Check for updates

Received: 1 September 2022 | Accepted: 26 November 2022

DOI: 10.1111/1462-2920.16290

RESEARCH ARTICLE

ENVIRONMENTAL MICROBIOLOGY Applied Microbiology International

Estimate of the degradation potentials of cellulose, xylan, and chitin across global prokaryotic communities

Dan-dan Li[1] | **Zheng Zhang**[1,2] | **Jia-ning Wang**[1] | **Peng Zhang**[1] | **Ya Liu**[1] | **Yue-zhong Li**[1]

[1]State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao, China

[2]Suzhou Research Institute, Shandong University, Suzhou, China

Correspondence
Zheng Zhang and Yue-zhong Li, State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao 266237, China.
Email: zhangzheng@sdu.edu.cn and lilab@sdu.edu.cn

Funding information
China Postdoctoral Science Foundation, Grant/Award Numbers: 2021M700084, 2021M701987; Fundamental Research Funds of Shandong University, Grant/Award Number: 2020GN113; Key Research and Development Program of Shandong Province, Grant/Award Number: 2018GSF121015; National Key Research and Development Program, Grant/Award Numbers: 2018YFA0900400, 2018YFA0901704; National Natural Science Foundation of China, Grant/Award Numbers: 32070030, 32270073; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20190199; Science & Technology Fundamental Resources Investigation Program, Grant/Award Number: 2022FY101100

Abstract

Complex polysaccharides (e.g. cellulose, xylan, and chitin), the most abundant renewable biomass resources available on Earth, are mainly degraded by microorganisms in nature. However, little is known about the global distribution of the enzymes and microorganisms responsible for the degradation of cellulose, xylan, and chitin in natural environments. Through large-scale alignments between the sequences released by the Earth Microbiome Project and sequenced prokaryotic genomes, we determined that almost all prokaryotic communities have the functional potentials to degrade cellulose, xylan, and chitin. The median abundances of genes encoding putative cellulases, xylanases, and chitinases in global prokaryotic communities are 0.51 (0.17–1.01), 0.24 (0.05–0.57), and 0.33 (0.11–0.71) genes/cell, respectively, and the composition and abundance of these enzyme systems are environmentally varied. The taxonomic sources of the three enzymes are highly diverse within prokaryotic communities, and the main factor influencing the diversity is the community's alpha diversity index rather than gene abundance. Moreover, there are obvious differences in taxonomic sources among different communities, and most genera with degradation potentials are narrowly distributed. In conclusion, our analysis preliminarily depicts a panorama of cellulose-, xylan-, and chitin-degrading enzymatic systems across global prokaryotic communities.

INTRODUCTION

The degradation of complex polysaccharides (e.g. cellulose, xylan, and chitin) is a key step in the global carbon cycle (Bomble et al., 2017; Schmidt & Dauenhauer, 2007). Of these polysaccharides, cellulose and xylan are the major carbon sources in terrestrial ecosystems. Cellulose, a globally abundant polysaccharide, and hemicellulose, with xylan as its main component, together constitute plant cell wall material (Gao et al., 2020; Scheller & Ulvskov, 2010). Chitin is produced by arthropods and fungi, dominates as a component of marine ecosystem biomass and is also an abundant structural polysaccharide (Chen et al., 2020; Stuart et al., 2016). In nature, many glycoside hydrolases (GHs) targeting cellulose, xylan, and chitin are derived from prokaryotic genomes (Berlemont & Martiny, 2016; Fabian et al., 2017; García-Palacios et al., 2016). It is important to reveal the composition of the enzymes and microorganisms responsible for degrading these complex polysaccharides in different environments, which will aid our understanding of degradation processes and will also provide a guide for the efficient exploration of these enzyme resources in nature.

Cellulose, as a representative complex polysaccharide, is currently the most abundant renewable biomass resource for generating materials and energy needed

© 2022 Applied Microbiology International and John Wiley & Sons Ltd.

for the sustainable development of human society (Ragauskas et al., 2006; Schubert, 2006). One example of a potential way to help solve the world's energy problems is converting cellulose into sugars that can be fermented by microorganisms to produce bioethanol (Lopez-Mondejar et al., 2019; Petridis & Smith, 2018). However, the natural resistance of plant cell walls to microbial and enzymatic deconstruction is largely responsible for the high cost of cellulose conversion (Bomble et al., 2017; Dragone et al., 2020; Himmel et al., 2007). The decomposition of complex polysaccharides (e.g. cellulose) in nature is accomplished by the concerted action of multiple GHs (Brune, 2014; Cragg et al., 2015; Li et al., 2011; Payne et al., 2015). Understanding the compositions of complex polysaccharide-degrading enzymatic systems in microbial communities will help guide the optimization of industrial production and can also provide opportunities for discovering novel industrial enzymes that resist the diverse biological stresses that occur in natural environments (Elleuche et al., 2015; Lopez-Mondejar et al., 2016; Tiwari et al., 2018).

Our previous studies showed that the genome-sequenced proportion of global major prokaryotic communities has reached a high level (Zhang et al., 2020). The Earth Microbiome Project (EMP) samples the Earth's microbial communities at an unprecedented scale to assess the distribution of prokaryotes in global environments (Thompson et al., 2017). Herein, by correlating 10,000 EMP samples with data on more than 200,000 prokaryotic genomes, we comprehensively investigated the compositions and abundances of cellulases, xylanases, and chitinases to provide a panorama of the functional potentials of diverse prokaryotic communities involved in complex carbohydrate degradation in various global environments.

EXPERIMENTAL PROCEDURES

Mapping EMP operational taxonomic units to prokaryotic genomes

The EMP aimed to sample the Earth's microbial communities at an unprecedented scale to advance our understanding of the organizing biogeographic principles that govern the microbial community structure on Earth (Thompson et al., 2017). Our analysis was conducted using the 10,000-sample subset released by the EMP, which is considered to be representative for studies of diverse environmental types. A modest sequencing depth of 5000 observations was randomly selected from each sample. A total of 262,011 OTUs were obtained from a set of 10,000 EMP samples using Deblur, an analysis method based on exact sequences rather than clustering, to remove suspected error sequences and provide single-nucleotide resolution 'sub-OTUs' (Amir et al., 2017). Chimera filtering relied on the EMP project.

The NCBI RefSeq database is a curated non-redundant collection of sequences representing whole or frame genomes (Pruitt et al., 2007). We obtained all 217,614 bacterial and archaeal genomes collected by the database. The 16S tags from the EMP data were mapped to those extracted from the prokaryotic genomes using BLASTn (Boratyn et al., 2013; Camacho et al., 2009) with a 100% identity threshold and 95% alignment coverage. If multiple genomes were found, those with the highest alignment identity and length were selected. In addition, considering that the sequenced genomes are mainly from cultivated strains, we also used a 16S rRNA-V4 region identity of not less than 97% as the criterion to associate a fraction of uncultivated OTUs for control studies.

Identification of cellulases, xylanases, and chitinases

CAZy is a specialized database dedicated to the display and analysis of genomic, structural, and biochemical information on carbohydrate-active enzymes (Lombard et al., 2014; Turnbaugh et al., 2009). In the CAZy database, enzyme activities that are experimentally demonstrated are assigned to individual enzyme modules whenever possible. We collected information on GH family classifications and experimentally characterized proteins from the CAZy database. Taxonomic classification information was also obtained from the NCBI Taxonomy database (Sayers et al., 2021).

We focused on GH families predominantly involved in cellulose, xylan, and chitin deconstruction. Most glycoside hydrolases from the families GH5, GH6, GH7, GH8, GH9, GH12, GH44, GH45, and GH48 act on cellulose (EC 3.2.1.4, EC 3.2.1.74, EC 3.2.1.91 and EC 3.2.1.176), while the families GH10, GH11, and GH30 are mostly xylanases (EC 3.2.1.8 and EC 3.2.1.32), and the families GH18, GH19, and GH85 are chitinases (EC 3.2.1.14, EC 3.2.1.96, EC 3.2.1.200 and EC 3.2.1.201) (Berlemont & Martiny, 2015; Lombard et al., 2014; Talamantes et al., 2016).

Most sequences from the GH families that are involved in cellulose, xylan, or chitin deconstruction can be identified using Pfam- and CDD-based annotations (Nguyen et al., 2018). For the obtained prokaryotic genome-encoded protein sequences, RPS-BLAST was used for domain annotation and family classification (E value < 0.01). Combining the CAZy annotations, characterized functional proteins, and Pfam (or CDD) family classifications, we identified putative cellulases, xylanases, and chitinases from the abovementioned 15 GH families on a large scale. Most of these families

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

are functionally monospecific, while the GH families GH5, GH8, GH30, and GH44 are polyspecific. For polyspecific families, we performed detailed sequence comparisons using BLAST sequence searches to identify close relatives of functionally characterized proteins (best matches, *E* value < 0.001).

