



OPEN ACCESS

EDITED BY

Reeta Goel,
G. B. Pant University of Agriculture and
Technology, India

REVIEWED BY

Kuan-ting Hsin,
National Taiwan University, Taiwan
Yanfeng Wang,
University of Groningen, Netherlands
Lina Baz,
King Abdulaziz University, Saudi Arabia

*CORRESPONDENCE

Gaurav Sharma
✉ sharmaG@bt.iith.ac.in

RECEIVED 03 January 2025

ACCEPTED 19 March 2025

PUBLISHED 07 May 2025

CITATION

Saraf N and Sharma G (2025) Comparative
genomic insight into the myxobacterial
carbohydrate-degrading potential.
Front. Microbiol. 16:1550287.
doi: 10.3389/fmicb.2025.1550287

COPYRIGHT

© 2025 Saraf and Sharma. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Comparative genomic insight into the myxobacterial carbohydrate-degrading potential

Niharika Saraf and Gaurav Sharma*

Department of Biotechnology, Indian Institute of Technology Hyderabad, Sangareddy, Telangana, India

Myxobacteria are an intriguing group of social-behavior-depicting microbes with unique physiological characteristics such as fruiting body formation, gliding motility, and predation, encompassing the largest genomes (>9 Mb) within the Eubacteria kingdom. These soil-dwelling organisms are crucial for lignocellulosic biomass degradation, which has both ecological and industrial significance. While previous studies have demonstrated polysaccharide deconstruction abilities in a few myxobacterial species, we aim to elucidate the distribution of their Carbohydrate Active Enzymes (CAZymes) domains per organism, with a focus on proteins involved in the catabolism of critical polysaccharides such as cellulose, lignin, xylan, starch, pectin, fructan, chitin, and dextran, across 61 high-quality sequenced myxobacterial genomes. Our findings reveal that 3.5% of the total genes at the median level have domains related to CAZyme functions across different myxobacterial families. Notably, family Archangiaceae (4.4%) and Myxococcaceae (3.7%) members exhibit the most significant genomic diversity and potential for degrading multiple substrates within lignocellulosic biomass. These plentiful CAZymes probably enable these majorly soil-harboring myxobacteria to breakdown various carbohydrate substrates into simpler biological molecules, which not only allow these organisms to sustain in poor-nutrient environments but also enable them to be critical players in carbon cycling and organic matter decomposition. Overall, it can be postulated that myxobacteria have an unexplored genomic potential that may play an integral role in the degradation of recalcitrant plant biomass, potentially influencing soil health and composition. This study further suggests the critical ecological importance of these CAZymes in sustaining the balance of terrestrial ecosystems and diverse industrial applications.

KEYWORDS

myxobacteria, comparative genomics, carbohydrate active enzymes (CAZyme), polysaccharide degradation, soil-dwelling bacteria, lignocellulosic biomass

Highlights

- Polysaccharides are the most abundant polymers, making up the Earth's biomass.
- Polysaccharide degradation is well-known to be carried out by diverse microorganisms; however, there is more to be explored concerning the novel organisms that can degrade these biomolecules efficiently along with understanding the reactions carried out in this process.
- Soil-dwelling myxobacteria, model organisms for our study, are unique and understudied social-behavior-depicting microbes.
- This research investigated the genomic potential of the 61 myxobacterial organisms to encode carbohydrate-active enzymes that break down various substrates, including lignocellulosic biomass, predominantly in their habitat.

Introduction

Carbohydrates, primarily polysaccharides such as cellulose, starch, xylan, chitin, lignin, pectin, fructan, dextran, etc., (Berlemont and Martiny, 2015) are the most abundant naturally occurring polymeric biomolecules with numerous imperative biological, pharmacological, and industrial applications (Mohammed et al., 2021) such as textile processing, detergent production, bleaching of paper and prebiotic production (Chettri et al., 2020). As per recent studies, the most abundant biologically renewable resource on earth is the lignocellulosic biomass (LCB; Yousuf et al., 2020), which is primarily composed of cellulose, hemicellulose, and lignin (Cortes-Tolalpa et al., 2017). LCB captures attention as it can serve as an alternative to fossil fuels and pave the way to become the source of biofuels, to meet the ever-increasing demand for energy (Zoghalmi and Paës, 2019; Inyang et al., 2022). Other important polysaccharides are xylan, a component of hemicellulose (Bajpai, 2009); chitin, which is abundant in the shells of arthropods like crabs, shrimps, and insects, and produced by bacteria and fungi (Mohammed et al., 2021); pectin, mostly found in plant cell walls and peels of citrus fruits (Mudgil, 2017); fructans, which are reserve carbohydrates (Stick and Williams, 2009) and dextrans, which are microbial water-soluble exopolysaccharides (Silva et al., 2017). The utilization of LCB or other polysaccharides requires physiochemical processing and degradation (Ezeilo et al., 2017; Devi et al., 2022) by diverse Carbohydrate-Active enZymes (CAZymes; Ezeilo et al., 2017; Zerva et al., 2021). Based on their mechanisms of action, these CAZymes can be classified into three categories, i.e., (1) enzymes involved in their assembly or glycosyltransferases (GT), and (2) enzymes involved in their breakdown—glycoside hydrolases (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), and (3) accessory functions that contain auxiliary activities (AAs) and carbohydrate-binding modules (CBMs). GHs carry out the hydrolysis of the glycosidic linkage, GTs help in the making of glycosidic bonds, PLs cause the breakdown of glycosidic bonds in a non-hydrolytic manner, CEs hydrolyse the carbohydrate esters, AAs are redox enzymes that play a role in association with other CAZymes, and CBMs help in adhesion to the carbohydrate molecule upon which the enzymes act (Lombard et al., 2014). Therefore, the understanding of enzymatic reactions behind these processes can help in improving strategies to cope with the recalcitrant nature of some polysaccharides toward degradation (Ezeilo et al., 2017), further helping in utilizing their potential more efficiently.

Using a diversity of enzymes specific to each carbohydrate substrate, microorganisms degrade miscellaneous carbohydrates in their ecological niches such as the soil (Yeager et al., 2017), oceans (Sichert and Cordero, 2021), or the human gut (Flint et al., 2012) and maintain equilibrium in those ecosystems. Based on the abundance of these enzymes, LCB decomposition potential has been reported for bacteria, fungi, as well as protozoa (Harindintwali et al., 2022). Recent metagenomic studies of rhizosphere microorganisms have shown a plenitude of carbohydrate-active enzymes in Actinobacteria, Proteobacteria, Bacteroidetes, Verrucomicrobia, Gemmatimonadetes and Firmicutes (Tashkandi and Baz, 2023). Bacteroidetes members have also been reported to encode a

lot of carbohydrate-degrading genes, playing a critical role in carbohydrate digestion in our gut (Flint et al., 2012). The cellulose-degrading potential of *Pseudomonas*, *Streptococcus*, and *Bacillus* species has been widely utilized with several applications in the paper, food, and biofuel industries (Zhang and Dong, 2022). However, myxobacteria, which are a notable group of large-genome size non-pathogenic soil-dwelling social microbes, have not been investigated extensively.

Myxobacteria belong to the order Myxococcales within the phylum Proteobacteria, however, recently, the International Committee on Systematics of Prokaryotes (ICSP) classified them under a new phylum, Myxococcota (Oren and Garrity, 2021; Panda et al., 2022). They are a fascinating group of prokaryotes depicting complex social behavior in the form of fruiting body formation, social motility or swarming, biofilm formation, and predation (Muñoz-Dorado et al., 2016; Zusman et al., 2007; Kaiser et al., 2010). Most of these organisms are aerobic with gigantic genomes ranging from 9 to 16 Mbps, with a few exceptions (Mohr, 2018). Myxobacteria are typically soil bacteria that can thrive in almost any kind of environment, be it temperate regions, tropical rain forests, arctic tundra, deserts, or marine/saline environments. They are known as ‘wolf-pack’ predators and can prey upon bacteria (both, Gram-positive and Gram-negative bacteria) and fungi through toxic secretions into the extracellular environment (Marshall and Whitworth, 2019). An industrially and clinically pertinent characteristic of myxobacteria is that they have a remarkable potential to produce a plethora of secondary metabolites showing a broad spectrum of biological activities such as anti-bacterial, anti-fungal, immunosuppressive effects, and insecticidal, anti-malarial, and herbicidal activities (Reichenbach, 2001; de Frias et al., 2018). Overall, these diverse aspects make myxobacteria quite extraordinary and distinct from other bacteria and, therefore, intriguing to study.

Order Myxococcales have been classified into three broad suborders, nine families, 20 genera, and > 200 species (Figure 1; Supplementary Table S1; Mohr, 2018). Most of the members are soil microbes, with a few being marine; however, only some organisms are well studied for their carbohydrate degradation activities, mainly in the context of predation, where these enzymes help in the breakdown of the cell wall of prey bacteria (Zhou et al., 2021; Li Y. et al., 2022; Zhang et al., 2023). *Sorangium cellulosum* of the suborder Sorangiineae decomposes cellulose and crystalline cellulose. Most of its strains can also degrade starch, xylan, and chitin (Li Y. et al., 2022; Reichenbach, 2005). *Sandaracinus amylolyticus*, another Sorangiineae member, degrades starch and has characterized proteins with domains involved in degrading chitin, agar, and cellulose reported in its genome (Sharma et al., 2016). The Archangiaceae family has several genera among which, *Cystobacter fuscus* and *Stigmatella aurantiaca* are well known to degrade chitin whereas *Archangium gephyra* does not degrade chitin (Sharma and Subramanian, 2017). Chitin degrading activity has also been reported in *Corallococcus* sp. *EGB* (Zhao et al., 2021) and *Myxococcus fulvus* (Shahbaz and Yu, 2020), both of which belong to the suborder Cystobacterineae.

Although the carbohydrate degrading potential of a few myxobacterial species has been previously studied, this study aims to perform a high-throughput pan-phylum genomic study to explore the distribution, architecture, and functions of these carbohydrate degrading enzymes (CAZymes). This study investigated 61 myxobacterial genomes (majorly complete and a few draft assemblies) and analyzed their CAZymes to untap their polysaccharide-degradation genomic potential. The study highlights varied CAZymes with diverse

Abbreviations: CAZy, Carbohydrate Active Enzymes; LCB, lignocellulosic biomass; GH, glycoside hydrolases; GT, glycosyl transferases; PL, polysaccharide lyases; CE, carbohydrate esterases; AA, auxiliary activities; CBM, carbohydrate-binding modules; NCBI, National Center for Biotechnology Information; CDS, Coding Sequence; GC%, Guanine-Cytosine percentage.

