Enrichable consortia of microbial symbionts degrade macroalgal polysaccharides in *Kyphosus* fish

Aaron Oliver[a#], Sheila Podell[a], Linda Wegley Kelly[b], Wesley J. Sparagon[c], Alvaro M. Plominsky[b], Robert S. Nelson[d], Lieve M. L. Laurens[d], Simona Augyte[e], Neil A. Sims[e], Craig E. Nelson[c], Eric E. Allen[a,f,g#]

[a] Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

[b] Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

[c] Daniel K. Inouye Center for Microbial Oceanography, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu, HI, USA

[d] National Renewable Energy Laboratory, Golden, CO, USA

[e] Ocean-Era Inc., Kailua-Kona, HI, USA

[f] Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA

[g] Department of Molecular Biology, School of Biological Sciences, University of California San Diego, La Jolla, CA, USA

To whom correspondence should be addressed (a3oliver@ucsd.edu, eallen@ucsd.edu)

Running title: *Kyphosid gut and inoculated bioreactor genomes*

Keywords: Kyphosus, fish gut microbiome, macroalgal polysaccharides, sulfatase

Abstract

Coastal herbivorous fishes consume macroalgae, which is then degraded by microbes along their digestive tract. However, there is scarce foundational genomic work on the microbiota that perform this degradation. This study explores the potential of *Kyphosus* gastrointestinal microbial symbionts to collaboratively degrade and ferment polysaccharides from red, green, and brown macroalgae through *in silico* study of carbohydrate-active enzyme and sulfatase sequences. Recovery of metagenome-assembled genomes (MAGs) reveals differences in enzymatic capabilities between the major microbial taxa in *Kyphosus* guts. The most versatile of the recovered MAGs were from the Bacteroidota phylum, whose MAGs house enzymes able to decompose a variety of algal polysaccharides. Unique enzymes and predicted degradative capacities of genomes from the *Bacillota* (genus *Vallitalea*) and *Verrucomicrobiota* (order Kiritimatiellales) suggest the potential for microbial transfer between marine sediment and *Kyphosus* digestive tracts. Few genomes contain the required enzymes to fully degrade any complex sulfated algal polysaccharide alone. The distribution of suitable enzymes between MAGs originating from different taxa, along with the widespread detection of signal peptides in candidate enzymes, is consistent with cooperative extracellular degradation of these carbohydrates. This study leverages genomic evidence to reveal an untapped diversity at the enzyme and strain level among *Kyphosus* symbionts and their contributions to macroalgae decomposition. Bioreactor enrichments provide a genomic foundation for degradative and fermentative processes central to translating the knowledge gained from this system to the aquaculture and bioenergy sectors.

Importance

Seaweed has long been considered a promising source of sustainable biomass for bioenergy and aquaculture feed, but scalable industrial methods for decomposing terrestrial compounds can struggle to break down seaweed polysaccharides efficiently due to their unique sulfated structures. Fish of the genus *Kyphosus* feed on seaweed by leveraging gastrointestinal bacteria to degrade algal polysaccharides into simple sugars. This study is the first to build genomes for these gastrointestinal bacteria to enhance our understanding of herbivorous fish digestion and fermentation of algal sugars. Investigations at the gene level identify *Kyphosus* guts as an untapped source of seaweed-degrading enzymes ripe for further characterization. These discoveries set the stage for future work incorporating marine enzymes and microbial communities in the industrial degradation of algal polysaccharides.

Introduction

The *Kyphosus* genus of herbivorous fish, commonly referred to as nenue or rudderfish, graze primarily on macroalgae (1). *Kyphosus* fish serve important ecological roles by controlling algal cover in Indo-Pacific (2) and Caribbean coral reefs (3), thereby mediating coral-algal competition and overall coral growth and benthic community composition (4). Their diverse diet includes macroalgae from the three major taxonomic groups: Rhodophyta (red), Chlorophyta (green) and Ochrophyta (brown) (1). Polysaccharides constitute as much as 60% of macroalgal cells by weight (5) and serve roles in both cell structure and energy storage (6). The complex network of linkages in structural polysaccharides resist degradation from chemical and enzymatic stressors and serves as a physical defense mechanism for algal cells (7).

Algal polysaccharides differ from common polysaccharides found in land plants due to the addition of sulfate ester groups (8). Structural polysaccharides from red algae include agar, carrageenan, porphyran, and xylan, which all contain such sulfate groups (9). Brown algae contains the sulfated polysaccharide fucoidan for structure as well as unsulfated alginate as a storage polysaccharide (9). Green algae contain sulfated polysaccharides such as xylan and ulvan but also contain large amounts of unsulfated cellulose common in land plants (9). Algal polysaccharides are depolymerized primarily through the enzymatic activity of bacterial glycoside hydrolases (GHs) and polysaccharide lyases (PLs) (10), two classes of carbohydrate-active enzymes (CAZymes) (11). Sulfated polysaccharides are particularly recalcitrant to digestion because an additional enzyme class, the sulfatases, is necessary for complete degradation. Full enzyme pathways for the breakdown of various algal polysaccharides have been proposed (9,12) that include both required CAZyme and sulfatase activities. However, not all algal polysaccharides have well-defined degradation pathways or unique associated CAZymes that enable a high-level connection between gene presence and catabolized substrates. Likewise, sulfatase classes within the SulfAtlas database (13) are primarily classified based on evolutionary history rather than substrate specificity or enzymatic activity, so our ability to evaluate pathway completeness *in silico* is limited.

Once complex carbohydrates are broken into subunits by CAZymes and sulfatases, they are utilized by gut microbiota in fermentation reactions to produce short-chain fatty acids (SCFAs) (14). The SCFAs acetate, propanoate, and butyrate have been previously measured in high quantities in *Kyphosus* hindguts (15) and are utilized by the host fish for energy (16). Previous work has suggested correlations between SCFA profiles and bacterial composition (15), but there is no genomic work in algivorous fish pinpointing which microbiota contribute to host nutrition in this way and what pathways are utilized to produce these essential SCFAs.

Our overall understanding of the role of gut microbiota in digestion is still limited in most fishes (17), including *Kyphosus*, in part due to a focus on gut composition and diversity rather than function. The genetic study of *Kyphosus* gut symbionts has been limited to 16S rRNA (15,18) and metabolomic (18) investigations until the incorporation of shotgun metagenomics in

a few recent studies (19,20). What functional profiling has been done in fish guts often relies on extrapolation from amplicon-based taxonomic distributions (21–24), and no study has yet generated a large collection of metagenome-assembled genomes (MAGs) from an algivorous fish gut. A *de novo* genomic investigation of *Kyphosus* symbionts has the potential to reveal degradative capacities that cannot be extrapolated from taxonomic lineage or relatedness to database representatives.

Discoveries from better studied human gut and terrestrial herbivore systems provide suggestions for how *Kyphosus* symbionts might gain and use such gene pathways. Human gut bacteria have acquired enzymes which degrade sulfated algal polysaccharides through horizontal gene transfer (25,26). Horizontal gene transfer of antibiotic resistance genes has also been observed in fish gut biofilms (27), but this phenomenon has not yet been reported for carbohydrate-active enzymes in any fish gut symbiont microbe. Once acquired, CAZymes and sulfatases potentially originating from one or multiple organisms may then decompose algal polysaccharides in complex, stepwise pathways. A cooperative division of labor strategy, in which partial breakdown products from one bacterial population serve as a degradative substrate for other bacteria in the community, has been proposed to occur in human gut microbiota (28) and has been suggested as a way to improve polysaccharide degradation in engineered communities (29). The degree to which collaboration may occur in the herbivorous fish gastrointestinal tract remains unknown.

Exploring functional diversity not only improves our understanding of herbivorous fish digestion but may also enable concrete applications in the fields of aquaculture and bioenergy. Most aquaculture is currently sustained through compound feeds that are composed of fishmeal and fish oils from wild-caught fish (30). Although innovations in aquaculture feed have lowered the trophic levels of captive carnivorous fish and improved overall feed efficiency (31), concerns about sustainability and food security remain. Wan *et al.* (2019) argue that the discovery of efficient methods to degrade complex polysaccharides and enhance nutrient digestibility is a key knowledge gap and barrier limiting macroalgae inclusion into commercial aquafeeds (32). Macroalgal feed additives are also known to counteract methanogenesis in terrestrial ruminants (33) and thus can be applied to reduce methane emissions from livestock husbandry. However, deficiencies in ruminant microbiome digestive capacities may influence the future development and long-term success of seaweed dietary supplementation strategies. Research on *Kyphosus* symbionts and their enzymes can inspire commercializable and scalable methods to break down these barriers in industry.

Innovations exploiting the experimental propagation of enrichment cultures with *Kyphosus* symbionts can harness these microbial communities for further study and experimentation with commercial outputs in the bioenergy sector as well as the development of macroalgal feed supplements. While a few bacterial isolates have been recovered and sequenced from kyphosid guts (34), no previous study has enriched entire communities from these fishes to investigate their hydrolytic and fermentative capabilities. Hydrolysis of carbohydrates, proteins,

and lipids into their monomeric components is a key step in biogas and bioethanol production from macroalgae (35,36), and the degradation of algal polysaccharides is often the rate limiting step in anaerobic digestion (37). Milledge *et al.* (2019) call for future studies to look beyond commercially available enzymes to discover candidates that can more efficiently degrade algal polysaccharides (38). The *Kyphosus* gut, with its understudied functional diversity and degradative pathways, offers an untapped source of such enzyme and inoculum candidates.

This study leverages metagenome-assembled genomes from *Kyphosus vaigiensis, Kyphosus cinerascens, and Kyphosus hawaiiensis* gut symbionts and inoculated bioreactor enrichments to connect whole genome degradative potential of algal polysaccharides to accurate taxonomic lineages and functional roles. The addition of genomes from bioreactor enrichments explores leveraging the metabolic capacities of *Kyphosus* gut consortia in industrial processes. This work extends previous studies of taxonomic-level biogeography (18) and contig-level gene associations (15,20) in this system using high-quality MAGs, which enables differentiation between processes that can potentially be executed within a single cellular compartment (individual microbial species/population) and those likely to require cooperative action by multiple cells from different species (community impacts). Discoveries in this study provide foundation for genome-level understanding of microbial contributions to herbivorous fish digestion and beget future investigations to apply these findings towards applications in the aquaculture and bioenergy sectors.