Global distribution of enzymes

Based on the corresponding relationship between RefSeq genomes and EMP OTUs, we analysed the environmental distributions and abundances of putative prokaryote-encoding cellulase, xylanase, and chitinase genes. The number of 16S rRNA genes and the number of three enzyme genes in the genome were counted for the OTUs corresponding to the genomes in each sample. Then, the gene abundance of the three enzymes was individually calculated based on OTUs, which are shown as the gene/cell value, that is, the average number of these enzyme genes carried by a prokaryotic cell in the community. Global distributions of cellulases, xylanases, and chitinases were constructed by using the 'maps' package in R.

EMPO is used to assign samples to habitats in a hierarchical framework that captures the major axes of microbial community diversity (Table S2) and classifies 17 microbial environments (Level 3) as free-living or host-associated (Level 1) and saline or non-saline (if free-living) or animal or plant (if host-associated) (Level 2) (Thompson et al., 2017). Based on the random forest classification method, the samples were classified into the corresponding environmental labels. The subset of 10,000 samples released by EMP gives an equal (as possible) number of samples in various EMPO_3 environments. The hypersaline (saline) environment was excluded as only 13 samples were included, and the environmental distributions of cellulase, xylanase, and chitinase genes were determined based on the other 16 EMPO_3 environments.

To weaken the effect of differentiated subsets of samples in similar environments, we also used Environment Ontology (ENVO) for analysis. ENVO is a resource and research target for the semantically controlled description of environmental entities, representing the biomes, environmental features, and environmental materials pertinent to genomic and microbiome-related investigations (Buttigieg et al., 2016). All the 24 ENVO environments studied each contained more than 20 samples. In addition, we specifically compared gut samples from herbivores (artiodactyla) with those from omnivores (primates). Based on the environmental parameters recorded for the EMP samples, we also analysed the enzyme distributions in five extreme environments, that is, high temperature (≥45°C), low temperature (≤10°C), acidic (pH ≤5), alkaline (pH ≥9), and saline (>50 psu).

RESULTS

Prokaryotic cellulases, xylanases, and chitinases are globally distributed

Based on the classification and annotation of the CAZy database, most glycoside hydrolases from the families GH5, GH6, GH7, GH8, GH12, GH44, GH45, and GH48 act on cellulose, while the families GH10, GH11, and GH30 are primarily xylanases, and the families GH18, GH19, and GH85 are chitinases (Berlemont & Martiny, 2015; Lombard et al., 2014; Talamantes et al., 2016). Most of the known members of these GH families are from prokaryotes; for example, more than 90% of the families GH6, GH8, GH30, GH44, GH48, and GH85 members are encoded by bacteria, whereas only GH7 and GH45 are mainly derived from eukaryotes. For the major cellulase families, we found that more than 90% of the functionally characterized members of the families GH6, GH7, GH9, GH45, and GH48 had cellulase activity (EC 3.2.1.4, EC 3.2.1.74, EC 3.2.1.91, and EC 3.2.1.176), which exhibited single-substrate specificity (Figure 1A and Table S1). More than 95% of the characterized GH12 family members from archaea and bacteria were also cellulases (Figure 1B). The families GH5, GH8, and GH44 displayed multisubstrate specificity, with 40%–70% of the characterized members being cellulases. In terms of xylanase families, more than 95% of the functionally characterized members of GH10 and GH11 showed xylanase activity (EC 3.2.1.8 and EC 3.2.1.32). Approximately 40% of the characterized members of the polyspecific GH30 family were xylanases. In addition, there were a small number of xylanases in families GH5 and GH8. Similarly, in GH18, GH19, and GH85, among the characterized proteins, more than 95% displayed chitinase activity (EC 3.2.1.14, EC 3.2.1.96, EC 3.2.1.200, and EC 3.2. 1.201).

To evaluate the distribution of cellulases, xylanases, and chitinases in global prokaryotic communities, we identified genes (belonging to the 15 GH families described above) potentially encoding these three enzymes in the sequenced prokaryotic genomes, correlated them with EMP sample data, and calculated the relative abundances of cellulases, xylanases, and chitinases in each sample (genes/cell, see Experimental procedures section). We found that under the most stringent criteria, that is, the corresponding relationships between the sequenced genomes and EMP operational taxonomic units (OTUs) were determined by 100% complete identity of the 16 S rRNA-V4 region, prokaryotic-encoded putative cellulases, xylanases, and chitinases were detected in at least 99.1%, 95.6%, and 98.1%, respectively, of the 10,000 EMP samples studied (Figure S1). This implies that almost all prokaryotic communities have functional potentials to degrade cellulose, xylan, and chitin. The median gene

14622920, 2023, 2, Downloaded from https://enviromicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

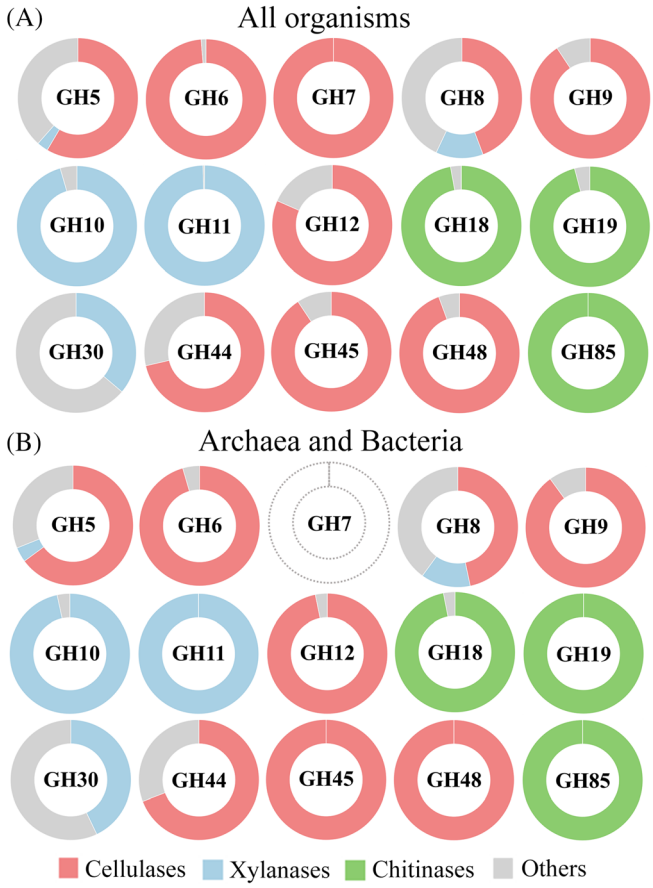


FIGURE 1 Distribution of substrate specificity and activities of experimentally characterized enzymes from GH families associated with the deconstruction of cellulose, xylan, and chitin. (A) Enzymatic proportions in 15 GH families from all organisms. (B) Enzymatic proportions in 15 GH families from archaea and bacteria. Red represents cellulases, blue represents xylanases, green represents chitinases, and grey represents enzymes with other functions. Statistics are based on the CAZy database, and Table S1 contains detailed information.

abundance of putative cellulases in the EMP samples was 0.51 (0.17–1.01) genes/cell, followed by chitinases (0.33 [0.11–0.71]) and xylanases (0.24 [0.05–0.57]) (Dataset S1). Thus, the prokaryotic-encoded cellulase, xylanase, and chitinase genes are widely distributed globally.

Using the 10,000 EMP samples, we previously determined a high proportion of genomic information revealed for global prokaryotic communities (Zhang et al., 2020). After 2 years, the median proportions of cells and taxa with known genomic information were updated to 40.2% (20.5%–88.1%) and 21.3% (11.4%–57.0%), respectively, with the criterion of 100% identity of the EMP 16S rRNA (V4 region) to that of sequenced genomes. If the 97% sequence identity was used as a criterion, the median proportions of cells and taxa with known genomic information further reached 65.2% (46.3%–94.7%) and 42.5% (29.7%–75.9%), respectively. Because the sequenced prokaryotic genomes were mainly from cultivated strains, the 100% sequence identity was thus mainly associated with cultivated strains only, and the 97% identity was associated with cultivated and partially uncultivated strains. For the relative abundances of cellulases, xylanases, and chitinases in prokaryotic communities, the values estimated from these two data subsets (with the 100% and 97% identities) were highly positively correlated ($p < 0.001$, Figure S2). Moreover, consideration of only cultivated strains did not significantly underestimate the relative abundances of these three enzymes in the community compared to the values obtained by correlating some uncultivated strains.

