



# Genomic and transcriptomic landscapes and evolutionary dynamics of molluscan glycoside hydrolase families with implications for algae-feeding biology



Jing Wang<sup>a,b,1</sup>, Hongwei Yu<sup>a,1</sup>, Lijie Yao<sup>a,1</sup>, Yuli Li<sup>a,b,\*</sup>, Fuyun Liu<sup>a</sup>, Wentao Han<sup>a</sup>,  
Cong Cui<sup>a</sup>, Jingjie Hu<sup>a,d</sup>, Zhenmin Bao<sup>a,c,d</sup>, Shi Wang<sup>a,b,d,\*</sup>

<sup>a</sup>MOE Key Laboratory of Marine Genetics and Breeding and Sars-Fang Centre, Ocean University of China, Qingdao 266003, China

<sup>b</sup>Laboratory for Marine Biology and Biotechnology, Pilot Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

<sup>c</sup>Laboratory for Marine Fisheries Science and Food Production Processes, Pilot Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

<sup>d</sup>Laboratory of Tropical Marine Germplasm Resources and Breeding Engineering, Sanya Oceanographic Institution, Ocean University of China, Sanya 572000, China

## ARTICLE INFO

### Article history:

Received 2 July 2020

Received in revised form 10 September 2020

Accepted 11 September 2020

Available online 28 September 2020

### Keywords:

Mollusc  
Glycoside hydrolase  
Gene family expansion  
Adaptive evolution  
Algae feeding

## ABSTRACT

The hydrolysis of sugar-containing compounds by glycoside hydrolases (GHs) plays essential roles in many major biological processes, but to date our systematic understanding of the functional diversity and evolution of GH families remains largely limited to a few well-studied terrestrial animals. Molluscs represent the largest marine phylum in the animal kingdom, and many of them are herbivorous that utilize algae as a main nutritional source, making them good subjects for studying the functional diversity and adaptive evolution of GH families. In the present study, we conducted genome-wide identification and functional and evolutionary analysis of all GH families across major molluscan lineages. We revealed that the remarkable expansion of the GH9, GH10, GH18 and GH20 families and the wide adoption of carbohydrate-binding modules in molluscan expanded GH families likely contributed to the efficient hydrolysis of marine algal polysaccharides and were involved in the consolidation of molluscan algae-feeding habits. Gene expression and network analysis revealed the hepatopancreas as the main organ for the prominent expression of approximately half of the GH families (well corresponding to the digestive roles of the hepatopancreas) and key or hub GHs in the coexpression gene network with potentially diverse functionalities. We also revealed the evolutionary signs of differential expansion and functional divergence of the GH family, which possibly contributed to lineage-specific adaptation. Systematic analysis of GH families at both genomic and transcriptomic levels provides important clues for understanding the functional divergence and evolution of GH gene families in molluscs in relation to their algae-feeding biology.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Carbohydrates including monosaccharides, oligosaccharides and polysaccharides are the most abundant and widely distributed organic compounds on Earth. Glycoside hydrolases (GHs) are the key enzymes participating in carbohydrate metabolism, which is a prerequisite for organisms to obtain energy [1–3]. Except for some Archaea and parasitic single-cell eukaryotes, GH genes are

widely present in almost all organisms [4,5]. They are widely involved in the hydrolysis of sugars and glycoconjugates and play essential roles in many important biological processes, such as energy acquisition and metabolism, degradation of cellular components, recognition and adhesion between cells and other physiological and biochemical reactions [6,7]. In recent decades, GHs have become hot research subjects in various fields of biology, involving various groups of microorganisms, plants and animals [1,8–11].

Molluscs are the largest marine phylum in the animal kingdom. Many of them are herbivorous that utilize algae as their main nutritional source, which could be a characteristic of molluscan ancestors and makes these organisms good subjects for studying GH family evolution and adaptive roles [12,13]. The polysaccha-

\* Corresponding authors at: MOE Key Laboratory of Marine Genetics and Breeding and Sars-Fang Centre, Ocean University of China, Qingdao 266003, China.