Materials and Methods

Sample description and metagenomic assembly

DNA was extracted from liquid samples from ten anaerobic bioreactors inoculated with gut contents from two *Kyphosus* fishes (**Table S1**) using methods previously described (18) and propagated to enrich degradative properties. Samples were taken 9-10 days after inoculation and incubation at 30°C. Anoxic cultures of 50ml were processed in a portable anaerobic chamber containing one-third strength sterile artificial seawater (Instant Ocean, Spectrum Brands, Blacksburg, VA) in 150ml serum bottles, crimp sealed with a rubber septum. Approximately 1g of fish gut section contents were placed in the bottles along with the indicated substrate (**Table S1**) and sealed, with no additional feedstock added before sequencing.

Samples were sequenced using Illumina NovaSeq 6000 technology (Illumina, San Diego, CA). Read trimming was performed using Trimmomatic v. 0.36 (39) with the following parameters: adapter-read alignment settings 2:30:10, LEADING:10, TRAILING:20, HEADCROP:12, SLIDINGWINDOW:4:15, MINLEN:200. Taxonomic composition of metagenomic reads was determined using Kraken v. 2.0.9 (40), with taxonomic assignment using a protein database based on all amino acid sequences in the NCBI nr database (41) as of April 2022. Cleaned reads were assembled in metaSPAdes v. 3.13 (42) with a minimum contig retention size of 2000 nucleotides.

Gene calling and functional annotation

Gene boundaries were predicted using prodigal v. 2.6.2 (43) and annotated using prokka v. 1.12 (44). Genes were assigned to CAZy classes from the dbCAN HMMdb v. 10 database (45) based on the CAZy database (11) and to sulfatases classes from the SulfAtlas v 2.3 database (13), using methods previously described (20). Signal peptides were identified using SignalP v. 6 (46) with default parameters.

Enzyme novelty was evaluated using DIAMOND blastp (47) searches against the NCBI nr database (41) as of April 2022. Some CAZyme classes were grouped into the category of "peptidoglycanases" using the division proposed by López-Mondéjar *et al.* (2022) (48). Distributions of annotated proteins were compared to free-living relatives from the OceanDNA database (49).

Metagenomic binning and biosynthetic gene cluster prediction

Metagenomic binning was performed using MetaWRAP v. 1.3.2 (50) with a minimum completeness cutoff of 0.7 and a maximum contamination cutoff of 0.05 as determined by CheckM v. 1.0.12 (51). MAG taxonomy was determined using GTDB-Tk v. 1.5.1 (52) with release 202 of the Genome Taxonomy Database (53).

Viral contigs and prophage were identified using DeepVirFinder v. 1.0 (54) using a q-score cutoff of 0.94. Viral sequence completeness was determined using Checkv v. 1.5 (55), we only retained regions marked as "high-quality" or "complete". Viral sequences were assigned host taxonomies using VPF-class (56).

Biosynthetic gene clusters (BGCs) were predicted for each MAG using antiSMASH v. 6.1 (57). Predicted products and BGC classes were annotated using BiG-SLiCE v. 1.1.1 (58). Gene cluster distances were calculated using the BiG-FAM webservice v. 1.0.0 (59), using a novelty distance cutoff of 900 following previous studies (59–61). Short chain fatty acid gene clusters were annotated using gutSMASH v. 5.0.0 (62).

Phylogenomics and enzyme phylogenetics

A phylogenetic tree of MAGs was generated using PhyloPhlAn v. 3.0.2 (63) using a concatenated universal set of 400 marker genes (64). MAGs containing at least 100 marker genes underwent concatenated alignment using mafft v. 7.505 (65). The phylogenetic tree was built using RaxML v. 8.2.12 (66) and visualized using R v. 4.2.0 (67) packages treeio v. 1.20.0 (68), ggtree v. 3.4.0 (69), and ggtreeExtra v. 1.6.0 (70).

Multiple sequence alignments for genes belonging to CAZy class GH86 were made using MUSCLE v 3.8.31 (71) and visualized using the R package ggmsa v. 1.2.0 (72). Gene trees were created using FastTree v. 2.1.10 (73). Additional reference genes were included in the tree based on DIAMOND blastp hits to the NCBI nr database as of April 2022. Protein domains were

analyzed with the CDD webservice (74). 3D protein structures for CAZymes were predicted using ColabFold v. 1.3.0 (75) and visualized using ChimeraX v. 1.3 (76). Residue conservation was visualized using the WebLogo (77) webservice.

Data availability

All custom code used for data analysis and visualization are available at https://github.com/AaronAOliver/KyphosusMAGs. Sequence reads are available under SRA bioproject numbers PRJNA819194 and PRJNA1023379. Complete MAG sequences and predicted proteins are available on Zenodo (https://zenodo.org) under DOI no. 10.5281/zenodo.8277654.

Results

A (meta)genome catalog of enrichable symbionts in the *Kyphosus* gut

New data derived from *K. cinerascens* and *K. hawaiiensis* enrichment cultures expands the diversity of previous *K. cinerascens*, *K. hawaiiensis*, and *K. vaigiensis* gut metagenomes (20). This more complete catalog of *Kyphosus* gut microbiota provides additional details on the metabolic potential of taxa that were rare in the *in vivo* gut metagenome samples and highlights potential challenges in harnessing gastrointestinal microbiota for industrial processes. The fish inoculum species, gut location, and feedstock that were combined to establish each enrichment sample are described in **Table S1**. The taxonomic classification of unassembled metagenomic reads revealed a surprising consistency between the *in vivo* gut microbiomes (20) and enrichment samples (**Figure 1**). *Bacillota*, *Bacteroidota*, and *Gammaproteobacteria* constitute the dominant bacterial lineages in most samples, although the *Desulfovibrionales* order (phylum *Thermodesulfobacteriota*) was highly abundant in two enrichment samples.

Figure 1. Taxonomic distribution of enrichment and fish gut samples. The taxonomic distribution of unassembled classified reads as determined using Kraken2. Any taxonomic lineages that are not associated with a binned MAG are grouped into “unbinned taxa.”

211 medium and high-quality MAGs were binned from the *in vivo* fish gut metagenomes and newly assembled enrichment metagenomes. These MAGs all met the minimum 70% completion, maximum 5% redundancy standards (78). The number of recovered MAGs per metagenome is shown in **Figure S1**. The assembly statistics for enrichment metagenomes are shown in **Table S2** and MIMAG-compliant (78) summary information are shown in **Table S3.** Consistent with the unassembled read-based taxonomic profiles of the metagenomes, most MAGs were assigned to the phyla *Bacillota* (78 MAGs), *Bacteroidota* (72), the class *Gammaproteobacteria* (31), the class *Desulfovibrionales* (13), or *Verrucomicrobiota* (6). The enrichments provide access to data on microbial members that were not as abundant in the fish gut metagenomes and vice versa. In one example, the *Verrucomicrobiota* class *Kiritimatiellales* was binned in fish gut samples but not in enrichment metagenomes. This novelty was reflected in

nucleotide similarities, as only 9 of the 74 (12%) enrichment MAGs match MAGs generated from *in vivo* fish gut metagenomes at the species level.

Viral sequences comprised less than 0.5% of all unassembled metagenomic reads, with 69 viral contigs and 3 prophages identified as either high quality or complete. With these viral elements, 30 auxiliary metabolic genes found on potential prophage regions were annotated as CAZymes, and 13 as sulfatases, suggesting a potential role for viral dissemination of these genes across the bacterial community. The taxa *Bacillota*, *Bacteroidota*, and *Gammaproteobacteria* were the most frequently predicted viral hosts, which is consistent with the taxonomic abundances of classified unassembled metagenomic reads and recovered MAGs (**Table S4**). Despite the presence of numerous auxiliary metabolic genes generally related to polysaccharide degradation, none of the viral sequences we detected appeared to specifically target large, complex sulfated macroalgal polysaccharides.

Genome capacities reveal metabolic specialization among gut symbionts of *Kyphosus* fish

The distribution of CAZymes and sulfatases was correlated with the phylogeny of fish gut and enrichment MAGs (as determined through a concatenated marker gene tree, **Figure 2a**). This assessment revealed that among the MAGs generated in this study, the *Bacteroidota* genomes contained the majority of CAZymes and sulfatases (**Figure 2b**). Algal degradation-specific CAZyme-rich genomes among the MAGs from other phyla were restricted either to a single order, *Kiritimatiellales* (*Verrucomicrobiota*), or a single genus, *Vallitalea* (*Bacillota*). Recovered *Gammaproteobacteria* and *Desulfovibrionales* genomes lacked enzymes required for digesting sulfated algal polysaccharides, despite the relatively high abundance of these taxonomic groups in classified unassembled reads and the recovered MAGs. However, the *Gammaproteobacteria* MAGs contained more peptidoglycanases than other taxa, suggesting a niche in digesting alternative dietary components. This analysis also showed that CAZymes targeting ulvan, a green algal polysaccharide, were less prevalent among the obtained symbiotic MAGs than CAZymes targeting red and brown algae-associated polysaccharides (**Figure 2b**), consistent with previous results quantifying relative amounts of these algae types consumed by the *Kyphosus* fish included in this study (20). The most abundant phyla all had binned MAGs from both *in vivo* and enrichment samples (**Figure 2c**).

Figure 2. Genomic CAZyme distributions reveal connections between metabolic strategies and taxonomic lineage. (A) The gene tree shows a concatenated alignment of 400 PhyloPhlAn universal marker genes for each recovered MAG, with branches colored by assigned MAG taxonomy. (B) The inner ring displays genomic gene counts for sulfatases and carbohydrate-active enzymes that specifically target algal polysaccharides or peptidoglycan. (C) Environmental source of each MAG.

An assessment of SCFA production gene pathways of recovered MAGs using gutSMASH (62) revealed that most of the *Kyphosus* gut symbiotic taxa (67% of fish gut MAGs,

77% of enrichment MAGs) can potentially contribute to host nutrition through the production of SCFAs (**Figure 3**). 139 genomes from analyzed kyphosid fish gut microbial communities contained pathways for producing acetate but only six genomes contained pathways for butyrate production. The pyruvate formate lyase and pyruvate:ferredoxin oxidoreductase pathways were the most abundant overall, present in 126 MAGs, while *Bacteroidota* contained the most gene clusters (39) related to propanoate production.