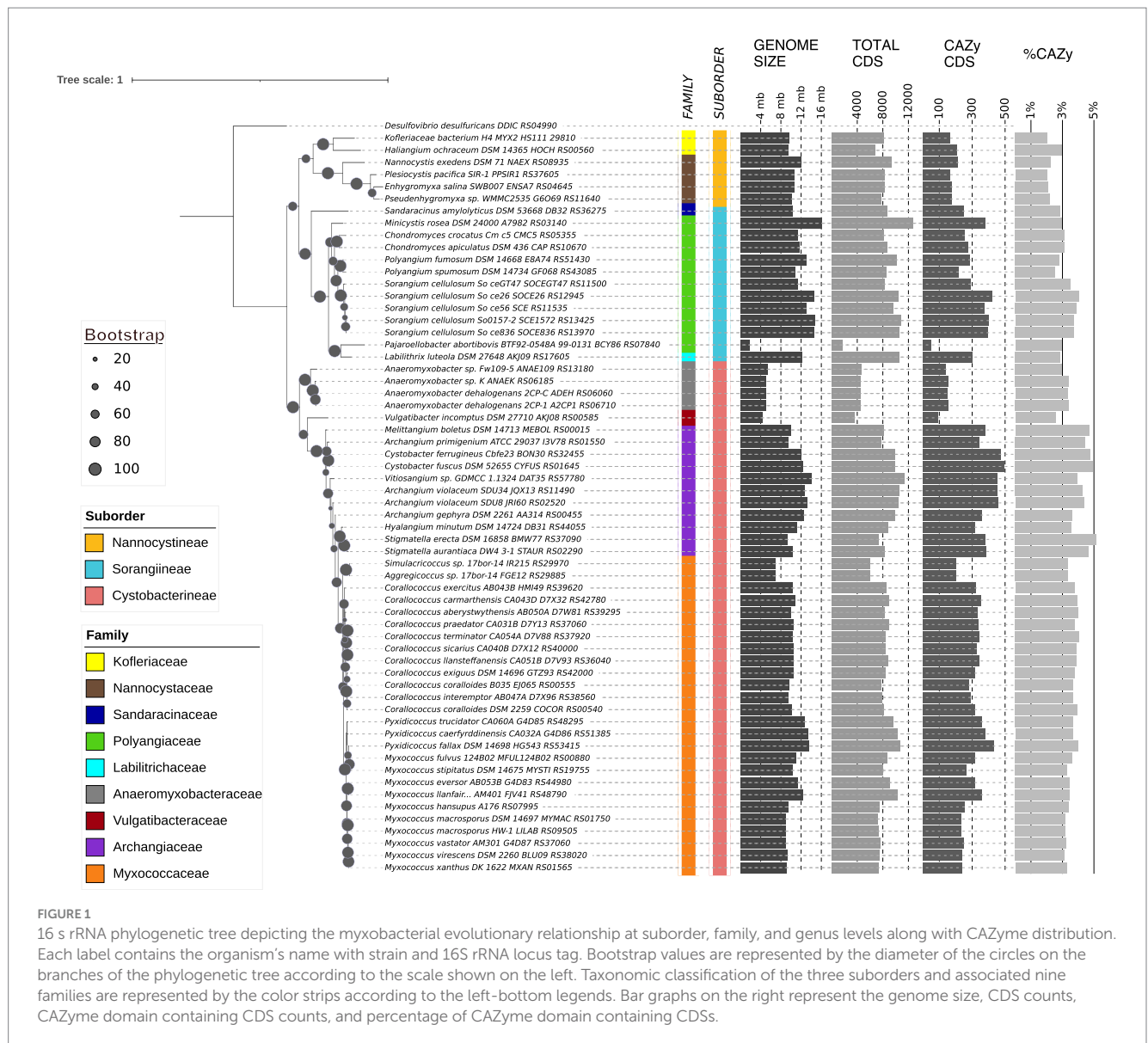


FIGURE 1

16 s rRNA phylogenetic tree depicting the myxobacterial evolutionary relationship at suborder, family, and genus levels along with CAZyme distribution. Each label contains the organism's name with strain and 16S rRNA locus tag. Bootstrap values are represented by the diameter of the circles on the branches of the phylogenetic tree according to the scale shown on the left. Taxonomic classification of the three suborders and associated nine families are represented by the color strips according to the left-bottom legends. Bar graphs on the right represent the genome size, CDS counts, CAZyme domain containing CDS counts, and percentage of CAZyme domain containing CDSs.

domain architecture among myxobacterial families, challenging assumptions about their genome size and CAZyme density correlation.

CDSs having at least one CAZyme domain divided by the total CDSs in that genome.

Methods

Selection of organisms

Sixty-one reference and representative myxobacterial organisms were selected from NCBI RefSeq based on representation of each suborder and family level taxonomy and best possible chromosome-level assembly (draft assemblies were selected only when chromosomal level assembly was not available). Based on assembly and annotation files, a genome statistics table is created to showcase their genomic features such as taxonomy, assembly, NCBI genome ID, genome size, GC%, contigs, CDS number, %CAZyme CDS as identified by custom shell scripts (Figure 1; Supplementary Table S1). %CAZyme CDS were calculated per genome from the number of

Phylogenetic analysis using 16 s rRNA sequences

16 s rRNA sequences of all myxobacterial organisms along with outgroup, *Desulfovibrio desulfuricans*, were extracted and multiple sequence alignment was performed using MUSCLE v3.8.1551. The GTR + GAMMA+I model of RAXML v8.2.12 was used to build the phylogeny tree with 100 bootstraps. Visualization of the resultant tree was done using iTOL v6 (Letunic and Bork, 2021),¹ along with suborder and family information mapping.

¹ <https://itol.embl.de/>

Domain-based homology detection and domain architecture construction

CAZyme domains were identified from the myxobacterial protein sequences by performing hmmscan (Finn et al., 2011; HMMER 3.3.2 installed in Nov 2020) against the dbCAN-HMMdb-V9 profile database downloaded from <http://www.cazy.org/>, using default parameters in domtblout output format. The output was parsed using a shell script to sort the output as per overlapped/redundant domains by retaining high-quality matches and calculating the covered fraction of hmm profiles. The parsed and sorted data was represented in the form of 'domain architecture' wherein the domains were arranged in order of their position in the protein sequence using a custom R script.

CAZyme distribution

The counts of proteins per organism containing domains specific to each CAZyme category were calculated to obtain the distribution values from the architecture. The distribution was obtained at two levels, i.e., CAZyme broad category level and CAZyme family/subfamily level. Custom shell and Python scripts were written to perform these tasks.

Curation of specific CAZyme domains, their architecture and distribution

The different types of CAZymes and their EC numbers were identified through literature search (Joshi et al., 2021; Hamid et al., 2013; Kumar and Chandra, 2020; Li et al., 2021; Krivorotova and Sereikaite, 2014; Khalikova et al., 2005; Ng and Randy, 2011). The list of CAZyme domains associated with each enzyme was identified by searching respective "EC Number" in CAZydb. The protein modular architecture specific to those domains was extracted using custom shell scripts. The enzyme-specific domains were also explored to showcase their occurrence in an alone/ duplicate/ combination manner with similar or different domains.

Correlation and regression analysis

Correlation analyses were carried out using the `lm` function in R and the plots for these analyses were made using `ggplot2`. We performed correlation analysis between genome size and the total number of CAZyme domains per organism and similarly between genome size and distribution of each of the six CAZyme categories individually.

Principal component analysis

The principal component analysis was carried out using the `prcomp` method and `ggplot2` package in R, further clustering using family and suborder information. The principal components used for this analysis were the counts of the six CAZyme categories-AA, GH, GT, CBM, PL, and CE per organism.

Results

Our studied dataset represents the whole myxobacterial taxonomy

Our dataset included 61 order Myxococcales organisms, which has ~70% (42/61), ~10% (6/61), and ~20% (13/61) representation in three suborders, i.e., Cystobacterineae, Nannocystineae and Sorangiineae, respectively (Figure 1; Supplementary Table S1). We selected these organisms to capture the vast diversity of myxobacteria from all taxonomic lineages with a priority of having a complete genome. As *Myxococcus* and *Coralloccoccus* spp. from family Myxococcaceae are the most studied and characterized myxobacteria (Garcia and Müller, 2014), family Myxococcaceae within the suborder Cystobacterineae have the highest representation (26/61; ~40%) in our dataset. Other predominant families, i.e., Archangiaceae and Polyangiaceae have 11 members each representing ~20% of the total dataset whereas the rest of the six families have 1–4 members each. Out of 26 unique genera, *Myxococcus* and *Coralloccoccus* have 10 and 11 members, respectively. Most of these organisms are soil-dwelling, isolated from diverse lignocellulose biomass present in different countries. However, all suborder Nannocystineae members are present in marine high-salt environments. It is also noteworthy to mention that *Myxococcus fulvus* HW-1 is the only member from the suborder Cystobacterineae that has been isolated from the marine environment (Albataineh and Stevens, 2018).

Apart from their habitat and taxonomy, most organisms in the studied dataset have a 9–16 Mb genome size (which is higher than the average bacterial genome size) suggesting their complex metabolic and regulatory implications (Whitworth et al., 2021). *Anaeromyxobacter* spp., *Vulgatibacter incomptus*, and *Pajaroellobacter abortibovis* are exceptions as they have smaller genome size (<5 Mb) and a few other physiological characteristics such as respiration, predation, etc., different from the rest of the myxobacteria (Sanford et al., 2002; Yamamoto et al., 2014). This dataset includes several largest genomes known so far including *Minicystis rosea* (16.04 Mb genome; Kellner et al., 2018; Pal et al., 2021), *Sorangium* spp. (11–14 Mb genomes; Land et al., 2015), followed by family Archangiaceae and Myxococcaceae members. We found that the myxobacterial family-wise mean genome size is almost 3–4 times greater than the previously reported average bacterial genome size, ~3.87 Mb (diCenzo and Finan, 2017). The highest count of total coding sequences (CDS) is present in *M. rosea* and the lowest in *P. abortibovis*. As anticipated, encoded genes per genome show a good correlation with genome size ($R^2 = 0.96$). The GC content ranges from ~66–75% in our dataset except for *P. abortibovis* with 47.45%. Bacterial GC content can range from 16 to 75% (Lightfield et al., 2011) with *Anaeromyxobacter dehalogenans* 2CP-C at the apex having 74.9% and the other myxobacteria are located in the upper quartile of this range.

Myxobacteria harbor a diverse CAZyme repertoire across their monophyletic taxonomic lineages

First, we employed 16S rRNA sequences of all 61 myxobacterial organisms to build a maximum-likelihood phylogenetic tree to comprehend their evolutionary relationship (Figure 1). This phylogeny

was able to perfectly cluster the organisms into distinct clades based on their taxonomic classification at suborder, family, and genus levels. This phylogeny depicts two major clades differentiating into suborder Cystobacterineae members and Nannocystineae and Sorangiineae members together, further separated into their subclades. All families but Archangiaceae show a monophyletic distribution, thereby making a single clade for all members at the genus and species level. However, one of the *Archangium* species within the family Archangiaceae is present at two different branches suggesting the requisite of better phylogenetic markers for their classification.

To understand the relative distribution of CAZyme domains and associated polysaccharide degradation capabilities of order Myxococcales, we functionally characterized all myxobacterial protein sequences based on CAZyme domains using the HMMER suite's hmmscan module against the dbCAN CAZyme database. We also mapped the organism-wise CAZyme CDS distribution to the phylogeny alongside the total CDSs (Figure 1) and we saw that some species have a higher CAZyme density. Our analysis reported that myxobacteria utilize almost 2–5% of the total CDSs in carbohydrate-degrading enzyme functions. This study further revealed that members of the suborder Cystobacterineae in the dataset we curated have a slightly higher proportion of CAZyme domains in their genomes as compared to suborder Sorangiineae with median values of 3.66 and

3.11%, respectively (Figure 2A), with the lowest, 2.16%, in suborder Nannocystineae. Pairwise Welch's t-test showed that the CAZyme CDSs are significantly higher in Cystobacterineae (p value = $1.3e^{-05}$) and Sorangiineae as compared to Nannocystineae (p -value = 0.0003). However, as we have more representation from Cystobacterineae members in the present dataset, future studies should include more members from other suborders to confirm this correlation. The lower abundance in marine organisms can be explained by their presence in the marine salt-tolerant ecosystem as the meso and bathypelagic layers of the ocean do not have as much availability of organic material for such bacterial species to break down (Baltar et al., 2009). Family-wise proportions (Figure 2A) depict that the highest %CAZyme CDSs are present in the family Archangiaceae with a median value of 4.44% and lowest in the family Nannocystaceae with 2.16%. However, more representation is needed from the marine organisms in future analyses.

