recalcitrant carbohydrates [30, 31], this reliance on hindgut fermentation varies with dietary composition [9, 32]. Higher species richness and lower β-diversity between animals can be observed within the ruminant hindgut [25, 27]. Like the small intestine, the lower GIT is primarily dominated by *Bacillota* [26]. In the rumen, there is a relative evenness among members of *Bacillota* - including *Ruminococcus*, *Clostridum*, and *Butyrivibro* spp. - whereas in the large intestine, *Clostridium* spp. are markedly more abundant [20, 25, 26]. There are conflicting reports between studies which suggest *Ruminococcaceae* family members are either more or less abundant in the rumen than the lower intestine [20, 25, 26], suggesting environmental or host influences can heavily influence the abundance of these species with the ruminant GIT. Likewise, although members of the *Bacteroidota* phyla are found throughout the GIT, the rumen is dominated by *Prevotella* spp., as opposed to *Bacteroides* and *Alistipes* spp. being the most abundant within the large intestine [20, 25, 26].

1.2.2 Impact of diet on methane production

Cattle feed can be divided into forage/roughage and concentrate. Forage is highly fibrous feed with high levels of complex polysaccharides including hemicellulose, cellulose, and pectins; whereas concentrate contains easily digestible, simple starches [33]. Forage:concentrate feed ratios have differential impacts on feed passage rate (*i.e.* the rate of transport of feed materials through the gastrointestinal tract), as forage has a slow passage rate and concentrate has a high passage rate [34]. Increasing concentrate quantity in feed is well documented to reduce CH_4, as concentrates increase the passage rate of feed and H_2 in the rumen, which inhibits H_2 producing pathways in microorganisms in the rumen, resulting in an overall decrease/flux of H_2 and CH_4. Forage based diets reduce feed passage time, allowing methanogens to grow more slowly, which balances H_2 levels, leading to increased CH_4 production. Therefore, inclusion of concentrates in the diet can reduce CH_4 production; however, too much concentrate puts the animal at risk for ruminal acidosis [33]. This delicate balance makes the reduction of CH_4 more challenging.

CH_4 mitigation research has focused on different aspects of beef production, including, but not limited to, breeding genetics, early life establishment of the intestinal microbiome, rumen microbiota adjustment through probiotic microorganisms, and inhibition of methanogenesis. 3-nitrooxypropanol is a competitive inhibitor of methyl-coenzyme M reductase [35], the enzyme that catalyzes the final step of CH_4 formation [36]. 3-nitrooxypropanol has shown to reduce CH_4 emissions up to 80% without any perceived negative effects to ruminants [37]. Alternatively, halogenated analogs of CH_4, including chloroform ($CHCl_3$) bromoform ($CHBr_3$), and bromochloromethane (CH_2BrCl), found in seaweeds are natural compounds that act in a similar manner to 3-nitrooxypropanol to inhibit CH_4 production [38, 39]. These halogenated compounds are most abundant in red seaweeds and have been intensively studied as a ruminant feed additive to reduce the production of CH_4 [40, 41]. Red seaweed *Asparagopsis taxiformis* has been shown to reduce CH_4 up to 98% [42, 43], making it an excellent candidate feed additive for CH_4 reduction. Furthermore, seaweeds contain more polysaccharides and less lignin than terrestrial plants [44], which may allow seaweeds to be digested more easily by the rumen microbiota. With such promise, new questions must now be posed over the commercial

viability of seaweed species as feed additives, including transport, cultivation, and their effects on animal health and meat quality [45].

1.3 Plant cell wall structure

Plant cell walls are composed of a primary cell wall organized into microfibrils, and thin membranes (lamellae) of polysaccharides and protein, and often a secondary cell wall which provide more rigidity to mature plant cells [46]. Plant cell wall structure and polysaccharides can vary drastically in structure between plants [47], which is further confounded by interactions within the plant cell wall including microfibrils generated from hydrogen bonding [47], and crosslinking between polysaccharides [48, 49], aromatic polymers including lignin [50], and proteins [51]. The foundation of these polysaccharide interactions is underpinned in the diversity of monosaccharides and glycosyl linkages between them.

1.3.1 Monosaccharides

Monosaccharides have the chemical formula $C_x(H_2O)_n$, with the simplest carbohydrates being the 3 carbon molecules glyceraldehyde and dihydroxyacetone ($C_3H_6O_3$; simplest aldehyde and ketone respectively) [52]. Apart from these, all monosaccharides have at least a single chiral carbon, of which the furthest from the carbonyl group (i.e. the aldehyde or ketose group) is used to denote D- vs L- configuration [52]. D- and L- versions of the same monosaccharide are enantiomers of one another, whereas diastereomers and epimers form distinct monosaccharide types [53]. For example, D-Glucose is a C3 C4 diastereomer of D-Gulose, but a C2 epimer of D-Mannose. Monosaccharides are in equilibrium between their open and closed ring forms, which can take the shape of 5-member rings (furanose), or 6-member rings (pyranose) [52]. Monosaccharide diversity is not just a result stereochemistry, but the size of the carbohydrate. Arabinose and xylose are 5 carbon aldopentose sugars which generally form a 5- or 6-member ring, respectively. These rings form between the hydroxyl group of the chiral C5 and the carbonyl carbon (anomeric carbon) [52]. The stereochemistry at the anomeric carbon is unique, as in the ring form it becomes another chiral carbon with a hydroxyl group cable of mutarotation (switching between two orientations) [52, 53]. These

configurations are denoted α or β depending on if the anomeric hydroxyl-group is on the same side as the CH_2OH group [52, 53].

Hexose sugars can often be observed with modifications, including deoxy forms or the addition of functional groups including carboxylic acids, amine groups, and acetyl, methyl, and sulfate ester groups (**Figure 1.1**). In uronic acids the aldehyde at the C6 carbon is oxidized to a carboxylic acid functional group, which can be further modified by a methyl ester [54, 55]. Galacturonic acid is the main substituent of pectin homogalacturonan and rhamnogalacturonan II backbones. The negative charge of the functional group is critical for gel formation of homogalacturonan, allowing metal chelation and cross-linking between homogalacturonan molecules [56]. Uronic acids can be found in numerous polysaccharides and ecosystems including terrestrial plants, seaweed polysaccharides, and host glycans. Deoxy sugars are distinct in that their C6 carbon does not contain a hydroxyl group and is instead replaced with a hydrophobic methyl group [55]. Commonly seen deoxy sugars, fucose and rhamnose, appear in milk and blood group oligosaccharides, and terrestrial plant and seaweed polysaccharides.

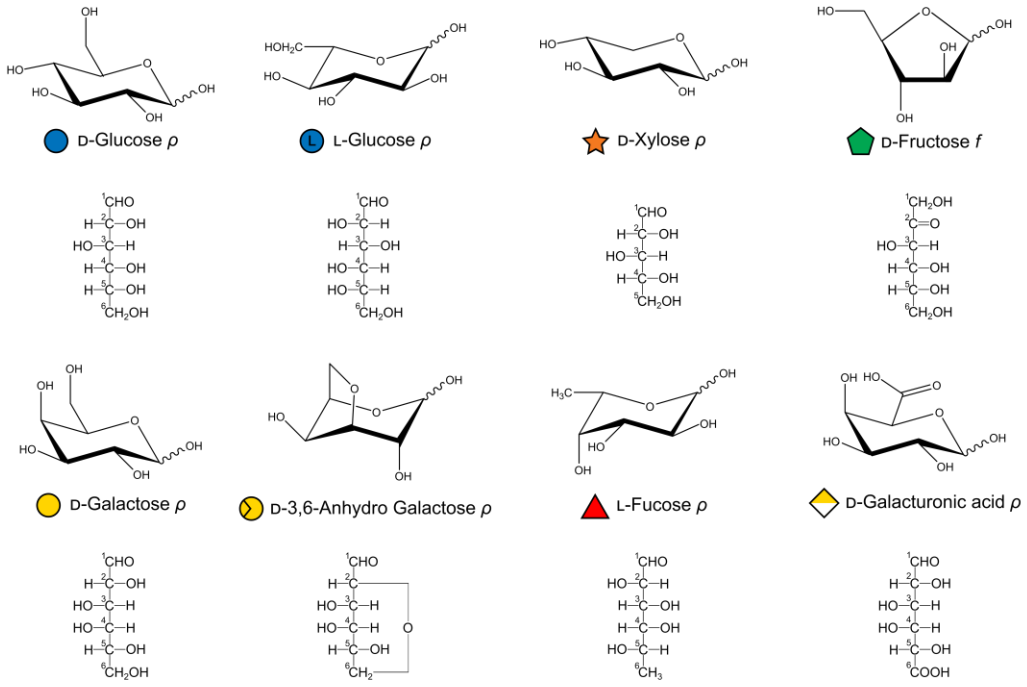

Figure 1.1: Monosaccharides and modifications. Cartoon schematic and Fischer projections of monosaccharides and potential modifications.

1.3.2 Polysaccharides

Monosaccharides form glycosidic linkages to form oligosaccharides, and polysaccharide structures. The plant cell wall is comprised predominantly of cellulosic, hemicellulosic, and pectic polysaccharides, of which cellulose predominates. Cellulose is a linear chain of 4-linked β-D-glucopyranoses existing abundantly in the form of hydrogen-bonded, cable-like microfibrils that contain a heterogeneous mixture of crystalline and amorphous regions that have a diameter ranging from 3 nm to 20 nm depending on cell wall types [57]. Non-cellulosic polysaccharides demonstrate great diversity in monosaccharide composition and linkage. Hemicelluloses are a group of plant polysaccharides consisting mostly of 4-linked neutral sugar backbone, with or without side chains or substituent groups (e.g. methyl group, acetyl group, ferulic acids). This includes mainly xyloglucans, xylan and heteroxylans (e.g. arabinoxylan, 4-*O*-methyl glucuronoxylan, glucuronoarabinoxylan), mannans and heteromannans (e.g. glucomannan, galactomannan, galactoglucomannan, and mixed-linkage glucans in higher plants (**Figure 1.2**) [57, 58]. Pectins are a group of galacturonic acid-rich polysaccharides, including homogalacturonan and rhamnogalacturonans. Homogalacturonan has a 4-linked galacturonic acid backbone that can be 6-*O*-methyl-esterified and 2,3-*O*-acetylated [57]. rhamnogalacturonan-I consists of a backbone of alternating galacturonic acids and rhamnoses and side chains of arabinan, galactan, and arabinogalactans; while rhamnogalacturonan-II is composed of a homogalacturonan backbone decorated with highly complex side chain structures built with more than 20 types of glycosidic linkages from 13 different monosaccharides (**Figure 1.2**) [57].

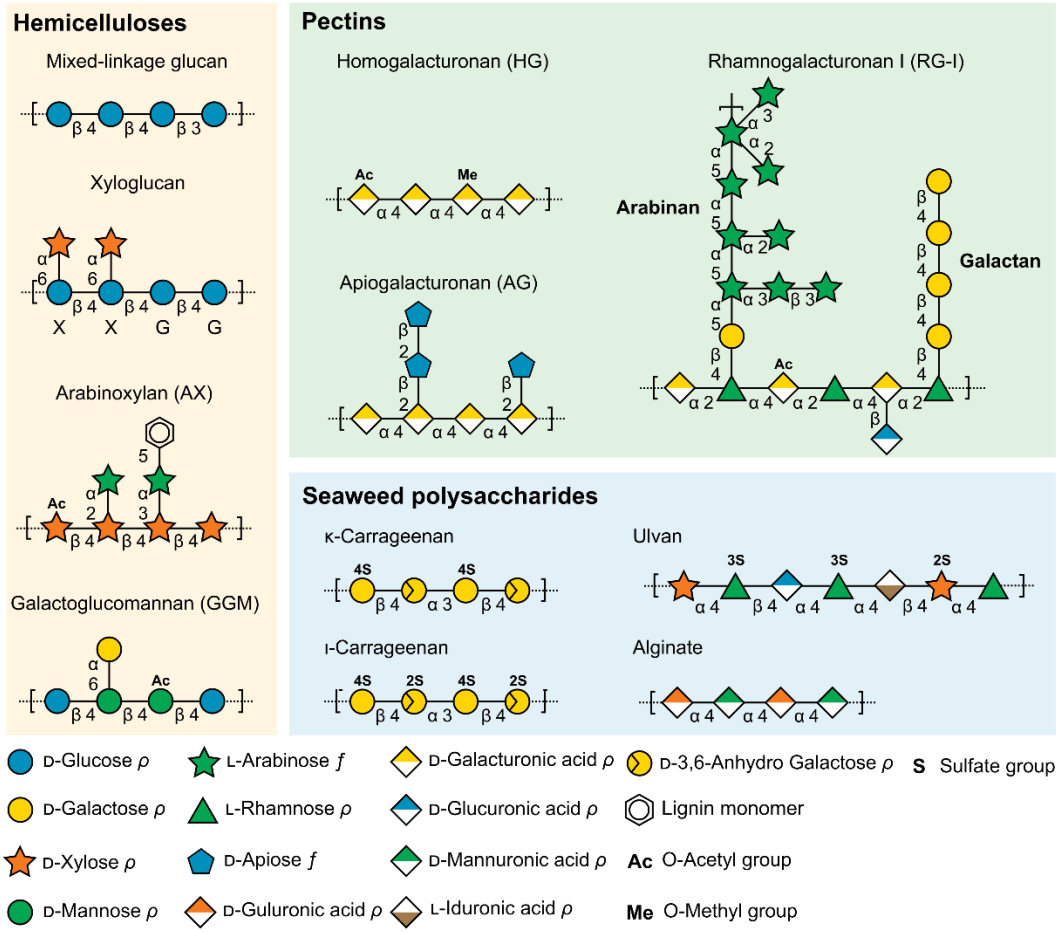

Figure 1.2: Overview of polysaccharide diversity. Cartoon schematics of non-cellulosic plant cell wall polysaccharides. Monosaccharide symbols follow the Symbol Nomenclature for Glycans [55].

1.3.2.1 Diversity of cell wall polysaccharides

Monocot (cereal crops, such as corn, wheat, barley) and dicot plants (legumes, oilseeds, soybeans) have similar cellulose content in primary and secondary cell walls, but differ greatly in the abundance and chemistry of hemicelluloses [47, 48, 59, 60]. Typically, monocots contain much more heteroxylans than dicots in both the primary cell wall (20-40% vs. 5%) and in the secondary cell wall (40-50% vs. 20-30%) [59, 61-65]. Heteroxylans can vary greatly in their substitution patterns, effecting interactions with cellulose and lignin, and in turn, biomass recalcitrance [50, 66]. Dicots generally contain more glucuronoxylan, whereas monocot heteroxylans contain arabinose sidechains (arabinoxylan, glucuronoarabinoxylan) [67]. This difference can be observed between common agricultural crops, including the dicotyledon canola [68] and cereal monocotyledons [69]. Mixed-linkage glucans are absent in most dicots, but represent 10-30% of total cell wall content in monocots [59, 65]. This contrasts with xyloglucan (20-35% vs. 5%) and pectins (20-25% vs. 1-5%) which are more prevalent in primary cell walls of dicots rather than monocots [59, 62, 65]. Although large differences in hemicellulose content and composition exist between monocot and dicot plants, variation can also be seen within a single group. For example, monocotyledon heteroxylans can vary in concentration, presence of glucuronoarabinoxylan or arabinoxylan, and xylan substitution level or arabinose: xylose ratio [69-72]. Furthermore, variation can be seen at the species level as xyloglucan sidechains were shown to differ between canola species *Brassica napus* and *Brassica campestris* [68], between plant anatomy (*e.g.* root vs. root hairs; sugarcane bagasse vs. straw) [73, 74], and between developmental stages in rice [75].

Seagrasses have adapted to marine environments from early diverging land plants and contain similar cell wall polysaccharides, albeit the fine structure of these polysaccharides is largely unknown [76]. Seagrass pectin is unique in its apiose substituent homogalacturonan backbone; this apiogalacturonan is also referred to as zosterin for its discovery in *Zostera* spp. [76]. It is unclear if apiose is an adaptation to saline environments, although it should be noted that boron is responsible for cross-linking amongst residues in terrestrial plant cell wall rhamnogalacturonan II structures [56]. However, it is suggested that other seagrass modifications, including sulfated galactans

[76, 77] and increased arabinogalactan proteins [51] are likely adaptations to saline environments.

1.4 Seaweed cell wall structure

Seaweed (or macroalgae) is a colloquial naming convention for a polyphyletic group which does not share a common ancestor. Therefore, it is unsurprising that the cell wall between seaweed types can vary greatly between not just terrestrial plants, but also traditional seaweed types: green, red, and brown. Green seaweeds are subdivided into 3 distinct clades: the *Prasinodermophyta*, *Chlorophyta*, and the *Charophyta* - the last of which includes the closest relatives to terrestrial plants [78, 79]. Many green, red, and brown seaweeds contain well-structured cell walls composed of cellulose microfibrils, proteins, and matrix polysaccharides [80-82]. Lignin or lignin-like molecules have been found to some extent in all seaweed types, although their role in certain species is unclear [83-85]. Marine seaweeds often contain sulfated polysaccharides which is thought to improve stability in the highly saline marine environment [86]. These polysaccharides, including alginate and fucoidan (brown seaweed), agarans and carrageenans (red seaweeds), and ulvans (green seaweed) have often been used as gelling and thickening agents, as well as therapeutics [87].

1.4.1 Brown algae

Brown seaweeds (*Phaeophyceae*) are known for complex fucose-containing sulfated polysaccharides and alginate cell wall matrix polysaccharides, and laminarin storage polysaccharides. Laminarins are composed of 3-linked β-D-glucopyranoses, with branches of 6-linked β-D-glucopyranoses and are more or less conserved across species of brown seaweed [88]. Fucose-containing polysaccharides are the most structurally diverse polysaccharides within brown seaweeds, which initially classified as fucoidans, are now broken up into fucoidans and sulfated fucans. Fucans contain an α-L-fucose linked backbone, either 3-linked, 4-linked, or alternating between 3-linked and 4-linked [89, 90], and are often heavily sulfated and substituted with acetyl or monosaccharide groups [91]. Fucoidans are a broad class of fucose-rich hemicellulose-like polysaccharides including sulfated xylofucoglucans, xylofucoglucuronans [92], and a range of poorly characterized

polysaccharides containing large amounts of mannose, galactose, and uronic acid linkages [91]. Alginates are composed of 4-linked D-mannuronic acid and its C5 epimer, L-guluronic acid (**Figure 1.2**) [88]. These residues are commonly seen in distinct patterns within alginate, generally denoted M (mannuronic acid) and G (guluronic acid), and appear in different ratios between species. L-Guluronic acid residues are the main substituent in calcium-dependent gelation, which form "egg-box" structures. Therefore, M:G ratios directly influence gel strength [93]. Brown seaweed alginate differs from bacterial alginate in its M:G ratio, and its absence of acetylation [94]. Interestingly It is theorized that alginate biosynthesis in brown seaweed arose from horizontal gene transfer from bacteria [95, 96], therefore it is unclear if the machinery to acetylate alginate was acquired and lost or never acquired.

1.4.2 Red algae

Red seaweed (*Rhodophyta*) is known for its abundance of sulfated galactans in the forms of agarose, porphyrans, and carrageenans. Agarans (agarose and porphyran) are best characterized by their alternating α-3-linked D-galactose and β-4-linked L-galactose units, whereas carrageenan subtypes alternate between α-3-linked D-galactose and β-4-linked D-galactose [88]. Agarans and carrageenans become increasingly complex through sulfation, most notably at the C6 carbon, which can undergo an elimination reaction in basic conditions to form a 3,6-anhydro bridge [88, 97]. Anhydro bridges enable gel formation, increasing the viscosity of sulfated galactans [98]. Sulfate and anhydro bridges delineate agaran and carrageenan subtypes, the simplest being between porphyran (4-linked 6-*O*-sulfo-α-L-galactose agaran) to agarose (4-linked 3,6-anhydro-α-L-galactose). Carrageenan consists of many subtypes depending on anhydro bridge formation and further sulfation patterns, often viewed in disaccharide repeating units (**Figure 1.2**). κ-carrageenan contains a 4-linked β-D-anhydro-galactose and 3-linked 4-*O*-sulfo-α-D-galactose, whereas it's "counterpart" μ-carrageenan, contains a 4-linked 6-*O*-sulfo-β-D-galactose and 3-linked 4-*O*-sulfo-α-D-galactose. Carrageenan subtypes can contain multiple sulfate additions at the 2,4, and 6 carbon positions [99]. Sulfated galactans are the primary polysaccharides in most red seaweed with the exceptions being within orders *Nemaliales* and *Palmariales* which

are composed of mixed-linkage xylans, both neutral and sulfated, and sulfated xylomannans [88, 100].

1.4.3 Green Algae

As previously mentioned, green seaweed is subdivided into 3 groups, of which not all have traditional cell walls. The *Prasinodermophyta* group produces individual scales that are primarily composed of acid sugars including uronic acid and ketodeoxyoctonic acid [83, 101]. In contrast, cell walls of *Charophyta* members, especially those closely related to land plants, share many similarities with terrestrial plant cell walls [83, 102]. Therefore, the variation between green seaweed cells walls and polysaccharide structure is vast. Within *Chlorophyta*, the *Ulvophyceae* class contains numerous marine members with distinctive cell wall polysaccharides including sulfated 4-link mannans, 3-linked xylans [83], 4-linked glucuronic acid (glucuronan) [103], 2(3),4-linked sulfated rhamnose chains (rhamnan sulfate) [104], and ulvans, a complex polysaccharide composed of D-glucuronic acid, L-iduronic acid, L-rhamnose, D-xylose and sulfate additions (**Figure 1.2**) [105]. Ulvan structures vary greatly between species but is often seen in repeating units as either ulvanobiuronic acid: β-4-D-glucuronic acid to α-3-L-rhamnose, or α-4-L-iduronic acid to α-3-L-rhamnose; or ulvanobioses: β-4-D-xylose to α-3-L-rhamnose [105]. In these forms rhamnose can contain one or two sulfate additions at the C2,3 positions, and xylose can be sulfated at the C2 position [106]. It should be noted that ulvan is not constrained to these repeating units, as glucuronic acid sidechains, and repeating α-4-L-rhamnose backbones has been observed [107].

1.5 Carbohydrate active enzymes

Microorganisms within ruminant communities employ a range of both catalytic and non-catalytic machinery to disassemble the plant cell wall matrix. The most notable of these enzymes are classified as Carbohydrate-Active enZymes (CAZymes). CAZymes are classified into classes based upon the catalytic mechanism by which they act, including glycoside hydrolases (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities (AA), glycosyl transferases (GTs), and non-catalytic carbohydrate-

binding modules (CBMs). Each of these classes are further divided into sequence-related families.

1.5.1 CAZyme classification

GHs hydrolyze glycosidic bonds between carbohydrates or a carbohydrate and aglycone moiety, such as lipids or proteins [108, 109]. For most GH-mediated hydrolysis, two residues are critical for this enzymatic mechanism, a proton donor and a nucleophile/base, and results in a mechanism that either retains or inverts the anomeric configuration (**Figure 1.3**) [108, 110]. With such diverse substrate potential existing in nature [111], it is unsurprising that GHs have been found to be active on carbohydrate polymers ranging from homopolymers, such as starch [112] and cellulose [113] to highly branched and chemically heterogenous substrates, such as pectins [114]. At the time of publication, there are over 190 sequence-based families of GHs in the CAZy database [115].

PLs cleave polysaccharide chains with a β-elimination reaction, resulting in a terminal unsaturated hexuronic acid at the reducing (**Figure 1.3A**) [110, 116]. PLs are typically involved in the cleavage of acidic substrates, such as: pectins (*e.g.* homogalacturonan), chondroitin, xanthan, and alginate [117]. At the time of publication, there are over 40 different PL families assigned within the CAZy database [115].

CE families are currently classified into 21 different families [115]. These enzymes catalyze the de-*O*- or de-*N*-acetylation of esterified sugars through a variety of mechanisms, whereby the sugar can either act as the acid (*e.g.* pectin methyl esters) or the alcohol (*e.g.* acetylated xylan) (**Figure 1.3A**) [110]. Removal of carbohydrate esters increases the access of GHs and PLs to their substrates, and therefore, is an important event in the catabolism of chemically complex polysaccharides.

AAs are the most recently described CAZyme class and deploy a redox reaction to fragment structural polysaccharide and lignin substrates (**Figure 1.3A**) [118]. AAs are currently divided into 18 families, encompassing 9 families of ligninolytic enzymes, and 6 families of lytic polysaccharide mono-oxygenases. While first discovered to target chitin [119], lytic polysaccharide mono-oxygenases have shown activity on common plant cell

wall polysaccharides including cellulose. Many AA enzymes are metalloenzymes, requiring copper to catalyze the digestion of lignocellulosic biomass [120, 121].

In contrast to the other CAZyme classes discussed, GTs are responsible for the synthesis of carbohydrates. GTs require an acceptor and phosphate substituted donor sugar [122]. GTs are classified into 4 folds, GT-A, GT-B, GT-C, and the uncommon lysozyme-fold found in GT family 51 [123] GT-A and B folds contain both inverting and retaining families, whereas members with GT-C folds and GT family 51 lysozyme folds have shown to be inverting.

Although non-catalytic, CBMs have large roles in 1) targeting enzymes towards their substrate and 2) concentrating enzymes within a region [124]. The majority of CBMs have a β-sandwich fold and often are directly attached to CAZymes [124]. CBMs were thought to be distinct from lectins based on their association with CAZymes [125] However, this is not without exceptions, including CBM32 [126] and CBM29 [127] members which are not attached to catalytic modules. To add to this confusion, type C CBMs including CBM18 and CBM47 share significant homology towards lectins and are classified as such [127].

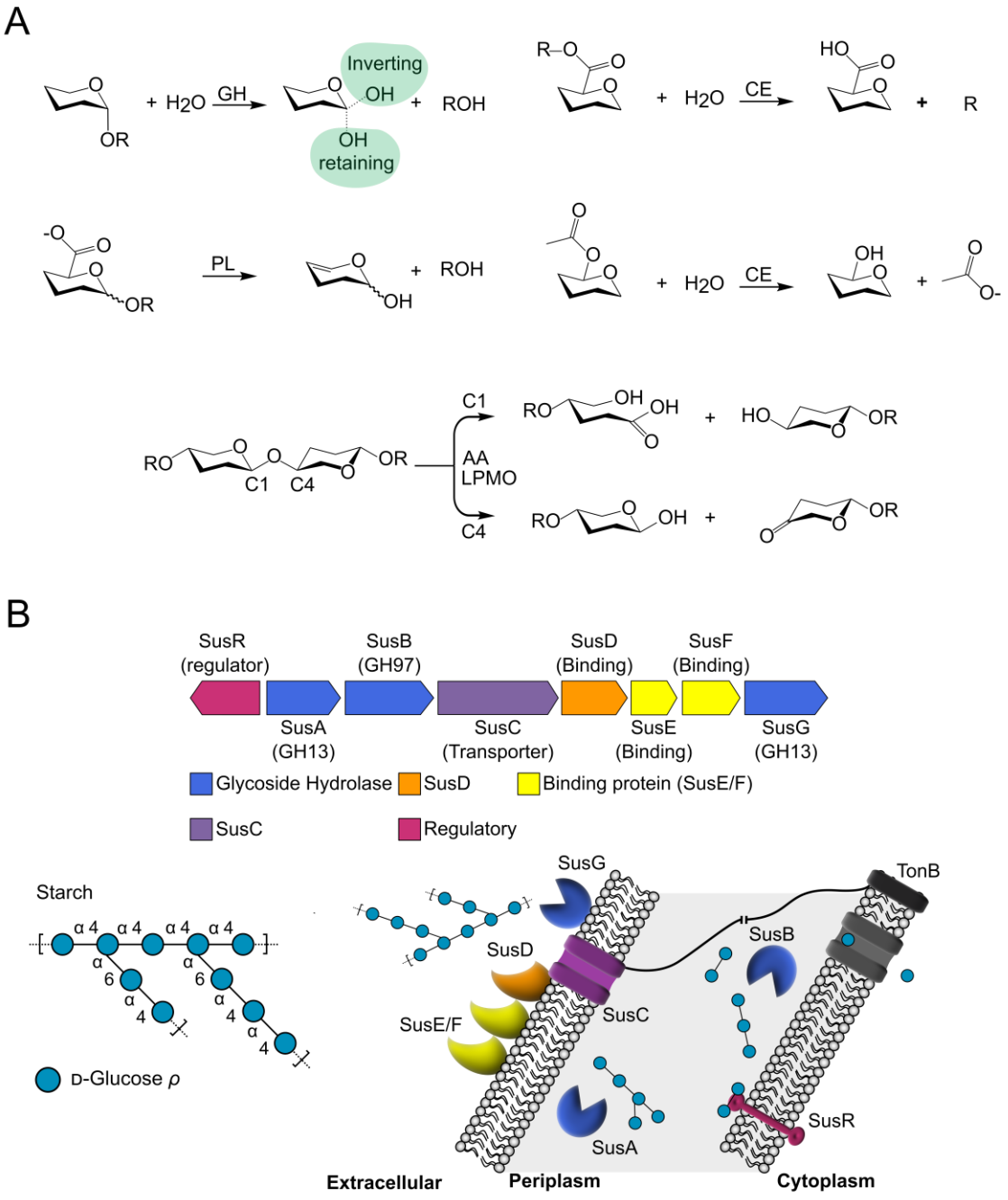

Figure 1.3: CAZyme class mechanisms and PUL organization. A) CAZyme depolymerization mechanisms and specificities. Simplified reaction schematics are shown of a GH – both retaining mechanism and inverting mechanism, PL, and CE, and the AA of lytic polysaccharide mono-oxygenases active on C1 and C4 carbons. **B)** Cartoon schematic of the sus operon genetic loci and mechanism from *B. thetaiotaomicron*.

1.5.1.1 Sulfatases

Although not classified as a CAZyme, sulfatases play an important role in the saccharification of sulfated polysaccharides. Characterized sulfatases have been identified for seaweed polysaccharides: carrageenan, ulvan, and fucoidan [128], and host glycans [129]. Sulfatases can be classified into 4 families (S1-4), of which have unique catalytic mechanisms [130, 131]. Most sulfatases predicted and characterized are assigned to family S1, which has further been divided into over 100 subfamilies. This family is often denoted as the formylglycine-dependent family, as it requires a post-translation modification to generate the formylglycine residue to facilitate an elimination reaction [131]. Similar to CAZymes, S1 sulfatases have been shown to be exo- or endo-active on their respective substrates [128]. Both actives have been discovered within seaweed targeting sulfates, including exo-active and endo-active ulvan members (S1 subfamily 25 and subfamily 7 respectively) [132] and carrageenan members (S1 subfamily 15 and subfamily 19) [133, 134].

1.5.2 Coordinated plant cell wall deconstruction

Primary carbohydrate degraders commonly employ a cascade of CAZymes and other machinery to digest polysaccharide structures. Strategies used to organize and secrete these machineries can vary greatly between species. These strategies include directly secreting CAZymes attached to CBMs into the environment, creating large “scaffolds” of enzymes (aka. Cellulosomes), producing outer membrane vesicles with catalytic machinery attached to the vesicle surface, or employing specialized operons to produce catalytic machinery for distinct polysaccharide structures [135, 136].

CAZymes and other machinery for bacterial saccharification of carbohydrates including transport channels, binding proteins, and other genes required for metabolism can often be seen co-expressed within the genome as operons. The most well-known example being the lac operon from *Escherichia coli* designed for lactose metabolism. Operonic carbohydrate utilization structures have been observed within gram positive *Lachnospiraceae* members [137], plant cell wall pathogens including *Xanthomonas campestris* [138], but most notably appear within *Bacteroidota* members which employs numerous Polysaccharide Utilization Loci (PULs) [137]. The Starch Utilization System

(Sus) was first identified within *Bacteroides thetaiotaomicron*, and as the name suggests, targets starch polysaccharides (**Figure 1.3B**) [139]. Although the Sus system is considered the first PUL identified, this term was not coined until later [140, 141]. PULs are denoted by their TonB-dependent transport proteins (SusC-like), binding proteins (SusD, E, F-like), regulatory genes, and a diverse array of CAZymes and other ancillary enzymes including sulfatases and proteases [136].

The cellulosome structure was first reported within *Clostridium thermocellum* [142] and consists of a network of scaffold and dockerin proteins which coordinate and concentrate cellulosic CAZymes on the plant cell wall surface [143]. Cellulosomes have since been identified in many *Oscillospiraceae* members including *Clostridium* and *Ruminococcus* species within the ruminant, and cellulosome-like structures have always been identified with fungal members [144]. Cellulosomes can vary greatly in size between species and containing multiple scaffolds with either anchor the cellulosome to the bacterial cell well (anchoring scaffold), act as the primary scaffold for substrate deconstruction (primary scaffold), or expand the cellulosome, increasing efficiency (adaptor scaffolds) [143]. Scaffolds and enzymes are held together by linker segments and, most importantly, cohesion and dockerin domains which form strong interactions with dissociation constant $< 10^{-11}$ M [143, 145]. The components of cellulosomes can be transcribed within larger operons, or in individual units [146].

Interestingly, *Ruminococcus albus* produces CAZymes which contain dockerin domains, however, no evidence of scaffolding and cohesion interactions has been observed suggesting these are freely secret CAZymes [143]. *Fibrobacter* species produce outer membrane vesicles which contain hundreds of proteins designed to digest complex plant cell wall [147]. This strategy may also be employed by *Bacteroidota* members, although little is known about the function and implications of microbial interactions between species and the generation of soluble carbohydrates used as “public goods” [148].

1.5.3 CAZyme functional prediction

The CAZy database was launched in 1999, and is the single source for CAZyme curation [115], in addition to providing links to relevant publications and other online resources, such as CAZypedia [149] and the PUL database PULDB [150]. These extensive

resources have led to other external developments to assist with CAZyme discovery and characterization. For example, the CAZyme annotation tool dbCAN [151] provides hidden markov models (HMMs) generated from the CAZy database to facilitate user sequence annotation. dbCAN identifies sequence boundaries to improve prediction accuracy, creating profile-HMMs based on homologous sequence alignments. Alternatively, the CAZyme analysis toolkit [152], currently unmaintained, implements Pfam-defined profile-HMMs which were recently shown to identify >98% of GHs in the CAZy database [153]. These profile-HMMs provide valuable protein domain prediction, especially helpful in determining boundaries in multi-modular CAZymes and/or attached CBM modules [154], and are currently used by an expanding list of pipelines and software tools [154-156]. However, it should be noted that due to differing thresholds between profile-HMMs there may be discrepancies between Pfam and dbCAN annotations when compared to those of CAZy [115].

The addition of subfamily designations to large, polyspecific families in the CAZy database and the subsequent profile-HMMs generated by dbCAN, have greatly improved functional prediction of novel sequences for CAZy families GH5 [157], GH13 [158], GH16 [159], GH30 [160], and GH43 [161]. However, there are still inherent limitations with family- and subfamily-based classifications. While CAZy families are reflective of structural similarities, resulting from sequence relatedness, assignment of a sequence to a CAZy family is not necessarily definitive of function. Functional differences between members of the same subfamily and polyspecific families without subfamily delineations, unfortunately, convolute the prediction of CAZyme activity. As well, sequence based CAZyme prediction is difficult due to the availability, abundance, and quality of biochemically characterized sequences within families. Fortunately, there is a growing list of novel software designed to aid in the annotation (PULpy [162], DRAM [163], dbCAN-PUL [164]), curation (dbCAN-PUL [164]) and high-resolution phylogeny (SACCHARIS v2.0 [154], CUPP [155]) of CAZymes.

Both PULpy and DRAM software packages use profile-HMMs sourced from both dbCAN and Pfam to identify CAZymes. PULpy focuses heavily on identifying PULs within metagenomes, demonstrated in ruminants [165], and DRAM extrapolates CAZyme annotation to predict carbohydrate utilization of identified taxonomic units. Recently,

dbCAN-PUL was developed for the curation of PULs by substrate, taxonomy, and characterization method. The repository can also be downloaded and used as a database to BLASTX against novel CAZymes. Alternatively, SACCHARIS v2.0 is a pipeline that streamlines identification and phylogenetic analysis of CAZyme sequences. Sequences collected from the CAZy database, as well as user inputted sequences, are trimmed to the predicted catalytic domain using dbCAN, aligned [166], and a best-fit newick tree is generated [167-169]. SACCHARIS is a real-time software which enables the functional prediction of CAZymes based upon tree topologies generated using the current state of knowledge [170-172]. The CUPP downloadable software uses peptide pattern recognition to find conserved peptide motifs within CAZyme families to develop strict CUPP groups or subfamilies, and a recent web server allows for annotation of user sequences [173]. CUPP has been used to elucidate sequence function in pectin and alginate lyase families [174, 175], as well as using fungal CAZyme secretomes to predict fungal phylogenies [176]. Together with -omics based technologies, CAZyme prediction tools will aid in the interpretation of sequence datasets at the microbe, community, and gene level. Ultimately, this interpretation is necessary to inform CAZyme discovery and characterization, which will further be used to improve biofuel production.

1.6 Nutritional adaptation

Due to the complexity of plant cell walls, including polysaccharide and cell wall matrix composition, the breakdown of the plant cell wall by ruminant microorganisms is rarely achieved by a single species. Environmental factors, host genetics, and inter- and intra-species interactions within an environment create defined niches for species to inhabit [177]. Distinct microbial communities adhere to fibrous feed particles, forming unique and dynamic microecosystems [135, 178]. These communities are temporally sensitive, and change throughout the digestive process [179] from initial colonization, the increased abundance of primary fibre degraders, to community members with less fibre-consuming abilities [135]. The different periods of digestion paired with the trophic levels observed within the rumen microbiome create complex microbial interactions and niche partitioning.

1.6.1 Microbial foraging strategies

Microbial interactions within the rumen ecosystem are complex, as apparent metabolic redundancy, microbial foraging strategies and priorities, and ecological interactions such as competition and cooperation influence microbial abundance. Microorganisms across phyla adopt generalist or specialist strategies to occupy niches and proliferate [180]. Although sometimes difficult to underpin which organisms are considered generalist and which are specialist, species that are cable of occupying multiple environments or utilizing multiple nutrients are said to have a larger niche breadth [180]. Generalists take on the metabolic cost and trade-offs to accommodate perturbations within their niche as opposed to specialists [181]. Generalist and strategies have evolved within multiple phyla independently for example, in the *Bacteroidales* order, *Bacteroides* spp., notably *B. thetaiotaomicron*, contain numerous PULs which allow them to metabolize multiple complex carbohydrates, whereas *Parabacteroides* spp. including *P. distasonis* specializes mostly on host glycan utilization [182].

Metabolic redundancy and competition for resources appear in multiple cases, most notably cellulose [135, 183]; However, despite this redundancy there exists numerous community members which successfully compete for nutrients. Although it is not always clear when there is direct competition, an observed metabolic overlap between species may not be the case in reality. Priority foraging within generalist microorganisms such as *Bacteroides* spp. has been reported [184], enabling numerous generalist organisms to reside within similar niches, leading to alternative stable states [185]. Even within cellulose specialists, *Fibrobacter suuccinogenes* and *Ruminococcus flavefaciens*, these species have been found to colonize different parts of hay (stem vs. leaf) [186]. It is unclear if these interactions are due to species exclusion [135], or substrate adherence-based niche [181]. Therefore, nutritional niches within both specialist and generalist foraging organisms are not always obvious. The absence or presence of primary degraders within the community also trickles down to the lower trophic levels, as competitive exclusion of one species towards another may lead complex networks of microbial communities in rock-paper-scissors like interaction [187]. Primary degraders can either degrade substrates on the outer surface of their cell, enabling cross-feeding by producing small mono and oligosaccharides (i.e. “public goods”), or saccharify polysaccharides through a selfish mechanism [136,

188]. The former allows for cooperation between *Bacteroides* spp. enabling cross-feeding between members which form complex networks of recipient and producer species [185, 189, 190]. These networks expand through the trophic levels, where secondary and tertiary producers digest the soluble carbohydrates and fermentation products of the primary producers [19, 191].

1.6.2 Acquisition of novel functions

The selective pressure to outcompete other community members within a niche or expand into new niches can often lead to the introduction of novel functions or speciation [192]. Splitting into novel species (cladogenesis) is more commonly seen in generalist species, which were predicted to have 19-fold higher speciation rates compared to specialist lineages [192]. Through gene duplications *Bacteroides* spp. have accumulated multiple copies of tandem SusC/D-like pairs which have undergone neofunctionalization to accommodate a variety of carbohydrates [182]. Due to PUL complexity within *Bacteroides* spp., it is often difficult to determine whether SusC/D-like pairs were a result of speciation or duplication, i.e. orthologs or paralogs within strains. Further, it is common for paralogs to cause incongruent gene and species trees, clouding the interpretation of speciation, duplication, and loss events [193]. Neofunctionalization of paralog CAZymes has not been well-studied apart from it being assumed within species. However, neofunctionalization within the pectin active PL2 family is believed to have created distinct endo- and exo- active subfamilies within plant pathogens [194]. In contrast to gene duplication and neofunctionalization, methanogens within the rumen are thought to have "streamlined" their genomes through mutation and gene loss, evolving from a free-living microorganism into the specialized oligotroph [195].

1.6.2.1 Horizontal gene transfer

Gene homology and ancestry within species can be further clouded through the presence of xenologs and the process of horizontal gene transfer (HGT). HGT is the exchange of genetic elements between organisms, whether between similar species or different phylogenetic domains [196]. HGT can occur through conjugation, transformation, and transduction, and recent studies suggest that transfer can be facilitated by outer

membrane vesicles [197]. Although it is not always clear what mediates the transfer of genetic material, environmental pressures including nutrient availability can impact the transmissibility [198]. HGT events have been well-studied for the transfer of antibiotic resistant genes, including work conducted by Abigail Salyers uncovering evidence of extensive transfer of erythromycin resistance genes within *Bacteroide*s spp. [199]. Moreover, recent work with *Bacteroides* spp. has suggested HGT can directly affect polysaccharide metabolism as well. The yeast mannan PUL-1 in *B. thetaiotaomicron* is found in multiple *Bacteroides xylanisolvens* strains in human [188] and rumen *B. thetaiotaomicron strains* [200]. Often the extent of transfers within a population has been correlated with diet, as red seaweed degrading CAZymes predicted to be transfer from marine bacteria are found to be more prevalent in East Asian cultures [201]. This was expanded on when a Japanese gut microbe, *Bacteroides plebeius,* was shown to degrade porphyran with the aid of CAZymes which have high homology to those in the marine *Bacteroidota*, *Zobellia galactanivorans*, providing evidence for HGT [202]. Since then, evidence for the transfer of CAZymes involved in degrading seaweed polysaccharides including alginate [203], agarose [204], and carrageenan [205] from marine microorganisms to human gut bacteria has strengthened. In ruminants, HGT of CAZymes from bacteria to fungi [206], and ciliates [207] has been reported, but little is known about the transfer of CAZymes dedicated to the saccharification of seaweed from marine bacteria donors to rumen/intestinal recipient bacteria. Isolated bacteria from sheep on North Ronaldsay Island, Scotland, which notably forage almost entirely on marine plant species, were found to contain CAZymes dedicated to saccharify marine plant cell wall polysaccharides [208]. However, the origin of these CAZymes is still in question, whether they evolved from sequence mutation or were acquired from marine microorganisms.

1.7 Project proposal

This thesis explores the adaptation and evolution of ruminant microorganisms towards the saccharification of seaweed cell wall polysaccharides. Three distinct populations of cattle or artificial rumen systems which, either through human intervention or native foraging patterns, have incorporated marine-based forages into their diet have been examined:

1. The first study (Chapter 2) focuses on the adaptation of ruminant microbiomes to brown seaweed *Saccharina latissima* using the rumen simulation technique (RUSITEC) and sheep trials. This study was a collaborative project between Agriculture and Agri-Food Canada and the Norwegian University of Life Sciences.

2. The second study (Chapter 3) was conducted in collaboration with Edgar Smith, the owner and operator of Beaver Meadows farm in Comox, B.C. - a natural beef pasture. Beaver Meadows collects and supplements the diet of cattle with local red seaweed *Mazzaella japonica*. Cattle faecal samples are then collected and composted, which is used to fertilize the pastures, creating a unique ecosystem fed by red seaweed.

3. The third study (Chapter 4) was a co-developed project with the Tla-o-qui-aht [*klaw-oh-kwee-awt*] First Nation on Meares Island, BC. Meares Island hosts a free-roaming herd of rewilded cattle that grazes the local shoreline, forest and grasses in the Opitsaht village. This herd has naturally adapted over the last ~125 years to readily available forages in the surrounding environment, including forest vegetation, and costal eelgrasses (*Zostera spp.*) and green seaweed *(Ulva intestinalis*).

RUSITEC, faecal and rumen samples from all studies (when available) underwent multi-omic analysis, bacterial enrichment and isolations, and characterization of putative seaweed polysaccharide saccharifying bacteria and CAZymes. The adaptation of ruminant microorganisms and the origin of CAZymes targeting exotic polysaccharides is explored. The thesis ends with a summary of results and conclusions, with a brief perspective on the impact of my work on the field and compelling avenues for future research (Chapter 5). Together, these three studies provide conclusive evidence towards the adaptation of the ruminant microbiome to exotic seaweed polysaccharides.

Chapter 2 Microbial alginate foraging is conserved in geographically and taxonomically distinct ruminant microbiome

2.1 Preface

Chapter 2 is presented in manuscript form that is accepted with the journal Nature Communications (Tingley & Ferrillo et al., 2025; [209]). This collaborative study examines the adaptation of simulated cattle rumen microbiomes and sheep rumen microbiomes to the brown seaweed *S. latissima*. As co-first author, I participated in the writing and structuring of the manuscript, revision and editing process. My contributions to the manuscript include analysis of metagenomic datasets, phylogenetic analysis of species and gene trees, and characterization of alginate-active CAZymes. Alessandra Ferrillo conducted sheep trials and sheep metagenomics and metaproteomic analysis. The full contribution list is presented below. My contributions are displayed in figures: 2.3B, 2.4, 2.5, 2.6; Appendix 1 figures A1.5-11, A1.3-14 and Appendix tables A1.4-5 and Extended Tables 1.1. The citation style of this manuscript has been updated from the original publication to maintain continuity in this thesis.

Authors:

Jeffrey P. Tingley[1,2]*, Alessandra Ferrillo[3]*, Marissa L. King[1], Alemayehu Kidane[4], Barinder Bajwa[1], Xiaohui Xing[1], Tina Johannessen[1], Alexsander Lysberg[1], Liv Torunn Mydland[4], Margareth Øverland[4], Greta Reintjes[5], Anna Y. Shearer[1,2], Leeann Klassen[1], Kristin E. Low[1], Trushar R. Patel[2,6,7,8], Stephanie A. Terry[1], Phillip B. Pope[3,9], D. Wade Abbott[1,2†], and Live H. Hagen[3†]

Affiliations:

[1] Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, 5403-1st Avenue South, Lethbridge, Alberta, Canada, T1J 4B1.

[2] Department of Biochemistry, University of Lethbridge, Lethbridge, AB, Canada, T1K 6T5.

[3] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1432 Ås, Norway.

[4] Faculty of Biosciences, Norwegian University of Life Sciences, 1432 Ås, Norway.

[5] Microbial-Carbohydrate Interactions Group, Faculty of Biology /Chemistry, University of Bremen, Bremen, Germany.

[6] Alberta RNA Research and Training Institute and Department of Chemistry and Biochemistry, Lethbridge Alberta, Canada
[7] Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary Alberta, Canada
[8] Li Ka Shing Institute of Virology and Discovery Lab, University of Alberta, Edmonton Alberta, Canada
[9] Centre for Microbiome Research, School of Biomedical Sciences, Queensland University of Technology (QUT), Translational Research Institute, Woolloongabba, QLD, Australia

* These authors contributed equally
† To whom correspondence should be addressed

Author Contributions:

A.F., J.P.T., D.W.A. and L.H.H. conceived the study, performed the primary analysis of the data, and generated figures. D.W.A., T.R.P., and L.H.H. supervised the work. S.A.T. and D.W.A. designed and performed the RUSITEC analysis. P.B.P. and L.H.H. proposed and performed the lamb omics analysis. L.T.M., M.Ø., and A.K. designed, performed and/or analyzed the lamb feeding trial. T.J. and A.L. extracted lamb rumen DNA. J.P.T., and M.L.K. contributed to the RUSITEC data analysis. B.B. and X.X. contributed to AIR preparation and linkage analysis for *S. latissima*. G.R., L.K., and M.L.K. generated and analyzed FLAPS results. J.P.T. and A.Y.S. characterized alginate lyases, K.E.L. generated and analyzed LC-MS. A.F., J.P.T., D.W.A., and L.H.H. wrote the paper, with valuable inputs and edits from P.B.P. and G.R. All authors reviewed and approved the final manuscript.

2.2 Abstract

Seaweed plays a crucial role in carbon cycling and is expected to be a valuable resource for sustainable biomass, with applications in biofuel production, human nutrition, and animal feed. Although seaweed has historically been used as a feed source for livestock grazing near coastlines, the process by which it is digested in the rumen remains unknown. Here, we show how the brown alga *S. latissima* is catabolized in the rumen ecosystem of two different species using *in vivo* and *in vitro* experimental systems. Evidence of digestion was obtained using a combination of animal models, bacterial imaging, multilayered meta-omics, and enzyme biochemistry. We demonstrate that alginate PULs found within geographically distinct ruminants share diverse alginate utilization systems. Core alginate enzymes have been maintained throughout populations, with ancillary enzymes predicted to be gained or lost through gene duplication or loss events, respectively. The conservation of alginate utilization systems within diverse ruminant populations suggests microbial populations maintain catalytic machinery for seaweed polysaccharide metabolism.

2.3 Introduction

The effects of climate change, such as extreme drought and flooding, are deleterious for food production and reduce access to cultivatable land. Thus, seeking solutions from the ocean to address the challenges associated with the harvest of terrestrial food and feed is garnering interest. Cultivated and wild-harvested macroalgae are emerging as a promising source of biomass for future application in biofuel production, human nutrition, and animal feed [210]. Certain varieties of marine macroalgae, such as kelp, grow faster than terrestrial feed crops [211, 212], and create highly productive coastal forests that sequester substantial amounts of carbon globally while improving habitat for other marine life [213]. Importantly, macroalgae cultivation does not rely on freshwater, nor compete with arable land or land areas dedicated to conservation, which makes it an attractive source of sustainable biomass. Algal biomass is also rich in micronutrients and proteins, thus has potential to supplement livestock diets [214].

The brown macroalga *S. latissima* (‘sugar kelp’) belongs to the *Laminariaceae* family and is widely distributed across temperate coastal waters, including the North Atlantic and North Pacific. Like most brown macroalgae, the cell wall of *S. latissima* is

largely comprised of structural polysaccharides such as alginate, cellulose, and fucoidans, which together account for 30-50% of the algal total dry weight [215]. Depolymerization of alginate is catalyzed by a highly specialized group of PL enzymes, called alginate lyases, that cleave glycosidic bonds through β-elimination, resulting in unsaturated products [116]. A number of alginate lyases have been studied from various organisms, primarily originating from marine environments, including marine algae [216], fungi [217], viruses [218], marine mollusks [219, 220] and human gut bacteria [221]. Intriguingly, previous studies have highlighted possible transmission of seaweed-degrading enzymatic machinery from marine bacteria to the human gut microbiome [205, 222, 223]. This includes alginate lyases [221] and PULs for red-algal galactan degradation in human gut bacteria [204, 222]. In most cases, these events have been observed to occur within *Bacteroidota*, a phylum known to harbor complex PULs or systems dedicated to saccharifying distinct polysaccharides. Exchange of PULs within *Bacteroidota*, promotes the acquisition of novel genetic machinery and the ability to subsist on available polysaccharide substrates [224].

In recent years, the practice of adding seaweed to livestock feed has gained momentum due to the potential of certain species to reduce enteric methane emissions [225, 226]. However, feeding seaweed to cattle is not a new strategy; farmers in coastal areas have historically collected kelp to supplement the livestock feed in years with poor harvests [227]. In some instances, animals have survived almost entirely on seaweed [208, 228]. While some ruminant populations, such as the North Ronaldsay sheep, have consumed seaweed for generations, the evolutionary events that endowed intestinal microorganisms with the potential to convert their chemically complex polysaccharides into a source of nutrition for their hosts, are as-yet unknown. In this study, we investigated the adaptation of two geographically distinct ruminant microbiomes to dietary kelp alginates. To evaluate microbiome responses, we first analyzed rumen samples from an *in vivo* feeding experiment, where naïve lambs were fed 2.5 and 5.0% *S. latissima* on dry matter (DM) basis. Through meta-omic analyses, we demonstrate the presence of functionally active rumen-associated *Prevotella* populations with the enzymatic machinery required to catabolize brown seaweed alginate through ‘alginate utilization loci’ (AULs). These results were validated in an unrelated rumen ecosystem using RUSITEC systems, where rumen content from cattle was incubated with *S. latissima* at low (2.0%) and high (50%) inclusion

rates. Interactions between alginate and microbial cells were confirmed using fluorescently labelled polysaccharides and extracted rumen communities. The activity of *Bacteroidota* AULs was further validated using recombinant enzymes derived from both lamb and cattle datasets. Alginate-consuming microbial populations were present across different production systems and different host species, yet with high identity between alginate lyases within the identified AULs. Our results underscore the remarkable ability of rumen microbial communities to rapidly adapt to novel and complex polysaccharides.

2.4 Results

2.4.1 Algal polysaccharides in *S. latissima*

A complete understanding of ruminal kelp digestion requires knowledge of the chemical and structural diversity of kelp polysaccharides. To determine the linkage composition of *S. latissima* kelp harvested off west coast of Canada was subjected to GC-MS based glycosidic linkage analysis using a recently established procedure optimized for unfractionated cell wall of brown seaweeds [91]. The identification of all permethylated alditol acetates (PMAAs), including those from 4-Gul*p*A and 4-Man*p*A, was based on their characteristic MS fragmentation patterns and retention times (**Figure 2.1B & 2.1C**). The relative abundance of linkages identified from *S. latissima* (**Table A1.1**; **Figure A1.1**) were used to estimate the relative compositions of total cell wall polysaccharides including cellulose, mixed-linkage glucan, laminarin, fucoidan, and alginate (**Figure 2.1D**). Cellulose was found to be the most abundant polysaccharide, as indicated by the dominance of 4-Glc*p*, which is consistent with our comparative analysis of various brown seaweeds [91]. Fucoidans, whose linkages were assigned based on previous study [89], exhibited a highly diverse linkage pattern, including t-Fuc*p* (2.1%), 2-Fuc*p* (1.2%), 3-Fuc*p* (3.2%), 4-Fuc*p* (1.7%), 2,3-Fuc*p* (1.1%), 2,4-Fuc*p* (0.8%), 3,4-Fuc*p* (2.0%), and 2,3,4-Fuc*p* (5.1%) linkages, together representing 17.3% of the cell wall polysaccharides in *S. latissima*. Alginate constituted 14.9% of the total relative abundance, with Gul*p*A linkages (10.1%) being more prevalent than Man*p*A linkages (4.6%), suggesting an M/G ratio of 0.46. The laminarin composition was calculated to be 1.3%, based on the sum of 3-Glc*p* and 3,6-Glc*p*, multiplied by two to account for the corresponding t-Glc*p* of 3,6-Glc*p* side

chains. Of note, mixed-linkage glucan linkages may be confounded with those found in laminarin and cellulose.

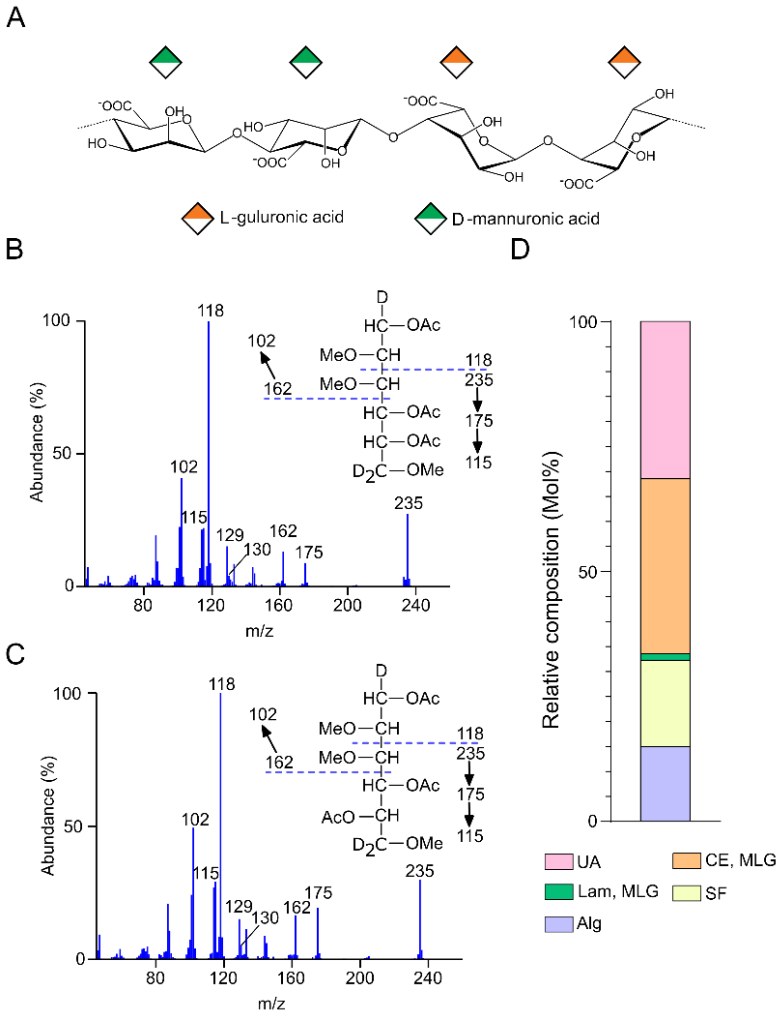

Figure 2.1: Linkage analysis and polysaccharide estimation of the brown algae *S. latissima* cell walls. A) Alginate is a high molecular weight, unbranched matrix polymer composed of two uronic acids: α-L-guluronic acid and β-D-mannuronic acid. These are linked by 1,4-glycosidic bonds and arranged in blocks of either homodimers (polyM or polyG) or heterodimers (polyMG). **B)** EI-MS spectra and ion fragmentation patterns for 4-GulpA and **C)** 4-ManpA from PMAAs of *S. latissima* alginate. **D)** Estimated polysaccharide composition (Mol%) of cell walls extracted from *S. latissima* (n=2). The following abbreviations were used for their respective polysaccharides and assigned according to the literature [229]: SF (fucoidans – All Fucp, All 2-Manp, All 6-Galp, t-Galp, 3,4-Galp, 4-GlcpA, 3-GlcpA, t-Xylp); CE/MLG (cellulose, mixed-linked glucan – 4-Glcp + 4,6-Glcp + (t-Glcp – 3,6-Glcp)); Alg (alginate – All GulpA and ManpA linkages); Lam/MLG (laminarin, mixed-linkage glucans – 3-Glcp + 3,6-Glcp*2); UA (unassigned – all remaining linkages).

2.4.2 Niche specialization dictates decomposition of brown algae polysaccharides in naïve ruminants

In our initial study of the rumen microbiome from Norwegian White lambs (n=24), 16S rRNA gene analysis was conducted on both fluid and fiber-attached rumen samples obtained from all animal dietary groups participating in the month-long feeding trial, including control, low seaweed (2.5% *S. latissima* on DM basis) and high seaweed (5.0% *S. latissima* on DM basis) (**Text A1.1**). As expected, notable differences in the community structure between the fluid and fiber-attached samples were observed across all three dietary groups. However, no definitive evidence indicating a macroalgae-induced alteration in the microbiome structure was identified (**Figure A1.2**). In the bovine RUSITEC experiment, where the range of seaweed inclusion was 2.0 and 50% *S. latissima* (DM basis), significant changes to rumen microbial communities were observed between time points (**Figure A1.3A**; ANOSIM and PERMANOVA, $P = 0.001$) and 50% *S. latissima* inclusion (**Figure A1.3B**; ANOSIM, $P = 0.005$; PERMANOVA, $P = 0.008$) compared to the control forage. RUSITEC communities supplemented with *S. latissima* also showed lower Shannon and inverse Simpson diversity values compared to control communities, although the Chao1 richness index was not significantly altered over time, or by seaweed supplementation (**Figure A1.3C**). Notably, an evident increase in the relative abundance of *Bacteroide*s was seen at the 8-day time point in the 50% *S. latissima* treatment (**Figure A1.3D**). The capability of the RUSITEC microbial communities to interact with *S. latissima* polysaccharides was investigated *in vivo* using fluorescent *S. latissima* extract (FLA-SLAT). Cells within the 2.0% and 50% *S. latissima*-treatments in the RUSITEC demonstrated uptake of FLA-SLAT, which overlayed with 4'6-diamidino-2-phenylindole (DAPI) DNA staining (**Figure A1.4**). In both rumen systems, digestibility of seaweed was observed with no detrimental impacts (**Text A1.1**).

Although the overall rumen community structure seemed unaffected by dietary supplementation of seaweed when given at biologically relevant doses, we wanted to further elucidate the genetic potential to utilize brown seaweed within the microbiomes. For this, shotgun metagenomic sequencing and metagenomic assembled genomes (MAG) reconstruction were carried out on lamb rumen samples from all three dietary groups, as well as for the control and 50% *S. latissima*-treated RUSITEC communities.

Reconstruction of the lamb rumen metagenome yielded 260 MAGs (named "$MAGX_{OV}$") of medium to high quality, defined by the MIMAG standard for completeness and contamination [230]. Most of these MAGs were affiliated with the *Bacteroidota* phylum (n=148), followed by *Bacillota* (n=36) and *Methanobacteriota* (n=7) (**Figure 2.2A**). When investigating the gene content of these MAGs, we identified enzymes annotated as alginate lyases encoded within 28 MAGs (**Extended Tables A1.1**). Most of these MAGs were associated with members of *Bacteroidota*, along with *Fibrobacterota* and *Verrucomicrobiota*. Due to the structural composition of alginate, alginolytic systems often appear in gene clusters, or AULs, in Gram-negative bacteria; these systems consist of Sus-like proteins and alginate lyases with complementary specificities [223, 231, 232]. We therefore extended our gene search to include the co-occurrence of neighboring SusC-like and SusD-like substrate-binding proteins, which are archetypical components of PULs [233-235]. This refinement led to the identification of ten *Bacteroidota* MAGs, of which four were classified as *Prevotella* species, five assigned to the *Bacteroidales* RC9 clade, and one to the *Paludibacteriaceae* family (**Figure 2.2A**).

The genomes recovered from the RUSITEC systems accounted for another 132 MAGs (named "$MAGX_{BOV}$"), with a noticeable fraction being *Bacillota* (n=64) and *Bacteroidota* (n=32) (**Figure 2.2A**). Like the lamb microbiome, the majority of alginate lyases were found within *Bacteroidota* phylum members originating from the *S. latissima*-treated RUSITEC system (**Extended Tables A1.1**); six *Bacteroidota* MAGs contained alginate lyases, five of which also harbored intact AULs (**Figure 2.2A**). Notably, an alginate utilization cluster (AUC) was also identified within $MAG32_{BOV}$ classified as *Brevimundis bullata* (87% complete; contamination <1%), which does not have the traditional SusC/D-like system.

