Herbivorous fish microbiome adaptations to sulfated dietary polysaccharides

Sheila Podell (spodell@ucsd.edu)
University of California, San Diego
Aaron Oliver
University of California, San Diego
Linda Wegley Kelly
University of California, San Diego
Wesley Sparagon
University of Hawaii at Manoa
Robert S. Nelson
National Renewable Energy Laboratory
Lieve M. L. Laurens
National Renewable Energy Laboratory
Neil A. Sims
Ocean-Era Inc
Craig E. Nelson
University of Hawaii at Manoa
Eric E. Allen
University of California, San Diego

Research Article

Keywords: fish gut microbiome, polysaccharide utilization, sulfatase, macroalgal digestion, Kyphosid

Posted Date: June 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1754689/v1

Abstract

Background.

Gut microbiota of kyphosid herbivorous fish, known for feeding almost exclusively on red and brown macroalgae rich in sulfated polysaccharides, contain highly compartmentalized, herbivore-specific taxonomic compositions. The current study explores protein functional activities of microbial taxa comprising this system in digestive compartment-specific metagenomes, identifying specific enzyme associations contributing to macroalgal decomposition.

Results.

Assembled metagenomes from proximal gut samples were enriched in taxa most closely related to environmental marine species, especially *Vibrio*-related Gammaproteobacteria, which progressively decline in hindgut compartments in favor of host-specific groups often found in terrestrial vertebrate microbiomes, including *Alistipes*-related Bacteroidota, Clostridia-related Bacillota, uncharacterized Verrucomicrobiota, and Desulfovibrio-related Deltaproteobacteria. Predicted degradative capacity for sulfated macroalgal polysaccharides is highly enriched in fish hindgut samples, concomitant with expansion of Bacteroidota, Bacillota, and Verrucomicrobiota clades. Extracellularly exported arylsulfatase (SulfAtlas) and carbohydrate active enzyme (CAZy) classes associated with red and brown macroalgal digestion are greatly expanded in kyphosid fish gut versus terrestrial ruminant metagenomes. Genomic mapping of these fish gut-enriched gene functions reveals quantitative signatures suggesting polysaccharide utilization loci associations and coordinated networks of extracellularly exported enzymes targeting specific macroalgal polysaccharides. Co-localization frequencies provide insight into potential gene duplication events and suggest preferred substrates for previously uncharacterized arylsulfatase classes.

Conclusions.

Metagenomic networks of co-localized polysaccharide hydrolysis and sulfatase gene functions have been used to demonstrate distinct patterns of fish gut compartment-specific microbial enzymes targeting complex sulfated polysaccharides. These patterns reveal previously undescribed potential for cooperative exported enzyme activities and highlight prospective contributions of multiple bacterial taxa to overall digestive capabilities.

Background

The digestive systems of herbivorous fish share a surprising number of features with those of terrestrial ruminants, including persistent as well as transient microbial communities that are highly impacted by diet (reviewed in [1–4]). Both animal groups have elongated gastrointestinal tracts divided into multiple

compartments [5–7], including distal hindgut sections where fermentative microbial taxa thrive under anoxic conditions [1, 8–12] to produce short chain fatty acids (SCFAs) that assist in host nutrition [7, 13, 14].

The diets of fish from the genus *Kyphosus* differ from those of terrestrial herbivores, as well as many other fish, in containing large quantities of highly sulfated polysaccharides, including structural carrageenans, porphyrans, and fucoidans from the cell walls of red and brown macroalgae (reviewed in [15–17]). The diverse galactose, mannose, xylose, fucose, rhamnose, arabinose, and uronic acid subunits of macroalgal polysaccharides occur in a wide variety of different ratios, sulfation states, and branching patterns [18–20]. Processing these complex carbohydrates into simpler sugars for incorporation into central metabolic pathways requires sulfate removal either before, concurrently with, or after glycosidic bond cleavage [21–23], but determining precise hydrolysis mechanisms can be difficult for substrates requiring coordinated action by multiple different enzymes. Enzyme characterization ambiguities can also arise in cases of natural substrate promiscuity, potentially confounding experimental determinations based on simplified oligosaccharide model compounds instead of full-length polymers [24]. Additional challenges arise in trying to predict functional specificities from genomic, metagenomic, and/or transcriptomic sequence data for protein families with few experimentally characterized examples.

Microbial taxa capable of hydrolyzing sulfated macroalgal polysaccharides are known to include numerous species from the Bacteroidota lineage, as well as Gammaproteobacteria, Bacillota (formerly known as Firmicutes [25]), and Verrucomicrobiota. Organisms with these capabilities have been found as epiphytic adherents to algal surfaces [26–32], detritovores in seawater and marine sediments [33–36], and in host-associated gut microbiota from humans [37], sea urchins [38], sea cucumbers [39], marine snails [23], and surgeonfishes [40]. Some bacteria may use related enzyme functions to break down sulfated glycosaminoglycans from eukaryotic host tissue [41, 42].

Genomic locations of glycohydrolase genes targeting sulfated polysaccharides are often clustered together within polysaccharide utilization loci (PULs) and CAZyme gene clusters (CGCs) that may also include arylsulfatases, sugar binding proteins, transporters, and regulatory elements [43, 44]. PULs and CGCs can be characterized bioinformatically by sequence comparison to the CAZy database of carbohydrate active enzymes [45], and associated database [46] using either bacterial isolate genomes or culture-independent metagenomes [31]. A large database of arylsulfatase sequences is also available [47], but includes only categories based on broad evolutionary relatedness, rather than experimentally determined substrate specificities.

The breakdown of naturally occurring sulfated polysaccharides via exported extracellular enzymes is known to occur in cultured bacteria isolated from macroalgal surfaces [48], but the extent to which these observations can be applied to the microbiomes of macroalgae-eating vertebrates is unknown. A better understanding of enzyme candidate diversity, species of origin, and subcellular localization might lead to practical applications in aquaculture and animal husbandry, for example by enhancing digestibility of macroalgal feed sources through probiotic supplements [49–53]. This information would also be useful

in identifying key organisms and enzyme variants for biotechnological processing of marine algal feedstocks for the production of value-added products such as biofuels [54], as well as pharmaceutical production of biologically active glycan compounds such as fucoidans, with medicinally valuable anti-inflammatory, anti-cancer, anti-viral, anti-oxidation, anticoagulant, antithrombotic, and anti-angiogenic effects [22, 55].

The current study uses metagenomic analysis to explore microbially encoded proteins in the gut microbiota of three species of macro-algivorous Hawaiian reef fishes, *Kyphosus vaigiensis, Kyphosus cinerascens*, and *Kyphosus hawaiiensis* [56], comparing predicted carbohydrate active enzymes and arylsulfatases to those found in gut metagenomes from terrestrial ruminants [57]. This work extends previous studies showing taxonomic variation of microbial communities across kyphosid gut compartments [14, 58], highlighting taxa likely to be involved in macroalgal carbohydrate digestion, identifying potentially extracellularly exported, cooperative networks of hindgut-specific macroalgal polysaccharide digestion enzymes, and providing foundational support for future work in kyphosid fish microbiome transcriptomics, targeted microbial culture, aquaculture optimization, and industrial processing of macroalgal products.

Methods

Metagenomic sequencing and assembly

DNA samples were isolated from lumen contents distributed over gut compartments from three different Kyphosid fish species (Table 1), using previously described collection and processing procedures [58]. Approximately 30 million paired-end reads were generated per sample using Illumina NovaSeq 6000 technology. Reads from each individual sample were quality filtered and trimmed using Trimmomatic version 0.36 [59], then assembled separately using metaSPAdes version 3.13 [60] with a minimum contig retention size of 2000 nt.

Table 1

Fish gut metagenome samples. Additional descriptive information for these samples is provided in [58]. Abbreviations: GI foregut, HG hindgut. Increasing gut section numbers indicate more distal relative positions within each compartment.

Fish ID number	Species	Description	total gut length (cm)	Gut sections sampled	total metagenomic reads
F5	*K. vaigiensis*	Female adult	108	GI2, GI3, HG2, HG3	33,830,381
F6	*K. cinerascens*	Male adult	123	GI3, GI4, HG2, HG3	32,414,982
F7	*K. hawaiiensis*	Unsexed adult	122	GI2, GI3, HG2, HG3	29,814,385
F8	*K. vaigiensis*	Unsexed juvenile	70	GI2, GI3, HG2, HG3	31,548,838

Taxonomic classification

Taxonomic assignments were made for unassembled reads in each sample using Kraken2, version 2.0.9 [61] using GenBank nr (accessed August 2021) as the search database. Assembled metagenomic contigs were taxonomically classified using DarkHorse version 2.0_rev09, as previously described [62]. DNA-directed RNA polymerase subunit beta (RpoB) protein sequences were retrieved from assembled contigs annotated with PROKKA as described below. Closest sequenced database relatives were identified by top matches in BLASTP searches against both GenBank nr and predicted proteins from terrestrial ruminant metagenome assembled genomes (MAGs) [57]. Multiple sequence alignments of all fish gut RpoB sequences > 900 amino acids in length plus their closest database relatives were obtained using MUSCLE version 3.8.31 [63]. These alignments were used to build phylogenetic trees using FastTree version 2.1.10 [64]. Trees were visualized using the R package ggtree, version 3.3.5 [65]. Fish microbiome 18S rRNA gene sequences were obtained by BLASTN search against the SILVA_138.1_SSURef_NR99 database [66], retaining only those matches with alignments having an e-value of 1e-5 or better, and covering at least > 30% of the reference sequence length.

Gene annotation

Assembled metagenomic contigs were annotated with PROKKA version 1.14 [67], using protein sequence prefixes keyed to source fish sample and gut compartment. Predicted proteins containing microbial extracellular export signals were identified using SignalP version 6.0f [68]. Carbohydrate enzyme classes were assigned using HMMSEARCH with CAZy version 10 database patterns downloaded from dbCAN2 [69]. Sulfatase enzyme categories were determined via BLASTP searches against the SulfAtlas version 1.3 database [47]. An e-value of 1e-15 or better and minimum alignment length covering at least 30% of the protein were used as cutoff values for both HMM patterns and BLASTP searches. Assembled nucleotide sequences for 391 terrestrial ruminant MAGs [57] were downloaded from NCBI BioProject PRJEB34458 and annotated using PROKKA, SignalP, CAZy, and SulfAtlas exactly as described above, to

enable direct comparisons with assembled fish gut metagenomes. PROKKA gene names for terrestrial ruminant genomes were prefixed using GenBank genome identifiers (e.g., RUMXXXX).

Metagenomic occurrence frequencies were tallied for CAZy and SulfAtlas database entries and selected PROKKA annotation keywords, then grouped into subsets according to species of origin (fish: *K. vaigiensis, K. cinerascens*, and *K. hawaiiensis*; ruminant: reindeer, red deer, sheep, goat, cattle, or mixed ruminant assembly) and normalized for total number of predicted protein sequences in each subset. Normalized values were subjected to 1-tailed, homoscedastic t-tests using Microsoft Excel version 16.54 to evaluate statistical significance (p-values) of observed differences between fish and ruminant data sets. Enzyme classes co-localized on the same metagenomic contig were identified using a Unix command-line pipeline of custom perl scripts available on GitHub [70]. Co-occurrence frequencies of enzyme pairs obtained using this pipeline were plotted as edge-weighted network diagrams using Cytoscape version 3.9.1.