The overall genomic potential for carbohydrate metabolism can be estimated by counting the number of translated CDSs having at least one CAZyme domain from any of the six categories (AA, CBM, CE, GH, GT, PL) and we found that the highest number of CAZyme domain containing CDSs is 493/9943 (Supplementary Table S1) in the Archangiaceae member *Cystobacter fuscus* DSM 52655 which thrives in soil and herbivorous animals' dung (Reichenbach, 2005). We further observed that the five topmost organisms with the highest

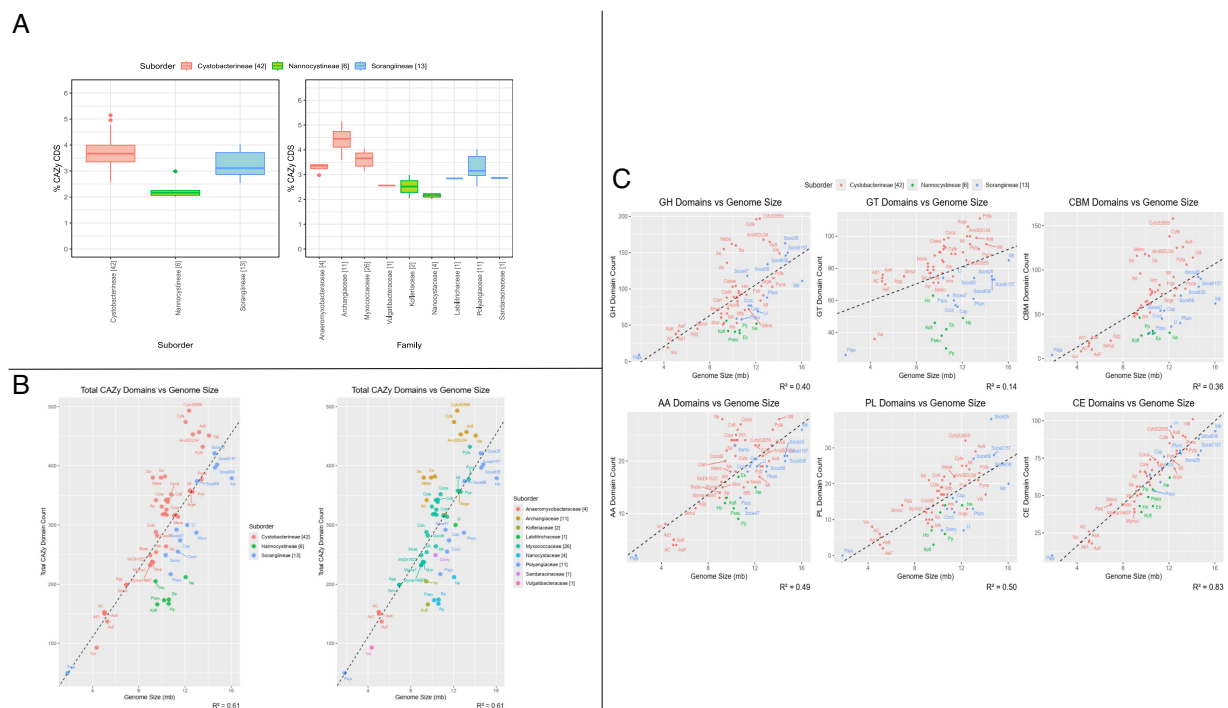


FIGURE 2

CAZymes distribution and correlation with genome size. **(A)** Box whisker plots depict %CDS containing CAZyme domains per organism, segregated at suborder level and family level. At the family level, the first four families belong to the suborder Cystobacterineae, the next two families belong to the suborder Nannocystineae, and the last three belong to the suborder Sorangiineae. **(B)** Organism-wise correlation analysis between genome size and CAZyme distribution counts (the sum of counts of all 6 CAZyme categories—GH, GT, AA, PL, CE, CBM)—showing (a) suborder & (b) family-wise classification. **(C)** Organism-wise correlation analysis between genome size and domain count of each CAZyme category (AA, CBM, CE, GH, GT, PL) showing suborder-wise classification. In all panels, the numbers mentioned in brackets in the legend and the x-axis depict the number of organisms in that suborder/family. The abbreviations correspond to the short tags we assigned to each organism. The full organism name corresponding to each tag is given in Supplementary Table S1.

CAZyme potential have been identified within the family Archangiaceae itself, whereas the lowest number of CAZymes is seen in the Polyangiaceae member, *P. abortibovis*, which is a pathogenic parasite with 1.8 Mb genome size probably having a limited requirement for carbohydrate degradation, and still, it has 2.94% CAZyme representation at the gene level. The highest and lowest proportions of CAZymes are present in *Stigmatella erecta* DSM 16858 at 5.14% and *Plesiocystis pacifica* SIR-1 with 2.01% CAZyme CDSs in its genome (Figure 2A). Based on this, we hypothesize that CAZyme density is not dependent on the number of CDSs the organism has. To further confirm this, we checked the correlation between genome size and CAZyme proteins (Figure 2B), which revealed that most of the family Archangiaceae members encode more CAZymes as compared to other families. Although we see this potential in Archangiaceae through our comparative genomics approach, experimental validation is required to confirm the expression of these CDSs. Overall, owing to several positive and negative outliers, we found a suboptimal/slightly positive correlation between genome size and CAZyme counts ($R^2 \sim 0.61$). Several organisms deviate from the trendline, and this highlights that despite their relatively smaller genome size, their immense CAZyme potential is an aspect that must be explored and utilized further.

This analysis provided an organism-wise distribution for all 61 organisms for all CAZyme categories, their families, and subfamilies (Supplementary Table S2). We found that the most abundant CAZyme category is GH followed by GT, CE, CBMs, AA, and PL with median values of 84, 77, 63, 62, 19, 14 and sample standard deviation 44.73, 18.52, 20.58, 34.76, 6.45, 7.64, respectively. Archangiaceae members have approximately two times more GHs (160) encoded in their genomes as compared to Polyangiaceae (111) and Myxococcaceae (83), at the median scale. Irrespective of having the largest genome size, *Minicystis rosea*, and other *Sorangium* spp. organisms have almost 30–45% lesser number of GHs, suggesting that Archangiaceae members have the largest genomic potential in terms of carbohydrate degradation as supported by their glycosyl hydrolases encoded proteins. To our surprise, we do not see any major difference in other categories such as GT, PL, CE, and AAs at the median level suggesting their less variation across the dataset. However, CBM numbers showed a large variation of up to two times in Archangiaceae (116) as compared to Polyangiaceae (54) and Myxococcaceae (63), at the median scale, which is almost in correlation with GH numbers. Correlation studies between each CAZyme category with the genome size suggested no correlation for GTs, low correlation for GHs and CBMs (R^2 value ~ 0.38), moderate correlation (R^2 value ~ 0.50) for AA and PL, and good correlation for CEs (R^2 value 0.83; Figure 2C).

In the process of seeing an overall picture of the distribution of the major CAZyme families and subfamilies across all taxonomic families, we created a ball and dot map of all categories which have an average total of >10 proteins at all taxonomic families (Figure 3). As expected from our above findings, the Archangiaceae family has a greater relative abundance in nearly all major GH families, which are the key domains in polysaccharide degradation. The most abundant subfamilies are CE1 with acetyl xylan esterase activity (Dilokpimol et al., 2022), and other major families include GT4, GT2, GH13, CBM50, and GH23 which are most abundant in either Archangiaceae, Myxococcaceae, or Labilitrachaceae. Among these, relevant to our study are GH13 which has dual roles with α -amylase starch degrading

activity as well as glucan-1,6- α -d-glucosidase dextran degrading activity, GH23 which has endochitinase activity, and CE4 involved in acetyl xylan esterase activity.

We further wanted to analyze if CAZyme density could be used as a factor to differentiate the taxonomic clusters at the suborder and family level and for that we further utilized the distribution counts of the six CAZyme categories to perform principal component analysis (Supplementary Figure 1). The two axes, PC1 and PC2, showed 72.72 and 10.86% variability, respectively. We observed that at both levels, some clusters were overlapping, justified to some extent by exceptional cases like *P. abortibovis* which have extremely low CAZyme counts, and some members of the family Archangiaceae which have relatively high CAZyme counts in most categories.

CAZymes could play an important role in the myxobacterial lifestyle

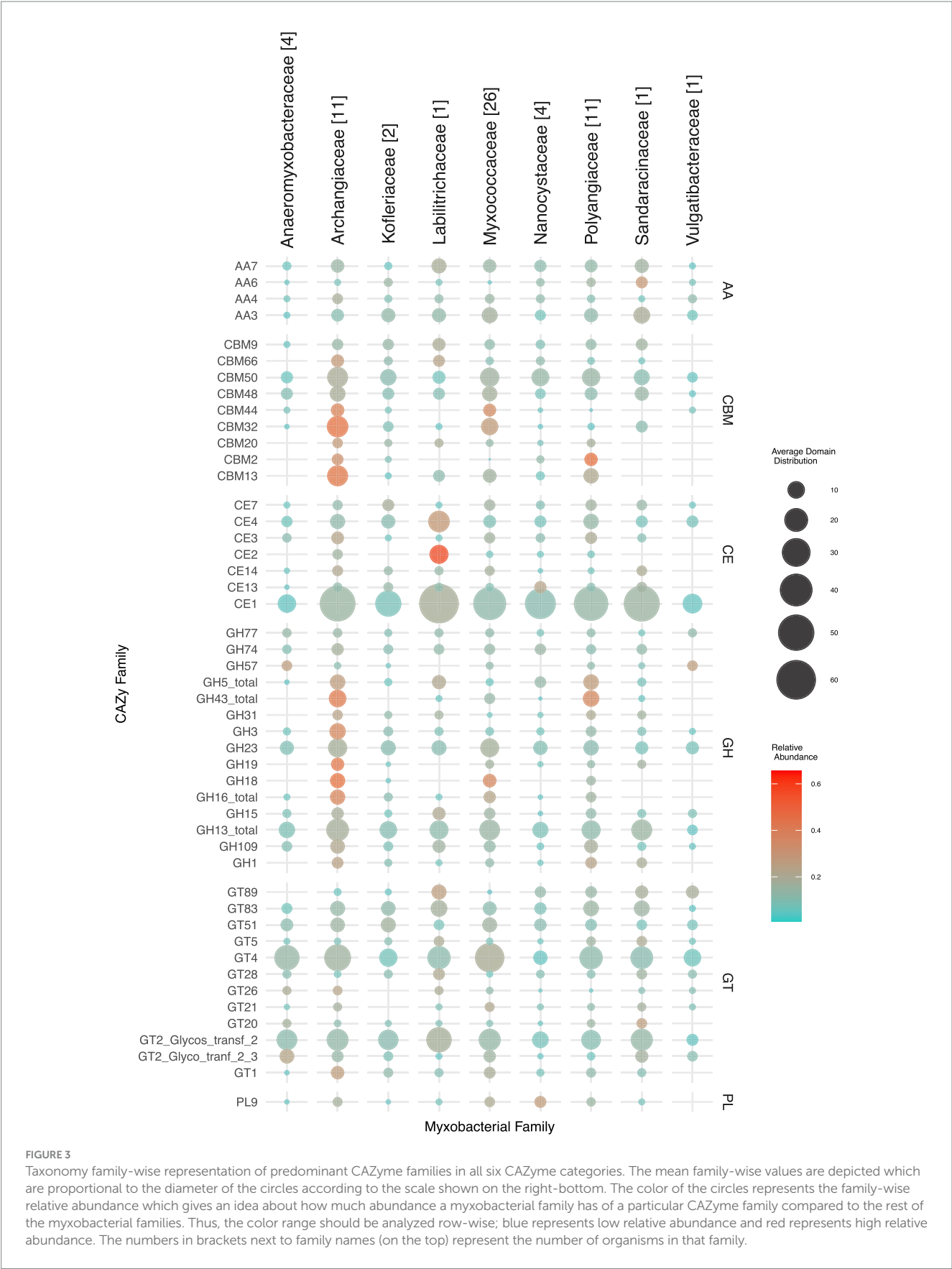
After having looked at the CAZyme domain distribution broadly, we emphasized on their functionality based on their actions against major carbohydrate substrates—cellulose, xylan, lignin, pectin, chitin, starch, dextran, and fructan (Supplementary Tables S3, S4). Being majorly soil bacteria, we anticipate that myxobacteria do have a role to play in the degradation of lignocellulosic biomass. Lignocellulosic biomass presents a complex and intricate matrix that is not readily accessible for enzymatic degradation. However, glycoside hydrolases (GHs) and related enzymes exhibit the capability to break down these structures (Ezeilo et al., 2017), a process that holds significant implications for biofuel production. Consequently, such investigations contribute to the expanding knowledge of these potentially economically valuable enzymes and their natural sources.