The potential of prokaryotes to degrade cellulose, xylan and chitin varies highly among environments

For different environment types of the Earth Microbiome Project Ontology (EMPO) classification, we calculated the gene abundances of prokaryotic cellulases, xylanases, and chitinases based on the 100% sequence identity of the 16 S rRNA (V4 region) (Figure 2 and Dataset S1). The results showed that the gene abundance of putative cellulases differed greatly among prokaryotic communities in different environments, which was highest in the soil (non-saline) environment, reaching 1.21 (0.92–1.63) genes/cell, moderate in the animal distal gut environment, 0.68 (0.28–0.97) genes/cell, and lowest in the water (saline) environment at 0.25 (0.13–0.49) genes/cell. For host-associated prokaryotic communities, the cellulase gene abundance in plant-associated communities was significantly higher than that in animal-associated communities (Wilcoxon rank test, $p < 0.001$). For free-living prokaryotic communities, the cellulase gene abundance in non-saline environmental communities was significantly higher than that in saline environmental communities (Wilcoxon rank test, $p < 0.001$). Generally, soil and plant rhizosphere environments are rich in cellulose, and plant hosts can also provide more cellulose substrates than animal hosts, suggesting that the gene abundance of cellulases in prokaryotic communities may be highly correlated with the supply of cellulose substrates in the environment.

Likewise, the gene abundances of xylanases and chitinases in prokaryotic communities may also be closely related to environmental substrate supplies (Figure 2 and Dataset S1). The putative xylanase gene abundance was highest in prokaryotic communities of the plant surface environment, at 0.78 (0.29–1.97) genes/cell, and similarly high in the soil (non-saline) environment, but very low in the animal distal gut and water (saline) environments. The putative chitinase genes were the most abundant in prokaryotic communities in the soil (non-saline) environment, reaching 0.87

14629920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

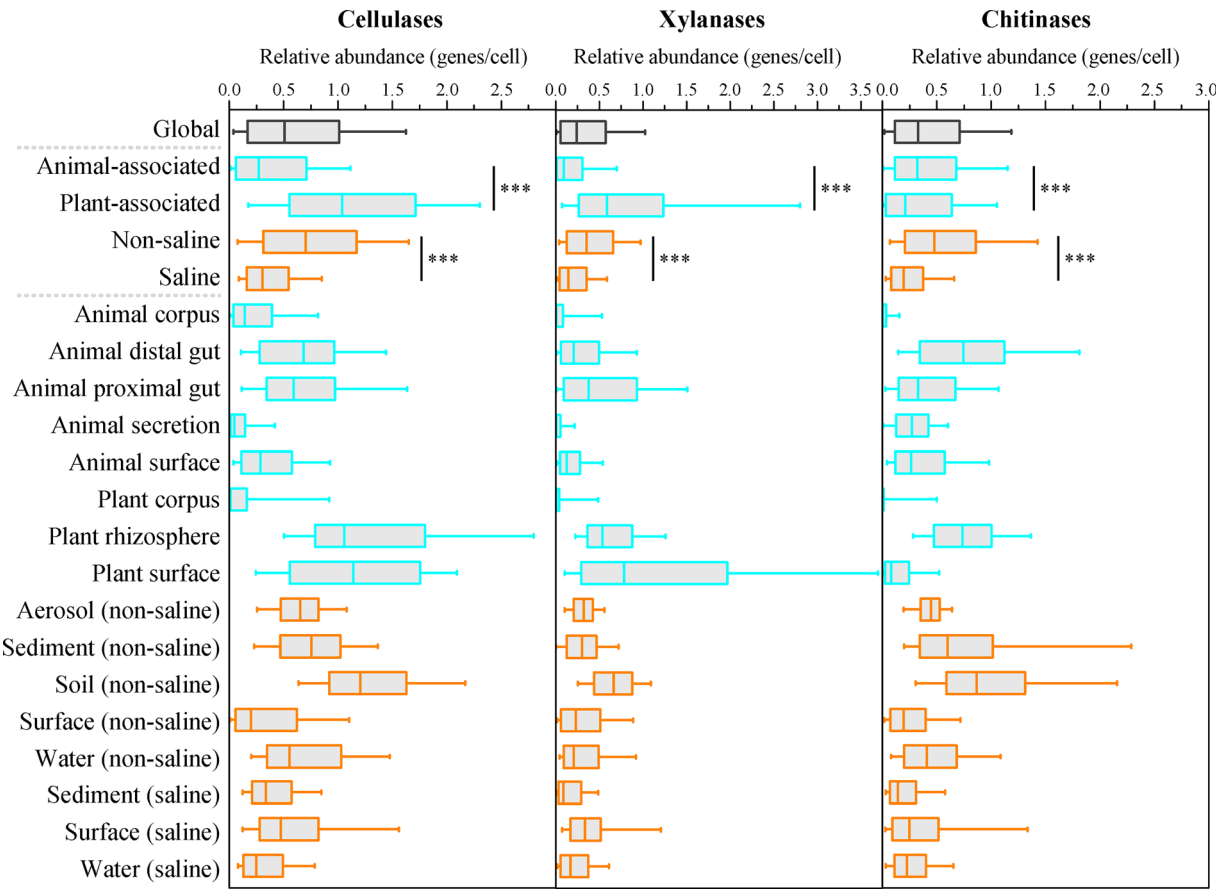


FIGURE 2 Different distributions of putative cellulases, xylanases, and chitinases in Earth Microbiome Project Ontology (EMPO) environments. Boxplots represent the gene abundance (genes/cell) of cellulases, xylanases, or chitinases for each sample in different environments at the global, EMPO_2 and EMPO_3 levels. The hypersaline (saline) environment was excluded due to few samples. Cyan represents host associated, and orange represents free living. For the boxplots, the middle line indicates the median, the box is the 25th to 75th percentile, and the whisker is the 10th to 90th percentile of the observed values. Comparisons between bins at the EMPO_2 level were analysed using the Wilcoxon rank test, ***$p < 0.001$. Table S2 provides a description of the EMPO environments and the number of samples included. Source data are provided in Dataset S1.

(0.59–1.31) genes/cell. Chitinase gene abundance was also high in the animal distal gut environment but low in the water (saline) environment. For host-associated prokaryotic communities, the gene abundance of xylanases was much higher in plant-associated communities than in animal-associated communities (Wilcoxon rank test, $p < 0.001$), but the opposite was true for chitinases (Wilcoxon rank test, $p < 0.001$). For free-living prokaryotic communities, the abundances of both the xylanase and chitinase genes were significantly higher in non-saline environmental communities than in saline environmental communities (Wilcoxon rank test, $p < 0.001$).

To explore potential relationships to the substrate supply, we further estimated the gene abundances of putative prokaryotic cellulases, xylanases, and chitinases based on the Environment Ontology (ENVO) classification that semantically represents environmental features (Buttigieg et al., 2016). Of the three enzymes, cellulase genes were the most abundant in tropical shrubland, at 1.83 (1.36–2.12) genes/cell and were also abundant in several types of grassland and forest environments but were less abundant in marine and anthropogenic terrestrial biome environments (except for cropland) (Figure S3). Similarly, the gene abundance of xylanases was high in forest, shrubland, and grassland environments and was the lowest in the village environment. There was a substantial abundance of chitinase genes in grassland, shrubland, and forest environments, with the peak in the montane grassland environment, but a low abundance in marine, desert, and anthropogenic terrestrial environments (except for cropland). Furthermore, for gut samples, the abundance of cellulase genes encoded by gut prokaryotic communities in herbivores (artiodactyla) was significantly higher than that in omnivores (primates) (Wilcoxon rank test, $p < 0.001$). Therefore, the potential for prokaryotic communities to degrade cellulose, xylan,

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and chitin is highly variable among environments and seems to be closely related to the supply of corresponding substrates in the environment.

Cellulose-, xylan-, and chitin-degrading enzymatic systems in prokaryotic communities are diverse across environments

The degradation of complex polysaccharides, including cellulose, xylan, and chitin, is typically not dependent on a single species but is achieved through the coordination of complex microbial communities in nature (Cragg et al., 2015; Li et al., 2011). We focused on the entire enzymatic system that potentially degraded cellulose, xylan, and chitin in prokaryotic communities in diverse environments. The occurrence frequency of each GH family in all the EMP samples was calculated (Figure S4). The results showed that the putative cellulases of families GH5, GH8, and GH9 appeared in more than 85% of the samples, followed by GH6, GH12, GH44, and GH48 with a frequency of approximately 30%–60%, whereas GH45 was only present in less than 5% of the samples. Xylanases of the GH10 family appeared in more than 95% of the samples, while the families GH5, GH8, GH11, and GH30 were present in approximately 30%–50% of the samples. For chitinases, the GH18 family was present in more than 95% of the samples, and the families GH19 and GH85 appeared in more than 70% of the samples.