Results

Taxonomic composition of metagenomic assemblies

The 16 fish gut metagenomic samples assembled using metaSPAdes yielded a total of 1.478 Tbp (terra base pairs) of assembled nucleotides in contigs > 2 kb, encoding 1,432,202 predicted proteins (**Addditional File 1**). Microbial community compositions of these samples were initially assessed using unassembled reads with Kraken2, enabling relative abundance estimates for eukaryotic, archaeal, and viral taxa as well as bacteria. High-level taxonomic classifications for individual samples (Fig. 1) were consistent with multi-sample 16S rRNA gene amplification averages previously reported for kyphosid fishes [58]. These observations included midgut dominance of Gammaproteobacteria and hindgut enrichment of Bacteroidota and Bacillota, with lower abundances of Alpha-, Beta–, and Deltaproteobacteria, Spirochaetota, Verrucomicrobiota, and Archaea sequences distributed across all compartments. More detailed examination of individual samples showed that Gammaproteobacteria levels varied widely among midgut compartments, and that eukaryotic-associated reads were present in much lower abundances in adult samples from *K. vaigiensis, K. cinerascens*, and *K. hawaiiensis* versus the juvenile *K. vaigiensis* sample (fish 8).

Divergence among fish gut-associated bacterial clades was further explored using predicted protein sequences encoding the single-copy DNA-directed RNA polymerase subunit beta (*rpoB*) gene on assembled contigs. An amino acid sequence tree (Fig. 2) shows the largest numbers of *rpoB* genes recovered were taxonomically classified as Bacillota (48), followed by Bacteroidota (37), Gammaproteobacteria (26), Spirochaetes (7), Deltaproteobacteria (6), Verrucomicrobiota (5), Alphaproteobacteria (4), and Melainabacteria (1). No *rpoB* sequences were detected for Archaea, consistent with the low coverage of this taxonomic group observed in unassembled reads.

Closest database relatives to kyphosid metagenome *rpoB* genes suggest that ingested environmental bacteria may dominate midgut sample microbial communities, with hindgut compartments containing closer relatives to gastrointestinal-associated taxa from terrestrial hosts (**Additional File 2**). Gammaproteobacteria RpoB protein sequences from fish midgut samples most closely resembled marine *Vibrio*, including one shallowly branching clade matching shrimp-associated *Vibrio harveyi*, at 99.9% identity (WP_005433641.1). Although Verrucomicrobiota RpoB sequences were quite divergent from previously described relatives, their closest database matches were to marine bacteria of the Kiritimatiellaceae family (*Kiritimatiella glycovorans*, 78% identity) and algae-associated bacteria from genus *Lentimonas* (WP_162027363.1, 63% identity). One Spirochaetes clade found in fish midgut samples was distantly related to mammalian pathogen *Brevinema andersonii* (SFB68422.1, 84% identity), but *Brevinema*-related taxa have also been observed in 16S rRNA gene amplification studies of trout [71], tilapia [72], salmon [73], seahorses [74], and kyphosids [58], as well as environmental metagenomes from sulfidic artesian boreholes [75]. Spirochaete sequences from fish hindguts clustered more closely with a terrestrial ruminant sequence (RUG30335_00745, 72% identity). The closest database relatives to fish gut *rpoB* sequences from Bacteroidota, Bacillota, Deltaproteobacteria, and Alphaproteobacteria were all MAGs from terrestrial ruminants, rather than marine environmental or bacterial isolate genomes.

Assembled fish gut sequences were also analyzed for the presence of 18S rRNA genes, as markers for eukaryotic organisms (**Additional Files 3–4)**. Teleost 18S rRNA genes were most abundant in samples from juvenile fish F8, consistent with overall taxonomic classifications based on unassembled reads. Non-teleost 18S rRNA gene sequences, found in all fish and distributed across all gut compartments, most frequently matched parasitic single-celled protozoa (*Trichomonadea, Entamoeba, Giardia, Plasmodium*) and multicellular worm taxa (*Enenterum, Acanthocephalus, Enoplus, Opisthadena*) often associated with fish pathology. 18S rRNA genes associated with marine algae (*Hydrolithon*, *Scytosiphon*), fungi (*Stereopsis*), and crustaceans (*Podocopa*) were also detected at low levels, but could potentially have originated from dietary sources.

Protein functional annotations

Although relative abundances of predicted protein functions in metagenomic assemblies do not necessarily correspond with biological activity, expanded numbers of metagenomically predicted genes from specific protein families can indicate potential adaptive capacity. Functional annotations of predicted proteins from fish gut metagenome assemblies (**Additional Files 5–10)** included both generalized and taxon-specific structural and metabolic features consistent with phylogenetic distributions observed in Fig. 1. Some illustrative examples include midgut enrichment of flagellar motility and chemotaxis genes characteristic of Gammaproteobacteria Vibrionaceae and hindgut expansion of genes encoding Bacteroidota-specific SusC/SusD extracellular attachment proteins, as well as glycoside hydrolases targeting galactose, fucose, and xylose-rich polysaccharides typically found in Bacteroidota and Bacillota species from terrestrial microbiomes. Predicted genes annotated as catalases

and superoxide dismutases, suggesting potential resistance to oxidative stress, were observed in bacterial contigs from fish metagenomes in both midgut and hindgut compartments.

The most frequently encountered non-hypothetical gene function descriptions in fish metagenomes were arylsulfatases (1.9% of all predicted proteins in adult fish samples), as might be anticipated from a diet rich in sulfated macroalgal polysaccharides. In contrast, numbers of predicted proteins annotated as dissimilatory sulfite reductases (< 0.007%) were quite low, consistent with low abundances of Desulfovibrio-related Deltaproteobacteria (Fig. 1). Low concentrations of sulfate-reducing bacteria were somewhat unexpected, considering their previously reported expansion under conditions of high sulfate availability in the digestive systems of terrestrial animals [76].

Although no direct oxygen concentration measurements are currently available for kyphosid fish gut compartments, low levels of genes encoding strictly anaerobic processes and the predominance of relatively oxygen-tolerant fermentative taxa such as Bacteroidota, Bacillota, and Gammaproteobacteria suggest the possibility of microaerophilic oxygen stress as a potential inhibitory factor for more strictly anaerobic species. Another striking difference from terrestrial ruminant microbiomes was the absence of key methanogenesis genes such as methyl-coenzyme M reductase, coupled with low read mapping frequencies to methanogenic archaea. This result is consistent with numerous studies demonstrating inhibition of methanogenesis and suppression of methanoarchaeal growth by red and brown macroalgae, both *in vitro* [77] and when used as a dietary supplement for terrestrial ruminants [78].

Macroalgal polysaccharide hydrolysis

High levels of structural micro-heterogeneity in naturally occurring marine macroalgal polysaccharides [15] might be expected to promote expansion and diversification of microbial enzyme families involved in their hydrolysis. This hypothesis was initially tested by comparing predicted protein annotations for kyphosid fish gut metagenomes to a set of 391 MAGs from terrestrial ruminant digestive systems [57], processed using an identical bioinformatic pipeline (Fig. 3, **Additional Files 6–10**). Although the annotations produced by this pipeline were not comprehensive (for example laminarinases, cellulases, and fucosidases were not explicitly labeled as such), these results did reveal significant enrichment of predicted proteins described as agarases, carrageenases, porphyranases, and arylsulfatases in the fish metagenomes (p-values < 0.05). Starch molecules, typically cleaved by enzymes annotated as amylases, occur in both plants and green algae, but not in the red and brown algae typically eaten by kyphosid fish. Amylase keywords were correspondingly more abundant in terrestrial ruminant metagenomes. In contrast, xylanase annotation frequencies did not differ significantly between the two data sets (p-value = .093), consistent with the observation that xylans of various types occur in red and green algae as well as terrestrial plants. Within fish gut compartments, proteins annotated as carrageenases, porphyranases, and agarases were observed at consistently higher frequencies in hindgut versus midgut samples, in contrast to more variable compartmental distributions for alginate lyases and arylsulfatases (**Additional File 11**).

Candidate enzymes for macroalgal digestion were further characterized by classification according to the CAZy reference database (Fig. 4, **Additional File 12)**. CAZy classes targeting red algal agars (GH117, GH50), porphyrans (GH86) and carrageenans (GH82, GH150), as well as non-sulfated alginates from brown algae (PL38, PL6_1, PL7_5) were particularly abundant in fish metagenomes. Class PL40 sequences, whose annotated activities include both exo-alginase and ulvanase, were also significantly expanded in fish gut samples, but other CAZy database classes described as ulvanases (PL24, PL25, PL28, PL37) were not. Abundant fish-specific classes annotated as chondroitinases (GH88, PL8_1, PL30, PL29) and heparin lyases (PL13) are most likely associated with breakdown of host-associated extracellular matrix components. Fish metagenomes were significantly enriched in classes annotated as galactosidases (GH110), iduronidases (GH39), and xylanases (GH10), as well as classes GH136, GH16_14, PL9_2, GH102, annotated as hydrolyzing polysacccharides containing multiple different monomers, but potential involvement of these classes in macroalgal decomposition could not be unambiguously determined. In agreement with annotation keyword results, the great majority of sequences from fish-enriched CAZy classes described as hydrolyzing sulfated macroalgal polysaccharides were obtained from hindgut compartments.

CAZy glycohydrolase classes are based on shared amino acid sequence similarity, protein structural features, and catalytic mechanisms, but often do not distinguish between enzymes acting on chemically similar substrates, especially those with different branching patterns [45, 79, 80]. Polyspecific CAZy classes are particularly common among glycohydrolases annotated as having 1,3 β-glucosidase and/or 1,4 β-glucosidase activities, including laminarinases and cellulases (**Additional File 13**). Classes GH16_3 and GH16_21, historically described as laminarinase subfamilies despite having a substantially broader range of actual substrates [81], were slightly more abundant in fish gut than terrestrial ruminant metagenomes, but these differences were not statistically significant. Other CAZy classes annotated as hydrolyzing non-sulfated, glucose-containing polysaccharides (e.g, GH3, GH5, GH8, GH9, and GH64) were more abundant in metagenomes from terrestrial ruminants than fish. However model enzymes in these classes include not only those hydrolyzing algal polysaccharides [82], but also examples that digest non-algal glycans from bacterial capsid polysaccharides [83] and the cell walls of plants [84] and fungi [85]. These ambiguities preclude confident substrate predictions for these particular CAZy classes based on sequence data alone.

Two recently described CAZy classes (GH107 and GH168) include enzymes experimentally demonstrated to hydrolyze sulfated brown algal fucans [22, 31]. Although neither of these two classes were detected in terrestrial ruminant metagenomes, kyphosid fish samples included 13 predicted sequences from class GH168. Additionally, fucosidase class GH141, present in multiple copies in PULs of known fucan-degrading bacteria [34], was highly expanded in fish gut metagenomes. In contrast, fucosidases from class GH29 were abundant in both kyphosid fish and terrestrial ruminant metagenomes, and class GH95 fucosidases were significantly more abundant in the terrestrial samples.