The overall prevalence of acetate pathways was lower than that found previously in human gut microbiota. The total absence of some alternate fermentation pathways from our MAGs, such as choline utilization, suggests that those processes are not core to dominant members of the *Kyphosus* gut microbiome. Only one genome from this study contained fermentation pathways involving the degradation of amino acids such as glycine, threonine, and lysine, suggesting that *Kyphosus* gut microbiota do not rely directly on dietary proteins for energy. Such lessened reliance on nitrogen-based substrates for fermentation is consistent with a low protein, algae-based diet rich in available polysaccharides and limited in available nitrogen.

Functional adaptations to life in the *Kyphosus* gut

Adaptations to environmental conditions in herbivorous fish guts studied here are reflected in the high abundance of CAZyme classes specifically targeting algal polysaccharides. **Figure 4a** shows that the amino acid sequences of selected CAZyme classes abundant in our assembled metagenomes are well conserved across *Kyphosus* gut symbiont genomes. However, such enzymes are poorly represented in both specialty and general databases of previously described sequences, with closest enzyme homologs averaging below 60% sequence similarity for most of the examined CAZyme classes. Similar trends are observed for the sulfatase subclasses in these *Kyphosus* gut symbiont genomes (**Figure 4b**). Both cases denote the extent that this study expands known sequence diversity within these enzyme classes, potentially suggests new subclasses, and highlights unusual domains that may not be captured by current databases.

Figure 4. *Kyphosus* gut symbionts encode CAZymes and sulfatases divergent from other datasets and environments. Percent identity of binned (A) CAZymes and (B) sulfatases to best blast matches found in the following databases: all genes from MAGs in this study (orange), the GenBank nr database (green), and either (A) the CAZy database or (B) the SulfAtlas database (blue). Each group is labeled by the number of genes with that enzyme annotation found in our MAGs.

The addition of novel enzymes sequences to each of these enzyme classes presents numerous opportunities to expand our understanding of marine polysaccharide degradation. One example using the phylogeny of CAZy class GH86, consisting of β-agarases and β-porphyranases, illustrates previously unappreciated cryptic variability within this enzyme family. A gene tree of class GH86 CAZyme examples from this study (**Figure 5**), that includes the closest GenBank homologs, shows that most of the genes are associated with *Bacteroidota* from

Kyphosus guts. This is consistent with the high abundance of CAZymes and sulfatases found among MAGS from the phylum (**Figure 2)**. Surprisingly, two GH86 genes recovered in *Bacillota* MAGs from bioreactor enrichments and two homologs from the NCBI nr database cluster together with two genes found among hindgut MAGS from the phylum *Verrucomicrobiota*, suggesting potential horizontal gene transfer from marine sediment communities into *Kyphosus* gut microbiota. Binned genes annotated as β-porphyranases all originate from hindgut or enrichment samples, consistent with previously reported physiological localization of polysaccharide degradation capabilities (20).

Figure 5. A β-agarase/β-porphyranase gene tree highlights an undescribed protein domain present in multiple phyla. (A) A gene tree of binned GH86 enzymes, with gene names colored by genome taxonomy. Nodes with black diamonds represent collapsed clades without the undescribed domain. A multiple sequence alignment is appended to the end of the tree, with colored vertical lines representing amino acid positions and white vertical lines representing gaps. (B) The predicted protein structure of GH86 enzyme R2_26_16226, with conserved CAZy domains highlighted in gray, the predicted signal peptide in green, and the conserved undescribed domain in pink. An uncollapsed version of the gene tree is included as **Figure S2** and a motif logo of the domain is included as **Figure S3**.

The multiple sequence alignment in **Figure 5a** highlights a unique pattern within the Bacillota genus *Vallitalea* and neighboring *Verrucomicrobiota* CAZyme sequences that has not been described in prior literature. This pattern might either extend the signal peptide or add an additional uncharacterized domain between the signal peptide and the porphyranase catalytic subdomain (79). Among NCBI nr homologs, only genes from an isolated *Vallitalea* genome (WP_212695143.1, WP_212695474.1) (80) contained this pattern. No other proteins in the GenBank nr database contained sequences matching this region at greater than 50% amino acid identity for this pattern of approximately 168 amino acids, with few conserved residues among our sequenced examples (**Figure S3**). Outside of the clade containing this novel domain, variability occurs primarily in the putative signal peptide region at the N-terminus of the protein, while the porphyranase domain itself is far more conserved. **Figure 5b** displays the predicted 3-dimensional structure of a *Kyphosus* symbiont GH86 enzyme, with the additional uncharacterized structure positioned between the predicted signal peptide and annotated catalytic β-agarase and β-porphyranase domains. This uncharacterized domain might influence an array of enzymatic properties such as a novel substrate specificity, concentration dependence, improved efficiency, or tolerance of different abiotic conditions. Although the function of this domain cannot be determined bioinformatically, this example is an interesting candidate for further enzymatic characterization and shows the promise of uncovering novel enzyme activity within the metabolic repertoire of the *Kyphosus* gut.

MAG sequences were interrogated using antiSMASH biosynthetic gene cluster detection software to determine whether *Kyphosus* gut-associated microbial taxa might encode any unusual secondary metabolites. The majority of *Bacillota*, *Bacteroidota*, *Verrucomicrobiota*, and

Gammaproteobacteria MAGs from both fish gut inocula and bioreactor enrichments encoded BGCs typical of taxonomic relatives found in other vertebrate gut environments, such as lanthipeptides, betalactone, and arylpolyene (81,82). However, BGCs were not particularly abundant in our MAG catalog relative to other similar genomes. Our recovered *Gammaproteobacteria*, *Bacillota*, and *Bacteroidota* average fewer BGCs per genome than a random set of seawater MAGs of each taxonomic group from the OceanDNA database. Thus, our host-associated MAGs contain fewer BGCs per genome than their free-living relatives.

A total of 307 BGCs were annotated within our MAGs (**Figure 6**). 23% of annotated BGCs were determined to be complete based on BiG-FAM. 20 BGCs represent putative novel gene cluster families as determined by BiG-FAM (**Figure 6b**). These novel gene cluster families may represent unique natural products or enzymes specialized to the *Kyphosus* gut environment. Complete biosynthetic gene cluster annotations, novelty assessment, and associated taxonomy are included in **Table S5**.

Figure 6. *Kyphosus* gut symbiont MAGs encode novel biosynthetic gene clusters. (A) On the positive y-axis, counts of binned BGCs grouped by BiG-SLiCE class and labeled by predicted product. On the negative y-axis, counts of binned BGCs grouped by BiG-SLiCE class and colored by associated MAG taxonomy. (B) Distance of binned BGCs to the nearest gene cluster family as determined by BiG-FAM. A distance above 900, marked by a dashed red line, suggests novelty and divergence from previously described gene cluster families. BGCs are colored orange if they are annotated as complete by BiG-FAM. Abbreviations used: RiPP, ribosomally synthesized and post-translationally modified peptides; RRE, RiPP recognition element; NRPS, non-ribosomal peptide synthetase; PKS, polyketide synthase.

Community digestion of complex algal polysaccharides

Polysaccharide digestive capabilities vary among MAGs from different microbial taxa in the *Kyphosus* fish gut community, as shown in **Figure 7**. Despite overall microbiome-wide diversity, the MAGs generated in this study show that few individual genomes contain all of the enzymes necessary to completely degrade even a single type of complex algal polysaccharide, let alone the huge variety of natural variants characteristic of marine macroalgae (83) that might be ingested by generalist herbivorous fishes. Each microbial genome instead contains a limited assortment of enzymes capable of partially degrading a selection of different carbohydrate moieties, including potentially incomplete breakdown products generated by other microbes. Thus, combined pangenomic capabilities of several taxonomic groups appear to contain complementary collections of exported CAZymes that might facilitate adaptation to unpredictable variability in available polysaccharide content.

Figure 7. *Kyphosus* gut symbiont MAGs encode the capacity to degrade various algal polysaccharides collaboratively, but not solitarily. Each row represents a single MAG from the annotated taxonomic lineage. Only MAGs from the four lineages with the highest concentration of CAZymes (Bacillota, Bacteroidota, Gammaproteobacteria, and

Verrucomicrobiota) are shown. MAGs with no applicable CAZy classes are not shown, and CAZy classes not associated with a single substrate or not found in any MAG are not shown. Green bars denote a signal peptide annotated to at least one of the appropriate CAZyme in a single MAG, while yellow bars mark the absence of a signal peptide on all appropriate CAZyme candidates within a MAG.

Levels of contribution to community-wide degradation of algal polysaccharides through extracellular enzymes are dependent on both cell taxonomy and targeted substrate. More than 90% of CAZymes that target macroalgal polysaccharides from *Bacteroidota* MAGs contain signal peptides that indicate export or integration into the cellular membrane. CAZymes in *Bacillota* MAGs largely lack these signal peptides in enzymes predicted to degrade fucoidan and agar, but the signal peptides are more abundant in the smaller set of CAZymes targeting xylan and alginates. Few *Bacillota* MAGs contain all the enzymes required to fully degrade complex algal polysaccharides such as porphyran, suggesting that cells from this taxonomic group might scavenge partial breakdown products degraded extracellularly by other taxa.

Verrucomicrobiota polysaccharide digestion enzymes appear to be more specialized towards red algae, with genomes consistently containing CAZymes predicted to digest agar, carrageenan, and porphyran. However, MAGs from this phylum seem to be lacking enzymes predicted to target green or brown algal polysaccharides. *Gammaproteobacteria* MAGs appear to have more enzymes involved in the digestion of non-sulfated polysaccharides such as alginate, and occasionally enzymes involved in agar degradation. Thus, the *Gammaproteobacteria* symbionts analyzed here have likely specialized in polysaccharide types that are easier to digest.