We found that for more than 50% of the samples, the putative cellulases in a single community contained at least five different GH families, and more than 95% of the samples had at least two coexisting cellulase GH families (Figure 3A). Similarly, at least two different GH

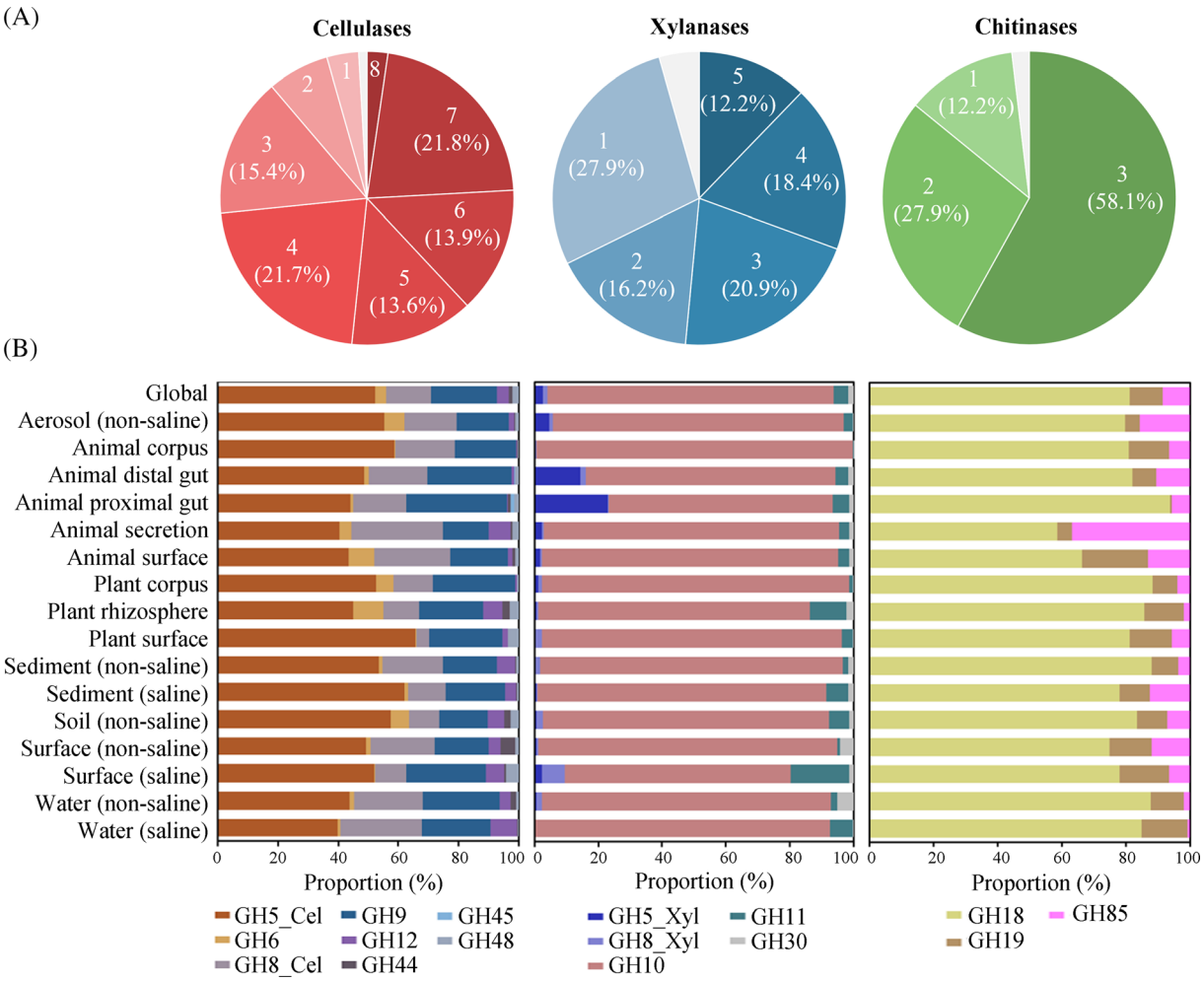


FIGURE 3 Compositions of potential cellulose-, xylan-, and chitin-degrading enzymatic systems in prokaryotic communities. (A) Statistics on the number of families encoding cellulases, xylanases or chitinases in a single community across the Earth Microbiome Project (EMP) samples. The plots indicated that a single prokaryotic community may encode 0–8 cellulase families, 0–5 xylanase families, or 0–3 chitinase families. Colours from dark to light represent the number decrease of coexisting families in a single community. (B) Compositions of putative cellulases, xylanases, and chitinases in different Earth Microbiome Project Ontology (EMPO) environments. The column chart displays the proportion of each GH family in the total gene abundance of such enzymes in each environment. Source data are provided in Dataset S1.

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

families for putative xylanases coexisted in more than two-thirds of the samples, and more than 10% of the samples had all the five xylanase GH families. For the chitinase genes, more than 85% of the samples contained at least two GH families, and nearly 60% of the samples contained all the three chitinase families. These results suggested that prokaryotic communities may generally have highly diverse enzymatic systems for the degradation of cellulose, xylan, and chitin.

Furthermore, we statistically analysed the compositions of potential cellulose-, xylan-, and chitin-degrading enzymatic systems encoded by the prokaryotic communities in different EMPO environment types. Globally, for the cellulose-degrading enzymatic system, the GH5 family members contributed more than 50% of the total cellulase gene abundance, GH8 contributed approximately 15%, GH9 contributed approximately 20%, and the other families collectively contributed approximately 10% (Figure 3B). Notably, the composition of the cellulose-degrading enzymatic system varied markedly across environments. For instance, the proportion of GH5 gene abundance changed from 39.8% (water (saline)) to 65.7% (plant surface), the GH8 family ranged from 4.1% (plant surface) to 30.3% (animal secretion), and the GH9 family ranged from 15.4% (animal secretion) to 33.5% (animal proximal gut). Additionally, the environment with the highest proportion of GH6 cellulase genes was the plant rhizosphere environment, reaching 10.3%. Comparably, the highest proportions of families GH12, GH44, GH45, and GH48 were in water (saline), surface (non-saline), animal proximal gut, and surface (saline) environments, with values of 8.5%, 5.0%, 1.4%, and 4.1%, respectively.

For the potential xylan-degrading enzymatic system, the GH10 family members occupied approximately 90% of the total gene abundance, while the other families together accounted for approximately 10% (Figure 3B). At the EMPO_3 level, the proportion of the GH10 family fluctuated between 70.0% (animal proximal gut) and 99.1% (animal corpus). For other xylanase families, the proportion of GH5 family members was high in the animal proximal gut (22.9%) and animal distal gut (14.4%) environments, while the GH8, GH11, and GH30 families accounted for the highest proportions in the surface (saline), surface (saline), and water (non-saline) environments, with values of 7.2%, 18.5%, and 5.0%, respectively.

For the potential chitin-degrading enzymatic system, the GH18 family members contributed approximately 80% of the total gene abundance, while the GH19 and GH85 families each accounted for approximately 10% (Figure 3B). The proportion of the GH18 family in different EMPO_3 environment types fluctuated from 58.5% (animal secretion) to 93.5% (animal proximal gut), while GH19 and GH85 had the highest proportions in the animal surface and animal secretion environments, with corresponding proportions of 20.6% and 37.0%, respectively. Consequently, the composition of cellulose-, xylan-, and chitin-degrading enzymes in prokaryotic communities is not only diverse but also highly variable among diverse environments.

The taxonomic sources of cellulase, xylanase and chitinase genes in prokaryotic biomes are diverse

Based on the sequenced prokaryotic genomes, a total of 933 genera of 10 bacterial groups and 5 genera of 3 archaeal groups encoded cellulases, 614 genera of 7 bacterial groups and 4 genera of 2 archaeal groups encoded xylanases, and 681 genera of 11 bacterial groups and 2 genera of 2 archaeal groups encoded chitinases in the 10,000 EMP samples (Dataset S2). In a single sample, the median numbers of prokaryotic genera encoding putative cellulases, xylanases, and chitinases were 14 (6–30), 7 (3–19), and 11 (5–23), respectively (Figure 4).

Across different EMPO environment types, the numbers of genera encoding the enzymes differed by more than 10 times (Figure 4). For example, the median numbers of prokaryotic genera encoding cellulases, xylanases, and chitinases in the plant rhizosphere environment were 52 (46–58), 36 (31–41), and 36 (32–40), but only 3 (2–6), 1 (0–2), and 2 (1–5) in the animal corpus environment, respectively. The diversity of taxonomic sources (number of genera) of cellulases, xylanases, and chitinases in prokaryotic communities was highly positively correlated with each other (Pearson correlation, $p < 0.001$, Figure S5). We found that the taxonomic source diversities of the three enzymes in the prokaryotic community were highly positively correlated with the community's alpha diversities but weakly correlated with the gene abundances of the three enzymes (Figure 5A,B). Thus, the alpha diversity index, rather than enzymatic gene abundance or the abundance of substrate supply, is the main factor affecting the diversity of the taxonomic sources of cellulases, xylanases, and chitinases in environments.