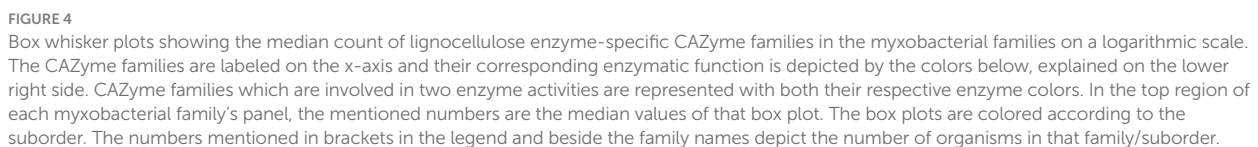
Using the enzyme-specific EC numbers (curated from literature) as a query in the CAZyme database, we first curated a list of CAZyme families that are specific to the enzymatic degradation of each of our chosen substrates (Supplementary Table S3), followed by their distribution across all myxobacteria (Figure 4). Many CAZyme families act on a single carbohydrate substrate but some are bifunctional and involved in the degradation of more than one substrate (the last few domains on the x-axis). For example, GH13 is involved in the degradation of two substrates—starch and dextran. This bifunctionality may be possible because amylopectin (a starch component) and dextran are both α -1,6 linked (Hondoh et al., 2008; Yang et al., 2022) chains and thus they both can get catabolized by the same enzymatic action, i.e., α -glycosidases (Vera-Ponce de León et al., 2020). Overall, for all enzymes again, the higher abundance is seen in Archangiaceae, Myxococcaceae, and Polyangiaceae members, which needs to be validated experimentally.

Myxobacteria are potent lignocellulosic biomass degraders

Cellulose degradation

The prime component of lignocellulosic biomass, cellulose is cleaved by different types of cellulases (See Supplementary Table S3 for





The Archangiaceae member *C. ferrugineus* Cbfe23 exhibits the highest number of cellulase-associated domain-containing proteins

The cellulase GH domains can occur alone, in duplicates, or paired with other CAZyme domains (Figure 6), which possibly means that they can perform their action independently and in association. GH6, GH9, GH48, and GH51 have both endocellulase and exocellulase (or cellobiohydrolase activity; Supplementary Table S3). Notably, GH6 is predominantly associated with CBM4, and occasionally with CBM3 and CBM37. Among these associated domains, CBM4 is implicated in the degradation of amorphous cellulose and xylan but not crystalline cellulose; CBM3 targets cellulose and, less frequently, chitin; while CBM37 has a broad polysaccharide binding potential, as documented in the CAZyme database. GH9 domains are majorly associated with CBM30, CBM4, and CMB3 which are well known to have cellulose, xylan, or chitin-binding function. It is interesting to note that the endocellulase domain GH74 domain is typically present alone in most identified proteins; however, when it co-occurs with other domains, it is linked with GH33, which has trans-sialidase activity. Additionally, GH10, GH1, GH116, GH2, GH3, GH30 (and its sub-families), GH39, GH51,

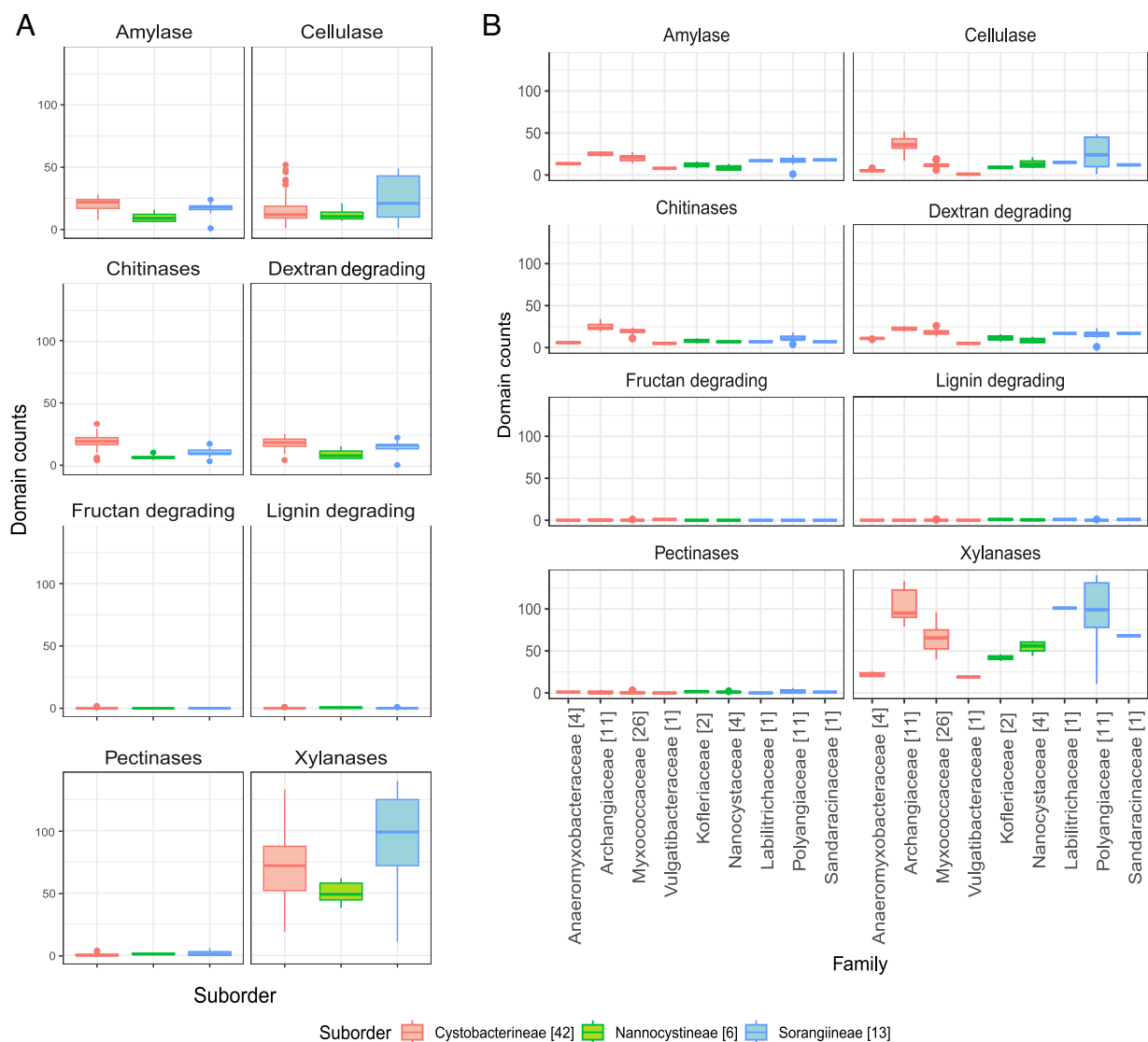


FIGURE 5
Distribution of specific lignocellulose enzymes. **(A)** Enzyme-wise box whisker plots representing the suborder level distribution of the respective CAZyme domains. **(B)** Enzyme-wise box whisker plots showing the family-level distribution of the respective CAZyme domains. The numbers mentioned in brackets in the legend depict the number of organisms in that family. The y-axis scales in all panels are kept the same for comparative visualization.

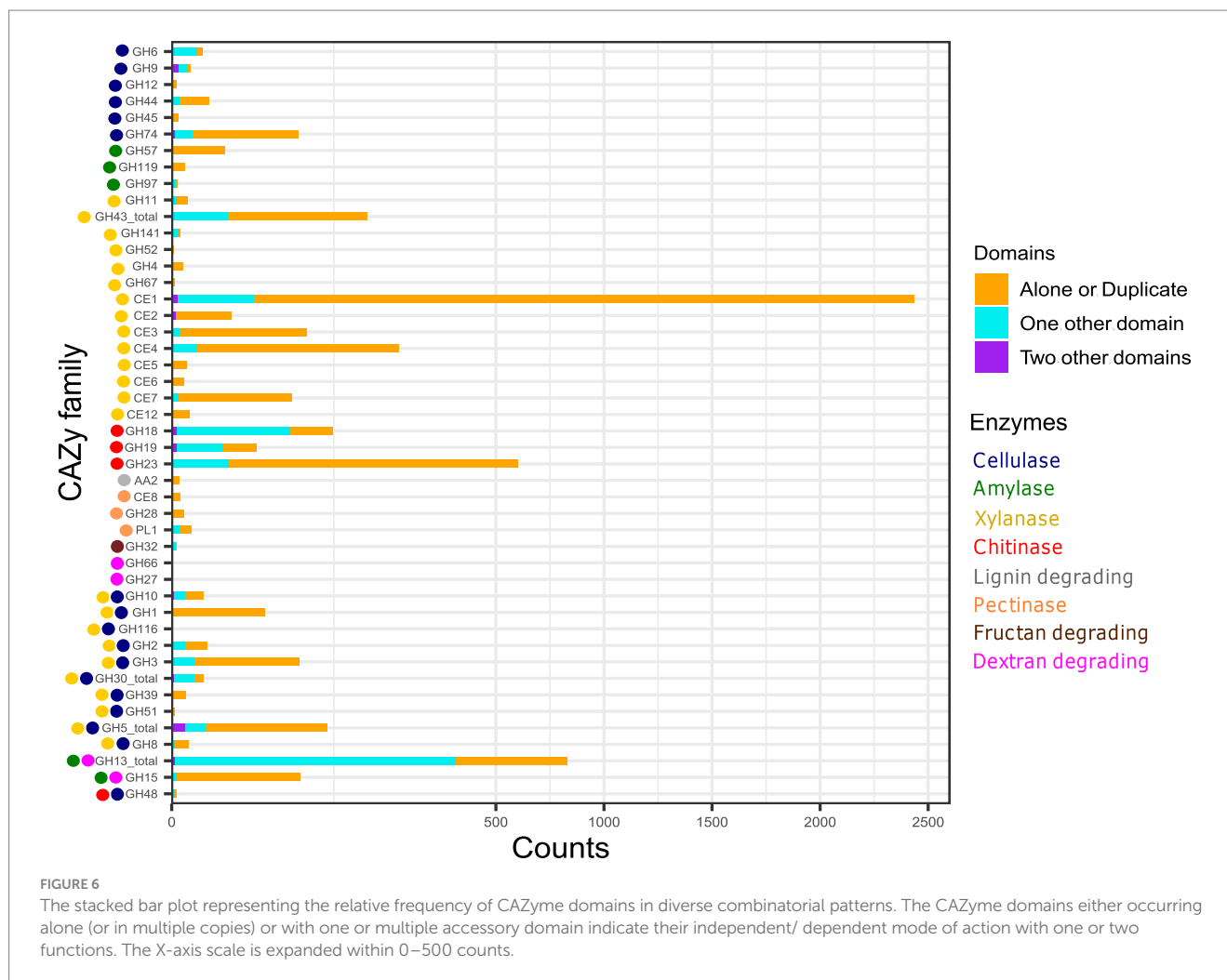
GH5 (and its sub-families), and GH8 are commonly involved in the degradation of both cellulose and xylan.