Bacterial export signal sequences were detected in the majority of sequences from fish-expanded CAZy classes annotated as degrading either red and brown macroalgae or host extracellular matrix

polysaccharides (Fig. 4a) with the notable exception of class GH95, whose members may act intracellularly. In Gram-positive bacteria such as Bacillota, proteins containing signal sequences are translocated across the cell membrane and exported to the extracellular environment. In Gram-negative bacteria such as Bacteriodota, signal sequences mediate initial translocation across the inner membrane to the periplasmic space, which may or may not be followed by secretion across the outer membrane. Percentages of signal sequence-containing macroalgal hydrolysis enzymes were generally lower in terrestrial ruminant gut metagenomes than kyphosid fishes, although N-terminal signal sequences could be missing in some cases due to incomplete metagenomic assembly. However, potential assembly truncation should not adversely affect reliability of the amino acid similarity comparisons presented in Fig. 4C, which show that median percent identities within classes GH95, PL38 and GH168 were much lower than for other classes (50–85% versus 95–99%). These relatively heterogeneous CAZy classes might need to be divided into narrower subclasses as more experimental data becomes available in the future.

Taxonomic distributions of fish gut-expanded CAZy enzyme classes were estimated by protein sequence comparisons to classified relatives in GenBank nr (Fig. 4b). Predicted proteins from classes GH168, GH165, and PL40, targeting brown algal polysaccharides, along with classes PL30, PL29 and PL13, most likely digesting host tissues, exclusively matched database sequences from Bacteroidota-related clades. Predicted carrageenan degrading class GH82 enzymes were confined to taxa identified as either Bacteroidota or Verrucomicrobiota, while other macroalgal-degrading classes also contained lower percentages of matches to proteins classified as Bacillota and Gammaproteobacteria. Although taxonomies cannot be assigned to all enzyme candidates, especially those from groups that are sparsely represented in public databases, these results confirm the dominant role of Bacteroidota as elite complex carbohydrate digesters, as previously described in other environments [25].

Polysaccharide desulfation

Negatively charged sulfate residues in polysaccharides can shield glycosidic bonds from enzymatic cleavage [34], stabilizing carbohydrate backbones against degradation and impeding transport of partially degraded external intermediates into bacterial cells. To identify fish gut-specific enzymes that might help overcome these barriers, all predicted metagenomic arylsulfatases were classified according to SulfAtlas database categories [47], and relative abundances of these sulfatase classes were compared between kyphosid fish and terrestrial ruminant metagenome samples.

More than half of all arylsulfatases detected in fish gut metagenomes fell into the five most abundant SulfAtlas classes, S1_16, S1_15, S1_17, S1_8, and S1_19 (Fig. 5a**).** Predicted proteins from these classes were highly concentrated in hindgut compartments, and greatly enriched in fish gut samples relative to terrestrial ruminants. Other sulfatase classes abundant in fish hindguts, such as S1_7, S1_14, S1_11, and S1_22, were present at similar (p-value > 0.05) or higher numbers in terrestrial ruminant metagenomes. The majority of fish gut metagenome sulfatase sequences contained bacterial export signal sequences, but this was not true of ruminant examples from the same classes, except for S1_13, S1_27 and S1_46, where the export signal pattern was reversed. These three classes were also the only ones not dominated

by database matches to Bacteroidota-related sequences (Fig. 5b). Class S1_13 matches consisted primarily of Gammaproteobacteria, and more than half of S1_27 sequences were most closely matched to Bacillota relatives. Figure 5c shows that intra-class sequence similarity levels were quite high for most SulfAtlas groups (93–100% median amino acid identity), except classes S1_46 and S1_27 (70–85% median amino acid identity).

Co-localization of sulfatase and polysaccharide degradation enzyme classes

CAZy and SulfAtlas classified genes occurring in close proximity to each other on the same contig are potential candidates for co-localization within a common PUL. Although not all metagenomic contigs are long enough to encompass full-sized PULs, which often include as many as 25 adjacent genes [46], median separation distance for the 1453 unique CAZy/SulfAtlas gene pairs detected in fish gut metagenomes was only 4 genes apart, with 98% separated by 25 or fewer genes (**Additional File 14).** In contrast, the median intervening distance for CAZy and SulfAtlas class pairs co-localized on terrestrial ruminant metagenomic contigs was 15 genes, with only 62% falling within a 25 gene distance limit.

CAZy-SulfAtlas class pairs identified by proximity screening can be visualized as a network of nodes, with edges depicting co-localization frequencies (Fig. 6**)**. Similarities and differences in potential PUL associations are highlighted by sub-networks, such as those including only fish-enriched CAZy (Fig. 6b) or SulfAtlas (Fig. 6c) classes, as well as those limited to connections associated with individual enzymes (Figs. 7–8**)**. Co-localization patterns for CAZy and SulfAtlas classes present in both fish gut and terrestrial ruminant metagenomes (**Additional File 16**) showed little similarity to each other, consistent with microbiome adaptation to different dietary inputs.

One striking feature of gene co-localization networks was the presence of highly concentrated links (thicker connecting lines) between some enzyme classes. Higher proximity frequencies associated with these links suggest more likely occurrence within a common PUL, potentially facilitating co-regulation of gene expression. However, both linkage frequencies and the diversity of co-localized nodes varied widely between individual enzyme classes. These variations were not necessarily proportional to overall metagenomic abundance, as illustrated by comparing the co-localization frequency network centered on CAZy class GH50 (94 metagenomic occurrences) with that of class GH150 (95 metagenomic occurrences) in Fig. 7, and SulfAtlas class S1_29 (56 metagenomic occurrences) with SulfAtlas class S1_30 (47 metagenomic occurrences) in Fig. 8. Network comparisons also show enormous variation in the number of self-loops formed by neighboring genes from the same class on a single metagenomic contig, suggesting differences in the expansion of particular classes via gene duplication. The most extreme example of gene duplication was observed in class PL38 alginate lyases, where tandem repeats of 7–10 closely related sequences were found in several assembled contigs predicted to originate from an unknown Spirochaete lineage.

Detailed comparisons of co-localized enzyme classes can reveal subtle relationships and provide hints about potential metabolic interactions that might benefit from co-regulation. As an example, sulfatase

classes S1_16 and S1_8 were frequently co-localized on the same contig, suggesting potential complementary PUL-linked functions within a single genome. Both classes were also frequently paired with polysaccharide-degrading CAZy classes that included those targeting highly sulfated iota- (GH82) and lambda- (GH150) carrageenans, as well as exo-α-galactosidase class GH110, α-1,3-(3,6)-anhydro-D-galactosidase class GH127, and generalized hexosidase class GH2. Links to CAZy classes targeting agarases (GH117, GH50) and porphyranases (GH16_11) occurred much less frequently. A different pattern was observed in terrestrial ruminant metagenomes, where network connections for sulfatase class S1_16 were more evenly distributed over a wider variety of partner classes (**Additional File 13**), while S1_8 was more strongly paired with sulfatases S1_27 and S1_4 instead of S1_16, along with glycosyltransferases, carboxylesterases, and general glycoside hydrolases like GH2 and GH3.

A matrix summarizing SulfAtlas/CAZy co-localization frequencies (Fig. 9) shows that nearly all SulfAtlas classes expanded in fish gut metagenomes included some examples of proximity with red algal polysaccharide-hydrolyzing CAZy classes, but some were also located near CAZy classes predicted to degrade brown algal and/or host chondroitin substrates. Classes S1_17, S1_14, S1_28, and S1_25, were predominantly linked to CAZy classes hydrolyzing brown algal substrates, while classes S1_19, S1_20, S1_27, and S1_30, were more narrowly associated with red-algal digesting enzyme sequences. These distinctively different co-localization patterns suggest promising avenues for future experimental determination of sulfatase class-specific substrate ranges.

One limitation of metagenomic co-localization analysis is that not all paired enzyme features are equally informative, particularly those including diverse glycohydrolase classes like GH2. At least a dozen different functional activities have been ascribed to this CAZy class, targeting not only β-galactosides, but also β-mannosides, β-glucuronosides, and β-galacturonosides. Another constraint is that predicted proteins from sparsely represented microbial taxa are often encoded on shorter, less completely assembled contigs, reducing the amount of available information about potential neighboring genes. Although this limitation can decrease detection sensitivity for rare enzyme class pairs, infrequent relationships could also be interpreted as less important to overall fish digestive processes than features showing elevated metagenomic linkage frequencies that potentially contribute more to successful metabolic strategies, resulting in higher community abundances for the organisms encoding them.

Discussion

Metagenomic sequencing and assembly of 16 different, fish gut compartment-specific samples has enabled the connection of specific taxonomic groups with predicted protein functional activities likely to contribute to efficient macroalgal digestion. Persistent, rather than transient host associations seem probable for hindgut-enriched Bacteroidota, Bacillota, and Verrucomicrobiota clades that are more similar to terrestrial herbivore-associated taxa than to those from the external marine environment. Midgut-enriched sequences, which are dominated by taxa most closely related to free-living marine Gammaproteobacteria, are likely to be replenished frequently from ingested food sources and surrounding seawater.

Consistently limited taxonomic diversity in fish gut microbiomes might be enforced by extracellular sequestration of macroalgal carbohydrate breakdown products by mechanisms previously demonstrated to be active in carbohydrate-utilizing Bacteroidota and Bacillota from other systems, for example SusC/SusD [86] and cellulosome [87] complexes. However, these structures would not be expected to limit diffusion or availability of free sulfate groups liberated by digestion of highly sulfated polysaccharides. The low levels of sulfate-reducing Deltaproteobacteria observed were therefore surprising, since this group has previously been shown to flourish under high sulfate conditions in host-associated microbial communities from terrestrial ruminants [76].

It is possible that sulfate-reducing bacteria could be inhibited in fish gut microbiomes by small quantities of oxygen that poison dissimilatory sulfate reduction. The constant flow of fresh seawater through kyphosid digestive systems might serve to partially replenish oxygen levels depleted by microbial metabolism, favoring fermentative Gammaproteobacteria, Bacteroidota and Bacillota species that are more aerotolerant [88, 89]. Relatively high metagenomic abundance of genes encoding catalases and superoxide dismutases throughout both midgut and hindgut compartments lend support to this hypothesis. Alternatively, the scarcity of sulfate-reducing bacteria, despite sulfate concentrations that can exceed 100 mM in the hindgut [90], may also be influenced by competitive exclusion as a result of host- or microbiome-derived factors in addition to abiotic determinants [7].

Despite general similarities in taxonomic composition between kyphosid fish and terrestrial ruminant microbiome communities, relative abundances of sulfatase and CAZy enzyme classes targeting macroalgal-specific polysaccharides are quite different, consistent with non-overlapping dietary sources. The reliance of previously described large-scale PUL detection methods on sequences from well-characterized taxa [36, 91] has been avoided in this study by detecting metagenomic enzyme class co-localizations that also include uncultured and poorly studied environmental taxa, regardless of whether or not they use typical PUL architectures. Co-localization-based associations can be informative even in cases where only partial PULs are assembled, by identifying statistically verifiable frequency patterns replicated in multiple independent samples.

The hundreds of new macroalgal degradation candidate sequences identified in this study provide valuable tools for exploring evolutionary mechanisms that may be responsible for acquisition and expansion of macroalgal degradation capabilities among individual taxa. The discovery of closely related enzyme class members positioned as neighbors on the same metagenomic contig suggests gene duplication as a likely mechanism for expanding substrate range without compromising current competency. This possibility could be explored further by constructing multiple sequence alignments for enzyme classes containing members of divergent taxonomic groups. These alignments might then be used to identify conserved amino acid residues associated with catalytic mechanisms and/or substrate binding activity, and for constructing gene trees highlighting potential acquisition of macroalgal breakdown capabilities through horizontal gene transfer.