Although the taxonomic sources of cellulase, xylanase, and chitinase genes within prokaryotic communities were highly diverse, they were also highly differentiated across communities. The number of cross-community prokaryotic genera encoding cellulase, xylanase and chitinase genes decayed exponentially with increasing community numbers (Figure 5C). On the one hand, between 40% (plant rhizosphere, xylan) and 90% (animal corpus, cellulose) of the genera with the corresponding substrate degradation potentials were present in no more than 10 EMP samples. On the other hand, widely distributed genera were very rare. Even if the statistical analysis was performed in different EMPO environment types, the genera that were present in more than 50% of the samples in an

14622920, 2023, 2, Downloaded from https://enviromicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

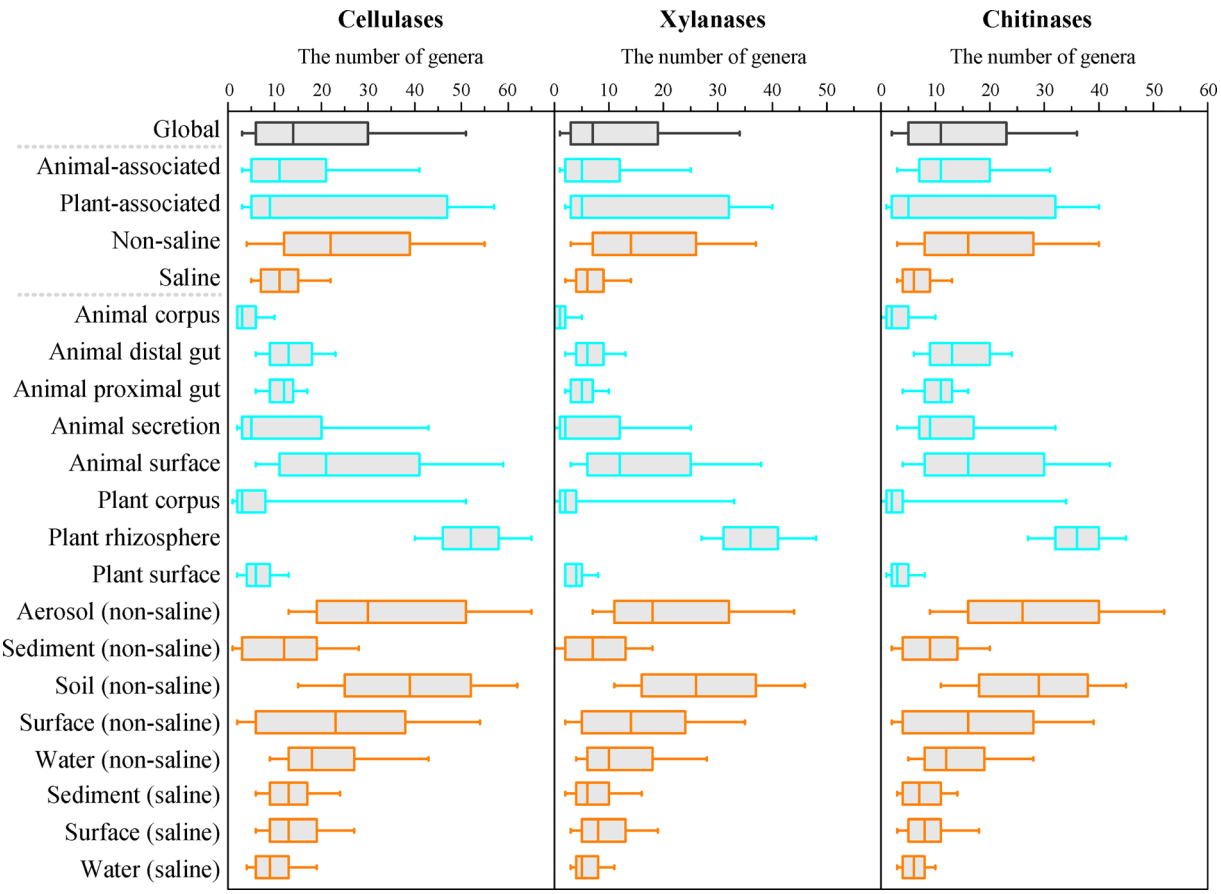


FIGURE 4 Boxplots represent the number of prokaryotic genera encoding putative cellulases, xylanases, or chitinases per sample in the different environments. Cyan represents host associated and orange represents free living. For the boxplots, the middle line indicates the median, the box represents the 25th to 75th percentiles, and the whisker indicates the 10th to 90th percentiles of observations. Table S2 provides a description of the Earth Microbiome Project Ontology (EMPO) environments and the number of samples included. Source data are provided in Dataset S2.

environment did not exceed 5% of the number of genera with the corresponding substrate degradation potential in that environment, with the exception of the plant rhizosphere, which reached approximately 10%. These results suggested that, even for the same type of environment, the genera with the corresponding substrate degradation potentials vary substantially across communities.

We highlighted the top 10 frequently occurring prokaryotic genera that encode cellulases, xylanases, or chitinases in each EMPO environment (Figures 6 and S6). The most widely distributed prokaryotic genera encoding cellulases included 10 genera of 3 bacterial groups and occurred in different frequencies across environments, from 54.6% (animal secretion, *Cronobacter*) to 99.6% (plant rhizosphere, *Bradyrhizobium*). Similarly, for xylanases, the most widely distributed prokaryotic genera included 11 genera of 4 bacterial groups, with an occurrence frequency ranging from 32.8% (animal secretion, *Paraprevotella*) to 99.6% (plant rhizosphere, *Bradyrhizobium*); for chitinases, 11 genera from 3 bacterial groups were the most widely distributed, with an occurrence frequency ranging from 50.3% (animal corpus, *Klebsiella*) to 99.6% (plant rhizosphere, *Cronobacter*). Notably, the widely distributed prokaryotic genera were often similar; for example, the most widely distributed prokaryotic genera were identical in seven environments for cellulases and xylanases or identical in three environments for cellulases and chitinases.

Microorganisms living in extreme environments are considered to be an important resource pool for exploring glycoside hydrolases with various special properties (Suleiman et al., 2020). The cellulases, xylanases and chitinases that can maintain catalytic activities under extreme temperature or pH conditions have the potential to be applied to biotechnological processes under harsh conditions (Lopez-Mondejar et al., 2019). According to the environmental parameters recorded in the EMP samples, we specifically focused on those widely distributed (top 10 in occurrence frequency) prokaryotic genera encoding cellulases, xylanases, or chitinases in high-temperature, low-temperature, acid, alkaline, and saline environments (Figure S7). We