We further observed that some cellulase CAZyme families were specifically present in a few organisms or families or suborders. GH10 (former cellulase family F), GH12 (former cellulase family H), GH8 (former cellulase family D), and GH9 (former cellulase family E) were predominantly present in genus all under study *Sorangium* spp. and a few members of family Archangiaceae (Supplementary Table S3). More importantly, these members were almost absent in the most abundant family Myxococcaceae except for sporadic distribution in a few organisms. Among the cellobiases (beta-glucosidases) function, we found a unique family, GH116, which is only present in *Sorangium cellulosum* So ce56 (locus_tag = SCE_RS22635); even the taxonomic distribution in CAZyme database shows its exceptionally low occurrence in Proteobacteria phylum. Blastp analysis reveals its closest

homologs in *Acetivibrio cellulolyticus*, a phylum Bacillota member known for its efficient crystalline cellulose degradation capabilities. Overall, this further suggests the novel origin and functional importance of GH116 protein in *Sorangium cellulosum* So ce56. GH48 (former cellulase family L) and GH51 are also interesting candidates to explore further as they are only present in several family Archangiaceae members and almost absent in other myxobacteria. Overall, cellulase protein distribution indicates the huge cellulose degradation genomic potential within *Sorangium* spp. and less-explored family Archangiaceae members.

Lignin degradation

To enhance the fermentation rate of cellulose and hemicellulose, lignin must first be removed which can be performed by bacteria, fungi, bacteria-fungal co-culturing, or enzymatic pre-treatment (Li



X. et al., 2022). Our analysis indicates that auxiliary activities (AA) families AA1 and AA2 are implicated in various ligninolytic processes. Specifically, AA1, which exhibits laccase activity, is absent in all organisms examined in this study. In contrast, AA2, which is associated with lignin peroxidase, manganese peroxidase, and versatile peroxidase activities, appears infrequently and cannot be correlated with their taxonomy. Overall, lignin-degradative domains are not prominently represented in our dataset (Figures 4, 5) and in every case we saw, AA2 domain does not co-occur with any other CAZyme domain (Figure 6) which shows its independent functions here- but this would be better understood in other organisms which have more representation of AA2 domain and its architecture combinations.

Xylan degradation

Plant hemicellulose predominantly comprises xylan, a polysaccharide structurally similar to cellulose but composed of D-xylose monomers (Bajpai, 2009). Cellulose hydrolysis is facilitated by hemicellulose-degrading enzymes, a critical factor for enhancing lignocellulosic biomass degradation in industrial applications (Busk et al., 2014). *xynA*, a xylanase-encoding gene has been thoroughly studied in *Sorangium cellulosum* So9733-1 (Wang et al., 2011). Xylan degradation is mediated by xylanases, the most diverse class of enzymes in our dataset, spanning 24 different CAZyme families.

Consequently, there are more xylanase domain-containing proteins per organism compared to those targeting other polysaccharide substrates. Nearly 60% of these CAZyme families are specific to xylan deconstruction, with carbohydrate esterases (CEs) being highly represented; no other polysaccharide-degrading enzyme category in our dataset encompasses as many as eight CE families (Supplementary Table S3).

The function of feruloyl esterases and acetyl xylan esterases is primarily mediated by CE domains. Notably, family CE1, involved in both enzymatic actions, is the most prevalent domain (Figure 4), ranging from five proteins in the Sorangiineae pathogenic member *P. abortibovis* to 64 proteins in the Archangiaceae member *Vitiosangium* sp. GDMCC 1.1324. At the suborder and family levels, significant xylan-degrading potential is observed in Sorangiineae (Figure 5A), particularly within the Polyangiaceae family and Archangiaceae members (Figures 4, 5B). *S. cellulosum* So0157-2 exhibits the highest xylanase distribution with 140 proteins, followed by *S. cellulosum* So ce836 and *S. cellulosum* So ce26 with 138 and 137 proteins, respectively.

Most specific xylan-degrading domains occur independently (Figure 6), though GH43 and its subfamilies, as well as CE1, can possess accessory domains. GH43 subfamilies co-occur with several carbohydrate-binding modules (CBMs): CBM6 which interacts with

amorphous cellulose, β -1,4-xylan, and glucan; CBM66, which binds with terminal fructoside residue of fructans; and CBM13, which binds mannose in plants and xylan in bacteria. CE1 domain is occasionally accompanied by a few other CAZyme modules such as CBM20, PL22, CE7, and CBM32.

One of the β -xylosidases, GH52 CAZyme family, is uniquely distributed only within four out of five *Sorangium* spp. members. These *Sorangium* proteins show maximum homology with *Agarivorans* spp. members from Gammaproteobacteria with 98% query coverage and ~50% identity. Similarly, the GH67 domain functioning as α -glucuronidase is also present only in *Sorangium* spp. with closest relatives in Gemmatimonadaceae bacterium having >80% query coverage and >60% identity.

Myxobacteria encode a diverse set of enzymes to degrade other important polysaccharide substrates

Starch degradation

At the suborder level, Cystobacterineae members have the highest median values for amylase domain-containing proteins (Figure 5A). Previously, α -amylases, AmyM (Li et al., 2015), and CoMA (Zhou et al., 2018) have been characterized in the Cystobacterineae member *Coralloccoccus* sp. EGB. Another multifunctional amylase, AmyC, has been characterized within the same species (Wu et al., 2015). Looking at the family level, Archangiaceae members have the highest median value of 25 proteins (Figures 4, 5B). The Archangiaceae members – *C. fuscus* DSM 52655, *A. violaceum* SDU34, and *A. violaceum* SDU8 have equal numbers and the highest counts with a value of 28 proteins (Supplementary Table S4). Starch-degrading domains are also found in Myxococcaceae, with a median count value of 21 proteins. The order Sorangiineae and family Sandaracinaceae member, *S. amylolyticus* DSM 53668, is a species known to degrade starch (Sharma et al., 2016) and it has 18 proteins with starch-degrading domains. Starch degradative domains include α -amylases associated GH13, GH57, GH119, β -amylases functioning GH14, γ -amylases associated GH15, GH97. GH13 and its subfamilies as well as GH15 are bifunctional domains with both dextran and starch degrading functions (Figures 4, 6). GH13 subfamilies GH13_11 and GH13_10, were found to be associated with CBM48 in 143 and 102 proteins, respectively which makes it the most frequent accessory domain for GH13 in our dataset. Other accessory domains include CBM20, CBM32, starch binding CBM25 the α -glucans amylose, amylopectin, pullulan binding CBM41, starch binding CBM69 which is also distantly related to CBM20 and CBM48.

GH119 domain is involved in α -amylase function, and it is specifically present in all family Archangiaceae members and only one genus in the family Myxococcaceae, i.e., *Coralloccoccus* members. GH119 is closely related to the previously well-known α -amylase family, GH57. These myxobacterial GH119 homologs have close homologs in *Cohnella* and *Paenibacillus* spp. of phylum Bacillota (Firmicutes) as indicated by >95% query coverage and >60% identity suggesting their putative evolutionary relationship with the latter.

Chitin degradation

Endochitinase activities are predominantly shown by GH18, GH19, GH23, and GH48 CAZyme families. A recent study has

reported a GH19 enzyme, C25GH19A, which is chitin-induced in *Coralloccoccus silvisoli* c25j2 (Zhou et al., 2024). An anti-fungal GH18 chitinase, CcCti1, was also identified and further studied in *Coralloccoccus* sp. EGB (Li et al., 2019). Based on the distribution of these families, the suborder Cystobacterineae members demonstrate encoding potential for chitin degradation (Figure 5A). Within this suborder, members of the family Archangiaceae exhibit the highest median number of chitinase domain-containing proteins, with a median value of 23 (Figures 4, 5B). *Vitiosangium* sp. GDMCC 1.1324 possesses the highest number of chitin-degrading domains, totalling 34 (Supplementary Table S4). Among Archangiaceae, previous studies have shown that members of the genus *Cystobacter* can degrade chitin but *Archangium* *gephyra* cannot (Sharma and Subramanian, 2017; Lang et al., 2015). Following the previous study, our study shows that *A. gephyra* DSM 2261 has a lower count of chitinase-associated domain-containing proteins as compared to a higher count (19 proteins) in the *Cystobacter* members *C. fuscus* DSM 52655 and *C. ferrugineus* Cbfe23 (29 and 30 proteins respectively). Among Myxococcaceae, chitin-degrading enzyme domains have been previously reported in *Coralloccoccus* species and *Myxococcus* species (Shahbaz and Yu, 2020; Li et al., 2019). With a median value of 20 proteins, Myxococcaceae members stand second highest after Archangiaceae. Chitinase domains GH18, GH19, and GH23 showcase endochitinase activity among which the first two are always majorly associated with other domains whereas the last one majorly works independently (Figure 6). GH18 plays a role in chitobiosidase activity as well and in our data, it is majorly associated with several CBMs, such as CBM12, CBM13, CBM2, CBM44, CBM5, CBM8, CBM13, CBM61, and CBM 37. Similarly, GH19 is present with CBM12, CBM13, CBM2, and CBM50. However, GH23 is mostly independent (515 proteins) but when associated with an accessory domain, it is one or two or three replicates of CBM50 itself. This suggests that CBM50 has some additional evolutionary relatedness with GH23. GH48 acts both on cellulose as well as chitin (endochitinase activity). Our study revealed that this domain is only present in a few members of the family Archangiaceae along with two members of the family Polyangiaceae, i.e., *Minicystis rosea* DSM 24000, *Polyangium fumosum* DSM 14668.

Dextran degradation

Suborder Cystobacterineae has the highest genomic potential for dextran degradation as well (Figure 5A). Here again the most potent are the members from the underlying Archangiaceae family (Figures 4, 5B) with a median of 22 proteins having dextran degrading domains. *A. violaceum* SDU8 has 26 proteins with dextranase domains which is the maximum value in this dataset. GH66 and GH49 play a role in dextranase activity, GH13 with its subfamilies and GH15 are involved in glucan-1,6- α -d-glucosidases functioning and GH27 in glucan-1,6- α -isomaltosidases. GH49 also has a role to play in dextran 1,6- α -isomaltotriosidase activity but GH49 is not identified in any of the genomes in this study.

The GH66 domain is only present in *Anaeromyxobacter* sp. Fw109-5, which depicts the closest homologs in *Thermoanaerobacter* spp. from phylum Bacillota (Firmicutes) with >95% query coverage and ~45% identity. Both of these are facultative Fe (III)-reducing bacteria, and such homology possibly suggests the shared evolutionary relationship between these two homologs. Similarly, the GH27 domain which is involved in glucan-1,6- α -isomaltosidases function occurs only in *Sorangium cellulosum* So ce836 and not in any other

genome. Its closest homologs are present in the *Verrucomicrobiia* bacterium from phylum Verrucomicrobiota, with >70% query coverage, >40% identity, and e-values of $<1e^{-135}$, showcasing good sequence homologs.

Fructan degradation

Fructan degradation is catalyzed by fructan exohydrolases which function with the help of the GH32 domain. GH32-containing proteins are not too abundant in this myxobacterial dataset (Figures 4, 5). This domain is only present in six proteins (Figure 6), out of which it occurs independently in two proteins and the other four, with two copies of domain CBM66 which is well-known for fructan binding.