Frequently observed extracellular export signals in fish gut-enriched CAZy and SulfAtlas classes, combined with the known complexity and diversity of naturally occurring macroalgal polysaccharides, suggest potential opportunities for cooperative activity on recalcitrant substrates. Exported enzyme cooperativity could include multiple genes originating from the same individual strain, variants comprising the pan-genomic repertoire of closely related strains, or even pan-microbiome diversity arising from widely different taxa within the gut microbial community. Future strategies for testing this hypothesis could include reconstruction of metagenome assembled genomes (MAGs) from binned metagenomic contigs, transcriptional mapping to determine *in vivo* gene expression levels, *in vitro* substrate hydrolysis measurements using combinations of purified enzymes obtained by expression cloning, and comparisons of macroalgal degradation performance by enrichments or defined mixed cultures of live bacteria.

Conclusions

Comparisons of marine and terrestrial herbivore microbiomes described in this study have revealed fish gut compartment-specific microbial adaptations to a diet rich in sulfated macroalgal polysaccharides, improving our understanding of the enzymes and organisms involved in digesting these molecules and providing resources for future investigations into their potential for cooperative processing.

List Of Abbreviations

PUL: polysaccharide utilization loci, CAZy: carbohydrate active enzyme, HG: hindgut, GI: gastrointestinal midgut, PL: polysaccharide lyase, GH: glycoside hydrolase, MAG: metagenome-assembled genome, HMM: hidden markov model.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

The datasets generated and/or analyzed during the current study are available through NCBI project id PRJNA819194, under SRA accession numbers SRR19136343-SRR19136358 for raw reads, WGS accession numbers JAMHIX000000000-JAMHIZ000000000 and JAMHJA000000000-JAMHJM000000000 for assembled contigs. Amino acid sequences for PROKKA predicted proteins from assembled contigs are available in fasta format under Zenodo DOIs 10.5281/zenodo.6635023 for

kyphosid fish gut metagenomes generated in this study and 10.5281/zenodo.6635166 for terrestrial ruminant MAGs from [57].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the United States Government, Department of Energy, Advanced Research Projects Agency–Energy grant ARPA-E DE-FOA-0001858 to RSN and LMLL with contracts to CEN, LWK, and EEA; National Science Foundation grants OCE-1837116 and EF-2025217 to EEA; and National Institutes of Health NIEHS grant R01-ES030316 to EEA.

Authors' contributions

SP, LWK, RSN, LMLL, NAS, CEN, and EEA conceived the study. WJS and LWK performed wet-lab fish gut sampling and DNA isolations. SP and AO designed and performed bioinformatic analyses and generated visualizations. Custom computer code was designed and written by SP, who also wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge Lucie Leonard, Keelee Martin, and the entire Ocean Era team for assistance in organizing and conducting fieldwork for sample collection. We also thank Helen Meigs and all of the local fisherfolk for assistance in field sampling.

References

1. Egerton S, Culloty S, Whooley J, Stanton C, Ross RP. The Gut Microbiota of Marine Fish. Front Microbiol. 2018;9:873; doi: 10.3389/fmicb.2018.00873.
2. Legrand TPRA, Wynne JW, Weyrich LS, Oxley APA. A microbial sea of possibilities: current knowledge and prospects for an improved understanding of the fish microbiome. Reviews in Aquaculture. 2019;12(2):1101–34; doi: 10.1111/raq.12375.
3. Scott JJ, Adam TC, Duran A, Burkepile DE, Rasher DB. Intestinal microbes: an axis of functional diversity among large marine consumers. Proc Biol Sci. 2020;287(1924):20192367; doi: 10.1098/rspb.2019.2367.
4. Wang AR, Ran C, Ringø E, Zhou ZG. Progress in fish gastrointestinal microbiota research. Reviews in Aquaculture. 2018;10(3):626–40; doi: 10.1111/raq.12191.
5. Choat JH, Clements KD. Vertebrate herbivores in marine and terrestrial environments: A nutritional ecology perspective. Annual Review of Ecology and Systematics. 1998;29:375–403; doi: DOI 10.1146/annurev.ecolsys.29.1.375.

6. Sturm E, Horn M. Food habits, gut morphology and pH, and assimilation effciency of the zebraperch Hermosilla azurea, an herbivorous kyphosid fsh of temperate marine waters. Marine Biology. 1998;132:515–22.
7. Mountfort DO, Campbell J, Clements KD. Hindgut fermentation in three species of marine herbivorous fish. Appl Environ Microbiol. 2002;68(3):1374–80; doi: 10.1128/AEM.68.3.1374-1380.2002.
8. Fidopiastis PM, Bezdek DJ, Horn MH, Kandel JS. Characterizing the resident, fermentative microbial consortium in the hindgut of the temperate-zone herbivorous fish, Hermosilla azurea (Teleostei: Kyphosidae). Marine Biology. 2005;148(3):631–42; doi: 10.1007/s00227-005-0106-2.
9. Nielsen S, Wilkes Walburn J, Verges A, Thomas T, Egan S. Microbiome patterns across the gastrointestinal tract of the rabbitfish Siganus fuscescens. PeerJ. 2017;5:e3317; doi: 10.7717/peerj.3317.
10. Jones J, DiBattista JD, Stat M, Bunce M, Boyce MC, Fairclough DV, et al. The Microbiome of the Gastrointestinal Tract of a Range-Shifting Marine Herbivorous Fish. Front Microbiol. 2018;9:2000; doi: 10.3389/fmicb.2018.02000.
11. Feng W, Zhang J, Jakovlic I, Xiong F, Wu S, Zou H, et al. Gut segments outweigh the diet in shaping the intestinal microbiota composition in grass carp Ctenopharyngodon idellus. AMB Express. 2019;9(1):44; doi: 10.1186/s13568-019-0770-0.
12. Ringø E, Zhou Z, Vecino JLG, Wadsworth S, Romero J, Krogdahl Å, et al. Effect of dietary components on the gut microbiota of aquatic animals. A never-ending story? Aquaculture Nutrition. 2015;22(2):219–82; doi: 10.1111/anu.12346.
13. Briggs JA, Grondin JM, Brumer H. Communal living: glycan utilization by the human gut microbiota. Environ Microbiol. 2021;23(1):15–35; doi: 10.1111/1462-2920.15317.
14. 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; doi: 10.1111/mec.16394.
15. 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. Ann Rev Mar Sci. 2021;13:81–108; doi: 10.1146/annurev-marine-032020-012810.
16. Zhang K, Zhang F, Wu YR. Emerging technologies for conversion of sustainable algal biomass into value-added products: A state-of-the-art review. Sci Total Environ. 2021;784:147024; doi: 10.1016/j.scitotenv.2021.147024.
17. 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. Frontiers in Marine Science. 2022;9; doi: 10.3389/fmars.2022.882196.
18. Hehemann JH, Boraston AB, Czjzek M. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. Curr Opin Struct Biol. 2014;28:77–86; doi: 10.1016/j.sbi.2014.07.009.

19. Michel G, Czjzek M. Polysaccharide-degrading enzymes from marine bacteria. Woodhead Publishing; 2013.
20. Lee WK, Ho CL. Ecological and evolutionary diversification of sulphated polysaccharides in diverse photosynthetic lineages: A review. Carbohydr Polym. 2022;277:118764; doi: 10.1016/j.carbpol.2021.118764.
21. Prechoux A, Genicot S, Rogniaux H, Helbert W. Enzyme-Assisted Preparation of Furcellaran-Like kappa-/beta-Carrageenan. Mar Biotechnol (NY). 2016;18(1):133–43; doi: 10.1007/s10126-015-9675-3.
22. Silchenko AS, Rasin AB, Zueva AO, Kusaykin MI, Zvyagintseva TN, Rubtsov NK, et al. Discovery of a fucoidan endo-4O-sulfatase: Regioselective 4O-desulfation of fucoidans and its effect on anticancer activity in vitro. Carbohydr Polym. 2021;271:118449; doi: 10.1016/j.carbpol.2021.118449.
23. Vuillemin M, Silchenko AS, Cao HTT, Kokoulin MS, Trang VTD, Holck J, et al. Functional Characterization of a New GH107 Endo-alpha-(1,4)-Fucoidanase from the Marine Bacterium Formosa haliotis. Mar Drugs. 2020;18(11); doi: 10.3390/md18110562.
24. Liu G, Wu S, Jin W, Sun C. Amy63, a novel type of marine bacterial multifunctional enzyme possessing amylase, agarase and carrageenase activities. Sci Rep. 2016;6:18726; doi: 10.1038/srep18726.
25. Garcia-Lopez M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, Kyrpides NC, et al. Analysis of 1,000 Type-Strain Genomes Improves Taxonomic Classification of Bacteroidetes. Front Microbiol. 2019;10:2083; doi: 10.3389/fmicb.2019.02083.
26. Hehemann JH, Correc G, Thomas F, Bernard T, Barbeyron T, Jam M, et al. Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium Zobellia galactanivorans. J Biol Chem. 2012;287(36):30571–84; doi: 10.1074/jbc.M112.377184.
27. Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T, et al. The genome of the alga-associated marine flavobacterium Formosa agariphila KMM 3901T reveals a broad potential for degradation of algal polysaccharides. Appl Environ Microbiol. 2013;79(21):6813–22; doi: 10.1128/AEM.01937-13.
28. Bakunina IY, Nedashkovskaya OI, Balabanova LA, Kukhlevsky AD, Belous OS. Activities of O-Glycoside Hydrolases and Other Polysaccharide-Degrading Enzymes of Cultivable Bacterial Isolates of the Pacific Red Alga Ahnfeltia tobuchiensis (Kanno et Matsubara, 1932) Makienko, 1970. Russian Journal of Marine Biology. 2020;46(5):387–9; doi: 10.1134/S1063074020050028.
29. Ficko-Blean E, Prechoux A, Thomas F, Rochat T, Larocque R, Zhu Y, et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nat Commun. 2017;8(1):1685; doi: 10.1038/s41467-017-01832-6.
30. Reisky L, Stanetty C, Mihovilovic MD, Schweder T, Hehemann JH, Bornscheuer UT. Biochemical characterization of an ulvan lyase from the marine flavobacterium Formosa agariphila KMM 3901(T). Appl Microbiol Biotechnol. 2018;102(16):6987–96; doi: 10.1007/s00253-018-9142-y.