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

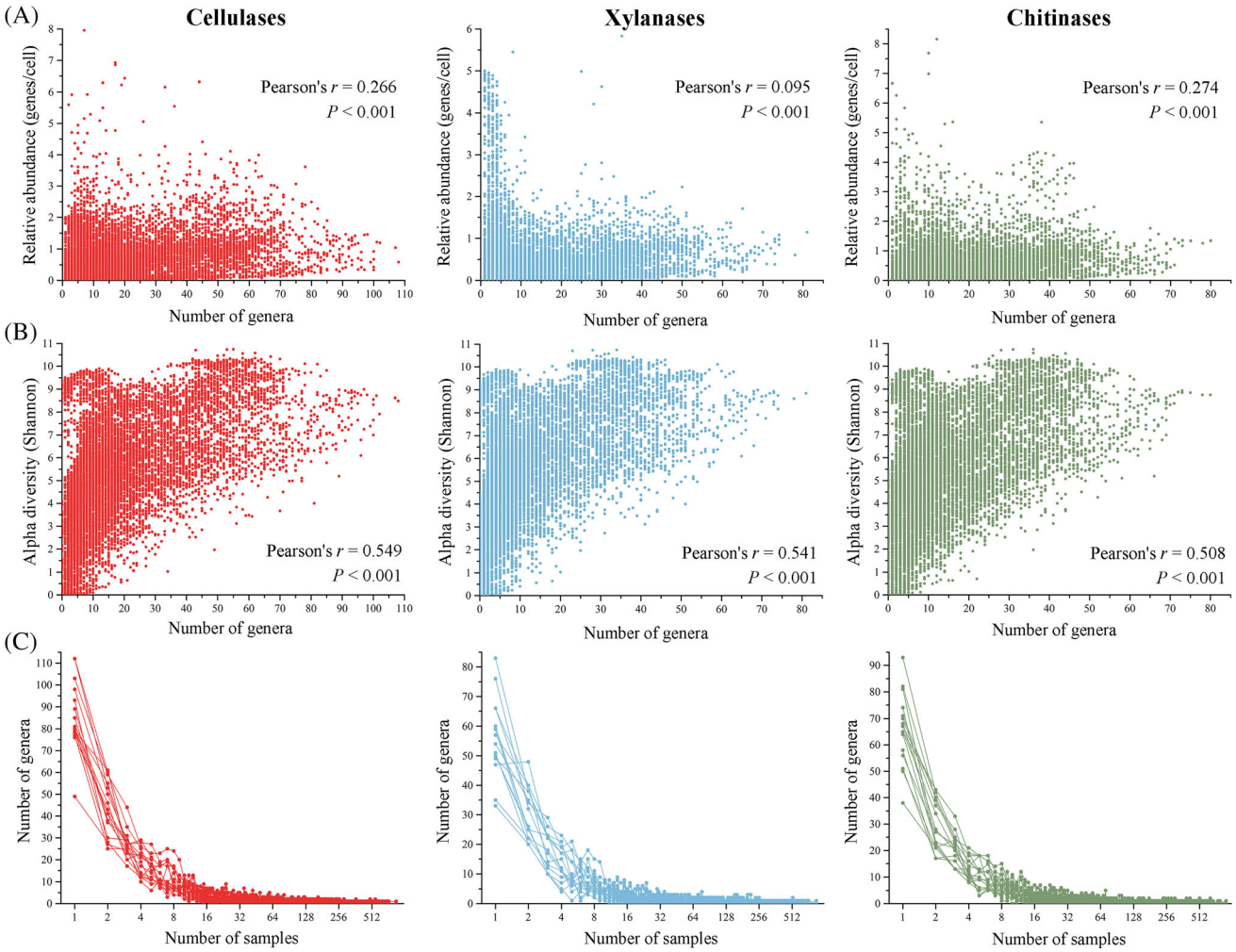


FIGURE 5 Taxonomic sources of cellulases, xylanases, and chitinases in prokaryotic communities. (A) Weak Pearson correlation between the taxonomic sources of degrading enzymes and the relative abundance of genes in the 10,000 Earth Microbiome Project (EMP) communities. (B) Strong Pearson correlation between the taxonomic sources and the community's alpha diversity. (C) The number of cross-community prokaryotic genera encoding cellulase, xylanase and chitinase genes decayed exponentially with increasing community numbers. Statistical analysis was performed for each of 16 EMPO_3 environments. The diversity of taxonomic sources encoding putative cellulases, xylanases and chitinases in the community is expressed as the number of genera, the gene abundance is calculated as genes/cell, and the alpha diversity is represented by the Shannon index. Dots represent individual prokaryotic communities. Source data are provided in Dataset S1 and Dataset S2.

found that there were 8, 10, and 6 endemic prokaryotic genera encoding cellulases, xylanases, and chitinases, respectively, that were common in high-temperature environments but did not belong to any of the common genera of EMPO environments. The common endemic prokaryotic genera for the three enzymes in saline environments also reached 7, 7, and 5, respectively. Thus, the enzyme-encoding prokaryotic genera in extreme environments are significantly different from those in the EMPO environment.

DISCUSSION

Prokaryotes are ubiquitously dispersed in diverse and extreme habitats on Earth and are considered to encode exceptionally diverse and multifunctional GHs, including enzymes for degrading complex polysaccharides such as cellulose, xylan and chitin (Li et al., 2021; Thapa et al., 2020). However, the environmental distribution of these complex polysaccharide-degrading enzymes, as well as potential microbial sources, has not been fully elucidated. Here, through large-scale alignments of 10,000 samples released by the EMP with more than 200,000 sequenced prokaryotic genomes, the distributions of putative cellulose-, xylan-, and chitin-degrading enzymes across global prokaryotic communities were thoroughly analysed.

The EMP solicited environmental samples and associated metadata across diverse environments that were collected by nearly 100 independent studies around the world and tracked individual sequences across diverse samples and studies with standardized environmental descriptors (Thompson et al., 2017).

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

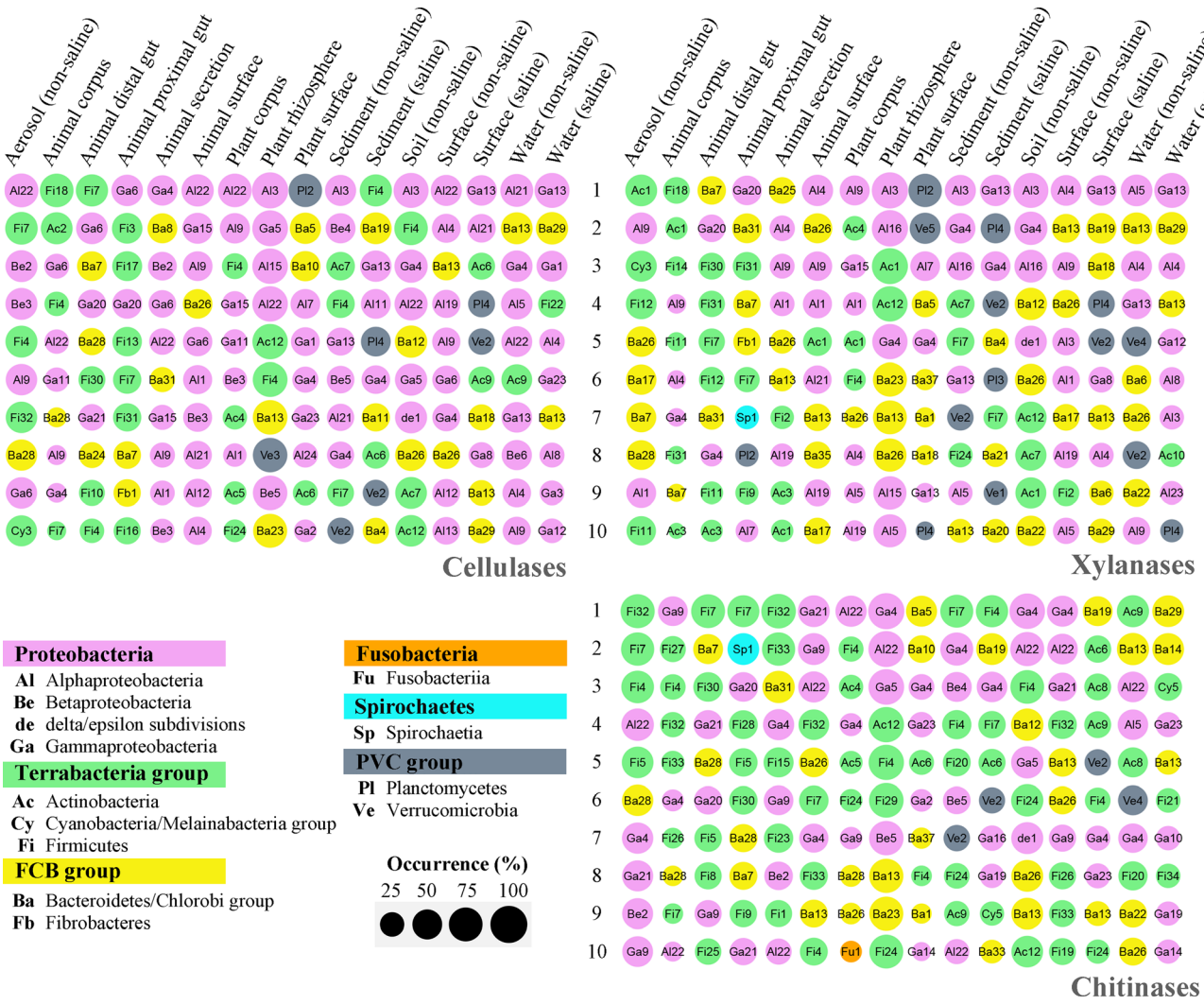


FIGURE 6 Common prokaryotic genera (top 10 in occurrence frequency) encoding putative cellulases, xylanases, and chitinases in various environments. Statistical analysis was performed for each of EMPO_3 environments. Distinct colours represent different taxonomic groups. The size of the circle indicates the occurrence frequency of the genera in all samples for each environment. Specific information of the prokaryotic genera can be found in Figure S6. Source data are provided in Dataset S2.