Pectin degradation

CE8, GH28, and PL1 are the domain families substantial for pectinase activity CE8 having pectin methylsterase activity, GH28 for endo-polygalacturonases, and PL1 having pectin lyase activity are the domain families substantial for pectinase activity. We could identify only a few pectinases in our dataset (Figures 4, 5), where the highest number is seen in *S. cellulorum* So ce26 with 6 domains. Pectin-degrading pathways have also been earlier reported in *S. cellulorum* (Schneiker et al., 2007). Overall, CE8 was found in only 3 proteins which are encoded by *Archangium violaceum* genomes (Figure 6) and our analysis reveals that all 3 of those proteins have varying architecture, one having CE8 in triplicates, one with the domain PL9_4, and the other with the pectinase specific PL1 itself. GH28 is found in 18 proteins overall, while mostly present alone, this domain could also be found associated with one accessory domain which is PL9_2, or two accessory domains PL9_2 and PL9_4 (Figure 6). The accessory PL9_4 plays a role in the peptidoglycan lyase activity, while no specific activity has been characterized for PL9_2 as per the CAZyme database. PL1 mostly occurs alone, or it can be with other accessory domains such as PL9, mannose/xylan binding CBM13, the rhamnose binding CBM67, and CBM37.

Overall, our data indicates that the Archangiaceae, Myxococcaceae, and Polyangiaceae families have significant carbohydrate-degrading genomic potential, warranting further exploration, experimental validation, and utilization. Conversely, the pathogenic *Pajaroellobacter* demonstrates the lowest distribution of CAZymes across all classes, which is consistent with expectations for a pathogenic organism that has reduced genome size. For all enzyme categories, the lower values of Nannocystineae could be attributed to their marine habitat (Ivanova et al., 2010; Figure 7) and thereby reduced carbohydrate degradation but more representation and experimental validation are required in future studies. This is supported by earlier reports during the initial characterization of *Nannocystis exedens*, which noted the absence of starch and cellulose degradation (Reichenbach, 1970). Collectively, these findings from the current dataset suggest that CAZyme potential is higher in soil-dwelling myxobacterial species compared to their marine counterparts, though further studies are needed to confirm this trend (Figure 7).

Discussion

Myxobacteria, ubiquitous across various habitats, exhibit their highest diversity in steppe and warm climatic zones. Notably, certain

species within the genera *Myxococcus*, *Coralloccoccus*, *Archangium*, and *Nannocystis* are also found in harsher environments such as deserts and high-altitude regions. These topsoil bacteria are frequently isolated from the dung of herbivorous animals and decaying plant material, which serve as well-known predominant sources of complex polysaccharides (Reichenbach, 1999). The polysaccharide catabolism potential of organisms using the carbohydrate-active enzymes (CAZyme) database mapped on their encoded genes can underscore their value for industrial applications. It has been previously shown that Gemmatimonadetes, Actinobacteria, and Proteobacteria members are crucial for nutrient cycling because of their high CAZyme potential (Zhou et al., 2023). *Streptomyces*, *Micromonospora*, a few *Mycobacterium* and *Propionibacterium* strains have enriched carbohydrate-active enzyme (CAZyme) domains, which have been found to help in lignocellulose biomass degradation (Lewin et al., 2016). The role of *Streptomyces* species in nutrient cycling has been elucidated in several studies (Book et al., 2016; Sidar et al., 2024; Takasuka et al., 2013; Book et al., 2014). Insect associated *Streptomyces* species with genome sizes in the range of 6.9–7.4 Mb encode 130–185 CAZyme proteins with percentage CAZyme proteins in the range of 2.3–2.9% (Book et al., 2014), which helps them in sustaining their relationship with insects. A metagenomic analysis of an anaerobic digestion microbiome exposed to acidosis revealed that Bacteroidetes MAGs constitute $8.8\% \pm 1.8\%$ of CAZyme genes whereas Firmicutes MAGs with $3.6\% \pm 1.1\%$, Spirochaetes MAGs with $2.9\% \pm 1.2\%$ (Bertucci et al., 2019) and a smaller percentage of myxobacterial MAGS was also present in such ecosystem.

In this study, we revealed the genomic potential of polysaccharide catabolism in a diverse representation of myxobacterial species by diving deep into analyzing the CAZyme domains involved in degrading the major substrates cellulose, starch, xylan, fructan, lignin, pectin, dextran, and chitin. Members of Sorangiineae, especially *Sorangium cellulosum* have been well-known as cellulose degraders as the name suggests, however, the genomic background of these activities was not much studied. Our research fills this gap and reveals that many other myxobacterial species possess considerable, yet underexplored, capacities for polysaccharide deconstruction. The degradation of polysaccharides, especially lignocellulosic biomass, holds immense potential for biofuel production. Previous research has shown that overexpression of certain plant-associated domains, like glycosyltransferases (GTs), can enhance biomass formation (Yang et al., 2017). This concept could be explored and applied to harness the converse catabolic activities of myxobacteria for increasing efficiency.

Our findings indicate substantial variability in genomic CAZyme potential among these large-genome harboring myxobacterial families and organisms, with some species exhibiting high potential for substrate-specific degradative capabilities despite having relatively small genomes. Notably, members of the soil-dwelling Archangiaceae family demonstrated a remarkable richness in CAZymes. It is important to recognize and mention that deriving biological conclusions from high-throughput studies such as ours, which involve multi-organism and multi-level comparisons, heavily relies on the quality of publicly available genome assemblies. Most of the assemblies used in our study have complete genomes with high-quality genome annotations, however, a few of them are incomplete, multi-contig assemblies that may not capture the total encoding potential. The inclusion of these draft assemblies was to avail the maximum taxonomic and genomic diversity available at the

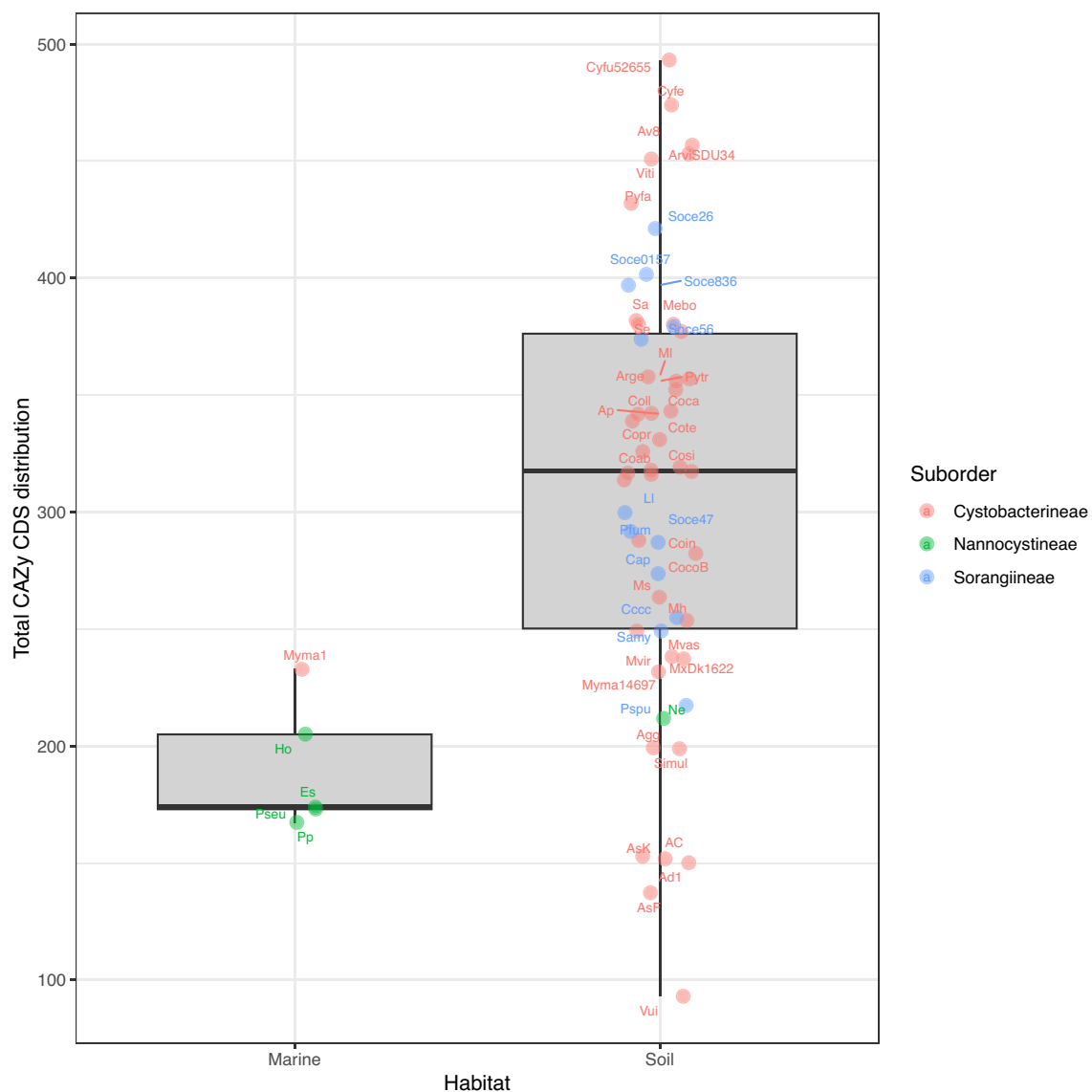


FIGURE 7

Boxplots showing the organism-wise preferred habitat and respective CAZyme domain distribution of the 61 myxobacteria in our dataset, categorized by suborder. The abbreviations correspond to the short tags, assigned to each organism. The full organism name corresponding to each tag has been provided in the [Supplementary Table S1](#).

time of data curation. Despite these limitations, our deep analysis has unveiled new avenues of knowledge.

While we performed this analysis at three levels- suborder, family, and organism, there may be cases where a bacterium/its family may have immense genomic potential but is not expressing the same when it comes to its expression in its life cycle. Genomic studies alone cannot address these limitations, necessitating extensive experimental validation. We urge experimental researchers to verify whether high counts of substrate-degrading CAZyme domains correlate with increased degradation efficiency. Nonetheless, the carbohydrate degradation potential highlighted in our study is substantial and provides a basis for designing experiments to harness this capability for practical applications. The dataset used in the study is taxonomically broad but still limited to certain well-studied myxobacterial groups. This could introduce phylogenetic bias, where

the observed patterns in CAZyme distribution might not be generalizable to other, less-studied phylogenetically close groups. We acknowledge this limitation and suggest the inclusion of more taxa in future studies to provide a more complete picture.

In conclusion, we present a comprehensive high-throughput genomic analysis of the Carbohydrate Active enZymes-encoding potential of 61 myxobacteria members (order Myxococcales or currently known phylum Myxococcota), examining their genomic diversity at the levels of suborder, family, and genus, as well as their functionality in terms of singular or multiple enzymatic activities. Comparative analysis among members of order Myxococcales revealed variations in carbohydrate utilization potential and recommended that lignocellulose biomass degradation is not universally conserved and similar within these groups of organisms. Such carbohydrate degradation analyses lay a preliminary foundation

for utilizing myxobacterial organisms in diverse paper-pulp, food-beverage, textile, and biofuel industries.

Data availability statement

Publicly available datasets were analyzed in this study. The accession numbers to these datasets are available in the “U” column of [Supplementary Table 1](#).