Discussion

The recovery and characterization of 211 MAGs from *Kyphosus* gut and enrichment metagenomes connect detailed taxonomic classification with the potential of the major microbial contributors to digest complex algal polysaccharides. Algal polysaccharide-targeting enzymes from this study are divergent in sequence from previously sequenced and characterized representatives from other environments, clarifying prior assumptions about the metabolic capacities of this system using 16S rRNA or community composition. This work confirms and expands earlier work showing that certain members of the *Bacillota* and *Verrucomicrobiota* lineages are unexpectedly richer in some CAZyme and sulfatase enzyme classes than their respective taxonomic relatives (20). Differences between source inocula and the metagenomes of bioreactor enrichments inoculated with *Kyphosus* gut bacteria highlight potential challenges in harnessing these microbiota for bioenergy preprocessing of macroalgal feedstocks.

This study is the first to describe specific genes encoding SCFA production pathways in the genomes of fish gut microbiota. Microbial fatty acids serve as a key metabolite in gut-brain communication (84) and are a major source of available carbon for the host (85). SCFA pathway diversity is unexpectedly low for a system previously shown to contain high SCFA concentrations *in vivo* (16). However, this observation is consistent with a few dominant

lineages, primarily the *Bacteroidota*, producing high amounts of SCFAs from the breakdown products of algal polysaccharides. Prior chemical work has observed that propanoate is more abundant than butyrate in *Kyphosus* guts (16), and our pathway enzyme abundance information at the genome level supports these observations (**Figure 3**). Likewise, observations in that same work noted rates of sulfate reduction were higher than methanogenesis, although both processes were negligible compared to SCFA production. This aligns with the low abundance of *Desulfovibrionales* and the near complete absence of Archaea in our metagenomes, which is consistent with repeated observations that dietary red macroalgae inhibit methanogenesis and thus the success of gut Archaea (33). Both sulfate reduction and methanogenesis appear to be minor sources of energy available for Kyphosid host absorption, compared to fermentation by *Bacteroidota* and *Bacillota*.

Herbivorous fish frequently contain visible amounts of sediment in their guts (86), which is thought to increase physical abrasion of gut contents and aid with degradation. Previous works have shown functional redundancy between the metabolic capacities fish gut and sediment communities (87), including carbon cycling. Gene flow has also been observed from fish feces to sediment microbiomes (88). Although the relative abundance of sediment-associated microbes in kyphosid fish microbiomes is low (20), one explanation for similar CAZymes and sulfatases between fish gut and sediment microbes could involve a circular loop of gene flow from fish guts to sediment through fecal pellets, and from sediments into fish guts through digestion. This hypothesis is supported by the fact that Kyphosid gut community *Bacillota* from the genus *Vallitalea* appear more closely related to marine sediment bacteria (80,89) than any previously reported examples from seawater or terrestrial gut microbiota (**Table S3).** Likewise, sediment-dwelling *Verrucomicrobiota* from the order *Kiritimatiellales* similar to those in *Kyphosus* fish guts have also been shown to degrade sulfated macroalgal polysaccharides (90), with genomes rich in both glycoside hydrolases and sulfatases (91). It is possible that consumption of sediment by *Kyphosus* fish improves polysaccharide digestion not only through physical breakdown of seaweed, but also by the contribution of additional enzyme capabilities originally derived from sediment bacteria that likely encounter highly diverse recalcitrant organic substrates including macroalgae biomass (92).

Although *Kiritimatiellales* MAGs recovered from fish guts contain more enzymes targeting algal polysaccharides than other members of their phyla, these taxa were not enriched in or recovered from enrichment metagenomes. This should not be problematic for enrichment processing if the dominant *Bacteroidota* contain CAZymes with overlapping specificities for the same substrates, as suggested in **Figure 7**. However future work will be needed to characterize detailed, sample-specific polysaccharide degradative chemistry using such a framework. *Vallitalea* and *Verrucomicrobiota* enzymes may also have some unique functionalities, as suggested by the extra domain present in their β-porphyranase sequences. Isolation and *in vitro* characterization of bioinformatically predicted enzyme activities will be necessary to integrate these discoveries into aquaculture and bioenergy applications.

Metagenomic data from the MAGs in this study suggest that few individual cells have the genomic potential to independently degrade all of the complex sulfated polysaccharide substrates present in marine macroalgae. However, secreted and extracellularly exposed transmembrane CAZymes may enable collaborative interactions between fish gut microbes to facilitate complete digestion of these molecules, without the high metabolic cost of encoding a complete, independent repertoire in every genome. A division of labor strategy cannot be fully confirmed without *in vitro* tests (93), although the first condition of functional complementarity appears to hold true between *Kyphosus* symbionts based on our bioinformatic investigations. In one similar study, gene based observations of complementarity for marine lignocellulose-degrading bacteria align with *in vitro* observations that support a division of labor hypothesis (94). Future work involving cultured representatives and enriched microcosms will be required to pin down the ecological strategies used by symbionts in this system.

This study provides a new baseline for *Kyphosus* microbiota at the genome level but begets a slew of new questions that require additional experimentation. Further work that connects enrichment composition, feedstock polysaccharide composition, and physical configuration to chemical measurements of degraded polysaccharides would help determine which phyla are required for complete polysaccharide breakdown. Isolation and characterization of divergent proteins with completely novel domains will determine what new enzymatic properties are unique to this system. Metatranscriptomic analyses utilizing the genome catalogs presented here will enable detailed analysis of substrate-specific metabolic pathway expression and species collaboration. *Kyphosus* digestive systems have long been studied as models for herbivorous fish gut fermentation and can now be explored further using these additional techniques to deliver a deeper understanding of their degradative and fermentative capabilities.

Conclusion

Among the first metagenome-assembled genomes recovered from herbivorous fish guts and corresponding bioreactors, a new genomic catalog of *Kyphosus* gut symbionts highlights untapped diversity in enzymatic and collaborative potential in the degradation of algal polysaccharides. The enzymes encoded within these symbiont genomes are divergent from the extent of sequenced CAZymes, supporting the promise of herbivorous fish guts as a source of novel and industrially relevant enzymes. Expansion of these discoveries will not only clarify ecological interactions but have the potential to improve the applicability of macroalgae in the bioenergy and aquaculture sectors.

Acknowledgements

This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (#S10 OD026929). Computational analyses were performed using the San Diego Supercomputer Center's Triton Shared Computing Cluster. This work was funded by the

United States Government, Department of Energy, Advanced Research Projects Agency—Energy grant ARPA-E DE-FOA-0001858 to R.S.N. and L.M.L.L. with contracts to C.E.N., L.W.K., and E.E.A., National Science Foundation grants OCE-1837116 and EF-2025217 to E.E.A., and National Institutes of Health NIEHS grant R01-ES030316 to E.E.A.

References

1. Wu P, Wang T, Liu Y, Li C, Xiao Y, Xu S, et al. Differences of Macroalgal Consumption by Eight Herbivorous Coral Reef Fishes From the Xisha Islands, China. Front Mar Sci [Internet]. 2022 [cited 2022 May 9];9. Available from: https://www.frontiersin.org/article/10.3389/fmars.2022.882196

2. Knudsen SW, Clements KD. World-wide species distributions in the family Kyphosidae (Teleostei: Perciformes). Mol Phylogenet Evol. 2016 Aug 1;101:252–66.

3. Dell CLA, Longo GO, Burkepile DE, Manfrino C. Few Herbivore Species Consume Dominant Macroalgae on a Caribbean Coral Reef. Front Mar Sci [Internet]. 2020 [cited 2022 May 17];7. Available from: https://www.frontiersin.org/article/10.3389/fmars.2020.00676

4. Pillans RD, Babcock RC, Thomson DP, Haywood MDE, Downie RA, Vanderklift MA, et al. Habitat effects on home range and schooling behaviour in a herbivorous fish (Kyphosus bigibbus) revealed by acoustic tracking. Mar Freshw Res. 2017 Feb 1;68(8):1454–67.

5. Augyte S, Sims NA, Martin K, Van Wychen S, Panczak B, Alt H, et al. Tropical Red Macroalgae Cultivation with a Focus on Compositional Analysis. Plants. 2023 Jan;12(20):3524.

6. Patel AK, Vadrale AP, Singhania RR, Michaud P, Pandey A, Chen SJ, et al. Algal polysaccharides: current status and future prospects. Phytochem Rev [Internet]. 2022 Feb 9 [cited 2022 May 17]; Available from: https://doi.org/10.1007/s11101-021-09799-5

7. 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. Annu Rev Plant Biol. 2011;62(1):567–90.

8. Helbert W. Marine Polysaccharide Sulfatases. Front Mar Sci [Internet]. 2017 [cited 2022 May 18];4. Available from: https://www.frontiersin.org/article/10.3389/fmars.2017.00006

9. Bäumgen M, Dutschei T, Bornscheuer UT. Marine Polysaccharides: Occurrence, Enzymatic Degradation and Utilization. ChemBioChem. 2021;22(13):2247–56.

10. Jagtap AS, Manohar CS. Overview on Microbial Enzymatic Production of Algal Oligosaccharides for Nutraceutical Applications. Mar Biotechnol. 2021 Apr 1;23(2):159–76.

11. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022 Jan 7;50(D1):D571–7.

12. Li J, He Z, Liang Y, Peng T, Hu Z. Insights into Algal Polysaccharides: A Review of Their Structure, Depolymerases, and Metabolic Pathways. J Agric Food Chem. 2022 Feb 16;70(6):1749–65.

13. Barbeyron T, Brillet-Guéguen L, Carré W, Carrière C, Caron C, Czjzek M, et al. Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for Sulfatases Reflecting Their Substrate Specificity. PLOS ONE. 2016 Oct 17;11(10):e0164846.

14. Rimmer DW, Wiebe WJ. Fermentative microbial digestion in herbivorous fishes. J Fish Biol. 1987;31(2):229–36.

15. Pardesi B, Roberton AM, Lee KC, Angert ER, Rosendale DI, Boycheva S, et al. Distinct microbiota composition and fermentation products indicate functional compartmentalization in the hindgut of a marine herbivorous fish. Mol Ecol. 2022;31(8):2494–509.

16. Mountfort DO, Campbell J, Clements KD. Hindgut Fermentation in Three Species of Marine Herbivorous Fish. Appl Environ Microbiol. 2002 Mar;68(3):1374–80.