The EMP also used a reference-free method, Deblur, to remove suspected error sequences and provide single-nucleotide resolution 'sub-OTUs.' On this basis, we aligned the EMP OTUs and prokaryotic genomes and determined their corresponding relationship based strictly on 100% identity of the 16S rRNA-V4 region sequences. Consistent with our previously reported results (Zhang et al., 2020), the genetic information of prokaryotic communities has been revealed to a high degree, with the updated median proportions of genome-sequenced cells and taxa in different biomes of 40.2% (20.5%–88.1%) and 21.3% (11.4%–57.0%), respectively. Using the sequenced genomic information corresponding to the EMP samples, we found that almost all prokaryotic communities have the functional potentials to degrade cellulose, xylan and chitin and estimated the abundance of relevant genes in the communities normalized on a gene/cell basis. Considering that the sequenced genomes were mainly from cultivated strains, we also associated a fraction of uncultivated OTUs with the criterion of not less than 97% identity in the 16S rRNA-V4 region to estimate the gene abundance, and these two results did not show significant differences. The method of mapping EMP OTUs to prokaryotic genomes to obtain community genetic information on a large scale has been adopted in some recent studies, which provides a new way to investigate the global distribution of prokaryotic genes (Li et al., 2021; Machado et al., 2021; Zhang et al., 2021).

Complex polysaccharide-degrading enzymes are produced by various microorganisms dwelling in a broad range of environmental niches, including ruminant rumens, extreme ecosystems, soil, insect/termite guts, and marine or saline environments (Berlemont & Martiny, 2016; DeAngelis et al., 2010). These microorganisms are mainly bacteria and fungi, and the

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

enzymes produced by these organisms can function synergistically to break down biomass (Payne et al., 2015). Both bacteria and fungi may be important in the degradation of polysaccharides in natural ecosystems, such as in forest topsoil microbial communities, in which fungi are mainly involved in the decomposition of recalcitrant biopolymers in summer, while bacteria replace them for this role in winter (Zifcakova et al., 2017). However, compared with fungi, less attention has been given to the complex polysaccharide-degrading enzymes encoded by prokaryotes. Our survey of typical global ecosystems revealed that almost all prokaryotic communities have the functional potentials to degrade cellulose, xylan and chitin, and at least 933, 614, and 681 prokaryotic genera (mostly bacteria) encoding putative cellulases, xylanases, and chitinases, respectively, were detected. Considering the presence of vast uncultivated strains, these numbers are obviously still underestimated. We found that genera with the potential to degrade cellulose, xylan and chitin in prokaryotic communities are not only highly diverse but also highly differentiated. These three degrading enzymes were encoded by multiple taxonomically distinct prokaryotes within a community and by taxa that were highly distinct between communities, exhibiting obvious functional redundancy.

The identification of complex polysaccharide-degrading enzymatic systems is a prerequisite for understanding polysaccharide deconstruction by complex communities of microorganisms and for the identification of new enzymes for biotechnological applications (Nguyen et al., 2018). Over the past decades, researchers have isolated and biochemically characterized thousands of related proteins. Based on these studies, the CAZy database classifies nine major cellulase families, three major xylanase families, and three major chitinase families, and these families are mostly monospecific in function (Berlemont & Martiny, 2015; Lombard et al., 2014; Talamantes et al., 2016). The vast majority of sequences from the GH families involved in cellulose, xylan, and chitin deconstruction can be identified using Pfam- and CDD-based annotations (Nguyen et al., 2018). Combining CAZy annotations, functionally characterized proteins, and Pfam (or CDD) family classification, we identified putative cellulases, xylanases, and chitinases belonging to the 15 GH families mentioned above on a large scale in the NCBI reference sequence (RefSeq) genomes. For a few functionally polyspecific families, we also utilized BLAST sequence searches to perform detailed sequence comparisons to identify close relatives of characterized proteins. Furthermore, we combined RefSeq genomes with EMP sample OTUs to display putative cellulases, xylanases, and chitinases encoded by prokaryotic communities on a global scale. Notably, GHs together with other carbohydrate-active enzymes (e.g. carbohydrate esterases [CEs], lytic polysaccharide monooxigenases [LPMOs], and polysaccharide lyases [PLs]) support the key carbon-cycling processes across ecosystems (Drula et al., 2022; Lombard et al., 2014). This study focused on putative cellulases, xylanases and chitinases, but this pipeline may provide a general approach to identify these related proteins in the genomes and characterize their environmental distributions.

Complex polysaccharides such as cellulose are considered one of the most promising and sustainable natural resources for biofuels and feedstocks for the development of useful organo-compounds, feed additives, and other value-added products (Lee et al., 2012). Enhanced cellulose or xylan processing is an important way to explore novel or more effective biocatalysts (Bala & Singh, 2019; Bugg et al., 2020). The inherent catalytic attributes, such as non-stability, inefficiency, and end-product inhibition of currently available cellulolytic enzymes, limit their utilization at an industrial scale (Champreda et al., 2019). The search for robust inherent and stable cellulolytic enzymes from diverse and unique habitats could be a primary strategy for the successful utilization of abundantly available cellulosic substrates (Thapa et al., 2019). Fungi have been traditionally exploited as producers of lignocellulose-degrading enzymes, while bacteria are now widely exploited and seen as good candidates for future strategies due to their wide functional diversity and versatility (Obeng et al., 2017). Despite the fact that only a small percentage of bacteria have been cultivated to date, our work has revealed the enormous and underexploited degradation potential of bacteria and their specific enzymes to degrade complex polysaccharides such as cellulose in natural ecosystems. Prokaryotic communities based on their habitats can be diverse and novel sources of highly efficient and robust complex polysaccharide-degrading enzymatic systems.

In conclusion, this study provides a panoramic view of the functional potentials of global prokaryotic communities for degrading cellulose, xylan, and chitin under various environmental conditions, showing the exceptionally diverse composition of enzymatic systems and unveiling the prospective sources of microbes encoding complex polysaccharide-degrading enzymes that are critical to answer the specific ecological question of by whom, where, and how complex polysaccharides such as cellulose are degraded in the environment.

AUTHOR CONTRIBUTIONS

Dan-dan Li: Conceptualization (equal); formal analysis (equal); funding acquisition (supporting); investigation (equal); methodology (equal); resources (equal); validation (equal); visualization (equal); writing – original draft (equal). **Zheng Zhang:** Conceptualization (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); validation (equal); visualization

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(equal); writing – review and editing (equal). **Jia-ning Wang:** Methodology (equal); resources (supporting); validation (supporting). **Peng Zhang:** Data curation (supporting); validation (supporting). **Ya Liu:** Data curation (supporting); funding acquisition (supporting); validation (supporting). **Yue-Zhong Li:** Conceptualization (equal); funding acquisition (equal); project administration (equal); supervision (lead); writing – review and editing (equal).

ACKNOWLEDGEMENTS

This work was financially supported by the National Natural Science Foundation of China (32070030), the National Key Research and Development Program (2018YFA0900400 and 2018YFA0901704), the Key Research and Development Program of Shandong Province (2018GSF121015) (Yue-zhong Li), the National Natural Science Foundation of China (32270073), the Science & Technology Fundamental Resources Investigation Program (2022FY101100), the Natural Science Foundation of Jiangsu Province (BK20190199), the Fundamental Research Funds of Shandong University (2020GN113) (Zheng Zhang), and the China Postdoctoral Science Foundation (2021M701987) (Dan-dan Li) and (2021M700084) (Ya Liu).

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The main data supporting the findings of this study are available within the article and its Supplementary Information. All other data supporting the findings of this study are available from the corresponding authors upon reasonable request.

ORCID

Zheng Zhang https://orcid.org/0000-0001-9971-6006

REFERENCES

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J. T., Zech Xu, Z. et al. (2017) Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems*, 2, 2.

Bala, A. & Singh, B. (2019) Cellulolytic and xylanolytic enzymes of thermophiles for the production of renewable biofuels. *Renewable Energy*, 136, 1231–1244.

Berlemont, R. & Martiny, A.C. (2015) Genomic potential for polysaccharide deconstruction in bacteria. *Applied and Environmental Microbiology*, 81, 1513–1519.

Berlemont, R. & Martiny, A.C. (2016) Glycoside hydrolases across environmental microbial communities. *PLoS Computational Biology*, 12, e1005300.

Bomble, Y.J., Lin, C.Y., Amore, A., Wei, H., Holwerda, E.K., Ciesielski, P.N. et al. (2017) Lignocellulose deconstruction in the biosphere. *Current Opinion in Chemical Biology*, 41, 61–70.

Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N. et al. (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, 41, W29–W33.

Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews. Microbiology*, 12, 168–180.

Bugg, T.D.H., Williamson, J.J. & Rashid, G.M.M. (2020) Bacterial enzymes for lignin depolymerisation: new biocatalysts for generation of renewable chemicals from biomass. *Current Opinion in Chemical Biology*, 55, 26–33.