E-mail addresses: [liyuli@ouc.edu.cn](mailto:liyuli@ouc.edu.cn) (Y. Li), [swang@ouc.edu.cn](mailto:swang@ouc.edu.cn) (S. Wang).

<sup>1</sup> These authors contributed equally to this work.

rides in marine algae, including cellulose, hemicellulose, chitin and laminarin, are the main carbohydrate source of herbivorous molluscs, and these compounds are substantially different from those associated with terrestrial higher plants, which include mainly cellulose and starch [14–24]. As a main source of energy, important cellular components and signaling molecules, the differences in polysaccharide types are closely related to the diet as well as morphological and physiological functions of the animals that consume these compounds. For example, most molluscs can encode cellulases to degrade and utilize cellulose [25,26], while vertebrates and most ecdysozoans are generally considered unable to encode cellulases by themselves, and their cellulose degradation depends on intestinal commensal microorganisms [27–29]. The organic matrix of the “skeleton” of ecdysozoans and molluscs is composed of chitin, and chitinase is involved in the metabolism of their “skeleton” [29,30]. While previous studies have reported multiple GHs (including laminarinase, cellulase, chitinase, amylase, mannanase, and fucosidases) in molluscs [11,31,32], their genomic distribution, expressional and evolutionary characteristics across major molluscan lineages remain poorly understood. Further study could provide important clues for understanding the functional divergence and adaptive roles of GHs in molluscs.

Due to the rapid development of advanced sequencing technologies in recent years, the molluscan genomics field has witnessed significant achievements by providing invaluable genome resources and unprecedented opportunities for addressing many important biological questions [33–39]. In this study, we conducted genome-wide identification, expressional and evolutionary analysis of all GH families in major molluscan lineages. Our findings provide important clues for understanding the evolution of GH gene families in molluscs in relation to their algae-feeding biology.

## 2. Methods

### 2.1. Genome-wide identification, classification and annotation of GH family genes

To identify the GH family genes, we first obtained all the coding genes of 7 molluscs (*Chlamys farreri*, *Crassostrea gigas*, *Pinctada fucata*, *Bathymodiolus platifrons*, *Modiolus philippinarum*, *Lottia gigantea*, *Octopus bimaculoides*) and 14 other metazoan species including the deuterostomes *Homo sapiens*, *Danio rerio*, *Branchiostoma floridae*, *Apostichopus japonicus*, the protostomes *Lingula anatina*, *Capitella teleta*, *Helobdella robusta*, *Tribolium castaneum*, *Cryptotermes secundus*, *Drosophila melanogaster*, *Daphnia pulex*, *Caenorhabditis elegans* and the non-bilaterians *Nematostella vectensis*, *Acropora digitifera* (Table S1). Since functional domains (catalytic domains) were the most important features for identification and classification of GH families [3,40], we used the command “hmmscan” in the software HMMER (v3.1b2) to search for GH domains, against the Pfam-A database [41] with the e-value threshold  $1e^{-5}$ . The reliability of the identified GH family genes was manually checked by their functional annotations with the expected hydrolase activity derived from functional annotation database, including the Carbohydrate active enzyme database (CAZy) [4], Uniprot database [42], the NCBI non-redundant proteins (Nr) database [43], Gene ontology (GO) [44] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database [45], according to the sequence similarities. The genomic distributions of GH18 family genes in the seven molluscan genomes were retrieved from genome annotation files released by previous studies (Table S1) [33–35,37,46,47]. The conserved domain structures of all molluscan GH family genes were identified using the Batch CD-Search tool in NCBI database by searching against the Pfam

Database (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>).

### 2.2. Phylogenetic and evolutionary analysis of GH families

To construct the phylogenetic trees of GH family genes, the sequences of the longest catalytic domain (>120 amino acids) for each gene were first obtained. The multiple sequence alignments were performed using the MEGA 6 software with default parameters [48]. The Neighbor-Joining (NJ) trees were constructed using the p-distance method with 2000 bootstrap replicates. The active centre of the catalytic domain of each family was extracted based on the sequence alignment according to previous studies [7,49] and the sequence logo was generated using WebLogo (<http://weblogo.berkeley.edu/>) [50].