17. Perry WB, Lindsay E, Payne CJ, Brodie C, Kazlauskaite R. The role of the gut microbiome in sustainable teleost aquaculture. Proc R Soc B Biol Sci. 2020 May 13;287(1926):20200184.

18. Sparagon WJ, Gentry EC, Minich JJ, Vollbrecht L, Laurens LML, Allen EE, et al. Fine scale transitions of the microbiota and metabolome along the gastrointestinal tract of herbivorous fishes. Anim Microbiome. 2022 May 23;4(1):33.

19. Stevenson SJR, Lee KC, Handley KM, Angert ER, White WL, Clements KD. Substrate degradation pathways, conserved functions and community composition of the hindgut microbiota in the herbivorous marine fish Kyphosus sydneyanus. Comp Biochem Physiol A Mol Integr Physiol. 2022 Oct 1;272:111283.

20. Podell S, Oliver A, Kelly LW, Sparagon WJ, Plominsky AM, Nelson RS, et al. Herbivorous Fish Microbiome Adaptations to Sulfated Dietary Polysaccharides. Appl Environ Microbiol. 2023 May 3;89(5):e02154-22.

21. Serag AM, Abdel-Sabour MS, El-Hadidi M, Maged M, Magdy M, Ramadan MF, et al. Comparative 16S Metabarcoding of Nile Tilapia Gut Microbiota from the Northern Lakes of Egypt. Appl Biochem Biotechnol. 2022 May 1;194(5):2168–82.

22. Liu H, Guo X, Gooneratne R, Lai R, Zeng C, Zhan F, et al. The gut microbiome and degradation enzyme activity of wild freshwater fishes influenced by their trophic levels. Sci Rep. 2016 Apr 13;6(1):24340.

23. Hu Z, Tong Q, Chang J, Yu J, Li S, Niu H, et al. Gut bacterial communities in the freshwater snail Planorbella trivolvis and their modification by a non-herbivorous diet. PeerJ. 2021 Feb 12;9:e10716.

24. Yang G, Jian SQ, Cao H, Wen C, Hu B, Peng M, et al. Changes in microbiota along the intestine of grass carp (Ctenopharyngodon idella): Community, interspecific interactions, and functions. Aquaculture. 2019 Jan 1;498:151–61.

25. 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 Mar 9;30(3):314-328.e11.

26. Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010 Apr;464(7290):908–12.

27. Abe K, Nomura N, Suzuki S. Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. FEMS Microbiol Ecol. 2020 May 1;96(5):fiaa031.

28. Romero Marcia AD, Yao T, Chen MH, Oles RE, Lindemann SR. Fine Carbohydrate Structure of Dietary Resistant Glucans Governs the Structure and Function of Human Gut Microbiota. Nutrients. 2021 Sep;13(9):2924.

29. Lindemann SR. A piece of the pie: engineering microbiomes by exploiting division of labor in complex polysaccharide consumption. Curr Opin Chem Eng. 2020 Dec 1;30:96–102.

30. Sofia F. The State of World Fisheries and Aquaculture 2018-Meeting the sustainable development goals. Fish Aquac Dep Food Agric Organ U N Rome. 2018;

31. Cottrell RS, Metian M, Froehlich HE, Blanchard JL, Sand Jacobsen N, McIntyre PB, et al. Time to rethink trophic levels in aquaculture policy. Rev Aquac. 2021;13(3):1583–93.

32. Wan AHL, Davies SJ, Soler-Vila A, Fitzgerald R, Johnson MP. Macroalgae as a sustainable aquafeed ingredient. Rev Aquac. 2019;11(3):458–92.

33. Sofyan A, Irawan A, Herdian H, Jasmadi, Harahap MA, Sakti AA, et al. Effects of various macroalgae species on methane production, rumen fermentation, and ruminant production: A meta-analysis from in vitro and in vivo experiments. Anim Feed Sci Technol. 2022 Dec 1;294:115503.

34. Pardesi B, Roberton AM, Wollmuth EM, Angert ER, Rosendale DI, White WL, et al. Tannockella kyphosi gen. nov., sp. nov., a member of the family Erysipelotrichaceae, isolated from the hindgut of the marine herbivorous fish Kyphosus sydneyanus. Int J Syst Evol Microbiol. 72(5):005374.

35. Maneein S, Milledge JJ, Nielsen BV, Harvey PJ. A Review of Seaweed Pre-Treatment Methods for Enhanced Biofuel Production by Anaerobic Digestion or Fermentation. Fermentation. 2018 Dec;4(4):100.

36. González-Gloria KD, Rodríguez-Jasso RM, Shiva, Aparicio E, Chávez González ML, Kostas ET, et al. Macroalgal biomass in terms of third-generation biorefinery concept:

Current status and techno-economic analysis – A review. Bioresour Technol Rep. 2021 Dec 1;16:100863.

37. Sutherland AD, Varela JC. Comparison of various microbial inocula for the efficient anaerobic digestion of Laminaria hyperborea. BMC Biotechnol. 2014 Jan 23;14(1):7.

38. Milledge JJ, Nielsen BV, Maneein S, Harvey PJ. A Brief Review of Anaerobic Digestion of Algae for Bioenergy. Energies. 2019 Jan;12(6):1166.

39. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

40. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019 Nov 28;20(1):257.

41. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2021 Dec 1;gkab1112.

42. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017 May;27(5):824–34.

43. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010 Mar 8;11(1):119.

44. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl. 2014 Jul 15;30(14):2068–9.

45. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018 Jul 2;46(W1):W95–101.

46. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022 Jan 3;1–3.

47. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021 Apr;18(4):366–8.

48. López-Mondéjar R, Tláskal V, da Rocha UN, Baldrian P. Global Distribution of Carbohydrate Utilization Potential in the Prokaryotic Tree of Life. mSystems. 2022 Nov 22;7(6):e00829-22.

49. Nishimura Y, Yoshizawa S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. Sci Data. 2022 Jun 17;9(1):305.

50. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018 Sep 15;6(1):158.

51. 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 Jul;25(7):1043–55.

52. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020 Mar 15;36(6):1925–7.

53. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022 Jan 7;50(D1):D785–94.

54. Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, et al. Identifying viruses from metagenomic data using deep learning. Quant Biol. 2020 Mar 1;8(1):64–77.

55. 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 May;39(5):578–85.

56. Pons JC, Paez-Espino D, Riera G, Ivanova N, Kyrpides NC, Llabrés M. VPF-Class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. Bioinformatics. 2021 Jul 1;37(13):1805–13.

57. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021 May 12;49(W1):W29–35.

58. Kautsar SA, van der Hooft JJJ, de Ridder D, Medema MH. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. GigaScience. 2021 Jan 29;10(1):giaa154.

59. Kautsar SA, Blin K, Shaw S, Weber T, Medema MH. BiG-FAM: the biosynthetic gene cluster families database. Nucleic Acids Res. 2021 Jan 8;49(D1):D490–7.

60. Sánchez-Navarro R, Nuhamunada M, Mohite OS, Wasmund K, Albertsen M, Gram L, et al. Long-Read Metagenome-Assembled Genomes Improve Identification of Novel Complete Biosynthetic Gene Clusters in a Complex Microbial Activated Sludge Ecosystem. mSystems. 2022 Nov 29;7(6):e00632-22.

61. Huang R, Wang Y, Liu D, Wang S, Lv H, Yan Z. Long-Read Metagenomics of Marine Microbes Reveals Diversely Expressed Secondary Metabolites. Microbiol Spectr. 2023 Jul 6;11(4):e01501-23.

686 62. Pascal Andreu V, Roel-Touris J, Dodd D, Fischbach MA, Medema MH. The gutSMASH
687 web server: automated identification of primary metabolic gene clusters from the gut
688 microbiota. Nucleic Acids Res. 2021 Jul 2;49(W1):W263–70.

689 63. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise
690 phylogenetic analysis of microbial isolates and genomes from metagenomes using
691 PhyloPhlAn 3.0. Nat Commun. 2020 May 19;11(1):2500.

692 64. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for
693 improved phylogenetic and taxonomic placement of microbes. Nat Commun. 2013 Aug
694 14;4(1):2304.

695 65. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
696 Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr 1;30(4):772–80.

697 66. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
698 phylogenies. Bioinformatics. 2014 May 1;30(9):1312–3.

699 67. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna,
700 Austria; 2022. Available from: https://www.R-project.org/

701 68. Wang LG, Lam TTY, Xu S, Dai Z, Zhou L, Feng T, et al. Treeio: An R Package for
702 Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol Biol
703 Evol. 2020 Feb 1;37(2):599–603.

704 69. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinforma.
705 2020;69(1):e96.

706 70. Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, et al. ggtreeExtra: Compact visualization of richly
707 annotated phylogenetic data. Mol Biol Evol. 2021 Jun;38(9):4039–42.

708 71. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
709 Nucleic Acids Res. 2004;32(5):1792–7.

710 72. Zhou L, Yu G. ggmsa: Plot multiple sequence alignment using “ggplot2” [Internet]. 2022.
711 Available from: http://yulab-smu.top/ggmsa/

712 73. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for
713 Large Alignments. PLOS ONE. 2010 Mar 10;5(3):e9490.

714 74. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD:
715 NCBI’s conserved domain database. Nucleic Acids Res. 2015 Jan 28;43(D1):D222–6.

716 75. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold -
717 Making protein folding accessible to all [Internet]. bioRxiv; 2022 [cited 2022 May 11]. p.
718 2021.08.15.456425. Available from:
719 https://www.biorxiv.org/content/10.1101/2021.08.15.456425v3

76. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci Publ Protein Soc. 2021 Jan;30(1):70–82.

77. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A Sequence Logo Generator. Genome Res. 2004 Jun 1;14(6):1188–90.

78. 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 Aug;35(8):725–31.

79. 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. Proc Natl Acad Sci U S A. 2012 Nov 27;109(48):19786–91.

80. Ben Aissa F, Postec A, Erauso G, Payri C, Pelletier B, Hamdi M, et al. Vallitalea pronyensis sp. nov., isolated from a marine alkaline hydrothermal chimney. Int J Syst Evol Microbiol. 2014 Apr;64(Pt 4):1160–5.