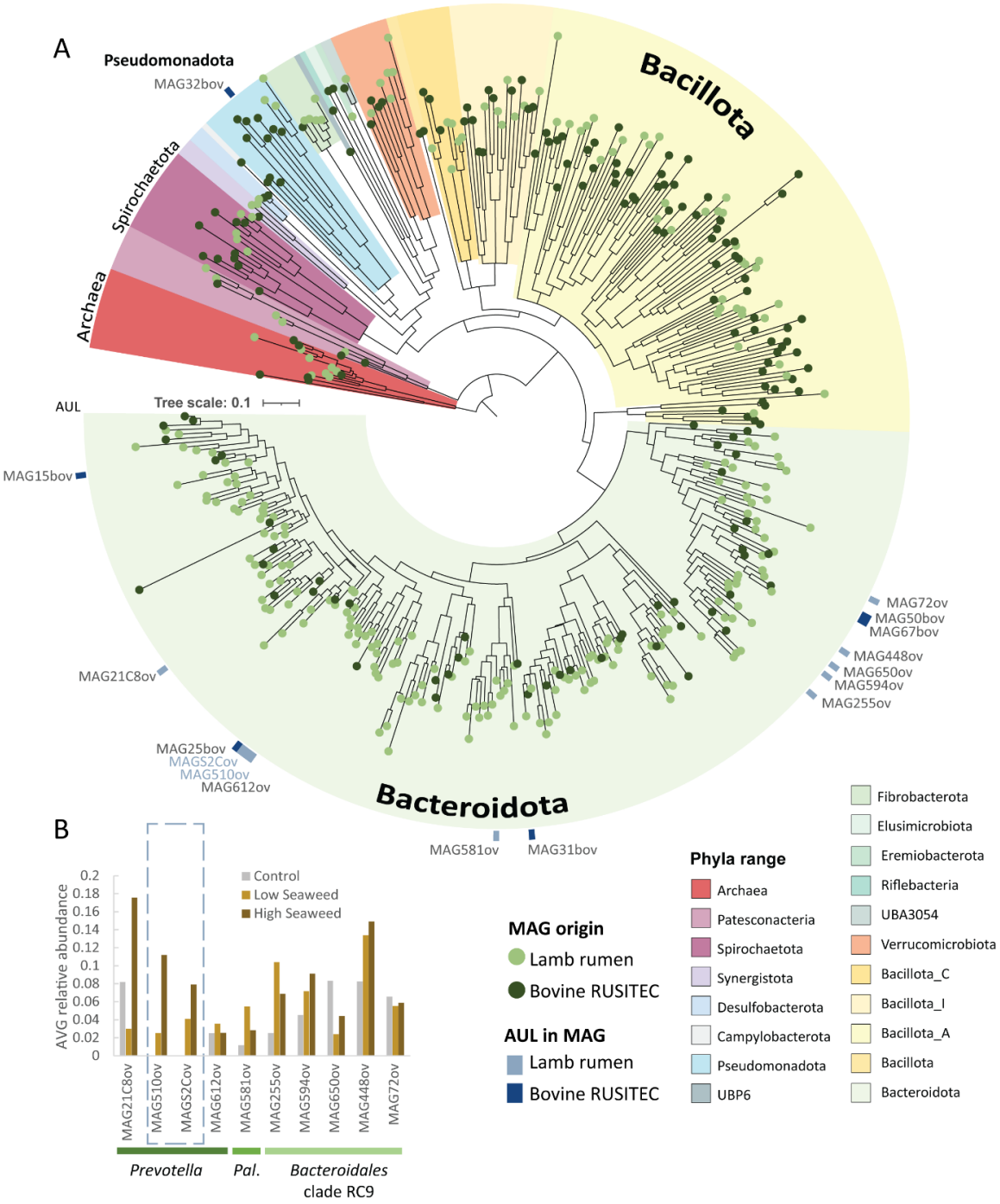

Figure 2.2: Rumen microbes encode enzymes for alginate utilization. A) Phylogenetic tree of MAGs recovered from lamb rumen (light green circle on branch) and the bovine RUSITEC (dark green circle on branch) microbiomes. In total 16 MAGs, of which all except one (MAG32BOV) are affiliated to the *Bacteroidota*-phylum, encoded AUL for alginate decomposition. **B)** Relative abundance of the lamb rumen MAGs encoding alginate lyases were calculated for each sample using CoverM and displayed as the average of each diet group. Two closely related *Prevotella* MAGs harboring AUL (MAGS2Cov

and MAG510$_{OV}$) demonstrated dose-dependent increase in relative abundance in response to inclusion *of S. latissima* in the diet. Pal.; *Paludibacteraceae*.

Intriguingly, among the lamb-originating MAGs encoding alginate lyases, two closely related *Prevotella*-affiliated MAGs, MAG510$_{OV}$ and MAGS2C$_{OV}$, demonstrated a pronounced increase in relative abundance in response to increasing levels of *S. latissima* in the diet. In contrast to the other potential alginate-degrading genotypes, none of these two *Prevotella* MAGs were detected in fiber-attached rumen samples from the control diet group (**Figure 2.2B**), further indicating that their presence correlates with *S. latissima* inclusion in the diet. The relative abundance of all MAGs can be found in **Extended Tables A1.1**.

2.4.3 Functionally active alginate consuming *Prevotella* populations

To investigate whether the putative alginate-degrading populations were functionally active, we analyzed the metaproteome of rumen samples from all lambs in the control and high *S. latissima* (5.0%) dietary groups. Our metagenome-centered metaproteomic analysis resulted in the detection of 10,312 protein groups that mapped back to our sample-specific database (**Extended Tables A1.2**). Among these protein groups, we detected two groups annotated as alginate lyases (**Figure 2.3A**), suggesting a role in the depolymerization of alginate. Notably, these two alginate lyases, both PL6, mapped back to the two *Prevotella* MAGs (MAG510$_{OV}$ and MAGS2C$_{OV}$) that demonstrated a dose-dependent increase in relative abundance in response to inclusion of *S. latissima* in the diet. While the peptides constituting one of the detected PL6 enzymes appeared to uniquely match MAGS2C$_{OV}$, the other PL6 undistinguishably mapped back to protein sequences found in both *Prevotella*-affiliated MAGs. The alignment of the full-length protein sequences within the shared PL6 protein group revealed a high degree of sequence similarity (BLASTp: 96% id), suggesting a common ancestral origin and functional redundancy in their substrate specificity. The label-free quantification (LFQ) of the detected proteins confirmed that the two PL6 enzymes were enriched in the animal group fed with high inclusion of *S. latissima* (**Figure 2.3B**). Of note, in both MAG510$_{OV}$ and MAGS2C$_{OV}$, the genes of the expressed PL6 enzyme were neighboring another gene encoding an enzyme from a known alginate lyase family; PL17 (no protein detection).

In addition to the detection of the PL6 enzyme, a proteome inspection of $MAG510_{OV}$ also unveiled differential expression of a SusC-like and its adjacent SusD-like substrate-binding protein in the 5% *S. latissima* dietary group compared to the control. Specifically, the $MAG510_{OV}$-affiliated SusD-like protein was the most upregulated protein with *S. latissima* supplementation in the entire dataset (**Figure 2.3A**). Another essential component of the Sus system, the SusE-like outer membrane protein, was also detected in the proteome of $MAG510_{OV}$, yet at a lower LFQ detection level. Within a PUL, the *susC*- and *susD*-like genes are typically situated in close proximity to the relevant CAZymes [235]. However, due to genome fragmentation, the susC/susD-like pair and the alginate lyases (*pl6* and *pl17*) were found towards the edges on two separate contigs, and thus, a complete AUL could not be determined. Despite their physical separation on different contigs, possibly due to incomplete genome reconstruction, the explicit co-expression of *susC*/*susD-like* and *pl6* implies that these genes may still function as a part of a co-regulated AUL.

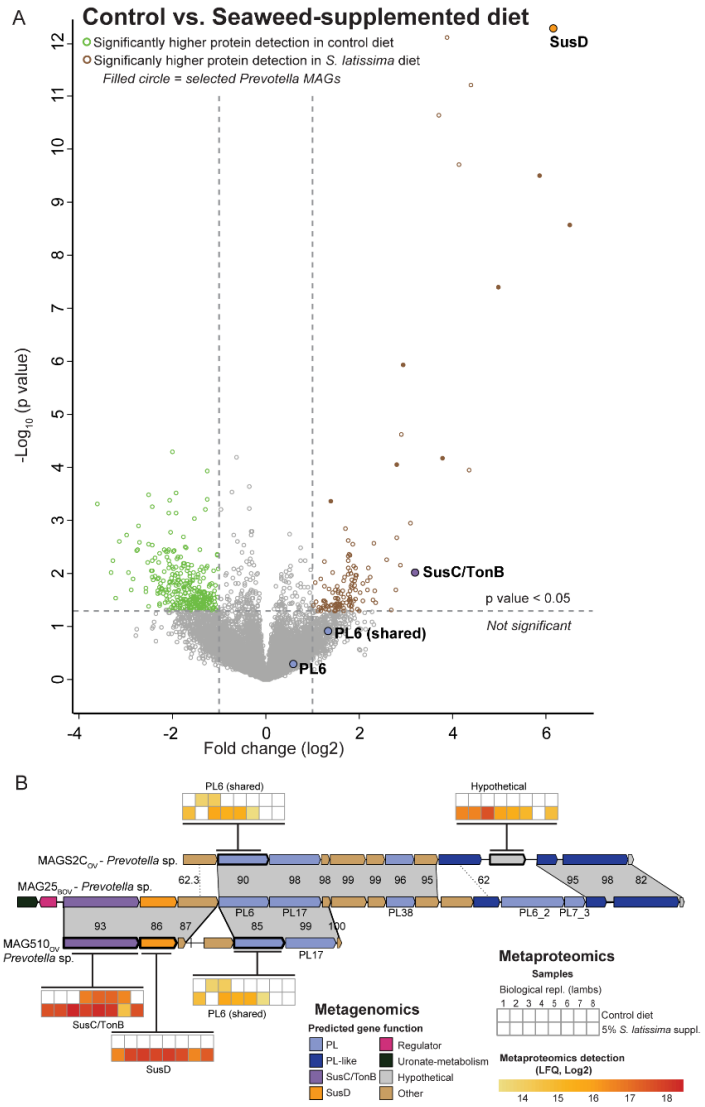

Figure 2.3: Detection of alginate lyases within AULs. A) Volcano plot of all protein groups identified in lamb rumen samples demonstrates detection of two alginate lyases (PL6) from *Prevotella* spp. (MAG510$_{OV}$ and MAGS2C$_{OV}$), along with a SusC- and SusD-like that were significantly higher expressed in the 5% seaweed-supplemented diet group compared to the control group. **B)** According to the predicted gene organization, the PL6 enzymes and SusC/SusD-like were located within AULs that showed high synteny to AULs extracted from *Prevotella* spp. MAG25$_{BOV}$ found in the cattle-based RUSITEC microbiome. Gray shading indicates homology and % identity values. Genes with bolded outlines were detected within the lamb metaproteome, and the heatmaps show the protein intensities (LFQ values for the eight replicates sampled in control and 5% *S. latissima*-fed lambs).

2.4.4 *Prevotella* spp. AULs in globally separated rumen ecosystems

We observed that AULs found in *Bacteroidota* originating from the lamb rumen microbiome and cattle-based RUSITEC system from two different continents were highly syntenic and contained highly similar sequences to the other AULs. In total, eight *Bacteroidota* phyla members had intact AULs, along with the AUC from the *B. bullata* MAG32$_{BOV}$. All MAGs contained at least one PL6 and PL17 pair, with some containing either additional PL6, PL7, or PL38 sequences within the AULs (**Figure 2.4**). Between RUSITEC and lamb *Bacteroidota* species MAGs, the highest similarity was discovered between *Prevotella* MAGs, specifically the cattle-based RUSITEC MAG25$_{BOV}$, and lamb rumen MAGs MAGS2C$_{OV}$ and MAG510$_{OV}$ showing amino acid identity as high as 99% (**Figure 2.3B, Extended Tables A1.1**). This similarity extends to the alginate lyases, as the PL17 members across the three AULs shared 98% amino acid identity, while the shared PL6 members exhibit 85% identity. The PL38 protein shared between MAG25$_{BOV}$ and MAGS2C$_{OV}$ displayed 96% sequence homology.

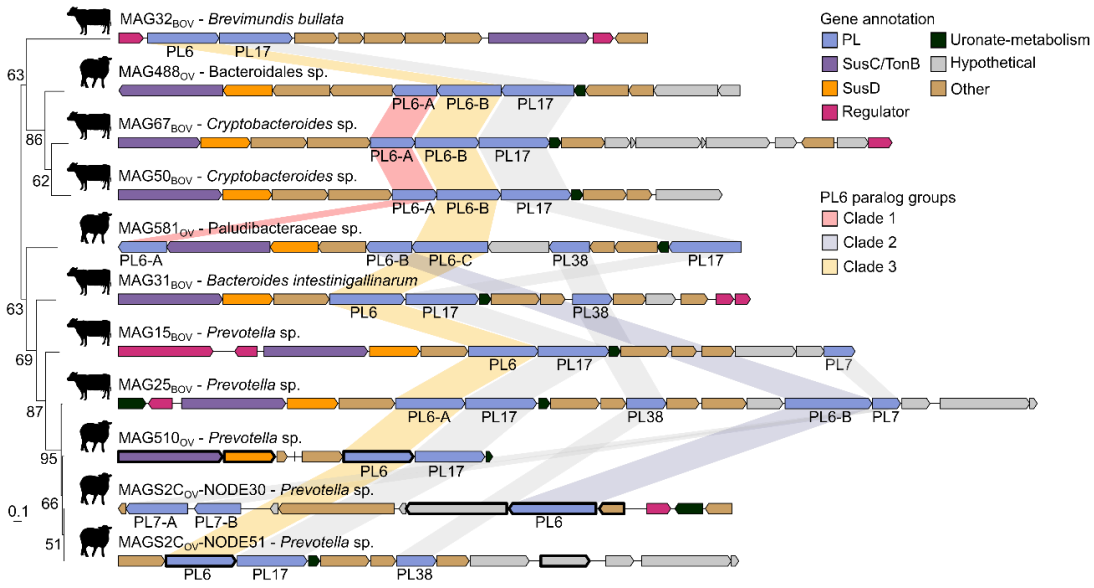

Figure 2.4: Synteny between lamb and cattle alginate AULs. OrthoFinder was used to generate a species tree composed of MAGs containing AUL/AUCs. ClustalOmega, BLAST, and SACCHARIS were used to identify orthologs and similarity between AUL/AUC genes. The AULs and AUC were created with the gggenes package. Bolded genes were identified within the lamb metaproteomics.

Using SACCHARIS [154], the PL6 enzymes within *Bacteroidota* AULs, along with homologous PL6s found on shorter contigs in both the RUSITEC and lamb MAGs, partitioned into three distinct clades based on amino acid identity. In addition, the PL6 enzyme found within *B. bullata* $MAG32_{BOV}$ formed a more distantly related group (Clade 4) compared to the other three clades. Aligning these PL6 members to all PL6 enzymes within the CAZy database revealed that Clade 1, 3, and 4 consisted entirely of PL6 subfamily 1 members, while Clade 2 was composed of subfamily 2 members (**Figure 2.5A; Figure A1.5; Figure A1.6**). Members of Clade 1 and 2 differed in structure from those of Clade 3 and 4, as they lacked the predicted inactive C-terminal domain present in Clade 3 and 4 members (**Figure A1.7**) [236]. Further inspection of the phylogeny showed that PL6 members of both Clade 1 and Clade 2 were most closely related to characterized sequences from *Rhodothermus marinus* DSM 4252. Specifically, Clade 1 members shared the highest sequence similarity (37-40% identity) with Rmar_1165, an enzyme active on poly-MG alginate blocks (consisting of both Guluronate: G and Manuronate: M). Similarly, Clade 2 members were most closely related to Rmar_1386 (38-42%), also active on poly MG alginate (**Figure A1.5**) [237]. A majority of Clade 3 members had the highest sequence similarity (41-48%) to an exo-type alginate lyase of human gut bacterium *Bacteroides clarus* [236], except $MAG581_{OV}$-C, which was more similar (33.43%) to a poly-MG-active PL6 enzyme characterized from the marine bacterium *Nonlabens ulvanivorans* [237]. Lastly, the $MAG32_{BOV}$ PL6 member shared 55% sequence similarity with an endo-poly-MG alginate active PL6 from *Stenotrophomas maltophilia* KJ-2 (**Figure A1.6**) [238]. The clustering of PL6 contrasted with PL17 members which shared high similarity and clustered together within subfamily 1 (**Figure 2.5A**).

The similarities between *Bacteroidota* PL6s from the gastrointestinal tracts of terrestrial animals and those found in marine/marine sediment microbiota suggested that some level of gene transfer may be involved in the acquisition of alginate saccharification. NOTUNG [239] was used to reconcile OrthoFinder species trees and PL17 and PL6s gene trees between 75 genomes from marine environments and terrestrial gastrointestinal tracts (**Table A1.4**). PL6 members were either separated into paralog clades or kept together as an ortholog group. The presence and absence of PL6 and PL17 members were predicted to be a result of gene duplication and loss events, rather than HGT events. The only HGT

event predicted was observed within clade 2 between terrestrial host MAGs: MAG581$_{OV}$ (*Paludibacteraceae* sp.) and the *Prevotellaceae* family (MAG25$_{BOV}$ and MAGS2C$_{OV}$) (**Figure A1.8; Table A1.5**). Amongst almost all species analyzed, PL17 and PL6 Clade 3 members are often associated and make up the core component of AULs in *Bacteroidaceae* (**Figure 2.5B**). Furthering this data, GC content compared between AUL/AUC PLs in this study was similar to the remainder of respective their genomes (**Figure 2.5C**), suggesting these genes were not part of a HGT event.

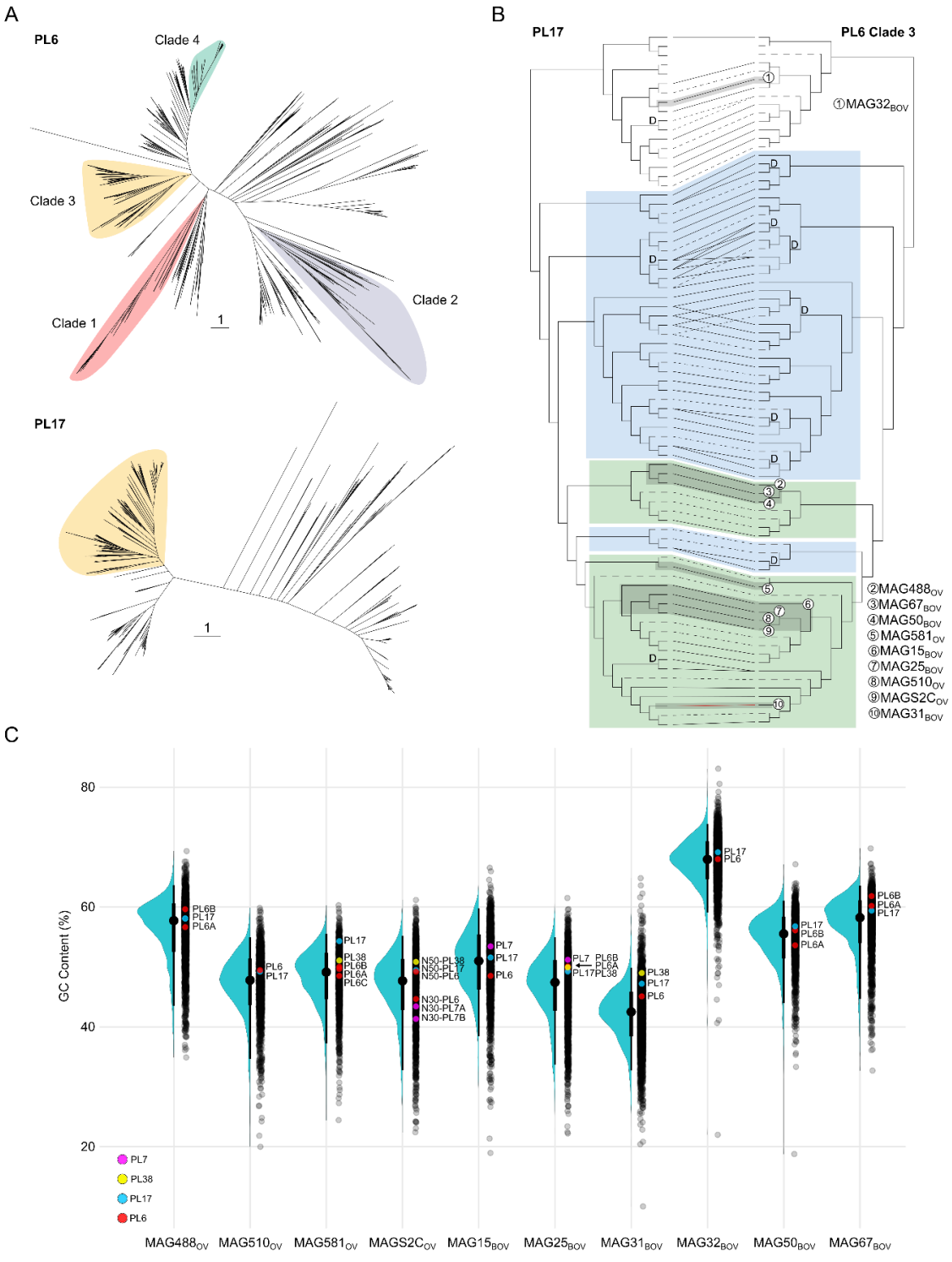
A
PL6
Clade 4
Clade 3
Clade 2
Clade 1
1
PL17
1
B
PL17
PL6 Clade 3
①MAG32BOV
D
②MAG488OV
③MAG67BOV
④MAG50BOV
⑤MAG581OV
⑥MAG15BOV
⑦MAG25BOV
⑧MAG510OV
⑨MAGS2COV
⑩MAG31BOV
C
GC Content (%)
80
60
40
20
PL7
PL38
PL17
PL6
PL6B
PL17
PL6A
PL6
PL17
PL17
PL38
PL6B
PL6A
PL6C
N50-PL38
N50-PL17
N50-PL6
N30-PL6
N30-PL7A
N30-PL7B
PL7
PL17
PL6
PL7
PL6B
PL6A
PL17
PL38
PL38
PL17
PL6
PL17
PL6
PL17
PL6B
PL6A
PL6B
PL6A
PL17
MAG488OV
MAG510OV
MAG581OV
MAGS2COV
MAG15BOV
MAG25BOV
MAG31BOV
MAG32BOV
MAG50BOV
MAG67BOV

Figure 2.5: Synteny between lamb and cattle alginate AULs. A) Phylogenetic trees generated from all sequences from PL6 and PL17 within the CAZy database, combined with those predicted from lamb and cattle MAGs. Phylogenetic clades with uncharacterized MAG sequences are shaded and given labels 1-4. **B)** Comparison of PL17 and PL6 clade 3 NOTUNG gene trees. Gene loss (dotted lines) and duplications "D" are marked per species, and lines are drawn between trees to connect sequences within the same genome. MAGs identified in this study are shaded gray, while terrestrial *Bacteroidota* species are shaded green and marine *Bacteroidota* species. are shaded blue. The full labeled tree is present in **Figure A1.9**. **C)** GC content of each gene with ruminant MAGs was plotted using half-eye plots made with ggdist [240]. Default quartiles were used (95% and 66%) and visualized by the lines on the density plot. Key alginate PL family members within AULs were highlighted.

Overall, PL6 members of Clade 3 were found in all *Bacteroidota* MAGs encoding AULs, but not within the AUC identified in $MAG32_{BOV}$. Near half of the PULs contained multiple PL6 members belonging to different clades, of which the combination of Clade 1 and Clade 3 were most prevalently occurring in both lamb rumen (*Bacteroidales* sp. $MAG488_{OV}$ and *Paludibacteraceae* sp. $MAG581_{OV}$) and cattle-based RUSITEC MAGs (*Cryptobacteriales* spp. $MAG50_{BOV}$ and $MAG67_{BOV}$) (**Figure 2.6A**). The *Prevotella* MAGs seemed to combine PL6 Clade 3 with Clade 2 within the same AUL ($MAG25_{BOV}$ and $MAG581_{OV}$), or at separate loci ($MAGS2C_{OV}$). Of note, the two *Prevotella*-affiliated PL6 proteins detected in the metaproteome belonged to Clade 2 and Clade 3 (shared PL6).

To confirm the predicted alginate activity of AUL/AUCs, PL6 members from each clade (Clade 1: $MAG581_{OV}$-A, $MAG50_{BOV}$-A; Clade 2: $MAG581_{OV}$-B, $MAG25_{BOV}$-B; Clade 3: $MAG581_{OV}$-C; Clade 4: $MAG32_{BOV}$) were selected for gene synthesis and functional characterization. All selected PL6 members contained SPI signal peptides, except for $MAG50_{BOV}$-A which had an SPII signal peptide (not shown, as signal peptides were ultimately removed for protein expression in this study). Enzymes were incubated with alginate from brown seaweed and observed after overnight digestion, as well as continuous direct measurement of unsaturated product formation at 232 nm. Direct measurement was used to determine the optimum pH and initial velocities of the selected PL6 enzymes in the linear range of the reaction (**Figure A1.10**). All PL6 members, except $MAG581_{OV}$-B, produced absorbance at 232 nm, indicating enzymatic activity. However, kinetics of most PL6 enzymes demonstrated inefficient digestion of alginate, with the exception of $MAG50_{BOV}$-A, and Michaelis-Menten could not be accurately fitted (**Figure**

A1.10F). Despite variations in catalytic efficiencies, all PL6 members produced oligosaccharides when incubated with alginate (**Figure 2.6B**), however, only $MAG50_{BOV}$-A sufficiently degraded poly-G & poly-M oligosaccharides (**Figure A1.11**). Using High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) and thin-layer chromatography (TLC), PL6 members from the same clade displayed overlapping activities. Liquid chromatography with electrospray ionization mass spectrometry (LC-ESI-MS) was used to further characterize the structure of unsaturated oligosaccharides (**Figure 2.6C & 2.6D**).

All PL6s produced alginate oligosaccharides of varying degrees of depolymerization. Clade 3 and Clade 4 members produced monosaccharides that appear to decompose through a keto-enol intermediate into 4-deoxy-L-erythro-5-hexoseulose uronate [241-243]. Both $MAG581_{OV}$-B and $MAG25_{BOV}$-B appeared less active on brown seaweed alginate, producing only small amounts of alginate oligosaccharides.

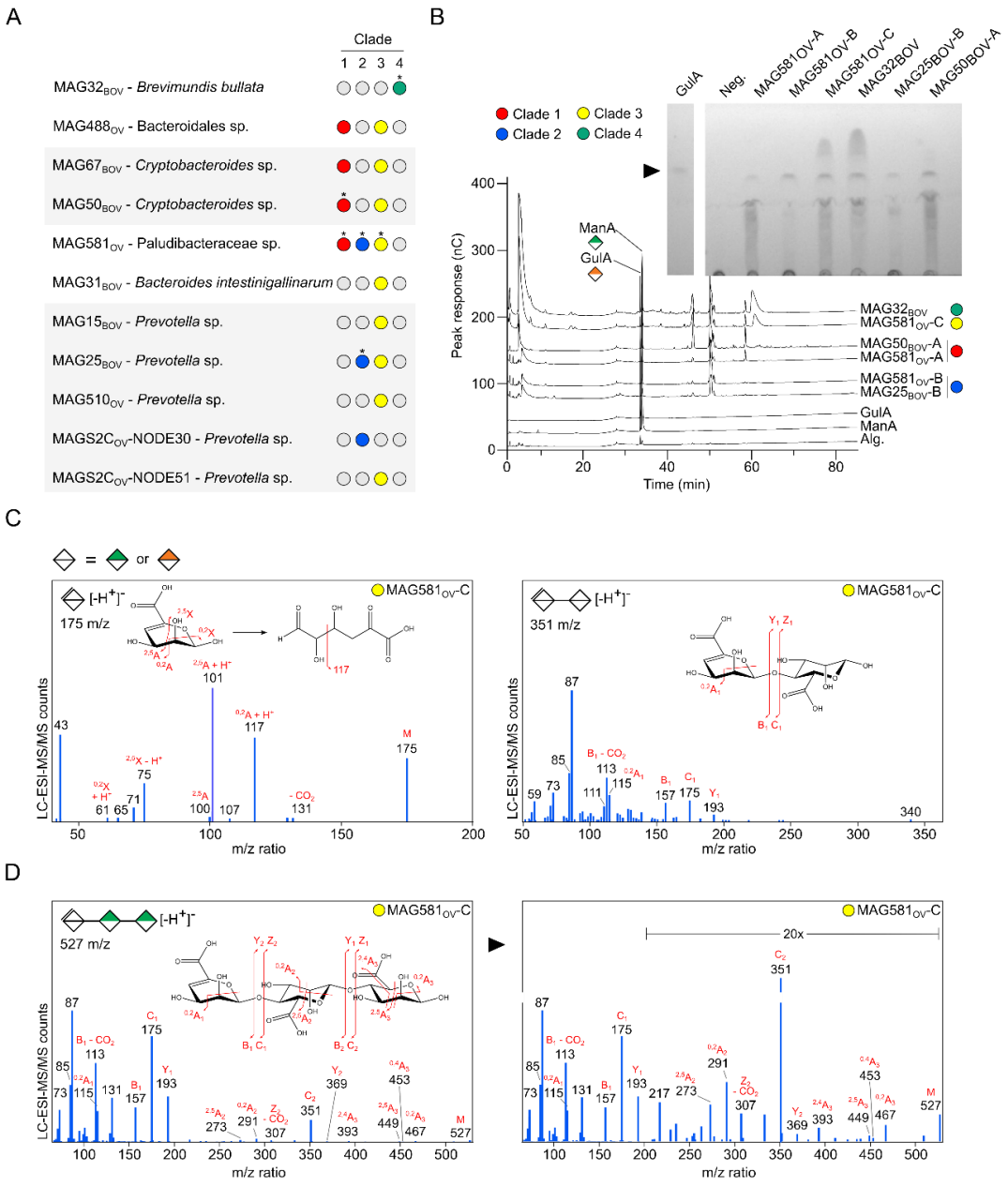

Figure 2.6: Characterization of alginate oligosaccharide products from *Bacteroidota* species PL6 alginate lyases. A) Summary of which lamb and rumen MAGs containing AUL/AUCs contain PL6 members within clades 1-4. Black stars indicate sequences which were selected for further characterization. **B)** TLC and HPAEC-PAD product analysis of brown seaweed alginate digested with lamb and cattle-associated MAG PL6 enzymes. Man*p*A and Gul*p*A represent monosaccharide standards. **C)** Alginate lyase oligosaccharide products were analyzed by LC-ESI-MS/MS. Ions of m/z 175 (left) and 351 (right) corresponding to unsaturated mono- and disaccharide products were observed, and spectra were extracted. ESI-MS/MS with HCD was performed in order to confirm identity of the extracted ions, and MS2 product ion spectra are shown with the carbohydrate

fragmentation depicted and identified. **D)** MS2 spectra of ion 527, corresponding to unsaturated trisaccharide, identified as ΔHexA-ManA-ManA based on the presence of signature ions. Monosaccharide symbols are displayed according to the Symbol Nomenclature for Glycans system. For monosaccharides unable to be discerned between Gul*p*A and Man*p*A by LC-ESI-MS, white symbols are used.

2.5 Discussion

2.5.1 *S. latissima* cell wall structure

The matrix cell wall polysaccharides of brown seaweeds are largely comprised of anionic polysaccharides, such as alginate and fucoidan [215]. Their charge density corresponds with the ionic nature of ocean water and enables metal-dependent, higher-order polymer cell wall dynamics, such as gelation. The polysaccharide composition of *S. latissima* was confirmed to consist of cellulose and mixed-linkage glucans, and the matrix polysaccharides fucoidan and alginate. Alginate comprised approximately 15% of the total plant cell wall polysaccharides in *S. latissima* (**Figure 2.1D**; **Table A2.1**), which is in agreement with previous reports [91]. Fucoidan comprised ~17% of the estimated polysaccharide content, almost equivalent to that of alginate. *S. latissima* fucoidan is known to be structurally complex, consisting of highly sulfated α-1,3-L-fucan backbones, with side chains composed of β-1,6-D-galactose or alternating β-D-glucuronic acid and α-D-mannuronic acid extensions [89]. Despite the abundance of fucoidan, we were unable to detect any putative fucanases from the lamb feeding experiment or bovine RUSITEC metagenomic and metaproteomics datasets, suggesting that fucoidan is likely not digested by either ruminant. To date, the discovery of a fucanase or fucoidan-specific sulfatase from a terrestrial gut microorganism has not been reported. Whether this stems from its structural complexity, bioavailability, or other factors remains to be determined.

2.5.2 Structure and abundance of AULs from ruminant datasets

Bacteroidota MAGs with the genetic capacity to degrade alginate were constructed from the lamb feeding experiment and bovine RUSITEC sequencing data. Intriguingly, MAGs that encoded the capacity to consume alginate included two closely related *Prevotella* strains that exhibited diet-induced dose responses in their relative abundance (**Figure 2.4**). Metaproteomics further confirmed these *Prevotella* species produced PLs specific for the digestion of alginate (**Figure 2.3**). Reconstruction of the pathway

architecture containing these alginate lyases indicated they were functional components of AULs, similar to those previously observed in marine and human gut bacteria [221, 223, 244] (**Figure 2.4**). Likewise, PL17 members co-localized with PL6 subfamily 1 members in Clade 3 [221]. However, it appears that members of PL6 subfamily 2 and subfamily 1 Clade 1 are ancillary members brought about through gene duplication. Furthermore, the detection of elevated levels of SusC-like and SusD-like proteins in the *S. latissima* diet group reinforces the relevance of these AULs for alginate digestion. Surprisingly, a large-scale diet-induced difference was not observed in the overall microbial community structure when evaluated by 16S rRNA gene sequencing. This suggested that a limited amount of specific rumen microbial species play roles by occupying a specialized niche for digestion of seaweed-derived polysaccharides. Notably, other PL enzymes associated with alginate consumption were also identified within the AULs, though their expression was not detected. This could be due to several technical factors, such as their amenability to protein extractions methods and their absolute levels being below proteomic detection thresholds. Moreover, a plethora of biological factors could plausibly explain an absence of detection, including alginate specificity and/or percent inclusion level required to necessitate expression, compartmentalized secretion, specificity of induction products (*i.e.*, which digest products induce AULs) [245, 246], or insufficient adaptation time for efficient utilization of seaweed. Others have suggested that alginate polymers decompose slowly [244, 247].

Increasing the overall alginate concentration to 50% of the total mixed ration (25-fold over the lamb feeding experiment) within the RUSITEC experiment, allowed us to discern multiple *Bacteroidota* organisms capable of digesting alginate (**Figure 2.4A**). In contrast, *Prevotella* spp. abundance was not significantly different between control and *S. latissima*-supplemented RUSITECs, and, although expected [248], we identified multiple *Prevotella* spp. within RUSITEC-derived MAGs containing AULs.

To measure direct interactions between polysaccharides and microorganisms in the rumen, we previously developed a technique using fluorescent polysaccharides, or FLAPS [200]. This enables the response of microorganisms to labeled substrates to be quantified at the single-cell level. Accumulation of fluorescent polysaccharide extracts from *S. latissima* (FLA-SLAT) (**Figure A1.4**) in microbial cells, provided further support that

alginate is consumed, and this occurs in response to incubation with *S. latissima*. SR-SIM imaging following FLA-SLAT incubation showed that most interactions with FLA-SLAT occurred in the periphery of cells, which is reminiscent of Gram-negative bacteria that employ a selfish mechanism of foraging [200, 249, 250].

2.5.3 Function of PLs from ruminant AULs

The operonic AUL structures within this study were identified using previously characterized lyase families: PL6, PL7, PL17, and PL38. While members of PL6 have been shown to be both endo- and exo-acting, they typically display an endo-mode of action, indiscriminately cleaving internal glycosidic bonds within the alginate polymer to produce oligosaccharides with C4 C5 unsaturations at their non-reducing ends. In contrast, PL17 enzymes are often categorized as exo-acting oligoalginate lyases that primarily act on the polyGM structure at the end of an alginate chain, releasing unsaturated mono- or disaccharides of mannuronate [221]. Previous studies of alginate consuming marine and human gut bacteria have shown that PL6 and PL17 members are frequently co-located, strongly suggesting that they may operate in a coordinated manner to efficiently depolymerize alginate [221, 223]. In addition to PL6 and PL17, some MAGs encoded alginate lyases belonging to the PL7 and PL38 families. To confirm the function of representative enzymes from these AULs, several PL6 members were selected for biochemical characterization. PL6 enzymes can possess “keystone” status by catalyzing the first stages of alginate depolymerization at the cell surface [251]. Several recombinant PL6 enzymes were shown to produce substantial amounts of unsaturated products by TLC, HPAEC-PAD, and LC-ESI-MS. Due to the complexity of M/G orientation and ratio, and the ability of PL6 enzymes to change mode of action (endo/exo) depending on either alginate substrate (poly-M or poly-G) [237], and the absence or presence of non-catalytic domains [236], precise activity was not demonstrated within this study. Notably, delineation between Man*p*A and Gul*p*A signatures are often indistinguishable using methods other than NMR; however, prediction of some oligosaccharide structures is possible based upon a previous study that defined patterns of differential MS2 fragmentation abundance of alginate oligosaccharides. For example, an identified unsaturated tri-saccharide, was found to contain signature ions (m/z of 307, and high ratio

of 449/453) suggesting the oligosaccharide product consists of a ΔHexA-ManA-ManA [252] (**Figure 2.5C**). Further, EI-MS and MS2 fragmentation revealed that some uronic acid residues of alginate may be sulfated (**Figure A1.12**). This unexpected result warrants further study.

2.5.4 Evolution of alginate metabolism within the rumen microbiome

Seaweed polysaccharides tend to be structurally and chemically complex, highly charged, and absent from terrestrial plants typically consumed by livestock. Members of the human microbiome are known to have catabolic pathways dedicated to digestion of polysaccharides from red [202, 205] and brown [221] seaweeds. Multiple hypotheses, including the 'sushi factor' [202, 222] and ancient transfer [221], have been proposed to explain the origin of these pathways in gut microbiomes. In contrast to humans, much less is known about the mechanisms by which livestock microbiomes adapt to the introduction of exotic dietary polysaccharides, such as alginate from brown seaweed. Previously, alginate-consuming *Prevotella* spp. were isolated from the North Ronaldsay sheep that live on the Orkney Islands, which have a diet augmented with brown seaweeds, likely *Laminaria*, *Fucus,* and *Ascophyllum* species [208]. Recent studies have also revealed that wild Svalbard reindeer populations are adapting to the warming Arctic climate by supplementing their diet with kelp, driven by an increasing prevalence of ice-locked pastures [253]. How these animals survived a radical change in their diet and lifestyle, and the evolutionary processes that drove microbial adaptation towards seaweed polysaccharide catabolism has not been thoroughly studied.

Alginate is produced by *Pseudomonas* spp. found natively within the rumen [94, 254]. However, it differs from seaweed alginates, containing lower M/G ratios and high levels of *O*-2, *O*-3 acetylation of mannuronic acid [94, 255]. Therefore, although we show that dietary *S. latissima* results in expression of AULs, it remains to be shown if alginate produced by autochthonous bacteria helps drive the evolution of these pathways. It has been previously speculated that xanthan gum utilization loci in human gut *Ruminococcaceae* evolved under selection of exopolysaccharides before being exposed to structural analogs found in food additives [256]. Likewise, it is possible that structurally related alginate substrates present a common niche for microbiota to fill in the rumen. One

distinguishing factor is the absence of carbohydrate esterases in the seaweed-alginate associated MAGs, which is a conceivable requirement for bacterial alginate metabolism, suggesting there is differential specificity [221, 257]. It is unclear from our data how adaptation to alginate effects other microorganisms, as lower Shannon diversity with the 50% RUSITEC trial (**Figure A1.3**) suggests overall lower microbial diversity and increasing algal content did not result in a noticeable shift in community structure. This suggests alginate does not function as a prebiotic in the rumen. It was found that carrageenan supplementation drastically altered the faecal microbiome composition, but not the rumen microbiome composition [258]. The role of the lower gut microbiota in alginate digestion should be explored in future work.

Here, we discover and characterize the structure of syntenic AULs responsible for alginate catabolism from both cattle RUSITEC and lamb microbial communities exposed to dietary *S. latissima*. Alginate catabolism appears to be a core functional trait performed by autochthonous members of the ruminant microbiome [259], indicating that these genomic loci were not recently transferred from the surface microbiome of *S. latissima* or another marine ecosystem. The conservation of PL6 and PL17 pairs within *Bacteroidota* species from geographically and taxonomically distinct ruminants, suggests that these genetic loci were acquired by one or a series of evolutionary events through speciation or duplication, however, there Is little evidence of horizontal gene transfer. In this light, perhaps AULs may have evolved in an ancestor common to humans and cattle. Despite conservation of function, there is noticeable structural variation in the gene content and organization (**Figure 2.4**). Regardless of how individual pathways assembled, the overall functional conservation of AULs within the ruminant microbiomes is clearly maintained throughout lineages and between populations. In this manner, the maintenance of seaweed alginate catabolism in the absence of substrate availability represents a “latent trait” within the rumen microbiome. Often, these pathways are encoded within the genomes of saccharolytic generalists, which likely provides a mechanism for retention of alginate competent strains when these substrates become scarce. Ultimately, this provides a nutritional benefit to the host and justifies the fitness cost required to maintain this unique catalytic machinery within *Bacteroidota* species. [260, 261]. Regardless, acquisition and maintenance of a diverse metabolic capacity appears to position the animal to rapidly adapt

to abrupt changes in its diet and provides tantalizing clues into the ancient foraging habits of its ancestors.

2.6 Methods

2.6.1 Preparation of unfractionated cell wall of *S. latissima*

S. latissima was harvested from the waters near Vancouver Island, B.C., Canada on May 1st, 2022. The cell wall extractions were prepared in accordance to Avci, Pattathil [262] with modifications according to previous reports [263]. Approximately 100 mg of ball-milled dry sample was extracted in 40 mL of absolute ethanol (EtOH) for 8 h, followed by three 8 h extractions in 40 mL of 80% (v/v) EtOH, and then two 20 min washes with absolute EtOH. After each alcohol extraction and wash, samples were centrifuged at 3,000 × *g* for 30 min. The final alcohol insoluble residue (AIR) was vacuum-dried by SpeedVac (Thermo Scientific, USA).

2.6.2 Glycosidic linkage analysis of cell wall

The AIR of *S. latissima* was analyzed as described [91]. Briefly, polysaccharides in dry AIR power (10 mg) were methylated by 1.2 mL of methyl iodide in 2 mL of dimethyl sulfoxide in the presence of excess sodium hydroxide [100, 264]. The methylation product was suspended in dichloromethane (DCM) and partitioned against 10% acetic acid in water (v/v) over ice once then against deionized water three times. After each partitioning, the upper phase was collected and pooled. Following the final partitioning, the lower phase was evaporated to dryness, and the pooled upper phase was mixed with the evaporated lower phase. The mixture was dialyzed with a molecular weight cut-off of 6,000 to 8,000 Da against 0.1 M triethylamine hydrochloride (TEAH) deionized water solution then against pure deionized water, followed by freeze-drying. Two additional rounds of methylation were conducted with the 0.1 M TEAH dialysis step omitted [91]. The permethylated sample was converted to (PMAAs by hydrolysis in 2 mL of 4 M trifluoroacetic acid (TFA) at 100 °C for 4 h, reduction with 20 mg of sodium borodeuteride ($NaBD_4$, 99 atom % D, Alfa Aesar) in 2 mL of deionized water for 16 h [265], and acetylation by heating in a mixture of acetic anhydride and TFA (5:1, v/v) at 60 °C for 1 h [91]. In a separate experiment, the dry AIR powder (10 mg) was subjected to weak

methanolysis (0.5 M methanolic HCl, 80 °C, 20 min), followed by $NaBD_4$ reduction to convert uronic acids to their corresponding 6,6'-dideuterated neutral sugars [266-268]. Excess $NaBD_4$ was neutralized by acetic acid, followed by evaporation of the product to dryness and repeated evaporation in 10% acetic acid in methanol (v/v) then in absolute methanol to remove borate, before the sample was converted to PMAAs [91]. All PMAAs were analyzed using an Agilent 7890A-5977B GC-MS system (Agilent Technologies, Santa Clara, CA, USA) coupled to a medium-polarity Supelco SP-2380 column (60 m × 0.25 mm × 0.2 μm; Sigma-Aldrich, USA) under a consistent flow of helium at 0.8 mL/min. Inlet temperature was maintained at 250 °C. Sample was injected with a 10:1 split ratio. The oven program started at 120 °C (hold 1 min) and then increased at 3 °C/min to 200 °C (hold 50 min) then at 3 °C min/min to 250 °C (hold 20 min). The PMAAs were identified by comparing their EI-MS fragmentation patterns and retention times to those of prepared PMAA standards [91] and also by referencing the literature (Carpita and Shea, 1989). Relative molar linkage compositions of the PMAAs were determined from the TIC chromatogram, following the principle that a PMAA's quantity correlates with the ratio of its TIC peak area to its molecular weight, as described in the published protocol [58]. The peak corresponding to the PMAA of 4-Glcp was used to normalize the relative abundances of neutral sugar PMAAs in the untreated sample and the PMAAs derived from uronic acid linkages in the sample pretreated with weak methanolysis-$NaBD_4$ reduction [91]. Three separate experiments were conducted to the AIR sample.

2.6.3 *In vivo* experimental design

A total of 24 Norwegian White lambs were used in the *in vivo* feeding experiment, as previously explained in detail by Grabež, Devle [269]. In brief, weaned ewe lambs with body weight 37.3 ± 1.6 kg were randomly assigned to three dietary treatment groups (n = 8 per group). All lambs had free access to clean drinking water and were fed the experimental diet ad-libitum twice a day (at 08:00 h and 14:00 h) in individual pens. The control group (no seaweed included) were fed a total mixed ration of grass silage and compound feed, rolled barley and mineral premix. The two other treatment groups were additionally supplemented with 2.5% and 5.0% *S. latissima* on a DM basis. The *S. latissima* used in the *in vivo* feeding experiment was harvested by Seaweed Solution AS (Trondheim,

Norway) at their cultivation site outside Frøya, Norway (collected 28/05/2018), and immediately frozen. Prior to the feeding experiment, the seaweed biomass was chopped to uniform particle size and wilted, before stored frozen again until final preparation of experimental diets. The experimental period lasted for 35 days, during which the lamb growth rate, daily feed intake, and feed dietary composition were regularly monitored. All animal procedures were approved by the committee overseeing the rules and regulations governing animal experiments in Norway under the surveillance of the Norwegian Food Safety Authority (FOTS-ID: 16406).

2.6.4 RUSITEC experimental design

The RUSITEC experiment followed a 2 x 2 + 1 factorial design with two experimental runs and two treatments. Each RUSITEC unit was equipped with eight 920 mL fermenters, designated with three replicates per treatment (n = 6) in each unit across both runs. Here, the basal substrate consisted of a 50:50 (DM basis) barley silage and barley straw diet. The two dietary treatments included control (no seaweed included) and *S. latissima* (collected from Vancouver Island, B.C., Canada, 01/05/2022) included at 2% dietary DM. An additional fermenter in each run was fed 50% *S. latissima* for microbial profiling. The seaweed replaced equal proportions of barley silage and barley straw, and treated was randomly allocated to a RUSITEC fermenter in each unit. Prior to the experiment, barley straw, barley silage, and seaweeds were dried at 55 °C for 48 h and then ground through a 4 mm screen using a Wiley Mill (standard model 4; Arthur H. Thomas Co., Philadelphia, PA, USA). A total of 10 g DM was included in nylon bags (R1020; 10 × 20 cm; 50 ± 10 µM porosity; 107 R1020, ANKOM Technology, Macedon, NY, USA) for incubation within the two RUSITEC units.

Rumen inoculum was obtained from three ruminally cannulated beef heifers previously adapted to a barley silage and barley straw-based diet. Donor heifers used in this experiment were cared for in accordance with the guidelines of the Canadian Council on Animal Care (2009) and were approved by the Institutional Animal Care and Use Committee (ACC2304). The rumen contents were collected 2 h pre-feeding from four different sites within the rumen and squeezed through PECAP mesh (mesh size 250 µm; PA66CG-250 136 cm, Sefar Nytal, Gilbert Saguenay, QC, CA). Both the solid and liquid

proportions were pooled in equal proportions from each heifer, and transported to the laboratory in an insulated thermos kept at 39 °C. All fermenters were maintained at 39 °C in a water bath before filling with 180 mL of pre-warmed McDougall's buffer (McDougall, 1948), and 720 mL of double strained rumen fluid added under a stream of CO_2 to maintain anaerobic conditions. One R1020 Ankom bag filled with 20 g (wet weight) of mixed rumen solids, and one bag with the dietary treatment were added to each fermenter. The fermenters were then fitted within each RUSITEC unit contained within a circulating water bath retained at 39 °C. After 24 h, the bag containing the rumen solids was replaced with a dietary treatment bag, such that each day the bag incubated for 48 h was replaced with a new bag. Artificial saliva was continuously infused into fermenters (26 mL h^{-1}) using a peristaltic pump set to achieve a dilution rate of 2.9% h^{-1}. The effluent and gas were collected within 2 L Erlenmeyer flasks, and 4 L reusable gas tight collection bags Curity®; Conviden Ltd., Mansfield, MA, USA), respectively, connected through a closed tubing system. The RUSITEC experimental period lasted for 15 days, with day 1-7 used for adaptation and day 8-15 used for sampling and measurements. The DM, OM, CP, and NDF degradability, and total VFA concentration was determined from day 8-15 and analysis was conducted as previously described [270]. The capability of the RUSITEC microbial communities to interact with *S. latissima* polysaccharides was investigated in vivo using FLA-SLAT, as described in **Text A1.2**.

2.6.5 Microbial sampling and DNA extraction

Samples for microbial analysis were collected from both the RUSITEC experiment and the *in vivo* feeding trial. These samples were subsequently used for 16S rRNA and shotgun metagenomics.

From each RUSITEC fermenter, 2 mL of liquor was collected at day 0, 1, 8 and 15. Samples were centrifuged at 20,000 × *g* for 20 min and supernatants were removed. Pellets were suspended in 1 mL of DNA/RNA Shield™ (Zymo Research, USA) before genomic extractions were performed using the DNeasy PowerSoil Pro kit (Qiagen, Canada) following the manufacturers protocols. DNA was stored at -80 °C until analysis.

In the *in vivo* feeding trial, rumen samples were collected at the end point of the experiment period. Here, all animals were slaughtered at a commercial slaughterhouse

(Rudshøgda, Nortura SA, Norway) and their intact gastrointestinal tracts were directly moved to a working bench. The stomach was opened, and the reticulo-rumen content was hand-mixed prior to sampling. The mixed sample was transferred into a sterile strainer blender bag (0.50 mm pore size; Stomacher® 400 Seward BA 6041, Worthing, UK) and gently squeezed to separate the fluid and particle phases. The phases were then sampled into cryotubes, transported in liquid nitrogen to the laboratory and stored at − 80 °C until analysis. Upon microbiome analysis, samples from both fluid and particle phase were thawed on ice and homogenized by vortexing prior cell lysis and DNA extraction. For cell lysis, 0.25 g homogenized biomass was weighted out for bead-beating using a FastPrep-24 Homogenizer (MP Biomedicals LLC., Ohio, USA) at 4 m s^{-1} for 45 seconds. DNA were extracted using DNeasy PowerLyzer PowerSoil kit (Qiagen, Hilden, Germany) following to the manufacturers protocol. The extracted DNA were quantified using a Qubit Fluorimeter and the Qubit dsDNA BR Assay Kit (ThermoFisher Scientific, Waltham, MA, USA) and stored at -80 °C until further use.

2.6.6 16S rRNA gene sequencing

In the RUSITEC experiment, the purified DNA samples were sent to Génome Québec (Montréal, QC, Canada) for Illumina MiSeq PE250 16S rRNA gene sequencing using the primers: 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') targeting the V4 region of the 16S rRNA gene. Paired end reads were quality trimmed using Trimmomatic [271] with a sliding window of 5:20, and adapters were automatically removed. Trimmed fasta files were merged and classified using Kraken 2 [272] according to the Silva 138.1 SSU database. Bracken [273] was used to estimate bacterial and archaeal abundance at the genus level.

Liquid and particle phase samples from all 24 lambs was collected for the amplification of the 16S rRNA gene. The 16S rRNA gene was PCR-amplified using Pro341F/Pro805R primer pair [274], which targets the V3-V4 region of both bacteria and archaea. The 25 µL PCR reactions consisted of 1× iProof High-Fidelity Master Mix (Bio-Rad, Hercules, CA, USA), 2.5 µL each primer, and 7.5 µL template DNA. PCR thermal cycling began with a hot start step at 98 °C for 3 minutes, and they were followed by 25 cycles of 98 °C denaturation for 30 s, 53 °C annealing for 30 s and 72 °C extension for 30

s, followed by a final extension step at 72 °C. The amplicon libraries were purified with AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA) and indexed with the Nextera XT Index Kit v2 (Illumina, San Diego, CA, USA). Equimolar concentrations of the libraries were pooled together and further purified with AMPure XP beads. The pooled library was quantified using Qubit Fluorometer, diluted and denatured before sequenced on an Illumina MiSeq sequencing system using 2 x 300 bp cycle runs with the MiSeq reagent kit v3. The generated 16S rRNA gene sequences were processed using the DADA2 pipeline [275] in R v4.3.2 [276] which included quality control, elimination of primers and adapters. The sequences were clustered into Amplicon Sequence Variance (ASVs) and the ASV identified were taxonomically assigned using SILVA Database release 138.1 [277]. Downstream analysis of lamb and RUSITEC communities are described in detail in **Text A1.2**.

2.6.7 Metagenomics

To reconstruct a core rumen microbial genome catalogue from each experimental group of lambs, extracted DNA from corresponding fluid and particle from four animals in each diet group (in total 24 samples) were subjected for metagenome sequencing. The selection of these representative samples was guided by diversity and abundance metrics from the 16S rRNA gene analysis. Shotgun metagenomic sequencing was performed at the Norwegian sequencing center (Oslo, Norway), where the libraries were prepared using PCR-free TruSeq chemistry and sequenced on two lanes of the Illumina HiSeq6000 to generate 2 x 150 bp pair-ended reads. From the RUSITEC system, DNA from samples collected on experimental day 15 was sent for shotgun metagenomics at Génome Québec (Montréal, QC, Canada). These samples were sequenced on an Illumina NovaSeq 6000 platform, also generating 2 x 150 bp pair-ended reads.

2.6.7.1 Generation of MAGs

RUSITEC sequences were trimmed with Trimmomatic (v0.39) [271] and assembled using metaSPAdes (v3.13.0) [278]. Assembled contigs were binned using the MetaWRAP (v1.3.2) [279] binning and bin-refinement pipeline, using CONCOCT (v1.1.0) [280], Maxbin2 (v2.2.6) [281], and Metabat2 (v2.12.1) [282]. Bins were dereplicated using

dRep (v3.2.2) [283] and filtered for 50% completion and under 10% contamination. CheckM (v1.1.3) [284] was used to evaluate the quality, while MAG abundance was estimated using CoverM (v0.6.1) (https://github.com/wwood/CoverM). MAGs were annotated using dbCAN (v3.0.7) [151] and Bakta (v1.9.3) [285]. GTDB-Tk (v2.4.0) [286] was used to infer taxonomic classification.

All metagenome raw reads from lamb samples were quality filtered using Trimmomatic (v0.36) [271] in pair end mode before the trimmed reads were assembled into contigs. Both individual assemblies and co-assemblies of all samples were carried out metaSPAdes (v3.13.0) [278] and MegaHIT (v1.2.9) [287], respectively. Contigs from the single assembly were binned using VAMB (v3.0.2) [288], while the contigs from the co-assembly was binned using both MetaBAT2 (v2.12.1) [282] and MaxBin2 (v2.2.7) [281]. Collectively, these three binning strategies generated 1,532 bins across all samples. These bins were re-dereplicated at 99% ANI using dRep (v3.2.2) [283], which resulted in 291 dereplicated bins/MAGs, with completeness above 50% and contamination below 25%. This collection of MAGs were used as genome database for metagenome-centric metaproteome analysis, allowing to retain genetic information without compromising the sensitivity and false discovery rate estimation [289]. For biological interpretation, only MAGs with contamination < 10% were considered. CheckM (v1.1.3) [284] was used to evaluate the quality parameters of each MAG, while CoverM v0.6.1 in genome mode was used to estimate the MAG abundance in each sample. Taxonomically classification of the MAGs was assigned using GTDB-Tk (v2.4.0) [286], and function annotation carried out using dbCAN, PFAM, and KEGG, integrated in the DRAM (v1.2.4) [163] annotation tool.

2.6.7.2 AUL and AUC phylogenetic analysis

AULs from lamb and RUSITEC data were manually screened for SusC- and SusD-like pairings. Syntenic AULs were compared using Diamond BLASTp (v2.1.8) [290] and Clustal Omega [291], and alignments were visualized using RStudio and the gggenes package [292]. MAG species trees were generated using OrthoFinder (v2.5.5) [293]. PL6, PL17 CAZyme phylogeny was completed using SACCHARIS v2.0.0.dev19 [154]. Rstudio packages ggdist [240], gghalves [294], and ggrepel [295] was used to generate GC content plots.

In total 75 genomes (**Table A1.4**) from marine environments and terrestrial gastrointestinal tracts, including AUL and AUC containing MAGs from this study, were selected based on presence of AULs (PULDB [150]) or PL6 and PL17 family members. In addition, NCBI reference strains without alginate-targeting CAZymes were included in the analysis. All genomes were run through OrthoFinder to generate a species tree (**Figure A1.13**). Additionally, genomes were analysed using dbCAN to identify putative PL6 and PL17 members, and SACCHARIS to identify ortholog clustering in the case of PL6 family members. Gene trees were generated by first generated alignments with MUSCLE (v5.1) [166], followed by trimming poorly aligned regions with trimAI (v1.5.0) [296] using the -automated1 option. The best-fit substitution model was determined with ModelTest—NG (v0.1.7) [297], and phylogenetic trees were subsequently generated with RAxML-NG (v1.2.2) [298]. NOTUNG v3.0-beta [239] was used to generate gene species tree alignments using default cost parameters, while RecPhyloXML [299] was used to generate a summary of species-gene tree reconciliations. All gene reconciliations are shown in **Figure A1.14.**

2.6.8 Metaproteomics

2.6.8.1 Protein extraction and sequencing

Sample preparation for metaproteomic analysis was carried out using the particle phase of rumen samples collected from all lambs in the control and 5% *S. latissima* experimental group. Lysis buffer (10 mM DTT, 100 mM Tris-HCl (pH=8) and 4% SDS) and 4 mm glass beads (≤ 160 μm) were added to 500 μL of rumen samples, followed by brief mix and resting on ice for 30 minutes. Mechanical cell lyses was performed using FastPrep-24 Classic Grinder (MP Biomedical, Ohio, USA) for 3 × 45 s at 6.5 m s^{-1}. Samples were then centrifuged at 16,000 × *g* for 15 minutes at 4 °C and lysate was transferred to a new tube, followed by a clean-up using Wessel-Flügge precipitation [300]. The precipitated proteins were dissolved in SDS-PAGE buffer, heated in the water bath for 5 min at 95 °C and ran on SDS-PAGE Any-kD Mini-PROTEAN TGX Stain-Free gels (Bio-Rad, California, USA) for 3 minutes. The gel was stained using Coomassie Blue R-250 and visual protein bands were carefully cut into 1×1 mm pieces. After washing the gel

pieces, reduction of disulfides was carried out by incubation proteins in 10 mM DTT for 30 min at 56 °C, followed by carbamidomethylation by incubation with 55 mM iodoacetamide for 30 min in the dark, at room temperature. Subsequently, proteins were digested into peptides using trypsin. Peptides were concentrated and eluted using C18 stage tips, before dried on a SpeedVac, re-suspended in 0.1% formic acid, and quantified using a Nanodrop One instrument. Finally, the peptide samples were processed using a nano LC-MS/MS coupled to a timTOF Pro mass spectrometer (Bruker, Germany), as described in detail in **Text A1.2**.

2.6.8.2 (Meta)genome-centric metaproteomics data analysis

The MS raw data were analyzed using FragPipe v16.3, with the search engine MSFragger v3.3 [301] embedded, in addition to Philospopher v4.0.0 [302]. Here, a protein sequence catalogue retrieved from the 291 recovered MAGs, viral scaffolds (described in **Text A1.2**), and publicly available rumen eukaryote genomes previously described in Andersen, Altshuler [24], were used as a sample specific database. Common contaminants such as human keratin, bovine serum albumin and host genome (*Ovis Aries*) were included in the database, as well as decoy sequence entries based on the reverse protein sequences. Carbamidomethylation was used as fixed modification, while oxidation of methionine and protein N-terminal acetylation were added as variable modifications. Trypsin was used as digestive enzyme, with maximum one missed cleavage allowed. For LFQ, IonQuant was applied with FDR-controlled match-between-runs (MBR) enabled, followed by global median normalization of intensities across all runs [303]. Protein groups detected from FragPipe were uploaded to Perseus v1.6.15.0 [304], for further downstream analyses. First, protein groups identified as contaminants and host proteins were removed, and the intensities were Log2(x) transformed. Furthermore, a protein group needed to be detected in minimum three of eight biological replicates in at least one of the experimental diet groups (control and/or 5% inclusion of *S. latissima*) to be considered. For statistical analysis in Perseus, missing values were imputed using normal distribution-based approach with a width of 0.3 and a downshift of 2.8 relative to the distribution of measured intensities. Protein groups with significant changes in LFQ intensities were identified using a two-sided Student's T-test (p-value < 0.05). To account for differences in organism

abundance, organism-level normalization was performed for target enzymes. For each MAG and sample, the linear LFQ intensities of the target protein was divided by the total LFQ intensities for all proteins assigned to that MAG. Normalized values are provided in **Extended Tables A1.2**.

2.6.9 Biochemical analysis of PLs

PL6 enzymes from *Bacteroidota* species alginate AULs, and an outgroup OTU, *Brevundimonas Bullata* that does not have a traditional SusC/D-like system, were selected for their taxonomic and phylogenetic diversity within the PL6 family. Selected PL6 genes were trimmed to remove the signal peptide and synthesized into pET28a with a C-terminal His_6-tag and codon optimized for *E. coli* production (Biobasic). Expression plasmids were transformed into BL21 (DE3) Tuner competent cells (Novagen) and grown to an OD_{600nm} 0.6-0.8 in LB Miller broth containing 50 μg mL^{-1} kanamycin before induction with 1 mM IPTG at 18 °C overnight. Induced cultures were pelleted at 6,500 × *g* for 20 min and lysed using a combination of lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 0.1 mg mL^{-1} lysozyme) and sonication (2 min of 1 s intervals of medium intensity sonic pulses at a power setting of 4.5 – Heat systems Ultrasonics Model W-225). Cell lysate was centrifuge at 17,500 × *g* for 1 h and filtrate was purified using Ni-NTA resin columns (Cytiva) and immobilized metal affinity chromatography. Recombinant protein was eluted with an imidazole gradient in 20 mM Tris, pH 8. Recombinant PL6 enzymes were concentrated and dialyzed into 20 mM Tris pH 8.0, 500 mM NaCl, and 2% glycerol using 30 kDa cut-off Amicon Ultra centrifugal concentrators (Millipore Sigma).

50 nM of recombinant PL6 enzymes was incubated with 5 mg mL^{-1} brown seaweed alginate (Sigma), mannuronate oligosaccharides DP20-35 (Elicityl) and guluronate oligosaccharides DP25-45 (Elicityl) in 20 mM of varying buffers to determine pH optima at 37 °C: citrate-phosphate (pH 4-8.5), CHES (pH 8.5-10), Glycine (pH 8.5-11), and CAPs (pH 10-11). PL6 β-elimination reaction and the generation of C4-C5 unsaturation can be monitored with absorbance at 232 nm. Reactions were run for 10 minutes while continuously monitoring absorbance at 232 nm at 30 s intervals using a SpectraMax ID3 plate reader. Initial velocities of PL6 enzyme activity on brown seaweed alginates were determined using eight concentrations of substrate at the optimal pH for each enzyme.

Rates of product formation were calculated in GraphPad Prism (9.0.2) using the extinction coefficient 6,150 M^{-1} cm^{-1} to convert absorbance to product concentration [305]. Due to substrate complexity in molecular weight and M/G ratio and organization, velocities were unable to be fit to the Michaelis Menten equation.

Reactions were ethanol precipitated by incubating in 50% EtOH overnight at -20 °C before centrifugation at 12,000 × *g* for 10 min and retrieving the supernatant to remove larger, undigested alginate polymers. Samples were dried on a SpeedVac to remove EtOH and resuspended in ultrapure 18 MΩ cm^{-1} H_2O. Digests were run on TLC using a mobile phase of 2:1:1 butanol: acetic acid: H_2O, and visualized using an orcinol solution (70:3 acetic acid: sulfuric acid, and 1% orcinol). HPAEC-PAD was performed on a Dionex ICS-3000 using a 3 x 150 mm CarboPac PA20 column (Thermo Scientific). Digests were eluted at a 0.35 mL min^{-1} flowrate at 30 °C in a background of 100 mM NaOH in an increase sodium acetate gradient (0 to 20 min, 0 to 0 M; 20 to 80 min, 0 to 1 M).

Liquid chromatography was performed on a Vanquish ultra-high performance liquid chromatography (UHPLC) system (Thermo Scientific). Alginate oligosaccharide samples were prepared in water and injected in a volume of 10 µL at a concentration of 200 µg/mL. Separation of the digested oligosaccharides was achieved using an Acquity UPLC BEH Amide (HILIC) Column, 130Å, 1.7 µm, 2.1 mm X 150 mm (Waters) at a flow rate of 300 µL min^{-1} at 30 °C, using a gradient as shown in **Table A1.6** [306]. Electrospray ionization mass spectrometry (ESI-MS) was performed on an Orbitrap Fusion Tribrid system (Thermo Scientific) in negative ion mode. Mass spectra parameters are shown in **Table A1.7**. To select ions for MS2 experiments, an intensity threshold filter was employed with a minimum intensity of 25,000 and maximum intensity of 1E+20, and a dynamic exclusion filter was used after 1 time for 2.5 s with default mass tolerance values. Higher-energy collisional dissociation (HCD) was employed to generate fragments. MS spectra were analyzed using Xcalibur and Freestyle software packages (Thermo Scientific), and product ions were identified based on both manual interpretation and comparison to previous studies [252].

2.7 Data availability

Raw metagenomic sequencing data have been deposited to the Sequence Read Archive (SRA) under the project accession numbers PRJEB83690 (lamb in vivo feeding experiment) and PRJNA1200888 (bovine RUSITEC experiment). The MAGs are publicly available via Figshare (lamb rumen MAGs doi: 10.6084/m9.figshare.28024343; bovine RUSITEC MAGs, doi: 10.6084/m9.figshare.28024394). The proteomics data, including the complete database, have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD059090.

2.8 Code availability

The codes generated in this study are available at Zenodo (doi:10.5281/zenodo.14515213).