### 2.3. Expression analysis of the GH family genes of *C. farreri* and *C. gigas*

The transcriptome data of adult tissues/organs and developmental stages of scallop *C. farreri* and oyster *C. gigas* used in the expression analysis were downloaded from NCBI database released by previous studies [35,37]. High-quality sequencing reads were retained by requiring 80% base with a quality value greater than 20. Then the high-quality reads were mapped to the corresponding reference genomes using the software STAR (version 2.4.1c) [51]. Reads mapped to exonic regions were counted for corresponding genes using the software featureCounts (v1.6.3) [52] to calculate the gene expression level in term of TPM (Transcripts Per Kilobase of exon model per Million mapped reads) values. The expression patterns of GH genes and GH families (the sum of the TPM values of GH genes in each family) of the *C. farreri* and the *C. gigas* were shown in heatmaps. For a given GH family, its highly expressed tissues/organs were defined as those with TPM values higher than 1.5 folds of the average TPM of the other tissues/organs.

### 2.4. Gene coexpression network analysis of the GH family genes

To ensure the reliability of network construction, the co-expression network was constructed using the expression profile of *C. farreri*, for which multiple replications of the main tissues/organs are available but not in *C. gigas*. A total of 22,228 genes that expressed at least in one tissues/organs of *C. farreri* with the TPM values higher than two were used for weighted gene co-expression network analysis (WGCNA) [53]. The coexpression network was constructed using the WGCNA package in R [54] with the parameters of ‘softPower = 15, cutHeight = 0.99 and minModuleSize = 450’. Each gene module was labeled in unique color and the unassigned genes were labeled in grey. The intramodular connectivity value ( $K_{within}$ ) was used to measure a gene’s connection strength to other genes in the specified module (i.e., the hubness of a gene) [54]. The over-representation analysis of the GH family genes was performed for each module using a hypergeometric test with p-values adjusted by the Benjamini-Hochberg method for multiple-test correction [54,55]. To evaluate the expression pattern of each module among different tissues/organs, TPM values of each gene were firstly normalized across the different tissues/organs using the scale function in R (version 3.6.1), and the expression level of each gene cluster was represented in the median value. GO and KEGG pathway enrichment analyses of the gene modules were performed using the EnrichPipeline with the FDR cutoff 0.05 [56,57]. Cytoscape [58] was used for visualization of the co-expression networks.

### 3. Results

#### 3.1. Genomic landscape and phylogenetic analysis of molluscan GH families

We identified the full sets of GH family genes from the genomes of seven molluscs covering major molluscan lineages, in comparison with those identified from 14 additional animal species representative of the major groups in the animal kingdom. Within the molluscan group, the number of GH genes ranged from 70 to 167 (Fig. 1a; Table 1). These genes belonged to 26 different GH families (Fig. 1a; Table 1), of which 17 families were present in all seven molluscs, and only a few families were unique to certain species (Fig. S1a). Across-Metazoa comparison revealed that the four groups, Mollusca, Annelida, Ecdysozoa, and Deuterostomia, shared 17 GH families, accounting for 55% of all GH families. Compared with Annelida and Ecdysozoa, Mollusca shared more GH families with Deuterostomia (Fig. S1b), presumably a reflection of their conserved genomes [37,59]. Compared with other animal groups, a notable feature of molluscs was that both GH9 (endoglucanases) and GH10 (endo-1,4- $\beta$ -xylosidase) families were present and expanded in most herbivorous molluscs (Fig. 1; Table 1). The expansion of these two families was prominent in gastropods and bivalves (5–14 for GH9 and 2–13 for GH10; Fig. 1). In contrast to herbivorous molluscs, the cephalopod contained only one GH9 and one GH10 family member. Compared with the shallow-water mussel *M. philippinarum*, the deep-sea mussel *B. platifrons* showed much reduced numbers of GH9 and GH10 family genes. Another remarkable feature of the molluscan GH family was the expansion of GH18 (chitinase) and GH20 (beta-hexosaminidase), which was also common in ecdyzoans (Fig. 1a). Although most GH families consist of more than one kind of enzymatic activities, approximately 77% molluscan GH families were identified as having only one enzymatic activity, including even the expanded GH9, GH10 and GH20 (Table 1; Fig. S2). For molluscan GH18, chitinase (EC 3.2.1.14) and di-N-acetylchitinase (also known as chitosanase, EC 3.2.1.-) were identified, but only the chitinase genes were expanded in molluscs (Table 1; Fig. 2b-c). For the other GH families with two or more predicted enzymatic activities (i.e., GH2, GH13, GH30, GH31 and GH38), family members were divided into several subclades corresponding to different kinds of enzymes with different catalytic unit sequences (Table 1; Fig. S2-3). Besides this major class of carbohydrate-active enzyme, we also identified other enzymes related to polysaccharide degradation in the twenty representative metazoan species, such as alginate lyase (family PL14) and Lytic Polysaccharide Mono-Oxygenase (LPMO) (Table S2). We found that the expansion of PL14 may be a unique genomic feature belonging to gastropods, while LPMO showed remarkable expansion in molluscan genomes.