81. Donia MS, Fischbach MA. Small molecules from the human microbiota. Science. 2015 Jul 24;349(6246):1254766.

82. Zhang XX, Lv QB, Yan QL, Zhang Y, Guo RC, Meng JX, et al. A Catalog of over 5,000 Metagenome-Assembled Microbial Genomes from the Caprinae Gut Microbiota. Microbiol Spectr. 2022 Nov 2;10(6):e02211-22.

83. Arnosti C, Wietz M, Brinkhoff T, Hehemann JH, Probandt D, Zeugner L, et al. The Biogeochemistry of Marine Polysaccharides: Sources, Inventories, and Bacterial Drivers of the Carbohydrate Cycle. Annu Rev Mar Sci. 2021;13(1):81–108.

84. Majumdar A, Siva Venkatesh IP, Basu A. Short-Chain Fatty Acids in the Microbiota–Gut–Brain Axis: Role in Neurodegenerative Disorders and Viral Infections. ACS Chem Neurosci. 2023 Mar 15;14(6):1045–62.

85. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes. 2016 May 3;7(3):189–200.

86. Kopp D, Bouchon-Navaro Y, Cordonnier S, Haouisée A, Louis M, Bouchon C. Evaluation of algal regulation by herbivorous fishes on Caribbean coral reefs. Helgol Mar Res. 2010 Sep 1;64(3):181–90.

87. Escalas A, Troussellier M, Yuan T, Bouvier T, Bouvier C, Mouchet MA, et al. Functional diversity and redundancy across fish gut, sediment and water bacterial communities. Environ Microbiol. 2017;19(8):3268–82.

88. Muziasari WI, Pitkänen LK, Sørum H, Stedtfeld RD, Tiedje JM, Virta M. The Resistome of Farmed Fish Feces Contributes to the Enrichment of Antibiotic Resistance Genes in

Sediments below Baltic Sea Fish Farms. Front Microbiol [Internet]. 2017 [cited 2023 Jun 14];7. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2016.02137

89. Sun YT, Zhou N, Wang BJ, Liu XD, Jiang CY, Ge X, et al. Vallitalea okinawensis sp. nov., isolated from Okinawa Trough sediment and emended description of the genus Vallitalea. Int J Syst Evol Microbiol. 2018;69(2):404–10.

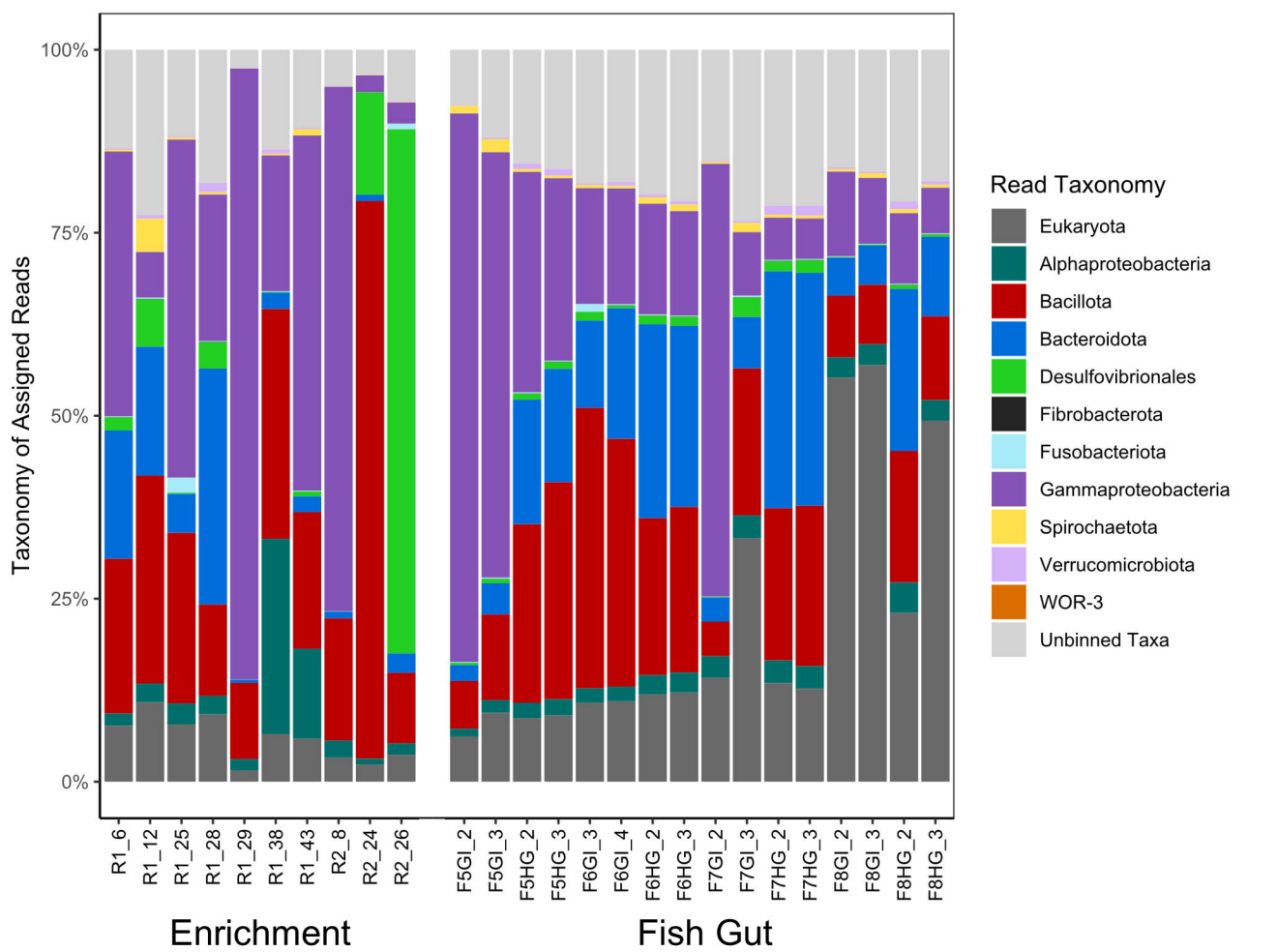
90. van Vliet DM, Palakawong Na Ayudthaya S, Diop S, Villanueva L, Stams AJM, Sánchez-Andrea I. Anaerobic Degradation of Sulfated Polysaccharides by Two Novel Kiritimatiellales Strains Isolated From Black Sea Sediment. Front Microbiol [Internet]. 2019 [cited 2022 May 20];10. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2019.00253

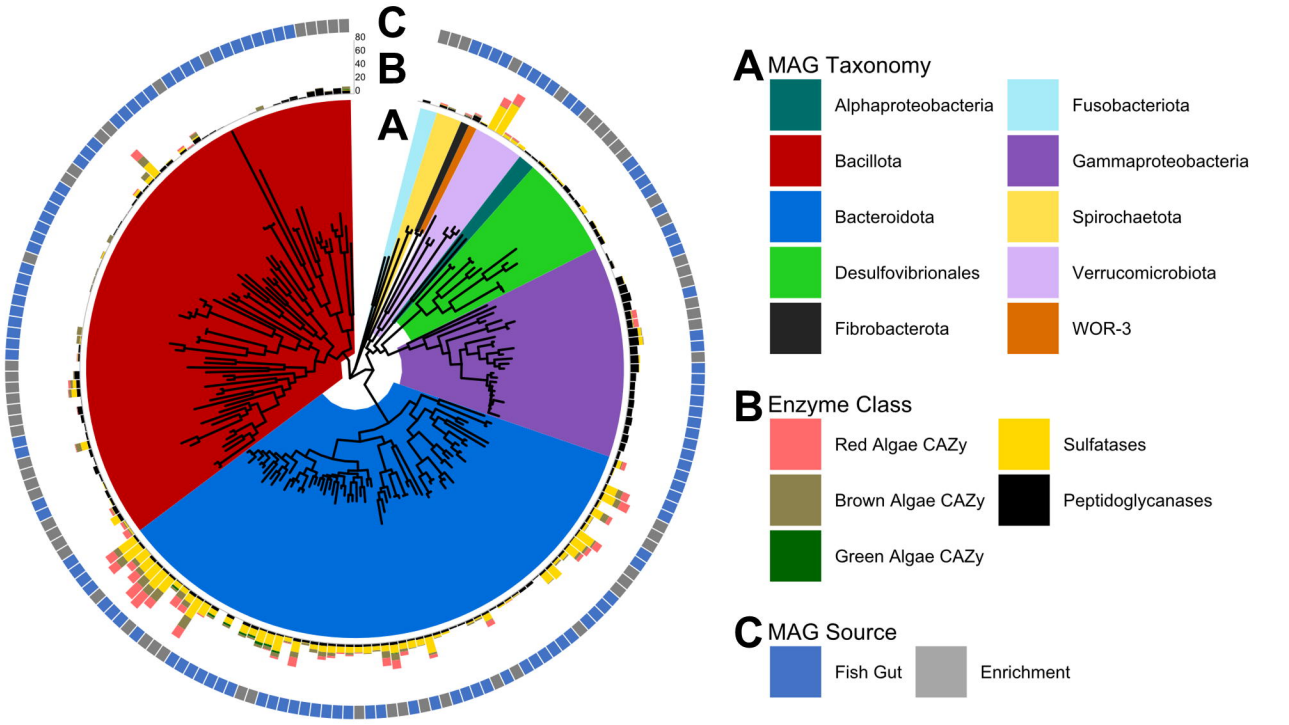
91. van Vliet DM, Lin Y, Bale NJ, Koenen M, Villanueva L, Stams AJM, et al. Pontiella desulfatans gen. nov., sp. nov., and Pontiella sulfatireligans sp. nov., Two Marine Anaerobes of the Pontiellaceae fam. nov. Producing Sulfated Glycosaminoglycan-like Exopolymers. Microorganisms. 2020 Jun;8(6):920.