2.9 Acknowledgements

We thank Cascadia Seaweed Corporation for the generous donation of *S. latissima*. Research at AAFC was supported through an AgriScience Project (ASP-207; J-002817) and the Prize for Outstanding Achievement in Science (AAFC) awarded to D.W.A. (J-003135). LC-ESI-MS data was collected at the University of Lethbridge Magnetic Resonance Facility with support from Tony Montina, Vincent Weiler, Maurice Needham, and Carl Holland. L.H.H. and A.F. are grateful for support from The Research Council of Norway (project no: 302639). P.B.B. is grateful for support from the Novo Nordisk Foundation (0054575—SuPAcow) and the Australian Research Council (Future Fellowship: FT230100560). The lamb feeding experiment was financed by The Research Council of Norway; BIOTEK2021/Havbruk Biofeed (project no. 229003) and Foods of Norway a Centre for Research Based Innovation (project no. 237841). The sequencing service of rumen samples was provided by the Norwegian Sequencing Centre (www. sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the 'Functional Genomics' and 'Infrastructure' programs of the Research Council of Norway and the Southeastern Regional Health Authorities. The authors acknowledge the Orion High Performance Computing Center at the Norwegian University

of Life Sciences and Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway for providing computational resources that have contributed to meta-omics computations reported in this paper. T.R.P. was supported by the Canada Research Chair program. Mass spectrometry-based proteomic analyses were performed by the MS and Proteomics Core Facility, Norwegian University of Life Sciences (NMBU). This facility is a member of the National Network of Advanced Proteomics Infrastructure (NAPI), which is funded by the Research Council of Norway INFRASTRUKTUR-program (project number: 295910). We are thankful for the generous access provided by the Molecular Ecology Department of the Max Planck Institute for Marine Microbiology for microscopy. G.R. was supported through funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; project number 496342779)

Chapter 3 Global distribution of microbial carrageenan foraging pathways reveals widespread latent traits within the genetic “dark matter” of ruminant intestinal microbiome

3.1 Preface

Chapter 3 is presented in manuscript form that is under revision with the journal Nature Communications (Tingley et al., 2025; [258]). This collaborative study examines the adaptation of cattle and non-cattle rumen microbiomes to red seaweed *M. japonica*. As first author, I participated in the writing and structuring of the manuscript, revision and editing process. My contributions to the manuscript include metagenomic analysis of faecal and rumen datasets, phylogenetic analysis of carrageen active CAZymes, and characterization of carrageenan active CAZymes. The full author contribution list is presented below. My contributions are displayed in figures: 3.1A-C, 3.2, 3.3A, 3.4, 3.5; Appendix 2 figures: A2.1, A2.4, A2.5E&F, A2.6, A2.8 & A2.9; and Appendix tables A2.1-2.11. The citation style of this manuscript has been updated from the original publication to maintain continuity in this thesis.

Authors:

Jeffrey P. Tingley[1,2], Thea O. Andersen[3,4], Liam G. Mihalynuk[5], Xiaohui Xing[1], Kristin E. Low[1], Douglas P. Whiteside[6,7], Ianina Altshuler[3,8], Nic Jujihara[1], Anna Y. Shearer[1,2], Leeann Klassen[1] Spencer Serin[9], Edgar Smith[10], Greta Reintjes[11], Trushar R. Patel[2,12,13], Alisdair B. Boraston[5], Live H. Hagen[4], Phillip B. Pope[3,4,14], D. Wade Abbott[1,2,¶]

Affiliations:

[1] Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

[2] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

[3] Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway

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

[5] Department of Biochemistry and Microbiology, University of Victoria, Victoria British Columbia, Canada

[6] Faculty of Veterinary Medicine, University of Calgary

[7] Animal Health Department, Wilder Institute/Calgary Zoo
[8] MACE Laboratory, ALPOLE-ENAC, École Polytechnique Fédérale de Lausanne, Sion, Switzerland
[9] Spoitz Enterprises Inc, Courtenay British Columbia, Canada
[10] Beaver Meadow Farms, Comox, British Columbia, Canada
[11] Microbial-Carbohydrate Interactions Group, Faculty 2 Biology / Chemistry, University of Bremen, Bremen, Germany
[12] Alberta RNA Research and Training Institute and Department of Chemistry and Biochemistry, Lethbridge Alberta, Canada
[13] Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary Alberta, Canada
[14] Centre for Microbiome Research, School of Biomedical Sciences, Queensland University of Technology (QUT), Translational Research Institute, Woolloongabba, QLD, Australia

[†] To whom correspondence should be addressed

Author contributions:

D.W.A and J.P.T, conceived and designed the study. T.O.A, I.A., L.H.H., P.B.P, conducted proteomics analysis. X.X. performed linkage analysis on *M. japonica*. J.P.T., G.R., L.K., A.Y.S, S.S., E.S., and D.P.W collected and enriched samples for isolations and metagenomics. J.P.T. and G.R. conducted 16S analysis. J.P.T performed shotgun metagenomic analyses and metagenomic surveys. J.P.T performed growth curves and isolations. J.P.T, L.M., K.E.L., and N.J. designed GH16 constructs, purified protein, and characterized CAZymes. L.M. and A.B.B. conducted crystallography. G.R., L.K., and A.Y.S conducted FISH and FLA-*Mj*Ex on enrichments and isolates. D.W.A. and T.R.P., supervised the work. J.P.T, T.O.A, D.W.A, and P.B.P drafted the manuscript. All authors wrote, edited, and approved the manuscript.

3.2 Abstract

Seaweed represents a promising source of sustainable, alternative feeds for livestock, yet despite their increasing popularity, the dietary fate of seaweed polysaccharides, such as carrageenan, is unknown. Here, we applied functional microbiome analyses of ruminant gastrointestinal tract microbiomes to discover catabolic enzymes specific for carrageenan digestion from the red seaweed *M. japonica*. *M. japonica* preferentially increased *Bacteroides* abundance within the distal gut over the rumen, and bacterial isolates had capacity to use carrageenans as a sole carbon source. Carrageenan-active polysaccharide utilization loci (CarPULs) were identified, and recombinant enzymes were characterized to provide insights into pathway specialization of divergent CarPULs. Selective enrichment and metagenomic mining revealed that carrageenan catabolism is widespread among geographically and taxonomically distinct ruminants, suggesting it is a globally distributed latent trait within the order Ruminantia and carried within microbiome as part of the microbial "dark matter". These pathways are structurally distinct from those found in marine bacteria, highlighting a complex and ancient evolutionary history of CarPULs in ruminant microbiomes.

3.3 Introduction

With growing global concerns about greenhouse gas emissions, the need for sustainable agricultural practices continues to escalate. Enteric methane is a by-product of ruminant fermentation [307-309] and exerts a stronger warming effect than carbon dioxide despite possessing a shorter half-life [310]. Some seaweeds, such as *Asparagopsis taxiformis* can act as inhibitors of methane emissions when provided as a feed additive to cattle [43, 311, 312]. Seaweed has been utilized for millennia to supplement cattle and sheep diets in herds raised near coastlines and represents a viable, but underutilized alternative feed source [214, 225].

Ruminants rely on established syntrophic relationships within their GIT microbial community to digest and extract energy from fibre in their diets. Many of these microorganisms are endowed with extensive inventories of CAZymes that modify and depolymerize chemically and structurally complex polysaccharides [115, 313]. These include "keystone" endo-acting enzymes [314, 315], which are surface associated and

catalyze initial polysaccharide fragmentation; and sulfatases [134], which remove obstinate sulfates preventing saccharification. Among the dominant phyla in ruminant GITs, gram-negative *Bacteroidota* have been intensively studied as saccharolytic generalists known for their metabolic plasticity [134]. *Bacteroidota* possess CAZyme collections, organized into independently regulated PULs, to catabolize discrete polysaccharide substrates [136, 313]. PUL inventories within the genome determines the diversity of substrates consumed by individual strains [141], and therefore, acquisition or loss of PULs are important considerations for determining responses to shifts in diet or exposure to exotic polysaccharides, such as those found in seaweeds [202, 205]. Elucidating the processes driving PUL acquisition and their timelines can be difficult. For example, the xyloglucan PUL found in *Bacteroides* spp. lies between ancestral genes shared by many strains with no evidence of lateral transfer [316]. Other PULs display residual genetic signatures consistent with horizontal gene transfer events. These include yeast mannan PULs found in western human GIT microbiomes [188], and dietary seaweed polysaccharide PULs, found predominantly in east Asian human GIT microbiomes [205], which can be specific for porphyran [202], agarose [204], or carrageenan [205].

Dietary selection and extinction pressures can drastically influence which metabolic capabilities are retained within the microbiome [317]. *Bacteroidota* have the capacity to digest multiple complex polysaccharides, yet for how long strains and PULs persist when their cognate substrates are limiting amounts or absent is not known. Previously, we have shown that alginate PULs are functionally conserved in two different ruminant species that were naïve to dietary seaweed and raised on two different continents [209]. We have proposed these “latent traits” maintained in GIT microbiomes represent a common nutritional strategy exploited by ruminants and other herbivores to rapidly adapt to dietary changes. Therefore, understanding how bacterial strains persist and unused genetic repositories are maintained within the genetic dark matter [318] of the microbiome warrants further study.

To understand the process of polysaccharide consumption by host microorganisms, it is important to know the chemical structure of the polysaccharide [264], which microorganisms and metabolic pathways respond to the substrate [319], and the function of enzymes within associated catabolic pathways [114]. Further, disentangling the

evolutionary origins of PULs can aid in understanding the timelines of pathway acquisition and the processes driving their functional divergence. Here we elucidated how *M. japonica* cell wall carrageenans are consumed by *Bacteroidota* that were enriched in the distal GIT microbiomes of cattle. Selective enrichment and primary sequence BLAST analysis revealed diverse carrageenan PUL (CarPUL) structures with functional specialization within ruminant members of suborder Artiodactyla. The implications for how naïve cattle and other herbivorous artiodactyls adapt to dietary carrageenan, and the origin of latent traits responsible for their digestion, are considered.

3.4 Results

3.4.1 Degradation of carrageenans within cattle is facilitated within the lower GIT by *Bacteroides* harbouring carrageenan-degrading PULs

Two cattle trials were conducted using *M. japonica* supplemented diets. In the first trial, *M. japonica* was provided to cattle *ad libitum* on pasture; in the second, cattle were fed 5% *M. japonica* as a silage in covered pens. Using 16S rRNA gene sequencing, few compositional differences were observed within the rumen microbiome of animals in the *ad libitum* trial; however, there was a minor, yet significant, increase in *Bacteroidaceae* observed in the silage trial (**Figure A2.1A**). In contrast, the faecal communities in both the *ad libitum* and silage trials displayed a significant increase in *Bacteroidaceae* when the animals were fed seaweed (**Figure 3.1A**). This is highlighted by an increase in *Bacteroides*, at the apparent expense of the uncultured *Bacillota* lineage UCG-010 (**Figure 3.1B**). Overall, *M. japonica* supplementation selectively influences the faecal microbiome over the rumen microbiome (**Figure A2.1C**).

Metagenome-guided metaproteomics was conducted to determine if *Bacteroidaceae* proteins were overexpressed within cattle fed *M. japonica*. A total of 244 MAGs with over 70% completion and under 10% contamination were generated within this study (**Figure 3.1C**), including 14 MAGs from *Bacteroidaceae* (**Table A2.1**). A *B. xylanisolvens* MAG (99.5% completion; 0.7% contamination), referred to as *Bx*MAG$_{BOV,}$ was identified in multiple silage-fed faecal samples and contained predicted CAZymes from families known to be active on carrageenan, including GH16 subfamily 17 members (GH16_17), GH82, GH150, GH167 and sulfatases. These CAZymes were organized into

a PUL denoted as *Bx*MAG$_{BOV}$ CarPUL. This MAG (and high identity reads mapped towards CarPULs) were only found in samples from cattle fed seaweed (**Table A2.2**).

Metaproteomics analysis revealed the largest differences in faecal samples from cattle fed 5% *M. japonica*, with more total proteins and higher overall detection levels as determined by summed LFQ intensity (**Figure A2.2A**). We found 771 proteins which showed significantly different ($p < 0.05$) protein detection (**Figure 3.1D**); of these, 733 had significantly higher protein detection in the faecal samples collected from cattle fed the seaweed silage diet compared to the control group. Identification of MAGs that were attributed to the significantly higher detected proteins in cattle fed *M. japonica*, revealed that *Bx*MAG$_{BOV}$ accounted for 45% of all proteins (n=331; **Figure 3.1D**). Also, *Bx*MAG$_{BOV}$ was among the most highly abundant populations in our analysis and exhibited a 27.7-fold increase in summed protein detected in faecal samples from cows fed the seaweed silage diet (**Figure A2.2B**). Of the total 4,420 unique protein groups recovered by our metaproteomic analysis, 549 proteins were taxonomically assigned to *Bx*MAG$_{BOV}$ (**Figure A2.2C**). Our metaproteomic analysis revealed that *Bx*MAG$_{BOV}$ CarPUL was almost exclusively detected in the faecal samples of cows fed *M. japonica* (**Figure 3.1E**).

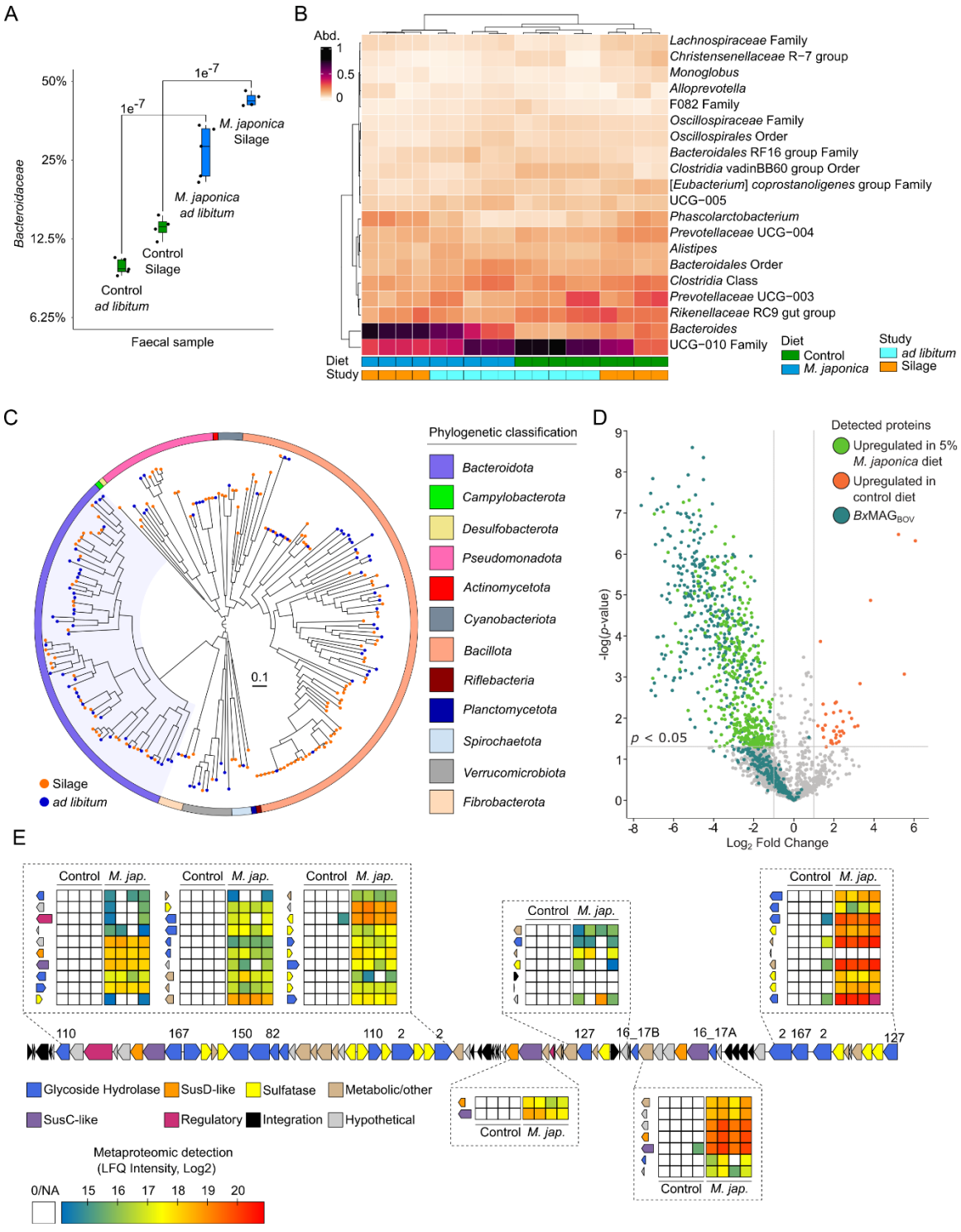

Figure 3.1: *M. japonica* supplementation increases *Bacteroides* within the faecal microbiota. A) Percentage abundance change of *Bacteroidaceae* between diet and study faecal samples. A linear regression model of % composition was generated using microViz [320], and a Tukey test used to measure significance (adjusted p-value 0.5) [276]. **B)** 16S rRNA gene relative abundance of genus' above 10,000 reads within faecal samples. Diet (Control: n=9 and *M. japonica*: n=9 supplemented) and Study (ad libitum: n=10 and Silage:

n=8) are denoted below the heatmap. **C)** MAGs generated and dereplicated between faecal and rumen metagenomes. Rumen MAGs are denoted with an orange circle and faecal with blue. **D)** Volcano plot showing proteins with significantly different protein expression recovered from faecal samples from cattle fed diet with 5% inclusion level of *M. japonica* vs those fed the control diet. Difference in protein expression is displayed as Log2 fold change in LFQ intensities. Proteins are colored based on their presence within *Bx*MAG$_{BOV}$ genome and the *Bx*MAG$_{BOV}$ CarPUL. Critical and highly expressed proteins are further labelled. **E)** *Bx*MAG$_{BOV}$ CarPUL detected within *M. japonica* supplemented cattle. Heatmaps indicate LFQ intensity for each *Bx*MAG$_{BOV}$ CarPUL gene for all four replicates of faecal samples of cattle fed control diet and diet supplemented with 5% *M. japonica* in the silage study. Genes that do not have a heatmap present were not identified within proteomes.

3.4.2 Rumen-associated *Bacteroides* strains display variation in CarPUL structures

Rumen samples collected from cattle in this study were enriched in minimalized media and either de-starched *M. japonica* water-soluble extract (*Mj*Ex) or commercially available carrageenans (κ, ι, and λ), allowing for isolation of cattle-derived bacterial species capable of growth on carrageenans. Five *Bacteroides* spp. isolates were successfully enriched on ι-carrageenan and *Mj*Ex, of which two were selected for further long-read genome sequencing, identification, and characterization based on sequencing quality: *Bx*Car5$_{BOV}$ and *Bx*Car17$_{BOV}$, isolated on ι-carrageenan and *Mj*Ex, respectively. *Bx*Car5$_{BOV}$ grew to a higher density than *Bx*Car17$_{BOV}$ on all commercial carrageenans (ι, λ, and κ) and the *Mj*Ex (**Figure 3.2A & 3.2B**), despite distorted optical densities from κ-carrageenan viscosity. *Bx*Car17$_{BOV}$ did not have noticeable growth on ι-carrageenan, but did grow on other carrageenan sources (λ, κ, and *Mj*Ex). However, growth on κ-carrageenan and *Mj*Ex was observed after a 12 h lag period. The selective uptake of *Mj*Ex by *Bx*CAR isolates was directly visualized using fluorescently labelled *Mj*Ex (FLA-*Mj*Ex). Both strains were found to internalize FLA-*Mj*Ex, further supporting that these strains can indeed consume polysaccharides from *M. japonica* (**Figure 3.2A; Figure A2.3**). FLA-*Mj*Ex import is consistent with a selfish-mode of foraging [200, 249], and higher in *Bx*Car5$_{BOV}$ than in *Bx*Car17$_{BOV}$ after 24 h, with 62% and 38% of cells showing staining respectively.

*Bx*Car17$_{BOV}$ possesses a single short CarPUL, while *Bx*Car5$_{BOV}$ possesses two CarPULs. One is highly similar to the *Bx*MAG$_{BOV}$ CarPUL (*Bx*Car5$_{BOV}$ CarPUL-1), sharing 99-100% identity between sequences (**Table A2.3**); the second is shorter and

unique to *Bx*Car5$_{BOV}$ (*Bx*Car5$_{BOV}$ CarPUL-2) (**Figure 3.2C**). Bovine CarPULs contain homologous CAZyme and sulfatase families, differing only in the number of genes present. For example, only *Bx*Car17$_{BOV}$ contains a single GH16_17; whereas *Bx*MAG$_{BOV}$ and *Bx*Car5$_{BOV}$ PULs contain two GH16_17s, all sharing between 40-90% identity. Likewise, CarPULs contain one or more copies of sulfatase subfamilies 1_8, 1_16, 1_30, and 1_81 (>80% shared identity; **Table A2.3**). The exception is *Bx*Car5$_{BOV}$ CarPUL-2, which is missing copies of S1_15 and S1_20.

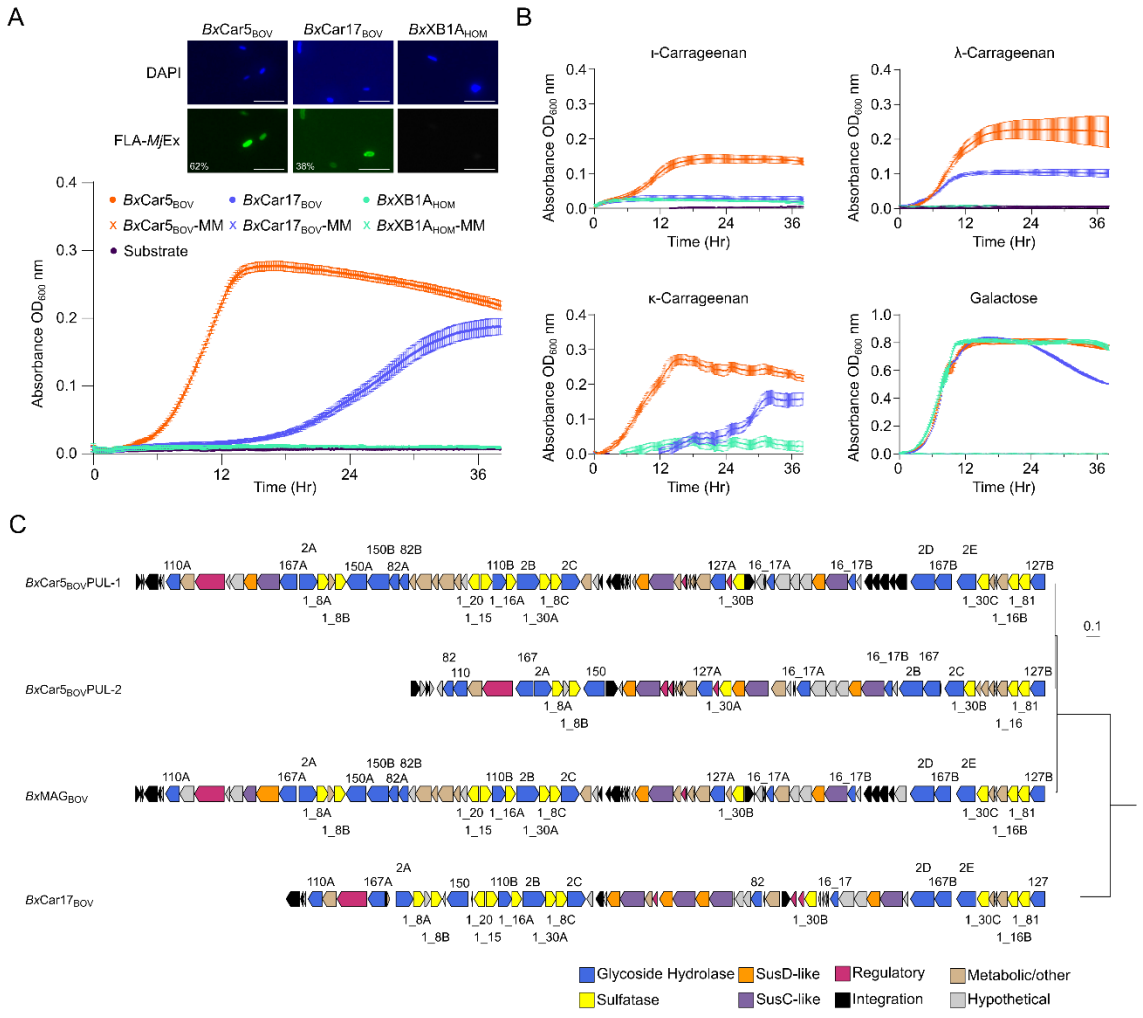

Figure 3.2: *B. xylanisolvens* enriched from *M. japonica* supplemented cattle contain variations on the *Bx*Car$_{BOV}$ PUL. A) Top: Isolates were incubated with 0.2% FLA-*Mj*Ex for 1d. Samples were co-stained with DAPI and imaged using an epifluorescence microscope (LED light cubes DAPI (EX: 385/30 EM: 450/50 DM: 425, ED light cubes FLA-MjEx (EX: 470/40 EM: 525/50 DM: 495). Isolates were primed for 24 h on *Mj*Ex.

Scale bar = 5 μm. Values in bottom left corner represent the percentage of cells showing FLA-*Mj*Ex uptake. **Bottom:** *B xylanisolvens* isolates and control *B. xylanisolvens* XB1A grown on 0.3% *Mj*Ex. **B)** *Bacteroides* isolates grown on 0.3% commercial carrageenans. OD_{600} nm was observed every 10 min. Negative OD values were excluded from the plots. **C)** Conservation of *Bx*Car5$_{BOV}$ CarPUL-1 & CarPul-2, *Bx*Car17$_{BOV}$ and *Bx*MAG$_{BOV}$ CarPULs aligned via OrthoFinder species tree.

3.4.3 Phylogenetically distinct GH16_17 members have differing carrageenan specificities

Keystone CAZymes play critical roles at the surface of microorganisms, and differences in their specificity or catalytic efficiency may dictate saccharification cascades of carrageenan subtypes and growth propensities. To elucidate functional differences in keystone GH16_17 members, primary sequences were aligned and visualized using SACCHARIS. Three distinct clades were observed (**Figure 3.3A**). *Bx*Car17$_{BOV}$ GH16_7A shares homology to a previously characterized λ-carrageenase from *B. thetaiotaomicron* 3731 (*Bt*3731$_{HOM}$) [205]; whereas *Bx*MAG$_{BOV}$ and *Bx*Car5$_{BOV}$ GH16_17 enzymes partitioned with members from *Bacteroides ovatus* CL02T12C04 (*Bo*12C04$_{HOM}$). Notably, catalytic residues were conserved in all sequences (**Figure A2.4A**). To determine if clades were associated with unique carrageenan subtype specificities, *Bx*MAG$_{BOV}$ GH16_17A and GH16_17B were selected for biochemical characterization.

The X-ray crystal structure of *Bx*MAG$_{BOV}$ GH16_17A was solved to 2.0 Å resolution (PDB ID: 9EFL). Superimposition with the GH16_17 *Pseudoalteromonas carrageenovora* CgkA in complex with a κ-carrageenan oligosaccharide [321], revealed conservation of catalytic residues in the -1 and -2 subsites. A discriminating feature was the substitution of G258 in CgkA with E280 of *Bx*MAG$_{BOV}$ GH16_17A. This results in elimination of a pocket responsible for 4S-Gal accommodation and introduces a steric clash with the active site surface (**Figure 3.3B**). This structural determinant is functionally conserved in the GH16_13 κ/β-carrageenase from *Wenyingzhuangia fucanilytica* [322], and unequivocally selects for unsulfated galactose in the -1 subsite (**Figure 3.3B**). To compare clade-dependent specificity, an AlphaFold 3 [323] generated model of *Bx*MAG$_{BOV}$ GH16_17B displayed the same active site architecture as CgkA complete with the 4S-binding pocket (**Figure 3.3C**). Therefore, we hypothesized that *Bx*MAG$_{BOV}$

GH16_17A may have activity on partially desulfated/hybrid carrageenans, similar to the GH16_13 enzymes *W. fucanilytica* and *Pseudoalteromonas fuliginea* PS47 (*Pf*GH16B). Specifically, the activity of *Pf*GH16B required prior removal of 4S from κ/ι-carrageenan for hydrolysis of the backbone [324]. Using FACE to detect the products of ι-carrageenan hydrolysis by *Bx*MAG$_{BOV}$ GH16_17A, we found that native ι-carrageenan was not hydrolyzed by *Pf*GH16B and was a poor substrate for *Bx*MAG$_{BOV}$ GH16_17A (**Figure 3.3D**). However, pretreatment of the carrageenan with a endo-4S-sulfatase, *Pf*S1_19B [134], rendered it an improved substrate for *Bx*MAG$_{BOV}$ GH16_17A, producing the same product profile as *Pf*GH16B. These results support the assignment of *Bx*MAG$_{BOV}$ GH16_17A as a carrageenase active on hybrid substrates with full or partial removal of the 4-sulfate groups.

In contrast, *Bx*MAG$_{BOV}$ GH16_17B demonstrated endo-activity on *Mj*Ex and all commercially available forms of carrageenan, including κ- and ι-carrageenan and κ/ι-carrageenan hybrids (**Figure 3.3E-F; Figure A2.5A-D**). Although digestion of λ-carrageenan was observed *via* TLC (**Figure A2.5E**), these products could not be confirmed by LC-ESI-MS and any products resulting from this commercial substrate were exclusively κ- and ι-carrageenan oligosaccharides. This supports that *Bx*MAG$_{BOV}$ GH16_17B is an endo-κ/ι-carrageenase. Treating *Mj*Ex with *Bx*MAG$_{BOV}$ GH16_17B also improved the growth of the *Bx*Car17$_{BOV}$ isolate, which lacks a homologous enzyme (**Figure A2.5F**), suggesting it has a biological role as a keystone enzyme in κ/ι-carrageenan utilization.

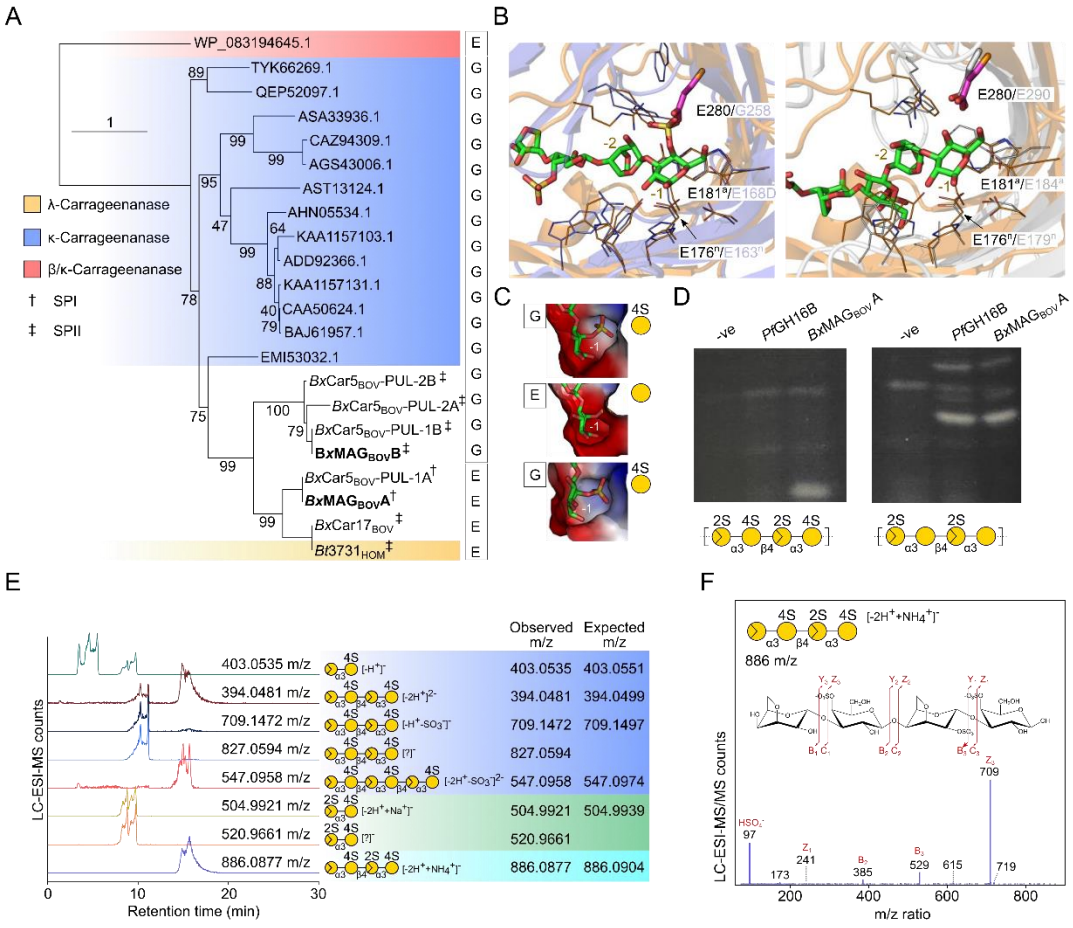

Figure 3.3: *Bx*MAG GH16_17 form functionally distinct clades to saccharify carrageenan subtypes. A) SACCHARIS v2.0 generated phylogeny of GH16_17 members from *Bx*MAG$_{BOV}$, and *Bx*Car5$_{BOV}$ and *Bx*Car5$_{BOV}$17 within this study aligned to characterized GH16_17 members from the CAZy database (www.cazy.org). Signal peptides of *Bx*Car enzymes are denoted by † (SPI) and ‡.(SPII) **B)** Structural analysis of *Bx*MAG$_{Bov}$ GH16A. **Left:** Overlay of the structure of *Bx*MAG$_{Bov}$ GH16A (orange) with the GH16_17 k-carrageenase from *Pseudoalteromonas carrageenovora* (blue; PDB ID 5OCQ). Key residues within only the -1 and -2 subsites are shown as sticks with the residues acting as nucleophile or acid/base labelled and indicated with superscript "n" or "a", respectively. The substitution in *Bx*MAG$_{Bov}$ GH16A blocking accommodation of the 4S group is shown as a magenta stick. **Right:** Overlay of the structure of *Bx*MAG$_{Bov}$ GH16A (orange) with the GH16_13 β/κ-carrageenase from *Wenyingzhuangia fucanilytica* (grey; PDB ID 9INT). **C)** Surface representations of the -1 subsites of 5OCQ (top), *Bx*MAG$_{Bov}$ GH16A (middle), and an AlphaFold 3 [323] generated model of *Bx*MAG$_{BOV}$ GH16B, respectively. The ligand in 5OCQ is shown; ligands in *Bx*MAG$_{Bov}$ GH16A and the *Bx*MAG$_{Bov}$ GH16B model are derived from overlays with 5OCQ. **D)** The activity of *Pf*GH16B (an α/β-carrageenan specific GH16_13) and *Bx*MAG$_{Bov}$ GH16A on ι-carrageenan **(Left)** and ι-carrageenan pretreated with *Pf*S1_19B to remove 4S modifications **(Right)** as examined by fluorophore-assisted carbohydrate electrophoresis.

E) *Bx*MAG$_{BOV}$ GH16B digests of *Mj*Ex were analyzed by LC-ESI-MS, and example extracted ion chromatograms are shown. **F)** ESI-MS/MS with HCD was performed in order to identify the extracted ions. The MS2 product ion spectra shown was identified neo-κ/ι-hybrid tetrasaccharide as majority enzymatic product from the *Mj*Ex. Monosaccharide symbols are displayed according to the Symbol Nomenclature for Glycans system [55].

3.4.4 CarPULs in other herbivorous artiodactyls

CarPULs have previously been identified within human and captive great ape GIT microbiomes [325], which showed as high as 100% sequence identity towards the bovine sourced CarPUL sequences (**Figure 3.4A; Table A2.4**). CAZyme BLAST hits for GH16_17 members performed here highlight that carrageenan utilization is widespread amongst geographically distinct human GIT datasets (**Figure 3.4A**); however, no ruminant homologs were identified during our database analysis. Therefore, to investigate if CarPULs were present in other ruminant microbiomes, or if introduction into bovines was a rare event, we analysed available artiodactyl metagenomic datasets. Approximately 17 Tb of shotgun metagenomic read datasets from buffalo, cattle, deer, goat, moose, sheep, and yak collected from public databases (**Table A2.5**) were searched against CarPUL protein sequences using BLAST to identify high similarity targets. Sequence similarity varied throughout the CarPULs and hosts; however, many regions demonstrated high conservation (>90%) in the keystone GH16_17 enzymes. Reads from buffalo, cattle, and deer metagenomes contained frequent hits with 90-100% sequence identity towards GH16_17 members within *Bx*MAG CarPUL and *Bx*Car5 CarPUL-1. Buffalo metagenomic read sets contained the highest identity hits towards all CarPULs, sharing homology with other carrageenan active families GH82 and GH150 (**Table A2.6**). Metagenomic datasets from Chinese Water Buffalo microbiomes [326] displayed sequences with 100% identity to GH16_17 members and were selected for co-assembly using MegaHIT to assemble full-length GH16_17 members (**Table A2.7**). Open reading frames originating from this assembly yielded multiple sequences (Meta$_{BUB}$) that were closely related to CarPUL GH16_17 enzymes identified in this study.

Faecal samples from herbivorous artiodactyls housed at the Wilder Institute/Calgary Zoo (**Table A2.8**) were used for culture-enriched metagenomics and selective isolations with *Mj*Ex. Culture-enriched metagenomics identified carrageenan-

active CAZyme members within musk deer and giraffe, which was further supported through isolation of a *B. zhangwenhongii* strain from giraffe (*Bz*Car$_{GIR}$) and *B. xylanisolvens* isolated from musk deer (*Bx*Car$_{MOS}$). The CarPULs and CAZymes from these hosts showed high synteny and sequence relatedness with CarPULs found in cattle from this study and human-associated strains (**Figure 3.4B**). Interestingly, the giraffe-derived *B. zhangwenhongii* CarPUL demonstrated near 100% identity with the *Bt*3731 CarPUL (**Table A2.4**).

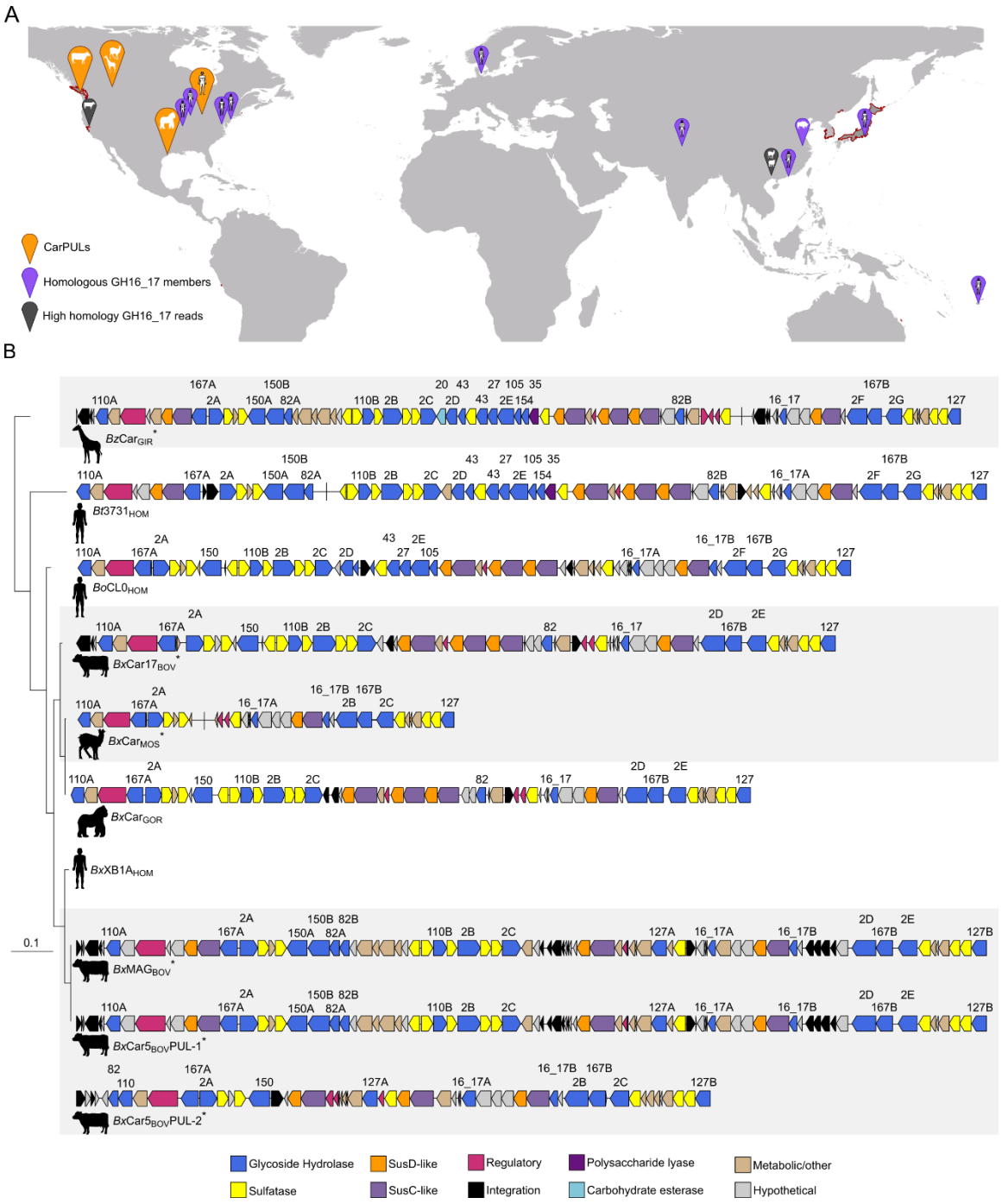

Figure 3.4: *Bacteroides* carrageenan PULs and CAZymes display homology between interspecies hosts and geographic regions. A) CarPULs and orphaned CarPUL genes identified within mammalian GIT metagenomes. Identified CarPULs, GH16_17 homologs, and highly homologous reads (100%) of putative GH16_17 members are marked on the map. Red outlines on the map indicate habitats of *M. japonica* [327]. The map was generated using ggOceanMaps [328]. **B)** CarPUL diagrams of bovine *B.*

xylanisolvens (this study), giraffe and musk deer *Bacteroides* (this study) and carrageenan PULs found within human- [205] and gorilla- [325] associated *Bacteroides* species.

3.4.5 Shared homology towards marine, sediment, and fish gut microorganisms

The wide distribution of CarPULs in terrestrial vertebrate GIT microbiomes raises an intriguing question about the origins of these CAZymes: did they evolve *de novo* in terrestrial vertebrate guts under dietary selection or originate from transfer from marine microorganisms, similar to what was described as the "sushi factor" for porphyran metabolism in the GIT microbiomes of Japanese people [202]. To explore this question, we compared sequence similarity between the mammalian CarPUL genes and NCBI-sourced marine and environmental genes and genomes and the *M. japonica* surface-associated microbiome (*Mj*SM) collected in this study. The top 100 BLAST hits and previously characterized CAZyme members [329] were aligned to CarPUL sequences. Homology was observed between CarPUL genes, and marine pelagic and sediment microorganisms within the GH2 family (**Table A2.9**) and sulfatases (between 70-80% identity for sulfatase 1_30 *vs. Mj*SM; **Table A2.10**). The highest identity with CarPUL sequences was consistently found between *Rikenellaceae* MAGs assembled from *Kyphosus sydneyanus* (Silver drummer fish) hindgut microbial communities [330]. SACCHARIS phylogenies revealed *Rikenellaceae* MAG GH16_17 members partitioned with CarPUL GH16_17 members, forming two distinct clades between *Bx*Car5$_{BOV}$ PUL-1 GH16_17A & B (**Figure 3.5A; Table A2.9**). Nearly the entire *Bx*Car5$_{BOV}$ CarPUL-2 shared homology with *R. alistipes* MAGs (**Figure 3.5B**), with the highest identity between putative sugar kinases (92%; **Table A2.11**). Genes surrounding the kinase (including metabolic genes, sulfatases, and GH127 members) were also highly syntenic with *R. alistipes* contigs (**Figure 3.5C**).

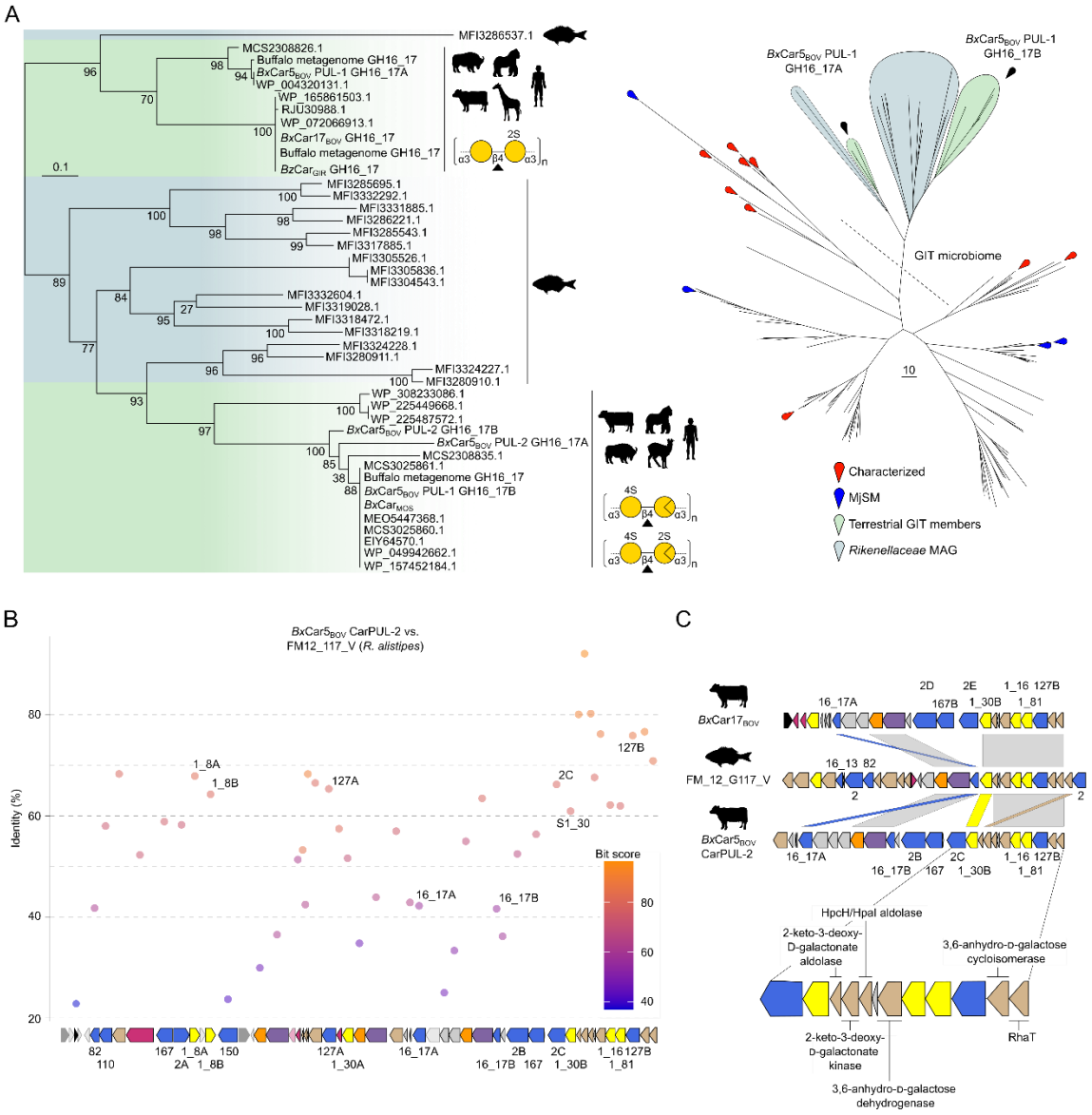

Figure 3.5: *Bx*Car PULs show homology with CarPULs from the marine fish GIT-associated *Rikenellaceaes* spp. A) The top 100 BLAST hits towards each CarPUL GH16_17 member (**Right**), predicted *Mj*SM GH16_17 members, and characterized GH16_17 members with the database were pooled alongside CarPUL GH16_17s and a phylogeny was created of aligned predicted catalytic domains. Clades are colored based on their host habitat; Green: terrestrial vertebrate *Bacteroides* spp. and teal: *K. sydneyanus* GIT members (*Rikenellaceae* MAGs; Facimoto et al. 2024). Individual nodes are denoted by a pin. BLAST hits are detailed in **Table A2.9**. A dotted line indicates clades exclusively containing GIT microbiome members (terrestrial vertebrate or *K. sydneyanus* sourced) which was further expanded on (**Left**) **B)** *Bx*Car5$_{BOV}$ CarPUL-2 genes were BLASTed against *Rikenellaceae* MAGs. Shown are the top hits towards FM_12_G117_V (*R. alistipes*). **C)** CarPUL synteny towards *Rikenellaceae* MAGs (shown FM_12_G117_V).

Regions of high homology were further expanded to highlight core functionality of CarPUL ancestors.

3.5 Discussion

3.5.1 CarPULs are structurally and functionally diverse latent traits

Although CarPULs were detected within many mammalian GIT microbiome datasets, they were not detected in control animals from the *ad libitum* or silage studies. This suggests that either CarPULs are not universal within herbivorous animals or sequencing depth may be insufficient to detect low-abundance PULs and CAZymes (*i.e.*, detection requires enrichment). Ruminants are notorious for the functional capacity of their rumen microbial communities. The role of latent traits in dietary adaptation extends this paradigm beyond what is detectable by low-depth, high-coverage sequencing methods of non-adapted metagenomes. Here, we saw a significant increase in *Bacteroidaceae* abundance within the faecal microbiome of cattle, underpinning carrageenan digestion may be a more pronounced trait within the lower GIT (**Text A2.1**). It is also possible that bespoke dietary enrichment strategies are required to detect cryptic latent traits. Red seaweed is primarily composed of sulfated galactans in the form of either agars and carrageenans, which can be modified with *O*-methylation, and discriminatory sulfation patterns [100]. Carrageenans are not chemically pure in nature and can exist as hybrids or mixtures, which influences their physiological properties [331]. Linkage analysis conducted here indicates that *M. japonica* matrix polysaccharides primarily consist of 'kappa family' (κ-, ι-, ν-, μ-) carrageenans [99, 100] (**Figure A2.7; Text A2.2 & Text A2.3**). Further, structural diversity in dietary seaweeds can be attributed to variations in algal species or seasonal factors [100].

Rikenellaceae CarPULs from the *K. sydneyanus* GIT likely saccharify κ/ι-carrageenans found within *Caulacanthus ustulatis* and other dietary seaweeds [332, 333]. Similarly, *M. japonica* may not have been a dietary component of ancestral ruminants and a primary source of carrageenans for the animals that were sampled. *M. japonica* is an introduced seaweed to the Pacific Northwest and originates from coastal waters of the Northwest Pacific (Korea, Japan, Russia) [334]; it is not present in all locations where

GH16_17 homologs were identified (**Figure 3.4A**). Mammalian GIT CarPULs can display different architectures and harbour putative CAZymes which target multiple carrageenan subtypes [133, 324], suggesting these pathways may be suited to target a diversity of carrageenans in nature, with strain host diversity contributing to individual catabolic responses [335].

3.5.2 GH16_17 phylogeny dictates specificity for carrageenan subtypes

CarPUL GH16_17 enzymes were observed to form phylogenetically distinct clades that correlated with substrate specificity and not host origin, with two distinct clades containing sequences from ruminant and non-ruminant hosts (**Figure 3.4; Figure 3.5**). *Bx*Car5$_{BOV}$ and *Bo*12C04$_{HOM}$ each contain two paralogs of GH16_17, one specific for κ/ι-carrageenan and one for α-carrageenan (**Figure 3.3**), a product of ι-carrageenan 4S-Gal*p* desulfation [324]. This was validated by their growth in pure culture on κ/ι-carrageenan [205] **(Figure 3.2B)**. Furthermore, the *Bx*Car17$_{BOV}$ GH16_17 is 100% identical to that found in the CarPUL from *Bt*3731$_{HOM}$ (**Table A2.4**), a strain which was isolated from humans and shown to metabolize λ-carrageenan [205]. In support, *Bx*Car17$_{BOV}$ grew on λ-carrageenan, and displayed compromised growth on *Mj*Ex and κ-carrageenan, requiring a 12 h lag period (**Figure 3.2A & 3.2B**).

A molecular determinant between κ/ι-carrageenan- and α-carrageenan-active GH16_17 enzymes results from mutations within the active-site pocket (**Figure 3.3C; Figure A2.4**). *Rikenellaceae* GH16_17 members did not display the glycine to glutamate mutation that excludes 4S-Gal*p* within the active site. In many cases, GH16_17 genes are paired with a GH16_13, which could provide a surrogate α-carrageenanse activity (**Figure 3.5C**). Intriguingly, the one *Rikenellaceae* MAG (FM_16_G121_IV) that did not contain a predicted GH16_13 member, possessed a GH16_17 with the highest similarity to *Bx*MAG$_{BOV}$ GH16_17A; however, in this case the glycine to glutamic acid mutation was not observed. This may provide a transitional snapshot into the evolutionary landscape of these partnered activities. The presence of multiple *Bx*MAG$_{BOV}$ GH16_17 members within this clade suggests that gene duplication and neofunctionalization of the GH16_17 subfamily may have occurred, similar to what was described for polysaccharide lyase

family 2 [336]. The presence of GH16_17 members with complimentary activities suggests it is a foraging strategy for complete saccharification of carrageenan-rich seaweeds.

3.5.3 CarPUL evolution and radiation into diverse microbial ecosystems

Previously, ancient HGT events have been implicated in the evolution of alginate utilization loci in human [221] and bovine [209] GIT metagenomes. Here, wide-spread radiation of CarPULs was observed between mammalian-associated metagenomes and isolates, and NCBI sequences across geographic barriers and host microbiomes. *Bacteroides* spp. compared herein contain highly syntenic CarPULs, irrespective of host source (giraffe, musk deer, human, cattle) and geographic region (Alberta, CA; Vancouver, CA; Michigan, US) (**Figure 3.4B**). Furthermore, high identity was observed between CarPUL CAZymes and GIT metagenomic reads from buffalo, sheep, and water buffalo from farms in Chinese provinces [26, 326], and cattle sampled on American farms [337] (**Figure 3.4A; Table A2.5; Table A2.6**). The geographic and host distribution of CarPULs suggests that these loci were acquired by ancestral seaweed-foraging terrestrial vertebrates and then maintained with low sequence divergence. Hierarchical responses to dietary seaweed polysaccharides and functional specialization of CarPULs at the strain or community level may help inform the origin and maintenance of latent traits. In this regard, we also discovered a porphyran/agarose PUL syntenic to those found within human gut *Bacteroides* spp. in the enriched giraffe faecal metagenome (**Figure A2.8**) [205]. As these catabolic abilities may be hidden in the genetic dark matter [318], assumptions of their presence or absence in the microbiome should be made cautiously. Of interest, CarPULs may have penetrated other ecosystems, as GH2 members discovered in Canadian landfills cluster with *K. sydneyanus* GIT and terrestrial vertebrate GIT members (**Figure A2.6**).

Bovine *Bacteroides* spp. CarPUL genes display high sequence identity with CarPULs from *R. alistipes* MAGs within the GIT of *K. sydneyanus.* This pattern was observed for all bovine CarPUL CAZyme families; and most notably for GH2, a well-studied, exo-acting polyspecific family. In the GH2 phylogenetic tree, nearly all CarPUL-associated sequences were present in clades primarily populated with GH2 members from *R. alistipes* and marine pelagic and sediment microorganisms (**Figure A2.6; Table A2.9**). These findings agree with previous reports that CarPULs in GIT-associated microbiomes

originated by HGT from marine microoganisms [205]. If a closer relative of bovine CarPULs exists within an as-yet-undiscovered marine bacterium, it may remain difficult to find. Marine environmental microbiomes contain great microbial diversity skewed by seasonal variation and other environmental stressors [338]. Therefore, the "ancestral" CarPUL may only be abundant during seasonally- or geographically constrained windows in extant bacteria, and absent from the available sequence datasets.

The origins of GIT-associated CarPULs appear complex. The mean % GC content within the bovine CarPULs was significantly different to the remainder of their genomes, suggesting the presence of these pathways resulted from HGT (**Figure A2.9; Text A2.2 & Text A2.3**). In contrast, *R. alistipes* MAG CarPUL genes were very similar in mean % GC content to the remainder of their genomes, suggesting CarPUL acquisition in the GIT of fish occurred much earlier than in terrestrial ruminants. This would correspond with *K. sydneyanus* feeding on seaweed well before the emergence of seaweed-consuming land animals. The differences in GC content also suggest acquisition occurred as separate events. However, it does not rule out that fish GIT microorganisms share a common origin with terrestrial mammalian CarPULs. CarPUL genes could have co-evolved throughout the vertebrate lineage or transferred into mammalian GIT communities during predation or faecal-oral transfer. In the latter case, HGT exchange between *Bacteroidales* order members has been observed within the human GIT [224] and marine ecosystems [339]. It is therefore likely that fish GIT *Bacteroidales* members, such as *Rikenellaceae*, may contain functional HGT genetic elements compatible with those found in terrestrial mammalian GIT members. Overall homology between putative integrative elements within bovine GIT CarPULs and fish GIT CarPULs was low; however, select genes do share up to 72% identity (IS4 family transposase; **Table A2.11**). Deeper investigation into the GIT microbiota of fish or other marine animals that forage on red seaweeds may hold subsequent clues into the mechanisms and origins of terrestrial mammalian GIT seaweed metabolism. Regardless of their origins, it is clear that CarPUL function is not host or geographically restricted and has proliferated within terrestrial mammalian GIT microbiomes as a latent trait within the genetic dark matter.

3.6 Methods

3.6.1 Animals and Environment sampling

Two animal trials were conducted as part of this study, *ad libitum* feed trial and a controlled silage trial. Both trials were conducted at a certified organic farm (Beaver Meadows farm - #03-16-198). For the *ad libitum* trial, Cattle (mixed heifers and steers) were separated into two groups on pasture. With one pasture offered *M. japonica* free choice (ca. 1 lb./cow). Fresh faecal samples were collected, and rumen samples were gathered from 10 animals (5 control, 5 treatment) slaughtered animals at Gunter Bros Meat Co. Ltd (6200 Ledington Road, Courtney, B.C.) For the silage trial, Angus X Hereford steers (n = 10) were feed a backgrounding diet (bait feed: alfalfa pellets) for two weeks, before switching to a grass-based silage. Cattle were fed a control grass silage diet for two weeks (**Table A2.12**). After the two-week period, fresh faecal samples for each animal were collected and rumen samples were collected *via* stomach tubing. The diet was then changed to include a 5% *M. japonica* supplement for another two-week period (**Table A2.12**). Fresh faecal samples were collected and rumen samples were collected via oral stomach tubing.

M. japonica samples were collected from Maple Guard Dr., BC. (Lat 49.441695, Long -124.676979) beachfront. *M. japonica* tufts were rinsed in artificial seawater to remove transient microbes, then finely scrapped with a sterile scalpel into 15 mL of artificial seawater and passed through a 0.2 μm filter. All samples were directly frozen in LN_2 and stored.

3.6.2 Metagenomic sequencing and data analysis

DNA from all samples were extracted using the DNeasy PowerSoil Pro kit (Qiagen, Canada) according to the manufacture's protocol. The DNA quality was analysed using Nanodrop 2000. 16S rRNA amplicon sequencing was achieved using Illumina MiSeq 2500 (Génome Québec). The 16S rRNA gene sequences were split into individual samples through the internal barcode. Sequences were trimmed using Trimmomatic (v0.39) [271] and passed through Kraken2 (v2.1.2 SILVA NR99 database) [272] for annotation. Kraken-biome [340] was used to create a biome file for phyloseq (v1.48.0) [341]. Statistical

analysis was done using the micoViz R package (v0.12.4) [320], dplyr (v1.1.4) [342], and ggplot2 (v3.5.1) [343].

DNA from faecal and rumen samples from the *ad libitum* study and the silage study, as well as the *Mj*SM were sequenced using Illumina Novaseq 6000 150 bp PE reads (Génome Québec). Raw reads were quality trimmed using Trimmomatic (SLIDINGWINDOW:5:20) and assembled individually using MetaSPAdes (v3.13.0) [278], and biological replicates together in a co-assembly using MEGAHIT (v1.1.3) [287] (**Supplementary Table 3.8**). Contigs were binned and refined using the MetaWRAP (v1.3.2) binning/refinement modules [279], with metaBAT2 (v2.12.1) [282], MaxBin2 (v2.2.6) [281], and CONCOCT (v1.1.0) [280]. Bins/MAGs between individual assemblies and co-assemblies were merged and de-replicated with dRep (v3.0.0) [283] at 70% completion and <10% contamination.

3.6.3 Metaproteomic data generation

To collect microbial cells from sample material with fibrous debris, faecal samples as well as the environmental samples (soil, compost, and filter paper of *M. japonica*) were gently washed prior to lysis and protein extraction. Microbial cells in 0.5 g of sample (faecal or environmental) were dissociated from the sample material through cycles of centrifugation and addition of dissociation buffer containing tert-butanol, Tween 80 and HCl to a pH of 2, and the cell containing supernatant collected and combined after each cycle [344]. The supernatant was centrifuged to a cell containing pellet, which was further washed in a cell wash buffer containing 10 mM Tris-HCl (pH=8) and 1 M NaCl. The remaining cell pellet was resuspended in lysis buffer (30 mM DTT, 150 mM Tris-HCl (pH=8), 0.3% Triton X-100, 12% SDS).

Similarly, 300 µl of the abovementioned lysis buffer was added to 0.5 g of rumen fluid. Glass beads (4 mm: ≤ 160 µm) were added to all samples (faecal, environmental and rumen samples), briefly vortexed and let rest on ice for 30 min. Cells were mechanically lysed using FastPrep-24 Classic Grinder for 3 cycles of 60 s at 4.0 m s^{-1}. Samples were further centrifuged at 16, 000 × g for 15 min at 4 °C and lysate transferred to a clean tube. For clean-up purposes, lysates were run on SDS-PAGE using Any-kD Mini-PROTEAN TGX Stain-Free gels and stained with Coomassie Blue R-250. Visible bands on gels were

carefully excised and reduced, alkylated, and digested to peptides with trypsin. Peptides were analysed by nano-LC/MSMS using a timsTOF Pro mass spectrometer (Trapped Ion Mobility Spectrometry quadrupole time-of-flight mass spectrometer) (Bruker, Germany).

3.6.3.1 Metaproteomic data analysis

MS raw data were analysed using FragPipe, powered by the proteomic search engine MSFragger [301] along with Philosopher [302] and IonQuant [303] for post-processing of MSFragger results including LFQ and false discovery rate (FDR) filtering. A sample specific sequence database used consisted of MAGs recovered from the silage study as well as MAGs from *Mj*SM. In total 443 MAGs (>70% completion and <10% contamination). The MAG database was supplemented with common contaminant protein entries, such as human keratin, and decoy protein entries based on reverse sequences for estimation of FDR. For LFQ, using IonQuant, the MBR feature was enabled. Oxidation of methionine and protein N-terminal acetylation were used as variable modifications, while carbomidomethylation of cysteine residues was used as fixed modifications. Trypsin was chosen as digestive enzyme and max missed cleavages allowed was set to one. FDR and MBR-FDR were both set to 1%. Further, detected protein groups were explored in Perseus (v. 2.0.7.0) [304]. Protein groups identified as possible contaminants were removed. Proteins were considered valid if they were detected in least 2/4 replicates, or 2/5 replicates for the environmental samples, in at least one sample type group (faecal samples from cows fed the control silage diet: FC, faecal samples from cows fed the silage diet supplemented with 5% *M. japonica*: FT, rumen samples from cows fed the control silage diet: RC, rumen samples from cows fed the silage diet supplemented with 5% *M. japonica*: RT, environmental samples, which included soil and compost samples as well as filter paper with *M. japonica*: ENV). One soil sample (Sample BM17-E – env_soil_1) mapped significantly fewer proteins in metaproteomic analysis compared to the rest of the data and was removed from further analysis as a technical outlier. Subsequently, this filtering resolved in a total of 4,420 unique protein groups across the 21 samples (**Table A2.13; Table A2.14**). Results from environmental samples were omitted from downstream analysis and interpretation of data. For statistical analysis, missing values were imputed by applying a 2.8 downshift from the normal distribution. Protein groups showing

significant changes in LFQ intensities ($p < 0.05$) were identified using a two-sided t-test. Dot plots for **Figure 3.1C** was created using ggplot2 [343] in R (v. 4.2.2) [276]. Volcano plot for **Figure 3.1D** was created using Perseus (v. 2.0.7.0). Predicted gene organisation map was created using the ggplot2 extension package gggenes [345], and protein expression heatmap was created using ggplot2.

3.6.4 *Bacteroides* culture enrichment and isolation

Rumen samples from cattle in the *ad libitum* feed trial were taken and inoculated directly into anaerobic media within sealed Hungate tubes (atmosphere: 85% N_2, 10% CO_2, 5% H_2). Anaerobic media was composed of *Bacteroides* minimal medium [200] (MM) supplemented with 1% Bacto™ Tryptone (BD), 0.5% Yeast Extract Bacteriological (VWR), and 0.5% *Mj*Ex. Cultures were grown until sufficient density was reached and re-inoculated into fresh media 3x before being glycerol stocked and placed into an anaerobic chamber with similar atmosphere. Cultures were diluted 10^{-3} and consecutively streaked on minimalized media agar plates containing 0.5% substrate (κ and ι-carrageenan, and *Mj*Ex) and 1% agar (4-5X) to obtain pure isolates. Isolates were grown in MM + 0.5% galactose.