#### 3.2. Functional domain diversity and its association with GH18 family evolution

Most of the molluscan GH families (42%) possessed a simple structural organization with only catalytic domains (Fig. 1b). Except for the catalytic domains, 31% of molluscan GH families had non-catalytic subdomains with  $\beta$ -sandwich fold (GH30\_C, GH59\_M, GH63\_N, GH81\_N, NtCtMGAM\_N, Gal\_mutarotas\_2), zincin-like fold (GH20b) or jelly-roll fold (GH2\_N) structures. Some of the GH2 (28.0%), GH9 (8.2%), GH13 (9.5%), GH16 (33.3%), GH18 (34.9%), GH20 (45.3%) and GH81 (25.0%) genes of the molluscan GH family contained one or more carbohydrate-binding modules (CBMs) located at the N/C-terminus or interspersed among catalytic domains, of which the most variable domain structures were present in the GH18 family (Fig. 1b). Moreover, some GH genes

also had an auxiliary domain, such as the fibronectin type 3-like (Fn3-like) domain in the GH3 family, the trefoil (P-type) domain in the GH31 family, the alpha mannosidase middle domain (Alpha-mann\_mid) in the GH38 family and the solute carrier family 3 member 2N-terminus domain (SLC3A2\_N) in the GH13 family (Fig. 1b). It is worth mentioning that GH3 has two kinds of catalytic domains: a triose phosphate isomerase (TIM) barrel-like domain (GH3\_N) and an  $\alpha/\beta$  sandwich domain (GH3\_C) with distinct catalytic residues (the catalytic nucleophile/base Asp in GH3\_N and the catalytic proton donor Glu in GH3\_C).

Most of the GH18 genes showed expansion and resided in tandem arrays (Fig. 2a; Fig. S4), while the other expanded families (GH9, GH10 and GH20) were mostly dispersed in the genome. Based on phylogenetic analysis, GH18 family genes in the phylum of Mollusca were classified into nine separate clusters (Group A-I) (Fig. 2b). The hydrolase activities of the GH18 members are identified as chitinases (EC:3.2.1.14) except for group A, which is di-N-acetylchitinase (Fig. 2b). For groups A-E, most genes had only catalytic domains but lacked CBMs. In contrast, genes from groups F-I mostly had an N-terminal catalytic domain followed by one or more CBMs (Fig. 2b). The conserved sequence logos showed the active centres of the molluscan GH18 catalytic domains conserved with the amino acid residues FDGLD(L/M)DWE(Y/F)P, and the most critical residues (glutamate residue E and aspartic acid residue D) in the GH18 catalytic domain were conserved among all groups except group A with LDGXNXDXEX (Fig. 2b-c). In addition, compared with group F-I, the active centres of groups B-E were less conserved for leucine and tyrosine residues (Fig. 2c). The GH18 genes in Group G were absent in the bivalves. As a group specific to Gastropoda and Cephalopoda, the active centres of Group G were mostly different from those of other CBM-containing groups (i.e., FDGLDMDWEFP instead of FDGLDLDEYFP) (Fig. 2b-c).