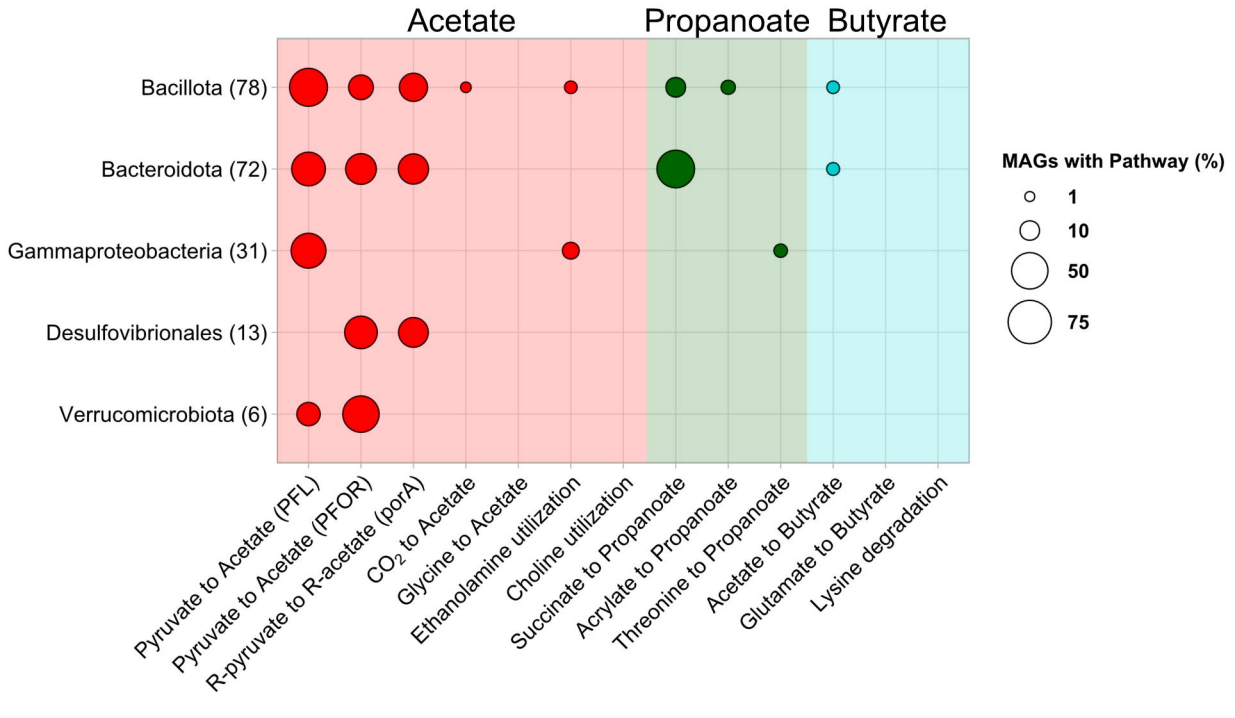
92. Arnosti C. Substrate specificity in polysaccharide hydrolysis: Contrasts between bottom water and sediments. Limnol Oceanogr. 2000;45(5):1112–9.

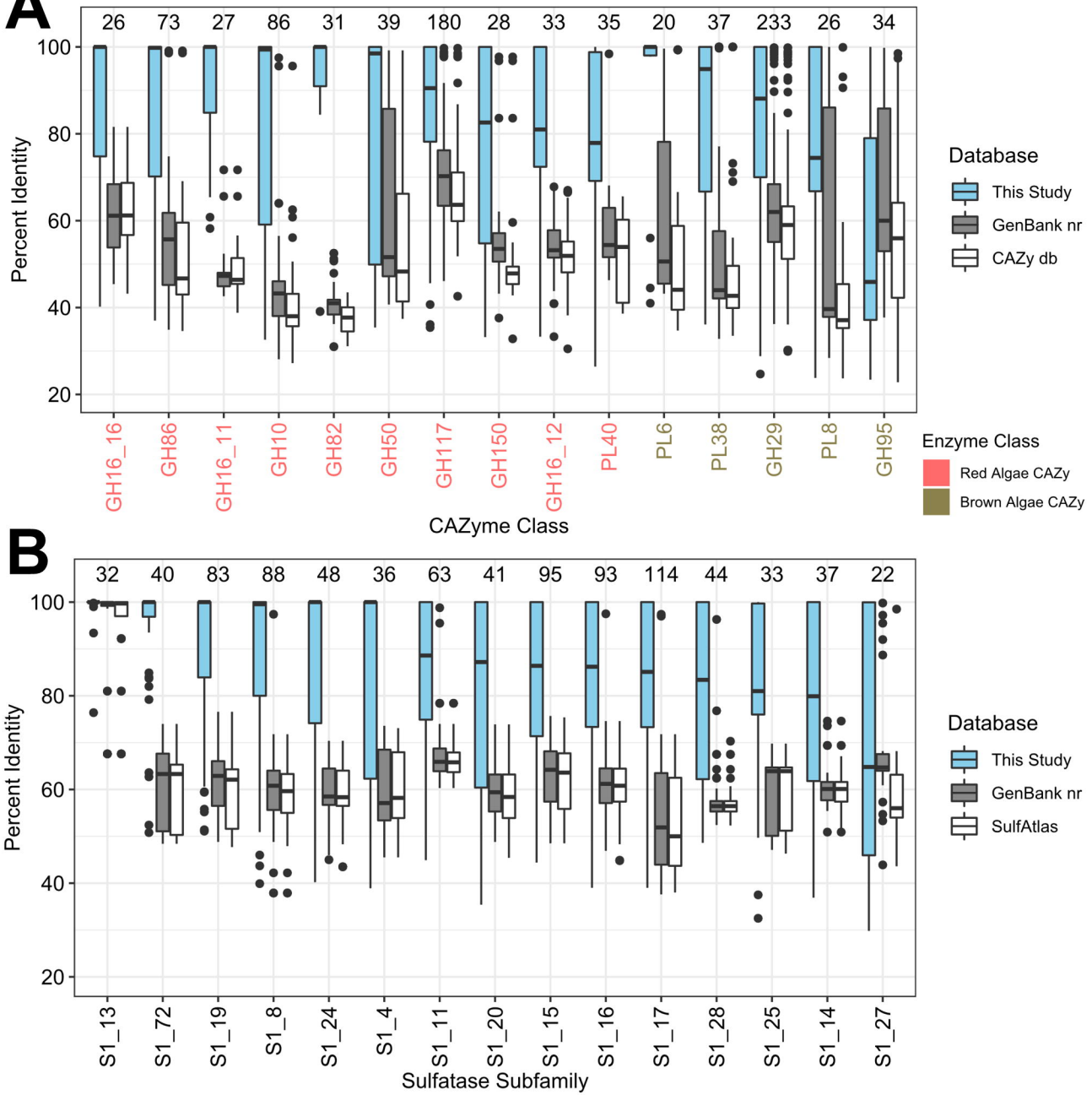
93. Giri S, Waschina S, Kaleta C, Kost C. Defining Division of Labor in Microbial Communities. J Mol Biol. 2019 Nov 22;431(23):4712–31.

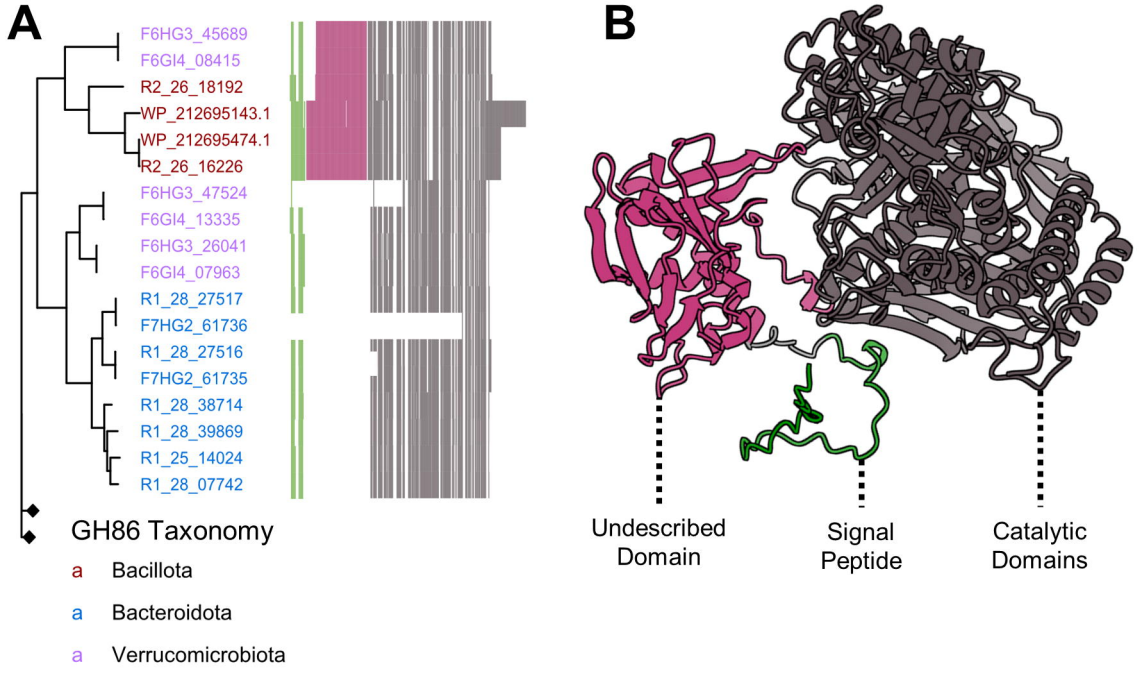
94. Peng Q, Lin L, Tu Q, Wang X, Zhou Y, Chen J, et al. Unraveling the roles of coastal bacterial consortia in degradation of various lignocellulosic substrates. mSystems. 2023 Jul 7;0(0):e01283-22.