31. Schultz-Johansen M, Bech PK, Hennessy RC, Glaring MA, Barbeyron T, Czjzek M, et al. A Novel Enzyme Portfolio for Red Algal Polysaccharide Degradation in the Marine Bacterium Paraglaciecola hydrolytica S66(T) Encoded in a Sizeable Polysaccharide Utilization Locus. Front Microbiol. 2018;9:839; doi: 10.3389/fmicb.2018.00839.
32. Gobet A, Barbeyron T, Matard-Mann M, Magdelenat G, Vallenet D, Duchaud E, et al. Evolutionary Evidence of Algal Polysaccharide Degradation Acquisition by Pseudoalteromonas carrageenovora 9(T) to Adapt to Macroalgal Niches. Front Microbiol. 2018;9:2740; doi: 10.3389/fmicb.2018.02740.
33. Gao B, Jin M, Li L, Qu W, Zeng R. Genome Sequencing Reveals the Complex Polysaccharide-Degrading Ability of Novel Deep-Sea Bacterium Flammeovirga pacifica WPAGA1. Front Microbiol. 2017;8:600; doi: 10.3389/fmicb.2017.00600.
34. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nat Microbiol. 2020;5(8):1026–39; doi: 10.1038/s41564-020-0720-2.
35. Wolter LA, Mitulla M, Kalem J, Daniel R, Simon M, Wietz M. CAZymes in Maribacter dokdonensis 62 – 1 From the Patagonian Shelf: Genomics and Physiology Compared to Related Flavobacteria and a Co-occurring Alteromonas Strain. Front Microbiol. 2021;12:628055; doi: 10.3389/fmicb.2021.628055.
36. Alejandre-Colomo C, Francis B, Viver T, Harder J, Fuchs BM, Rossello-Mora R, et al. Cultivable Winogradskyella species are genomically distinct from the sympatric abundant candidate species. ISME Communications. 2021;1(1); doi: 10.1038/s43705-021-00052-w.
37. Li M, Shang Q, Li G, Wang X, Yu G. Degradation of Marine Algae-Derived Carbohydrates by Bacteroidetes Isolated from Human Gut Microbiota. Mar Drugs. 2017;15(4); doi: 10.3390/md15040092.
38. Silchenko AS, Rasin AB, Zueva AO, Kusaykin MI, Zvyagintseva TN, Kalinovsky AI, et al. Fucoidan Sulfatases from Marine Bacterium Wenyingzhuangia fucanilytica CZ1127(T). Biomolecules. 2018;8(4); doi: 10.3390/biom8040098.
39. Mikkelsen MD, Cao HTT, Roret T, Rhein-Knudsen N, Holck J, Tran VTT, et al. A novel thermostable prokaryotic fucoidan active sulfatase PsFucS1 with an unusual quaternary hexameric structure. Sci Rep. 2021;11(1):19523; doi: 10.1038/s41598-021-98588-3.
40. Ngugi DK, Miyake S, Cahill M, Vinu M, Hackmann TJ, Blom J, et al. Genomic diversification of giant enteric symbionts reflects host dietary lifestyles. Proc Natl Acad Sci U S A. 2017;114(36):E7592-E601; doi: 10.1073/pnas.1703070114.
41. Luis AS, Jin C, Pereira GV, Glowacki RWP, Gugel SR, Singh S, et al. A single sulfatase is required to access colonic mucin by a gut bacterium. Nature. 2021;598(7880):332–7; doi: 10.1038/s41586-021-03967-5.
42. Ndeh D, Basle A, Strahl H, Yates EA, McClurgg UL, Henrissat B, et al. Metabolism of multiple glycosaminoglycans by Bacteroides thetaiotaomicron is orchestrated by a versatile core genetic locus. Nat Commun. 2020;11(1):646; doi: 10.1038/s41467-020-14509-4.

43. Lapebie P, Lombard V, Drula E, Terrapon N, Henrissat B. Bacteroidetes use thousands of enzyme combinations to break down glycans. Nat Commun. 2019;10(1):2043; doi: 10.1038/s41467-019-10068-5.
44. 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. 2021;49(D1):D523-D8; doi: 10.1093/nar/gkaa742.
45. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490-5; doi: 10.1093/nar/gkt1178.
46. Terrapon N, Lombard V, Drula E, Lapebie P, Al-Masaudi S, Gilbert HJ, et al. PULDB: the expanded database of Polysaccharide Utilization Loci. Nucleic Acids Res. 2018;46(D1):D677-D83; doi: 10.1093/nar/gkx1022.
47. Barbeyron T, Brillet-Gueguen L, Carre W, Carriere 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;11(10):e0164846; doi: 10.1371/journal.pone.0164846.
48. Brunet M, Le Duff N, Barbeyron T, Thomas F. Consuming fresh macroalgae induces specific catabolic pathways, stress reactions and Type IX secretion in marine flavobacterial pioneer degraders. ISME J. 2022; doi: 10.1038/s41396-022-01251-6.
49. Yukgehnaish K, Kumar P, Sivachandran P, Marimuthu K, Arshad A, Paray BA, et al. Gut microbiota metagenomics in aquaculture: factors influencing gut microbiome and its physiological role in fish. Reviews in Aquaculture. 2020;12(3):1903–27; doi: 10.1111/raq.12416.
50. Wang A, Ran C, Wang Y, Zhang Z, Ding Q, Yang Y, et al. Use of probiotics in aquaculture of China-a review of the past decade. Fish Shellfish Immunol. 2019;86:734–55; doi: 10.1016/j.fsi.2018.12.026.
51. Ghosh K, Ray AK, Ringø E. Applications of plant ingredients for tropical and subtropical freshwater finfish: possibilities and challenges. Reviews in Aquaculture. 2019;11(3):793–815; doi: 10.1111/raq.12258.
52. Hoseinifar SH, Dadar M, Ringø E. Modulation of nutrient digestibility and digestive enzyme activities in aquatic animals: The functional feed additives scenario. Aquaculture Research. 2017;48(8):3987–4000; doi: 10.1111/are.13368.
53. Zhao Y, Zhao G. Decreasing ruminal methane production through enhancing the sulfate reduction pathway. Animal Nutrition. 2022; doi: 10.1016/j.aninu.2022.01.006.
54. Kumar M, Sun Y, Rathour R, Pandey A, Thakur IS, Tsang DCW. Algae as potential feedstock for the production of biofuels and value-added products: Opportunities and challenges. Sci Total Environ. 2020;716:137116; doi: 10.1016/j.scitotenv.2020.137116.
55. Lin Y, Qi X, Liu H, Xue K, Xu S, Tian Z. The anti-cancer effects of fucoidan: a review of both in vivo and in vitro investigations. Cancer Cell Int. 2020;20:154; doi: 10.1186/s12935-020-01233-8.
56. Knudsen SW, Choat JH, Clements KD. The herbivorous fish family Kyphosidae (Teleostei: Perciformes) represents a recent radiation from higher latitudes. Journal of Biogeography.

2019;46(9):2067–80; doi: 10.1111/jbi.13634.

57. Glendinning L, Genc B, Wallace RJ, Watson M. Metagenomic analysis of the cow, sheep, reindeer and red deer rumen. Sci Rep. 2021;11(1):1990; doi: 10.1038/s41598-021-81668-9.
58. 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;4(1):33; doi: 10.1186/s42523-022-00182-z.
59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20; doi: 10.1093/bioinformatics/btu170.
60. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27(5):824–34; doi: 10.1101/gr.213959.116.
61. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257; doi: 10.1186/s13059-019-1891-0.
62. Podell S, Ugalde JA, Narasingarao P, Banfield JF, Heidelberg KB, Allen EE. Assembly-driven community genomics of a hypersaline microbial ecosystem. PLoS One. 2013;8(4):e61692; doi: 10.1371/journal.pone.0061692.
63. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7; doi: 10.1093/nar/gkh340.
64. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490; doi: 10.1371/journal.pone.0009490.
65. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinformatics. 2020;69(1):e96; doi: 10.1002/cpbi.96.
66. 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 Res. 2013;41(Database issue):D590-6; doi: 10.1093/nar/gks1219.
67. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9; doi: 10.1093/bioinformatics/btu153.
68. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022; doi: 10.1038/s41587-021-01156-3.
69. 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;46(W1):W95-W101; doi: 10.1093/nar/gky418.
70. Podell S. Metagenomic Co-localization Analysis Pipeline. https://github.com/spodell/MCoLA (2022). Accessed 06/09/2022.
71. Brown RM, Wiens GD, Salinas I. Analysis of the gut and gill microbiome of resistant and susceptible lines of rainbow trout (Oncorhynchus mykiss). Fish Shellfish Immunol. 2019;86:497–506; doi: 10.1016/j.fsi.2018.11.079.

72. Li M, Li L, Huang T, Liu Y, Lei A, Ma C, et al. Effects of Attenuated S. agalactiae Strain YM001 on Intestinal Microbiota of Tilapia Are Recoverable. Front Microbiol. 2018;9:3251; doi: 10.3389/fmicb.2018.03251.
73. Li Y, Bruni L, Jaramillo-Torres A, Gajardo K, Kortner TM, Krogdahl A. Differential response of digesta- and mucosa-associated intestinal microbiota to dietary insect meal during the seawater phase of Atlantic salmon. Anim Microbiome. 2021;3(1):8; doi: 10.1186/s42523-020-00071-3.
74. Wang R, Pan X, Xu Y. Altered Intestinal Microbiota Composition Associated with Enteritis in Yellow Seahorses Hippocampus kuda (Bleeker, 1852). Curr Microbiol. 2020;77(5):730–7; doi: 10.1007/s00284-019-01859-6.
75. Karnachuk OV, Lukina AP, Kadnikov VV, Sherbakova VA, Beletsky AV, Mardanov AV, et al. Targeted isolation based on metagenome-assembled genomes reveals a phylogenetically distinct group of thermophilic spirochetes from deep biosphere. Environ Microbiol. 2021;23(7):3585–98; doi: 10.1111/1462-2920.15218.
76. Dordevic D, Jancikova S, Vitezova M, Kushkevych I. Hydrogen sulfide toxicity in the gut environment: Meta-analysis of sulfate-reducing and lactic acid bacteria in inflammatory processes. J Adv Res. 2021;27:55–69; doi: 10.1016/j.jare.2020.03.003.
77. 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(1):e85289; doi: 10.1371/journal.pone.0085289.
78. 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. Anim Microbiome. 2019;1(1):3; doi: 10.1186/s42523-019-0004-4.
79. Barras D, Moore A, BA S. Enzyme-Substrate Relationships Among β-Glucan Hydrolas. In: Cellulases and Their Applications. Washington, DC: American Chemical Society; 1969. p. 105–38.
80. Nguyen STC, Freund HL, Kasanjian J, Berlemont R. Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy. Appl Microbiol Biotechnol. 2018;102(4):1629–37; doi: 10.1007/s00253-018-8778-y.
81. 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(44):15973–86; doi: 10.1074/jbc.RA119.010619.
82. Salmean AA, Duffieux D, Harholt J, Qin F, Michel G, Czjzek M, et al. Insoluble (1 –> 3), (1 –> 4)-beta-D-glucan is a component of cell walls in brown algae (Phaeophyceae) and is masked by alginates in tissues. Sci Rep. 2017;7(1):2880; doi: 10.1038/s41598-017-03081-5.
83. McIntosh M, Stone BA, Stanisich VA. Curdlan and other bacterial (1–>3)-beta-D-glucans. Appl Microbiol Biotechnol. 2005;68(2):163–73; doi: 10.1007/s00253-005-1959-5.
84. Kuge T, Nagoya H, Tryfona T, Kurokawa T, Yoshimi Y, Dohmae N, et al. Action of an endo-beta-1,3(4)-glucanase on cellobiosyl unit structure in barley beta-1,3:1,4-glucan. Biosci Biotechnol Biochem. 2015;79(11):1810–7; doi: 10.1080/09168451.2015.1046365.

85. Chaari F, Chaabouni SE. Fungal beta-1,3 – 1,4-glucanases: production, proprieties and biotechnological applications. J Sci Food Agric. 2019;99(6):2657–64; doi: 10.1002/jsfa.9491.
86. Foley MH, Cockburn DW, Koropatkin NM. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. Cell Mol Life Sci. 2016;73(14):2603–17; doi: 10.1007/s00018-016-2242-x.
87. Fontes CM, Gilbert HJ. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem. 2010;79:655–81; doi: 10.1146/annurev-biochem-091208-085603.
88. Aldunate M, De la Iglesia R, Bertagnolli AD, Ulloa O. Oxygen modulates bacterial community composition in the coastal upwelling waters off central Chile. Deep-Sea Research Part Ii-Topical Studies in Oceanography. 2018;156:68–79; doi: 10.1016/j.dsr2.2018.02.001.
89. Dufourny S, Everaert N, Lebrun S, Didelez M, Wavreille J, Froidmont E, et al. Oxygen as a key parameter in in vitro dynamic and multi-compartment models to improve microbiome studies of the small intestine? Food Res Int. 2020;133:109127; doi: 10.1016/j.foodres.2020.109127.
90. Evans DH, Claiborne JB, Currie S. Chapter 5: Intestinal Transport. In: The Physiology of Fishes, Fourth Edition. Taylor & Francis; 2013. p. 175–204.
91. 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(8):953–61; doi: 10.1038/s41587-019-0202-3.