#### 3.3. Comprehensive transcriptome profiling of molluscan GH families

Gene expression profiling facilitates the understanding of the function and evolution of GH families. Although many molluscan genomes have been sequenced, comprehensive transcriptome data remain largely scarce for molluscs. Here, we chose *C. farreri* and *C. gigas*, the two molluscs with the best availability of comprehensive transcriptome datasets (Table S3) [35,36], for analyzing the expression profiles of GH genes in adult tissues/organs and across various developmental stages. Molluscan GH family genes generally showed diverse expression patterns across adult tissues/organs and developmental stages (Fig. 3a-b; Fig. S5). In particular, approximately half of the GH families showed highly restricted expression in the hepatopancreas (Fig. 3a-b; Fig. S5). During development, the highly expressed GH families in the hepatopancreas started to be expressed after D-larval period and reached peak expression levels in the later umbo larval period (Fig. 3a-b). The composition of highly expressed GH families in the hepatopancreas was similar between scallop and oyster (Table 1), probably reflecting similar preferences for algae between the two molluscs. The highly expressed GH families in the hepatopancreas of *C. farreri*, were also previously reported as the main types of scallop polysaccharide hydrolases through mass spectrometry analysis [31].

#### 3.4. Coexpression gene network analysis of molluscan GH families

To identify coexpression modules and the key GH genes in the hydrolysis of polysaccharides and other biological processes, we further investigated the gene coexpression network of GH family genes across the main tissues/organs of *C. farreri* using the weighted gene coexpression network analysis (WGCNA) approach [54]. Network construction revealed 14 gene modules (M1-14) with different expression patterns among the main tissues/organs