Further *Bacteroides* enrichments were carried out on faecal samples from Wilder Institute//Calgary Zoo ruminants. Samples were collected from 7 ruminants, 3 pseudoruminants, and 1 nonruminant animals (**Table A2.8**). 1 g of faecal samples were diluted in 10 mL of anaerobic PBS and vortexed to homogenize on-site. 100 µL of homogenized mixture was inoculated in 1:1 anaerobic MM and *Mj*Ex and incubated for ~48 h.

Isolate DNA was extracted using the Qiagen Powersoil kit and was sequenced using Illumina (NovaSeq 6000 PE 150 bp), and Oxford Nanopore (R9.4.1; SQK-RBK110-96) sequencing. Reads from bacterial isolates from cattle and zoo animal samples were trimmed as above using Trimmomatic, and hybrid assembled using Unicyler (v0.4.8) [346]. Isolate genomes were checked for quality using Quast (v5.0.2) [347].

3.6.5 Carrageenan PUL analysis

Bovine MAGs and ruminant isolates were taxonomically classified using GTDB-Tk (v2.3.2 - R202) [286] and CheckM2 (v1.1.3) [348], and were functionally annotated with DRAM (v1.3.4) [163], prodigal (v2.6.3) [349] and dbCAN3 [151]. Protein sequence from CarPULs: *Bx*Car5$_{BOV}$ CarPUL-1, CarPUL-2, *Bx*MAG$_{BOV}$ PUL, and *Bx*Car17$_{BOV}$ were used as a protein database for Diamond BLASTx (v2.1.8) [290] against 17 TB of ruminant metagenomic datasets from cattle, buffalo, deer, moose, yak, and sheep collected from the literature (**Table A2.5**). Data was collected via NCBI-vdb (v3.1.0). Hits below a Bit score 50 and identity of 50% were filtered out, and the top hit for each SRR was selected for (**Table A2.6**). Buffalo metagenomic datasets were selected for co-assembly with MEGAHIT due to multiple high identity hits against carrageenan PUL sequences (**Table A2.7**).

CarPUL CAZyme protein sequences were BLASTed against the NCBI database and *Mj*SM taken in this study. The top 100 hits against CAZymes were collected and ran through SACCHARIS v2.0 (v2.0.1) [154] with CarPUL sequences. *Rikenellaceae* MAGs were downloaded from BioProject PRJNA1029302 and annotated as above. All phylogenies were visualized and annotated in ITOL [350]. OrthoFinder (v2.5.5) [293] was used to create species trees to align genome PULs in **Figure 3.2 & Figure 3.4**. All PUL diagrams were created using gggenes [345].

3.6.6 *Bacteroides* carrageenan growth profiling

*Bx*Car5$_{BOV}$ and *Bx*Car17$_{BOV}$ and wild-type *B. xylanisolvens* XB1A were cultured anaerobically at 37 °C overnight minimalized media + 0.5% galactose overnight. The overnight cultures ($OD_{600\ nm}$ 1.0–1.4) were diluted to an $OD_{600\ nm}$ of 0.025 in 1X MM. Wells of a 96-well microtiter plates (Falcon) were filled with 100 µL inoculant (n = 4) and 100 µL 0.3% (w/v) substrate. *Mj*Ex, κ, λ, and ι-carrageenan were autoclaved, the subsequently dialyzed against 6 kDa to remove any free monosaccharides. Galactose was used as a positive growth control. Negative control wells consisted of 100 µL 2X MM combined with 100 µL 0.3% (w/v) substrate and were used to normalize growth curves. Plates were sealed with polyurethane Breathe-Easy gas-permeable membranes (Sigma; Z390059). $OD_{600\ nm}$ of each well was measured with a Biotek Eon microplate reader and

recorded on Biotek Gen5 software every 10 min for 48 h. Mean (± standard deviation) of each condition (n = 4) was visualized using GraphPad Prism (v10.2.3).

3.6.7 *Bx*MAG$_{BOV}$ GH16_17B production and characterization

*Bx*MAG$_{BOV}$ GH16_17A & B were codon optimized and synthesized with Biobasic. Expression plasmids were transformed into BL21 (DE3) Tuner competent cells (Novagen) and grown in LB Miller broth containing kanamycin (50 ug mL^{-}). Cell cultures were grown to an optical density of 0.6-0.8 at $OD_{600\ nm}$ and induced with isopropyl β-D-1-thiogalactopyranoside at a final concentration of 1 mM at 16 °C overnight while shaking at 200 rpm. Cells were lysed using a combination of lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 0.1 mg mL^{-1} lysozyme) and sonication (2 min of 1 s intervals of medium intensity sonic pulses at a power setting of 4.5 – Heat systems Ultrasonics Model W-225). Cell lysate was centrifuged at 17,500 × *g* for 45 min and purified using Ni-NTA resin columns (Cytiva) and immobilized metal affinity chromatography. Samples were buffer exchange with 20 mM Tris, 500 mM NaCl, and 1% glycerol using ultrafiltration column, and further purification by size exclusion chromatography HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare). All fractions were confirmed using sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

Optimum pH was determined using a copper bicinchoninic acid reducing sugar assay measured with a SpectraMax ID3 (Molecular Devices) at 565 nm. Briefly, reducing sugar was measured from the release of substrate (3 mg mL^{-1} k-carrageenan) in the presence of 1 μM enzyme and 15 mM buffer. (citrate phosphate pH 3-8; bicine 8-9, CHES 10) after 5 minutes. Activity was plotted using GraphPad Prism (v10.2.3). Future reaction digests were completed at optimum pH. GH16 digests of κ, λ, and ι-carrageenan (carbosynth), and *Mj*Ex was conducted overnight with 3 mg mL^{-1} substrate, and were boiled at 90 °C for 10 min to complete the reaction. Digests were separated using thin layer chromatography on silica gel plates (Millipore) in a 1-butanol: distilled water: acetic acid (2:1:1 v/v/v) mobile phase and visualized with an ethanol: sulfuric acid (70:3 v/v) solution containing 1% (w/v) orcinol monohydrate (Sigma) and heated at 105 °C for 3 min. 1 mM κ -carrageenan standards (Dextra) were used as reference.

LC-ESI-MS was performed on a Vanquish UHPLC system (Thermo Scientific). Separation of the carrageenan oligosaccharides was achieved using an Acquity HILIC Column, 130 Å, 1.7 µm, 2.1 mm × 150 mm (Waters) at a flow rate of 300 µL min^{-1} at 30 °C, using a gradient over 20 min from 5-60% 10 mM ammonium formate pH 4.5 and 95-40% 10 mM ammonium formate pH 4.5 in acetonitrile.

Carrageenan oligosaccharide samples were prepared in water and injected in a volume of 10 µL at a concentration of 200 µg mL^{-1} for ESI-MS on an Orbitrap Fusion Tribrid system (Thermo Scientific) in negative ion mode. Mass spectra parameters are shown in (**Table A2.15**). To select ions for MS2 experiments, an intensity threshold filter was employed with a minimum intensity of 25,000 and maximum intensity of 1E+20, and a dynamic exclusion filter was used after 1 times for 2.5 s with default mass tolerance values. HCD was employed to generate fragments. MS spectra were analyzed using Xcalibur and Freestyle software packages (Thermo Scientific).

3.6.8 *Bx*MAG$_{BOV}$ GH16A protein production and purification

*Pf*S1_19B and *Pf*GH16B were produced and purified as described previously [245]. *Bx*MAG$_{BOV}$ GH16_17A for enzyme assays and crystallography was produced as follows. A preculture of BL21 (DE3) transformed with pET28-*Bx*MAG$_{BOV}$ GH16_17A was prepared by combining 50 µL of thawed glycerol stock with 40 mL of LB media containing the construct-specific antibiotics. The preculture was incubated overnight at 37 °C with shaking (170 rpm). The preculture was added to 2 L of 2x YT media containing the same construct-specific antibiotics and any necessary cofactors. The mixture was incubated at 37 °C with shaking until OD_{600} nm ~ 0.6 was reached. The cell mixture was cooled to 16 °C, expression of the construct was induced, and incubated overnight at 16 °C with shaking. The cell mixture was centrifuged at 5,000 rpm for 15 min at 4 °C. The pellet was resuspended in 10 mL of 25% sucrose and lysed by adding 10 mg of lysozyme, which was left to stir for 20 min. The lysis mixture was combined with 30 mL of deoxycholate and was left to stir for 10 minutes. The lysed cells were centrifuged at 5,400 × g for 60 min at 4 °C. *Bx*MAG$_{BOV}$ GH16_17A was purified from clarified cell lysate via immobilized metal affinity chromatography using solutions buffered with 20 mM TRIS, pH 8.0, containing 100 mM NaCl. Imidazole used for elution was removed by dialysis into 20 mM

TRIS pH 8.0, 100 mM NaCl overnight at 4 °C in. Protein was concentrated using a using a stirred ultrafiltration cell with a 10 kDa molecular weight cutoff filter. *Bx*MAG$_{BOV}$ GH16_17A was further purified by gel filtration chromatography using an S100 Sephacryl column and 20 mM TRIS, pH 8.0, 100 mM NaCl. Protein was again concentrated by ultrafiltration and quantified by absorbance at 280 nm.

3.6.9 *Bx*MAG$_{BOV}$ GH16A enzyme assays

Prior to performing the enzymes assays, a 1% (w/v) solution of ι -carrageenan was prepared in 20 mM TRIS, pH 8.0, containing 100 mM NaCl. A 5 mL aliquot was taken, placed into a 15 mL falcon tube, and incubated with 5 mM *Pf*S1_19B for 7 days while shaking at room temperature. A matched control with no *Pf*S1_19B was also incubated for 7 days. Both solutions were centrifuged for 3 min at 5,400 × g, the supernatant was collected, then filtered through a sterile syringe filter (0.22 μM). The untreated and *Pf*S1_19B treated carrageenan was then used for digests with *Pf*GH16B and *Bx*MAG$_{BOV}$ GH16A, both at 5 μM. Reactions were incubated for 3 days while shaking at room temperature.

Samples were labeled for fluorophore-assisted carbohydrate electrophoresis. Each digestion reaction was killed by adding 1 mL 95% EtOH and dried for 3 hours in the speed vacuum. Samples were resuspended in 5 μL of each 0.02 M ANTS (8-aminonaphthalene-1,3,6- trisulfonic acid) and 0.1 M $NaCNBH_3$, vortexed, spun down for 30 s at 5,000 rpm, and was left to incubate overnight at 37 °C wrapped in foil. Samples were dried for 4 hours in the speed vacuum. Once dried, each sample was resuspended in 500 μL of FACE loading dye and 5 μL of each sample was loaded into a FACE gel. The FACE gel was composed of a 35% acrylamide resolving gel and a 10% acrylamide stacking gel. The gel was visualized using a MultiImage™ Light Cabinet and AlphaImager software.

3.6.10 *Bx*MAG$_{BOV}$ GH16A crystallography

Crystals were grown at 18 °C using sitting-drop vapor diffusion for screening and hanging drop vapor diffusion for optimization. *Bx*MAG$_{BOV}$ GH16A at 5 mg mL^{-1} in 20 mM Tris-HCl, pH 8.0, with 0.5 M NaCl was crystallized in a 1:1 ratio with a solution of 0.2 M NaCl, 0.1 M Na_2HPO_4:Citric acid, pH 4.2, and 20% PEG8000. A single crystal of

*Bx*MAG$_{BOV}$ GH16A in crystallization solution supplemented with 20% (v/v) ethylene glycol as cryoprotectant was flash cooled by looping and plunging into liquid nitrogen. Diffraction data was collected on CMCF-ID beamline at the Canadian Light Source (CLS, Saskatoon, Saskatchewan) as indicated in (**Table A2.16**). All diffraction data were processed using XDS [351] and Aimless [352]. Data collection and processing statistics are shown in (**Table A2.16**). Initial phases were determined by molecular replacement using PHASER [353] and a model of *Bx*MAG$_{BOV}$ GH16A generated with AlphaFold 2 [354]. This initial model was manually corrected with COOT [355] and refinement of atomic coordinates was performed with Phenix.refine [356]. Water molecules were added in COOT with FINDWATERS and manually checked after refinement. In all datasets, refinement procedures were monitored by flagging 5% of all observations as ‘‘free’’ [357]. Model validation was performed with MOLPROBITY [358].

3.6.11 Production of fluorescent polysaccharides

FLA-*Mj*Ex and fluorescently labelled amylopectin (FLA-AP) were produced following a described protocol [359] with slight modifications. Prior to labelling, *Mj*Ex and amylopectin were hydrolysed by mild enzyme hydrolysis. Autoclaved *Mj*Ex (10 mg mL^{-1}) was treated with 2 µM *Bx*MAG$_{BOV}$ GH16B in 0.02 M citrate phosphate buffer, pH 5.5 for 3 h at 37 °C with mild shaking (50 rpm). The enzyme was then heat by boiling the solution in a hot water bath for 10 min. The sample was centrifuged and supernatant collected to run through a 5 kDa MWCO Vivaspin column (Sartorius) to remove low MW products. AP from maize (Sigma) was lightly digested with alpha-amylase from *Bacillus licheniformis* (Sigma) for 30 min, before heat killing and spinning out the enzyme. Low molecular weight products were removed using a Vivaspin column. The digested products were activated using CNBr and cleaned up with a Sephadex® G-50 gel filtration medium in a column connected to an ÄKTA Start™ chromatography system before incubating with fluoresceinamine isomer II (FLA; Sigma) overnight. Excess FLA was removed using a Vivaspin column as above until filtrate was clear. Retentate was collected, lyophilized, and stored at 4 °C.

3.6.12 Fluorescent polysaccharide incubations and imaging

*Bx*Car5$_{BOV}$, *Bx*Car17$_{BOV}$, and *Bx*XB1A$_{HOM}$ glycerol stocks were used to inoculate 0.5% Gal-MM and cultures were grown anaerobically. Overnight cultures were used to inoculate Gal-MM (not primed) or *Mj*Ex-MM (primed). Instead of *Mj*Ex-MM, *Bx*XB1A was subcultured into 50:50 Gal:*Mj*Ex -MM due to lack of grown on *Mj*Ex as the sole carbon source. Cultures were centrifuged for 5 min at 5,000 × *g* and pellets were resuspended in 2X MM. All cultures were diluted to 0.8 OD_{600nm}. For 0 h controls, 50 µL of resuspended cultures were aliquoted into 500 µL 1% formaldehyde (FA). 20 µL of resuspended cultures were then subcultured into tubes containing 50 µL 2X MM and 50 µL 0.4% FLA-*Mj*Ex, FLA-AP, *Mj*Ex, AP, or FLA + *Mj*Ex. 50 µL of culture was aliquoted into 500 µL of FA at 1 h and 1 d. Samples were fixed in FA overnight at 4 °C before centrifuging (5,000 × *g*, 10 min), removing the supernatant, and resuspending in 500 µL 1X phosphate saline buffer (PBS). Cells were imaged as previously described [360]. Briefly, samples were diluted 1 in 5 with 1X PBS buffer and fixed onto an Isopore™ filter using gentle vacuum. A 4:1 Citifluor™ AFI mountant medium to Vectashield™ vibrance antifade mounting medium solution containing DAPI was used to mount and counterstain the filter pieces. A Leica DMRBE epifluorescence microscope coupled to a Leica DFC 7000T camera and a X-cite 110 LED illumination system was used to image the samples. DAPI and FITC filter cubes were used during imaging.

Microscopy was conducted with an ECHO Microscope RVL2-K4 (A BICO COMPANY, USA) equipped with LED light cubes DAPI (EX: 385/30 EM: 450/50 DM: 425) and FITC (EX: 470/40 EM: 525/50 DM: 495) using a Plan X Apo oil; 1,42 NA, 60× oil emersion objective. For the FLA-*Mj*Ex imaging fixed LED strength of 46% and fixed exposure times of 40ms were applied. The threshold of 40ms was applied to prevent false positive analysis of cell autofluorescence. The images were processed using the Echo Pro software version 6.4.2 (Echo Laboratories). For the quantification of total FLA-*Mj*Ex manual counting was performed of a minimum of 15 fields of view from each sampling time point. Positive FLA-*Mj*Ex cells were identified by a co-localised DAPI and FLA-*Mj*Ex signal.

3.7 Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD060679. All sequencing data have been deposited to the NCBI SRA under the project accession numbers PRJNA1227608. The functionally annotated MAGs are available at Figshare (DOI: 10.6084/m9.figshare.28464551). The coordinates and Crystal structure factors for *Bx*MAG$_{BOV}$ GH16_17A have been deposited in submitted to the Protein Data Bank under the identifier (ID: 9EFL).

3.8 Ethics approval

All procedures and protocols involving cattle were reviewed and approved by the Thompson Rivers University animal care committee (award #: 101948). with cattle cared for following the guidelines of the Canadian Council on Animal Care. Faecal collection from animals at the Wilder Institute/Calgary Zoo was approved by the Wilder Institute/Calgary Zoo Animal Welfare, Ethics, and Research Review Committee (Protocol 2023-06).

3.9 Acknowledgements

Research at AAFC was supported through an AgriScience Project (ASP-207; J-002817) and Prize for Outstanding Achievement in Science (AAFC) awarded to D.W.A. (J-003135). LC-ESI-MS data was collected at the University of Lethbridge, Lethbridge Central Analytical Facility, with support from Tony Montina, Vincent Weiler, Maurice Needham, and Carl Holland. Insightful bioinformatic suggestions were provided by Rodrigo Ortega Polo at Agriculture & Agri-Food Canada Lethbridge. L.H.H. is grateful for support from The Research Council of Norway (302639 - SeaCow). P.B.B. is grateful for support from the Novo Nordisk Foundation (0054575—SuPAcow) and the Australian Research Council (Future Fellowship: FT230100560). Research at the University of Victoria was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant (FRN 04355). T.R.P. was supported by the Canada Research Chair program. Mass spectrometry-based proteomic analyses were performed by The MS and Proteomics Core Facility, Norwegian University of Life Sciences (NMBU). This facility

is a member of the National Network of Advanced Proteomics Infrastructure (NAPI), which is funded by the Research Council of Norway INFRASTRUKTUR-program (project number: 295910). We are thankful for the generous access provided by the Molecular Ecology Department of the Max Planck Institute for Marine Microbiology. G.R. was supported through funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; project number 496342779). We acknowledge the assistance of Dr. John Church (Thompson Rivers University) in assisting to organize the farm trial. We thank Arthur Valdivieso for technical assistance with *Bx*MAG$_{BOV}$ GH16_17A activity experiments. We also greatly appreciate the Animal Care, Health and Welfare staff at the Wilder Institute/Calgary Zoo for facilitating access to and collecting faeces for the study.

Chapter 4 Rewilding of cattle exhibit adaptation to a browser-grazer lifestyle through intestinal microbiome plasticity

4.1 Preface

Chapter 4 is presented in manuscript form that has been formatted for consideration in the journal Science (Tingley et al. 2025). This study examines the faecal microbiota of a wild herd of cattle "*Muus Muus*" located on Meares Island. As first author, I participated in the writing and structuring of the manuscript, revision and editing process. My contributions to the manuscript include metagenomics analysis of *Muus Muus* faecal samples, 16S rRNA and *tnL-P6* amplicon sequencing analysis of faecal samples, and the production and characterization of ulvan degrading CAZymes. The full author contribution list is presented below. My contributions are displayed in figures 4.1-4.4, 4.5A & B; Appendix 3 figures A 3.1-3.9; and Appendix tables A3.1-A3.10. The citation style of this manuscript has been updated from the original manuscript to maintain continuity in this thesis.

Authors: J. P. Tingley[1,2], T. O. Andersen[3,4], Kristin E. Low[1], Shawn Quick[5], Nic Jujihara[1], Leeann Klassen[1], Liam Mihalynuk[6], Greta Reintjes[7], Rahat Zaheer[1], Tim A. McAllister[1], Naomi Hanna[1,2], Mark Sumarah[8], Trushar R. Patel[9,10], Alisdair B. Boraston[6], Live H. Hagen[4], Phillip B. Pope[3,4,11], D. Wade Abbott[1,2]*

Affiliations:

[1] Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

[2] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

[3] Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway

[4] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway[5]

[5] Public works, Tla-o-qui-aht First Nation, Tofino, BC, Canada

[6] Department of Biochemistry and Microbiology, University of Victoria, Victoria British Columbia, Canada

[7] Microbial-Carbohydrate Interactions Group, Faculty 2 Biology / Chemistry, University of Bremen, Bremen, Germany
[8] London Research and Development Centre, Agriculture and Agri-Food Canada, London, ON, Canada
[9]Alberta RNA Research and Training Institute and Department of Chemistry and Biochemistry, Lethbridge, AB, Canada
[10] Department of Microbiology, Immunology and Infectious Disease, Cumming School of Medicine, University of Calgary, Calgary Alberta, Canada
[11] Centre for Microbiome Research, School of Biomedical Sciences, Queensland University of Technology (QUT), Translational Research Institute, Woolloongabba, QLD, Australia

*To whom correspondence should be addressed: wade.abbott@agr.gc.ca

Author Contributions:

D.W.A and J.P.T, conceived and designed the study. T.O.A, L.H.H., P.B.P, conducted proteomic analysis. J.P.T., K.E.L., N.H., and D.P.W collected and enriched samples for isolations and metagenomics. J.P.T. conducted 16S rRNA and *trnl-P6* analysis. J.P.T performed shotgun metagenomic analyses and metagenomic surveys. J.P.T performed growth curves and isolations. J.P.T, L.M., K.E.L., and N.J. designed ulvan lyases constructs, purified protein, and characterized CAZymes. L.M. and A.B.B. conducted crystallography. J.P.T, T.A.M and R.Z. conducted AMR analysis. M.S. conducted water analysis. D.W.A. and T.R.P., supervised the work. J.P.T. and D.W.A drafted the manuscript. All authors edited and approved the manuscript.

4.2 Abstract

Since the domestication of livestock by agrarian societies (~10,500 years ago), management practices have been used to optimize animal health and performance. The impact on these practices on microbial diversity and inheritable traits, such as catabolic potential and antimicrobial resistance (AMR), is not understood. Anthropocentric antimicrobial use in livestock production has led to the accumulation of AMR genes in bovine intestinal microbiomes, which is of concern for transmission to humans. Rewilding represents a contemporary approach for studying how animal microbiomes can be returned to a more natural state, potentially contributing to animal resilience and agricultural sustainability. Here, we present a transdisciplinary multi-layered omics study of faecal-associated microbiomes from of the "*Muus Muus*", a herd of rewilded cattle that live in the traditional territory of the Tla-o-qui-aht Peoples on Meares Island, British Columbia. These cattle have adopted a browser-grazer lifestyle foraging on forest plants and native grasses, which is reminiscent of their extinct, forest-dwelling ancestors, the auroch. In addition, the *Muus Muus* graze along the shoreline of Meares Island, consuming eelgrass and seaweeds. Through paired 16S rRNA, shotgun metagenomics, and *trnL-P6* sequencing, we reveal seasonal microbial-plant interactions in faecal samples, which presents a promising approach for investigating the foraging behaviors of wild herbivores. High resolution mapping of polysaccharide utilization loci using shotgun metagenomics and metaproteomics confirmed microbial digestion of exotic browses, including green seaweed ulvans. The resistome of the rewilded *Muus Muus* shows vastly less abundance and higher diversity of AMR genes, offering a unique glimpse into profiles of non-selected ruminant microbiome and potential solutions for the prevailing AMR crisis in animal agriculture.

4.3 Introduction

The diverse grazing and browsing lifestyles of wild ruminants, such as the extinct auroch (*Bos primigenius*), enabled their adaptation to changing landscapes and the emergence of agrarian societies [361]. How these impacted the structural and functional diversity of their intestinal microbiomes remains unknown. De-extinction [362] and rewilding [363] are two contemporary approaches that illuminate ancient interactions that existed between ancestral hosts, their microbiomes, and habitats. Rewilding efforts have

been theorized to reduce the impacts of climate change [364] and restore natural ecosystems [365]; however, very little is understood about the impact of rewilding on prevailing issues for animal health, such as AMR and inheritable catabolic potential.

Therapeutic and non-therapeutic antimicrobial use (AMU) in livestock has resulted in an AMR crisis [366] and an escalation in livestock morbidity [367]. Of upmost concern is the transmission of AMR genes to bacterial pathogens of importance for human health, and AMU is now viewed through the One Health lens [368]. Although global reform has curtailed AMU in agriculture, resistance towards anthropomorphic-introduced antimicrobials is persistent, being found within "antimicrobial-free" production systems [369].

The metabolic plasticity of bovine GIT microbiomes across different species and in response to altered diets has been well recognized [14, 15, 135]. The full catalytic potential of the bovine GIT, however, remains unknown. In the GIT, dietary fibres, such as cellulose, hemicellulose and pectins are saccharified by microbial CAZymes [15]. The chemical diversity of structural polysaccharides influences the expression and abundance of CAZymes [264]. Microorganisms from phylum *Bacteroidota* typically harbour a multitude of CAZymes housed within PULs dedicated to the saccharification of specific polysaccharides [137]. PULs have been described for common and exotic dietary polysaccharides, such as those found in seaweeds, persisting within intestinal microbiomes as "latent traits" [209, 258]. These relationships hold promise for developing alternative feeds to mitigate the impacts of a changing climate and landscapes [214].

The *Muus Muus* are a herd of wild cattle (n=10-20) that live in the old growth temperate rainforest of Meares Island, which lies in the traditional territory (haahuulthii) of the Tla-o-qui-aht First Nations. According to oral history, the herd was brought to the island in the late 19th century where they were a source of meat and milk for the Christie residential school (1900-1983; **Figure 4.1**) [370]. The herd therefore predates the introduction of antimicrobial use in animal agriculture (1935) [366]. For over a century, the *Muus Muus* have adopted a free-living lifestyle, patrolling the village of Opitsaht and surrounding land, grazing on native grasses within the village, and forest vegetation including salal, cedar and berry bushes, and marine eelgrass and seaweeds. Consequently,

unique CAZyme inventories would be required to digest the exotic diet consumed by the *Muus Muus*.

In this study we employed a two-eyed seeing approach ("Etuaptmumk") [371] based around the Tla-o-qui-aht Nation belief of "*his-shuk-nish-tsa-waak*", meaning "Everything is One" [372]. By weaving traditional ways of knowing with multi-layered omics, we sought to learn how the bovine microbiome adapted to exotic diets associated with a browser-grazer lifestyle, and how the lifestyle of this isolated herd contributes to animal health and resilience in the absence of AMU.

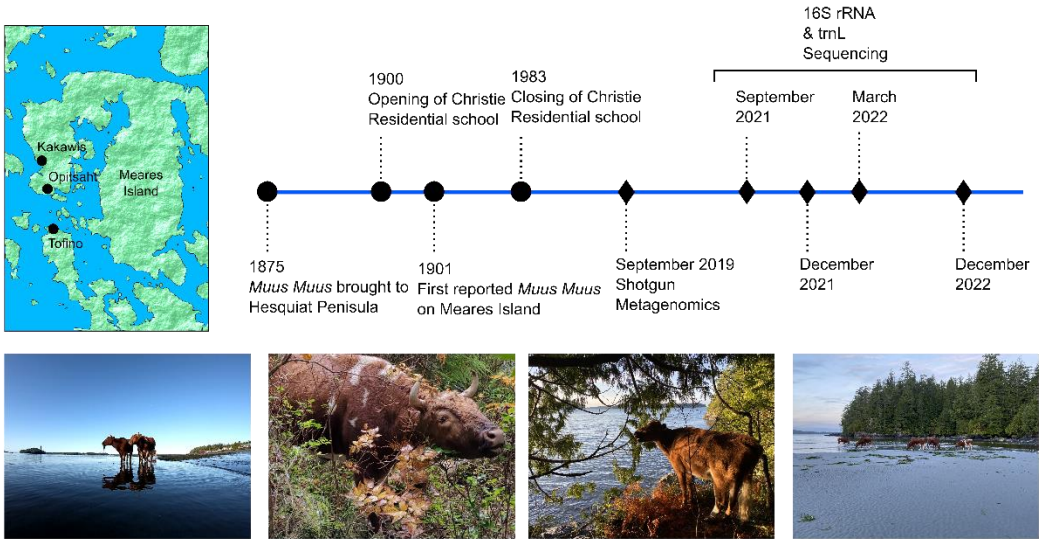

Figure 4.1: Rewilded *Muus Muus* have adopted unique foraging habits to graze in the old growth forest of Meares Island. The *Muus Muus* were introduced to Meares Island in the late 19th century by missionaries. Since, the herd have adapted to browser-grazer dietary habits in the absence of conventional livestock management practices.

4.4 Results

4.4.1 *Muus Muus* faecal microbiomes display distinct AMR profiles to conventional cattle systems

To collect independent data, the *Muus Muus* were categorized using sex, horn shape and size, and hide colors and patterns (**Figure A3.1**). Shotgun metagenomics analysis was performed on 19 faecal samples collected from the herd (**Table A3.1**), and the reads were

compared to GIT datasets from other ruminant species. The abundance of AMR genes within the *Muus Muus* were compared to datasets from conventional [369, 373], "natural" [369, 373], and "organic" beef and dairy cattle [369, 373]. The most apparent difference between *Muus Muus* faecal samples and traditional cattle was the low abundance of tetracycline resistant genes in the rewilded animals. The most significantly different gene across all comparisons is *tetQ* **(Figure 4.2A; Table A3.2**), encoding a ribosomal protection protein. The only tetracycline resistance genes detected within *Muus Muus* metagenomes were *tet(40)*, which encodes an efflux pump from *Clostridium* spp.. Further, numerous β-lactam and MLS resistance genes commonly seen within conventional dairy systems were noticeably less abundant or completely absent in *Muus Muus* samples (**Figure 4.2A; Figure A3.2**). Notably, there was not a single read to MEFA/B/C/E, an erythromycin efflux pump, within *Muus Muus* metagenomes. MEFA/B/C/E is a prominent marker found in conventional, natural, and organic systems (**Table A3.3**) [374]. The only evidence of β-lactam resistance genes was found in 4 animals, whereas two animals displayed reads for BLAEC β-lactamase for penicillin [375]. The most abundant AMR genes detected within the *Muus Muus* metagenomes were predicted to be from the aminoglycoside family, primarily APH2-DPRIME and AAC6-PRIME genes (**Figure 4.2B**). *Muus Muus* AMR gene counts were low in all but four of the nineteen *Muus Muus* samples (**Table A3.3**). Within these four samples, AMR relative abundance was predominantly AMR genes associated with metal and multi-compound efflux pumps (**Figure 4.2A**). These genes were present within all samples (conventional, natural, and organic), and were often not significantly different (**Figure 4.2B; Table A3.2**).

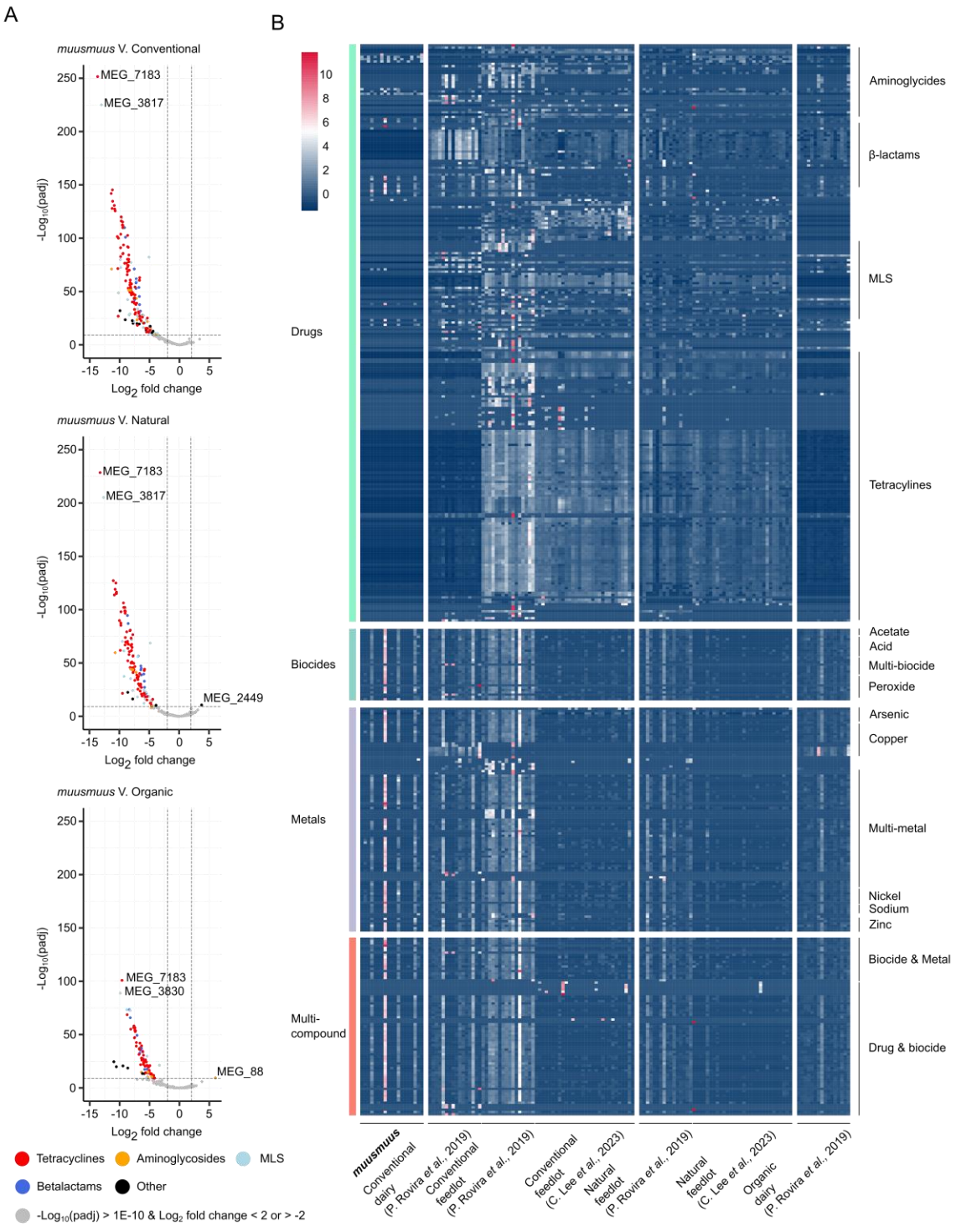

Figure 4.2: *Muus Muus* metagenome AMR profiles are distinct from traditional and organic cattle systems. A) Volcano plots of DESeq2 analysis between *Muus Muus* faecal samples, and samples from conventional, natural, and organic feedlot systems [369, 373], and conventional and organic dairy systems. Significant genes are colored by AMR class.

B) DESeq2 normalized abundance of biocide, metal and multi-compound AMR categories within all cattle production systems.

4.4.2 Seasonal mapping of consumed plants and microbial communities

Muus Muus faecal samples were collected from four animals between September 2021 and November 2022, with multiple faecal samples collected from each animal over week long sampling periods. Chloroplast *trnL*-P6 sequencing was conducted to determine if faecal DNA could be used to identify plant consumption and foraging patterns [376]. 183 unique ASVs were detected in at least one *Muus Muus faecal* sample, as opposed to the 16 found within barley fed (52% barley silage, 44% barley straw) control animals (**Table A3.4**). Observed α-diversity further pruned the data to 21-50 ASVs in the *Muus Muus* faecal samples, and 5-9 in the control animals (**Figure A3.3**). Tla-o-qui-aht traditional ecological knowledge was used to select control plants known to be consumed by the *Muus Muus*, which determined the taxonomic position of 16 plants (**Table A3.4 & A3.5; Figure 4.3A**). This analysis identified 4 of the top 8 plants (including the 3 most abundant plants by average count): *Gaultheria* sp. ("Salal"), *Rubus* spp. ("Himalayan blackberry", "Cutleaf blackberry" and "Salmon berry"), and *Thuja* sp. ("Cedar"), and *Ranunculus* ("Buttercup") (**Figure 4.4A**). Alongside other abundant plants, including *Plantaginaceae, Poaceae*, and *Asteraceae* family members, these forages were present in significantly different amounts between sampling periods (**Figure A3.4; Table A3.6**). Shrubs (*Physocarpus* sp.), moss (*Brachytheciaceae* sp.), and both control and eelgrasses (*Zostera* spp). were also significantly different between sampling periods, but far less abundant within datasets (**Figure 4.3A; Figure A3.4**). Clustering of significantly distinct ASVs revealed faecal samples from September 2021 separated from the other collection periods due to a greater abundance of *Zosteracea*, *Plantaginaceae*, *Asteraceae,* and *Rubus spp.* members, which was further supported by redundancy analysis (RDA) (**Figure 4.3B**).

16S rRNA sequencing was done on the same samples to identify if dietary patterns correlated with microbial profiles. DESeq2 revealed 134 differentially abundant operational taxonomic units (OTUs), including multiple *Methanobacteria* and *Clostrida* members (**Table A3.7**). Of note, September 2021 samples contained lower abundances of *Methanobacteria* then other collection periods (**Figure 4.3B; Figure A3.5**). Furthermore, Clostridia members varied between seasonal sampling with *Oscillospiraceae* UCG-010

family & *Lachnospiraceae gauvreauii* members being less present in the December 2022 collection, with the opposite trend being seen in *Oscillospiraceae Colidextribacter* members (**Figure A3.5**). Super imposed distance matrixes using procrustes analysis revealed moderate similarity between 16S rRNA and *trnL-P6* datasets ($M^2 = 0.533$; $r = 0.684$; $p<0.001$) with the most correlated datasets (lowest residual difference – **Table A3.8**) from March 2022 and December 2022 faecal samples (**Figure 4.3C**). SpiecEasi was used to identify associations between 16S rRNA and trnL datasets. The network was highly connected (Estrada class III) [377] with low commute times between nodes (0.060). Louvian clustering was used to detect community membership which demonstrated moderate observed modularity (0.583; $p<0001$ compared to a 1000 permutation null model). 19 communities were observed, with the largest containing 1,308 members (**Figure A3.6; Table A3.9**), and the greatest % of *trnL-P6* sequences within a community found at ~20%. Between significantly different taxa identified within the individual DESeq2 analyses, most taxa fell within the largest community (**Table A3.9**).

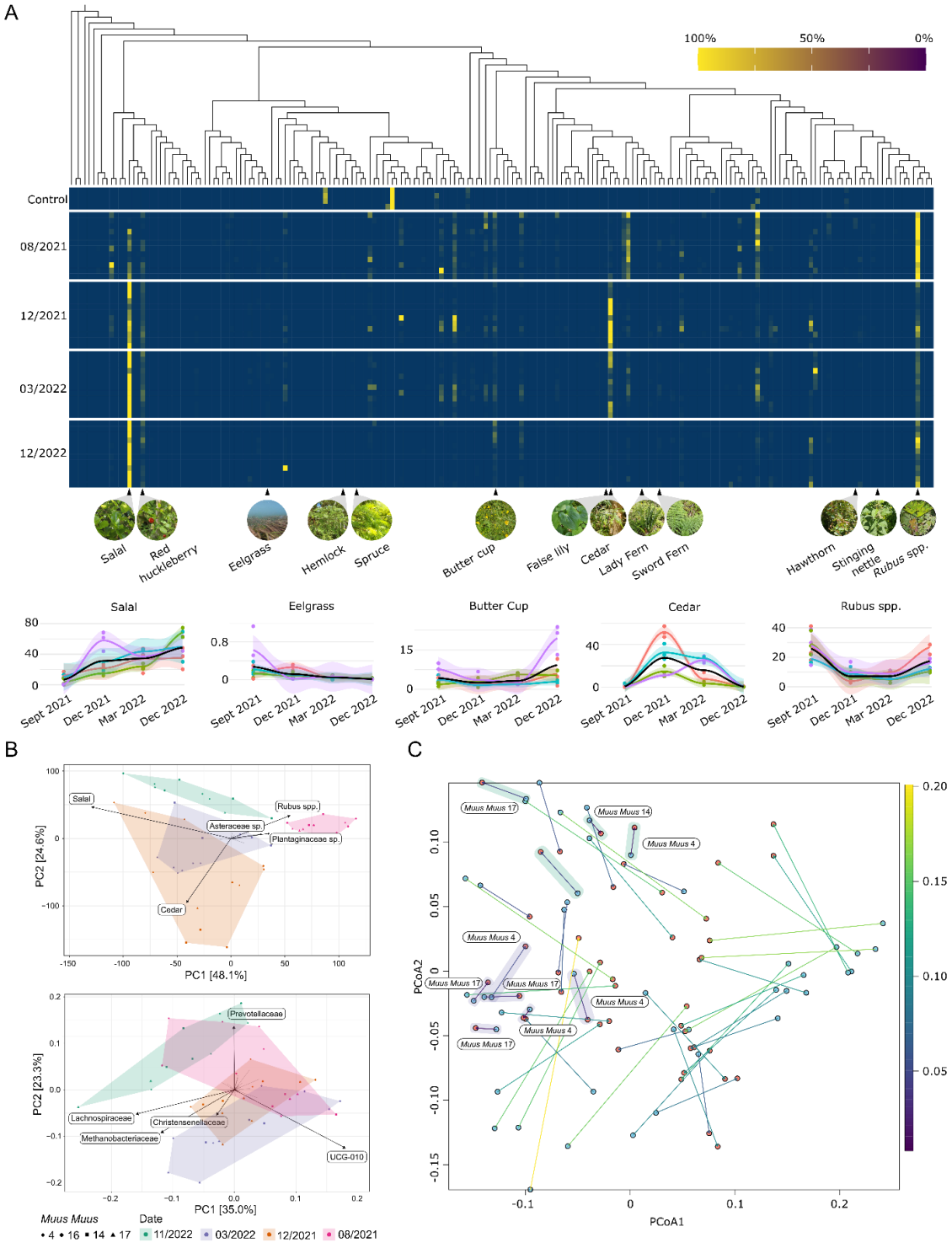

Figure 4.3: The *Muus Muus* consume a diverse range of browses and grazes. A) Relative abundance heatmap of *trnL-P6* sequences identified within Meares Island faecal sample. Phylogeny was created after aligning and trimming extract ASVs, finding a best-fit model using ModelTest-NG [297] and a phylogeny using RAxML-NG [298]. Heatmaps were separated by sampling period and generated with ITOL [350]. Plants collected and

sequenced from Meares Island are shown below the heatmap. Relative abundance line graphs illustrate significantly different taxa between sampling periods, with the black line denoting the average between the 4 animals sampled. **B)** RDA plots generated using microViz [320] of the *trnL-P6* (**top**) and 16S rRNA (**bottom**). The top 5 taxa resolved in the RDA analysis are shown. **C)** Procrustes analysis between trnL-P6 and 16S rRNA datasets. Residual length is denoted via colored gradient and similar colored lines. The top 10 shortest residuals are highlighted in the sample period color as denoted in B.

4.4.3 *Muus Muus* faecal MAGs harbour unique CAZymes, while retain core ruminant polysaccharide utilization systems

Shotgun metagenomic reads were assembled and binned to generate MAGs for analysis. After dereplication, 807 MAGs were generated above 70% completion and under 10% contamination (**Table A3.10**). Of these, 227 MAGs fall in the high-quality category (>90% completion <5% contamination) [230]. These MAGs consist of 151 *Bacteroidota*, 534 *Bacillota*, 21 *Pseudomonadota*, and 64 *Verrucomicrobiota* members alongside *Actinobacteriota, Chloroflexota, Cyanobacteria, Desulfobacterota,* Myxococcota*, Patescibacteria, Spirochaetota* members. *Muus Muus* MAGs were compared to a collection of ruminant GIT genomes including buffalo [26, 326], cattle [26, 165, 378-380], deer [26, 379], sheep [379], and yak [381] (n=18,702), and clustered using dRep compare. 553 Meares MAGs were not found within a secondary cluster with a ruminant GIT dataset genome (~95% FastANI), while 471 MAGs were not found within a primary cluster (~90% FastANI; **Figure 4.4A**). Of the MAGs, *Bacillota* represented the largest taxonomic group (n=356), followed by *Bacteroidota* (n=44), and *Verrucromicrobia* (n=31).

The ruminant genomes were annotated with dbCAN and putative CAZymes, alongside CAZymes downloaded from the CAZy database, were compared to the *Muus Muus* MAGs. Further, surface-associated microbiomes from *Zostera marina* and *Ulva intestinalis* were sequenced and run through dbCAN to detect potential horizontal gene transfer targets from environmental sources. Within each of the datasets, full sequences were dereplicated with CD-HIT at 95% and were blasted against the *Muus Muus* MAG CAZome. Blast hits against the *Muus Muus* dataset above 60% identity and 50 bitscore were treated as relevant hits. Of the over 70 thousand CAZymes identified, roughly 11 thousand showed no hits above 60% towards any dataset tested (~10%), while ~18% showed similarity above 90% identity (**Figure A3.7A**). The largest amount of hits was

found between the buffalo GIT and cattle GIT; this was expected as they are the largest databases consisting of lower GIT collected MAGs. Most blast hits were against GHs, followed by GTs and CEs. The highest identity observed between *Muus Muus* faecal MAGs and the environmental dataset was observed between *Bacteroidota* PL family 1 enzymes (71%). High identity blast hits were observed with numerous endo-active fibre and starch degrading CAZymes including GH5_4 members (cellulose, mixed linkage, xyloglucan), GH16_3 members (mixed-linkage glucans), GH28 (pectins), and GH13 subfamilies (starch, glycogen) (**Figure 4.4B**). Interestingly, only a single alginate targeting PL6 member shared identity with the environmental dataset (64.7%), whereas all but two PL6 members within the Meares dataset shared high identity with ruminant GIT MAGs (mean 90%; max >99% identity). PL17 genes, which are often clustered with PL6 [221], also displayed near 100% identity with ruminant GIT MAGs. In contrast, PL family members targeting animal glycans were more numerous and contained many divergent sequences. PL12 (n=111), PL8 (n=90), PL33 (n=267), PL35 (n=248) were highly abundant within the *Muus Muus* MAGs, a majority of which were assigned to *Bacillota* (70-85%). Of these, PL8 members observed the highest amount of sequence divergence.

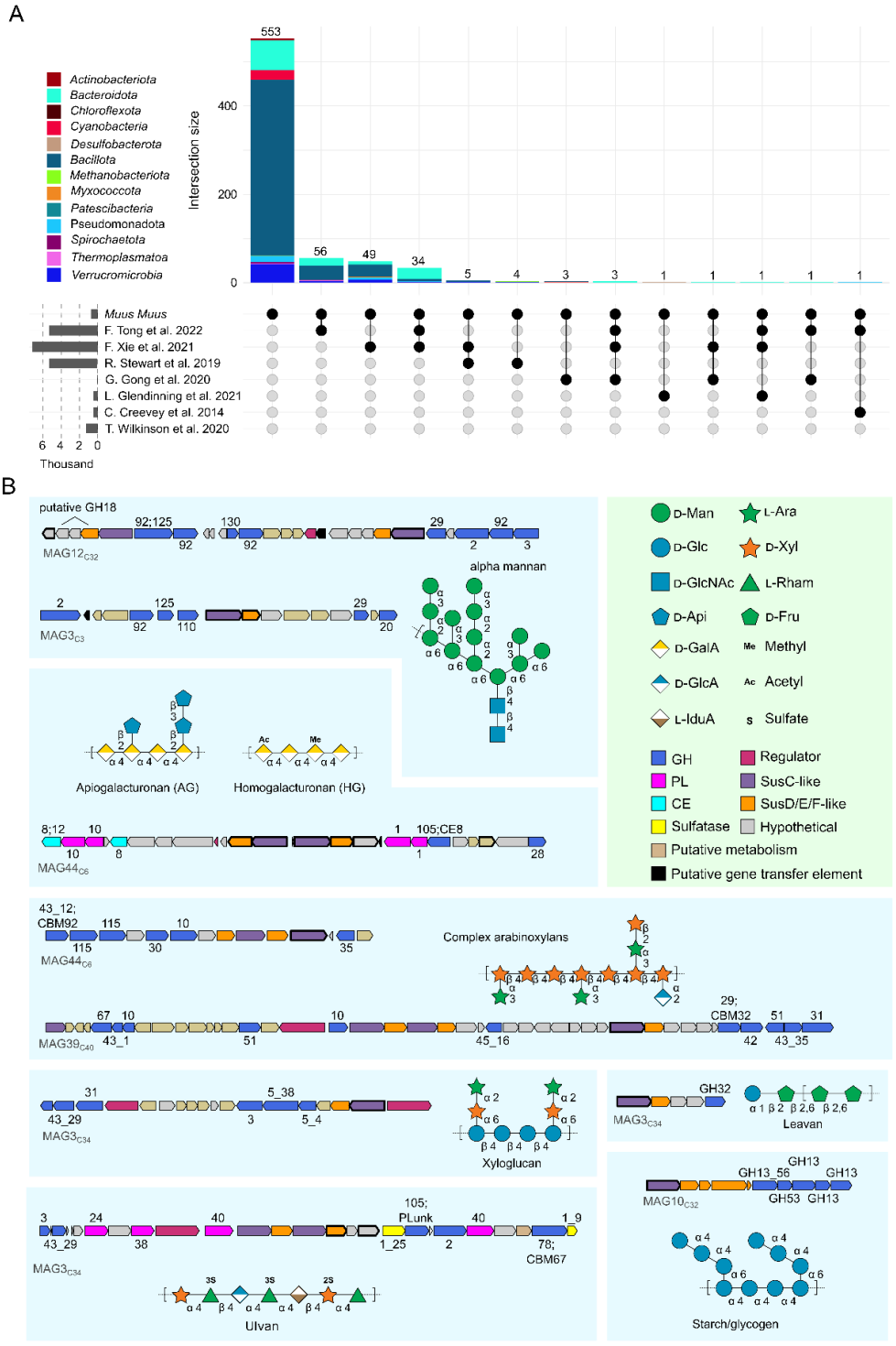

Figure 4.4: Unique *Muus Muus* MAGs contain high identity CAZymes to ruminant datasets. A) UpSetR plot of *Muus Muus* MAGs which fell within a secondary dRep cluster (95% fastANI) [382]. Bar charts are colored based on the GTDB-tk [286] predicted phylogeny of the MAGs. Only relationships between *Muus Muus* MAGs are shown for clarity and relevance. **B)** *Bacteroidota* PULs within *Muus Muus* faecal metagenomes

confirmed via proteomic detection of at least one sentinel SusC/SuD-like member. PUL proteins detected within the proteome are bolded. The predicted substrate based on characterized CAZymes and PULs is illustrated.

Bacteroidota spp. were investigated for PULs using PULpy [162] and manual inspection of SusC/D-like pairings. These areas were further annotated with sulfatase atlas [139] to identify putative sulfatases. 871 SusC/D-like parings were observed, with 304 containing at least 1 CAZyme. Vast substrate diversity was predicted based on observed PULs and metaproteomic confirmation of sentinel SusC/D-like proteins targeting traditional terrestrial plant polysaccharides: cellulose/xyloglucan, starch, fructans, hemicellulose, and pectins; and yeast α-mannans (**Figure 4.4C; Table A3.11**). Host glycan PULs, putatively targeting chondroitin sulfate, heparin, and hyaluronan identified by shotgun sequencing were not detected by proteomics (**Figure A3.7B**). Seaweed polysaccharide targeting PULs, including alginate (brown seaweed), agarose and porphyran (red seaweed), and ulvan (green seaweed) were also identified (**Figure 4.4C; Figure A3.7B**), with the ulvan PUL (UlvPUL) confirmed by proteomics detection of the SusD-like protein in 5 samples (from 3 different animals; **Table A3.11**). The majority of *Muus Muus* CAZymes (>60% identity towards ruminant GIT datasets) did not cluster into PULs, aligned with CAZymes present in truncated contigs, or were associated with structurally and functionally distinct PULs from other rumen datasets (often observed within host glycan PULs). Regardless, selected *Muus Muus* faecal PULs did align to ruminant dataset PULs with high identity, especially between MAGs from the buffalo dataset [326]. This includes nearly identical starch/glycogen PULs, α-mannan PULs, and notably an alginate PUL, which was shared between two *Paludibacteraceae* RF16 members (**Figure A3.8**). Although not found within other ruminant datasets, a porphyran/agarose PUL highly similar to the *Phocaeicola plebeius* DSM 17135 PUL identified with Japanese human guts was identified within the *Muus Muus* MAGs (**Figure A3.8**). Notably, the PUL found within *Muus Muus* faecal PUL was missing copies of GH50, GH105, GH154, and GH86 members.

4.4.4 *Phocaeicola* spp. MAGs contain the first characterized terrestrial GIT ulvan PUL

Ulvan degrading microorganisms have not been previously identified within the gut of terrestrial mammals. To enrich for genomes with ulvan degrading potential, fresh faecal samples were selectively cultivated in anaerobic *Bacteroides* minimalized media containing ulvan for 48 hours. Cultures were pooled and sequenced using Illumina and Nanopore sequencing. Enrichments confirmed the same UlvPUL from *Phocaeicola* spp. identified in the original metagenomes. Further, another potential ulvan-targeting region initially truncated due to the contig boundaries, was improved in the enriched sequencing (**Figure A3.9A**). This region contained an ulvan lyase (PL40), putative rhamnosidases (GH78, GH106, GH39), iduronidases (GH39), sulfatases and rhamnose metabolism genes. Intriguingly, SusC/D-like transport proteins were not found associated with this PUL in any of the sequence datasets, suggesting they are encoded upstream with the original ulvan PUL or ulvan oligosaccharides are enter through a promiscuous SusC/D-like transport system. This region shared high identity to a *Bacteroides mediterraneensis* MAG from a Korean human GIT metagenome (**Figure A3.9A**). The human-sourced PUL did not contain homologs for PspPL24, PspPL38, or PspPL40A enzymes; however, both PULs contained a GH105 module appended to an unknown Bacterial Neuraminidase Repeat domain (BNR). Curiously, within the CAZy database there are no known PL38 and PL42 members active on ulvan [115]. PspPL38 and PspPL42 partitioned away from all characterized sequences, forming distinct clades (**Figure A3.9B**).

Due to polyspecificity within the GH105 family [115], a SACCHARIS phylogeny was generated with all available GH105 sequences from the CAZy database [115] and PspGH105. PspGH105 clustered with GH105 sequences from marine bacteria shown to be active on unsaturated ulvan oligosaccharides (**Figure 4.5A**). The BNR domain aligned most closely with characterized PL24 and PL42 members (**Figure 4.5B**). Therefore, the full-length sequence is referred to as PspGH105;PL_{unk} here on out. We determined the X-ray crystal structure of PspGH105;Plunk at 2.45 Å resolution with two molecules in the asymmetric unit. Each monomer comprised a fusion of an (α/α_6) toroid domain and 7-bladed β-propeller (**Figure 4.5C**). The extensive interface between the two domains suggests a stable and rigid structure. The most similar structural homolog to the PspGH105

domain was the unsaturated β-glucuronyl hydrolase GH105 enzyme from *Nonlabens ulvanovorans* (PDB ID: 4CE7) [383]. A pairwise comparison gave a root mean square deviation of 2.5 Å while key residues in the active site are conserved, consisted with shared selectivity for removing unsaturated β-glucuronyl resides from the non-reducing terminus of ulvan chains (**Figure 4.5C**). While the PspPL$_{unk}$ domain shows fold similarity to ulvan active enzymes in PL families 24 and 42 the similarity ends at the shared 7-bladed β-fold. Prediction of the PspPL$_{unk}$ active site is unclear at this time.

To confirm the UlvPUL contained functionally active ulvanases, PspPL24, PspPL40 (PspPL40A & B), PspPL38, PspGH105;PL$_{unk}$, and the isolated PspPL$_{unk}$ domain were selected for biochemical characterization. All PLs and the PL$_{unk}$ module were confirmed by TLC & LC-ESI-MS to be active on ulvan (**Figure 4.5D; Figure A3.9C**). Each PL produced a distinct product profile with PspPL40A and PspPL38, and PspPL24 and PspPL$_{unk}$ displaying the most similarity to each other. The predicted L-rhamnose-α-1,4-D-glucuronate lyase (PspPL42) found within the upstream ulvan region was also shown to digest ulvan (**Figure A3.9D**), confirming novel catalytic specificities for PL42 and PL38 families.

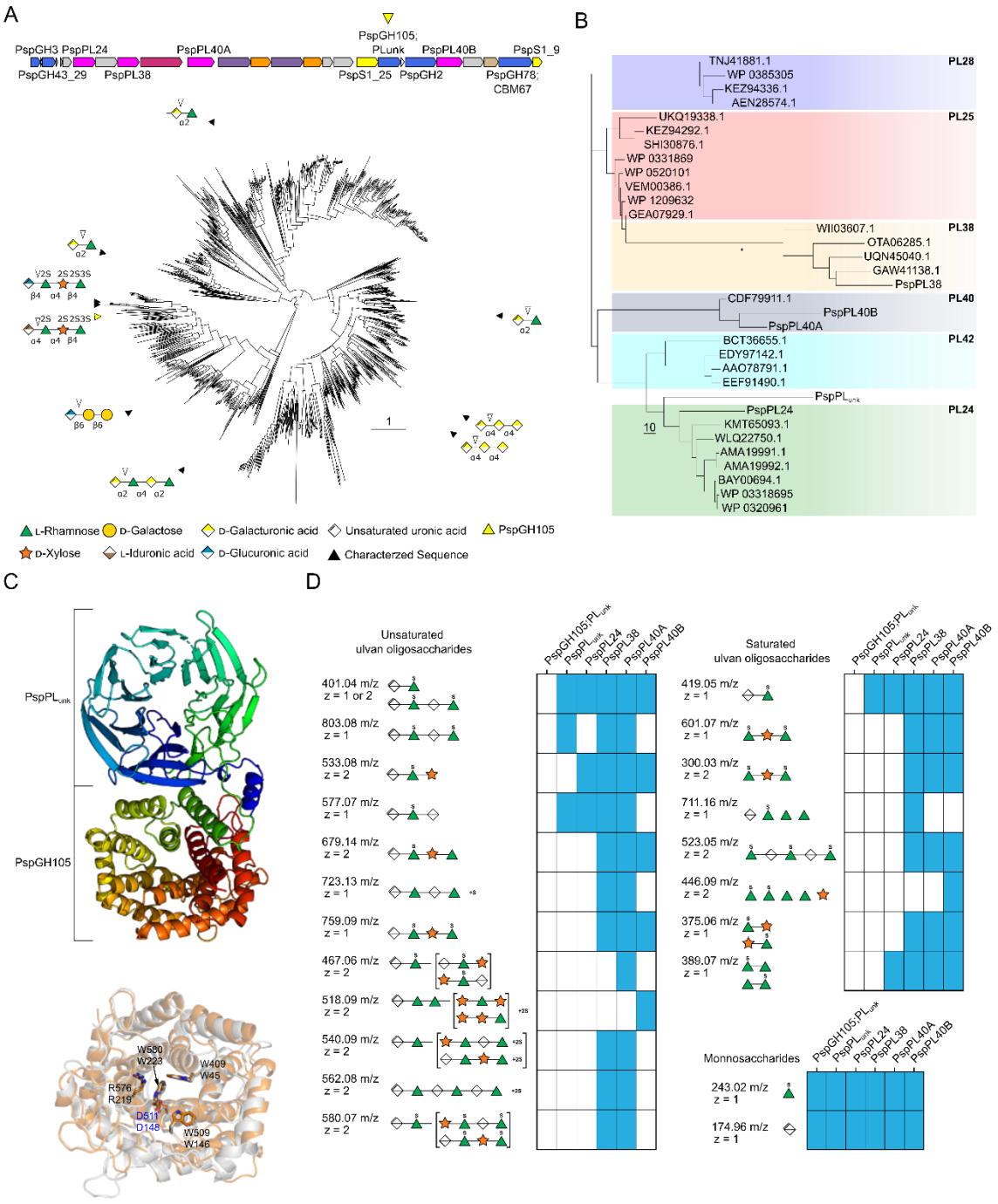

Figure 4.5: ***Phocaeicola*** **spp. UlvPUL employs numerous PL families for ulvan saccharification. A)** Structure of the UlvPUL (**top**). PspGH105 is indicated with an yellow arrow. **Bottom:** Phylogeny of all known GH105 members within the CAZy database. Characterized members are indicated using black arrows around the tree with their confirmed activities. Phylogeny was generated using SACCHARIS v2.0 [154]. **B)** RAxML-NG generated phylogeny of UlvPUL ulvan lyases, including PL_{unk}, and characterized ulvan lyases. The phylogeny was generated with 1,000 bootstraps. **C) Top:**

The structure of PspGH105;PL_{unk}. A color ramped cartoon illustration showing the fusion of (α/α6) toroid and 7-bladed β-propeller folds. **Bottom:** Overlap of the (α/α6) toroid GH105 domain with the known unsaturated β-glucuronyl hydrolase GH105 enzyme from *Nonlabens ulvanovorans* (NuGH105, PDB ID 4CE7). Residues in the GH105 domain of PspGH105;PL_{unk} are the top label and those from NuGH105 on the bottom. The catalytic acid base is labeled in blue. D) Product profiles of PspPL24, PspPL38, PspPL40A, PspPL40B, PspPL_{unk} and PspGH105 characterized by LC-ESI-MS. The table indicates products in the top 10% of detected ulvan oligosaccharides, including unsaturated, saturated, and monosaccharide products (**Table A3.12**).

4.5 Discussion

4.5.1 Adaptation of rewilded cattle within a temperate rainforest habitat

The Neolithic transition from subsistence foraging to industrialized societies has resulted in loss of human GIT microbial diversity. Hunter-gatherer and semi-nomadic tribes, such as the Hadza of Tanzania [384] and Yanomami of the Amazon rainforest in Venezuela and Brazil [385] have provided a window into ancient humans and the co-evolution of their microbiomes. In industrialized societies, microorganisms appear to be vanishing [386], likely as a result of diet, and exposure to modern medicines and sterilizing agents. In a similar manner, domestication of dairy and beef cattle involves formulated diets at different stages of production, and other restrictive management practices to optimize productivity. As a result, microbial diversity and function may have been lost during the transition from the free-roaming lifestyle of the Auroch. The rewilded *Muus Muus* have a grazer-browser lifestyle, which provides a unique glimpse into how natural forages shape the intestinal communities of cattle.

To explore the relationship between the *Muus Muus* and the land, a two-eyed seeing approach was employed to study herd dynamics and diet. Observing the unfettered grazing patterns in forest and shoreline pastures helped to identify unique plants and seaweeds that are differentially, and in some cases, seasonally consumed. Tla-o-qui-aht traditional ecological knowledge provided ancient insights into the benefits of foraging on many of these beneficial plants. These observations were weaved with western science using blinded *trnL-P6* sequencing for the first time in wild herbivores to catalog the traditional plant DNA detected in the feces of the *Muus Muus* (**Figure 4.3**). Together, these approaches confirmed the diversity and novelty of the *Muus Muus* diet, which includes a

mixture of exotic forest and marine plant forages not typically associated with cattle feeds (**Figure 4.4B, Figure A3.7B**). Interestingly, plant consumption was not always correlated with specific bacterial OTUs. This may result from the structural similarity of terrestrial plant cell walls and the plasticity of many GIT microorganisms endowed with a diverse inventory of catabolic pathways suitable for digesting of structural polysaccharides from both monocot and dicot species [1]. Interestingly, indirect correlations between Rubus species and methanogen abundance suggested some plants may have anti-methanogenic properties. This study has provided seminal evidence that trnL-P6 chloroplast marker sequencing is useful for monitoring herbivorous diets in the wild and could assist in future rewilding efforts.

4.5.2 The *Muus Muus* faecal microbiome contains rich catalytic potential and unique PUL inventories

The browser-grazer adaptations of the *Muus Muus* have resulted in enrichment of unique microorganisms and associated CAZomes. Both highly conserved PULs, such as starch and yeast α-mannans (**Figure A3.8**), and novel PULs, such as ulvan (**Figure 4.5**), were observed within the faecal MAGs. Unexpectedly, some of the most divergent CAZyme families were found in host glycan PULs that are predicted to be active on chondroitin sulfate, heparin sulfate, and hyaluronan. It is unclear if this observation results from poor sequence depth reported in previous ruminant datasets, whereby host-glycan catabolism genes were overlooked, or if host factors in the *Muus Muus* contribute to higher abundance of host-glycan foraging organisms. Host-glycan turnover catalyzed by the GIT microbiome is an important process for host health [387] and this interaction warrants further study.

The similarity of *Muus Muus* CAZymes and PULs to characterized and putative seaweed polysaccharide PULs including alginate, porphyran, and carrageenan, was most notable (**Figure A3.8**). The origins of these alginate and porphyran PULs are complex and distinct. AULs are highly conserved in the mammalian gut and proposed to have evolved by through an ancient horizontal gene transfer event [209, 221]. However, functional divergence in AULs has occurred through gene duplication and loss events [209]. Evolution of the porphyran systems, referred to as the “sushi factor” hypothesis, was

proposed to originate by HGT from marine bacteria [202]. Here we observe a PorPUL in the *Muus Muus* dataset that is highly conserved with human metagenomes (**Figure A3.8**) [205]. CarPULs can remain concealed as "latent traits" within the microbiomes of taxonomically and geographically distinct ruminants at low abundance, requiring enrichment for detection [258].

In contrast to what has been reported for brown and red seaweed polysaccharides, digestion of ulvan, the primary structural polysaccharide of green seaweed, had not been observed within a terrestrial GIT microbiome. The UlvPUL reported here, including the novel PspGH105;PL_{unk} enzyme, does not share any homology with other ruminant datasets. Interestingly, a portion of the UlvPUL showed high identity to an uncharacterized *B.* mediterraneensis MAG from a Korean human faecal sample (**Figure A3.9**) [388]. Determining the origins of the UlvPULs in the mammalian gut and whether they act as latent traits (requiring enrichment to detect) in diverse mammalian species will be the basis of future studies.