Buttigieg, P.L., Pafilis, E., Lewis, S.E., Schildhauer, M.P., Walls, R. L. & Mungall, C.J. (2016) The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of Biomedical Semantics*, 7, 57.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421.

Champreda, V., Mhuantong, W., Lekakarn, H., Bunterngsook, B., Kanokratana, P., Zhao, X.Q. et al. (2019) Designing cellulolytic enzyme systems for biorefinery: from nature to application. *Journal of Bioscience and Bioengineering*, 128, 637–654.

Chen, W., Jiang, X. & Yang, Q. (2020) Glycoside hydrolase family 18 chitinases: the known and the unknown. *Biotechnology Advances*, 43, 107553.

Cragg, S.M., Beckham, G.T., Bruce, N.C., Bugg, T.D., Distel, D.L., Dupree, P. et al. (2015) Lignocellulose degradation mechanisms across the tree of life. *Current Opinion in Chemical Biology*, 29, 108–119.

DeAngelis, K.M., Gladden, J.M., Allgaier, M., D'haeseleer, P., Fortney, J.L., Reddy, A. et al. (2010) Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities. *Bioenergy Research*, 3, 146–158.

Dragone, G., Kerssemakers, A.A.J., Driessen, J., Yamakawa, C.K., Brumano, L.P. & Mussatto, S.I. (2020) Innovation and strategic orientations for the development of advanced biorefineries. *Bioresource Technology*, 302, 122847.

Drula, E., Garron, M.L., Dogan, S., Lombard, V., Henrissat, B. & Terrapon, N. (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research*, 50, D571–D577.

Elleuche, S., Schafers, C., Blank, S., Schroder, C. & Antranikian, G. (2015) Exploration of extremophiles for high temperature biotechnological processes. *Current Opinion in Microbiology*, 25, 113–119.

Fabian, J., Zlatanovic, S., Mutz, M. & Premke, K. (2017) Fungal-bacterial dynamics and their contribution to terrigenous carbon turnover in relation to organic matter quality. *The ISME Journal*, 11, 415–425.

Gao, Y., Lipton, A.S., Wittmer, Y., Murray, D.T. & Mortimer, J.C. (2020) A grass-specific cellulose-xylan interaction dominates in sorghum secondary cell walls. *Nature Communications*, 11, 6081.

García-Palacios, P., McKie, B.G., Handa, I.T., Frainer, A. & Hättenschwiler, S. (2016) The importance of litter traits and decomposers for litter decomposition: a comparison of aquatic and terrestrial ecosystems within and across biomes. *Functional Ecology*, 30, 819–829.

Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W. et al. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 315, 804–807.

Lee, C.C., Kibblewhite, R.E., Wagschal, K., Li, R., Robertson, G.H. & Orts, W.J. (2012) Isolation and characterization of a novel GH67 alpha-glucuronidase from a mixed culture. *Journal of Industrial Microbiology & Biotechnology*, 39, 1245–1251.

Li, D.D., Wang, J.L., Liu, Y., Li, Y.Z. & Zhang, Z. (2021) Expanded analyses of the functional correlations within structural classifications of glycoside hydrolases. *Computational and Structural Biotechnology Journal*, 19, 5931–5942.

Li, L.L., Taghavi, S., McCorkle, S.M., Zhang, Y.B., Blewitt, M.G., Brunecky, R. et al. (2011) Bioprospecting metagenomics of

14622920, 2023, 2, Downloaded from https://enviromicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

decaying wood: mining for new glycoside hydrolases. *Biotechnology for Biofuels*, 4, 23.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42, D490–D495.
Lopez-Mondejar, R., Algora, C. & Baldrian, P. (2019) Lignocellulolytic systems of soil bacteria: a vast and diverse toolbox for biotechnological conversion processes. *Biotechnology Advances*, 37, 107374.
Lopez-Mondejar, R., Zuhlke, D., Vetrovsky, T., Becher, D., Riedel, K. & Baldrian, P. (2016) Decoding the complete arsenal for cellulose and hemicellulose deconstruction in the highly efficient cellulose decomposer Paenibacillus O199. *Biotechnology for Biofuels*, 9, 104.
Machado, D., Maistrenko, O.M., Andrejev, S., Kim, Y., Bork, P., Patil, K.R. et al. (2021) Polarization of microbial communities between competitive and cooperative metabolism. *Nature Ecology & Evolution*, 5, 195-+.
Nguyen, S.T.C., Freund, H.L., Kasanjian, J. & Berlemont, R. (2018) Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy. *Applied Microbiology and Biotechnology*, 102, 1629–1637.
Obeng, E.M., Adam, S.N.N., Budiman, C., Ongkudon, C.M., Maas, R. & Jose, J. (2017) Lignocellulases: a review of emerging and developing enzymes, systems, and practices. *Bioresources and Bioprocessing*, 4, 16.
Payne, C.M., Knott, B.C., Mayes, H.B., Hansson, H., Himmel, M.E., Sandgren, M. et al. (2015) Fungal cellulases. *Chemical Reviews*, 115, 1308–1448.
Petridis, L. & Smith, J.C. (2018) Molecular-level driving forces in lignocellulosic biomass deconstruction for bioenergy. *Nature Reviews Chemistry*, 2, 382–389.
Pruitt, K.D., Tatusova, T. & Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 35, D61–D65.
Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A. et al. (2006) The path forward for biofuels and biomaterials. *Science*, 311, 484–489.
Sayers, E.W., Cavanaugh, M., Clark, K., Pruitt, K.D., Schoch, C.L., Sherry, S.T. et al. (2021) GenBank. *Nucleic Acids Research*, 49, D92–D96.
Scheller, H.V. & Ulvskov, P. (2010) Hemicelluloses. *Annual Review of Plant Biology*, 61, 263–289.
Schmidt, L.D. & Dauenhauer, P.J. (2007) Chemical engineering: hybrid routes to biofuels. *Nature*, 447, 914–915.
Schubert, C. (2006) Can biofuels finally take center stage? *Nature Biotechnology*, 24, 777–784.
Stuart, R.K., Mayali, X., Lee, J.Z., Craig Everroad, R., Hwang, M., Bebout, B.M. et al. (2016) Cyanobacterial reuse of extracellular organic carbon in microbial mats. *The ISME Journal*, 10, 1240–1251.
Suleiman, M., Kruger, A. & Antranikian, G. (2020) Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnology for Biofuels*, 13, 153.
Talamantes, D., Biabini, N., Dang, H., Abdoun, K. & Berlemont, R. (2016) Natural diversity of cellulases, xylanases, and chitinases in bacteria. *Biotechnology for Biofuels*, 9, 133.
Thapa, S., Li, H., OHair, J., Bhatti, S., Chen, F.C., Al Nasr, K. et al. (2019) Biochemical characteristics of microbial enzymes and their significance from industrial perspectives. *Molecular Biotechnology*, 61, 579–601.
Thapa, S., Mishra, J., Arora, N., Mishra, P., Li, H., O'Hair, J. et al. (2020) Microbial cellulolytic enzymes: diversity and biotechnology with reference to lignocellulosic biomass degradation. *Reviews in Environmental Science and Bio/Technology*, 19, 621–648.
Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J. et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551, 457–463.
Tiwari, R., Nain, L., Labrou, N.E. & Shukla, P. (2018) Bioprospecting of functional cellulases from metagenome for second generation biofuel production: a review. *Critical Reviews in Microbiology*, 44, 244–257.
Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E. et al. (2009) A core gut microbiome in obese and lean twins. *Nature*, 457, 480–484.
Zhang, Z., Liu, Y., Zhang, P., Wang, J.N., Li, D.D. & Li, Y.Z. (2021) PAAR proteins are versatile clips that enrich the antimicrobial weapon arsenals of prokaryotes. *mSystems*, 6, e00953-21.
Zhang, Z., Wang, J., Wang, J., Wang, J. & Li, Y. (2020) Estimate of the sequenced proportion of the global prokaryotic genome. *Microbiome*, 8, 134.
Zifcakova, L., Vetrovsky, T., Lombard, V., Henrissat, B., Howe, A. & Baldrian, P. (2017) Feed in summer, rest in winter: microbial carbon utilization in forest topsoil. *Microbiome*, 5, 122.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li, D., Zhang, Z., Wang, J., Zhang, P., Liu, Y. & Li, Y. (2023) Estimate of the degradation potentials of cellulose, xylan, and chitin across global prokaryotic communities. *Environmental Microbiology*, 25(2), 397–409. Available from: https://doi.org/10.1111/1462-2920.16290

14622920, 2023, 2, Downloaded from https://environmicro-journals.onlinelibrary.wiley.com/doi/10.1111/1462-2920.16290 by Los Alamos National Laboratory, Wiley Online Library on [07/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License