**Table 1**  
Summary of function, substrate, expression of the GH families in 7 molluscs.

GH	Family member							Function	Enzymatic Activities	Substrate	Highly Expressed Tissue/Organs
	<i>C. farreri</i> (CF)	<i>C. gigas</i> (CG)	<i>P. fucata</i> (PF)	<i>B. platifrons</i> (BP)	<i>M. philippinarum</i> (MP)	<i>L. gigantea</i> (LG)	<i>O. bimaculoides</i> (OB)				
GH1	6	1	4	5	10	7	1	lactase-phlorizin hydrolase	EC:3.2.1.108; EC3.2.1.62	lactose, phlorizin	Hepatopancreas, Fgonad(CG)
GH2	2	4	5	3	3	6	2	$\beta$ -glucuronidase; $\beta$ -mannosidase	EC:3.2.1.31; EC:3.2.1.25	glucuronoside, mannosides	Hepatopancreas(CF), Hemolymph(CF), Kidney(CF), Hemolymph(CG)
GH3	2	2	4	4	5	21	2	$\beta$ -glucosidase	EC:3.2.1.21	$\beta$ -D-glucoside	Hepatopancreas, Kidney(CF), Hemolymph(CG)
GH9	7	5	11	9	14	13	1	endoglucanase	EC:3.2.1.4	cellulose	Hepatopancreas
GH10	8	3	7	2	6	13	1	<i>endo</i> -1,4- $\beta$ -xylanase	EC 3.2.1.8	xylan	Hepatopancreas, Mgonad(CG)
GH13	10	15	8	12	9	10	4	$\alpha$ -amylase; $\alpha$ -glucosidase; 1,4- $\alpha$ -glucan branching enzyme	EC:3.2.1.1; EC:3.2.1.20; EC:2.4.1.18	starch, etc	Kidney(CF), Hepatopancreas(CG)
GH16	4	4	2	3	4	10	0	laminarinase	EC:3.2.1.39	laminarin, etc	Hepatopancreas
GH18	23	18	20	33	31	24	24	chitinase; di-N-acetylchitobiase	EC:3.2.1.14; EC:3.2.1.-	chitin; chitobiose	Hepatopancreas, Mantle(CF)
GH20	18	19	13	18	9	11	7	$\beta$ -N-acetylhexosaminidase	EC:3.2.1.52	hexosamine	Smooth_muscle(CF), Foot(CF), Hepatopancreas(CG)
GH26	1	0	1	0	0	4	0	<i>endo</i> -1,4- $\beta$ -mannosidase	EC:3.2.1.78	mannans, galactomannans, glucomannans	Hepatopancreas(CF)
GH28	0	0	0	0	0	0	1	polygalacturonase	EC:3.2.1.15	pectate and other galacturonans	/
GH30	3	4	3	3	3	2	1	glucosylceramidase; $\beta$ -1,6-glucanase	EC:3.2.1.45; EC:3.2.1.75	glucosylceramide, lutean, pustulan, $\beta$ -glucan	Hepatopancreas
GH31	10	11	12	11	19	19	7	$\alpha$ -glucosidase; maltase-glucoamylase; sucrase-isomaltase; mannosyl-oligosaccharide alpha-1,3-glucosidase; sulfoquinovosidase; myogenesis-regulating glycosidase	EC:3.2.1.20; EC:3.2.1.3; EC:3.2.1.48; EC:3.2.1.10; EC:3.2.1.207; EC:3.2.1.199; EC:3.2.1.-	Broad Substrate (maltose, starch, etc)	Hepatopancreas
GH35	6	4	3	4	3	2	2	$\beta$ -galactosidase	EC:3.2.1.23	lactose, etc	Hepatopancreas, Kidney(CF)
GH38	9	10	10	11	9	6	5	$\alpha$ -mannosidase; mannosyl-oligosaccharide 1,3-1,6- $\alpha$ -mannosidase	EC:3.2.1.24; EC:3.2.1.114	mannoside	Hemolymph(CF)
GH39	1	1	1	1	1	1	1	$\alpha$ -L-iduronidase	EC:3.2.1.76	dermatan sulfate	Gill, Hemolymph(CF), Fgonad(CF), Hepatopancreas(CG)
GH47	2	2	2	2	4	2	2	$\alpha$ -mannosidase	EC:3.2.1.113	mannosides	Hemolymph(CF), Kidney(CF)
GH56	1	2	0	1	1	2	2	hyaluronidase	EC:3.2.1.35	hyaluronan	Hepatopancreas(CF), Kidney(CF), Labial_palp(CG)
GH59	1	1	1	2	2	6	1	galactocerebrosidase	EC:3.2.1.46	galactosylceramide, etc	Hemolymph, Kidney (CG), Gill(CF), Adductor_muscle(CG), Mgonad(CG)
GH63	1	1	2	1	1	1	1	mannosyl-oligosaccharide glucosidase	EC:3.2.1.106	Specific oligosaccharide precursor	Fgonad, Foot(CF), Mgonad(CF)
GH65	2	2	1	2	2	2	1	protein-glucosylgalactosylhydroxylysine glucosidase	EC:3.2.1.107	Specific glucose from the disaccharide unit, trehalose	Hepatopancreas

Table 1 (continued)

GH	Family member							Function	Enzymatic Activities	Substrate	Highly Expressed Tissue/Organs
	C. farreri (CF)	C. gigas (CG)	P. fucata (PF)	B. platifrons (BP)	M. philippinarum (MP)	L. gigantea (LG)	O. bimaculoides (OB)				
GH76	1	1	1	1	1	0	0	$\alpha$ -1,6-mannanase	EC:3.2.1.101	mannosides	Hepatopancreas, Mgonad(CG)
GH79	1	2	1	2	4	3	2	heparanase	EC:3.2.1.166	heparan sulfate proteoglycans, etc	Hemolymph, Gill(CF), Eye(CF), Labial_palp (CG)
GH81	1	1	0	0	1	1	0	endo-1,3(4)- $\beta$ -glucanase	EC:3.2.1.6	laminarin, lichenin, cereal D-glucans	Hepatopancreas, Mgonad(CG)
GH85	1	1	0	1	2	1	1	endo- $\beta$ -N-acetylglucosaminidase	EC:3.2.1.96	Specific glycoproteins	Gill, Hemolymph(CF), Kidney(CF), Labial_palp (CG)
GH99	1	2	1	1	1	0	1	glycoprotein endo- $\alpha$ -1,2-mannosidase	EC:3.2.1.130	mannosides	Hepatopancreas(CF), Kidney(CF), Gill(CF), Eye(CF), Fgonad(CG), Hemolymph(CG)
Total	70	/	/	/	/	122	116	113	132	145	167