4.5.3 Rewilded cattle have a lower abundance and higher diversity of AMR gene content

Cattle raised in commercial production settings, including feedlot animals and those from natural and organic farms, displayed an abundance of AMR genes; most commonly, those associated with tetracycline and β-lactam resistance (**Figure 4.2A**) [369]. In comparison, the faecal metagenomes of the *Muus Muus* had radically different AMR gene profiles (**Figure 4.2A**). Overall, the *Muus Muus* metagenomes displayed much higher relative diversity and orders of magnitude less abundance of AMR genes. According to the oral tradition of the Tla-o-qui-aht peoples, the *Muus Muus* have been isolated on Meares Island since the 19th century, which is before the advent of AMU in cattle production (circa 1935) [366], and since, the herd has not been exposed to antimicrobials through medical intervention or their environment. Testing of freshwater samples surrounding the grazing sites of the herd confirmed there were no anthropocentric sources of antimicrobial compounds detectable (**Table A3.13**). An emerging trend is that the cattle GIT resistome is stable and there are longitudinal consequences for production animals as a result of vertical (dam-calf) and horizonal (herd behaviors) transmission [389]. This explains why

animals raised in natural and organic settings displayed AMR gene profiles more similar to conventional production settings (**Figure 4.2**) [369], despite not being exposed to antimicrobials during management. Persistence of the resistome, therefore, does not appear to require selection and can be exacerbated by animal history. Currently, multigenerational rearing of cattle in the absence on antimicrobials and reducing exposure to manure contaminated with AMR genes remains the primary approach for restoring the resistome to a more natural state [368]. Alternatively, finding sources of non-adulterated intestinal microbiomes, such as those observed in the *Muus Muus*, may accelerate attempts to rewild the microbiomes of production cattle through inter-herd faecal transplants.

4.6 Methods

4.6.1 Animals and Environment sampling

Cattle were identified through sexing and coloration patterns (**Fig. A3.1**). Faecal samples were collected from Opitsaht and the surrounding area through five sampling periods (September 2019 n=19, September 2021 n=12, December 2021 n=12, March 2022 n=12, and December 2022 n=10). Faecal samples were immediately frozen In LN_2 in the field and stored in LN2 vapor until stored at -80 °C. Faecal samples were randomly collected for metagenomic and metaproteomic analysis from the September 2019 collection, and 4 cattle were selected for longitudinal 16S rRNA and *trnL-P6* analysis. DNA from all samples were extracted using the DNeasy PowerSoil Pro kit (Qiagen, Canada) according to the manufacture's protocol.

On October 4th, 2024. Plant leaves and small stem attachments were collected fresh, and frozen immediately in liquid nitrogen on-site and shipped back to Lethbridge in a liquid nitrogen dry-shipper. Plants remained frozen at -40 °C until freeze-dried. Plants were freeze dried for 1 week and then stored in airtight tubes until further processing. Dried plant material was then ball milled for 4 min at 30 Hz, 3 times with a 5 minute cooling interval, for a finely milled material. Ball milled material was DNA extracted using the Qiagen Plant Mini Kit. Extracted DNA concentrations were checked using the Qubit Flex Fluorometer and the Invitrogen Qubit dsDNA HS Assay Kit and sent for sequencing at a minimum concentration of 6 ng uL^{-1}.

NovaSeq 6000 PE 150 bp was implemented for shotgun metagenomic sequencing, and amplicon sequencing was achieved using MiSeq 2500 PE 250 bp (16S rRNA; 515F/806R primers) and NextSeq PE 300 bp (16S rRNA; *trnL-P6* sequencing) via Génome Québec.

4.6.2 Amplicon sequencing analysis

The 16S rRNA gene sequences were split into individual samples through the internal barcode. Sequences were trimmed using Trimmomatic (v0.39) [271] and passed through Kraken2 (v2.1.2 SILVA NR99 database) [272] for annotation. Kraken-biome was used to create a biome file for Phyloseq (v1.48.0) [341]. Statistical analysis was done using the micoViz R package (v0.12.4) [320] and DESeq2 (1.48.1) [390], and visualized using dplyr (v1.1.4) [342], ggplot2 (v3.5.1) [343], and pheatmap (v1.0.12) [391].

Plant *trnL-P6* sequencing was done using primers previously characterized [376]. Alongside Meares cattle faecal samples, control animal faecal samples (n=4) and plants collected from Meares Island in duplicate (n=15x2; **Table A3.4**). Amplicon Sequence Variants (ASV) were generated using previously described methods [376] and QIIME2 [392]. ASVs were classified using the custom script Assign-Taxonomy-with-BLAST [https://github.com/Joseph7e/Assign-Taxonomy-with-BLAST] and NCBI BLAST+ [393]. Statistical analysis on the *trnL-P6* sequences were performed as above. The *trnL-P6* tree was generated by first generating alignments with MUSCLE (v5.1) [166], trimming bad alignments with trimAI (v1.5.0) [296] (-automated), generating a best model with ModelTest—NG (v0.1.7) [297] and a finally phylogenetic tree with RAxML-NG (v1.2.2) [298]. Trees were visualized and annotated using ITOL [350]. 16S rRNA and *trnL-P6* distance matrixes were generated using vegan (v2.6.10) [394]. and used for procrustes analysis. SPIEC-EASI was used to create a combined 16S rRNA and trnL network [395] which was visualized using Igraph [396] for **Figure A3.6**.

4.6.3 Shotgun sequencing analysis

Raw reads were quality trimmed using Trimmomatic (SLIDINGWINDOW:5:20) and assembled individually using MetaSPAdes (v3.13.0) [278], and biological replicates together in a co-assembly using MEGAHIT (v1.1.3) [287]. Contigs were binned and

refined using the MetaWRAP (v1.3.2) [279] binning/refinement modules, with metaBAT2 (v2.12.1) [282], MaxBin2 (v2.2.6) [281], and CONCOCT (v1.1.0) [280]. Bins/MAGs between individual assemblies and co-assemblies were merged and de-replicated with dRep (v3.0.0) [283] at 70% completion and <10% contamination. Bovine MAGs and ruminant isolates were taxonomically classified using GTDB-Tk (v2.3.2 - R202) [286] and CheckM2 (v1.1.3) [284], and were functionally annotated with DRAM (v1.3.4) [163], prodigal (v2.6.3) [349] and dbCAN3 [151]. PULs were identified through manual inspection of SusC/D-like pairings and PULpy [162].

MAGs were collected from selected ruminant GIT studies and dereplicated with the Meares Island cattle using dRep (compare -pa 0.9 -sa 0.95 -nc 0.3 –S_algorithm fastANI). The alignment of Meares Island MAGs with other ruminant MAGs within primary and secondary clusters was converted to a true/false table and used to generate the UpsetR [397] plot in **Figure 4.4**. Previously published metagenomic assembled genomes from the ruminant GIT and the CAZy database (downloaded 04-2022) [115] were run through prodigal and dbCAN3 and all duplicates within each dataset were removed using CD-HIT (95% identity) [398]. CAZymes were blasted against the Meares Island CAZome (dereplicated in a similar manner) using Diamond [290]. BLAST hits above 60% identity and 50 Bitscore were selected for the resulting true/false table and UpsetR plot in **Figure A3.7**.

Metagenomic reads from Meares Island cattle and AMR-based metagenomic studies (PRJNA379303 [369] & PRJNA420682 [373]) were run through AMR++ and the MEGARes v2.0 [399] database to identify AMR genes within the datasets. Single nucleotide polymorphisms and regulators were removed from the analysis [373]. Statistics were run using DESeq2 (mention normalization method) [390], while adding a single count to each AMR hit to avoid zero values. Pheatmap and EnhancedVolcano (v1.26.0) [400] were used to generate figures.

4.6.4 Ulvan PUL CAZyme production and characterization

All PLs and GHs were codon optimized and synthesized with Biobasic. Expression plasmids were transformed into BL21 (DE3) Tuner competent cells (Novagen) and grown in LB Miller broth containing kanamycin (50 ug mL^{-1}). Cell cultures were grown to an

optical density of 0.6-0.8 at OD_{600} nm and induced with isopropyl β-D-1-thiogalactopyranoside at a final concentration of 1 mM at 16 °C overnight while shaking at 200 rpm. Cells were lysed using a combination of lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 0.1 mg mL-1 lysozyme) and sonication (2 min of 1 s intervals of medium intensity sonic pulses at a power setting of 4.5 – Heat systems Ultrasonics Model W-225). Cell lysate was centrifuged at 17,500 × g for 45 min and purified using Ni-NTA resin columns (Cytiva) and immobilized metal affinity chromatography. Samples were buffer exchange with 20 mM Tris, 500 mM NaCl, and 1% glycerol using ultrafiltration column, and further purification by size exclusion chromatography HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare). All fractions were confirmed using sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

Optimum pH for the PLs was determined by measuring the absorption of the C4-C5 unsaturation generated via the PL β-elimination reaction using the SpectraMax ID3 (Molecular Devices) at 232 nm. Reactions were run for 10 min while continuously monitoring absorbance at 232 nm at 30 s intervals for 5 min in selected buffers (citrate phosphate pH 3-8; bicine 8-9, CHES 10). Activity was plotted using GraphPad Prism (v10.2.3). Future reaction digests were completed at an optimum pH. Digests of ulvan (elycityl) were conducted overnight with 5 mg mL^{-1} substrate and were boiled at 90 °C for 10 min to complete the reaction.

The $PspPL_{unk}$ domain of the UlvPUL PL_{unk};GH105 was subcloned into PET28a using Nde1 and Xho1 subsites and expressed as stated above. Subcloning of the PspGH105 domain revealed that this domain is entirely insoluble even after refolding attempts.

Digests were separated using thin layer chromatography on silica gel plates (Millipore) in a 1-butanol: distilled water: acetic acid (2:1:1 v/v/v) mobile phase and visualized with an ethanol: sulfuric acid (70:3 v/v) solution containing 1% (w/v) orcinol monohydrate (Sigma) and heated at 150 °C for ~3 min. LC-ESI-MS was performed on a Vanquish UHPLC system (Thermo Scientific). Separation of the carrageenan oligosaccharides was achieved using an Acquity HILIC Column, 130 Å, 1.7 µm, 2.1 mm × 150 mm (Waters) at a flow rate of 300 µL min^{-1} at 30 °C, using a gradient as described in **Table A3.14**.

Ulvan digests were purified through solid phase extraction after elution in 50% acetonitrile and 0.1% trifluoroacetic acid, dried and resuspended in LC-MS grade water, and injected in a volume of 10 µL at a concentration of 200 µg mL^{-1} for ESI-MS on an Orbitrap Fusion Tribrid system (Thermo Scientific) in negative ion mode. Mass spectra parameters are shown in (**Table A3.15**). To select ions for MS2 experiments, an intensity threshold filter was employed with a minimum intensity of 2,500 and maximum intensity of 1E+20, relative intensity threshold of 20%, and a dynamic exclusion filter was used after 1 times for 2.5 s with default mass tolerance values. HCD was employed to generate fragments. Raw data were processed using Xcalibur and Freestyle software packages (Thermo Scientific) and the R packages xcms (version 4.6.0) [401-403], MSnbase (version 2.34.0) [404, 405], and BiocParallel (version 1.42.0) [406]. Due to the complexity of the samples, MS2 spectra were filtered based on the presence of any number of diagnostic fragments to identify species most likely to correspond to ulvan oligosaccharides within a 20 ppm tolerance: 96.9594, 138.9697, 153.9126, 173.4474, 85.0288, 150.1312, or 157.0131 m/z. Precursor species that passed the filter and also corresponded to the top 10% of species by integrated peak intensities were identified by manual interpretation and comparison to theoretical precursor and fragment masses calculated using GlycoWorkBench [407].

4.6.5 Crystallography

4.6.5.1 Protein Expression

The plasmid encoding PspGH105;PL_{unk} was transformed into *E. coli* BL21 competent cells. A single colony was used to inoculate precultures for overnight growth at 37 °C and shaking at 170 rpm in 25 mL LB containing 50 ug mL^{-1} kanamycin. The precultures were used to inoculate 2 L of 2xYT media, which were incubated with shaking at 37 °C and 170 rpm until an optical density of 0.6 at OD_{600} nm was reached. This was followed by one hour incubation at 16 °C and 170 rpm. Protein production was induced with the addition of isopropyl β-L-1-thioglalctopyranoside to a final concentration of 500 µM with overnight incubation at 16 °C and 170 rpm.

The cultures were pelleted at by centrifugation in an Avanti J-25I Centrifuge (Beckman Coulter) at 5000 × g for 10 min. Chemical lysis was performed according to the following protocol. The cell pellet was resuspended in 10 ml sucrose solution (35% sucrose, 50mM Tris pH 8) followed by the addition of 10 mg lysozyme and stirred for 20 minutes with a magnetic stirrer. 20 mL of 500 mM Deoxycholate solution (1% deoxycholate, 1% TritonX100, 50 mM Tris-HCl pH 8, 100mM NaCl) was added and stirred for 15 min. Lastly 150 µL 1 M $MgCl_2$ and 90 µL 2mg mL^{-1} DNase I was added. The solution was centrifuged at 4 °C at 15000 rpm for 45 min and the resulting clarified supernatant was retained.

4.6.5.2 Protein Purification

PspGH105;PL_{unk} was purified from clarified cell extracts by immobilized metal affinity chromatography with Ni^{2+}-NTA resin equilibrated with binding buffer (20 mM TrisHCl pH 8.0, 500 mM NaCl, 10% glycerol). The lysate was loaded onto the column resin, and washed with binding buffer and 20 mM imidazole were performed to eliminate non-target protein binding. The His-tagged target protein was eluted with 500 mM imidazole. Eluted protein solutions were dialyzed overnight in binding buffer to reduce the imidazole content. Both proteins were additionally dialyzed with thrombin and 2.5 mM $CaCl_2$ present to cleave the histidine tag. The dialyzed proteins were concentrated to 2 mL using a 10 kDa amicon membrane and further purified by size exclusion chromatography was performed using a Sephacryl S-200 HR column (GE Healthcare). Absorbances at 280 nm were used to calculate the protein concentration with calculated molar extinction coefficients.

4.6.5.3 Crystallization and data collection

Seven commercial crystal screens (PEG-ION, Index, Crystal Screen, MCSG-I, MCSG-II, MCSG-III, MCSSG-IV) from Hampton Research were utilized to test crystallization conditions using the sitting drop method. PspGH105;PL_{unk} at 20 g L^{-1} was placed in each of 96 wells at a 1:1 ratio with the 0.5 µl of the screen condition. Optimization of initial hits were performed using the hanging drop method. Crystals of PspGH105;PL_{unk} was optimized in 5% pentaerythritol propoxylate (vol/vol), 24% PEG MME 2000 (w/v),

0.05 M KBr. 25% Ethylene glycol (vol/vol) in crystallization solution was used as a cryoprotectant. Crystals were cryo-cooled directly in a nitrogen stream at 100K prior to data collection. Diffraction data for PspGH105;PL_{unk} was collected on CMCF-ID beamline at the Canadian Light Source (CLS, Saskatoon, Saskatchewan). This data was processed using XDS (PMID: 20124692) and Aimless (PMID: 23793146). Data statistics are given in **Table A3.16**.

4.6.5.4 Structure solution and refinement

The AlphaFold 2 [408] structure of PspGH105;PL_{unk} was used for molecular replacement search models for initial structure determination using PHASER [353]. Solution for PspGH105;PL_{unk} was optimized by splitting the search model into separate N- and C-terminal domains and searching independently with these models. The CCP4 Refmac was used for initial refinement [409]. Further structure refinement was performed manually in COOT [355]. Additional refinement was performed using Phenix refine, including the addition of waters. Structure validation was confirmed using MolProbity [410]. Final model statistics are given in **Table A3.16**.

4.6.6 *Bacteroides* culture enrichment and isolation

Faecal samples were taken and inoculated directly into anaerobic media within sealed Hungate tubes (atmosphere: 85% N2, 10% CO2, 5% H2). 1 g of faecal samples were diluted in 10 mL of anaerobic PBS and vortexed to homogenize on-site. 100 µL of homogenized mixture was inoculated in 1:1 *Bacteroides* MM [200] supplemented with 1% Bacto™ Tryptone (BD), 0.5% Yeast Extract Bacteriological (VWR), and 0.5% ulvan. Cultures were grown until sufficient density was reached and re-inoculated into fresh media twice before being glycerol stocked and placed into an anaerobic chamber with similar atmosphere. Cultures were diluted 10-3 and consecutively streaked on minimalized media agar plates containing 0.5% ulvan and 1% agar (4-5 times) to obtain pure isolates. Isolate DNA was extracted using the Qiagen Powersoil kit and was sequenced using Illumina (NovaSeq 6000 PE 150 bp), and Oxford Nanopore (R9.4.1; SQK-RBK110-96) sequencing. Reads were trimmed as above using Trimmomatic, and hybrid assembled

using Unicyler (v0.4.8) [346]. Isolate genomes were checked for quality using Quast (v5.0.2) [347].

4.7 Acknowledgements

Research at AAFC was supported through an AgriScience Project (ASP-207; J-002817) and Prize for Outstanding Achievement in Science (AAFC) awarded to D.W.A. (J-003135). LC-ESI-MS data was collected at the University of Lethbridge, Lethbridge Central Analytical Facility, with support from Tony Montina, Vincent Weiler, Maurice Needham, and Carl Holland. Insightful bioinformatic suggestions were provided by Rodrigo Ortega Polo at Agriculture & Agri-Food Canada Lethbridge.

Chapter 5 Conclusions

5.1 Outlook

The adaptation of livestock to novel feedstocks may provide a solution to pressures associated with climate change and land use competition with food crops. Foraging on wild or cultivated seaweeds represent novel sources of feed ingredients for ruminant livestock; however, how and at what rate ruminant microbiomes adapt to the introduction of exotic feeds and polysaccharides may dictate the overall feasibility of their incorporation into production systems.

Interestingly, although the systems studied here represent isolated datasets, the similarity between alginate, carrageenan, and ulvan PULs within ruminant and non-ruminant microbiomes suggest that these saccharification pathways are ubiquitous even amongst the landlocked *Muus Muus* living on Meares Island. One of the initial objectives of this thesis was to discern the timescale of saccharification pathway gene transfer events within ruminant microorganisms; however, it is likely that these events did not happen *ab initio* within the span of these studies nor on human timescales. Importantly, however, differential evolutionary mechanisms were observed in carrageenan and alginate studies. AULs within ruminant sourced *Bacteroidota* appear to be ubiquitous (**Figure 2.4**) and well ameliorated into the genome in regards to GC content (**Figure 2.5**). The origin of these pathways is unclear; for example, they may have been transferred during an ancient event that is no longer detectable through conventional gene-species analysis or arose originally to digest bacterial alginate. On the contrary, although there appears to be evidence of CarPULs present within geographically distinct populations (**Figure 3.4**), this evidence was often observed only through read mapping of CarPUL genes. In this study, the enrichment of CarPULs through controlled feed supplementation or bacterial culture enrichments enabled the detection of these complex systems; otherwise, it is likely that CarPULs are less prevalent within ruminant microbiomes than AULs. I conclude that these systems are latent traits present within microorganisms or collections of microorganisms, especially generalist *Bacteroidota* species, that are maintained within populations and present at low levels. CarPUL origins, although clouded in mystery, appear to involve marine herbivorous fish *Rikenellaceae* spp. (**Figure 3.5**), supported by the distinct GC in comparison to the remainder of the *Bacteroidota* genome (**Figure A2.9**). It is unclear if latent traits first originated within a single terrestrial mammal (ruminants, humans) and

transferred due to cross-species interactions including farming [411], or if these regions occurred during separate and distinct events. The UlvPUL found within Meares Island faecal metagenomes is far less ubiquitous than either the AUL or CarPULs identified within this study. However, the identification of similar genetic signatures between the UlvPUL and MAGs assembled within the Korean human gut microbiome (**Figure A3.9**) suggests that this region may also act as a latent trait. In this case, deeper sequencing of geographically distinct microbiomes or dietary enrichment will be needed to confirm this hypothesis.

The *Muus Muus* have adapted to grazing traditional plants growing in the old growth rainforest and shorelines of Meares Island. These unique foraging behaviors have enriched their faecal microbial communities to harbour vast amounts of unique CAZymes targeting seaweed polysaccharides. However, despite the unique diet of the Muus Muus, several highly conserved PULs were also detected that appear to be core polysaccharide utilization pathways maintained through the *Bovidae family*, other ruminants like buffalo, and even non-ruminant terrestrial mammals. PUL conservation extended from terrestrial plant cell wall polysaccharides, like starch and pectins and hemicellulose, to fungal cell wall polysaccharides (yeast α-mannan), and seaweed polysaccharides including alginate and porphyran/agarose. The latter of which was initially found within Japanese human gut microbiomes [222]. The identity that exists within these systems found in different hosts species suggests their origin predates the arrival of the *Muus Muus* on Meares Island (1901). The ancestor of bovines, the auroch, were predicted to have a diet similar to that of *Muus Muus*, including shoreline foraging and grazing [412], with browsing during times dietary constraints [361, 413]. It is likely the microbiome composition of the extinct auroch may never be determined, and therefore, other approaches like de-extinction and rewilding help to provide clues into these ancient host-microbiome interactions. In this manner, studying the microbiome of the *Muus Muus*, in comparison to production cattle GIT microbiomes, is helpful for comparing terrestrial and marine polysaccharide degradation pathways and to determined which saccharification pathways are formed due to recent events, or are considered “core” metabolism pathways.

5.2 Future directions

In these studies, it became abundantly clear that marine and marine-sourced microbiomes such as those seen in marine sediment [414], water [414], seaweed [415], and herbivorous animals including fish [330] house vastly diverse, distinct, and temporally variant microbiomes and degradation systems; all of which may host homologous genes to those found within terrestrial microbiomes. This study was largely unable to identify the exact origins of the terrestrial seaweed degrading genes within the expansive marine-sourced microbiomes. Although this maybe a result of sequence divergence due to long ameliorated gene transfer events observed within terrestrial microbiomes, it is also likely that the expansive and diverse nature of marine-sourced microbiomes still harbor highly identical degradation systems. These environments are largely under-sequenced and thorough sequencing of marine microbiomes exposed to different classes of seaweed may lead to the uncovering of regions of homology between marine and terrestrial microbiomes.

The implementation of seaweed or other exotic forages in ruminant diets requires further study to determine their environmental, health, and financial impacts. Currently the environmental impacts of exotic forages is a developing area of research. *S. latissima* and *M. japonica* have been studied for the potential methane reducing properties at various concentrations. *M. japonica* showed no significant reducing in methane production at 2% inclusion *in vivo* [416] while *S. latissima* showed reduction of methane, albeit at a much higher 25% inclusion *in vitro* [417], *Ulva* spp. have been shown to reduce methane *in vitro* [40] but it is unclear if *U. intestinalis* would provide the same results. It should be noted that *U. intestinalis* does contain bromoform [418]; therefore, future research on *U. intestinalis* may show promise. Importantly, the aforementioned studies did not demonstrate detrimental effects on rumen fermentation, suggesting that regardless of their potential, or lack thereof, to reduce methane, they still posses nutritional value. Future directions towards this end would be to expand feeding trials to alternative seaweeds and other unique dietary supplements. Terrestrial GIT microbiomes have an incredible ability to adapt to novel feedstuffs, including seaweed, artificial feed additives like xanthan gum [256], and non-polysaccharide structures like lignin [419], and synthetic polyesters [420]. Currently ongoing research will aid in determining if other traditional forages from Meares

Island, including *Rubus* spp. which indirectly correlated with reduced *Methanobacteriaceae* spp., can be used to reduce methane emissions (**Figure A3.5**).

Other impacts of seaweed and exotic forage supplementation must be addressed including health impacts of seaweed polysaccharides and other seaweed compounds. Numerous antioxidants and beneficial compounds have been identified within seaweeds; however, it is unclear if these have a significant impact on animal health [225]. For instance, there are conflicting reports on the effects of sulfated red seaweed polysaccharides on inflammation, with studies showing both pro- [421] and anti-inflammatory effects [422]. These studies will need to be validated in animal models at standardized inclusion percentages. Further, the use of seaweed as natural antimicrobials may hold promise in the reduction of medicinal antimicrobials used in conventional cattle systems [423]. The rewilding of domesticated animals holds promise to improve biodiversity and help mitigate mounting pressures from climate change and the AMR. In this regard, the *Muus Muus* herd represents a remarkable example of health and resilience in the absence of antimicrobial intervention and conventional cattle management. Opening natural foraging habitats to domesticated livestock may provide an unexplored solution to the cattle industry to improve animal health and welfare and lower the requirement for antimicrobial use. However, these studies are in their infancy and are a much bigger initiative than feed supplementation. Further weaving of traditional knowledge, rewilding practices and traditional cattle production may prove challenging.

Finally, the viability of using seaweed and other exotic forages as alternative feedstocks will require tailored formulation and feeding approaches for distinct production systems. Mariculture techniques and costs, transportation costs, mariculture and transportation emissions, and any potential financial benefits (market value and subsidies) will all play important factors on the viability of seaweed use in livestock production systems [424]. A key contribution of this thesis is that I demonstrate ruminant microorganisms can digest to distinct seaweed polysaccharides, suggesting that seaweed mariculture can be finetuned to geographic constraints and local mariculture crops.

5.3 Final thought

This thesis has explored in detail the adaptation and evolution of ruminant microbiomes to saccharify diverse and exotic forages, particularly seaweed polysaccharides. This was achieved using multiple experimental systems including 1) artificial RUSITEC trials 2) controlled cattle trials and 3) faecal collections from wild cattle. Detailed metagenomic analysis was performed on each system to detect putative saccharification pathways for the corresponding exotic forages –brown, red, or green seaweed – alongside comparative analysis to ruminant and non-ruminant databases to speculate about the origins of these pathways. These predictive analyses were supported through biochemical analysis of putative CAZymes active on alginate (brown seaweed), carrageenans (red seaweed), and ulvans (green seaweed), and involved critical collaborator contributions, including, but not limited to, glycosidic linkage analysis, metaproteomics analysis, and x-ray protein crystallography. Together these results demonstrate using complementary approaches and comprehensive evidence that the cattle microbiome can saccharify a diverse range of seaweed cell wall polysaccharides.

References

[1] Tingley JP, Low KE, Xing X, Abbott DW. Combined whole cell wall analysis and streamlined in silico carbohydrate-active enzyme discovery to improve biocatalytic conversion of agricultural crop residues. Biotechnology for Biofuels. 2021;14:16.

[2] Russell JB, Rychlik JL. Factors That Alter Rumen Microbial Ecology. Science. 2001;292:1119-22.

[3] Van Soest PJ. Nutritional ecology of the ruminant: Cornell university press; 1994.

[4] Liu K, Zhang Y, Yu Z, Xu Q, Zheng N, Zhao S, et al. Ruminal microbiota-host interaction and its effect on nutrient metabolism. Anim Nutr. 2021;7:49-55.

[5] Caballero-Flores G, Pickard JM, Núñez G. Microbiota-mediated colonization resistance: mechanisms and regulation. Nature Reviews Microbiology. 2023;21:347-60.

[6] Arshad MA, Hassan FU, Rehman MS, Huws SA, Cheng Y, Din AU. Gut microbiome colonization and development in neonatal ruminants: Strategies, prospects, and opportunities. Anim Nutr. 2021;7:883-95.

[7] Malmuthuge N, Liang G, Guan LL. Regulation of rumen development in neonatal ruminants through microbial metagenomes and host transcriptomes. Genome Biology. 2019;20:172.

[8] Li RW, Connor EE, Li C, Baldwin Vi RL, Sparks ME. Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. Environ Microbiol. 2012;14:129-39.

[9] STEVENS CE, HUME ID. Contributions of Microbes in Vertebrate Gastrointestinal Tract to Production and Conservation of Nutrients. Physiological Reviews. 1998;78:393-427.

[10] Moraïs S, Mizrahi I. The Road Not Taken: The Rumen Microbiome, Functional Groups, and Community States. Trends in Microbiology. 2019;27:538-49.

[11] Weimer PJ, Cox MS, Vieira de Paula T, Lin M, Hall MB, Suen G. Transient changes in milk production efficiency and bacterial community composition resulting from near-total exchange of ruminal contents between high- and low-efficiency Holstein cows. Journal of Dairy Science. 2017;100:7165-82.

[12] Zhou M, Peng Y-J, Chen Y, Klinger CM, Oba M, Liu J-X, et al. Assessment of microbiome changes after rumen transfaunation: implications on improving feed efficiency in beef cattle. Microbiome. 2018;6:62.

[13] Wallace RJ, Sasson G, Garnsworthy PC, Tapio I, Gregson E, Bani P, et al. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci Adv. 2019;5:eaav8391.

[14] Henderson G, Cox F, Ganesh S, Jonker A, Young W, Abecia L, et al. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Scientific Reports. 2015;5:14567.

[15] Terry SA, Badhan A, Wang Y, Chaves AV, McAllister TA. Fibre digestion by rumen microbiota — a review of recent metagenomic and metatranscriptomic studies. Canadian Journal of Animal Science. 2019;99:678-92.

[16] Mizrahi I, Wallace RJ, Moraïs S. The rumen microbiome: balancing food security and environmental impacts. Nature Reviews Microbiology. 2021;19:553-66.

[17] De Mulder T, Goossens K, Peiren N, Vandaele L, Haegeman A, De Tender C, et al. Exploring the methanogen and bacterial communities of rumen environments: solid adherent, fluid and epimural. FEMS Microbiology Ecology. 2016;93.

[18] Stockler RM, Hallowell H, Higgins KV, Groover ES, Hiltbold EM, Newcomer B, et al. Characterization and Comparison of the Rumen Luminal and Epithelial Microbiome Profiles Using Metagenomic Sequencing Technique. Front Vet Sci. 2022;9:799063.

[19] Culp EJ, Goodman AL. Cross-feeding in the gut microbiome: Ecology and mechanisms. Cell Host Microbe. 2023;31:485-99.

[20] Mao S, Zhang M, Liu J, Zhu W. Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. Scientific Reports. 2015;5:16116.

[21] Beauchemin KA, Ungerfeld EM, Eckard RJ, Wang M. Review: Fifty years of research on rumen methanogenesis: lessons learned and future challenges for mitigation. Animal. 2020;14:s2-s16.

[22] Patra A, Park T, Kim M, Yu Z. Rumen methanogens and mitigation of methane emission by anti-methanogenic compounds and substances. Journal of Animal Science and Biotechnology. 2017;8:13.

[23] Sharp R, Ziemer CJ, Stern MD, Stahl DA. Taxon-specific associations between protozoal and methanogen populations in the rumen and a model rumen system. FEMS Microbiology Ecology. 1998;26:71-8.

[24] Andersen TO, Altshuler I, Vera-Ponce de León A, Walter JM, McGovern E, Keogh K, et al. Metabolic influence of core ciliates within the rumen microbiome. The ISME Journal. 2023;17:1128-40.

[25] Lin L, Lai Z, Zhang J, Zhu W, Mao S. The gastrointestinal microbiome in dairy cattle is constrained by the deterministic driver of the region and the modified effect of diet. Microbiome. 2023;11:10.

[26] Xie F, Jin W, Si H, Yuan Y, Tao Y, Liu J, et al. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. Microbiome. 2021;9:137.

[27] Kamenova S, de Muinck EJ, Veiberg V, Utsi TA, Steyaert SMJG, Albon SD, et al. Gut microbiome biogeography in reindeer supersedes millennia of ecological and evolutionary separation. FEMS Microbiology Ecology. 2023;99.

[28] Xue D, Chen H, Luo X, Guan J, He Y, Zhao X. Microbial diversity in the rumen, reticulum, omasum, and abomasum of yak on a rapid fattening regime in an agro-pastoral transition zone. J Microbiol. 2018;56:734-43.

[29] Kastl AJ, Jr., Terry NA, Wu GD, Albenberg LG. The Structure and Function of the Human Small Intestinal Microbiota: Current Understanding and Future Directions. Cell Mol Gastroenterol Hepatol. 2020;9:33-45.

[30] Popova M, McGovern E, McCabe MS, Martin C, Doreau M, Arbre M, et al. The Structural and Functional Capacity of Ruminal and Cecal Microbiota in Growing Cattle Was Unaffected by Dietary Supplementation of Linseed Oil and Nitrate. Frontiers in Microbiology. 2017;Volume 8 - 2017.

[31] Gressley TF, Hall MB, Armentano LE. RUMINANT NUTRITION SYMPOSIUM: Productivity, digestion, and health responses to hindgut acidosis in ruminants1. Journal of Animal Science. 2011;89:1120-30.

[32] Staaland H, White RG. Influence of foraging ecology on alimentary tract size and function of Svalbard reindeer. Canadian Journal of Zoology. 1991;69:1326-34.

[33] González LA, Manteca X, Calsamiglia S, Schwartzkopf-Genswein KS, Ferret A. Ruminal acidosis in feedlot cattle: Interplay between feed ingredients, rumen function and feeding behavior (a review). Animal Feed Science and Technology. 2012;172:66-79.

[34] Janssen PH. Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. Animal Feed Science and Technology. 2010;160:1-22.

[35] Duin EC, Wagner T, Shima S, Prakash D, Cronin B, Yáñez-Ruiz DR, et al. Mode of action uncovered for the specific reduction of methane emissions from ruminants by the small molecule 3-nitrooxypropanol. Proceedings of the National Academy of Sciences. 2016;113:6172-7.

[36] Lieber DJ, Catlett J, Madayiputhiya N, Nandakumar R, Lopez MM, Metcalf WW, et al. A Multienzyme Complex Channels Substrates and Electrons through Acetyl-CoA and Methane Biosynthesis Pathways in Methanosarcina. PLOS ONE. 2014;9:e107563.

[37] Alemu AW, Shreck AL, Booker CW, McGinn SM, Pekrul LKD, Kindermann M, et al. Use of 3-nitrooxypropanol in a commercial feedlot to decrease enteric methane emissions from cattle fed a corn-based finishing diet. Journal of Animal Science. 2020;99.

[38] Wood JM, Kennedy FS, Wolfe RS. Reaction of multihalogenated hydrocarbons with free and bound reduced vitamin B12. Biochemistry. 1968;7:1707-13.

[39] Roque BM, Venegas M, Kinley RD, de Nys R, Duarte TL, Yang X, et al. Red seaweed (Asparagopsis taxiformis) supplementation reduces enteric methane by over 80 percent in beef steers. PLOS ONE. 2021;16:e0247820.

[40] Machado L, Magnusson M, Paul NA, de Nys R, Tomkins N. Effects of Marine and Freshwater Macroalgae on In Vitro Total Gas and Methane Production. PLOS ONE. 2014;9:e85289.

[41] Kinley RD, Martinez-Fernandez G, Matthews MK, de Nys R, Magnusson M, Tomkins NW. Mitigating the carbon footprint and improving productivity of ruminant livestock agriculture using a red seaweed. Journal of Cleaner Production. 2020;259:120836.

[42] Cabrita MT, Vale C, Rauter AP. Halogenated Compounds from Marine Algae. Marine Drugs. 2010;8:2301-17.

[43] Roque BM, Brooke CG, Ladau J, Polley T, Marsh LJ, Najafi N, et al. Effect of the macroalgae Asparagopsis taxiformis on methane production and rumen microbiome assemblage. Animal Microbiome. 2019;1:3.

[44] Osman MEH, Abo-Shady AM, Elshobary ME, Abd El-Ghafar MO, Abomohra AE-F. Screening of seaweeds for sustainable biofuel recovery through sequential biodiesel and bioethanol production. Environmental Science and Pollution Research. 2020;27:32481-93.

[45] Vijn S, Compart DP, Dutta N, Foukis A, Hess M, Hristov AN, et al. Key Considerations for the Use of Seaweed to Reduce Enteric Methane Emissions From Cattle. Frontiers in Veterinary Science. 2020;7.

[46] Cosgrove D, Jarvis m. Comparative structure and biomechanics of plant primary and secondary cell walls. Frontiers in Plant Science. 2012;Volume 3 - 2012.

[47] Carpita NC, Gibeaut DM. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J. 1993;3:1-30.

[48] Carpita NC. Structural and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol. 1996;47:445-76.

[49] Caffall KH, Mohnen D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydrate Research. 2009;344:1879-900.

[50] Terrett OM, Dupree P. Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. Curr Opin Biotech. 2019;56:97-104.

[51] Pfeifer L, Shafee T, Johnson KL, Bacic A, Classen B. Arabinogalactan-proteins of Zostera marina L. contain unique glycan structures and provide insight into adaption processes to saline environments. Scientific Reports. 2020;10:8232.

[52] Seeberger PH. Monosaccharide Diversity. 3rd ed: Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 2015.

[53] El Khadem H. Carbohydrate chemistry: monosaccharides and their oligomers: Elsevier; 2012.

[54] Cumpstey I. Chemical modification of polysaccharides. ISRN Org Chem. 2013;2013:417672.

[55] Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lütteke T, et al. Updates to the Symbol Nomenclature for Glycans guidelines. Glycobiology. 2019;29:620-4.

[56] O'Neill MA, Ishii T, Albersheim P, Darvill AG. Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide. Annu Rev Plant Biol. 2004;55:109-39.

[57] Anderson CT, Kieber JJ. Dynamic construction, perception, and remodeling of plant cell walls. Annu Rev Plant Biol. 2020;71:39-69.

[58] Pettolino FA, Walsh C, Fincher GB, Bacic A. Determining the polysaccharide composition of plant cell walls. Nat Protoc. 2012;7:1590-607.

[59] Vogel J. Unique aspects of the grass cell wall. Curr Opin Plant Biol. 2008;11:301-7.

[60] Pattathil S, Hahn MG, Dale BE, Chundawat SPS. Insights into plant cell wall structure, architecture, and integrity using glycome profiling of native and AFEXTM-pre-treated biomass. J Exp Bot. 2015;66:4279-94.

[61] Mitchell RAC, Dupree P, Shewry PR. A novel bioinformatics approach identifies Candidate genes for the synthesis and feruloylation of arabinoxylan. Plant Physiol. 2007;144:43-53.

[62] Ishii T. Structure and functions of feruloylated polysaccharides. Plant Sci. 1997;127:111-27.

[63] Ebringerová A, Hromádková Z, Heinze T. Hemicellulose. In: Heinze T, editor. Polysaccharides I Adv Polym Sci. Berlin, Heidelberg: Springer; 2005.

[64] Hatfield RD, Wilson JR, Mertens DR. Composition of cell walls isolated from cell types of grain sorghum stems. J Sci Food Agric. 1999;79:891-9.

[65] O'Neill MA, York WS. The composition and structure of plant primary cell walls. In: J.A. R, editor. Annual Plant Reviews. Boca Raton, FL: CRC Press; 2003. p. 1-54.

[66] Grantham NJ, Wurman-Rodrich J, Terrett OM, Lyczakowski JJ, Stott K, Iuga D, et al. An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. Nat Plants. 2017;3:859-65.

[67] Smith PJ, Wang HT, York WS, Pena MJ, Urbanowicz BR. Designer biomass for next-generation biorefineries: leveraging recent insights into xylan structure and biosynthesis. Biotechnol Biofuels. 2017;10:286.

[68] Pustjens AM, Schols HA, Kabel MA, Gruppen H. Characterisation of cell wall polysaccharides from rapeseed (Brassica napus) meal. Carbohydr Polym. 2013;98:1650-6.

[69] Wang J, Bai J, Fan M, Li T, Li Y, Qian H, et al. Cereal-derived arabinoxylans: Structural features and structure–activity correlations. Trends Food Sci Tech. 2020;96:157-65.

[70] Knudsen KEB. Fiber and nonstarch polysaccharide content and variation in common crops used in broiler diets1. Poult Sci. 2014;93:2380-93.

[71] Malunga LN, Beta T. Isolation and identification of feruloylated arabinoxylan mono- and oligosaccharides from undigested and digested maize and wheat. Heliyon. 2016;2:e00106.

[72] Huisman MMH, Schols HA, Voragen AGJ. Glucuronoarabinoxylans from maize kernel cell walls are more complex than those from sorghum kernel cell walls. Carbohydr Polym. 2000;43:269-79.

[73] Muszyński A, O'Neill MA, Ramasamy E, Pattathil S, Avci U, Peña MJ, et al. Xyloglucan, galactomannan, glucuronoxylan, and rhamnogalacturonan I do not have identical structures in soybean root and root hair cell walls. Planta. 2015;242:1123-38.

[74] Morais de Carvalho D, Martínez-Abad A, Evtuguin DV, Colodette JL, Lindström ME, Vilaplana F, et al. Isolation and characterization of acetylated glucuronoarabinoxylan from sugarcane bagasse and straw. Carbohydr Polym. 2017;156:223-34.

[75] Liu L, Paulitz J, Pauly M. The presence of fucogalactoxyloglucan and its synthesis in rice indicates conserved functional importance in plants. Plant Physiol. 2015;168:549-60.

[76] Pfeifer L, Classen B. The Cell Wall of Seagrasses: Fascinating, Peculiar and a Blank Canvas for Future Research. Frontiers in Plant Science. 2020;Volume 11 - 2020.

[77] Olsen JL, Rouzé P, Verhelst B, Lin Y-C, Bayer T, Collen J, et al. The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. Nature. 2016;530:331-5.

[78] Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, et al. Phylogeny and molecular evolution of the green algae. Critical reviews in plant sciences. 2012;31:1-46.

[79] Li L, Wang S, Wang H, Sahu SK, Marin B, Li H, et al. The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. Nature Ecology & Evolution. 2020;4:1220-31.

[80] Ye J, Yang C, Xia L, Zhu Y, Liu L, Cao H, et al. Protoplast Preparation for Algal Single-Cell Omics Sequencing. Microorganisms. 2023;11:538.

[81] Mazéas L, Yonamine R, Barbeyron T, Henrissat B, Drula E, Terrapon N, et al. Assembly and synthesis of the extracellular matrix in brown algae. Seminars in Cell & Developmental Biology. 2023;134:112-24.

[82] Baghel RS, Reddy CRK, Singh RP. Seaweed-based cellulose: Applications, and future perspectives. Carbohydrate Polymers. 2021;267:118241.

[83] Domozych DS, Ciancia M, Fangel JU, Mikkelsen MD, Ulvskov P, Willats WG. The Cell Walls of Green Algae: A Journey through Evolution and Diversity. Front Plant Sci. 2012;3:82.

[84] Alzate-Gaviria L, Domínguez-Maldonado J, Chablé-Villacís R, Olguin-Maciel E, Leal-Bautista RM, Canché-Escamilla G, et al. Presence of Polyphenols Complex Aromatic "Lignin" in Sargassum spp. from Mexican Caribbean. Journal of Marine Science and Engineering. 2021;9:6.

[85] Martone PT, Estevez JM, Lu F, Ruel K, Denny MW, Somerville C, et al. Discovery of Lignin in Seaweed Reveals Convergent Evolution of Cell-Wall Architecture. Current Biology. 2009;19:169-75.

[86] Lee W-K, Ho C-L. Ecological and evolutionary diversification of sulphated polysaccharides in diverse photosynthetic lineages: A review. Carbohydrate Polymers. 2022;277:118764.

[87] Carina D, Sharma S, Jaiswal AK, Jaiswal S. Seaweeds polysaccharides in active food packaging: A review of recent progress. Trends in Food Science & Technology. 2021;110:559-72.

[88] Usov AI, Zelinsky ND. 2 - Chemical structures of algal polysaccharides. In: Domínguez H, editor. Functional Ingredients from Algae for Foods and Nutraceuticals: Woodhead Publishing; 2013. p. 23-86.

[89] Bilan MI, Grachev AA, Shashkov AS, Kelly M, Sanderson CJ, Nifantiev NE, et al. Further studies on the composition and structure of a fucoidan preparation from the brown alga Saccharina latissima. Carbohydrate Research. 2010;345:2038-47.

[90] Pomin VH, Mourão PAS. Structure, biology, evolution, and medical importance of sulfated fucans and galactans. Glycobiology. 2008;18:1016-27.

[91] Bajwa B, Xing X, Serin SC, Hayes M, Terry SA, Gruninger RJ, et al. Characterization of Unfractionated Polysaccharides in Brown Seaweed by Methylation-GC-MS-Based Linkage Analysis. Mar Drugs. 2024;22.

[92] Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG, et al. Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. Annual Review of Plant Biology. 2011;62:567-90.

[93] Cao L, Lu W, Mata A, Nishinari K, Fang Y. Egg-box model-based gelation of alginate and pectin: A review. Carbohydrate Polymers. 2020;242:116389.

[94] Urtuvia V, Maturana N, Acevedo F, Peña C, Díaz-Barrera A. Bacterial alginate production: an overview of its biosynthesis and potential industrial production. World Journal of Microbiology and Biotechnology. 2017;33:198.

[95] Chi S, Liu T, Wang X, Wang R, Wang S, Wang G, et al. Functional genomics analysis reveals the biosynthesis pathways of important cellular components (alginate and fucoidan) of Saccharina. Current Genetics. 2018;64:259-73.

[96] Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, et al. Saccharina genomes provide novel insight into kelp biology. Nature Communications. 2015;6:6986.

[97] Lawson CJ, Rees DA. An enzyme for the metabolic control of polysaccharide conformation and function. Nature. 1970;227:392-3.

[98] Ropartz D, Lissarrague A, Jam M, Jouanneau D, Le Gall S, Annic B, et al. Exploration of the extracellular matrix of the red alga Chondrus crispus reveals unprecedented insights into carrageenan structures. Carbohydrate Polymers. 2025;348:122737.

[99] Álvarez-Viñas M, Souto S, Flórez-Fernández N, Torres MD, Bandín I, Domínguez H. Antiviral Activity of Carrageenans and Processing Implications. Marine Drugs. 2021;19:437.

[100] Bajwa B, Xing XH, Terry SA, Gruninger RJ, Abbott DW. Methylation-GC-MS/FID-based glycosidic linkage analysis of unfractionated polysaccharides in red seaweeds. Marine Drugs. 2024;22:192.

[101] Becker B, Marin B, Melkonian M. Structure, composition, and biogenesis of prasinophyte cell coverings. Protoplasma. 1994;181:233-44.

[102] Sørensen I, Pettolino FA, Bacic A, Ralph J, Lu F, O'Neill MA, et al. The charophycean green algae provide insights into the early origins of plant cell walls. The Plant Journal. 2011;68:201-11.

[103] Redouan E, Cedric D, Emmanuel P, Mohamed EG, Bernard C, Philippe M, et al. Improved isolation of glucuronan from algae and the production of glucuronic acid oligosaccharides using a glucuronan lyase. Carbohydrate Research. 2009;344:1670-5.

[104] HARADA N, MAEDA M. Chemical Structure of Antithrombin-active Rhamnan Sulfate from Monostrom nitidum. Bioscience, Biotechnology, and Biochemistry. 1998;62:1647-52.

[105] Kidgell JT, Magnusson M, de Nys R, Glasson CRK. Ulvan: A systematic review of extraction, composition and function. Algal Research. 2019;39:101422.

[106] Kidgell JT, Glasson CRK, Magnusson M, Sims IM, Hinkley SFR, de Nys R, et al. Ulvans are not equal - Linkage and substitution patterns in ulvan polysaccharides differ with Ulva morphology. Carbohydrate Polymers. 2024;333:121962.

[107] Tako M, Tamanaha M, Tamashiro Y, Uechi S. Structure of ulvan isolated from the edible green seaweed, Ulva pertusa. Advances in Bioscience and Biotechnology. 2015;6:645.

[108] Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure. 1995;3:853-9.

[109] Davies GJ, Henrissat B. Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. Biochem Soc Trans. 2002;30:291-7.

[110] Davies GJ, Gloster TM, Henrissat B. Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr Opin Struct Biol. 2005;15:637-45.

[111] Lapébie P, Lombard V, Drula E, Terrapon N, Henrissat B. Bacteroidetes use thousands of enzyme combinations to break down glycans. Nat Commun. 2019;10:2043-.

[112] Asatsuma S, Sawada C, Itoh K, Okito M, Kitajima A, Mitsui T. Involvement of alpha-amylase I-1 in starch degradation in rice chloroplasts. Plant Cell Physiol. 2005;46:858-69.

[113] Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, et al. Fungal cellulases. Chemical reviews. 2015;115:1308-448.

[114] Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. Nature. 2017;544:65-70.

[115] Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490-5.

[116] Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J. 2010;432:437-44.

[117] Garron ML, Cygler M. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. Glycobiology. 2010;20:1547-73.

[118] Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels. 2013;6:41.

[119] Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, et al. An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. Science. 2010;330:219-22.

[120] Vermaas JV, Crowley MF, Beckham GT, Payne CM. Effects of lytic polysaccharide monooxygenase oxidation on cellulose structure and binding of oxidized cellulose oligomers to cellulases. The journal of physical chemistry B. 2015;119:6129-43.

[121] Hemsworth GR, Johnston EM, Davies GJ, Walton PH. Lytic Polysaccharide Monooxygenases in Biomass Conversion. Trends Biotechnol. 2015;33:747-61.

[122] Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem. 2008;77:521-55.

[123] Taujale R, Zhou Z, Yeung W, Moremen KW, Li S, Kannan N. Mapping the glycosyltransferase fold landscape using interpretable deep learning. Nature Communications. 2021;12:5656.

[124] BORASTON Alisdair B, BOLAM David N, GILBERT Harry J, DAVIES Gideon J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochemical Journal. 2004;382:769-81.

[125] Abbott DW, van Bueren AL. Using structure to inform carbohydrate binding module function. Current Opinion in Structural Biology. 2014;28:32-40.

[126] Abbott DW, Eirín-López JM, Boraston AB. Insight into Ligand Diversity and Novel Biological Roles for Family 32 Carbohydrate-Binding Modules. Molecular Biology and Evolution. 2007;25:155-67.

[127] Freelove AC, Bolam DN, White P, Hazlewood GP, Gilbert HJ. A novel carbohydrate-binding protein is a component of the plant cell wall-degrading complex of Piromyces equi. J Biol Chem. 2001;276:43010-7.

[128] Hettle AG, Vickers CJ, Boraston AB. Sulfatases: Critical Enzymes for Algal Polysaccharide Processing. Frontiers in Plant Science. 2022;Volume 13 - 2022.

[129] Luis Ana S, Yates Edwin A, Cartmell A. Functions and specificity of bacterial carbohydrate sulfatases targeting host glycans. Essays in Biochemistry. 2023;67:429-42.

[130] Stam M, Lelièvre P, Hoebeke M, Corre E, Barbeyron T, Michel G. SulfAtlas, the sulfatase database: state of the art and new developments. Nucleic Acids Research. 2022;51:D647-D53.

[131] Hanson SR, Best MD, Wong C-H. Sulfatases: Structure, Mechanism, Biological Activity, Inhibition, and Synthetic Utility. Angewandte Chemie International Edition. 2004;43:5736-63.

[132] Reisky L, Préchoux A, Zühlke M-K, Bäumgen M, Robb CS, Gerlach N, et al. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. Nature Chemical Biology. 2019;15:803-12.

[133] Vickers CJ, Hettle AG, Hobbs JK, Shapiro-Ward S, Pluvinage B, Medley BJ, et al. The complete λ-carrageenan depolymerization cascade from a marine Pseudoalteromonad revealed by structural analysis of the enzymes. bioRxiv. 2025:2025.01.09.632246.

[134] Hettle AG, Vickers C, Robb CS, Liu F, Withers SG, Hehemann JH, et al. The Molecular Basis of Polysaccharide Sulfatase Activity and a Nomenclature for Catalytic Subsites in this Class of Enzyme. Structure. 2018;26:747-58.e4.

[135] Moraïs S, Mizrahi I. Islands in the stream: from individual to communal fiber degradation in the rumen ecosystem. FEMS Microbiology Reviews. 2019;43:362-79.

[136] Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide Utilization Loci: Fueling Microbial Communities. J Bacteriol. 2017;199.

[137] P OS, Martin JC, Lawley TD, Browne HP, Harris HMB, Bernalier-Donadille A, et al. Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. Microb Genom. 2016;2:e000043.

[138] Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, et al. Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS One. 2007;2:e224.

[139] Tancula E, Feldhaus MJ, Bedzyk LA, Salyers AA. Location and characterization of genes involved in binding of starch to the surface of Bacteroides thetaiotaomicron. J Bacteriol. 1992;174:5609-16.

[140] Bjursell MK, Martens EC, Gordon JI. Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, Bacteroides thetaiotaomicron, to the Suckling Period*. Journal of Biological Chemistry. 2006;281:36269-79.

[141] Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, et al. Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. PLOS Biology. 2011;9:e1001221.

[142] Bayer EA, Kenig R, Lamed R. Adherence of Clostridium thermocellum to cellulose. J Bacteriol. 1983;156:818-27.

[143] Artzi L, Bayer EA, Moraïs S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nature Reviews Microbiology. 2017;15:83-95.

[144] Haitjema CH, Gilmore SP, Henske JK, Solomon KV, de Groot R, Kuo A, et al. A parts list for fungal cellulosomes revealed by comparative genomics. Nature Microbiology. 2017;2:17087.

[145] Stahl SW, Nash MA, Fried DB, Slutzki M, Barak Y, Bayer EA, et al. Single-molecule dissection of the high-affinity cohesin-dockerin complex. Proc Natl Acad Sci U S A. 2012;109:20431-6.

[146] Bayer EA, Belaich JP, Shoham Y, Lamed R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol. 2004;58:521-54.

[147] Arntzen MØ, Várnai A, Mackie RI, Eijsink VGH, Pope PB. Outer membrane vesicles from Fibrobacter succinogenes S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. Environmental Microbiology. 2017;19:2701-14.

[148] Sartorio MG, Pardue EJ, Scott NE, Feldman MF. Human gut bacteria tailor extracellular vesicle cargo for the breakdown of diet- and host-derived glycans. Proceedings of the National Academy of Sciences. 2023;120:e2306314120.

[149] Consortium TC. Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. Glycobiology. 2017;28:3-8.

[150] Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, et al. PULDB: the expanded database of Polysaccharide Utilization Loci. Nucleic Acids Res. 2018;46:D677-D83.

[151] Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Research. 2023;51:W115-W21.

[152] Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. 2010;20:1574-84.

[153] Nguyen SN, Flores A, Talamantes D, Dar F, Valdez A, Schwans J, et al. GeneHunt for rapid domain-specific annotation of glycoside hydrolases. Sci Rep. 2019;9:10137-.

[154] Fraser ASC, Low KE, Tingley JP, Reintjes G, Thomas D, Brumer H, et al. SACCHARIS v2: Streamlining Prediction of Carbohydrate-Active Enzyme Specificities Within Large Datasets. Methods Mol Biol. 2024;2836:299-330.

[155] Barrett K, Lange L. Peptide-based functional annotation of carbohydrate-active enzymes by conserved unique peptide patterns (CUPP). Biotechnol Biofuels. 2019;12:102.

[156] Kultima JR, Coelho LP, Forslund K, Huerta-Cepas J, Li SS, Driessen M, et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. Bioinformatics. 2016;32:2520-3.

[157] Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol. 2012;12:186.

[158] Stam MR, Danchin EGJ, Rancurel C, Coutinho PM, Henrissat B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. Protein Eng Des Sel. 2006;19:555-62.

[159] Viborg AH, Terrapon N, Lombard V, Michel G, Czjzek M, Henrissat B, et al. A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). J Biol Chem. 2019;294:15973-86.

[160] St John FJ, González JM, Pozharski E. Consolidation of glycosyl hydrolase family 30: A dual domain 4/7 hydrolase family consisting of two structurally distinct groups. FEBS Letters. 2010;584:4435-41.

[161] Mewis K, Lenfant N, Lombard V, Henrissat B. Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: a Motivation for Detailed Enzyme Characterization. Appl Environ Microbiol. 2016;82:1686-92.

[162] Stewart RD, Auffret MD, Roehe R, Watson M. Open prediction of polysaccharide utilisation loci (PUL) in 5414 public Bacteroidetes genomes using PULpy. Preprint at bioRxiv.https://www.biorxiv.org/content/10.1101/421024v1.full (2018).

[163] Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Preprint at bioRxiv.https://www.biorxiv.org/content/10.1101/2020.06.29.177501v1.full (2020).

[164] Ausland C, Zheng J, Yi H, Yang B, Li T, Feng X, et al. dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. Nucleic Acids Res. 2020.

[165] Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat Biotechnol. 2019;37:953-61.

[166] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

[167] Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 2005;21:2104-5.

[168] Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE. 2010;5:e9490.

[169] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312-3.

[170] Jones DR, Xing X, Tingley JP, Klassen L, King ML, Alexander TW, et al. Analysis of active site architecture and reaction product linkage chemistry reveals a conserved cleavage substrate for an endo-alpha-mannanase within diverse yeast mannans. J Mol Biol. 2020;432:1083-97.

[171] Jones DR, Uddin MS, Gruninger RJ, Pham TTM, Thomas D, Boraston AB, et al. Discovery and characterization of family 39 glycoside hydrolases from rumen anaerobic fungi with polyspecific activity on rare arabinosyl substrates. J Biol Chem. 2017;292:12606-20.

[172] Bhandari P, Tingley JP, Abbott DW, Hill JE. Characterization of an α-glucosidase enzyme conserved in Gardnerella spp. isolated from the human vaginal microbiome. Preprint at bioRxiv.https://www.biorxiv.org/content/10.1101/2020.05.11.086124v2 (2020).

[173] Barrett K, Hunt CJ, Lange L, Meyer AS. Conserved unique peptide patterns (CUPP) online platform: peptide-based functional annotation of carbohydrate active enzymes. Nucleic Acids Res. 2020;48:W110-W5.

[174] Zeuner B, Thomsen TB, Stringer MA, Krogh KBRM, Meyer AS, Holck J. Comparative Characterization of Aspergillus Pectin Lyases by Discriminative Substrate Degradation Profiling. Front Bioeng Biotechnol. 2020;8:873.

[175] Pilgaard B, Wilkens C, Herbst F-A, Vuillemin M, Rhein-Knudsen N, Meyer AS, et al. Proteomic enzyme analysis of the marine fungus Paradendryphiella salina reveals alginate lyase as a minimal adaptation strategy for brown algae degradation. Sci Rep. 2019;9:12338.

[176] Barrett K, Jensen K, Meyer AS, Frisvad JC, Lange L. Fungal secretome profile categorization of CAZymes by function and family corresponds to fungal phylogeny and taxonomy: Example Aspergillus and Penicillium. Sci Rep. 2020;10:5158.

[177] Machovsky-Capuska GE, Senior AM, Simpson SJ, Raubenheimer D. The Multidimensional Nutritional Niche. Trends in Ecology & Evolution. 2016;31:355-65.

[178] McAllister T, Bae H, Jones G, Cheng K-J. Microbial attachment and feed digestion in the rumen. Journal of animal science. 1994;72:3004-18.

[179] Huws SA, Edwards JE, Creevey CJ, Rees Stevens P, Lin W, Girdwood SE, et al. Temporal dynamics of the metabolically active rumen bacteria colonizing fresh perennial ryegrass. FEMS Microbiol Ecol. 2016;92.

[180] von Meijenfeldt FAB, Hogeweg P, Dutilh BE. A social niche breadth score reveals niche range strategies of generalists and specialists. Nature Ecology & Evolution. 2023;7:768-81.

[181] Bell TH, Bell T. Many roads to bacterial generalism. FEMS Microbiology Ecology. 2020;97.

[182] Xu J, Mahowald MA, Ley RE, Lozupone CA, Hamady M, Martens EC, et al. Evolution of Symbiotic Bacteria in the Distal Human Intestine. PLOS Biology. 2007;5:e156.

[183] Weimer PJ. Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. Front Microbiol. 2015;6:296.

[184] Tuncil YE, Xiao Y, Porter NT, Reuhs BL, Martens EC, Hamaker BR. Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence. mBio. 2017;8.

[185] Goyal A, Dubinkina V, Maslov S. Multiple stable states in microbial communities explained by the stable marriage problem. The ISME Journal. 2018;12:2823-34.

[186] Shinkai T, Kobayashi Y. Localization of ruminal cellulolytic bacteria on plant fibrous materials as determined by fluorescence in situ hybridization and real-time PCR. Appl Environ Microbiol. 2007;73:1646-52.

[187] Allesina S, Levine JM. A competitive network theory of species diversity. Proc Natl Acad Sci U S A. 2011;108:5638-42.

[188] Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, et al. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature. 2015;517:165-9.

[189] Rakoff-Nahoum S, Coyne Michael J, Comstock Laurie E. An Ecological Network of Polysaccharide Utilization among Human Intestinal Symbionts. Current Biology. 2014;24:40-9.

[190] Rakoff-Nahoum S, Foster KR, Comstock LE. The evolution of cooperation within the gut microbiota. Nature. 2016;533:255-9.

[191] Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, et al. Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. Nature Microbiology. 2018;3:1274-84.

[192] Sriswasdi S, Yang C-c, Iwasaki W. Generalist species drive microbial dispersion and evolution. Nature Communications. 2017;8:1162.

[193] Yan Z, Smith ML, Du P, Hahn MW, Nakhleh L. Species Tree Inference Methods Intended to Deal with Incomplete Lineage Sorting Are Robust to the Presence of Paralogs. Syst Biol. 2022;71:367-81.

[194] McLean R, Hobbs JK, Suits MD, Tuomivaara ST, Jones DR, Boraston AB, et al. Functional Analyses of Resurrected and Contemporary Enzymes Illuminate an Evolutionary Path for the Emergence of Exolysis in Polysaccharide Lyase Family 2. J Biol Chem. 2015;290:21231-43.

[195] Khairunisa BH, Heryakusuma C, Ike K, Mukhopadhyay B, Susanti D. Evolving understanding of rumen methanogen ecophysiology. Front Microbiol. 2023;14:1296008.

[196] Brown JR. Ancient horizontal gene transfer. Nature Reviews Genetics. 2003;4:121-32.

[197] Brito IL. Examining horizontal gene transfer in microbial communities. Nature Reviews Microbiology. 2021.

[198] Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Molecular Ecology. 2013;22:1503-17.

[199] Shoemaker NB, Vlamakis H, Hayes K, Salyers AA. Evidence for extensive resistance gene transfer among Bacteroides spp. and among Bacteroides and other genera in the human colon. Appl Environ Microb. 2001;67:561-8.

[200] Klassen L, Reintjes G, Tingley JP, Jones DR, Hehemann J-H, Smith AD, et al. Quantifying fluorescent glycan uptake to elucidate strain-level variability in foraging behaviors of rumen bacteria. Microbiome. 2021;9:23.

[201] Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, et al. Comparative Metagenomics Revealed Commonly Enriched Gene Sets in Human Gut Microbiomes. DNA Research. 2007;14:169-81.

[202] Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010;464:908-12.

[203] Thomas F, Barbeyron T, Tonon T, Génicot S, Czjzek M, Michel G. Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides. Environmental Microbiology. 2012;14:2379-94.

[204] Pluvinage B, Grondin JM, Amundsen C, Klassen L, Moote PE, Xiao Y, et al. Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. Nature Communications. 2018;9:1043.

[205] Pudlo NA, Pereira GV, Parnami J, Cid M, Markert S, Tingley JP, et al. Diverse events have transferred genes for edible seaweed digestion from marine to human gut bacteria. Cell Host Microbe. 2022;30:314-28.e11.

[206] Garcia-Vallvé S, Romeu A, Palau J. Horizontal Gene Transfer of Glycosyl Hydrolases of the Rumen Fungi. Molecular Biology and Evolution. 2000;17:352-61.

[207] Ricard G, McEwan NR, Dutilh BE, Jouany J-P, Macheboeuf D, Mitsumori M, et al. Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. BMC Genomics. 2006;7:22-.

[208] Williams AG, Withers S, Sutherland AD. The potential of bacteria isolated from ruminal contents of seaweed-eating North Ronaldsay sheep to hydrolyse seaweed components and produce methane by anaerobic digestion in vitro. Microb Biotechnol. 2013;6:45-52.

[209] Ferrillo A, Tingley JP, King ML, Kidane A, Bajwa B, Xing X, et al. Microbial alginate foraging is conserved in geographically and taxonomically distinct ruminant microbiomes. bioRxiv. 2024:2024.12.17.628917.

[210] Ahmar S, Hensel G, Gruszka D. CRISPR/Cas9-mediated genome editing techniques and new breeding strategies in cereals - current status, improvements, and perspectives. Biotechnol Adv. 2023;69:108248.

[211] Hughes AD, Kelly MS, Black KD, Stanley MS. Biogas from Macroalgae: is it time to revisit the idea? Biotechnol Biofuels. 2012;5:86.

[212] Kraan S. Mass-cultivation of carbohydrate rich macroalgae, a possible solution for sustainable biofuel production. Mitig Adapt Strat Gl. 2013;18:27-46.

[213] Krause-Jensen D, Duarte CM. Substantial role of macroalgae in marine carbon sequestration. Nat Geosci. 2016;9:737-+.

[214] Makkar HPS, Tran G, Heuzé V, Giger-Reverdin S, Lessire M, Lebas F, et al. Seaweeds for livestock diets: A review. Animal Feed Science and Technology. 2016;212:1-17.

[215] Lafeuille B, Tamigneaux E, Berger K, Provencher V, Beaulieu L. Variation of the Nutritional Composition and Bioactive Potential in Edible Macroalga Cultivated from Atlantic Canada Subjected to Different Growth and Processing Conditions. Foods. 2023;12.

[216] Inoue A, Ojima T. Functional identification of alginate lyase from the brown alga Saccharina japonica. Scientific Reports. 2019;9:4937.

[217] Pilgaard B, Vuillemin M, Holck J, Wilkens C, Meyer AS. Specificities and Synergistic Actions of Novel PL8 and PL7 Alginate Lyases from the Marine Fungus Paradendryphiella salina. J Fungi (Basel). 2021;7.