Figures

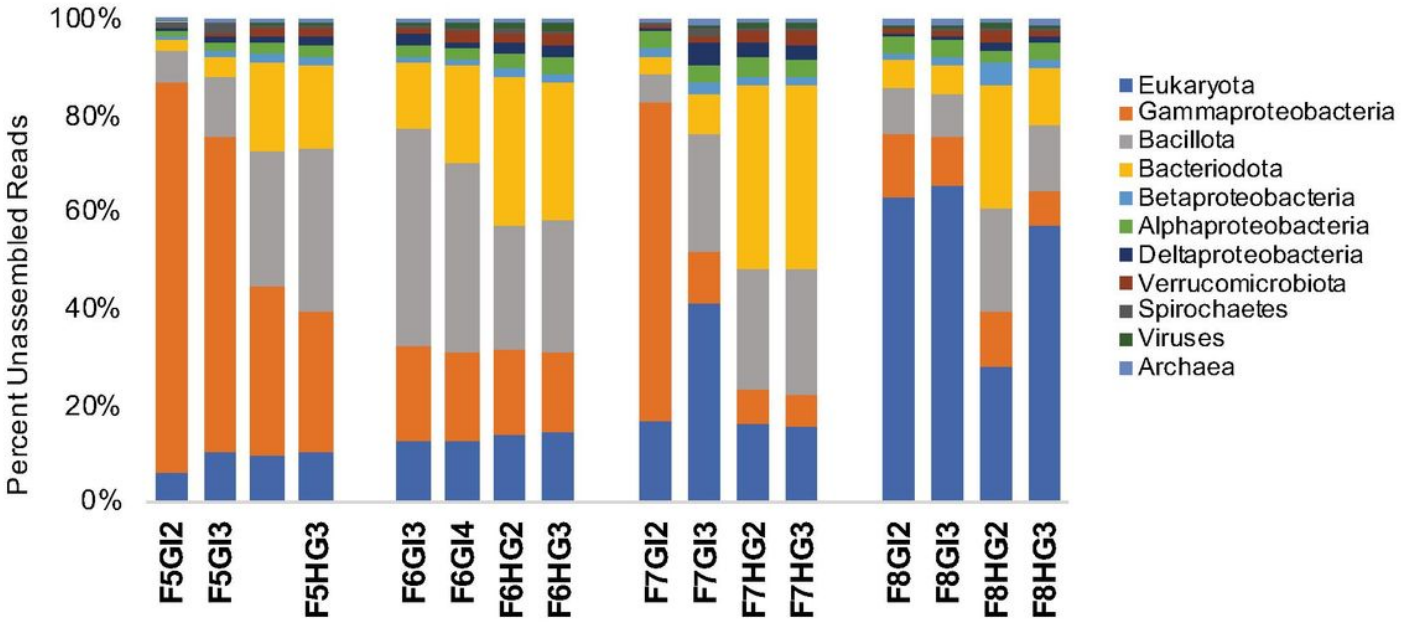

Figure 1

Compartmental Abundance of Microbial Taxa.

Relative abundances are based on Kraken2 analysis of unassembled Illumina NovaSeq reads.

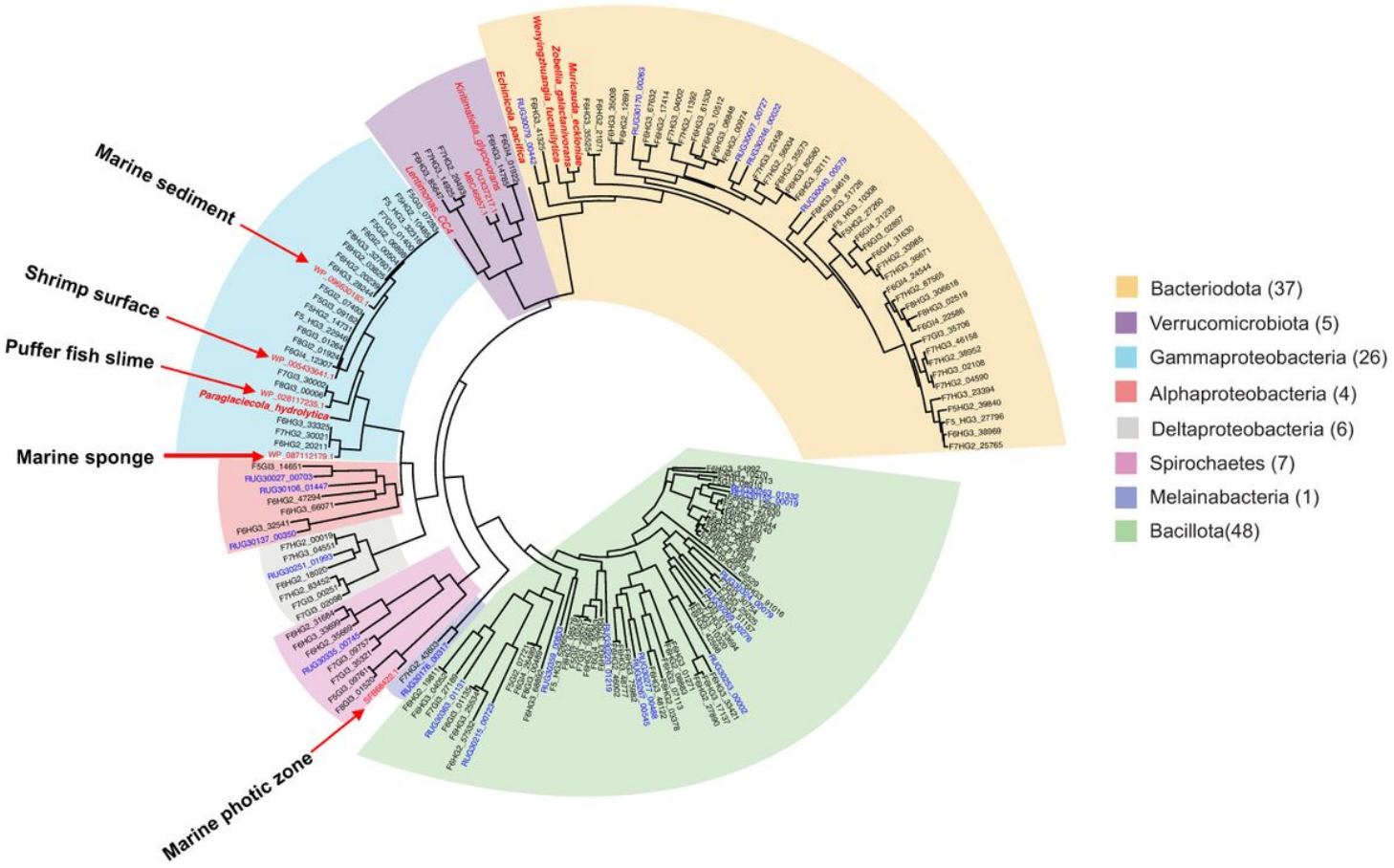

Figure 2

Single-copy rpoB gene bacterial taxonomy.

Predicted protein sequences from this study are shown in black, identified by fish number, gut compartment, and prokka annotation number. Sequences in red are from Genbank nr, with species known to digest sulfated macroalgal polysaccharides highlighted in bold. Blue text indicates predicted rpoB sequences from terrestrial ruminant MAGs (not currently included in Genbank as protein sequences) labeled with genome identifier codes (RUGXXXXX) from (Glendinning et al.), followed by prokka annotation numbers from this study. Additional information about these sequences is provided in Additional File 2.

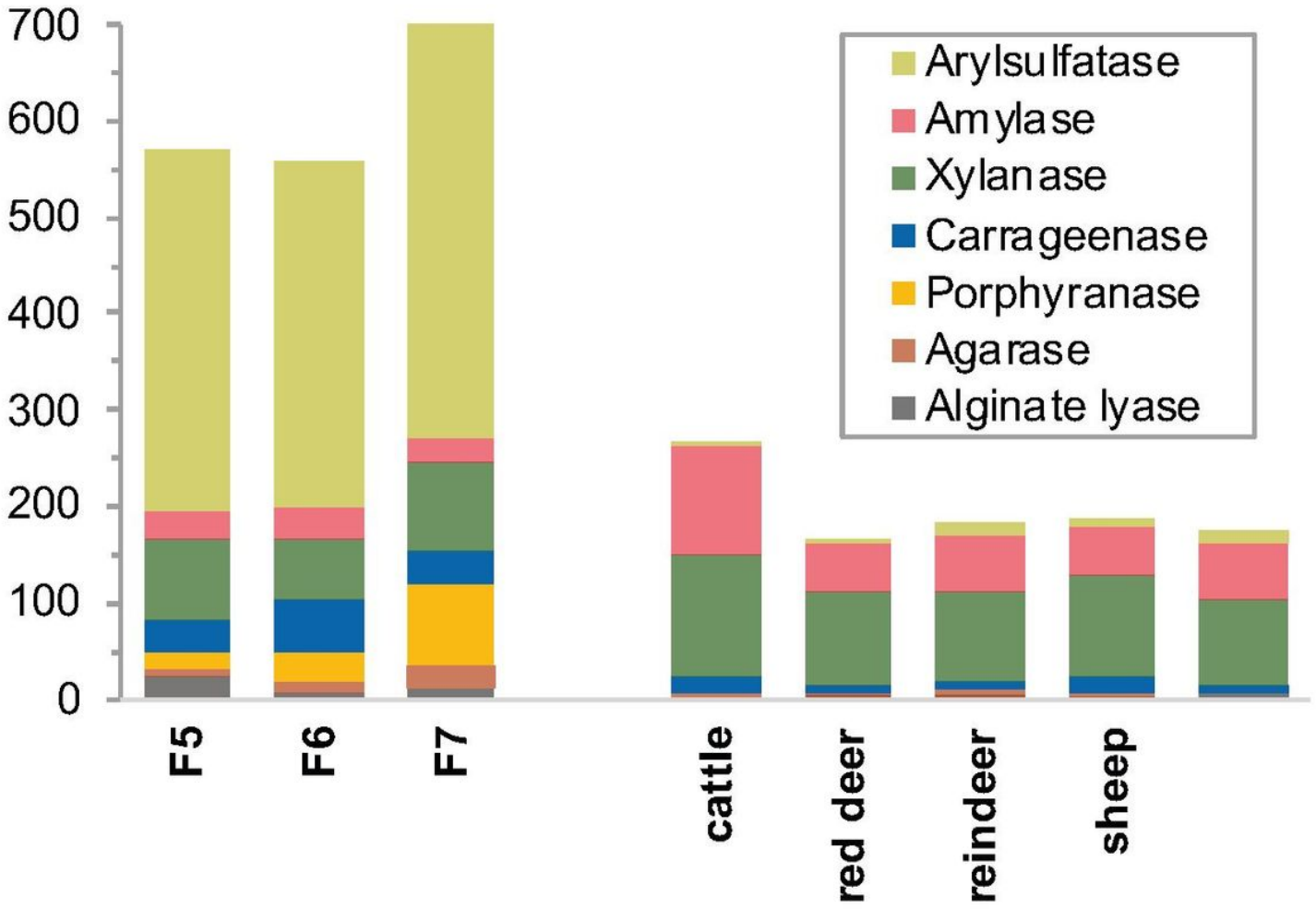

Figure 3

Polysaccharide-degrading activities annotated in kyphosid fish versus terrestrial ruminant metagenomes.