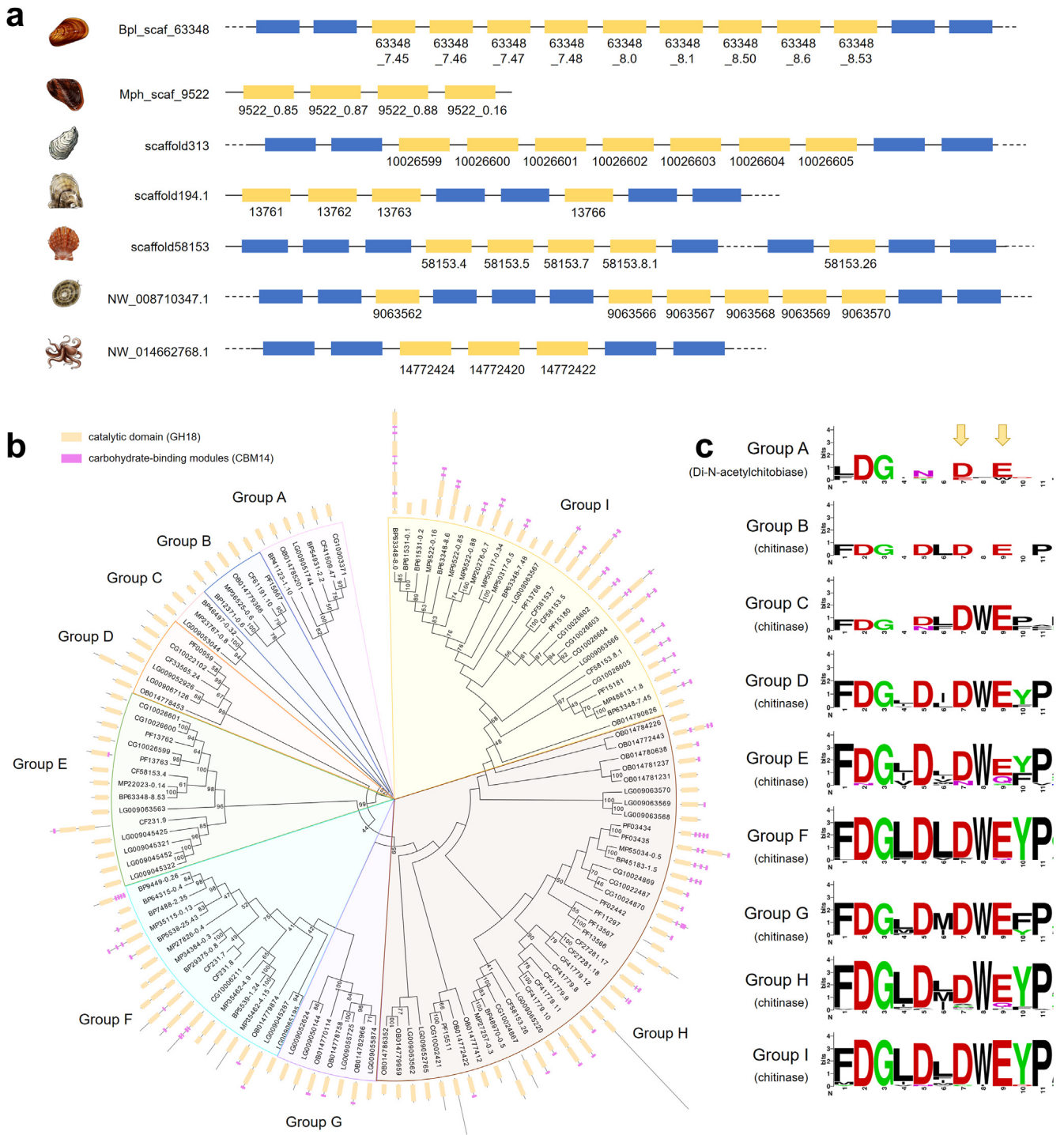
of each module suggested their putative roles in tissue/organ functions (Fig. 3d). The 3D-line graphs systematically presented the overall view of the expression pattern of the GH family in different tissues/organs of *C. farreri*. The module M2 (hepatopancreas-related module) comprised the most diverse categories of the GH family (fifteen GH families), followed by M14 (ten GH families; hemolymph-related module) and M12 (eight GH families; kidney-related module), which was clearly different from other modules containing one-to-four GH families (Figs. 3c-d; 4; Table S5). The GH family genes in M2 started to be expressed after the D-larval period and reached peak expression levels in the later umbo larval period (Fig. S7). GH20, GH18 and GH31 were the top 3 widespread GH families distributed in seven modules, five modules and five modules, respectively, while most hub genes were in the GH18 family (seven GH18 genes in contrast with three GH20 genes and no GH31 gene) (Figs. 3c; 4; Table S5). It is worth noting that in addition to four copies of the GH18 family in M2, four tandem-arranged genes were specifically expressed in the eyes of the scallop *C. farreri* (Fig. 3) and were also hub genes of module M10 (Fig. 3c). The genes of M10 were highly expressed in the eye and mantle (Fig. 3d). These genes were involved in chitin binding, chitinase activity, pigment binding and symporter activity according to the GO analysis (Fig. S8).