[218] Suda K, Tanji Y, Hori K, Unno H. Evidence for a novel Chlorella virus-encoded alginate lyase. FEMS Microbiology Letters. 1999;180:45-53.

[219] Hata M, Kumagai Y, Rahman MM, Chiba S, Tanaka H, Inoue A, et al. Comparative study on general properties of alginate lyases from some marine gastropod mollusks. Fisheries Sci. 2009;75:755-63.

[220] Nisizawa K, Fujibayashi S, Kashiwabara Y. Alginate Lyases in the Hepatopancreas of a Marine Mollusc, Dolabella auricula Solander*. The Journal of Biochemistry. 1968;64:25-37.

[221] Mathieu S, Touvrey-Loiodice M, Poulet L, Drouillard S, Vincentelli R, Henrissat B, et al. Ancient acquisition of "alginate utilization loci" by human gut microbiota. Scientific Reports. 2018;8.

[222] Hehemann JH, Kelly AG, Pudlo NA, Martens EC, Boraston AB. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. P Natl Acad Sci USA. 2012;109:19786-91.

[223] Ronne ME, Tandrup T, Madsen M, Hunt CJ, Myers PN, Moll JM, et al. Three alginate lyases provide a new gut isolate with the ability to grow on alginate. Appl Environ Microb. 2023.

[224] Coyne MJ, Zitomersky NL, McGuire AM, Earl AM, Comstock LE. Evidence of extensive DNA transfer between bacteroidales species within the human gut. mBio. 2014;5:e01305-14.

[225] Abbott DW, Aasen IM, Beauchemin KA, Grondahl F, Gruninger R, Hayes M, et al. Seaweed and Seaweed Bioactives for Mitigation of Enteric Methane: Challenges and Opportunities. Animals-Basel. 2020;10.

[226] Meo-Filho P, Ramirez-Agudelo JF, Kebreab E. Mitigating methane emissions in grazing beef cattle with a seaweed-based feed additive: Implications for climate-smart agriculture. Proceedings of the National Academy of Sciences. 2024;121:e2410863121.

[227] Evans FD, Critchley AT. Seaweeds for animal production use. J Appl Phycol. 2014;26:891-9.

[228] Hansen HR, Hector BL, Feldmann J. A qualitative and quantitative evaluation of the seaweed diet of North Ronaldsay sheep. Animal Feed Science and Technology. 2003;105:21-8.

[229] Bilan MI, Grachev AA, Shashkov AS, Kelly M, Sanderson CJ, Nifantiev NE, et al. Further studies on the composition and structure of a fucoidan preparation from the brown alga Saccharina latissima. Carbohydr Res. 2010;345:2038-47.

[230] Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35:725-31.

[231] D'Souza G, Schwartzman J, Keegstra J, Schreier JE, Daniels M, Cordero OX, et al. Interspecies interactions determine growth dynamics of biopolymer-degrading populations in microbial communities. Proc Natl Acad Sci U S A. 2023;120:e2305198120.

[232] Lu D, Zhang Q, Wang S, Guan J, Jiao R, Han N, et al. Biochemical characteristics and synergistic effect of two novel alginate lyases from Photobacterium sp. FC615. Biotechnol Biofuels. 2019;12:260.

[233] Terrapon N, Lombard V, Gilbert HJ, Henrissat B. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. Bioinformatics. 2014;31:647-55.

[234] Naas AE, Mackenzie AK, Mravec J, Schückel J, Willats WGT, Eijsink VGH, et al. Do Rumen Bacteroidetes Utilize an Alternative Mechanism for Cellulose Degradation? mBio. 2014;5:10.1128/mbio.01401-14.

[235] Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. Journal of Biological Chemistry. 2009;284:24673-7.

[236] Wang B, Dong S, Li FL, Ma XQ. Structural basis for the exolytic activity of polysaccharide lyase family 6 alginate lyase BcAlyPL6 from human gut microbe Bacteroides clarus. Biochem Biophys Res Commun. 2021;547:111-7.

[237] Mathieu S, Henrissat B, Labre F, Skjåk-Bræk G, Helbert W. Functional Exploration of the Polysaccharide Lyase Family PL6. Plos One. 2016;11.

[238] Lee SI, Choi SH, Lee EY, Kim HS. Molecular cloning, purification, and characterization of a novel polyMG-specific alginate lyase responsible for alginate MG block degradation in Stenotrophomas maltophilia KJ-2. Appl Microbiol Biotechnol. 2012;95:1643-53.

[239] Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics. 2012;28:i409-i15.

[240] Kay M. ggdist: Visualizations of Distributions and Uncertainty in the Grammar of Graphics. IEEE Trans Vis Comput Graph. 2024;30:414-24.

[241] Kawai S, Hashimoto W. 4-Deoxy-l-erythro-5-hexoseulose Uronate (DEH) and DEH Reductase: Key Molecule and Enzyme for the Metabolism and Utilization of Alginate. Molecules. 2022;27.

[242] Mori T, Takahashi M, Tanaka R, Miyake H, Shibata T, Chow S, et al. Falsirhodobacter sp. alg1 Harbors Single Homologs of Endo and Exo-Type Alginate Lyases Efficient for Alginate Depolymerization. PLoS One. 2016;11:e0155537.

[243] Hobbs JK, Lee SM, Robb M, Hof F, Barr C, Abe KT, et al. KdgF, the missing link in the microbial metabolism of uronate sugars from pectin and alginate. Proc Natl Acad Sci U S A. 2016;113:6188-93.

[244] Koch H, Dürwald A, Schweder T, Noriega-Ortega B, Vidal-Melgosa S, Hehemann JH, et al. Biphasic cellular adaptations and ecological implications of degrading a mixture of algal polysaccharides. Isme J. 2019;13:92-103.

[245] Robb CS, Hobbs JK, Pluvinage B, Reintjes G, Klassen L, Monteith S, et al. Metabolism of a hybrid algal galactan by members of the human gut microbiome. Nat Chem Biol. 2022;18:501-10.

[246] Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. Nat Microbiol. 2018;3:210-9.

[247] Zhu YM, Chen P, Bao YJ, Men Y, Zeng Y, Yang JG, et al. Complete genome sequence and transcriptomic analysis of a novel marine strain reveals the mechanism of brown algae degradation. Scientific Reports. 2016;6:38248.

[248] Fu T, Pan L, Shang Q, Yu G. Fermentation of alginate and its derivatives by different enterotypes of human gut microbiota: Towards personalized nutrition using enterotype-specific dietary fibers. Int J Biol Macromol. 2021;183:1649-59.

[249] Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. Isme J. 2017;11:1640-50.

[250] Hehemann JH, Reintjes G, Klassen L, Smith AD, Ndeh D, Arnosti C, et al. Single cell fluorescence imaging of glycan uptake by intestinal bacteria. Isme J. 2019;13:1883-9.

[251] Briggs JA, Grondin JM, Brumer H. Communal living: glycan utilization by the human gut microbiota. Environ Microbiol. 2021;23:15-35.

[252] Zhang Z, Yu G, Zhao X, Liu H, Guan H, Lawson AM, et al. Sequence analysis of alginate-derived oligosaccharides by negative-ion electrospray tandem mass spectrometry. J Am Soc Mass Spectrom. 2006;17:621-30.

[253] Hansen BB, Lorentzen JR, Welker JM, Varpe Ø, Aanes R, Beumer LT, et al. Reindeer turning maritime: Ice-locked tundra triggers changes in dietary niche utilization. Ecosphere. 2019;10:e02672.

[254] Kibegwa FM, Bett RC, Gachuiri CK, Machuka E, Stomeo F, Mujibi FD. Diversity and functional analysis of rumen and fecal microbial communities associated with dietary changes in crossbreed dairy cattle. PLoS One. 2023;18:e0274371.

[255] Remminghorst U, Rehm BH. Bacterial alginates: from biosynthesis to applications. Biotechnol Lett. 2006;28:1701-12.

[256] Ostrowski MP, La Rosa SL, Kunath BJ, Robertson A, Pereira G, Hagen LH, et al. Mechanistic insights into consumption of the food additive xanthan gum by the human gut microbiota. Nat Microbiol. 2022;7:556-69.

[257] Cha Q-Q, Wang X-J, Ren X-B, Li D, Wang P, Li P-Y, et al. Comparison of Alginate Utilization Pathways in Culturable Bacteria Isolated From Arctic and Antarctic Marine Environments. Frontiers in Microbiology. 2021;Volume 12 - 2021.

[258] Tingley JP, Andersen TO, Mihalynuk LG, Xing X, Low KE, Whiteside DP, et al. Global distribution of microbial carrageenan foraging pathways reveals widespread latent traits within the genetic “dark matter” of ruminant intestinal microbiomes. bioRxiv. 2025:2025.03.04.641501.

[259] Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH, et al. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Scientific Reports. 2015;5:14567.

[260] Feng J, Qian Y, Zhou Z, Ertmer S, Vivas EI, Lan F, et al. Polysaccharide utilization loci in Bacteroides determine population fitness and community-level interactions. Cell Host Microbe. 2022;30:200-15.e12.

[261] Wu M, McNulty NP, Rodionov DA, Khoroshkin MS, Griffin NW, Cheng J, et al. Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. Science. 2015;350:aac5992.

[262] Avci U, Pattathil S, Hahn MG. Immunological approaches to plant cell wall and biomass characterization: immunolocalization of glycan epitopes. Methods Mol Biol. 2012;908:73-82.

[263] Wood JA, Tan HT, Collins HM, Yap K, Khor SF, Lim WL, et al. Genetic and environmental factors contribute to variation in cell wall composition in mature desi chickpea (Cicer arietinum L.) cotyledons. Plant Cell Environ. 2018;41:2195-208.

[264] Badhan A, Low KE, Jones DR, Xing X, Milani MRM, Polo RO, et al. Mechanistic insights into the digestion of complex dietary fibre by the rumen microbiota using combinatorial high-resolution glycomics and transcriptomic analyses. Comput Struct Biotechnol J. 2022;20:148-64.

[265] Low KE, Xing X, Moote PE, Inglis GD, Venketachalam S, Hahn MG, et al. Combinatorial Glycomic Analyses to Direct CAZyme Discovery for the Tailored Degradation of Canola Meal Non-Starch Dietary Polysaccharides. Microorganisms. 2020;8.

[266] Muhidinov ZK, Bobokalonov JT, Ismoilov IB, Strahan GD, Chau HK, Hotchkiss AT, et al. Characterization of two types of polysaccharides from roots growing in Tajikistan. Food Hydrocolloid. 2020;105:105768.

[267] Hosain NA, Ghosh R, Bryant DL, Arivett BA, Farone AL, Kline PC. Isolation, structure elucidation, and immunostimulatory activity of polysaccharide fractions from Boswellia carterii frankincense resin. International Journal of Biological Macromolecules. 2019;133:76-85.

[268] Chong HH, Cleary MT, Dokoozlian N, Ford CM, Fincher GB. Soluble cell wall carbohydrates and their relationship with sensory attributes in Cabernet Sauvignon wine. Food Chem. 2019;298:124745.

[269] Grabež V, Devle H, Kidane A, Mydland LT, Øverland M, Ottestad S, et al. Sugar Kelp (Saccharina latissima) Seaweed Added to a Growing-Finishing Lamb Diet Has a Positive Effect on Quality Traits and on Mineral Content of Meat. Foods. 2023;12:2131.

[270] Terry SA, Garcia V, Beauchemin KA. Assessing the methane mitigation potential of Canadian red seaweeds using in vitro batch culture. Canadian Journal of Animal Science. 2023;103:394-405.

[271] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20.

[272] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biology. 2019;20:257.

[273] Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. Peerj Comput Sci. 2017:e104.

[274] Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a Prokaryotic Universal Primer for Simultaneous Analysis of and Using Next-Generation Sequencing. Plos One. 2014;9.

[275] Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581-+.

[276] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.

[277] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 2012;41:D590-D6.

[278] Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824-34.

[279] Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6:158.

[280] Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11:1144-6.

[281] Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605-7.

[282] Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.

[283] Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. Isme J. 2017;11:2864-8.

[284] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043-55.

[285] Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment- free sequence identification. Microb Genomics. 2021;7:000685.

[286] Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36:1925-7.

[287] Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674-6.

[288] Nissen JN, Johansen J, Allesoe RL, Sonderby CK, Armenteros JJA, Gronbech CH, et al. Improved metagenome binning and assembly using deep variational autoencoders. Nature Biotechnology. 2021;39:555-60.

[289] Van Den Bossche T, Kunath BJ, Schallert K, Schäpe SS, Abraham PE, Armengaud J, et al. Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows. Nat Commun. 2021;12:7305.

[290] Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18:366-8.

[291] Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. Nucleic Acids Res. 2024;52:W521-w5.

[292] Juggins S. rioja: Analysis of Quaternary Science Data. CRAN: Comprehensive R Archive Network (CRAN); 2022.

[293] Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 2019;20:238.

[294] Tiedemann F. gghalves: Compose Half-Half Plots Using Your Favourite Geoms. 2025.

[295] Slowikowski K. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. 12 ed2024.

[296] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972-3.

[297] Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution. 2019;37:291-4.

[298] Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35:4453-5.

[299] Duchemin W, Gence G, Arigon Chifolleau A-M, Arvestad L, Bansal MS, Berry V, et al. RecPhyloXML: a format for reconciled gene trees. Bioinformatics. 2018;34:3646-52.

[300] Wessel D, Flugge UI. A Method for the Quantitative Recovery of Protein in Dilute-Solution in the Presence of Detergents and Lipids. Anal Biochem. 1984;138:141-3.

[301] Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat Methods. 2017;14:513-20.

[302] Leprevost FD, Haynes SE, Avtonomov DM, Chang HY, Shanmugam AK, Mellacheruvu D, et al. Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nat Methods. 2020;17:869-70.

[303] Yu FC, Haynes SE, Nesvizhskii AI. IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs. Mol Cell Proteomics. 2021;20.

[304] Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13:731-40.

[305] Iwamoto Y, Araki R, Iriyama K-i, Oda T, Fukuda H, Hayashida S, et al. Purification and Characterization of Bifunctional Alginate Lyase from Alteromonas sp. Strain No. 272 and Its Action on Saturated Oligomeric Substrates. Bioscience, Biotechnology, and Biochemistry. 2001;65:133-42.

[306] Remoroza C, Cord-Landwehr S, Leijdekkers AG, Moerschbacher BM, Schols HA, Gruppen H. Combined HILIC-ELSD/ESI-MS(n) enables the separation, identification and quantification of sugar beet pectin derived oligomers. Carbohydr Polym. 2012;90:41-8.

[307] Bezner Kerr R, Hasegawa T, Lasco R. Food, fibre, and other ecosystem products Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change ed HO Pörtner et al. Cambridge: Cambridge University Press; 2022.

[308] Bryant M. Microbial methane production—theoretical aspects. Journal of animal science. 1979;48:193-201.

[309] Eisler MC, Lee MR, Tarlton JF, Martin GB, Beddington J, Dungait JA, et al. Agriculture: Steps to sustainable livestock. Nature. 2014;507:32-4.

[310] Montzka SA, Dlugokencky EJ, Butler JH. Non-CO2 greenhouse gases and climate change. Nature. 2011;476:43-50.

[311] Terry SA, Krüger AM, Lima PMT, Gruninger RJ, Abbott DW, Beauchemin KA. Evaluation of Rumen Fermentation and Microbial Adaptation to Three Red Seaweeds Using the Rumen Simulation Technique. Animals-Basel. 2023;13:1643.

[312] O'Hara E, Terry SA, Moote P, Beauchemin KA, McAllister TA, Abbott DW, et al. Comparative analysis of macroalgae supplementation on the rumen microbial community: Asparagopsis taxiformis inhibits major ruminal methanogenic, fibrolytic, and volatile fatty acid-producing microbes in vitro. Front Microbiol. 2023;14:1104667.

[313] La Rosa SL, Ostrowski MP, Vera-Ponce de León A, McKee LS, Larsbrink J, Eijsink VG, et al. Glycan processing in gut microbiomes. Current Opinion in Microbiology. 2022;67:102143.

[314] Cartmell A, Munoz-Munoz J, Briggs JA, Ndeh DA, Lowe EC, Basle A, et al. A surface endogalactanase in Bacteroides thetaiotaomicron confers keystone status for arabinogalactan degradation. Nat Microbiol. 2018;3:1314-26.

[315] Hehemann JH, Smyth L, Yadav A, Vocadlo DJ, Boraston AB. Analysis of keystone enzyme in Agar hydrolysis provides insight into the degradation (of a polysaccharide from) red seaweeds. J Biol Chem. 2012;287:13985-95.

[316] Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, et al. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature. 2014;506:498-502.

[317] Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL. Diet-induced extinctions in the gut microbiota compound over generations. Nature. 2016;529:212-5.

[318] Jiao J-Y, Liu L, Hua Z-S, Fang B-Z, Zhou E-M, Salam N, et al. Microbial dark matter coming to light: challenges and opportunities. National Science Review. 2021;8:nwaa280.

[319] Andersen TO, Kunath BJ, Hagen LH, Arntzen MØ, Pope PB. Rumen metaproteomics: Closer to linking rumen microbial function to animal productivity traits. Methods. 2021;186:42-51.

[320] Barnett DJM. microViz: an R package for microbiome data visualization and statistics. Journal of Open Source Software. 2021;6:3201.

[321] Michel G, Chantalat L, Duee E, Barbeyron T, Henrissat B, Kloareg B, et al. The κ-carrageenase of P. carrageenovora Features a Tunnel-Shaped Active Site: A Novel Insight in the Evolution of Clan-B Glycoside Hydrolases. Structure. 2001;9:513-25.

[322] Cao S, Zhang Y, Chen G, Shen J, Han J, Chang Y, et al. Cloning, Heterologous Expression, and Characterization of a βκ-Carrageenase From Marine Bacterium Wenyingzhuangia funcanilytica: A Specific Enzyme for the Hybrid Carrageenan-Furcellaran. Front Microbiol. 2021;12:697218.

[323] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630:493-500.

[324] Hettle AG, Hobbs JK, Pluvinage B, Vickers C, Abe KT, Salama-Alber O, et al. Insights into the κ/ι-carrageenan metabolism pathway of some marine Pseudoalteromonas species. Communications Biology. 2019;2:474.

[325] Nishida AH, Ochman H. Origins and Evolution of Novel Bacteroides in Captive Apes. bioRxiv. 2023.

[326] Tong F, Wang T, Gao NL, Liu Z, Cui K, Duan Y, et al. The microbiome of the buffalo digestive tract. Nat Commun. 2022;13:823.

[327] G.BIF.Org_User. Occurrence Download. The Global Biodiversity Information Facility; 2024.

[328] Vihtakari M. ggOceanMaps: Plot Data on Oceanographic Maps using 'ggplot2'. . R package version 2.2.0 ed2024.

[329] Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research. 2021;50:D571-D7.

[330] Facimoto CT, Clements KD, White WL, Handley KM. Bacteroidia and Clostridia are equipped to degrade a cascade of polysaccharides along the hindgut of the herbivorous fish Kyphosus sydneyanus. ISME Communications. 2024;4.

[331] Hilliou L, Larotonda FDS, Abreu P, Abreu MH, Sereno AM, Gonçalves MP. The impact of seaweed life phase and postharvest storage duration on the chemical and rheological properties of

hybrid carrageenans isolated from Portuguese Mastocarpus stellatus. Carbohydrate Polymers. 2012;87:2655-63.

[332] Moran D, Clements KD. Diet and endogenous carbohydrases in the temperate marine herbivorous fish Kyphosus sydneyanus. Journal of Fish Biology. 2002;60:1190-203.

[333] Miller IJ. The Chemical Structure of Galactans from Some New Zealand Red Algae. 2003;46:572-7.

[334] Holden JJ, MacNeill SK, Juanes F, Dudas SE. Beach-cast deposition and commercial harvesting of a non-indigenous alga, Mazzaella japonica: Implications for macrofauna communities in Baynes Sound, British Columbia. Estuarine, Coastal and Shelf Science. 2018;210:162-71.

[335] Mandelli F, Martins MP, Chinaglia M, Lima EAd, Morais MAB, Lima TB, et al. A functionally augmented carbohydrate utilization locus from herbivore gut microbiota fueled by dietary β-glucans. npj Biofilms and Microbiomes. 2024;10:105.

[336] Jones DR, McLean R, Hobbs JK, Abbott DW. A surrogate structural platform informed by ancestral reconstruction and resurrection of a putative carbohydrate binding module hybrid illuminates the neofunctionalization of a pectate lyase. Journal of Structural Biology. 2019;207:279-86.

[337] Liu J, Taft DH, Maldonado-Gomez MX, Johnson D, Treiber ML, Lemay DG, et al. The fecal resistome of dairy cattle is associated with diet during nursing. Nature Communications. 2019;10:4406.

[338] Ghaderiardakani F, Quartino ML, Wichard T. Microbiome-Dependent Adaptation of Seaweeds Under Environmental Stresses: A Perspective. Frontiers in Marine Science. 2020;7.

[339] Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y, et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nature Communications. 2017;8:1685.

[340] Dabdoub S. kraken-biom: Enabling interoperative format conversion for Kraken results (Version 1.2) [Software]. 2016.

[341] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one. 2013;8:e61217.

[342] Wickham H. dplyr: A Grammar of Data Manipulation. R package version 114, https://githubcom/tidyverse/dplyr, https://dplyrtidyverseorg. 2023.

[343] Wickham H. ggplot2: elegant graphics for data analysis Springer-Verlag New York; 2009. Preprint at. 2016.

[344] Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. Biotechniques. 2004;36:808-12.

[345] Wilkins D. gggenes: Draw Gene Arrow Maps in 'ggplot2'. R package version 0.5.1 ed2023.

[346] Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13:e1005595.

[347] Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34:i142-i50.

[348] Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods. 2023;20:1203-12.

[349] Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

[350] Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic acids research. 2024;52:W78-W82.

[351] Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66:125-32.

[352] Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013;69:1204-14.

[353] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40:658-74.

[354] Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020;577:706-10.

[355] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60:2126-32.

[356] Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol. 2019;75:861-77.

[357] Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature. 1992;355:472-5.

[358] Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 2018;27:293-315.

[359] Reintjes G, Klassen L, Abbott DW. Visualization of Carbohydrate Uptake Using Fluorescent Polysaccharides. Methods Mol Biol. 2023;2657:241-9.

[360] King ML, Xing X, Reintjes G, Klassen L, Low KE, Alexander TW, et al. In vitro and ex vivo metabolism of chemically diverse fructans by bovine rumen Bifidobacterium and Lactobacillus species. Anim Microbiome. 2024;6:50.

[361] Hofman-Kamińska E, Bocherens H, Drucker DG, Fyfe RM, Gumiński W, Makowiecki D, et al. Adapt or die-Response of large herbivores to environmental changes in Europe during the Holocene. Glob Chang Biol. 2019;25:2915-30.

[362] McCauley DJ, Hardesty-Moore M, Halpern BS, Young HS. A mammoth undertaking: harnessing insight from functional ecology to shape de-extinction priority setting. Functional Ecology. 2017;31:1003-11.

[363] Mills JG, Weinstein P, Gellie NJC, Weyrich LS, Lowe AJ, Breed MF. Urban habitat restoration provides a human health benefit through microbiome rewilding: the Microbiome Rewilding Hypothesis. Restoration Ecology. 2017;25:866-72.

[364] Cromsigt J, Te Beest M, Kerley GIH, Landman M, le Roux E, Smith FA. Trophic rewilding as a climate change mitigation strategy? Philos Trans R Soc Lond B Biol Sci. 2018;373.

[365] Cromsigt JPGM, Kemp YJM, Rodriguez E, Kivit H. Rewilding Europe's large grazer community: how functionally diverse are the diets of European bison, cattle, and horses? Restoration Ecology. 2018;26:891-9.

[366] Kirchhelle C. Pharming animals: a global history of antibiotics in food production (1935–2017). Palgrave Communications. 2018;4:96.

[367] He Y, Yuan Q, Mathieu J, Stadler L, Senehi N, Sun R, et al. Antibiotic resistance genes from livestock waste: occurrence, dissemination, and treatment. npj Clean Water. 2020;3:4.

[368] Sucena Afonso J, El Tholth M, McIntyre KM, Carmo LP, Coyne L, Manriquez D, et al. Strategies to reduce antimicrobials in livestock and aquaculture, and their impact under field conditions: a structured scoping literature review. J Antimicrob Chemother. 2024;79:11-26.

[369] Rovira P, McAllister T, Lakin SM, Cook SR, Doster E, Noyes NR, et al. Characterization of the Microbial Resistome in Conventional and “Raised Without Antibiotics” Beef and Dairy Production Systems. Frontiers in Microbiology. 2019;Volume 10 - 2019.

[370] Horsfield M. Voices from the sound: Chronicles of clayoquot sound and Tofino, 1899-1929: Salal Books; 2008.

[371] Bartlett C, Marshall M, Marshall A. Two-Eyed Seeing and other lessons learned within a co-learning journey of bringing together indigenous and mainstream knowledges and ways of knowing. Journal of Environmental Studies and Sciences. 2012;2:331-40.

[372] Atleo ER. Tsawalk: A Nuu-chah-nulth worldview: UBC press; 2007.

[373] Lee C, Zaheer R, Munns K, Holman DB, Van Domselaar G, Zovoilis A, et al. Effect of Antimicrobial Use in Conventional Versus Natural Cattle Feedlots on the Microbiome and Resistome. Microorganisms. 2023;11.

[374] Portillo A, Ruiz-Larrea F, Zarazaga M, Alonso A, Martinez Jose L, Torres C. Macrolide Resistance Genes inEnterococcus spp. Antimicrobial Agents and Chemotherapy. 2000;44:967-71.

[375] Schmidt J, Zdarska V, Kolar M, Mlynarcik P. Analysis of BlaEC family class C beta-lactamase. FEMS Microbiol Lett. 2023;370.

[376] Petrone BL, Aqeel A, Jiang S, Durand HK, Dallow EP, McCann JR, et al. Diversity of plant DNA in stool is linked to dietary quality, age, and household income. Proceedings of the National Academy of Sciences. 2023;120:e2304441120.

[377] Estrada E. Topological structural classes of complex networks. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics. 2007;75:016103.

[378] Wilkinson T, Korir D, Ogugo M, Stewart RD, Watson M, Paxton E, et al. 1200 high-quality metagenome-assembled genomes from the rumen of African cattle and their relevance in the context of sub-optimal feeding. Genome Biology. 2020;21:229.

[379] Glendinning L, Genç B, Wallace RJ, Watson M. Metagenomic analysis of the cow, sheep, reindeer and red deer rumen. Scientific Reports. 2021;11:1990.

[380] Creevey CJ, Kelly WJ, Henderson G, Leahy SC. Determining the culturability of the rumen bacterial microbiome. Microb Biotechnol. 2014;7:467-79.

[381] Gong G, Zhou S, Luo R, Gesang Z, Suolang S. Metagenomic insights into the diversity of carbohydrate-degrading enzymes in the yak fecal microbial community. BMC Microbiol. 2020;20:302.

[382] Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications. 2018;9:5114.

[383] Collén PN, Jeudy A, Sassi J-F, Groisillier A, Czjzek M, Coutinho PM, et al. A Novel Unsaturated β-Glucuronyl Hydrolase Involved in Ulvan Degradation Unveils the Versatility of Stereochemistry Requirements in Family GH105*. Journal of Biological Chemistry. 2014;289:6199-211.

[384] Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, et al. Gut microbiome of the Hadza hunter-gatherers. Nature Communications. 2014;5:3654.

[385] Conteville LC, Oliveira-Ferreira J, Vicente ACP. Gut Microbiome Biomarkers and Functional Diversity Within an Amazonian Semi-Nomadic Hunter-Gatherer Group. Front Microbiol. 2019;10:1743.

[386] Carter MM, Olm MR, Merrill BD, Dahan D, Tripathi S, Spencer SP, et al. Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. Cell. 2023;186:3111-24.e13.

[387] Luis AS, Hansson GC. Intestinal mucus and their glycans: A habitat for thriving microbiota. Cell Host & Microbe. 2023;31:1087-100.

[388] Kim CY, Ma J, Lee I. HiFi metagenomic sequencing enables assembly of accurate and complete genomes from human gut microbiota. Nature Communications. 2022;13:6367.

[389] Bava R, Castagna F, Lupia C, Poerio G, Liguori G, Lombardi R, et al. Antimicrobial Resistance in Livestock: A Serious Threat to Public Health. Antibiotics (Basel). 2024;13.

[390] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15:1-21.

[391] Kolde R, Kolde MR. Package 'pheatmap'. R package. 2018;1.

[392] Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology. 2019;37:852-7.

[393] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215:403-10.

[394] Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, et al. Package 'vegan'. Community ecology package, version. 2013;2:1-295.

[395] Tipton L, Müller CL, Kurtz ZD, Huang L, Kleerup E, Morris A, et al. Fungi stabilize connectivity in the lung and skin microbial ecosystems. Microbiome. 2018;6:12.

[396] Csardi G, Nepusz T. The igraph software. Complex syst. 2006;1695:1-9.

[397] Krassowski M. Krassowski/complex-upset. Zenodo Available at. 2020;10.

[398] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150-2.

[399] Bonin N, Doster E, Worley H, Pinnell LJ, Bravo JE, Ferm P, et al. MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing. Nucleic Acids Research. 2022;51:D744-D52.

[400] Blighe K, Rana S, Lewis M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.2. 0. R package version 116. 2019.

[401] Benton HP, Want EJ, Ebbels TMD. Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. Bioinformatics. 2010;26:2488-9.

[402] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Analytical Chemistry. 2006;78:779-87.

[403] Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics. 2008;9:504.

[404] Gatto L, Gibb S, Rainer J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. Journal of Proteome Research. 2021;20:1063-9.

[405] Gatto L, Lilley KS. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics. 2011;28:288-9.

[406] Morgan M, Wang J, Obenchain V, Lang M, Thompson R, Turaga N. BiocParallel: Bioconductor facilities for parallel evaluation. R package version 1.42.0 ed2025.

[407] Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. Journal of Proteome Research. 2008;7:1650-9.

[408] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583-9.

[409] Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67:355-67.

[410] Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010;66:12-21.

[411] Trinh P, Teichman S, Roberts MC, Rabinowitz PM, Willis AD. A cross-sectional comparison of gut metagenomes between dairy workers and community controls. BMC Genomics. 2024;25:708.

[412] de Carvalho CN, Muñiz F, Cáceres LM, Belaústegui Z, Rodríguez-Vidal J, Belo J, et al. Aurochs roamed along the SW coast of Andalusia (Spain) during Late Pleistocene. Scientific Reports. 2022;12:9911.

[413] Hofman-Kamińska E, Bocherens H, Drucker DG, Fyfe RM, Gumiński W, Makowiecki D, et al. Adapt or die—Response of large herbivores to environmental changes in Europe during the Holocene. Global Change Biology. 2019;25:2915-30.

[414] Hoshino T, Doi H, Uramoto G-I, Wörmer L, Adhikari RR, Xiao N, et al. Global diversity of microbial communities in marine sediment. Proceedings of the National Academy of Sciences. 2020;117:27587-97.

[415] Lachnit T, Meske D, Wahl M, Harder T, Schmitz R. Epibacterial community patterns on marine macroalgae are host-specific but temporally variable. Environ Microbiol. 2011;13:655-65.

[416] Terry SA, Coates T, Gruninger R, Abbott DW, Beauchemin KA. Evaluation of the red seaweed Mazzaella japonica as a feed additive for beef cattle. Frontiers in Animal Science. 2023;Volume 4 - 2023.

[417] Maia MRG, Fonseca AJM, Oliveira HM, Mendonça C, Cabrita ARJ. The Potential Role of Seaweeds in the Natural Manipulation of Rumen Fermentation and Methane Production. Scientific Reports. 2016;6:32321.

[418] Carpenter LJ, Liss PS. On temperate sources of bromoform and other reactive organic bromine gases. Journal of Geophysical Research: Atmospheres. 2000;105:20539-47.

[419] Gharechahi J, Vahidi MF, Sharifi G, Ariaeenejad S, Ding X-Z, Han J-L, et al. Lignocellulose degradation by rumen bacterial communities: New insights from metagenome analyses. Environmental Research. 2023;229:115925.

[420] Quartinello F, Kremser K, Schoen H, Tesei D, Ploszczanski L, Nagler M, et al. Together Is Better: The Rumen Microbial Community as Biological Toolbox for Degradation of Synthetic Polyesters. Frontiers in Bioengineering and Biotechnology. 2021;Volume 9 - 2021.

[421] Borsani B, De Santis R, Perico V, Penagini F, Pendezza E, Dilillo D, et al. The Role of Carrageenan in Inflammatory Bowel Diseases and Allergic Reactions: Where Do We Stand? Nutrients. 2021;13.

[422] Chen X, Fu X, Huang L, Xu J, Gao X. Agar oligosaccharides: A review of preparation, structures, bioactivities and application. Carbohydrate Polymers. 2021;265:118076.

[423] Ford L, Stratakos AC, Theodoridou K, Dick JTA, Sheldrake GN, Linton M, et al. Polyphenols from Brown Seaweeds as a Potential Antimicrobial Agent in Animal Feeds. ACS Omega. 2020;5:9093-103.

[424] DeAngelo J, Saenz BT, Arzeno-Soltero IB, Frieder CA, Long MC, Hamman J, et al. Economic and biophysical limits to seaweed farming for climate change mitigation. Nature Plants. 2023;9:45-57.

[425] Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596:590-6.

[426] Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, et al. Symbol Nomenclature for Graphical Representations of Glycans. Glycobiology. 2015;25:1323-4.

[427] Bennke CM, Reintjes G, Schattenhofer M, Ellrott A, Wulf J, Zeder M, et al. Modification of a High-Throughput Automatic Microbial Cell Enumeration System for Shipboard Analyses. Appl Environ Microbiol. 2016;82:3289-96.

[428] Oksanen JS, G.; Blanchet, F.; Kindt, R.; Legendre, P.; Minchin, P.; O'Hara, R.; Solymos, P.; Stevens, M.; Szoecs, E.; Wagner, H.; Barbour, M.; Bedward, M.; Bolker, B.; Borcard, D.; Carvalho, G.; Chirico, M.; De Caceres, M.; Durand, S.; Evangelista, H.; FitzJohn, R.; Friendly, M.; Furneaux, B.; Hannigan, G.; Hill, M.; Lahti, L.; McGlinn, D.; Ouellette, M.; Ribeiro Cunha, E.; Smith, T.; Stier, A.; Ter Braak, C. Vegan: community ecology package. version 0.1.4 ed. CRAN: Comprehensive R Archive Network (CRAN); 2023.

[429] Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26:1463-4.

[430] Hervé M. RVAideMemoire: Testing and Plotting Procedures for Biostatistics. CRAN: Comprehensive R Archive Network (CRAN); 2023.

[431] Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021;9:37.

[432] Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39:578-85.

[433] Despres J, Forano E, Lepercq P, Comtet-Marre S, Jubelin G, Chambon C, et al. Xylan degradation by the human gut Bacteroides xylanisolvens XB1A(T) involves two distinct gene clusters that are linked at the transcriptional level. BMC Genomics. 2016;17:326.

[434] Meale SJ, Li S, Azevedo P, Derakhshani H, Plaizier JC, Khafipour E, et al. Development of Ruminal and Fecal Microbiomes Are Affected by Weaning But Not Weaning Strategy in Dairy Calves. Frontiers in Microbiology. 2016;7.

[435] apos, Hara E, Neves ALA, Song Y, Guan LL. The Role of the Gut Microbiome in Cattle Production and Health: Driver or Passenger? Annual Review of Animal Biosciences. 2020;8:199-220.

[436] Charoensiddhi S, Conlon MA, Methacanon P, Franco CMM, Su P, Zhang W. Gut health benefits of brown seaweed Ecklonia radiata and its polysaccharides demonstrated in vivo in a rat model. Journal of Functional Foods. 2017;37:676-84.

[437] Stevenson TT, Furneaux RH. Chemical methods for the analysis of sulphated galactans from red algae. Carbohydrate Research. 1991;210:277-98.

[438] Nagasawa K, Inoue Y, Kamata T. Solvolytic desulfation of glycosaminoglycuronan sulfates with dimethyl sulfoxide containing water or methanol. Carbohydrate Research. 1977;58:47-55.

[439] Carpita NC, Shea EM. Linkage structure of carbohydrates by gas chromatography-mass spectrometry (GC-MS) of partially methylated alditol acetates. In: C.J. B, McGinnis GD, editors. Analysis of Carbohydrates by GLC and MS. Boca Raton, Florida: CRC Press, Inc.; 1989. p. 157-216.

[440] Xing X, Cui SW, Nie S, Phillips GO, Goff HD, Wang Q. Study on Dendrobium officinale O-acetyl-glucomannan (Dendronan®): Part I. Extraction, purification, and partial structural characterization. Bioactive Carbohydrates and Dietary Fibre. 2014;4:74-83.

Appendix 1 Supplemental Material to Chapter 2

Appendix 1 Figures

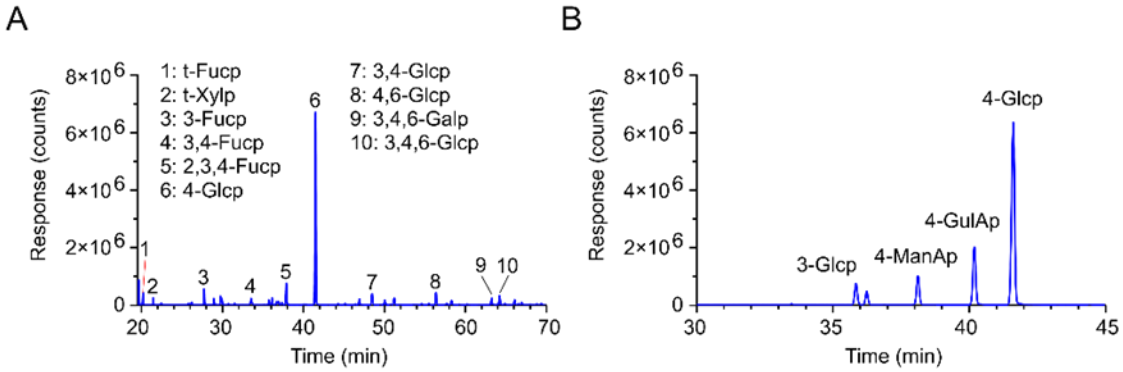

Figure A1.1: Linkage analysis chromatographs of *S. latissima*. Total ion current chromatograms of *S. latissima* with (**A**) standard methanolysis and (**B**) carboxyl reduced methanolysis with major PMMAs labelled.

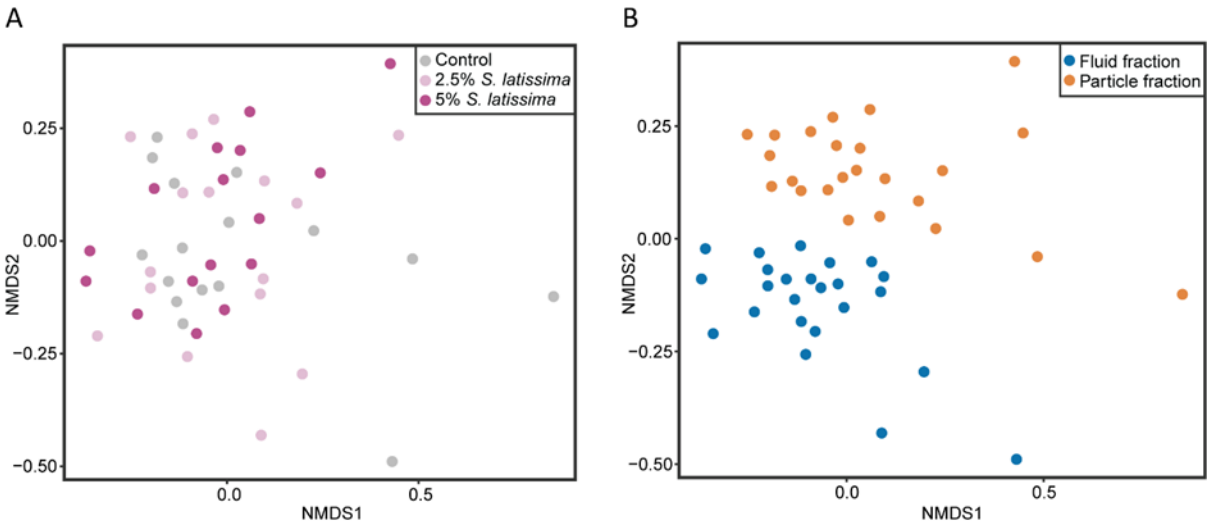

Figure A1.2: In vivo community analysis of lamb rumen microbial communities (16S rRNA gene) from dietary groups supplemented with 2.5% and 5% *S. latissima*. A) Non-metric multi-dimensional scaling (NMDS) plot comparing the community composition in samples taken from control (grey), 2.5% (light purple) and 5% (dark purple) *S. latissima* dietary groups. **B)** The NMDS plot colored by sample fraction, either fluid (blue) or particle (orange).

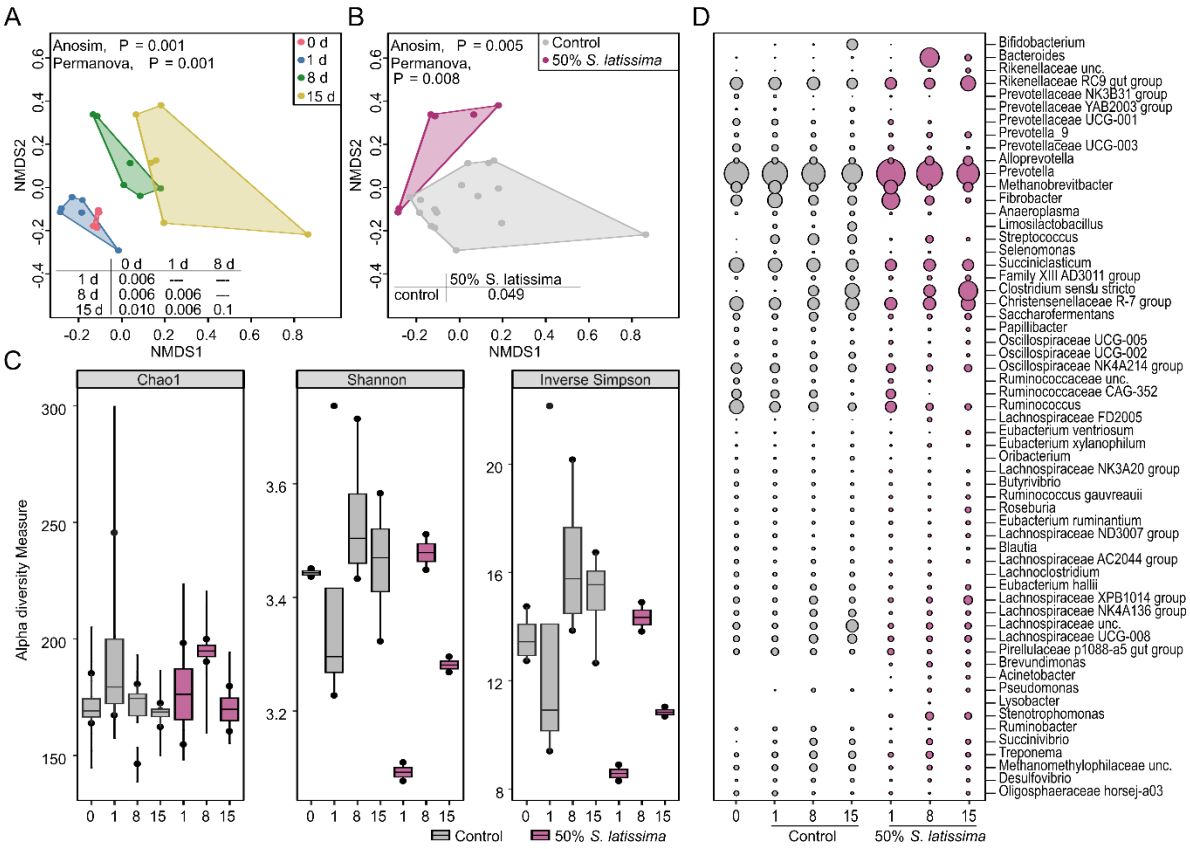

Figure A1.3: Ex vivo community analysis of RUSITEC rumen microbial communities (16S rRNA gene) supplemented with 50% *S. latissima*. A) Non-metric multi-dimensional scaling (NMDS) plot comparing the community composition at different time points: 0 d (pink), 1 d (blue), 8 d (green), and 15 d (yellow). **B)** NMDS plot comparing the effect each treatment has on the community composition where control (grey) and 50% *S. latissima* (purple). **C)** Alpha diversity (Chao1, Shannon, and inverse Simpson indices) in RUSITEC rumen microbial communities supplemented with 50% *S. latissima* over time. **D)** Bubble plot of different bacterial and archaeal genera that reached a minimum relative read abundance of 1% in control (grey) and 50% *S. latissima* (purple) RUSITEC rumen microbial communities.

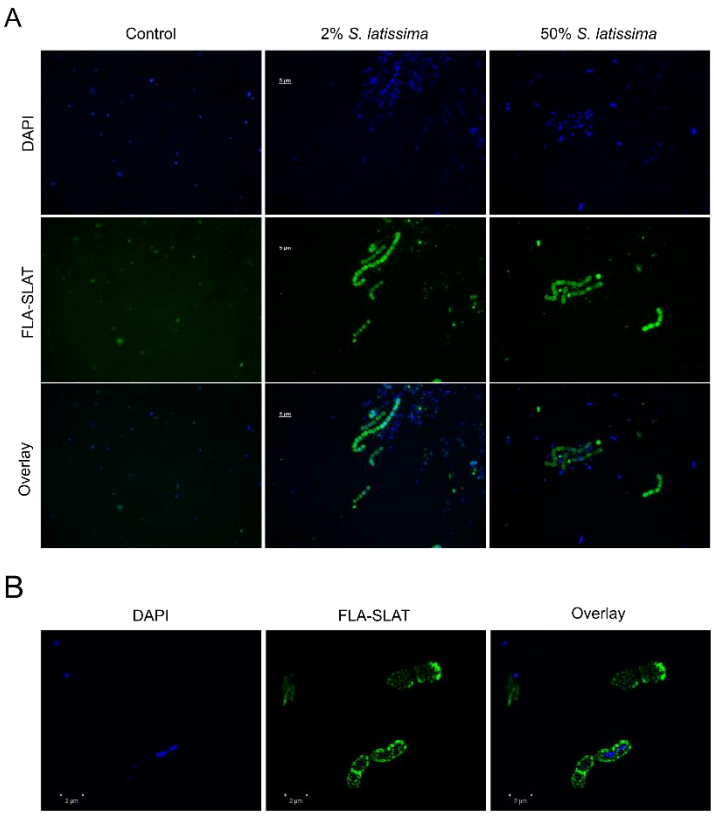

Figure A1.4: Epifluorescence visualization of FLA-SLAT interactions in RUSITEC rumen microbial communities that were enriched with 50% *S. latissima* for 15 days. **A)** FLA-SLAT interactions in control, 2% *S. latissima*, and 50% *S. latissima* RUSITEC microbial communities sampled from RUSITEC vessels after 15 days, stained with DAPI (blue) and incubated with 0.2 % FLA-SLAT (green) for 1 day. **B)** SR-SIM of cells from 50% *S. latissima* experiment after 1 day FLA-SLAT incubation. Cells were stained by DAPI, FLA-SLAT, and an overlay of DAPI and FLA-SLAT is shown.

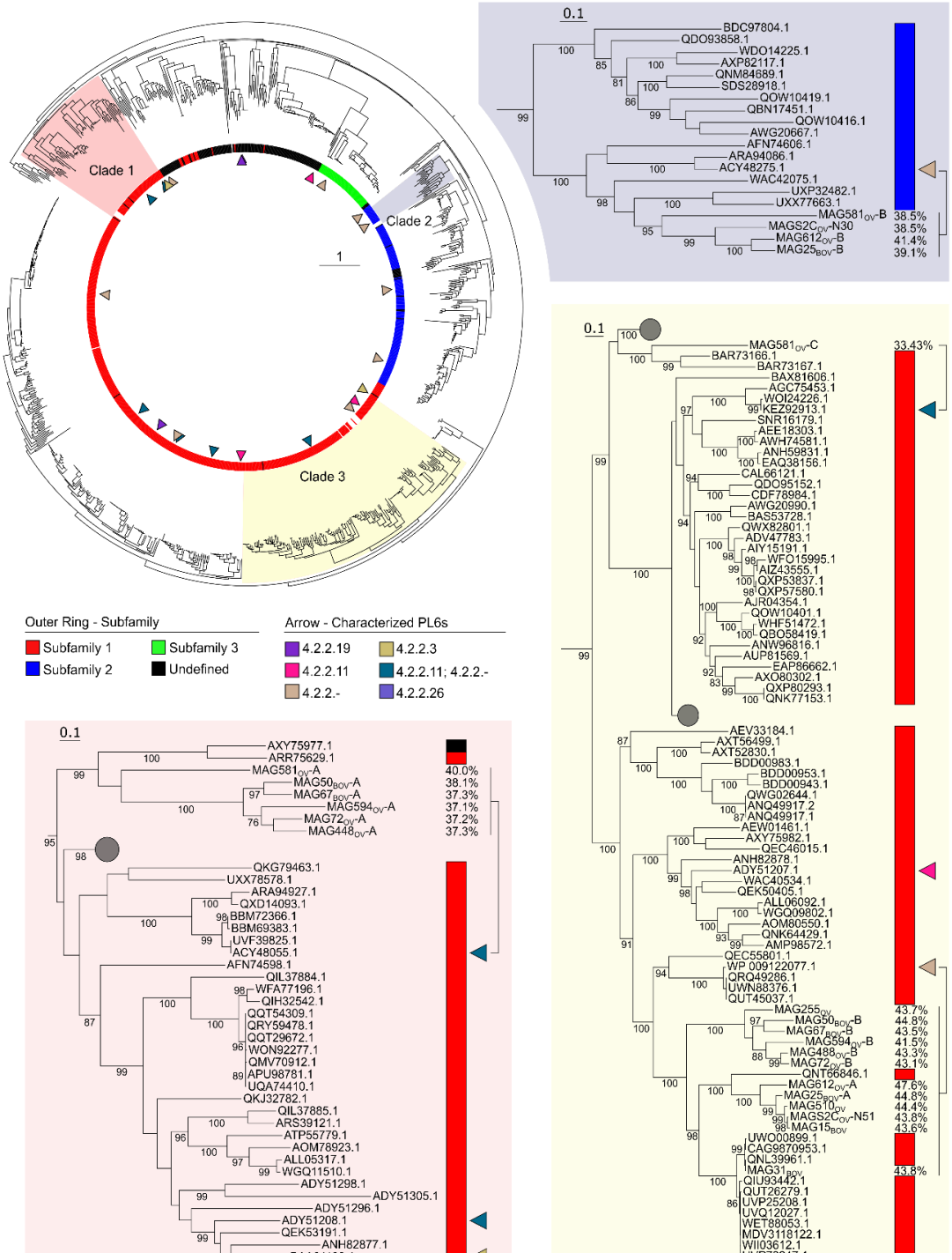
Clade 1
Clade 2
Clade 3
1
Outer Ring - Subfamily
Subfamily 1
Subfamily 2
Subfamily 3
Undefined
Arrow - Characterized PL6s
4.2.2.19
4.2.2.11
4.2.2.-
4.2.2.3
4.2.2.11; 4.2.2.-
4.2.2.26
0.1
BDC97804.1
QDO93858.1
WDO14225.1
AXP82117.1
QNM84689.1
SDS28918.1
QOW10419.1
QBN17451.1
QOW10416.1
AWG20667.1
AFN74606.1
ARA94086.1
ACY48275.1
WAC42075.1
UXP32482.1
UXX77663.1
MAG581OV-B
MAGS2COV-N30
MAG612OV-B
MAG25BOV-B
38.5%
38.5%
41.4%
39.1%
0.1
MAG581OV-C
BAR73166.1
BAR73167.1
BAX81606.1
AGC75453.1
WOI24226.1
KEZ92913.1
SNR16179.1
AEE18303.1
AWH74581.1
ANH59831.1
EAQ38156.1
CAL66121.1
QDO95152.1
CDF78984.1
AWG20990.1
BAS53728.1
QWX82801.1
ADV47783.1
AIY15191.1
WFO15995.1
AIZ43555.1
QXP53837.1
QXP57580.1
AJR04354.1
QOW10401.1
WHF51472.1
QBO58419.1
ANW96816.1
AUP81569.1
EAP86662.1
AXO80302.1
QXP80293.1
QNK77153.1
33.43%
AEV33184.1
AXT56499.1
AXT52830.1
BDD00983.1
BDD00953.1
BDD00943.1
QWG02644.1
ANQ49917.2
ANQ49917.1
AEW01461.1
AXY75982.1
QEC46015.1
ANH82878.1
ADY51207.1
WAC40534.1
QEK50405.1
ALL06092.1
WGQ09802.1
AOM80550.1
QNK64429.1
AMP98572.1
QEC55801.1
WP 009122077.1
QRQ49286.1
UWN88376.1
QUT45037.1
MAG255QV
MAG50BOV-B
MAG67BOV-B
MAG594OV-B
MAG488OV-B
MAG72OV-B
QNT66846.1
MAG612OV-A
MAG25BOV-A
MAG510OV
MAGS2COV-N51
MAG15BOV
UWO00899.1
CAG9870953.1
QNL39961.1
MAG31BOV
QIU93442.1
QUT26279.1
UVP25208.1
UVQ12027.1
WET88053.1
MDV3118122.1
WII03612.1
43.7%
44.8%
43.5%
41.5%
43.3%
43.1%
47.6%
44.8%
44.4%
43.8%
43.6%
43.8%
0.1
AXY75977.1
ARR75629.1
MAG581OV-A
MAG50BOV-A
MAG67BOV-A
MAG594OV-A
MAG72OV-A
MAG448OV-A
40.0%
38.1%
37.3%
37.1%
37.2%
37.3%
QKG79463.1
UXX78578.1
ARA94927.1
QXD14083.1
BBM72366.1
BBM69383.1
UVF39825.1
ACY48055.1
AFN74598.1
QIL37884.1
WFA77196.1
QIH32542.1
QQT54309.1
QRY59478.1
QQT29672.1
WON92277.1
QMV70912.1
APU98781.1
UQA74410.1
QKJ32782.1
QIL37885.1
ARS39121.1
ATP55779.1
AOM78923.1
ALL05317.1
WGQ11510.1
ADY51298.1
ADY51305.1
ADY51296.1
ADY51208.1
QEK53191.1
ANH82877.1

Figure A1.5: Similarity of *Bacteroidota* spp. PL6 members. SACCHARIS phylogeny of PL6 members within CAZy and *Bacteroidota* spp. members within this study. The full phylogeny (top left) highlights the three clades where *Bacteroidota* spp. PL6s were found (red – clade 1; blue – clade 2; yellow – clade 3). Clades were further pruned to identify closest related PL6 members (color highlighted as above), and the PL6 members within each clade was compared to the closest related characterized PL6 member. Clade 1 – ACY48055.1 (*Rhodothermus marinus* DSM 4252), clade 2 – ACY48275.1 (*Rhodothermus marinus* DSM 4252), and clade 3 - KEZ92913.1 (*Nonlabens ulvanivorans* PLR) and WP_009122077.1 (*B. clarus* YIT 12056).

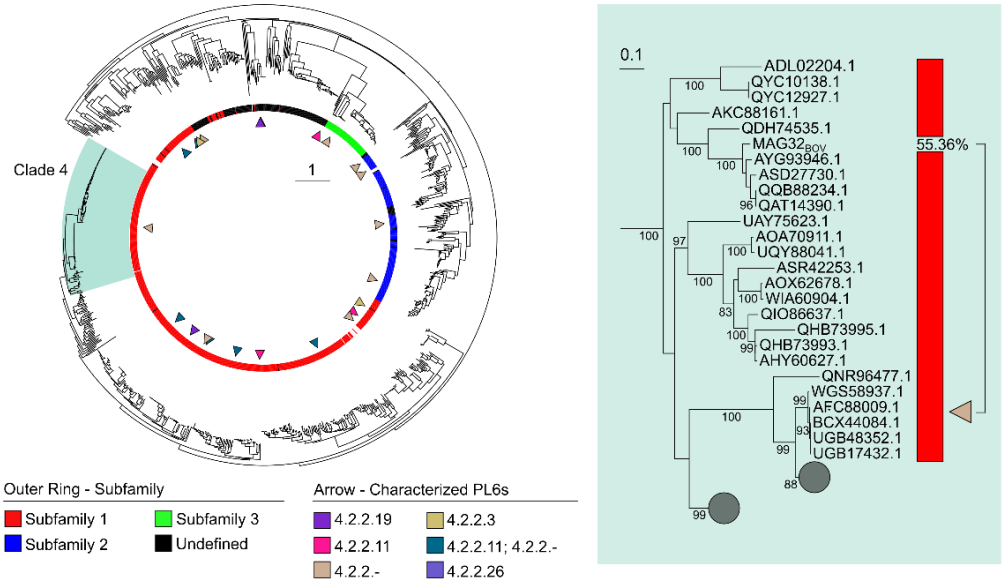

Figure A1.6: Similarity of *B. bullata* PL6 to other PL6 members. SACCHARIS phylogeny [154] of PL6 members within CAZy and MAG32BOV PL6 member. The full phylogeny (**left**) highlights the $MAG32_{BOV}$ clade, which is further pruned (**right**) to display closest related PL6 members. The identity between $MAG32_{BOV}$ PL6 and the closest characterized PL6 member (AFC88009.1 - *Stenotrophomonas maltophilia* KJ-2) is displayed.

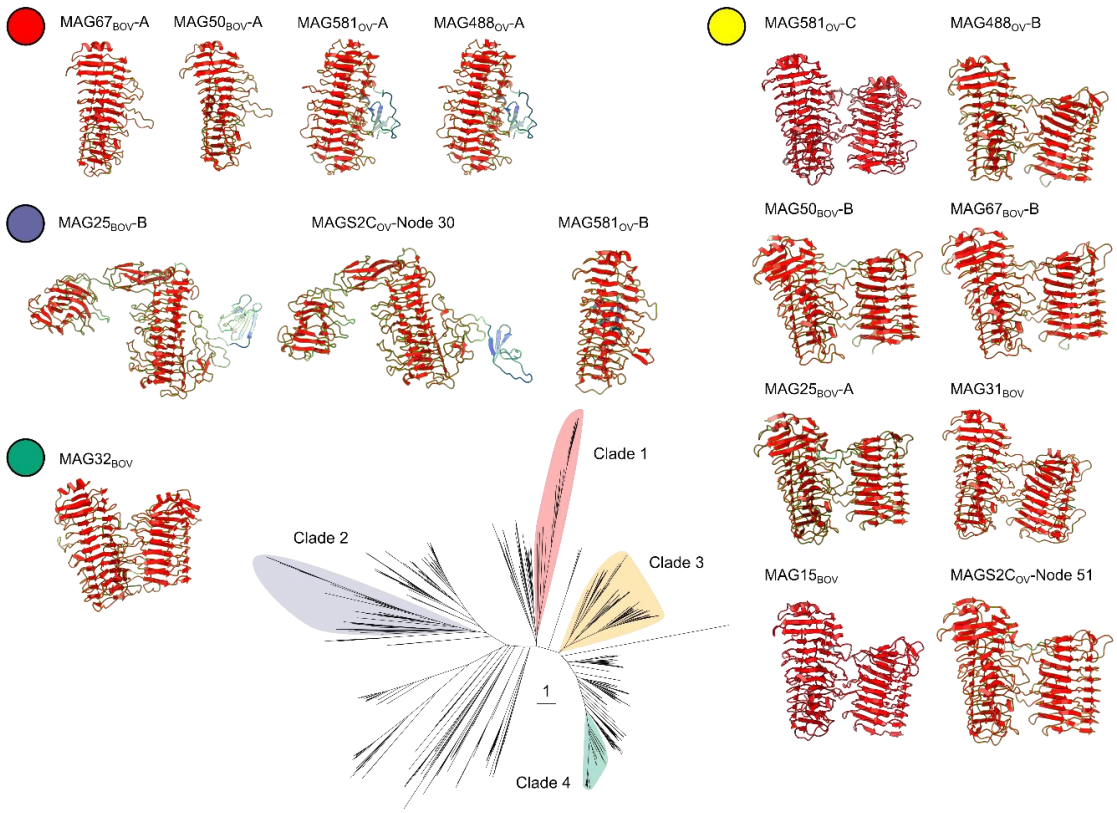

Figure A1.7: Predicted structure of alginate utilization loci (AUL) and alginate utilization cluster (AUC) PL6 members. AlphaFold [425] predicted structures of AUL and AUC PL6 members mapped to the PL6 phylogeny in **Figure 2.5**. Structures are colored based on b factor (confidence). Structures are grouped by their corresponding clade.

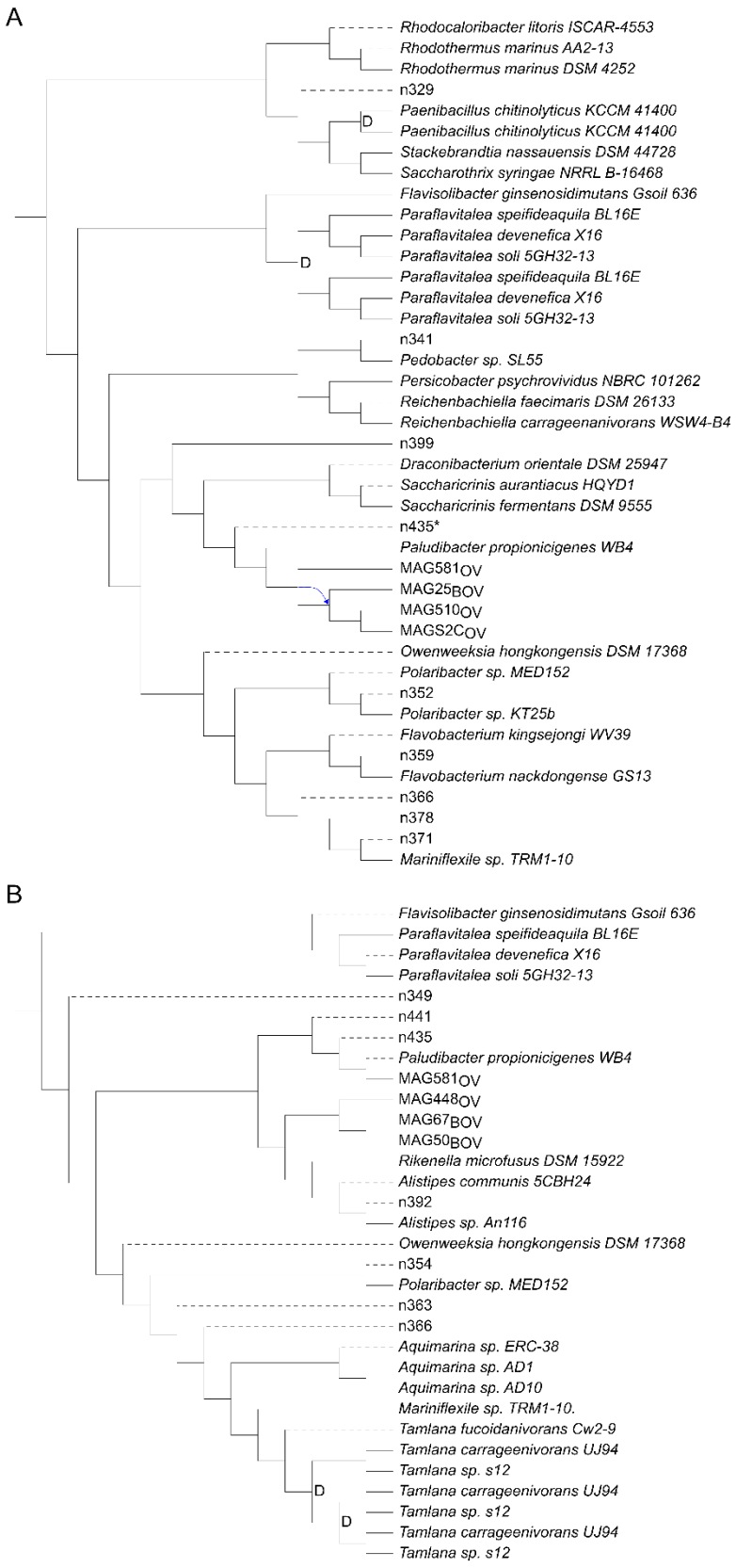
A
Rhodocaloribacter litoris ISCAR-4553
Rhodothermus marinus AA2-13
Rhodothermus marinus DSM 4252
n329
Paenibacillus chitinolyticus KCCM 41400
D
Paenibacillus chitinolyticus KCCM 41400
Stackebrandtia nassauensis DSM 44728
Saccharothrix syringae NRRL B-16468
Flavisolibacter ginsenosidimutans Gsoil 636
Paraflavitalea speifideaquila BL16E
Paraflavitalea devenefica X16
Paraflavitalea soli 5GH32-13
D
Paraflavitalea speifideaquila BL16E
Paraflavitalea devenefica X16
Paraflavitalea soli 5GH32-13
n341
Pedobacter sp. SL55
Persicobacter psychrovividus NBRC 101262
Reichenbachiella faecimaris DSM 26133
Reichenbachiella carrageenanivorans WSW4-B4
n399
Draconibacterium orientale DSM 25947
Saccharicrinis aurantiacus HQYD1
Saccharicrinis fermentans DSM 9555
n435*
Paludibacter propionicigenes WB4
MAG581OV
MAG25BOV
MAG510OV
MAGS2COV
Owenweeksia hongkongensis DSM 17368
Polaribacter sp. MED152
n352
Polaribacter sp. KT25b
Flavobacterium kingsejongi WV39
n359
Flavobacterium nackdongense GS13
n366
n378
n371
Mariniflexile sp. TRM1-10
B
Flavisolibacter ginsenosidimutans Gsoil 636
Paraflavitalea speifideaquila BL16E
Paraflavitalea devenefica X16
Paraflavitalea soli 5GH32-13
n349
n441
n435
Paludibacter propionicigenes WB4
MAG581OV
MAG448OV
MAG67BOV
MAG50BOV
Rikenella microfusus DSM 15922
Alistipes communis 5CBH24
n392
Alistipes sp. An116
Owenweeksia hongkongensis DSM 17368
n354
Polaribacter sp. MED152
n363
n366
Aquimarina sp. ERC-38
Aquimarina sp. AD1
Aquimarina sp. AD10
Mariniflexile sp. TRM1-10.
Tamlana fucoidanivorans Cw2-9
Tamlana carrageenivorans UJ94
Tamlana sp. s12
D
Tamlana carrageenivorans UJ94
Tamlana sp. s12
D
Tamlana carrageenivorans UJ94
Tamlana sp. s12

Figure A1.8: NOTUNG PL6 gene trees. PL6 clade 2 (**A**) and clade 1 (**B**) gene trees. Gene loss (dotted lines) and duplications "D" are marked per species and lines are drawn between clades represent predicted horizontal gene transfer events.