Keyword abundances have been normalized for total number of predicted proteins from each animal species. Abbreviations for adult fish samples (F5, F6, F7) are defined in Table 1. Full metagenomic annotation data and statistical calculations are provided in Additional Files 5-10.

A

predicted substrate type	CAZy Class ID	avg norm F567	avg norm rum	t-test p-value	predicted activitiy	pct fish in HG compart-ment	pct fish matches w/export signal	pct rum matches w/export signal
red	GH117	553	93	0.021	neoagarobiase	93%	70%	31%
red	GH50	114	121	0.482	agarase	83%	50%	3%
red	GH16_14	75	23	0.027	agarase	97%	88%	64%
red	GH16_11	385	4	0.012	porphyranase	96%	87%	67%
red	GH86	181	0	0.011	porphyranase	92%	74%	na
red	GH16_12	51	18	0.060	porphyranase	98%	80%	14%
red	GH82	125	8	0.002	iota-carrageenase	86%	92%	0%
red	GH150	121	2	0.003	lambda-carrageenase	85%	53%	50%
brown	GH29	747	512	0.272	fucosidase	94%	68%	29%
brown	GH95	126	275	0.047	fucosidase	94%	26%	31%
brown	GH141	230	45	0.027	fucosidase / xylanase	95%	77%	17%
brown	GH168	18	0	0.007	fucanase	94%	72%	na
brown	PL38	116	2	0.000	alginate lyase	54%	64%	50%
brown	PL6_1	116	8	0.002	alginate lyase / chondroitinase	82%	76%	38%
brown	GH165	40	0	0.000	alginate lyase	90%	73%	na
brown/green	PL40	205	1	0.011	exo-alginate lyase / ulvan lyase	95%	66%	100%
non-algal	GH88	250	37	0.005	chondroitin disaccharide hydrolase	91%	64%	30%
non-algal	PL8_1	68	0	0.000	chondroitinase	93%	71%	na
non-algal	PL30	53	7	0.028	chondroitinase	98%	76%	33%
non-algal	PL29	41	10	0.004	chondroitin-sulfate ABC endolyase	94%	85%	11%
non-algal	PL13	27	0	0.002	heparin lyase	95%	85%	na
ambig	GH10	281	114	0.008	xylanase	88%	71%	47%
ambig	GH110	251	2	0.000	α-galactosidase	87%	64%	50%
ambig	GH39	146	37	0.005	xylosidase	90%	62%	19%
ambig	GH136	155	9	0.010	lacto-N-biosidase	90%	82%	38%
ambig	PL9_2	68	1	0.000	exopolygalacturonate lyase	86%	61%	100%

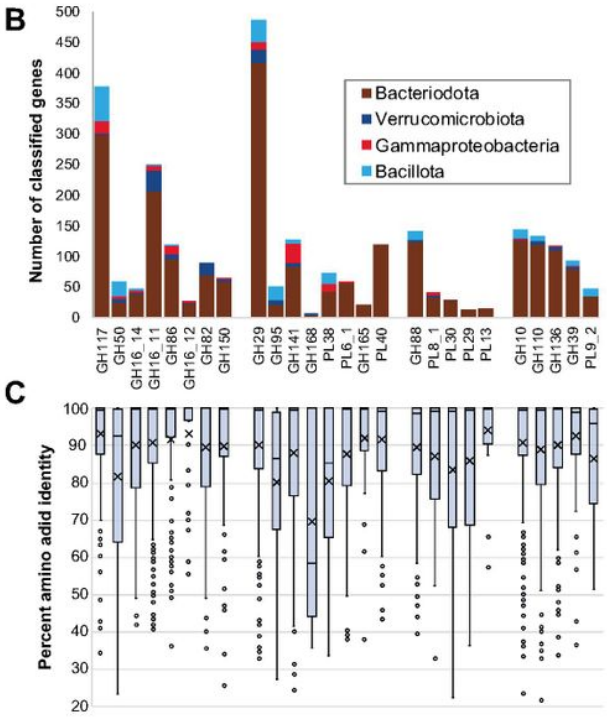

Figure 4

Polysaccharide hydrolase class enrichment.

A) Text colors denote predicted substrates based on CAZy annotations: red, red algae; brown, brown algae; purple, non-algae; grey, ambiguous. Abbreviations: avg norm, group average normalized per 1 million predicted proteins; F567, adult kyphosid fish; rum, terrestrial ruminant MAGs; HG, hindgut. Full data and significance calculations are provided in Additional File 12. B) taxonomic distribution of enriched classes based on top blast matches to database relatives (note that not all candidates had database matches). C) Intra-CAZy class diversity within fish gut metagenomes. Top of box = 75% percentile boundary, Bottom of box = 25% percentile boundary; Horizontal line = median; X = mean value; whisker = standard deviation; dots = outlier points.

A

SulfAtlas class	avg norm F567	avg norm rum	t-test p-value	pct fish matches in HG compart-ment	pct fish matches w/export signal	pct rum matches w/export signal
S1_16	510	253	0.037	89%	81%	9%
S1_15	465	67	0.001	91%	81%	11%
S1_17	455	27	0.001	91%	72%	21%
S1_8	360	140	0.003	92%	78%	15%
S1_19	259	1	0.003	91%	81%	11%
S1_11	223	246	0.418	94%	68%	28%
S1_14	222	181	0.315	93%	77%	17%
S1_20	209	69	0.000	87%	81%	7%
S1_28	164	35	0.028	92%	78%	15%
S1_25	134	36	0.012	94%	71%	24%
S1_4	133	35	0.010	90%	72%	20%
S1_24	131	6	0.008	85%	81%	5%
S1_23	114	2	0.003	93%	85%	9%
S1_72	111	0	0.011	91%	76%	na
S1_7	95	416	0.043	91%	68%	25%
S1_29	72	5	0.015	86%	84%	2%
S1_27	64	23	0.029	94%	43%	54%
S1_30	57	2	0.045	78%	62%	21%
S1_62	55	0	0.006	88%	90%	na
S1_13	50	1	0.070	50%	100%	100%
S1_46	45	2	0.028	90%	17%	81%
S1_N.C.	44	0	0.000	91%	45%	na
S1_51	43	0	0.000	85%	75%	na
S1_22	32	433	0.016	96%	87%	9%

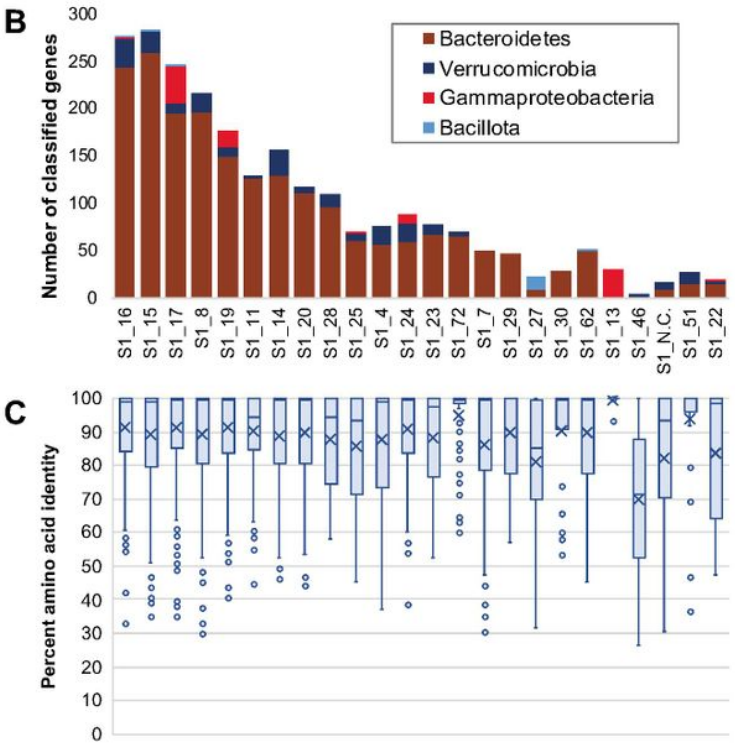

Figure 5

SulfAtlas enzyme class enrichment.

Comparisons of most abundant classes in adult fish metagenomes with terrestrial ruminant MAGs. Full data and significance calculations are provided in Additional File 12. Abbreviations: avg norm, group average normalized per 1 million predicted proteins; F567, adult kyphosid fish; rum, terrestrial ruminant MAGs; HG, hindgut. B) Taxonomic assignments of fish sulfatase class genes matching Genbank nr sequences with minimum BLASTP e-value scores < 1e-5 and alignments covering > 30% of database protein length. C) Intra-SulfAtlas class diversity within fish gut metagenomes. Box-whisker plot elements defined in Figure 4C.

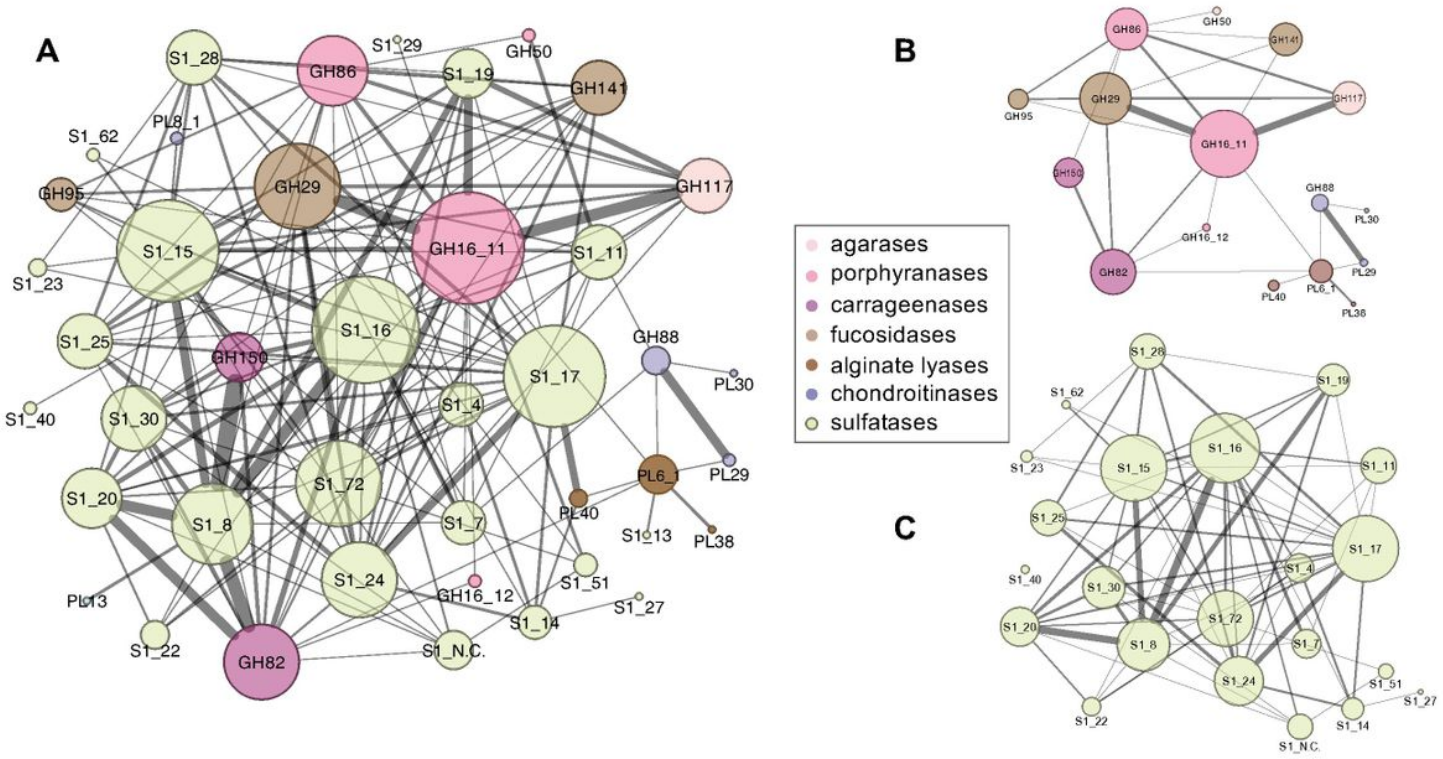

Figure 6

SulfAtlas and CAZy class co-localization networks.