Author contributions

NS: Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing, Data curation, Formal analysis, Visualization. GS: Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Project administration, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. NS acknowledges the fellowship by IIT Hyderabad. GS is supported by the IIT Hyderabad seed grant and the DST-INSPIRE Faculty Grant by the Government of India.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

References

- Albataineh, H., and Stevens, D. C. (2018). Marine Myxobacteria: a few good halophiles. *Mar. Drugs* 16:209. doi: 10.3390/md16060209
- Bajpai, P. (2009). “Xylanases” in *Encyclopedia of microbiology* (third edition). ed. M. Schaechter (Oxford: Academic Press), 600–612.
- Baltar, F., Aristegui, J., Sintes, E., van Aken, H. M., Gasol, J. M., and Herndl, G. J. (2009). Prokaryotic extracellular enzymatic activity in relation to biomass production and respiration in the meso- and bathypelagic waters of the (sub)tropical Atlantic. *Environ. Microbiol.* 11, 1998–2014. doi: 10.1111/j.1462-2920.2009.01922.x
- Berlemont, R., and Martiny, A. C. (2015). Genomic potential for polysaccharide deconstruction in Bacteria. *Appl. Environ. Microbiol.* 81, 1513–1519. doi: 10.1128/AEM.03718-14
- Bertucci, M., Calusinska, M., Goux, X., Rouland-Lefèvre, C., Untereiner, B., Ferrer, P., et al. (2019). Carbohydrate hydrolytic potential and redundancy of an anaerobic digestion microbiome exposed to acidosis, as uncovered by metagenomics. *Appl. Environ. Microbiol.* 85, e00895–e00819. doi: 10.1128/AEM.00895-19
- Book, A. J., Lewin, G. R., McDonald, B. R., Takasuka, T. E., Doering, D. T., Adams, A. S., et al. (2014). Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. *Appl. Environ. Microbiol.* 80, 4692–4701. doi: 10.1128/AEM.01133-14
- Book, A. J., Lewin, G. R., McDonald, B. R., Takasuka, T. E., Wendt-Pienkowski, E., Doering, D. T., et al. (2016). Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of expanded gene content and coordinated gene expression. *PLoS Biol.* 14:e1002475. doi: 10.1371/journal.pbio.1002475
- Busk, P. K., Lange, M., Pilgaard, B., and Lange, L. (2014). Several genes encoding enzymes with the same activity are necessary for aerobic fungal degradation of cellulose in nature. *PLoS One* 9:e114138. doi: 10.1371/journal.pone.0114138
- Chettri, D., Verma, A. K., and Verma, A. K. (2020). Innovations in CAZyme gene diversity and its modification for biorefinery applications. *Biotechnology Reports*. 28:e00525. doi: 10.1016/j.btre.2020.e00525
- Cortes-Talalpa, L., Salles, J. F., and van Elsas, J. D. (2017). Bacterial synergism in lignocellulose biomass degradation – complementary roles of degraders as influenced by complexity of the carbon source. *Front. Microbiol.* 8:1628. doi: 10.3389/fmicb.2017.01628

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2025.1550287/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Principal component analysis based on the organism-wise distribution of the six CAZyme Categories—GH, GT, AA, PL, CE, CBM—showing (a) suborder & (b) family-wise clustering. The numbers mentioned in brackets in the legend and X-axis depict the number of organisms in that suborder/family.

SUPPLEMENTARY TABLE S1

Detailed genome statistics information for all 61 myxobacterial genomes showcases their taxonomy information, abbreviated tags used throughout the study, genome size, GC content, CDS information, CAZy-specific CDS density, RNA information, source file RefSeq links, and individual category and total CAZyme domain distribution counts. Columns represented as heatmaps range from red as the lowest, yellow in the middle, and green as the highest value in that column. Columns (B, C, and D) are color-coded as per their family taxonomy.

SUPPLEMENTARY TABLE S2

Distribution of identified encoded proteins per organism per CAZyme family. Distribution implies the total number of CDSs having that unique domain per protein.

SUPPLEMENTARY TABLE S3

Distribution of proteins per organism per CAZyme family per lignocellulose enzyme. Distribution implies the total number of CDSs having that unique domain per protein.

SUPPLEMENTARY TABLE S4

Organism-wise distribution of each broad lignocellulose enzyme along with organism taxonomy.

SUPPLEMENTARY TABLE S5

Details of myxobacterial strains under study and their ecological niche from where the respective strain has been isolated.

- de Frias, U. A., Pereira, G. K. B., Guazzaroni, M.-E., and Silva-Rocha, R. (2018). Boosting secondary metabolite production and discovery through the engineering of novel microbial biosensors. *Biomed. Res. Int.* 2018:7021826. doi: 10.1155/2018/7021826
- Devi, A., Bajar, S., Kour, H., Kothari, R., Pant, D., and Singh, A. (2022). Lignocellulosic biomass valorization for bioethanol production: a circular bioeconomy approach. *Bioenergy Res.* 15, 1820–1841. doi: 10.1007/s12155-022-10401-9
- diCenzo, G. C., and Finan, T. M. (2017). The divided bacterial genome: structure, function, and evolution. *Microbiol. Mol. Biol. Rev.* 81, e00019–e00017. doi: 10.1128/MMBR.00019-17
- Dilokpimol, A., Verkerk, B., Li, X., Bellemare, A., Lavalley, M., Frommhagen, M., et al. (2022). Screening of novel fungal carbohydrate esterase family 1 enzymes identifies three novel dual feruloyl/acetyl xylan esterases. *FEBS Lett.* 596, 1932–1943. doi: 10.1002/1873-3468.14322
- Ezeilo, U. R., Zakaria, I. I., Huyop, F., and Wahab, R. A. (2017). Enzymatic breakdown of lignocellulosic biomass: the role of glycosyl hydrolases and lytic polysaccharide monoxygenases. *Biotechnol. Biotechnol. Equip.* 31, 1–16. doi: 10.1080/13102818.2017.1330124
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3, 289–306. doi: 10.4161/gmic.19897
- Garcia, R., and Müller, R. (2014). “The family Myxococcaceae” in *The prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*. eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt and F. Thompson (Berlin, Heidelberg: Springer), 191–212.
- Hamid, R., Khan, M. A., Ahmad, M., Ahmad, M. M., Abdin, M. Z., Musarrat, J., et al. (2013). Chitinases: an update. *J. Pharm. Bioallied Sci.* 5, 21–29. doi: 10.4103/0975-7406.106559
- Harindintwali, J. D., Wang, F., Yang, W., Zhou, J., Muhoza, B., Mugabowindekwe, M., et al. (2022). Harnessing the power of cellulolytic nitrogen-fixing bacteria for biovalorization of lignocellulosic biomass. *Ind. Crop. Prod.* 186:115235. doi: 10.1016/j.indcrop.2022.115235
- Hondoh, H., Saburi, W., Mori, H., Okuyama, M., Nakada, T., Matsuura, Y., et al. (2008). Substrate recognition mechanism of alpha-1,6-glucosidic linkage hydrolyzing enzyme, dextran glucosidase from *Streptococcus mutans*. *J. Mol. Biol.* 378, 913–922. doi: 10.1016/j.jmb.2008.03.016
- Inyang, V., Laseinde, O. T., and Kanakana, G. M. (2022). Techniques and applications of lignocellulose biomass sources as transport fuels and other bioproducts. *Int. J. Low-Carbon Technol.* 17, 900–909. doi: 10.1093/ijlct/ctac068
- Ivanova, N., Daum, C., Lang, E., Abt, B., Kopitz, M., Saunders, E., et al. (2010). Complete genome sequence of *Haliangium ochraceum* type strain (SMP-2T). *Stand. Genomic Sci.* 2, 96–106. doi: 10.4056/sigs.69.1277
- Kaiser, D., Robinson, M., and Kroos, L. (2010). Myxobacteria, polarity, and multicellular morphogenesis. *Cold Spring Harb. Perspect. Biol.* 2:a000380. doi: 10.1101/cshperspect.a000380
- Kellner, S., Spang, A., Offre, P., Szöllösi, G. J., Petitjean, C., and Williams, T. A. (2018). Genome size evolution in the Archaea. *Emerg. Topics in Life Sci.* 2, 595–605. doi: 10.1042/ETLS20180021
- Khalikova, E., Susi, P., and Korpela, T. (2005). Microbial dextran-hydrolyzing enzymes: fundamentals and applications. *Microbiol. Mol. Biol. Rev.* 69, 306–325. doi: 10.1128/MMBR.69.2.306-325.2005
- Krivorotova, T., and Sereikaite, J. (2014). Determination of fructan exohydrolase activity in the crude extracts of plants. *Electron. J. Biotechnol.* 17, 329–333. doi: 10.1016/j.ejbt.2014.09.005
- Kumar, A., and Chandra, R. (2020). Ligninolytic enzymes and its mechanisms for degradation of lignocellulosic waste in environment. *Heliyon*. 6:e03170. doi: 10.1016/j.heliyon.2020.e03170
- Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., et al. (2015). Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15, 141–161. doi: 10.1007/s10142-015-0433-4
- Lang, E., Schumann, P., Tindall, B. J., Mohr, K. I., and Spröer, C. (2015). Reclassification of *Angiococcus disciformis*, *Cystobacter minus* and *Cystobacter violaceus* as *Archangium disciforme* comb. nov., *Archangium minus* comb. nov. and *Archangium violaceum* comb. nov., unification of the families *Archangiaceae* and *Cystobacteraceae*, and emended descriptions of the families *Myxococcaceae* and *Archangiaceae*. *Int. J. Syst. Evol. Microbiol.* 65, 4032–4042. doi: 10.1099/ijsem.0.000533
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., et al. (2016). Evolution and ecology of Actinobacteria and their bioenergy applications. *Ann. Rev. Microbiol.* 70, 235–254. doi: 10.1146/annurev-micro-102215-095748
- Li, F., Foucat, L., and Bonnin, E. (2021). Effect of solid loading on the behaviour of pectin-degrading enzymes. *Biotechnol. Biofuels* 14:107. doi: 10.1186/s13068-021-01957-3
- Li, X., Shi, Y., Kong, W., Wei, J., Song, W., and Wang, S. (2022). Improving enzymatic hydrolysis of lignocellulosic biomass by bio-coordinated physicochemical pretreatment—a review. *Energy Rep.* 8, 696–709. doi: 10.1016/j.egy.2021.12.015
- Li, Z., Wu, J., Zhang, B., Wang, F., Ye, X., Huang, Y., et al. (2015). AmyM, a novel Maltohexaose-forming α -amylase from *Coralococcus* sp. strain EGB. *Appl. Environ. Microbiol.* 81, 1977–1987. doi: 10.1128/AEM.03934-14
- Li, Z., Xia, C., Wang, Y., Li, X., Qiao, Y., Li, C., et al. (2019). Identification of an endo-chitinase from *Coralococcus* sp. EGB and evaluation of its antifungal properties. *Int. J. Biol. Macromol.* 132, 1235–1243. doi: 10.1016/j.ijbiomac.2019.04.056
- Li, Y., Zhou, X., Zhang, X., Xu, Z., Dong, H., Yu, G., et al. (2022). A myxobacterial GH19 lysozyme with bacteriolytic activity on both gram-positive and negative phytopathogens. *AMB Express* 12:54. doi: 10.1186/s13568-022-01393-y
- Lightfield, J., Fram, N. R., and Ely, B. (2011). Across bacterial Phyla, distantly-related genomes with similar genomic GC content have similar patterns of amino acid usage. *PLoS One* 6:e17677. doi: 10.1371/journal.pone.0017677
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Marshall, R. C., and Whitworth, D. E. (2019). Is “wolf-pack” predation by antimicrobial Bacteria cooperative? Cell behaviour and predatory mechanisms indicate profound selfishness, even when working alongside kin. *Bioessays*. 41:e1800247. doi: 10.1002/bies.201800247
- Mohammed, A. S. A., Naveed, M., and Jost, N. (2021). Polysaccharides; classification, chemical properties, and future perspective applications in fields of pharmacology and biological medicine (a review of current applications and upcoming potentialities). *J. Polym. Environ.* 29, 2359–2371. doi: 10.1007/s10924-021-02052-2
- Mohr, K. I. (2018). Diversity of Myxobacteria—we only see the tip of the iceberg. *Microorganisms*. 6:E84. doi: 10.3390/microorganisms6030084
- Mudgil, D. (2017). “The interaction between insoluble and soluble Fiber” in *Dietary Fiber for the prevention of cardiovascular disease*. ed. R. A. Samaan (Oxford: Academic Press), 35–59.
- Muñoz-Dorado, J., Marcos-Torres, F. J., García-Bravo, E., Moraleda-Muñoz, A., and Pérez, J. (2016). Myxobacteria: moving, killing, feeding, and surviving together. *Front. Microbiol.* 7:7. doi: 10.3389/fmicb.2016.00781
- Ng, T. B., and Randy, C. (2011). Cellulase: types, actions, mechanisms and uses. *Cellulase: Types and Action, Mechanism and Uses.*, 251–263.
- Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71. doi: 10.1099/ijsem.0.005056
- Pal, S., Sharma, G., and Subramanian, S. (2021). Complete genome sequence and identification of polyunsaturated fatty acid biosynthesis genes of the myxobacterium *Minicystis rosea* DSM 24000T. *BMC Genomics* 22:655. doi: 10.1186/s12864-021-07955-x
- Panda, A., Islam, S. T., and Sharma, G. (2022). Harmonizing prokaryotic nomenclature: fixing the fuss over phylum name flipping. *MBio* 13, e00970–e00972. doi: 10.1128/mbio.00970-22
- Ramírez-Ramírez, N., Romero-García, E. R., Calderón, V. C., Avitia, C. I., Téllez-Valencia, A., and Pedraza-Reyes, M. (2008). Expression, characterization and synergistic interactions of Myxobacter Sp. AL-1 Cel9 and Cel48 Glycosyl hydrolases. *Int. J. Mol. Sci.* 9, 247–257. doi: 10.3390/ijms9030247
- Reichenbach, H. (1970). *Nannocystis exedens* gen. Nov., spec. Nov., a new myxobacterium of the family Sorangiaceae. *Arch. Mikrobiol.* 70, 119–138. doi: 10.1007/BF00412203
- Reichenbach, H. (1999). The ecology of the myxobacteria. *Environ. Microbiol.* 1, 15–21. doi: 10.1038/sj.jim.7000025
- Reichenbach, H. (2001). Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol.* 27, 149–156.
- Reichenbach, H. (2005). “Order VIII. Myxococcales” in: D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (eds.), *Bergey’s manual of systematic bacteriology*, second edition, (The Proteobacteria), part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria). (New York: Springer), 2, 1059–1072.
- Sanford, R. A., Cole, J. R., and Tiedje, J. M. (2002). Characterization and description of *Anaeromyxobacter dehalogenans* gen. Nov., sp. nov., an aryl-Halo-respiring facultative anaerobic Myxobacterium. *Appl. Environ. Microbiol.* 68, 893–900. doi: 10.1128/AEM.68.2.893-900.2002
- Schneider, S., Perlova, O., Kaiser, O., Alici, A., Altmeyer, M. O., et al. (2007). Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* 25, 1281–1289. doi: 10.1038/nbt1354
- Shahbaz, U., and Yu, X. (2020). Cloning, isolation, and characterization of novel chitinase-producing bacterial strain UM01 (*Myxococcus fulvus*). *J. Genet. Eng. Biotechnol.* 18:45. doi: 10.1186/s43141-020-00059-1