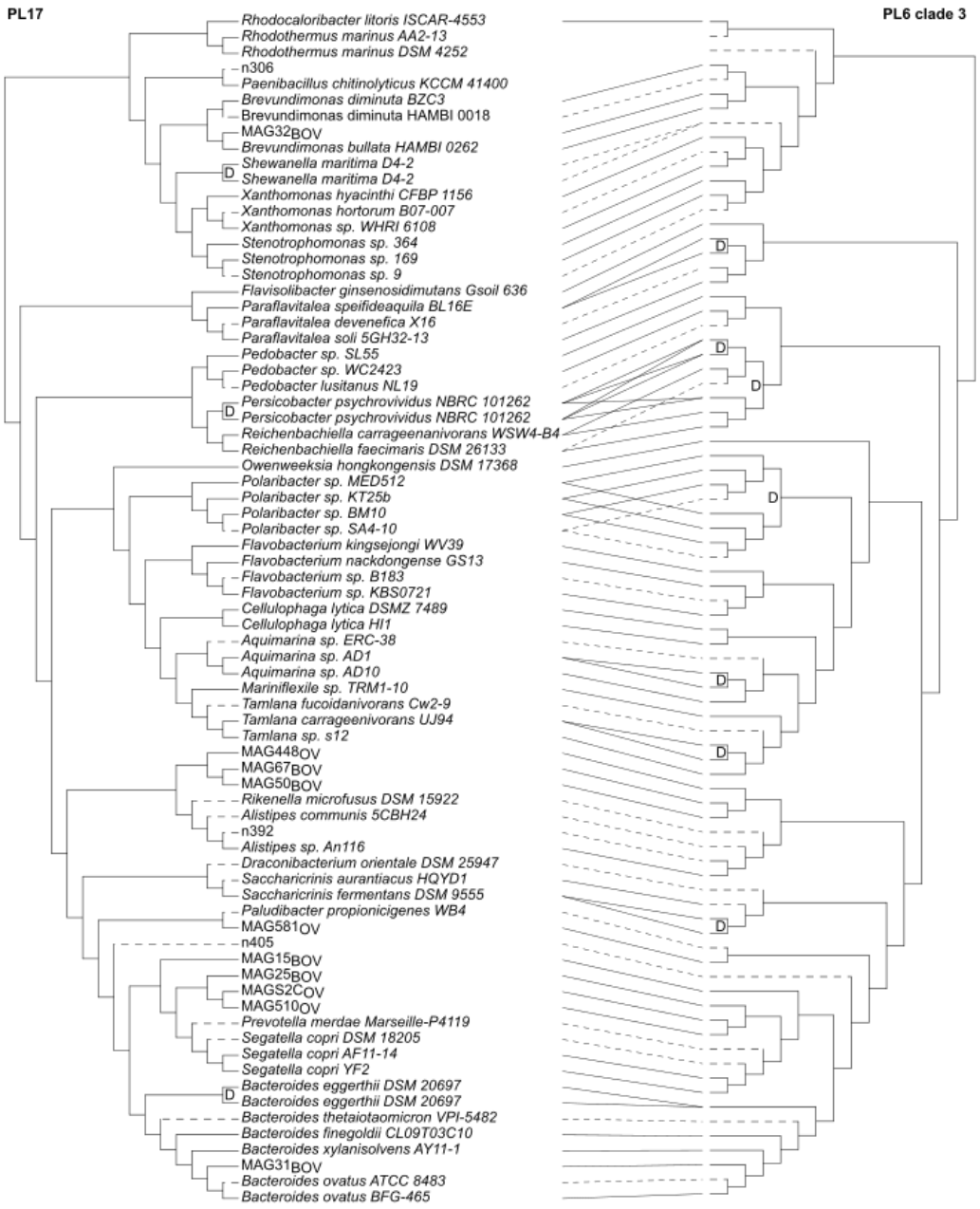

Figure A1.9: NOTUNG PL17 and PL6 Clade 3 gene trees. Comparison of PL17 and PL6 Clade 3 NOTUNG gene trees. Gene loss (dotted lines) and duplications "D" are marked per species and lines are drawn between trees to connect sequences within the same genome.

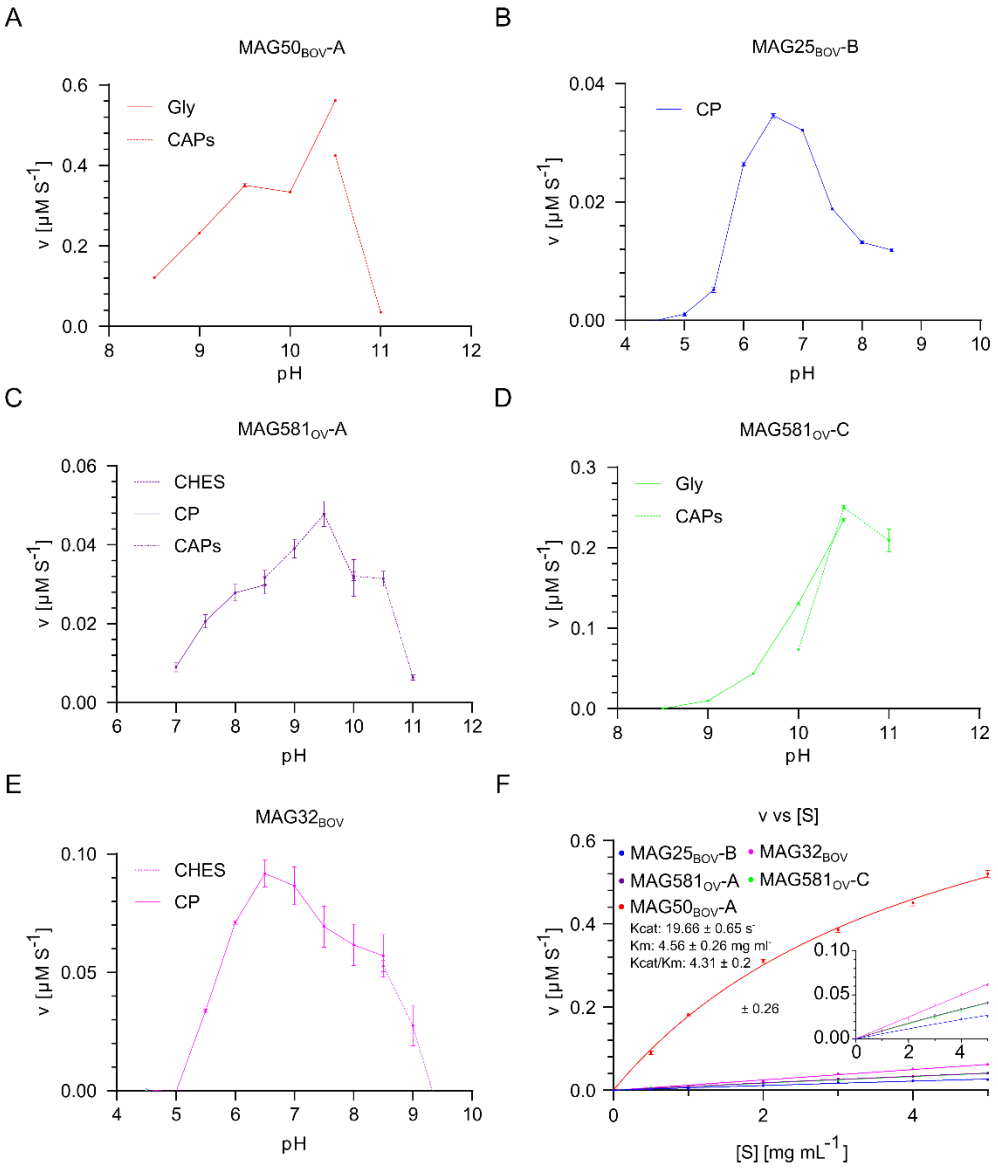

Figure A1.10: Activity of PL6 members of alginate types. A-E) pH optima of PL6 members within this study on brown seaweed alginate. Optima was determined using product formed (µM – calculated from absorbance at 232 nm) per second for 10 min reactions. Buffers: Gly – Glycine, CAPs - 3-(Cyclohexylamino)propane-1-sulfonic acid, CP – Citrate-phosphate, CHES - N-cyclohexyl-2-aminoethanesulfonic acid. **F)** initial velocities of PL6 members within study on brown seaweed alginate. Velocities were determined using product formed (µM – calculated from absorbance at 232 nm) per second against substrate concentration (mg mL^{-1}) for 10-minute reactions. Enzyme catalytic efficiency was calculated for $MAG50_{BOV}$-A.

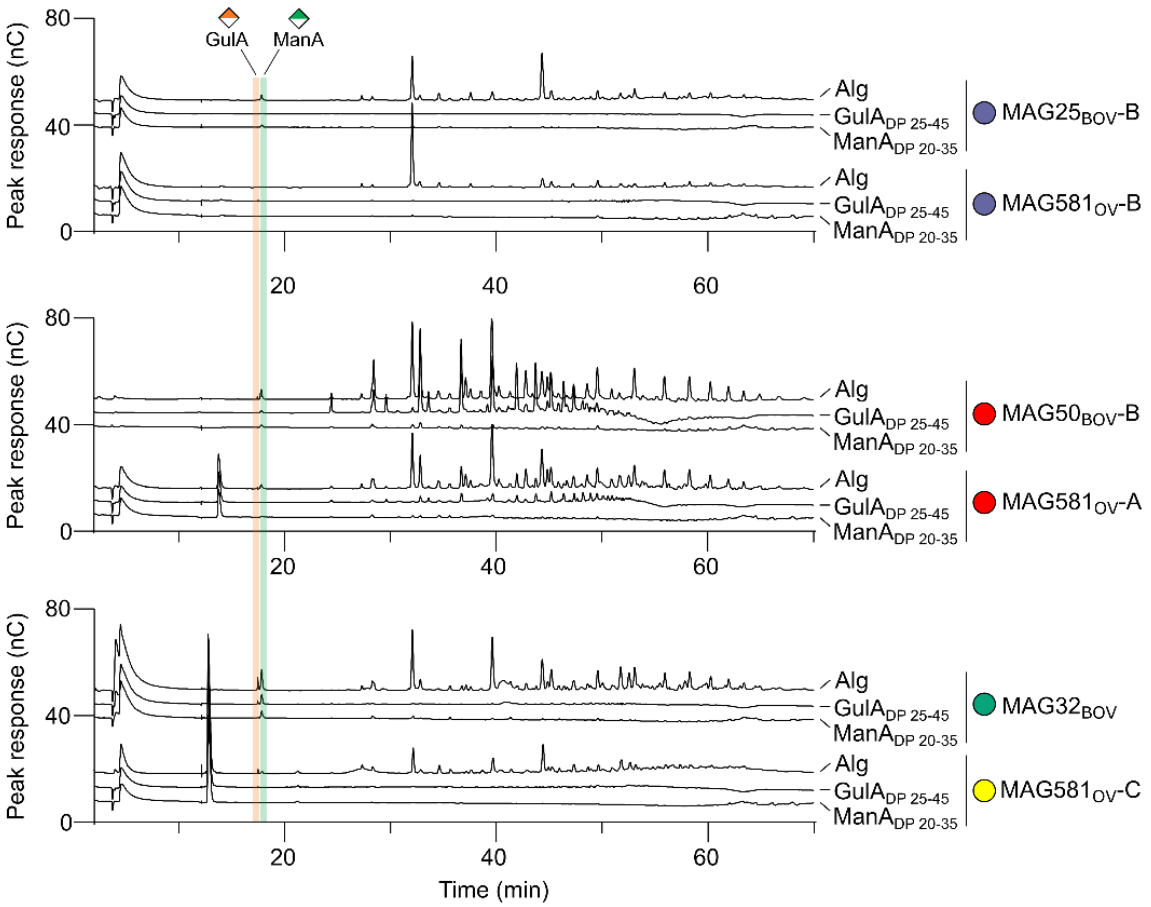

Figure A1.11: Characterization of PL6 members on alginate oligosaccharides. HPAEC-PAD product analysis of poly-GulpA (GulADP 25-45) and poly-ManpA (ManADP 20-35) digested with lamb and cattle-associated MAG PL6 enzymes. PL6 digests of alginate (Alg) not precipitated via EtOH were included as positive controls. Traces were separated into PL6 clade 2 (**top**), clade 1 (**middle**), and clade 3 and 4 (**bottom**) members. ManpA and GulpA represent monosaccharide standards.

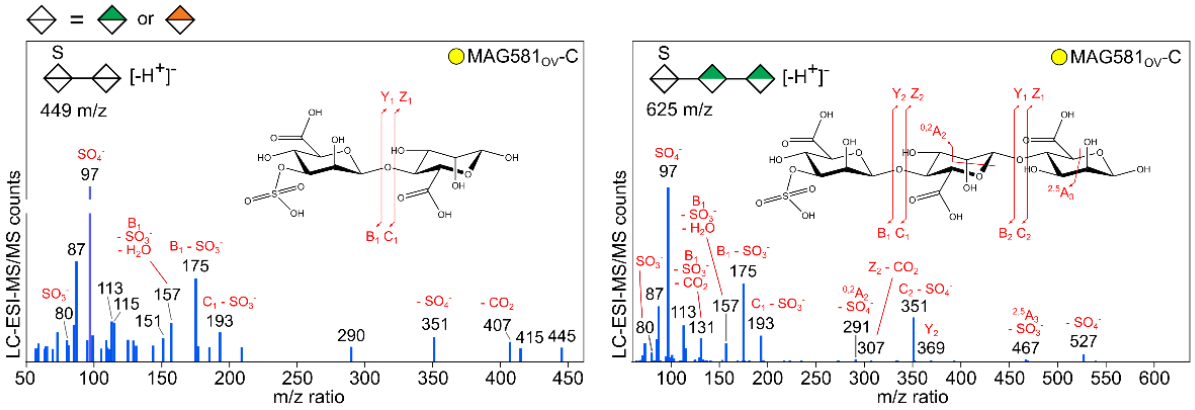

Figure A1.12: Sulfated alginate oligosaccharide enzyme products as identified by LC-ESI-MS/MS. Commercial alginic acid was incubated with MAG581-C enzyme and analyzed by LC-ESI-MS/MS. Oligosaccharide species for ions of m/z 449 and 625 were identified based on MS2 fragmentation data. Monosaccharide symbols are displayed according to the Symbol Nomenclature for Glycans system [426]. For monosaccharides unable to be discerned by LCESI-MS between mannuronate and guluronate, white symbols are used.

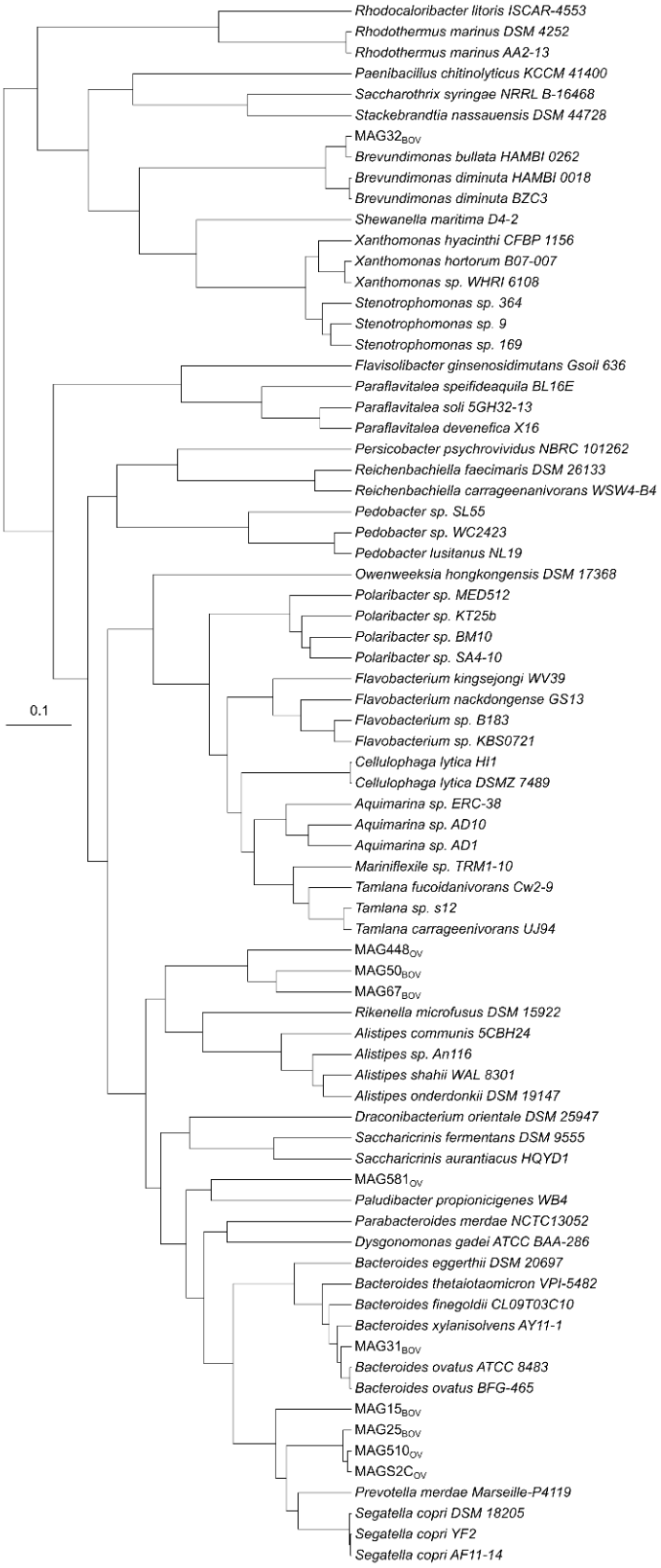
Rhodocaloribacter litoris ISCAR-4553
Rhodothermus marinus DSM 4252
Rhodothermus marinus AA2-13
Paenibacillus chitinolyticus KCCM 41400
Saccharothrix syringae NRRL B-16468
Stackebrandtia nassauensis DSM 44728
MAG32BOV
Brevundimonas bullata HAMBI 0262
Brevundimonas diminuta HAMBI 0018
Brevundimonas diminuta BZC3
Shewanella maritima D4-2
Xanthomonas hyacinthi CFBP 1156
Xanthomonas hortorum B07-007
Xanthomonas sp. WHRI 6108
Stenotrophomonas sp. 364
Stenotrophomonas sp. 9
Stenotrophomonas sp. 169
Flavisolibacter ginsenosidimutans Gsoil 636
Paraflavitalea speifideaquila BL16E
Paraflavitalea soli 5GH32-13
Paraflavitalea devenefica X16
Persicobacter psychrovividus NBRC 101262
Reichenbachiella faecimaris DSM 26133
Reichenbachiella carrageenanivorans WSW4-B4
Pedobacter sp. SL55
Pedobacter sp. WC2423
Pedobacter lusitanus NL19
Owenweeksia hongkongensis DSM 17368
Polaribacter sp. MED512
Polaribacter sp. KT25b
Polaribacter sp. BM10
Polaribacter sp. SA4-10
Flavobacterium kingsejongi WV39
Flavobacterium nackdongense GS13
Flavobacterium sp. B183
Flavobacterium sp. KBS0721
Cellulophaga lytica HI1
Cellulophaga lytica DSMZ 7489
Aquimarina sp. ERC-38
Aquimarina sp. AD10
Aquimarina sp. AD1
Mariniflexile sp. TRM1-10
Tamlana fucoidanivorans Cw2-9
Tamlana sp. s12
Tamlana carrageenivorans UJ94
MAG448OV
MAG50BOV
MAG67BOV
Rikenella microfusus DSM 15922
Alistipes communis 5CBH24
Alistipes sp. An116
Alistipes shahii WAL 8301
Alistipes onderdonkii DSM 19147
Draconibacterium orientale DSM 25947
Saccharicrinis fermentans DSM 9555
Saccharicrinis aurantiacus HQYD1
MAG581OV
Paludibacter propionicigenes WB4
Parabacteroides merdae NCTC13052
Dysgonomonas gadei ATCC BAA-286
Bacteroides eggerthii DSM 20697
Bacteroides thetaiotaomicron VPI-5482
Bacteroides finegoldii CL09T03C10
Bacteroides xylanisolvens AY11-1
MAG31BOV
Bacteroides ovatus ATCC 8483
Bacteroides ovatus BFG-465
MAG15BOV
MAG25BOV
MAG510OV
MAGS2COV
Prevotella merdae Marseille-P4119
Segatella copri DSM 18205
Segatella copri YF2
Segatella copri AF11-14
0.1

Figure A1.13: OrthoFinder generated species tree. The species tree was composed of AUL ruminant MAGs as well as NCBI genomes selected via their presence or absence (in the case of reference strains) of alginate lyases.

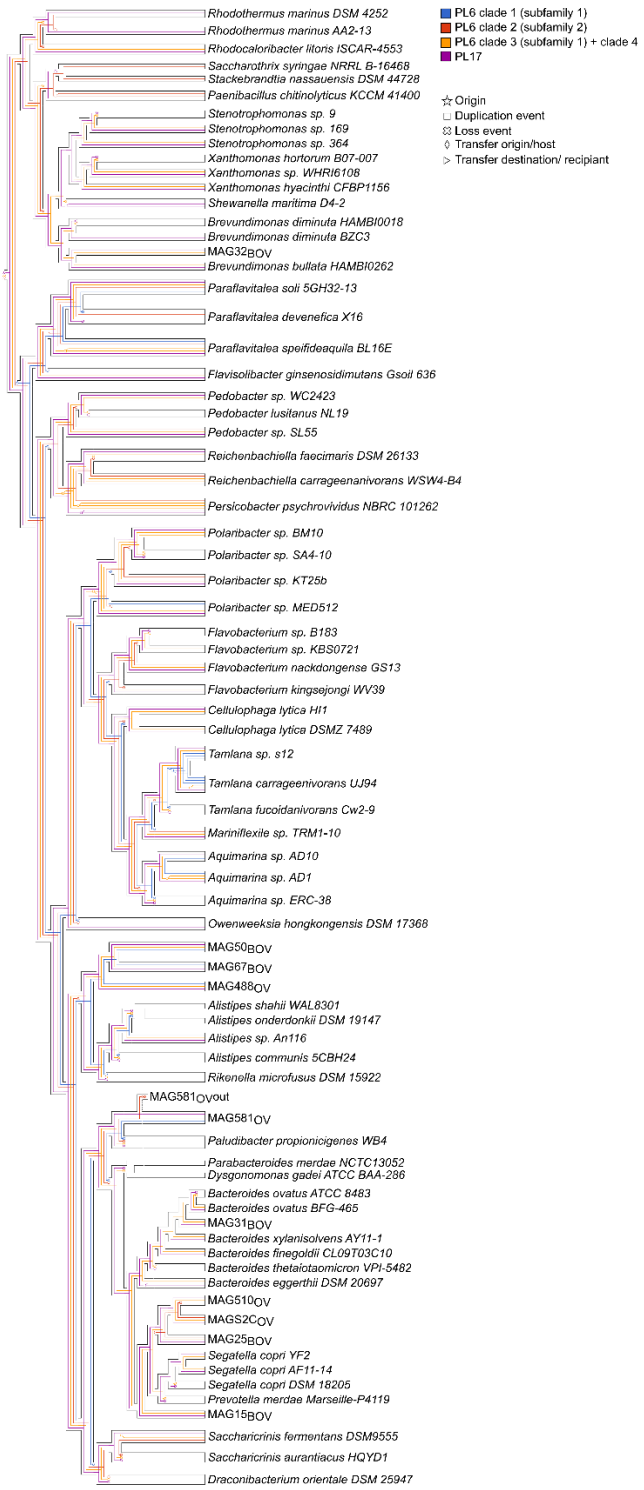
PL6 clade 1 (subfamily 1)
PL6 clade 2 (subfamily 2)
PL6 clade 3 (subfamily 1) + clade 4
PL17
Origin
Duplication event
Loss event
Transfer origin/host
Transfer destination/ recipiant
Rhodothermus marinus DSM 4252
Rhodothermus marinus AA2-13
Rhodocaloribacter litoris ISCAR-4553
Saccharothrix syringae NRRL B-16468
Stackebrandtia nassauensis DSM 44728
Paenibacillus chitinolyticus KCCM 41400
Stenotrophomonas sp. 9
Stenotrophomonas sp. 169
Stenotrophomonas sp. 364
Xanthomonas hortorum B07-007
Xanthomonas sp. WHRI6108
Xanthomonas hyacinthi CFBP1156
Shewanella maritima D4-2
Brevundimonas diminuta HAMBI0018
Brevundimonas diminuta BZC3
MAG32BOV
Brevundimonas bullata HAMBI0262
Paraflavitalea soli 5GH32-13
Paraflavitalea devenefica X16
Paraflavitalea speifideaquila BL16E
Flavisolibacter ginsenosidimutans Gsoil 636
Pedobacter sp. WC2423
Pedobacter lusitanus NL19
Pedobacter sp. SL55
Reichenbachiella faecimaris DSM 26133
Reichenbachiella carrageenanivorans WSW4-B4
Persicobacter psychrovividus NBRC 101262
Polaribacter sp. BM10
Polaribacter sp. SA4-10
Polaribacter sp. KT25b
Polaribacter sp. MED512
Flavobacterium sp. B183
Flavobacterium sp. KBS0721
Flavobacterium nackdongense GS13
Flavobacterium kingsejongi WV39
Cellulophaga lytica HI1
Cellulophaga lytica DSMZ 7489
Tamlana sp. s12
Tamlana carrageenivorans UJ94
Tamlana fucoidanivorans Cw2-9
Mariniflexile sp. TRM1-10
Aquimarina sp. AD10
Aquimarina sp. AD1
Aquimarina sp. ERC-38
Owenweeksia hongkongensis DSM 17368
MAG50BOV
MAG67BOV
MAG488OV
Alistipes shahii WAL8301
Alistipes onderdonkii DSM 19147
Alistipes sp. An116
Alistipes communis 5CBH24
Rikenella microfusus DSM 15922
MAG581OVout
MAG581OV
Paludibacter propionicigenes WB4
Parabacteroides merdae NCTC13052
Dysgonomonas gadei ATCC BAA-286
Bacteroides ovatus ATCC 8483
Bacteroides ovatus BFG-465
MAG31BOV
Bacteroides xylanisolvens AY11-1
Bacteroides finegoldii CL09T03C10
Bacteroides thetaiotaomicron VPI-5482
Bacteroides eggerthii DSM 20697
MAG510OV
MAGS2COV
MAG25BOV
Segatella copri YF2
Segatella copri AF11-14
Segatella copri DSM 18205
Prevotella merdae Marseille-P4119
MAG15BOV
Saccharicrinis fermentans DSM9555
Saccharicrinis aurantiacus HQYD1
Draconibacterium orientale DSM 25947

Figure A1.14: NOTUNG gene tree reconciliation. All NOTUNG gene trees were compiled and viewed via RecPhyloXML. Gene trees are represented via colored lines and the species tree is represented by the large white outline.

Appendix 1 Tables

Table A1.1: Relative abundances (Mol%) of glycosidic linkages identified from the cell walls of *Saccharina latissima* via GC-MS.

PMAA	Avg. Mol %	PMAA	Avg. Mol %	PMAA	Avg. Mol %
3-Araf	Trace	4-Glcp	35.0±6.4	t-Rhap	Trace
5-Araf	Trace	6-Glcp	Trace	2-Rhap	Trace
t-Fucp	2.1±0.2	2,3-Glcp	Trace	3-Rhap	Trace
2-Fucp	1.2±0.1	3,4-Glcp	2.7±0.2	4-Rhap	Trace
3-Fucp	3.2±0.2	3,6-Glcp	Trace	2,3-Rhap	Trace
4-Fucp	1.7±0.4	4,6-Glcp	2.5±0.4	2,4-Rhap	Trace
2,3-Fucp	1.1±0.2	2,3,6-Glcp	Trace	3,4-Rhap	Trace
2,4-Fucp	0.8±0.1	2,4,6-Glcp	Trace	2,3,4-Rhap	Trace
3,4-Fucp	2.0±0.6	3,4,6-Glcp	2.6±0.8	t-Xylp	1.5±0.3
2,3,4-Fucp	5.1±1.0	2,3,4,6-Glcp	0.8±0.3	2-Xylp	Trace
t-Galp	Trace	t-Manp	Trace	3-Xylp	Trace
2-Galp	Trace	2-Manp	1.2±0.2	4-Xylp	Trace
4-Galp	Trace	3-Manp	Trace	2,4-Xylp	Trace
6-Galp	Trace	4-Manp	Trace	3,4-Xylp	Trace
3,4-Galp	1.4±0.2	2,3-Manp	Trace	2,3,4-Xylp	Trace
3,6-Galp	1.1±0.2	2,4-Manp	1.2±0.2	t-GalpA	Trace
4,6-Galp	Trace	2,6,Manp	Trace	2,4-GlcpA+2,4-GalpA	Trace
2,3,6-Galp	Trace	3,4-Manp	Trace	t-GlcpA	Trace
2,4,6-Galp	Trace	3,6-Manp	Trace	3-GlcpA	2.6±0.1
3,4,6-Galp	1.8±0.4	4,6-Manp	Trace	4-GlcpA	2.4±0.1
2,3,4,6-Galp	Trace	2,3,6-Manp	0.7±0.2	t-GulpA	Trace
2,4-Glcp+2,4-Galp	1.6±0.3	2,4,6-Manp	Trace	4-GulpA	10.1±1.8
t-Glcp	Trace	3,4,6-Manp	Trace	t-ManpA	Trace
3-Glcp	0.9±0.0	2,3,4,6-Manp	Trace	4-ManpA	4.6±0.6

(n=3)

Table A1.2: The effect of 2.5% and 5% *S. latissima* on dry matter intake and apparent total tract digestibility of nutrients with Norwegian White lambs.

	Control	2.5% *S. latissima*	5% *S. latissima*	SEM	P-value
DMI, kg/d	1.55	1.53	1.55	0.04	0.66
Digestibility (ATTD), %					
DM	81.1	80.0	78.9	2.63	0.27
OM	81.7	80.5	79.4	2.66	0.19
CP	83.5	81.6	79.6	2.57	0.012
NDF	78.5	77.7	76.2	3.13	0.36
Starch	99.3	99.1	99.0	0.28	0.13
Ash	74.4	74.2	74.0	2.56	0.96

DMI, Dry matter intake; ATTD, apparent total tract digestibility of nutrients, DM, dry matter; OM, organic matter; CP, crude protein; NDF, neutral detergent fiber.

Table A1.3: The effect of 2.0% *S. latissima* on the degradability of nutrients and total volatile fatty acid production in the RUSITEC system.

	Control	2% *S. latissima*	SEM	P-value
Degradability, %				
DM	57.1	58.4	0.80	<0.001
OM	58.0	57.3	1.67	0.45
CP	69.3	69.0	0.85	0.77
NDF	36.1	35.9	2.75	0.87
Total VFA, mmol	68.4	68.3	1.82	0.95

DM, dry matter; OM, organic matter; CP, crude protein; NDF, neutral detergent fiber; VFA, volatile fatty acids

Table A1.4: Species - gene tree reconciliation genomes.

GeneBank ID	Species name	GeneBank ID	Species name
GCA_006542665.1	*Alistipes communis* 5CBH24	GCA_040026395.1	*Pedobacter lusitanus* NL19
GCA_025145285.1	*Alistipes onderdonkii* DSM 19147	GCA_026625705.1	*Pedobacter* sp. SL55
GCA_025145845.1	*Alistipes shahii* WAL 8301	GCA_040822065.1	*Pedobacter* sp. WC2423
GCA_002161005.1	*Alistipes* sp. An116	GCA_036492425.1	*Persicobacter psychrovividus* NBRC 101262
GCA_003443695.1	*Aquimarina* sp. AD1	GCA_002005425.1	*Polaribacter* sp. BM10
GCA_003443715.1	*Aquimarina* sp. AD10	GCA_900105145.1	*Polaribacter* sp. KT25b
GCA_026222555.1	*Aquimarina* sp. ERC-38	GCA_000152945.2	*Polaribacter* sp. MED152
GCA_025146565.1	*Bacteroides eggerthii* DSM 20697	GCA_002163835.1	*Polaribacter* sp. SA4-10
GCA_040687835.1	*Bacteroides finegoldii* CL09T03C10	GCA_900290275.1	*Prevotella merdae Marseille*-P4119
GCA_001314995.1	*Bacteroides ovatus* ATCC 8483	GCA_025639805.1	*Reichenbachiella carrageenanivorans* WSW4-B4
GCA_024758985.1	*Bacteroides ovatus* BFG-465	GCA_900176375.1	*Reichenbachiella faecimaris* DSM 26133
GCA_022453665.1	*Bacteroides thetaiotaomicron* VPI-5482	GCA_011682235.2	*Rhodocaloribacter litoris* ISCAR-4553
GCA_029369765.1	*Bacteroides xylanisolvens* AY11-1	GCA_009936255.1	*Rhodothermus marinus* AA2-13
GCA_034424665.1	*Brevundimonas bullata* HAMBI 0262	GCA_000024845.1	*Rhodothermus marinus* DSM-4252
GCA_002205555.1	*Brevundimonas diminuta* BZC3	GCA_000427365.1	*Rikenella microfusus* DSM 15922
GCA_034424705.1	*Brevundimonas diminuta* HAMBI 0018	GCA_001660705.2	*Saccharicrinis aurantiacus* HQYD1

GCA_000190595.1	*Cellulophaga lytica* DSMZ 7489	GCA_000517085.1	*Saccharicrinis fermentans* DSM 9555
GCA_000750195.1	*Cellulophaga lytica* HI1	GCA_009498035.1	*Saccharothrix syringae* NRRL B-16468
GCA_900111425.1	*Draconibacterium orientale* DSM 25947	GCA_003465445.1	*Segatella copri* AF11-14
GCA_000213555.1	*Dysgonomonas gadei* ATCC BAA-286	GCA_020735445.1	*Segatella copri* DSM 18205
GCA_007970805.1	*Flavisolibacter ginsenosidimutans* Gsoil 636	GCA_015074785.1	*Segatella copri* YF2
GCA_003076475.1	*Flavobacterium kingsejongi* WV39	GCA_004295345.1	*Shewanella maritima* D4-2
GCA_004355225.1	*Flavobacterium nackdongense* GS13	GCA_000024545.1	*Stackebrandtia nassauensis* DSM 44728
GCF_023539155.1	*Flavobacterium* sp. B183	GCA_014621775.1	*Stenotrophomonas* sp. 169
GCF_002007065.3	*Flavobacterium* sp. KBS0721	GCA_009832905.1	*Stenotrophomonas* sp. 364
GCA_003425985.1	*Mariniflexile* sp. TRM1-10	GCA_031582865.1	*Stenotrophomonas* sp. 9
GCA_000236705.1	*Owenweeksia hongkongensis* DSM 17368	GCA_002893765.1	*Tamlana carrageenivorans* UJ94
GCA_004117095.1	*Paenibacillus chitinolyticus* KCCM 41400	GCA_045788535.1	*Tamlana fucoidanivorans* Cw2-9
GCA_000183135.1	*Paludibacter propionicigenes* WB4	GCA_016767215.1	*Tamlana* sp. s12
GCA_900445495.1	*Parabacteroides merdae* NCTC13052	GCA_002285515.1	*Xanthomonas hortorum* B07-007
GCA_011759375.1	*Paraflavitalea devenefica* X16	GCA_009769165.1	*Xanthomonas hyacinthi* CFBP 1156

GCA_003555545.1	*Paraflavitalea soli* 5GH32-13	GCA_040202385.2	*Xanthomonas* sp. WHRI 6108
GCA_032395985.1	*Paraflavitalea speifideaquila* BL16E		

Table A1.5: NOTUNG event scores.

Gene	Event score	Duplication	Transfers	Losses
PL6 Clade 1	34.5	7	0	24
PL6 Clade 2	25	2	1	19
PL6 Clade 3	19	2	0	16
PL17	25.5	3	0	21

Table A1.6: LC-ESI-MS Gradient conditions for separation of alginate oligosaccharides.

Time (min)	A (%) 10 mM ammonium formate 50 mM formic acid 80% acetonitrile 20% water	B (%) 10 mM ammonium formate 50 mM formic acid 20% acetonitrile 80% water
0	100	0
1	70	30
60	20	80
61	20	80
67	100	0
68	100	0
75	100	0

Table A1.7: Parameters for ESI-MSn on the Orbitrap Fusion Tribrid.

Parameter (units)	Value
ESI	
Spray voltage: negative ion (V)	2500
Sheath Gas (Arb)	35
Aux Gas (Arb)	10
Sweep Gas (Arb)	1
Ion Transfer Tube Temp (°C)	325
Vaporizer Temp (°C)	250
MS	
Detector Type	Orbitrap
Orbitrap Resolution	120K
Mass Range	Normal
Scan Range (m/z)	150-2000
RF Lens (%)	50
MS2	
Collision Energy Type	Normalized
Isolation Mode	Quadrupole
Activation Type	HCD
Collision Energy Mode	Stepped
Collision Energies (%)	30,45,60,80
Detector Type	Orbitrap
Orbitrap Resolution	30K

Appendix 1 Text

Text S1: Results

Digestibility and degradability impacts of *S. latissima* in diets

Evaluating the digestibility of seaweed in animal diet is crucial for understanding its effect on the overall ruminal digestive activity and nutrient utilization. Thus, ruminal kelp digestion was assessed both in the *in vivo* lamb feeding and the *in vitro* bovine RUSITEC experiments.

For the lamb experiment, mean dry matter intake (DMI, kg/d) was not affected by the inclusion of *S. latissima* to diets. However, *S. latissima* inclusion at 5% (DM, basis) reduced apparent total tract digestibility of crude protein by 4.67% relative to the control diet (**Table A1.2**). Digestibility of other tested dietary components were not significantly affected despite numerical differences. For the RUSITEC experiment, inclusion of 2% *S. latissima* increased the *in vitro* degradability of DM by 2.24%. There were no effects of *S. latissima* inclusion on organic matter, crude protein, neutral detergent fiber, or production of total volatile fatty acids (**Table A1.3**).

Text S2: Material and Methods

Fluorescent Polysaccharides (FLAPS)

A hot water extraction (HWE) was performed on the dried, ball-milled *S. latissima* sample by incubating 1 g of sample in 40 mL distilled water for 8 h at 70 °C. Samples were centrifuged (3,000 × *g*, 10 min) and the supernatant was transferred to a new tube, and the hot water incubation was repeated twice more. Supernatants from the same sample were pooled and freeze-dried. The dried *S. latissima* HWE was de-starched by incubation with α-amylase (E-BLAAM, Megazyme) in 100 mM maleic acid buffer (pH 6.0) containing 100 mM NaCl and 3 mM $CaCl_2$ for 8 h at 70 °C, followed by an incubation with amyloglucosidase (E-AMGDPD, Megazyme) for 4 h at 50 °C. Resulting de-starched samples were extensively dialyzed (3,500 Da MWCO) against distilled water, before being freeze-dried. Fluorescently labelled *S. latissima* HWE (FLA-SLAT) was produced using a previously defined protocol [249], where the purified FLA-SLAT was freeze-dried, covered in aluminum foil and stably stored at -20 °C until further use. For FLA-SLAT

incubations, 1 mL from each RUSITEC vessel was collected at the day 15 time point, and were centrifuged (5,000 × *g*, 10 min). Pellets were suspended in 1 mL phosphate buffered saline (PBS; pH 7.4) and washed twice with PBS before a final resuspension in 2 mL PBS. 50 µL of each resuspended RUSITEC microbial community sample was incubated with 50 µL of 0.4% FLA-SLAT for 1 day within an anaerobic chamber (atmosphere: 85% N_2, 10% CO_2, 5% H_2, at 37 °C), where 40 µL aliquots were taken at the 1 h and 1 day time points. The 1 h and 1 day aliquots were immediately centrifuged (5,000 × *g*, 10 min), and the pellet was fixed with 4% (v/v) formaldehyde overnight at 4 °C. Fixed samples were centrifuged (5,000 × *g*, 10 min) and pellets were washed twice with PBS before being resuspended in 1 mL PBS and stored at 4 °C. Samples were diluted 1:10 and filtered onto a 25 mm, 0.2 µm pore size Isopore™ filter (Sigma, USA) using a gentle vacuum of <200 mbar. Dried filter pieces were counterstained with 4'6-diamidino-2-phenylindole (DAPI) and mounted on a glass microscope slide using a 4:1 mixture of Citiflour™ AFI mountant solution to Vectashield® vibrance antifade mounting medium (Vector Laboratories, USA). The fluorescence images were taken using an Echo Revolve R4K (Upright & Inverted Capability) microscope equipped with motorized LED fluorescence light, 5 MP CMOS monochrome camera, a Plan X Apo oil; 1,42 NA, Revolve, 60x oil immersion objective, with LED light cubes DAPI (EX: 385/30 EM: 450/50 DM: 425) and LED light cubes FITC (EX: 470/40 EM: 525/50 DM: 495). SR-SIM images for RUSITEC samples were visualized on a Zeiss ELYRA PS.1 (Carl Zeiss) using 561 and 488m lasers and BP 573-613, BP 502-538 and BP 420-480 + LP 750 optical filters. Z-stack images were taken with a Plan-Apochromat 63 Å~ /1.4 Oil objective and processed with the software ZEN2011 (Carl Zeiss). Images were exported to ACMEtool software (M. Seder, Technology GmbH, http://www.technobiology.ch and Max Planck Institute for Marine Microbiology, Bremen), where signals were evaluated according to Bennke, Reintjes [427]. Cell enumeration counts were plotted in GraphPad Prism (8.0.2) and displayed as relative abundance values.

16S rRNA gene sequencing

Taxonomic assigned ASV from the lamb rumen microbiome were combined with variant abundance table and processed using phyloseq (v1.46.0) [341]. Reads flagged as

Eukaryota, Chloroplast or Mitochondria were removed before the sequences were normalized for downstream analysis. The beta diversity was investigated using a Non-metric Multidimensional scaling (NMDS) analysis based on Bray-Curtis dissimilarity in vegan [428].

RUSITEC community analysis, statistics, and plotting of the Bracken output files were performed using R v4.2.2 [276] with the packages: phyloseq [341], ggplot2 [343], picante [429], rioja [292] and vegan [428]. Each sample was rarefied to 28,530 random reads using the "rarefy_even_depth(sample.size = 0.9*raremax)" function of phyloseq [341] prior to alpha- and beta-diversity analyses. NMDS analyses based on Bray-Curtis dissimilarity indices were performed using the "vegdist()" function of vegan [428] to visualize the separation of RUSITEC microbial communities between the different treatments and time points. Stress was determined to be 0.097 with the "metaMDS()" function [428], and a Shepard diagram was used to determine the non-metric (R2= 0.991) and linear fit (R2= 0.968) between the observed dissimilarity. Statistical evaluation of the RUSITEC microbial communities was performed with an analysis of similarity (ANOSIM) using distance matrices ["anosim()" function, distance = bray, 999 permutations] and a permutational multivariate analysis of variance (PERMANOVA) using distance matrices ["vegdist()" function, method = bray, 999 permutations] [428]. This analysis was followed by a pairwise comparison of the treatment or time point effect [pairwise.perm.manova()" function, method = Euclidian, "999 permutations] of RVAideMemoire [430]. Observed OTUs, Shannon, and inverse Simpson indices were calculated using the "estimate_richness()" function of phyloseq, as a measure of microbial community alpha-diversity. Changes in alpha-diversity were assessed in GraphPad Prism v8.0.2 using the Kruskal-Wallis test with a post-hoc Dunn's multiple comparisons.

Metaproteomics

After protein extraction, peptide samples were processed using a nano LC-MS/MS coupled to a timTOF Pro mass spectrometer (Bruker, Germany). The peptides were separated by an Aurora C18 reverse-phase (1.6 µm, 120 Å) 25 cm x 75 µm analytical column with an integrated emitter (IonOpticks, Melbourne, Australia). The temperature of the column was kept at 50 °C using the integrated oven. Equilibration of the column was

performed before the samples were loaded (equilibration pressure 800 bar). The flow rate was set to 300 nl min^{-1} and the samples were separated using a solvent gradient from 2% to 25% solvent B over 70 minutes, and to 37% over 9 minutes. The solvent composition was then increased to 95% solvent B over 10 min and maintained at that level for an additional 10 min. In total, a run time of 99 min was used for the separation of the peptides. Solvent A is 0.1% (v/v) formic acid in milliQ water, while solvent B is 0.1% (v/v) formic acid in LC-MS grade acetonitrile. The timsTOF Pro was run in positive ion data dependent acquisition PASEF mode with the control software Compass Hystar v5.1.8.1 and timsControl v1.1.19 68. The acquisition mass range was set to 100 – 1700 m z^{-1}. The TIMS settings were: 1 $K0^{-1}$ Start 0.85 V·s cm^{-2} and 1 $K0^{-1}$ End 1.4 V·s cm^{-2}, ramp time 100 ms, ramp rate 9.42 Hz, and duty cycle 100%. The capillary voltage was set at 1400 V, dry gas at 3.0 l min^{-1}, and dry temp at 180 ℃. The MS/MS settings were the following: number of PASEF ramps 10, total cycle time 0.53 sec, charge range 0-5, scheduling target intensity 20,000, intensity threshold 2,500, active exclusion release after 0.4 min, and CID collision energy ranging from 27-45 eV.

Recovery of virome scaffolds from lamb rumen

In addition to the prokaryote population, we attempted to recover viral content from the lamb metagenome co-assembly. This was done following the viral sequence identification SOP v3 (DOI:dx.doi.org/10.17504/protocols.io.bwm5pc86), with VirSorter2 v2.2.3 [431] and CheckV v0.8.1 [432]. DRAM-v.py was ran to annotate the sequences identified as viral, followed by manual curation according to the SOP. Annotated virome scaffolds are available in FigShare via doi:10.6084/m9.figshare.28925894. As the virome scaffolds were included in the sequence database for metaproteomics analysis, those with protein detection are provided in **Extended Tables 2**. Further exploration of the rumen virome is beyond the scope of this study.

Appendix 1 Extended data

The following tables are submitted as external tables due to their size:

Extended Tables 1 & Extended Tables 2

Appendix 2 Supplemental Material to Chapter 3

Appendix 2 Figures

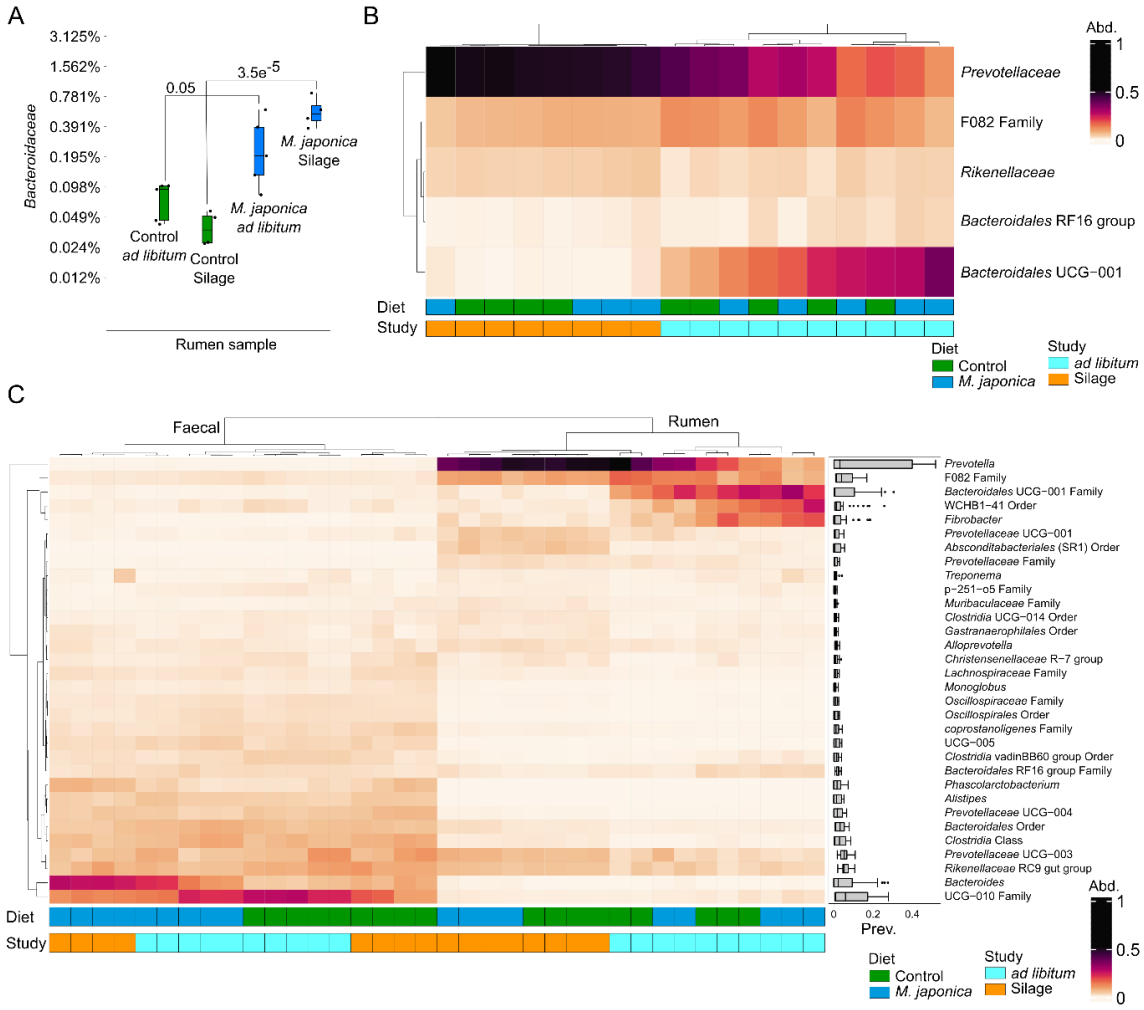

Figure A2.1: *M. japonica* supplementation influences the faecal microbiome greater than the rumen microbiome. A) % abundance change of *Bacteroidaceae* between diet and study rumen samples. A linear regression model of % composition was generated using microViz [320], and a Tukey test was used to measure significance (Adjusted P value) [276] **B)** 16S rRNA gene relative abundance of Genus' above 10,000 reads within rumen samples. **C)** 16S rRNA gene relative abundance of Genus' above 10,000 reads within rumen and faecal samples. Diet (Control: n=9 and *M. japonica*: n=9 supplemented) and Study (ad libitum: n=10 and Silage: n=8) are denoted below the heatmap.

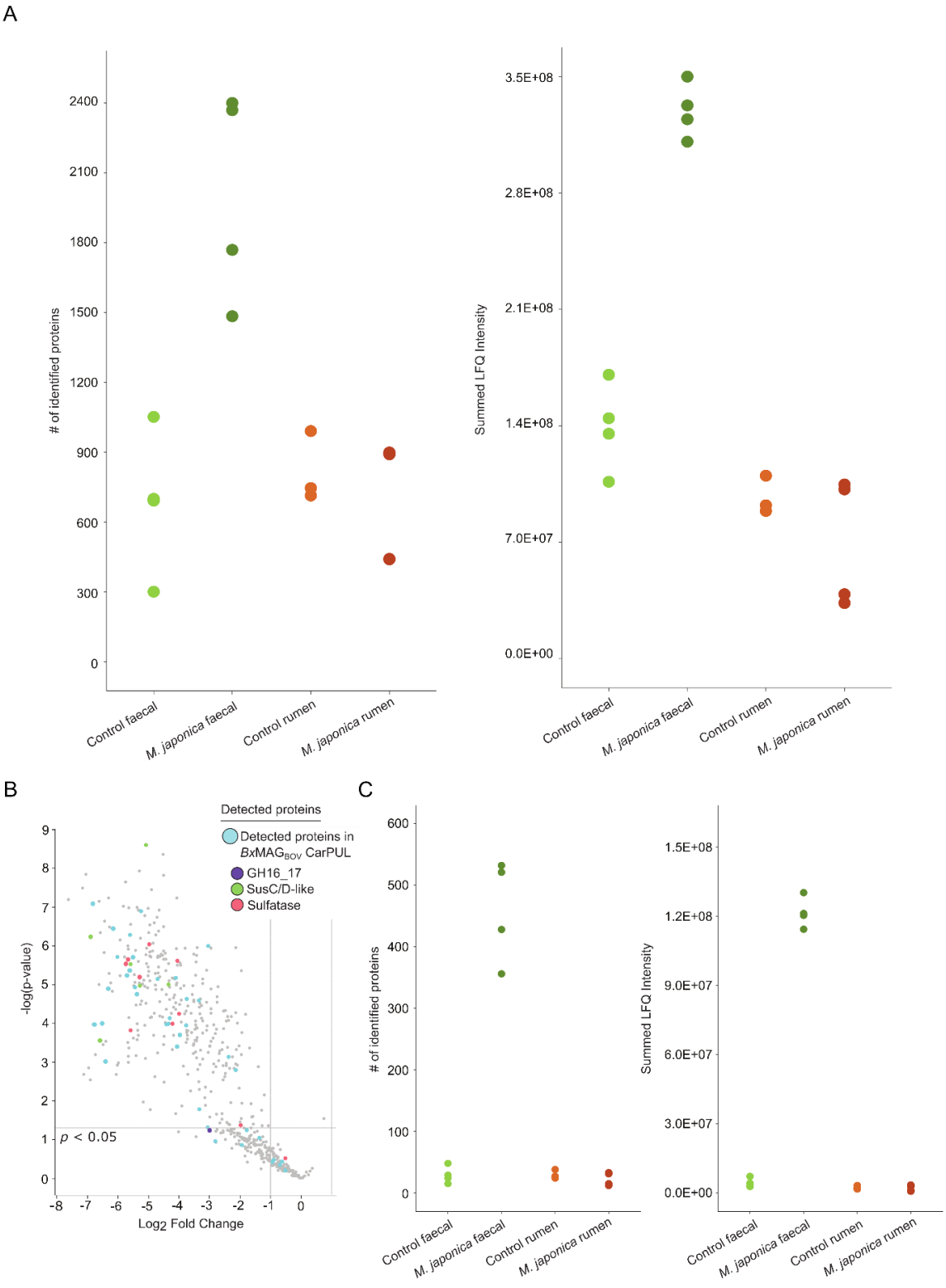
A
# of identified proteins
2400
2100
1800
1500
1200
900
600
300
0
Summed LFQ Intensity
3.5E+08
2.8E+08
2.1E+08
1.4E+08
7.0E+07
0.0E+00
Control faecal
M. japonica faecal
Control rumen
M. japonica rumen
B
Detected proteins
Detected proteins in BxMAG_BOV CarPUL
GH16_17
SusC/D-like
Sulfatase
-log(p-value)
p < 0.05
Log2 Fold Change
C
# of identified proteins
Summed LFQ Intensity
1.5E+08
1.2E+08
9.0E+07
6.0E+07
3.0E+07
0.0E+00

Figure A2.2: Proteomic expression of rumen and fecal samples supplemented with *M. japonica*. A) Dot plots for left: total proteins identified, and right: summed metaproteomic expression, presented as summed LFQ intensity in rumen and faecal samples from cattle (n=4) fed a control diet and a silage diet supplemented with 5% *M. japonica*, in addition to environmental samples from the sample site. Detected LFQ intensities for identified proteins can be found in **Table A2.13 B)** Volcano plot showing *Bx*MAG$_{BOV}$ proteins with significantly different protein expression recovered from faecal samples from cattle fed diet with 5% inclusion level of *M. japonica* vs those fed the control diet. Difference in protein expression is displayed as Log2 fold change in LFQ intensities. **C)** Dots plots above were condensed to show total proteins identified (**left**) and LFQ intensity (**right**) from only proteins within Bx.

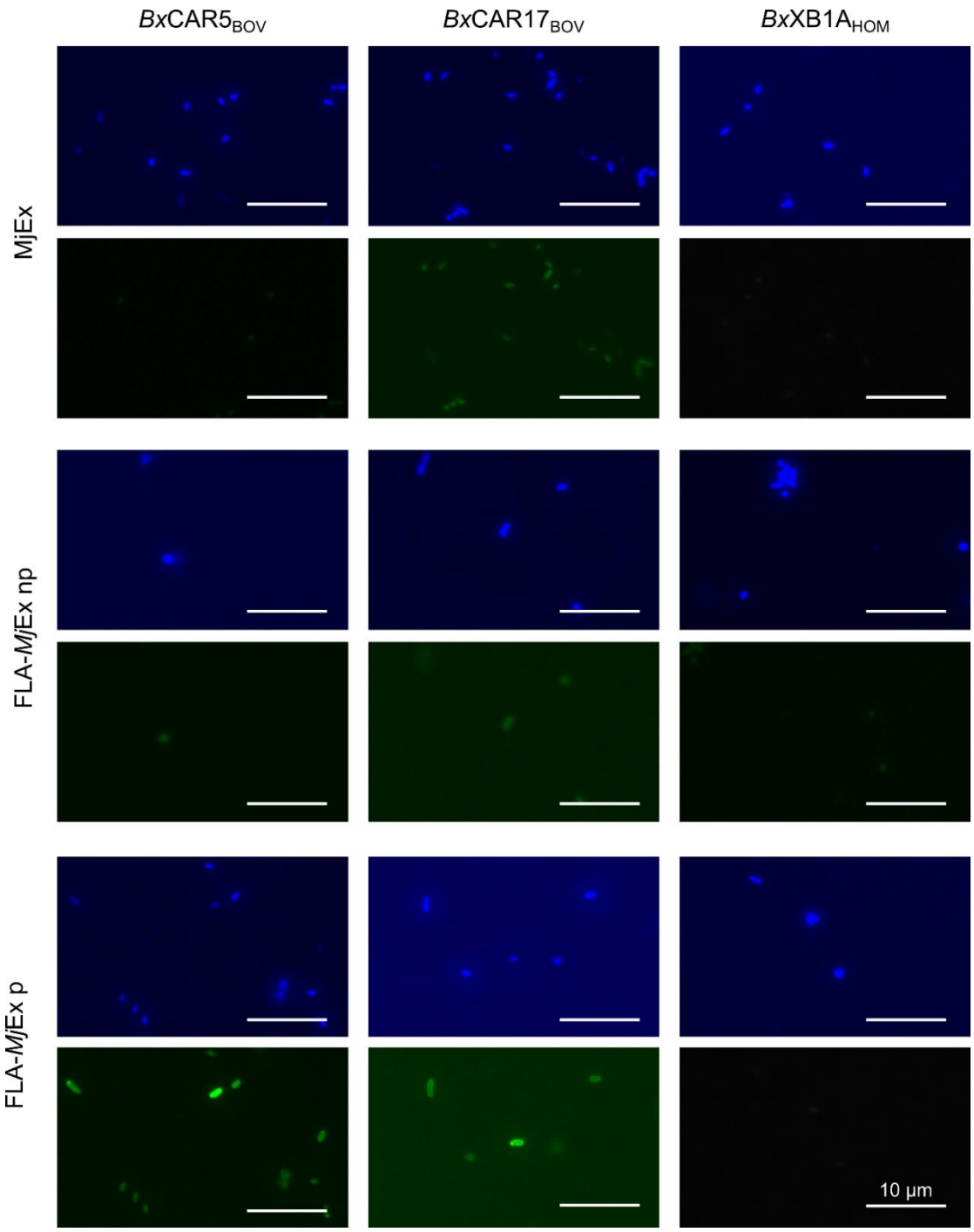
*Bx*CAR5$_{BOV}$
*Bx*CAR17$_{BOV}$
*Bx*XB1A$_{HOM}$
MjEx
FLA-*Mj*Ex np
FLA-*Mj*Ex p
10 µm

Figure A2.3: Visualization of *Bacteroides* spp. interacting with fluorescently labelled *M. japonica* extract. BxCAR5$_{BOV}$, BxCAR17$_{BOV}$, and BxXB1A$_{HOM}$, were incubated with 0.2% of unlabelled *Mj*Ex and FLA-*Mj*Ex for 1d. Samples were co-stained with DAPI and imaged using an epifluorescence microscope (LED light cubes DAPI (EX: 385/30 EM: 450/50 DM: 425, ED light cubes FLA-*Mj*Ex (EX: 470/40 EM: 525/50 DM: 495). np – not primed; isolate inoculum grown on galactose for 24 h, not primed for growth on *Mj*Ex; p – primed; isolate inoculum grown for 24 h on *Mj*Ex. Scale bar = 10 μm

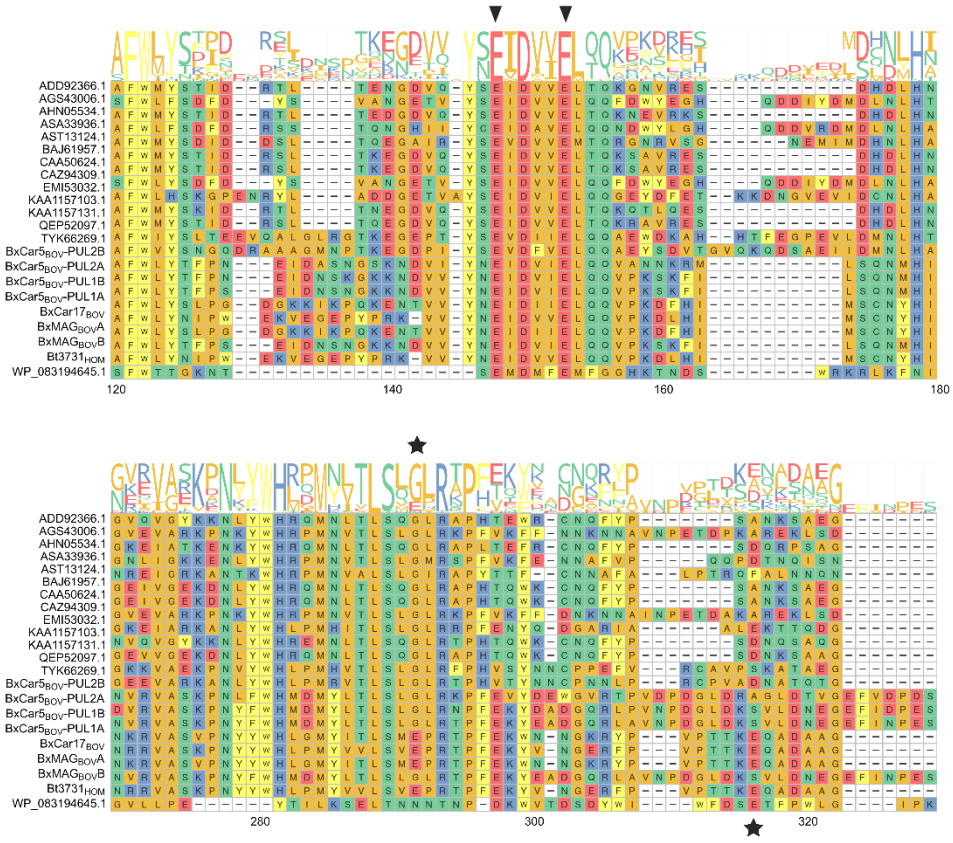

Figure A2.4: Sequence alignment of GH16_17 members. Sequence alignment of characterized GH16_17 members, and BxCarPUL$_{BOV}$ GH16s. GH16_13 member (WP_083194645.1; Cgbk16A_Wf) was added as a outgroup. Catalytic acid and nucleophiles glutamic acids (E) are denoted with black triangles. The black star denotes the amino acid, glycine – G or glutamic acid – E, which allows for or occludes 4S-modified anhydro-galactose in the -1 subsite, respectively.