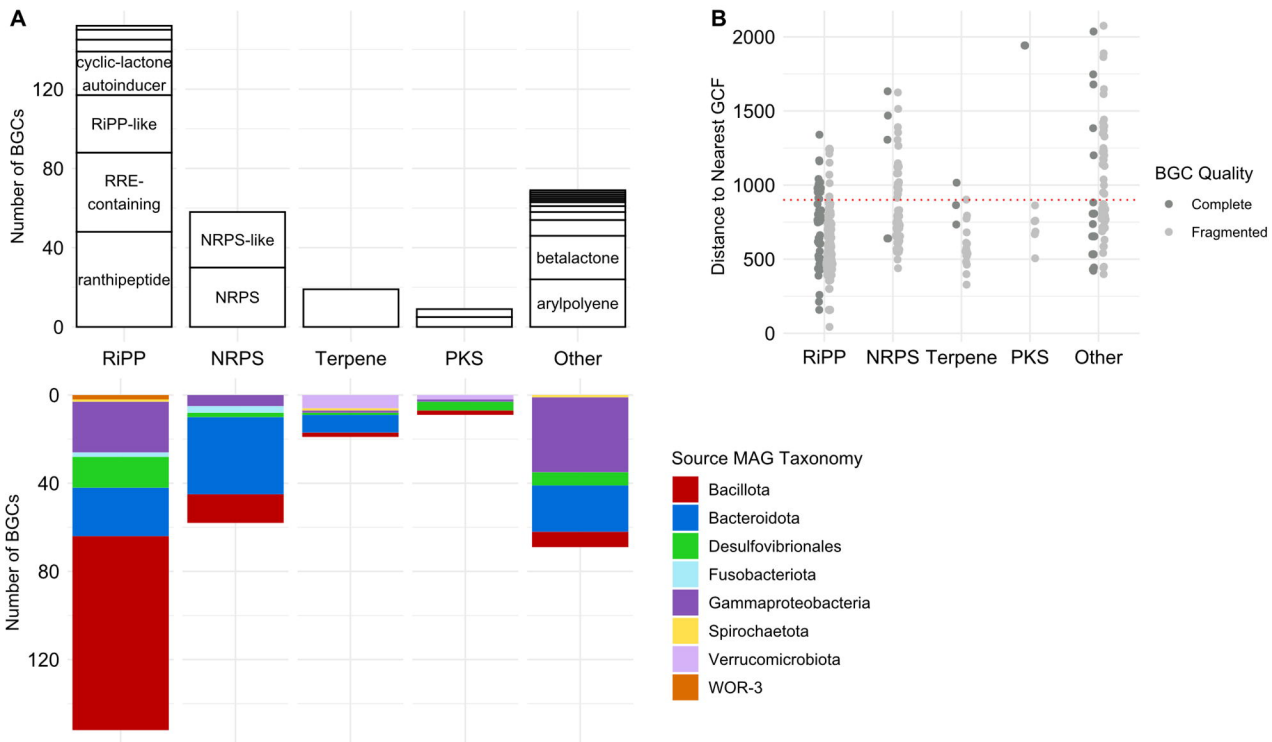

Taxonomy of Assigned Reads
100%
75%
50%
25%
0%
R1_6
R1_12
R1_25
R1_28
R1_29
R1_38
R1_43
R2_8
R2_24
R2_26
F5GI_2
F5GI_3
F5HG_2
F5HG_3
F6GI_3
F6GI_4
F6HG_2
F6HG_3
F7GI_2
F7GI_3
F7HG_2
F7HG_3
F8GI_2
F8GI_3
F8HG_2
F8HG_3
Enrichment
Fish Gut
Read Taxonomy
Eukaryota
Alphaproteobacteria
Bacillota
Bacteroidota
Desulfovibrionales
Fibrobacterota
Fusobacteriota
Gammaproteobacteria
Spirochaetota
Verrucomicrobiota
WOR-3
Unbinned Taxa

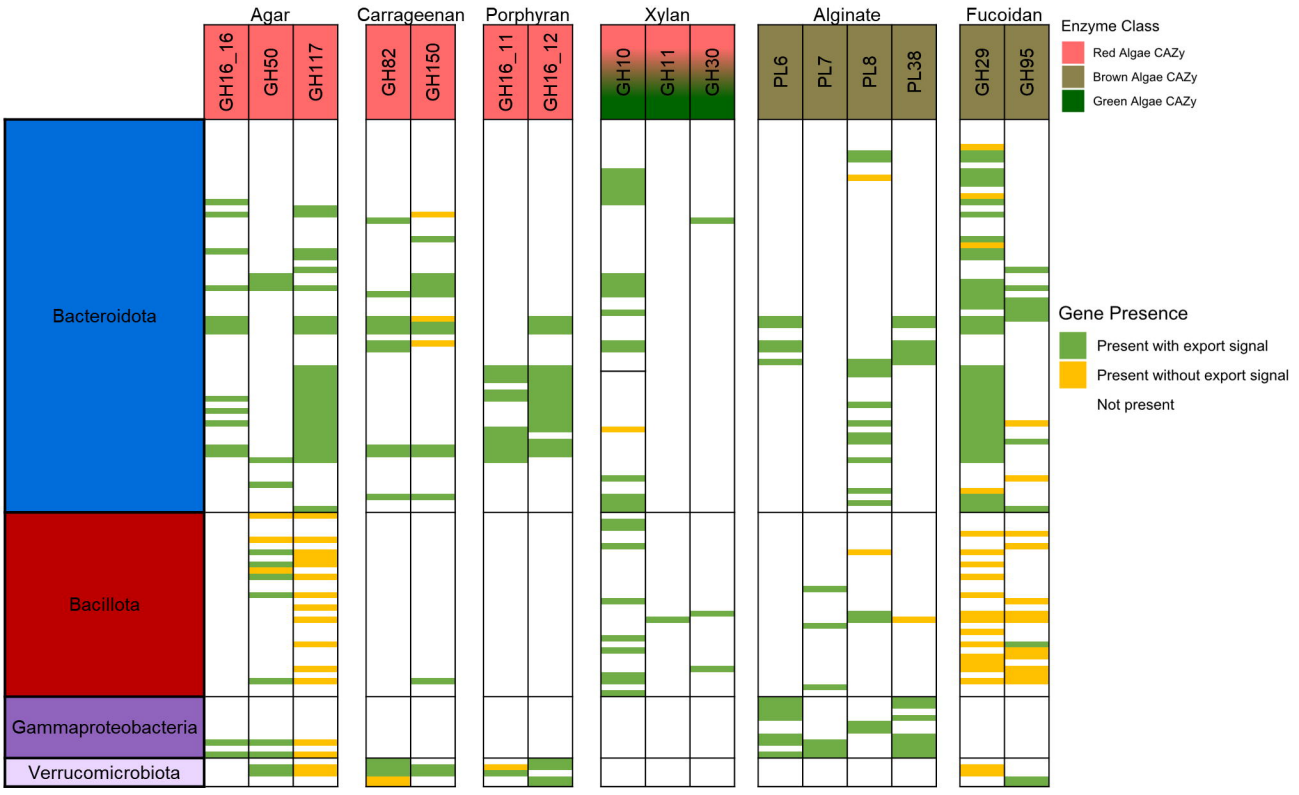

C
B
A
80
60
40
20
0
A MAG Taxonomy
Alphaproteobacteria
Bacillota
Bacteroidota
Desulfovibrionales
Fibrobacterota
Fusobacteriota
Gammaproteobacteria
Spirochaetota
Verrucomicrobiota
WOR-3
B Enzyme Class
Red Algae CAZy
Brown Algae CAZy
Green Algae CAZy
Sulfatases
Peptidoglycanases
C MAG Source
Fish Gut
Enrichment


Acetate
Propanoate
Butyrate
Bacillota (78)
Bacteroidota (72)
Gammaproteobacteria (31)
Desulfovibrionales (13)
Verrucomicrobiota (6)
Pyruvate to Acetate (PFL)
Pyruvate to Acetate (PFOR)
R-pyruvate to R-acetate (porA)
CO2 to Acetate
Glycine to Acetate
Ethanolamine utilization
Choline utilization
Succinate to Propanoate
Acrylate to Propanoate
Threonine to Propanoate
Acetate to Butyrate
Glutamate to Butyrate
Lysine degradation
MAGs with Pathway (%)
1
10
50
75


A
26
73
27
86
31
39
180
28
33
35
20
37
233
26
34
Percent Identity
100
80
60
40
20
GH16_16
GH86
GH16_11
GH10
GH82
GH50
GH117
GH150
GH16_12
PL40
PL6
PL38
GH29
PL8
GH95
CAZyme Class
Database
This Study
GenBank nr
CAZy db
Enzyme Class
Red Algae CAZy
Brown Algae CAZy
B
32
40
83
88
48
36
63
41
95
93
114
44
33
37
22
Percent Identity
100
80
60
40
20
S1_13
S1_72
S1_19
S1_8
S1_24
S1_4
S1_11
S1_20
S1_15
S1_16
S1_17
S1_28
S1_25
S1_14
S1_27
Sulfatase Subfamily
Database
This Study
GenBank nr
SulfAtlas


A
F6HG3_45689
F6GI4_08415
R2_26_18192
WP_212695143.1
WP_212695474.1
R2_26_16226
F6HG3_47524
F6GI4_13335
F6HG3_26041
F6GI4_07963
R1_28_27517
F7HG2_61736
R1_28_27516
F7HG2_61735
R1_28_38714
R1_28_39869
R1_25_14024
R1_28_07742
GH86 Taxonomy
a Bacillota
a Bacteroidota
a Verrucomicrobiota
B
Undescribed Domain
Signal Peptide
Catalytic Domains


A
Number of BGCs
0
40
80
120
cyclic-lactone autoinducer
RiPP-like
RRE-containing
ranthipeptide
NRPS-like
NRPS
betalactone
arylpolyene
RiPP
NRPS
Terpene
PKS
Other
B
Distance to Nearest GCF
0
500
1000
1500
2000
RiPP
NRPS
Terpene
PKS
Other
BGC Quality
Complete
Fragmented
Number of BGCs
0
40
80
120
Source MAG Taxonomy
Bacillota
Bacteroidota
Desulfovibrionales
Fusobacteriota
Gammaproteobacteria
Spirochaetota
Verrucomicrobiota
WOR-3


Agar
Carrageenan
Porphyran
Xylan
Alginate
Fucoidan
GH16_16
GH50
GH117
GH82
GH150
GH16_11
GH16_12
GH10
GH11
GH30
PL6
PL7
PL8
PL38
GH29
GH95
Bacteroidota
Bacillota
Gammaproteobacteria
Verrucomicrobiota
Enzyme Class
Red Algae CAZy
Brown Algae CAZy
Green Algae CAZy
Gene Presence
Present with export signal
Present without export signal
Not present