- Sharma, G., Khatri, I., and Subramanian, S. (2016). Complete genome of the starch-degrading Myxobacteria *Sandaracinus amylolyticus* DSM 53668T. *Genome Biol. Evol.* 8, 2520–2529. doi: 10.1093/gbe/evw151
- Sharma, G., and Subramanian, S. (2017). Unravelling the complete genome of *Archangium gephyra* DSM 2261T and evolutionary insights into Myxobacterial Chitinases. *Genome Biol. Evol.* 9, 1304–1311. doi: 10.1093/gbe/evx066
- Sichert, A., and Cordero, O. X. (2021). Polysaccharide-Bacteria interactions from the Lens of evolutionary ecology. *Front. Microbiol.* 12:705082. doi: 10.3389/fmicb.2021.705082
- Sidar, A., Voshol, G. P., El-Masoudi, A., Vijgenboom, E., and Punt, P. J. (2024). Highly variable domain architecture in carbohydrate-active enzymes highlights *Streptomyces* as promising resource for rice straw bioconversion. *Bioresour. Technol. Reports.* 25:101775. doi: 10.1016/j.biteb.2024.101775
- Silva, S. S., Fernandes, E. M., Pina, S., Silva-Correia, J., Vieira, S., Oliveira, J. M., et al. (2017). “2.11 polymers of biological origin” in *Comprehensive biomaterials II*. ed. P. Ducheyne (Oxford: Elsevier), 228–252.
- Stick, R. V., and Williams, S. J. (2009). “Disaccharides, oligosaccharides and polysaccharides” in *Carbohydrates: The essential molecules of life*. eds. R. V. Stick and S. J. Williams. *Second ed* (Oxford: Elsevier), 321–341.
- Takasuka, T. E., Book, A. J., Lewin, G. R., Currie, C. R., and Fox, B. G. (2013). Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci. Rep.* 3:1030. doi: 10.1038/srep01030
- Tashkandi, M., and Baz, L. (2023). Function of CAZymes encoded by highly abundant genes in rhizosphere microbiome of *Moringa oleifera*. *Saudi J. Biol. Sci.* 30:103578. doi: 10.1016/j.sjbs.2023.103578
- Joshi, N., Upadhyay, D., and Andhare, P. (2021). A study on amylase: review. *Int. J. Biol., Pharmacy Allied Sci.* 10, 333–340. doi: 10.31032/IJBPA/2021/10.4.1037
- Vera-Ponce de León, A., Jahnes, B. C., Duan, J., Camuy-Vélez, L. A., and Sabree, Z. L. (2020). Cultivable, host-specific Bacteroidetes symbionts exhibit diverse Polysaccharolytic strategies. *Appl. Environ. Microbiol.* 86, e00091–e00020. doi: 10.1128/AEM.00091-20
- Wang, S.-Y., Hu, W., Lin, X.-Y., Wu, Z.-H., and Li, Y.-Z. (2011). A novel cold-active xylanase from the cellulolytic myxobacterium *Sorangium cellulosum* So9733-1: gene cloning, expression, and enzymatic characterization. *Appl. Microbiol. Biotechnol.* 93, 1503–1512. doi: 10.1007/s00253-011-3480-3
- Whitworth, D. E., Sydney, N., and Radford, E. J. (2021). Myxobacterial genomics and post-genomics: a review of genome biology, genome sequences and related 'omics studies. *Microorganisms.* 9:2143. doi: 10.3390/microorganisms9102143
- Wu, J., Xia, B., Li, Z., Ye, X., Chen, Q., Dong, W., et al. (2015). Molecular cloning and characterization of a novel GH13 saccharifying α -amylase AmyC from *Corallococcus* sp. EGB. *Starch - Stärke.* 67, 810–819. doi: 10.1002/star.201400258
- Yamamoto, E., Muramatsu, H., and Nagai, K. (2014). *Vulgatibacter incomptus* gen. Nov., sp. nov. and *Labilithrix luteola* gen. Nov., sp. nov., two myxobacteria isolated from soil in Yakushima Island, and the description of *Vulgatibacteraceae* fam. Nov., *Labilithricaceae* fam. Nov. and *Anaeromyxobacteraceae* fam. Nov. *Int. J. Syst. Evol. Microbiol.* 64, 3360–3368. doi: 10.1099/ijs.0.063198-0
- Yang, W., Su, L., Wang, L., Wu, J., and Chen, S. (2022). Alpha-glucanotransferase from the glycoside hydrolase family synthesizes α (1–6)-linked products from starch: features and synthesis pathways of the products. *Trends Food Sci. Technol.* 128, 160–172. doi: 10.1016/j.tifs.2022.08.001
- Yang, Y., Yoo, C. G., Guo, H.-B., Rottmann, W., Winkler, K. A., Collins, C. M., et al. (2017). Overexpression of a domain of unknown function 266-containing protein results in high cellulose content, reduced recalcitrance, and enhanced plant growth in the bioenergy crop *Populus*. *Biotechnol. Biofuels* 10:74. doi: 10.1186/s13068-017-0760-x
- Yeager, C. M., Gallegos-Graves, L. V., Dunbar, J., Hesse, C. N., Daligault, H., and Kuske, C. R. (2017). Polysaccharide degradation capability of Actinomycetales soil isolates from a semiarid grassland of the Colorado plateau. *Appl. Environ. Microbiol.* 83, e03020–e03016. doi: 10.1128/AEM.03020-16
- Yousuf, A., Pirozzi, D., and Sannino, F. (2020). (eds.). Chapter 1 - Fundamentals of lignocellulosic biomass. Lignocellulosic biomass to liquid biofuels, Academic Press, 1–15. Available at: <https://www.sciencedirect.com/science/article/pii/B9780128159361000010>
- Zerva, A., Pentari, C., Ferousi, C., Nikolaivits, E., Karnaouri, A., and Topakas, E. (2021). Recent advances on key enzymatic activities for the utilisation of lignocellulosic biomass. *Bioresour. Technol.* 342:126058. doi: 10.1016/j.biortech.2021.126058
- Zhang, G., and Dong, Y. (2022). Design and application of an efficient cellulose-degrading microbial consortium and carboxymethyl cellulase production optimization. *Front. Microbiol.* 13:957444. doi: 10.3389/fmicb.2022.957444
- Zhang, L., Dong, C., Wang, J., Liu, M., Wang, J., Hu, J., et al. (2023). Predation of oomycetes by myxobacteria via a specialized CAZyme system arising from adaptive evolution. *ISME J.* 17, 1089–1103. doi: 10.1038/s41396-023-01423-y
- Zhao, Y., Wang, Y., Xia, C., Li, X., Ye, X., Fan, Q., et al. (2021). Whole-genome sequencing of *Corallococcus* sp. strain EGB reveals the genetic determinants linking taxonomy and predatory behavior. *Genes (Basel)* 12:1421. doi: 10.3390/genes12091421
- Zhou, X., Chen, X., Qi, X., Zeng, Y., Guo, X., Zhuang, G., et al. (2023). Soil bacterial communities associated with multi-nutrient cycling under long-term warming in the alpine meadow. *Front. Microbiol.* 14:14. doi: 10.3389/fmicb.2023.1136187
- Zhou, J., Li, Z., Zhang, H., Wu, J., Ye, X., Dong, W., et al. (2018). Novel Maltogenic amylase CoMA from *Corallococcus* sp. strain EGB catalyzes the conversion of Maltooligosaccharides and soluble starch to maltose. *Appl. Environ. Microbiol.* 84, e00152–e00118. doi: 10.1128/AEM.00152-18
- Zhou, Y., Yi, S., Zang, Y., Yao, Q., and Zhu, H. (2021). The predatory Myxobacterium *Citricoccus inhibens* gen. Nov. sp. nov. showed antifungal activity and Bacteriolytic property against Phytopathogens. *Microorganisms* 9:2137. doi: 10.3390/microorganisms9102137
- Zhou, X., Zhou, X., Zhang, X., Dong, H., Dong, Y., and Zhu, H. (2024). Secretory CAZymes profile and GH19 enzymes analysis of *Corallococcus silvisolis* c25j21. *Front. Microbiol.* 15:1324153. doi: 10.3389/fmicb.2024.1324153
- Zoghalmi, A., and Paës, G. (2019). Lignocellulosic biomass: understanding recalcitrance and predicting hydrolysis. *Front. Chem.* 7:7. doi: 10.3389/fchem.2019.00874
- Zusman, D. R., Scott, A. E., Yang, Z., and Kirby, J. R. (2007). Chemosensory pathways, motility and development in *Myxococcus xanthus*. *Nat. Rev. Microbiol.* 5, 862–872. doi: 10.1038/nrmicro1770