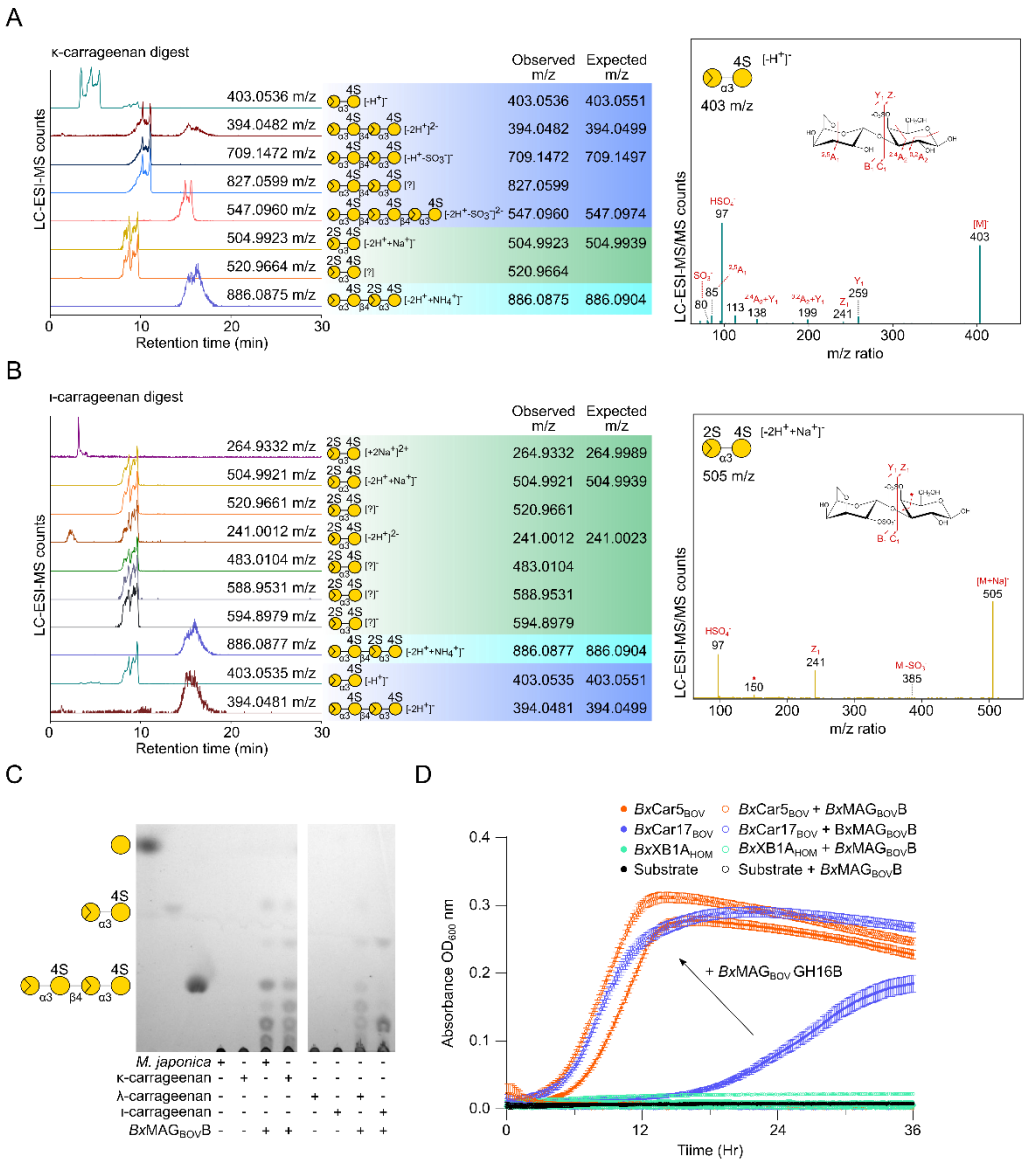

Figure A2.5: Analysis of *M. japonica*, κ-carrageenan, and ι-carrageenan enzyme digests. Carrageenan substrates were incubated with *Bx*MAG$_{BOV}$ GH16B enzyme and analyzed by LC-ESI-MS. ESI-MS/MS with HCD was performed in order to identify the extracted ions, and MS2 product ion spectra are shown with the carbohydrate fragmentation depicted and identified. **A)** Extracted ion chromatograms are shown locally normalized for **left:** κ-carrageenan digest and **right:** the identified ion neo-κ-carrabiose. **B)** Extracted ion chromatograms are shown locally normalized for **left:** ι-carrageenan digest and **right:** the identified ion neo-ι-carrabiose. Both neo-κ-carrabiose and neo-ι-carrabiose as majority enzymatic products from their respective digests. **C)** TLC of commercially available carrageenans and *Mj*Ex digested with *Bx*MAG$_{BOV}$ GH16_17B. Products compared to galactose, neo-κ-carrabiose, and neo-κ-carratetraose standards. Monosaccharide symbols are displayed according to the Symbol Nomenclature for Glycans system [55]. **D)** Growth cultures of *Bx*Car isolates on 0.3% *M. japonica*

supplemented with 1 mM *Bx*MAG$_{BOV}$ GH16_17B. OD_{600} nm was observed every 10 minutes. Negative ODs were excluded from the plot.

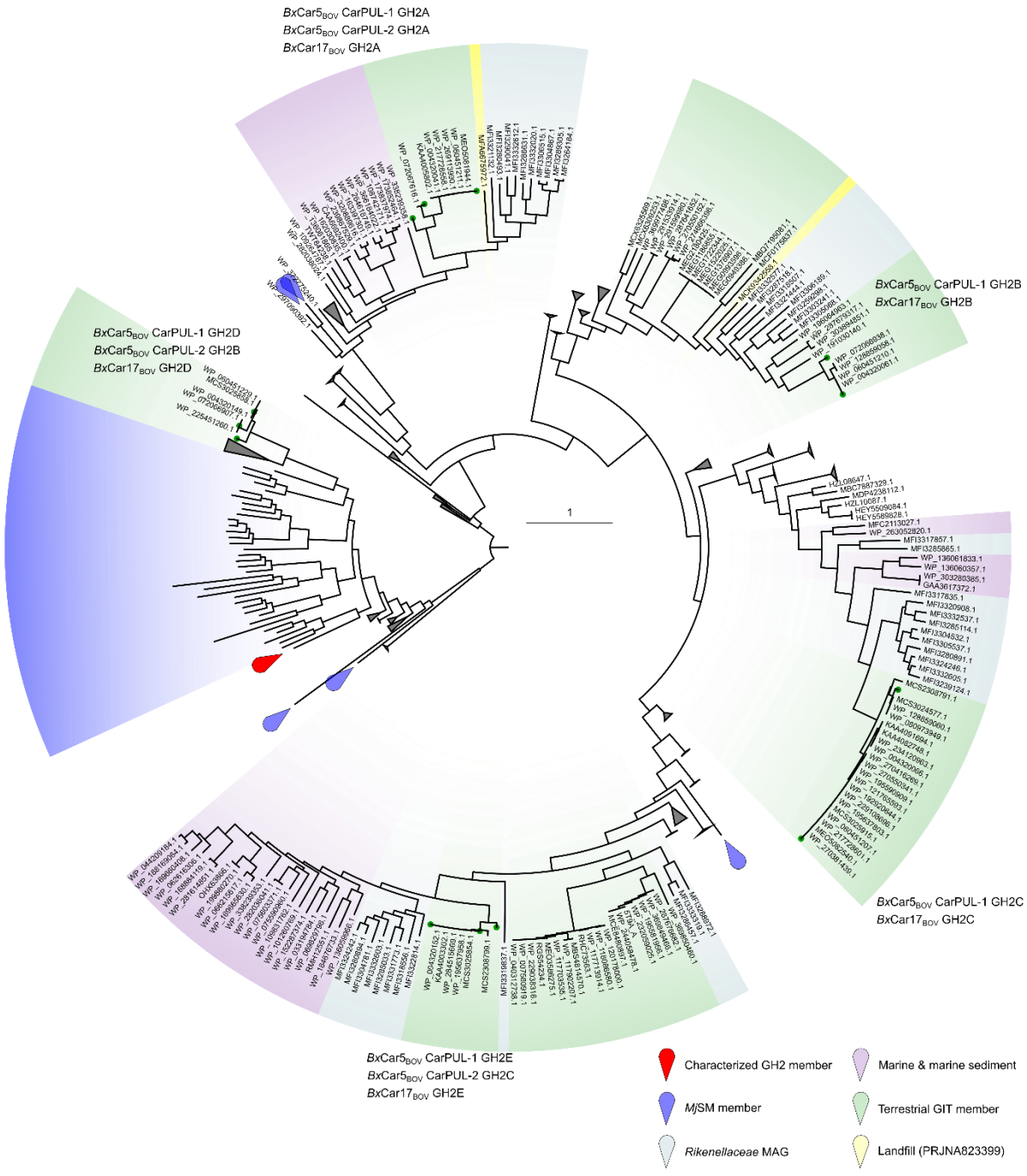
BxCar5BOV CarPUL-1 GH2A
BxCar5BOV CarPUL-2 GH2A
BxCar17BOV GH2A
BxCar5BOV CarPUL-1 GH2B
BxCar17BOV GH2B
BxCar5BOV CarPUL-1 GH2D
BxCar5BOV CarPUL-2 GH2B
BxCar17BOV GH2D
BxCar5BOV CarPUL-1 GH2C
BxCar17BOV GH2C
BxCar5BOV CarPUL-1 GH2E
BxCar5BOV CarPUL-2 GH2C
BxCar17BOV GH2E
1
Characterized GH2 member
MjSM member
Rikenellaceae MAG
Marine & marine sediment
Terrestrial GIT member
Landfill (PRJNA823399)

Figure A2.6: Homology of GH2s from CarPULs bacteria in the NCBI database. SACCHARIS phylogenies [154] were generated between the top 100 BLAST hits towards BxCar$_{BOV}$ CarPUL GH2 members, characterized GH2 members from the CAZy database, predicted GH2 members from *Mj*SM metagenomes, and *Bx*Car$_{BOV}$ CarPUL GH2s. Clades are colored based on their host environment; Green: terrestrial vertebrate *Bacteroides* spp., blue: $_{Mj}$SM members, teal: *K. sydneyanus* GIT members, purple: marine/sediment microorganisms, and yellow: landfill. Individual nodes are denoted by a pin, and CarPUL members are denoted with labelling and green dots at the node. BLAST hits are detailed in **Table A2.9**. Clades absent of CarPUL members were collapsed for phylogeny clarity.

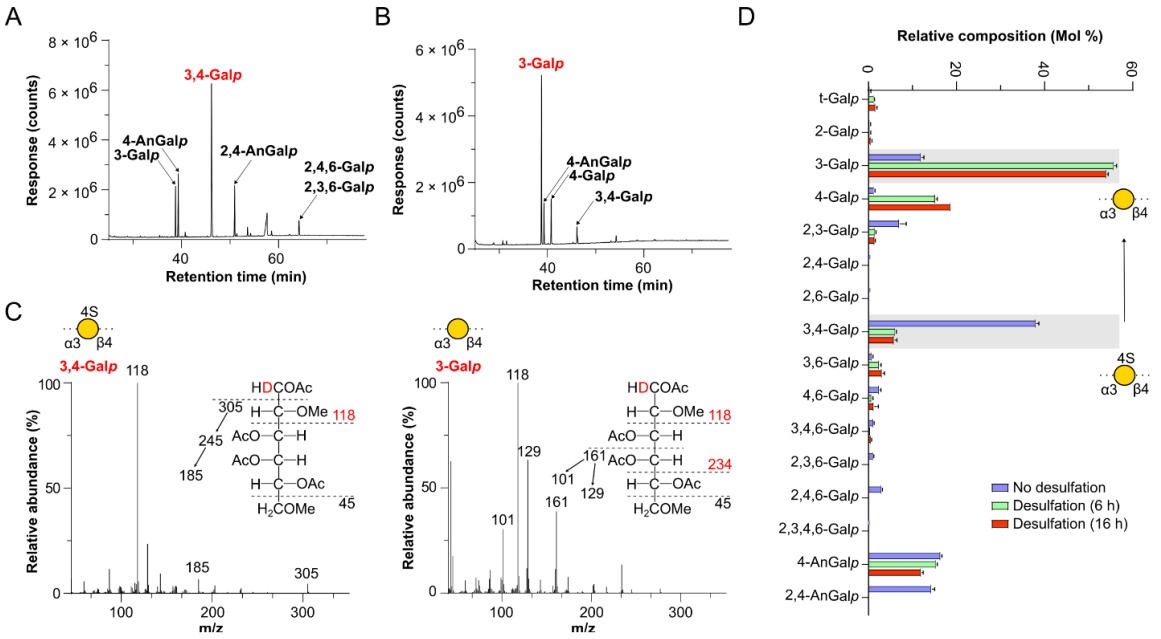

Figure A2.7: Glycosidic linkage analysis (methylation-GC-MS analysis) of *M. japonica* galactan and its desulfation product. GC-TIC chromatograms of PMAAs from destarched, water-soluble *Mj*Ex **A)** before solvolytic desulfation and **B)** after 16 h of solvolytic desulfation. **C)** EI-MS spectra and ion fragmentation patterns of PMAAs from the largest peaks within GC-TIC chromatograms. Top: no desulfation (3,4-Galp). Bottom: 16 h desulfation (3-Galp). **D)** Comparison of relative composition (Mol%) of observed PMAAs from samples with no desulfation and those subjected to 6 h and16 h of desulfation.

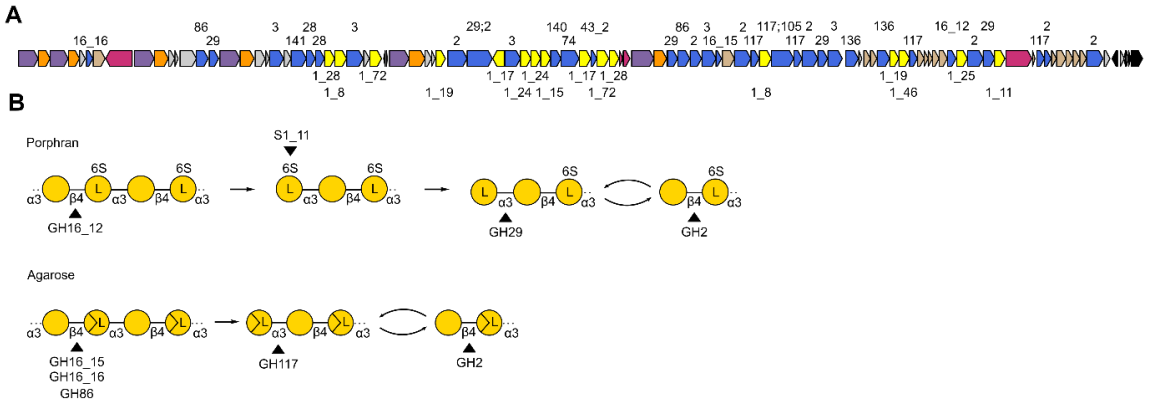

Figure A2.8: Porphyran/agarose PUL identified within giraffe metagenome. A) PUL diagram of the porphyran/agarose PUL within *Bz*Car$_{GIR}$. B) Predicted depolymerization pathway of porphyran and agarose in *Bz*Car$_{GIR}$.

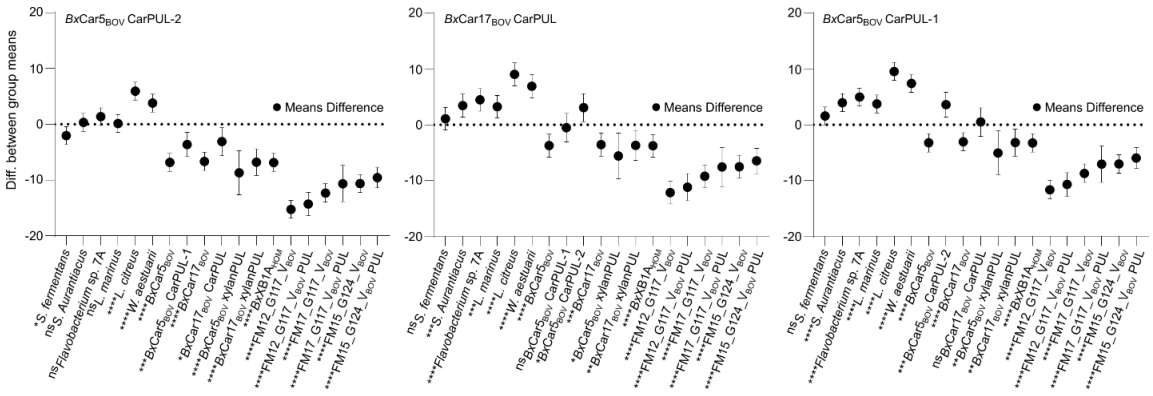

Figure A2.9: Mean % GC content between *Bx*Car CarPULs and *Bx*Car genomes is significantly different. Mean %GC content of CarPULs was compared to that of their genome (whole genome and terrestrial xylan PULs [433] within their genome), marine pelagic and sediment microbiota gene %GC content, and gene content of 3 *R. Alistipes* MAGs isolated from the GIT of *K. sydneyanus* (FM12, FM15, FM17). From left to right: *Bx*Car5$_{BOV}$ PUL-2, *Bx*Car17$_{BOV}$ CarPUL*,* *Bx*Car5$_{BOV}$ PUL-1. Mean %GC abundance between *Bx*Car PULs and control and marine datasets were compared using Graphpad prism. One-Way Anova with Games-Howell multiple comparisons were calculated against each PUL to identify the difference between the means. Along with previously stated controls, a native xylan PUL from each isolate was used. ns: $P > 0.05$; *: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$; ****: $P \leq 0.0001$.

Appendix 2. Tables

Table A2.5: Ruminant read sets for Table A3.8 & A3.9.

Cattle		
	PRJNA631951	doi.org/10.1038/s41396-020-00837-2
	PRJEB21624	doi.org/10.1038/s41467-018-03317-6
	PRJEB31266	doi.org/10.1038/s41587-019-0202-3
	PRJNA60251	doi.org/10.1126/science.1200387
	PRJEB23561	doi.org/10.1093/gigascience/giaa057
	PRJEB39057	doi.org/10.1186/s13059-020-02144-7
	PRJNA485488	doi.org/10.1128/MRA.01279-18
	PRJNA419351	doi.org/10.1128/AEM.00610-18
	PRJNA657455; PRJNA657473	doi.org/10.1186/s40168-021-01078-x
	PRJEB34458	doi.org/10.1038/s41598-021-81668-9
	PRJNA438833	doi.org/10.1038/s41467-019-12111-x
	PRJNA379303	doi.org/10.3389/fmicb.2019.01980
Moose		
	PRJNA301235	doi.org/10.1038/s41564-018-0225-4
	PRJEB12797	doi.org/10.1038/ismej.2017.108
Yak		
	PRJNA624740	doi.org/10.1186/s12866-020-01993-3
	PRJNA657455; PRJNA657473	doi.org/10.1186/s40168-021-01078-x
Sheep		
	PRJNA595610	doi.org/10.1038/s41587-021-01130-z
	PRJEB34458	doi.org/10.1038/s41598-021-81668-9
	PRJNA202380	doi.org/10.1038/s41598-020-61942-y
	PRJNA657455; PRJNA657473	doi.org/10.1186/s40168-021-01078-x
Water buffalo		
	PRJNA656389	doi.org/10.1038/s41467-022-28402-9
	PRJNA657455; PRJNA657473	doi.org/10.1186/s40168-021-01078-x
Deer (water, roe, red)		
	PRJEB34458	doi.org/10.1038/s41598-021-81668-9
	PRJNA657455; PRJNA657473	doi.org/10.1186/s40168-021-01078-x

Table A2.7: Buffalo reads sets chosen for MEGAHIT assembly.

SRA	Sample	Breed	Source
SRX9190418	NM8	Guang Xi swamp buffalo	Cecum
SRX9189964	LCN20	Guang Xi swamp buffalo	Feces
SRX9189955	LCN11	Guang Xi swamp buffalo	Feces
SRX9189916	LCN9	Guang Xi swamp buffalo	Feces
SRX9189915	LCN8	Guang Xi swamp buffalo	Feces
SRX9189912	LCN5	Guang Xi swamp buffalo	Feces
SRX9189909	LCN2	Guang Xi swamp buffalo	Feces
SRX9189908	LCN1	Guang Xi swamp buffalo	Feces
SRX9190312	GBS4_S182_L002	Guang Xi swamp buffalo	Feces
SRX9190363	NB8	Guang Xi swamp buffalo	Omasum
SRX9190124	NL8	Guang Xi swamp buffalo	Rumen
SRX9190127	S-1_BDME192032754-1a	Murrah buffalo	Rumen

Table A2.12: Bovine diets in this study.

Ingredient	AF kgs	DM kgs
Mineral	0.1	0.1
Whey	23	1.5
Grass Hat	2	1.74
Oat Hulls	1	0.75
Tofu Blend	3	1.27
Hydrogreen	14.5	3.34
3rd Cut Grass Silage	5.75	2.67
*Mazzaella japonica**	0.13	0.11

*silage diet

Table A2.15: Parameters for ESI-MSn on the Orbitrap Fusion Tribrid.

Parameter (units)	Value
ESI	
Spray voltage: negative ion (V)	2500
Sheath Gas (Arb)	45
Aux Gas (Arb)	10
Sweep Gas (Arb)	1
Ion Transfer Tube Temp (°C)	325
Vaporizer Temp (°C)	200
MS	
Detector Type	Orbitrap
Orbitrap Resolution	120K
Mass Range	Normal
Scan Range (m/z)	150-2000
RF Lens (%)	50
MS2	
Collision Energy Type	Normalized
Isolation Mode	Quadrupole
Activation Type	HCD
Collision Energy Mode	Stepped
Collision Energies (%)	30,45,60,80
Detector Type	Orbitrap
Orbitrap Resolution	30K

Table A2.16: Data collection and refinement statistics of *Bx*MAG$_{BOV}$ GH16_17A.

Parameter (units)	*Bx*MAG$_{BovA}$
Data Collection	
Beamline	CMCF-ID (CLS)
Wavelength	0.95372
Space Group	$P22_12_1$
a, b, c (Å)	56.72, 69.63, 109.06
Resolution (Å)	30.00-2.00 (2.05-2.00)
R_{meas}	0.162 (1.140)
R_{pim}	0.071 (0.493)
CC1/2	0.994 (0.586)
<I/sI>	7.5 (1.6)
Completeness (%)	99.5 (99.6)
Redundancy	4.7 (4.9)
No. of Reflections	139297 (10590)
No. Unique	29675 (2176)
Refinement	
Resolution (Å)	2
Reflections used in refinement	29642 (2644)
Reflections used for R-free	1473 (130)
R_{work}	0.175 (0.271)
R_{free}	0.212 (0.320)
No. of Atoms	
Protein	2562 (310 residues)
Non-hydrogen atoms	2873
Ligands	43
Water	268
***B*-factors**	
Average B-factor	32.55
Protein	31.71
Ligand	43.47
Water	38.85
RMS Bond Lengths (Å)	0.007
RMS Bond Angles (°)	0.896
Ramachandran Preferred (%)	95.5

Ramachandran Allowed (%)	4.5
Ramachandran Disallowed (%)	0
Rotamer outliers (%)	0
Clashscore	2.36
PDB ID	9EFL

Appendix 2 Text

Text A2.1: Supplemental Note

Spatial digestion of *M. japonica* in bovine GIT

Bacteroidota, primarily *Prevotella*, are widely recognized as one of the most abundant fibre degrading bacterial phyla in the rumen [264, 434]. Often overlooked, however, is the prowess of *Bacteroides* within the hindgut of ruminants that forage on undigested feed residues that bypass the rumen [264], such as cellulose and hemicelluloses [31, 435]. This can be attributed to increased bulk within the digestive tract, which could be impacted by seaweed polysacchrides as observed in the digesta of mice [436]. Metagenome-guided metaproteomics analysis presented here suggests that *Bacteroides*-affiliated MAGs are one of the primary mechanisms for microbial degradation and utilization of *M. japonica* carrageenans in the lower GIT of cattle **(Figure 3.1; Figure A2.2C)**. There was a minor increase in ruminal *Bacteroidaceae* under *M. japonica* supplementation, which paled in comparison to the increase in fecal samples. This is likely due to *Bacteroides* spp. being more abundant within the lower gastrointestinal tract [264] (**Figure A2.1A**). These findings suggest the disparity between research on the rumen microbiome compared to the lower GIT microbiome, may hamper the discovery of seaweed degrading pathways within ruminants. Further, the impact carrageenans, and potentially other seaweed polysaccharides, has on the lower GIT microbiome may provide targeted prebiotic benefits or function as delivery systems to the hindgut of ruminants.

Text A2.2: Supplemental Results

Confirmation of *M. japonica* well wall polysaccharides

Sulfated galactans were extracted from *M. japonica* and studied by linkage analysis via methylation-GC-MS analysis of PMAAs prepared from the native and desulfated products [437]. Methylation-GC of the sulfated galactan without desulfation showed a high level of 3,4-Gal*p* at 38.1%, followed by 4-AnGal*p* (16.5 %), 2,4-AnGal*p* (14.4 %), 3-Gal*p* (12.1 %), 2,3-Gal*p* (7.1 %), 2,4,6-Gal*p* (3.1 %), 4,6-Gal*p* (2.5 %) (**Figure A2.7A**), and various other minor Gal*p* linkages (**Table A2.17**). Compared to the sample without desulfation, the sample subjected to 6 h of solvolytic desulfation [438] was dominated by

3-Gal*p* (55.8 %), followed by 4-AnGal*p* (15.5 %), 4-Gal*p* (15.3 %), 3,4-Gal*p* (6.2 %), 3,6-Gal*p* (2.5 %), and other minor Gal*p* linkages (**Table A2.17**). This was not appreciably different than desulfation at 16 h (**Figure A2.7B; Table A2.17**). Glycosidic linkages of galactoses were confirmed based on the EI-MS ion fragmentation patterns of their PMAAs (**Figure A2.7C**), with reference to published patterns in the literature [439]. The results suggested that the majority of 3,4-Gal*p* was converted to 3-Gal*p* due to desulfation at the *O*-4 position (**Figure A2.7D**). A considerable amount of 4-Gal*p* was also detected in the desulfated galactan, indicating that the 4-linked galactose residues were sulfated as well. There was no 2,4-AnGal*p* detected in the desulfated samples, indicating the complete conversion of 2,4-AnGal*p* to 4-AnGal*p* by desulfation at the *O*-2 position. The total level of AnGal in the desulfated sample was less than half of that in the untreated sample, indicating partial degradation of AnGal caused by solvolytic desulfation [438]. This was further supported by the lower level of anhydro sugar in the 16 h treatment compared to the 6 h one (**Table A2.17**).

*Bx*Car CarPUL %GC content is distinct *Bx*Car genomes

Mean GC content percentage between the *Bx*Car$_{BOV}$ genomes and CarPULs was compared to that of control *Bx*XB1A$_{HOM}$, xylan PULs (terrestrial plant polysaccharide) from *Bx*Car5$_{BOV}$ and 17, and marine/sediment microorganisms: *L. citreus* and *S fermentans, L. marinus, Saccharicrinis aurantiacus, Flavobacterium sp. 7A,* and *Wenyingzhuangia aestuarii*. These were selected based on the high homology observed in CarPUL CAZyme BLAST results. The mean GC% from *Bx*CAR$_{BOV}$ CarPUL genes was more similar to those of marine pelagic and sediment organisms genes than to the remainder of the BxCAR$_{BOV}$ genome (**Figure A2.9**). Through Games Howell's multiple comparisons tests, the mean GC% of *Bx*Car5$_{BOV}$ CarPUL-1 (39.3 ± 4.8) was more closely aligned with the *S. fermentans* genome (37.7 ± 3.8^{ns}) than to itself (42.5 ± 5.3****), a terrestrial xylan PUL [433] within the *Bx*Car5$_{BOV}$ genome (44.4 ± 5.0**), and the reference *Bx*XB1A$_{HOM}$ genome (42.6 ± 5.0****); this pattern is repeated between the *Bx*Car17$_{BOV}$ CarPUL and similar datasets. *Bx*Car5$_{BOV}$ CarPUL-2 (35.7 ± 3.6) showed the highest similarity towards marine pelagic and sediment organisms, with GC% values showing no

significant difference to ORFs within *S. aurantiacus* (35.4 ± 3.4^{ns}), *Flavobacterium* sp. 7A (34.3 ± 3.6^{ns}), and *L. marinus* (35.6 ± 3.1^{ns}).

Text A2.3: Material and Methods

Comparative Analysis of GC Content

*Bx*MAG$_{BOV}$, *Bx*Car5$_{BOV}$, *Bx*Car17$_{BOV}$, control strain *B. xylanisolvens Bx*XB1A$_{HOM}$, and marine environment bacterium sharing the highest homology with carrageenan PUL CAZymes including *R. Alistipes* MAGs: FM_12_G117_V, FM_17_G117_V, FM_15_G124_V, *Wenyingzhuangia aestuarii* (GCF_011927765.1), *Lutibacter citreus* (GCF_003260195.1), *Labilibacter marinus* (GCF_001659685.2), *Flavobacterium* sp. 7A (GCF_025960985.1), *Saccharicrinis aurantiacus* (GCF_947489045.1), and *Saccharicrinis fermentans* (GCF_000517085.1) were annotated with prodigal (v2.6.3) to obtain GFF files. *R. Alistipes* MAGs were chosen based on their synteny with CarPULs. GC usage was calculated as a % for each gene and GraphPad Prism was used to calculate means and deviation, and Brown-Forsythe and Welch ANOVAs. Games-Howell's comparisons was done between carrageenan PULs and genomes to compare differences between GC means. Asterisks denote P value cutoff's (ns: $P > 0.05$; *: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$; ****: $P \leq 0.0001$).

Extraction of crude polysaccharide

Ball-milled dry powder of *M. japonica* (137.6 g) was soaked in 1.2 L of hexane under magnetic stirring for 2 h in a 2 L glass beaker covered with aluminum foil to reduce evaporation. The mixture was left undisturbed overnight to allow the powder to settle. The supernatant was carefully decanted using a glass pipette, leaving approximately 1 cm of liquid above the interface to avoid disturbing the precipitate. An additional 1.2 L of hexane was added, and the extraction process was repeated. The precipitate was then resuspended in 1.2 L of 95% (v/v) ethanol, magnetically stirred for 8 h, and centrifuged at 3000 g for 30 min at room temperature. The resulting pellet was resuspended in 1.2 L of 80% (v/v) ethanol, with the same stirring and centrifugation conditions applied, and the extraction process was repeated. The resulting pellet was evaporated to dryness in 50 mL centrifuge tubes using SpeedVac (Thermo Fisher Scientific, MA, USA). The dried sample was then

transferred to a glass beaker and underwent two rounds of water extraction at room temperature with constant magnetic stirring, followed by three rounds of hot water extraction at 70 °C in an incubator. Each extraction used 2 L of deionized water and lasted 8 h, with the beaker covered with aluminum foil. After each extraction, centrifugation (3,000 × *g*, 30 min, room temperature) was conducted, and the supernatant was collected while the pellet was carried forward to the next extraction. Supernatants from all the water extractions were pooled, poured into 40 L of absolute ethanol, and left at room temperature overnight, followed by centrifugation (3,000 × *g*, 30 min, room temperature). The precipitate was evaporated to dryness in 50 mL centrifuge tubes using the SpeedVac, redissolved in 2 L of deionized water by incubating at 70 °C overnight, and freeze-dried (56.7 g).

Amylase treatment of crude polysaccharide

Dry crude polysaccharide (14.4 g) was dissolved in 1 L of deionized water by incubating at 70 °C overnight. After that, 2 mL of thermostable α-amylase (3000 units/mL, Megazyme, Ireland) was added, and the mixture was incubated at 70 °C for 8 h [440]. The solution was then poured into 4 L of absolute ethanol, left standing at 4 °C overnight, and centrifuged (3,000 × *g*, 30 min, room temperature). The residue was evaporated to dryness using the SpeedVac, redissolved in 500 mL of deionized water by incubating at 70 °C overnight, extensively dialyzed with molecular weight cut-off (MWCO) of 6,000-8,000 Da against deionized water at 4 °C, and freeze-dried (13.8 g).

Purification of sulfated galactan by gradient ethanol precipitation

Amylase-treated polysaccharide (915 mg) was dissolved in 400 mL of deionized water by incubating at 70 °C overnight. The resulting solution was cooled to room temperature, left standing at 4 °C overnight, and then centrifuged (3,000 × *g*, 30 min, room temperature). The supernatant was vigorously stirred magnetically to create a water tunnel in a 2 L glass beaker. Absolute ethanol was slowly added dropwise to achieve a 15% (w/w) ethanol concentration. The mixture was kept at 4 °C for 8 h, followed by centrifugation (3,000 × *g*, 30 min, room temperature). Absolute ethanol was then added dropwise to the vigorously stirred supernatant until the ethanol concentration reached 30% (w/w), followed

by standing at 4 °C and centrifugation as described above. The resulting supernatant underwent two additional cycles of ethanol precipitation, with ethanol concentrations gradually increased to 45% and 60% (w/w), respectively. The final supernatant was evaporated to dryness in 50 mL centrifuge tubes using the SpeedVac. The dry sample was redissolved in 100 mL of deionized water by incubating at 70 °C overnight, and the resulting solution was freeze-dried (0.9 g, designated as fraction F60).

Glycosidic linkage analysis of sulfated galactan and its desulfation products

F60 (10 mg) was dissolved in 10 mL of deionized water by magnetic stirring at 70 °C overnight, followed by 24 h of dialysis (MWCO 6,000-8,000 Da) against 4 L of 0.1 M pyridine hydrochloride, then another 24 h of dialysis against 4 L of deionized water, and freeze-dried [245]. The resulting pyridinium salt of the sulfated galactan was subjected to solvolytic desulfation by heating in 10 mL of a DMSO/methanol mixture (9:1, v/v) at 80 °C for 6 h with magnetic stirring, followed by extensive dialysis (MWCO 6,000-8,000 Da) against deionized water [438]. The sample was permethylated using 1.2 mL of methyl iodide in 2 mL of DMSO in the presence of approximately 200 mg of NaOH powder [91]. An aliquot (2 mg) of the permethylated product was subjected to reductive hydrolysis and acetylation to generate partially methylated alditol acetates (PMAAs) [100]. In another experiment, F60 was permethylated, desulfated, and converted into PMAAs in the same manner, except that the desulfation treatment lasted for 16 h instead of 6 h. In a separate experiment, F60 without desulfation treatment was changed into its triethylammonium salt form by dialysis again 0.1 M triethylamine hydrochloride, permethylated, and then converted into C-1 deuterium-labeled PMAAs by 2 M TFA hydrolysis, $NaBD_4$ reduction, and acetylation, and also into PMAAs without deuterium labeling by reductive hydrolysis and acetylation [100]. PMAA derivatives were tested on an Agilent 7890A-5977B GC-MS system (Agilent Technologies, CA, USA) equipped with a Supelco SP-2380 column (100 m × 0.25 mm × 0.2 μm; Sigma-Aldrich, MA, USA), with oven temperature programmed to start at 100 °C (hold 1 min), followed by increases of 15 °C min^{-1} to 200 °C, and then 1 °C/min to 250 °C (hold 20 min). Inlet temperature was 250°C, and the constant column helium flow rate was 1.2 mL min^{-1}. The PMAAs were identified based on their EI-MS ion fragmentation patterns [439]. Relative molar linkage compositions of the PMAAs were

estimated from the total ion current (TIC) chromatogram, based on the principle that the quantity of a PMAA is proportional to the ratio of its TIC peak area to its molecular mass [58]. For each sample, two separate experiments were conducted.

Appendix 2 Extended data

The following tables are submitted as external tables due to their size:

Appendix Tables 2.1-4, 2.6, 2.7-11, 2.13-14, 2.17

Appendix 3 Supplemental Material to Chapter 4

Appendix 3 Figures

Figure A3.1: *Muus Muus* sampling guide used throughout the study. Sexing is indicated in the bottom right of each photograph (M – male; F - female).

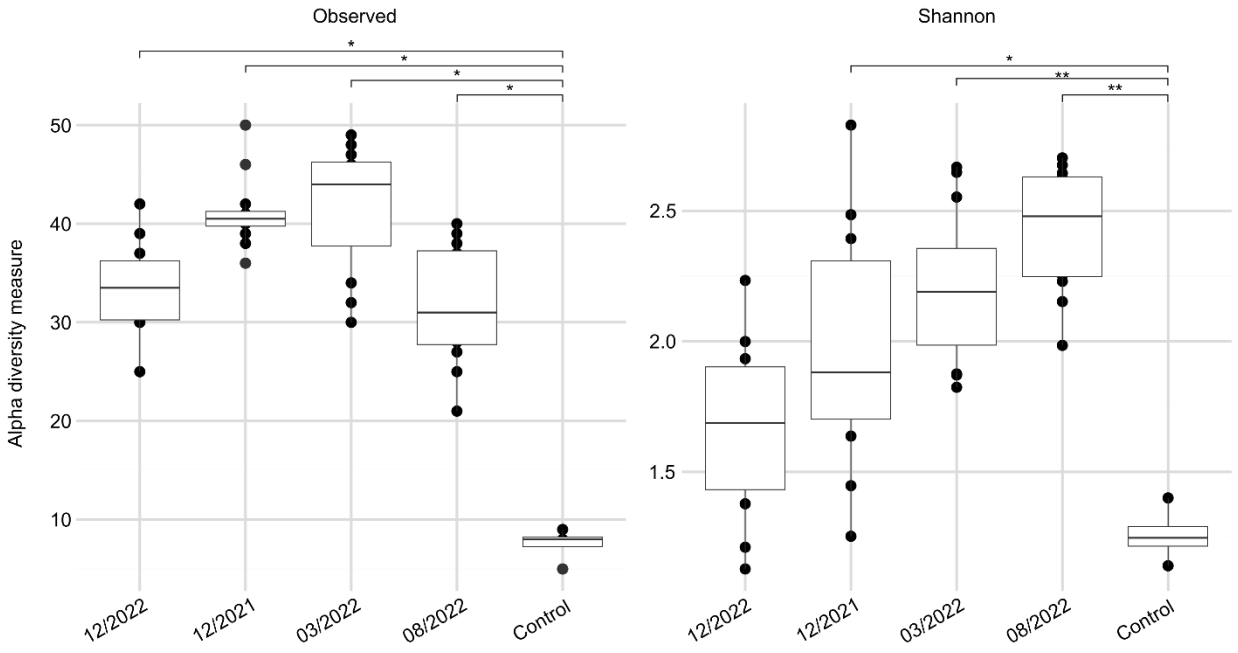

Figure A3.3: *trnL-P6* α-diversity between *Muus Muus* sampling periods. Observed and Shannon diversity α-diversity of *trnL-P6* amplicon marker between *Muus Muus* sampling dates and the control cattle. Wilcox significance ($p < 0.05$) is marked above the bar graphs.

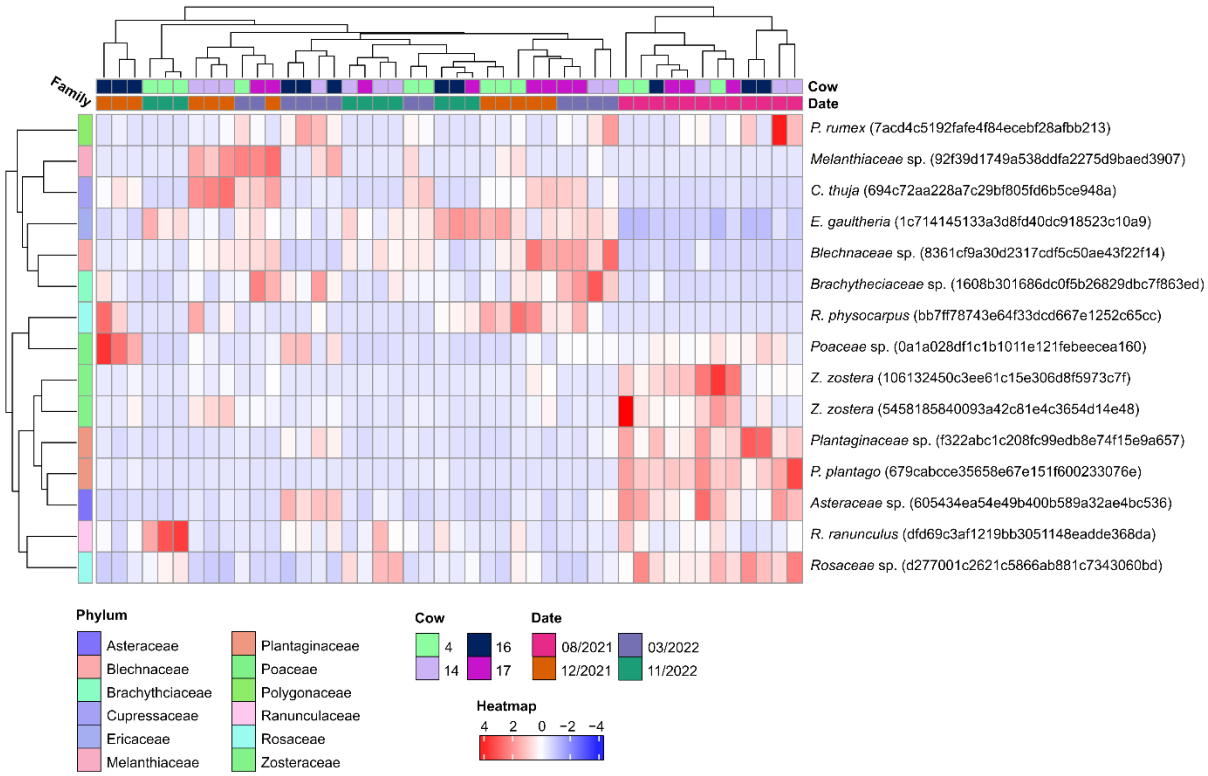

Figure A3.4: DESeq2 analysis of *trnL-P6* amplicons between *Muus Muus* sampling dates. Heatmaps were generated using Pheatmap (euclidean clustering). Amplicons are labelled to the right of the heatmap with the highest predicted taxonomy and the corresponding ASV identifier.

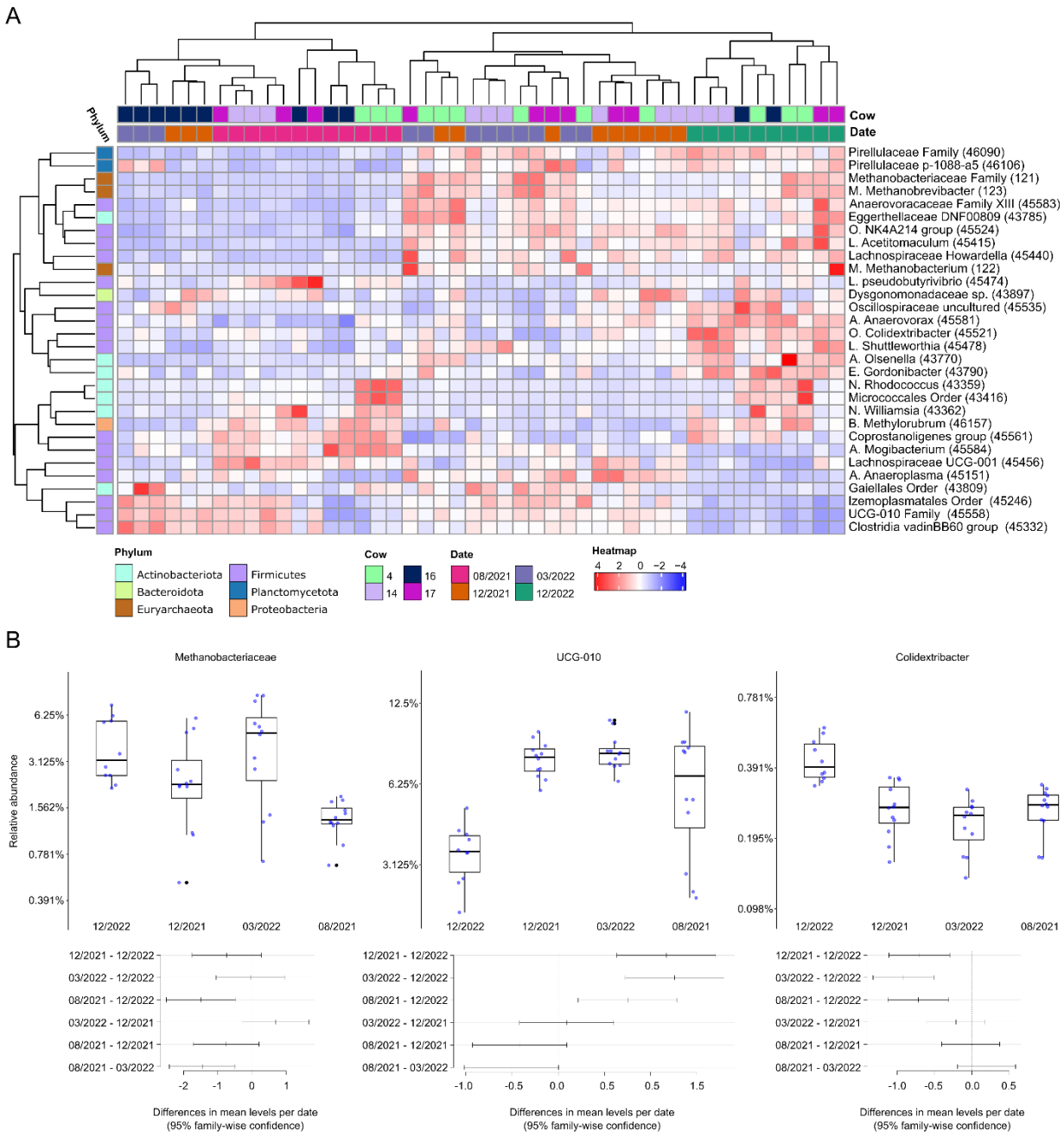

Figure A3.5: Differentially abundance of 16S rRNA OTUs between sampling dates. **A)** DESeq2 analysis of 16S rRNA amplicon between Muus Muus sampling dates. Heatmaps were generated using Pheatmap (euclidean clustering). Amplicons are labelled to the right of the heatmap with the highest predicted taxonomy and the corresponding OTU identifier. **B)** Relative abundance of selected significantly different OTUs from the DESeq2 analysis: *Methanobacteriaceae*, *Clostridia* UCG-010 Family, and *Colidextribacter*. Tukey tests between each sampling date was conducted and illustrated below each corresponding plot.

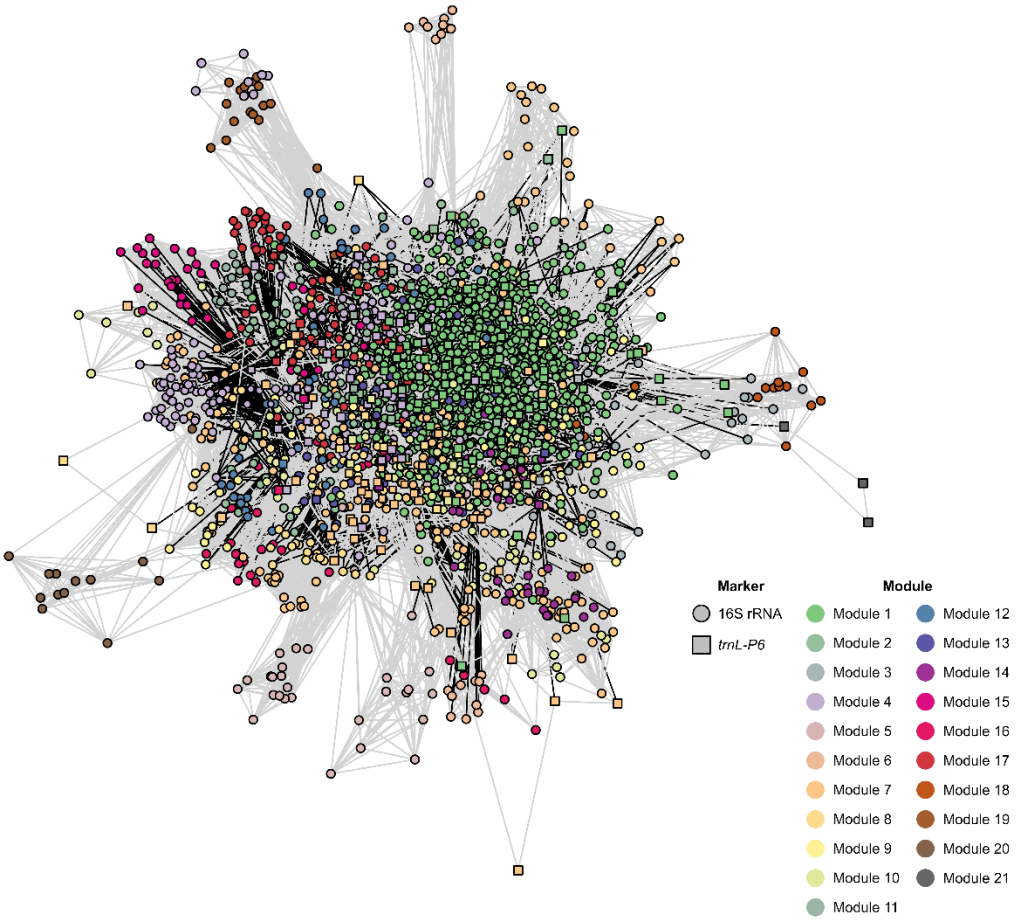

Figure A3.6: Cross-domain connections between *trnl-P6* ASVs and 16S rRNA OTUs. Igraph representation of the Spiec-Easi cross-domain network between *trnL-P6* and 16S rRNA amplicon datasets. Cross network interactions are marked with black lines, whereas intra-domain interactions are denoted in gray lines. ASVs/OTUs are marked by a square (*trnL-P6*) or circle (16S rRNA) and colored via module membership (levian clustering; **Table A3.9**).

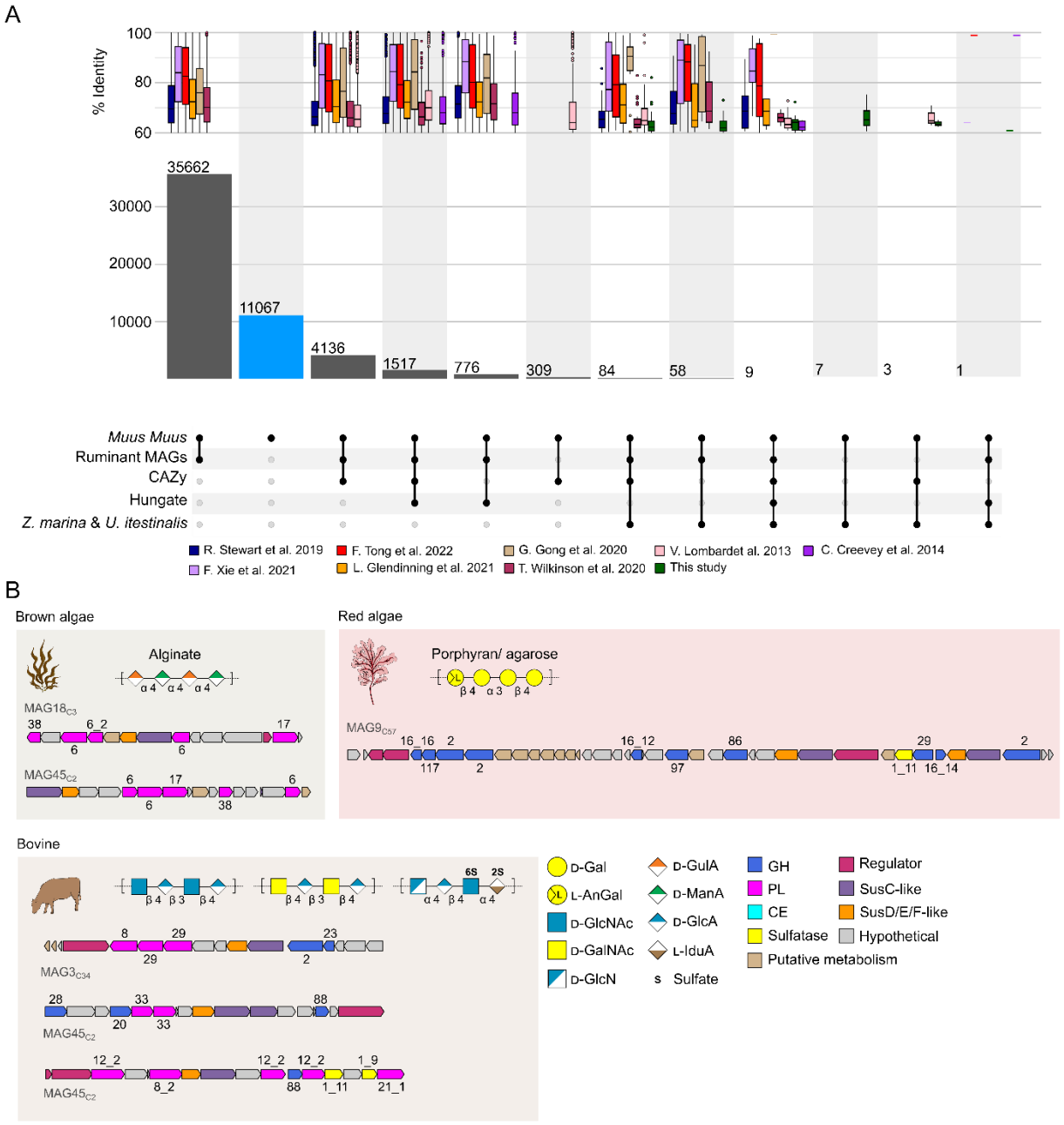

Figure A3.7: Unique *Muus Muus* faecal microbiome CAZymes. A) UpSetR comparison of Blast results done between dbCAN annotated CAZymes. For clarity, ruminant datasets were grouped into a single set, alongside environmental genomes (this study), the CAZy database, and the Hungate collection of genomes. Average identity was marked above the UpSetR plot for each dataset. *Muus Muus* fecal CAZymes that did not share at least 60% identity were treated as completely novel and are marked in blue. **B)** PULs not identified within the proteome targeting unique substrates, including brown seaweed alginate, red seaweed porphyrin/agarose, and host glucans.

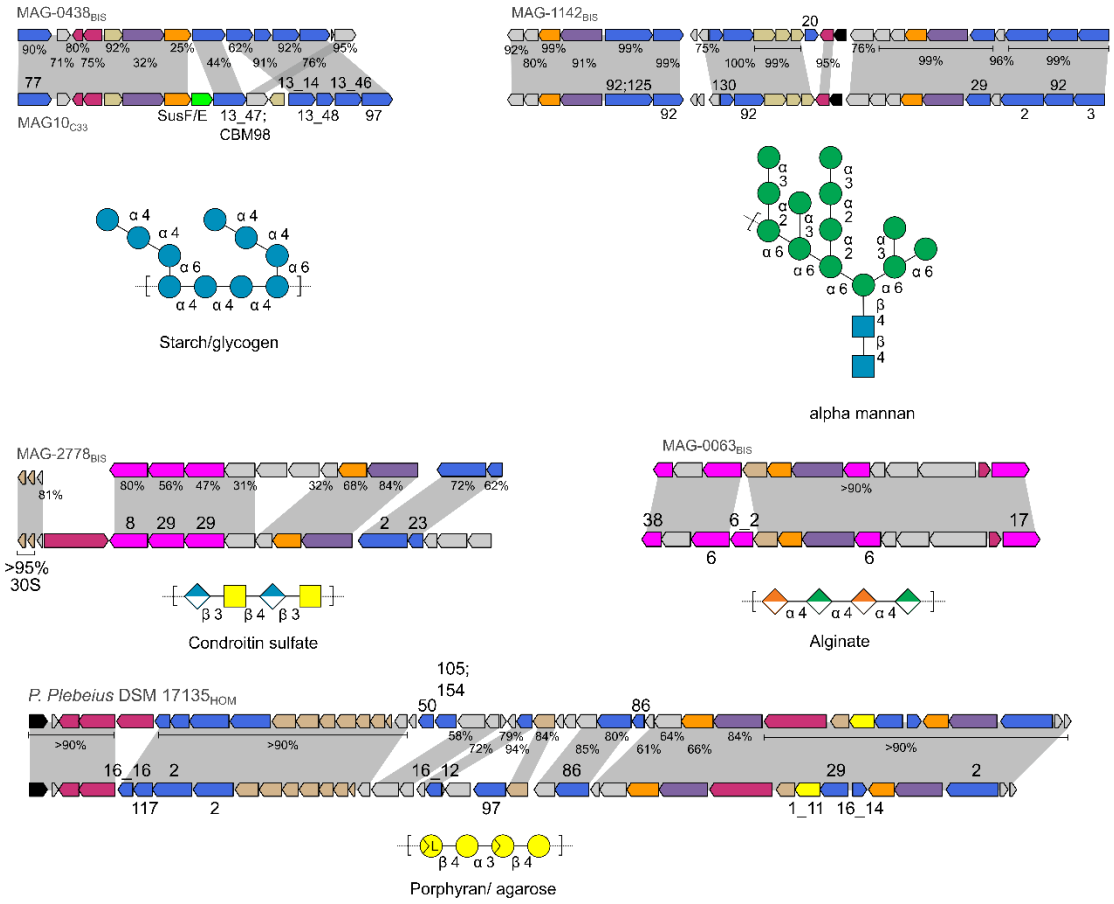

Figure A3.8: PUL similarity of *Muus Muus Bacteroidota* with PULs identified with ruminant and non-ruminant *Bacteroidota*. PULs are illustrated alongside their respective substrates: glycogen/starch, α-mannan, chondroitin sulfate, alginate, and porphyran. Shading marks significant homology with percent identity marked below each gene.

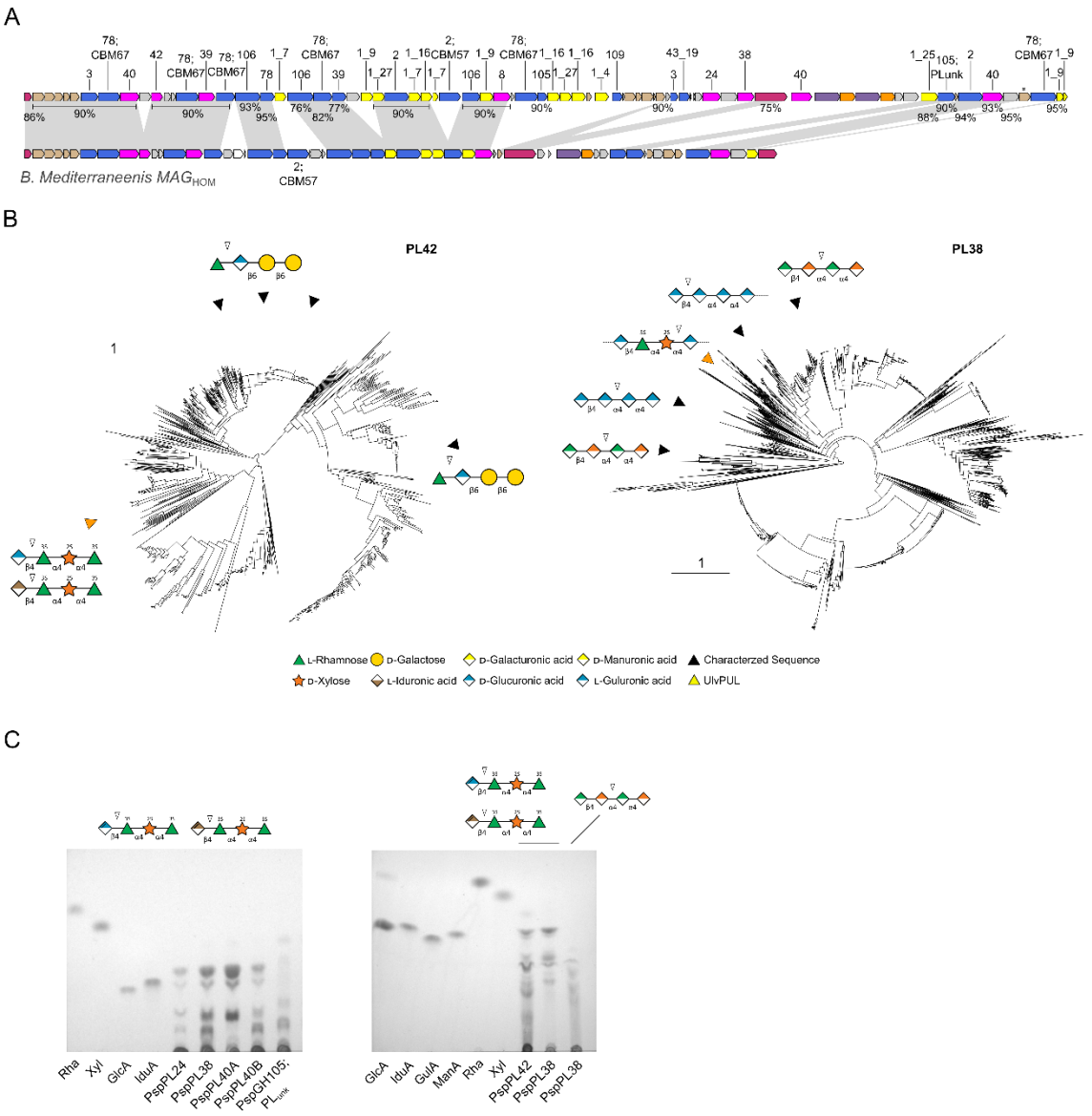

Figure A3.9: UlvPUL CAZymes are distinct from previously characterized family members. A) UlvPUL similarity to an uncharacterized *Bacteroides mediterraneensis* MAG found within humans [388]. Gray bars indicate similarity above 60%. **B)** SACCHARIS [154] trees of UlvPUL PL38 (**top**) and PL42 (**bottom**) family members with all members within the CAZy database. Characterized members are denoted with a black arrow and their confirmed activity, while UlvPUL members are denoted with and orange arrow. **C)** TLC product analysis of PspPL24, PspPL38, PspPL40A, PspPL40B, and PspGH105;PL$_{unk}$ treated with ulvan (**left**) and the polyspecific family members PspPL42 active on ulvan and PspPL38 active on ulvan and alginate (**right**).

Appendix 3 Tables

Table A3.1: Metagenomic assembly and binning statistics from *Muus Muus* fecal samples.

Sample	Id	Number of Reads before trimming	N50 (metaspades)	Bins (70% completetion; <10% contamination	Merged Bins
MuusMuus6_day1	Meares_1	71,234,964	3695	40	
MuusMuus6_day4	Meares_2	82,060,060	5631	60	
MuusMuus13_day5	Meares_3	75,753,375	5595	61	
MuusMuus13_day6	Meares_4	72,745,185	4253	59	
MuusMuus2_day1	Meares_5	73,063,619	2169	20	
MuusMuus2_day4	Meares_6	80,883,544	3553	53	
MuusMuus11_day4	Meares_7	73,770,849	6946	47	
MuusMuus11 day6	Meares_8	78,499,477	4012	61	
MuusMuus7_day4	Meares_9	77,702,887	2012	28	
MuusMuus7_day7	Meares_10	71,291,789	3995	41	808
MuusMuus3_day_4	Meares_32	80,422,458	6548	48	
MuusMuus3_day6	Meares_33	70,520,556	5204	77	
MuusMuus9_day4	Meares_34	84,137,309	4174	42	
MuusMuus9_day6	Meares_35	82,603,669	4160	81	
MuusMuus10_day4	Meares_36	72,482,189	4036	48	
MuusMuus10_day6	Meares_37	67,249,527	3640	31	
MuusMuus8_day1	Meares_38	66,406,770	3160	31	
MuusMuus8_day4	Meares_39	66,105,091	3447	31	
MuusMuus6_day6	Meares_40	75,714,061	4365	46	
Meares environment	Meares_12	307,336,002	5607	120	

Table A3.14: Gradient conditions for chromatographic separation of ulvan digest products.

Time (min)	A (%) 80% MeCN 20% H2O 10 mM ammonium formate 50 mM formic acid	B (%) 20% MeCN 80% H2O 10 mM ammonium formate 50 mM formic acid
0	100	0
1	70	30
20	55	45
22	55	45
22.1	0	100
25	0	100
25.1	100	0
35	100	0

Table A3.15: Parameters for ESI-MSn on the Orbitrap Fusion Tribrid.

Parameter (units)	Value
ESI	
Spray voltage: negative ion (V)	2500
Sheath Gas (Arb)	45
Aux Gas (Arb)	10
Sweep Gas (Arb)	1
Ion Transfer Tube Temp (°C)	325
Vaporizer Temp (°C)	0
MS	
Detector Type	Orbitrap
Orbitrap Resolution	120K
Mass Range	Normal
Scan Range (m/z)	150-2000
RF Lens (%)	50
MS2	
Collision Energy Type	Normalized
Isolation Mode	Quadrupole
Activation Type	HCD
Collision Energy Mode	Stepped
Collision Energies (%)	15,30,45,60,80
Detector Type	Orbitrap
Orbitrap Resolution	30K

Table A3.16: X-ray diffraction and structure refinement.

Data Collection	PspGH105;PLunk
Beamline	Home Beam
Wavelength (Å)	1.54178
Space Group	P212121
Cell Dimensions	
a,b,c (Å)	50.90, 142.19, 198.3
□,□, □ (□)	90.0., 90.0, 90.0
Resolution	48.80 – 2.45 (2.53 – 2.45)
Rmeas	0.130 (1.144)
Rpim	0.040 (0.369)
CC1/2	0.998 (0.686)
<I/σI>	13.8 (2.1)
Completeness (%)	96.8 (98.5)
Redundancy	9.8 (9.1)
No. of reflections	507818 (41077)
No. Unique	51960 (4527)
Refinement	
Resolution (Å)	2.45
Rwork/Rfree	0.23/0.28
No. of atoms	10708
Protein	4995 (Chain A), 5328 (Chain B)
Water	381
B-factors	
Protein	53.28 (Chain A), 47.95 (Chain B)
Water	45.54
r.m.s.d.	
Bond lengths (Å)	0.002
Bond angles (□)	0.472
Ramachandran (%)	
Preferred	96.84
Allowed	6.80
Disallowed	0.0

Appendix 3 Extended data

The following tables are submitted as external tables due to their size:

Appendix Tables 3.2-13