Fish-enriched enzyme classes are shown as nodes, with edges denoting co-localization on the same contig within an intervening distance of 25 or fewer genes. Node size is proportional to total number of edges, and edge thickness to connection frequency. Self-loops have been omitted to simplify visualization. A) Full network diagram for classes described in Figures 4 and 5. B) Sub-network including CAZy classes only. C) Sub-network including SulfAtlas classes only. Raw data tables are provided in Additional file 15.

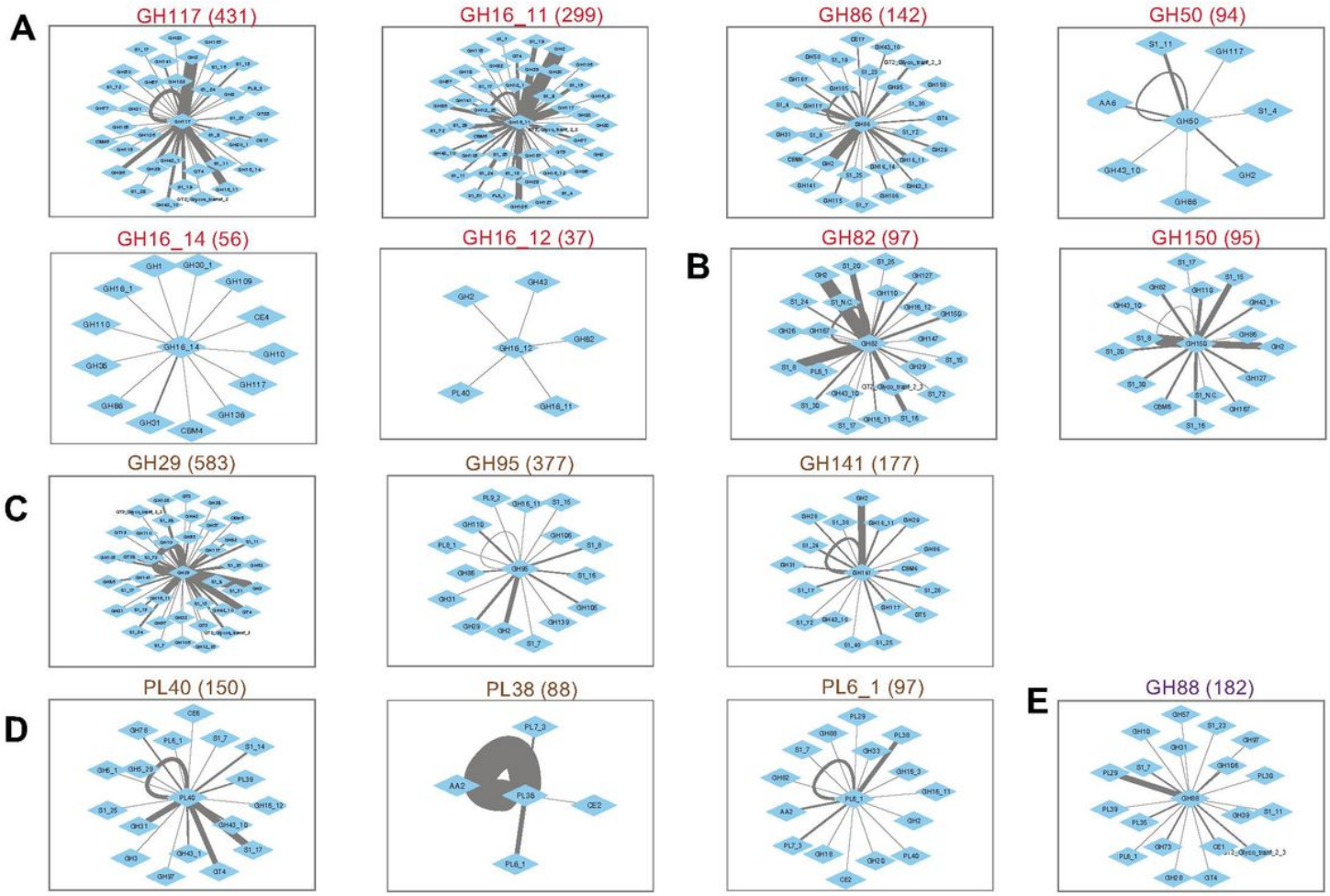

Figure 7

Co-localization network diagrams for individual fish-enriched CAZy classes.

A) Agarases and porphyranases; B) carrageenases C) fucosidases; D) alginate lyases; E) chondroitinases. Maximum intervening distance between nodes is is 25 genes. Numbers in parentheses indicate relative gene frequencies per 1 million proteins in adult fish metagenomes. Text colors denote predicted macroalgal substrate sources, as defined in Figure 4.

Figure 8

Co-localization network diagrams for individual fish-enriched sulfatase classes.

Numbers in parentheses indicate relative gene frequencies per 1 million proteins in adult fish metagenomes. Maximum intervening distance between nodes is 25 genes.

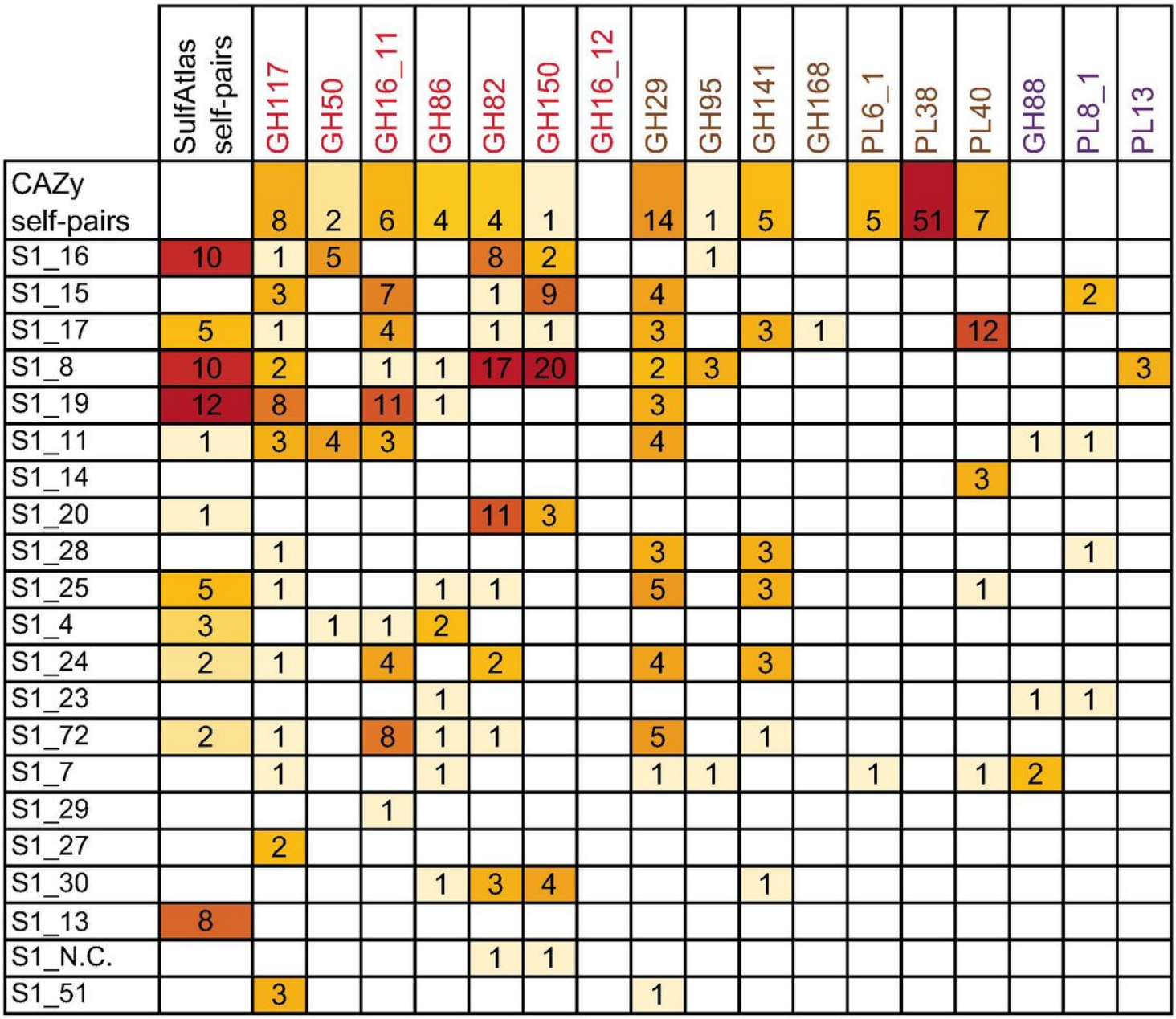

Figure 9

Compartmental Abundance of Microbial Taxa.

Text colors denote predicted macroalgal sources, as defined in Figure 4. Darker colors indicate higher values for the number of times each intersecting SulfAtlas-CAZy class pair occurred on the same metagenomic contig with an intervening distance of 25 genes or less.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Addfile01assemblystats.xlsx

- Addfile02rpoBblastmatches.xlsx
- Addfile03euke18Schart.pdf
- Addfile04euke18Sblastmatches.xlsx
- Addfile05compannotkeywords.xlsx
- Addfile06Fish5prokkaCAZySulfSignalP.csv.gz
- Addfile07Fish6prokkaCAZySulfSignalP.csv.gz
- Addfile08Fish7prokkaCAZySulfSignalP.csv.gz
- Addfile09Fish8prokkaCAZySulfSignalP.csv.gz
- Addfile10ruminantprokkaCAZySulf.csv.gz
- Addfile11CHOdegradenzcomparts.pdf
- Addfile12compCAZySulfcalcs.xlsx
- Addfile13glucanaseCAZyclasses.xlsx
- Addfile14separationdist.pdf
- Addfile15colocationstats.xlsx
- Addfile16rumnetworkcomposite.pdf
- AdditionalFilesLegends.docx