### 3.5. Functional analysis of the hepatopancreas-related module M2

The highly expressed GH genes were mostly found in the gene module M2 (42 GH genes), with a p-value of 5.20e-15 (adjusted by Benjamini-Hochberg) (Fig. 5a; Table S5), indicating significant enrichment. The M2 genes were highly expressed in the hepatopancreas (Fig. 5b). Functional analysis showed that these genes were enriched in the GO terms of carbohydrate binding, hydrolase activity, transferase activity and oxidoreductase activity (Fig. 5c), and the KEGG pathways of fat digestion and absorption, protein digestion and absorption, PPAR signaling pathway, vitamin digestion and absorption and starch and sucrose metabolism (Fig. 5d; Table S6), which correlated well with the digestive function of the hepatopancreas.

### 3.6. Expansion of the GH18 family with functional divergence

To understand molluscan GH family expansion in terms of functional divergence, we focused on the GH18 family, as it is the largest GH family expanded among molluscan species. The di-N-acetylchitinase genes are widely expressed in different organs, while most of the expanded chitinase genes are preferentially expressed in the hepatopancreas, mantle or eye (Fig. 6a). These chitinase genes are located in tandem arrays with variable domain structures and are widespread in different coexpression modules as hub genes, making them a good subject for studying GH family evolution and functional divergence. When focusing on the expression profile and phylogenetic relationship of GH18 family genes, we found evidence of within-cluster expressional divergence of the GH18 family. For example, for the largest, four clustered GH18 chitinase genes in the *C. farreri* genome (CF58153.4, CF58153.5, CF58153.7, CF58153.8.1), CF58153.5 and CF58153.7 were located in the middle of the array and were specifically highly expressed in the hepatopancreas, while the remaining two flanking copies were specifically highly expressed in the mantle (Figs. 2a; 6a). A similar phenomenon was also observed in oyster, but its hepatopancreas-related copies were more pronounced (five in contrast to two in scallop; Fig. 6a). This suggested that both hepatopancreas-related and mantle-related GH18 copies were already present in the common ancestor of scallop and oyster (Fig. 6b), and the biased duplication of hepatopancreas-related copies would be evolutionarily advantageous or adaptive for their algae-feeding biology.

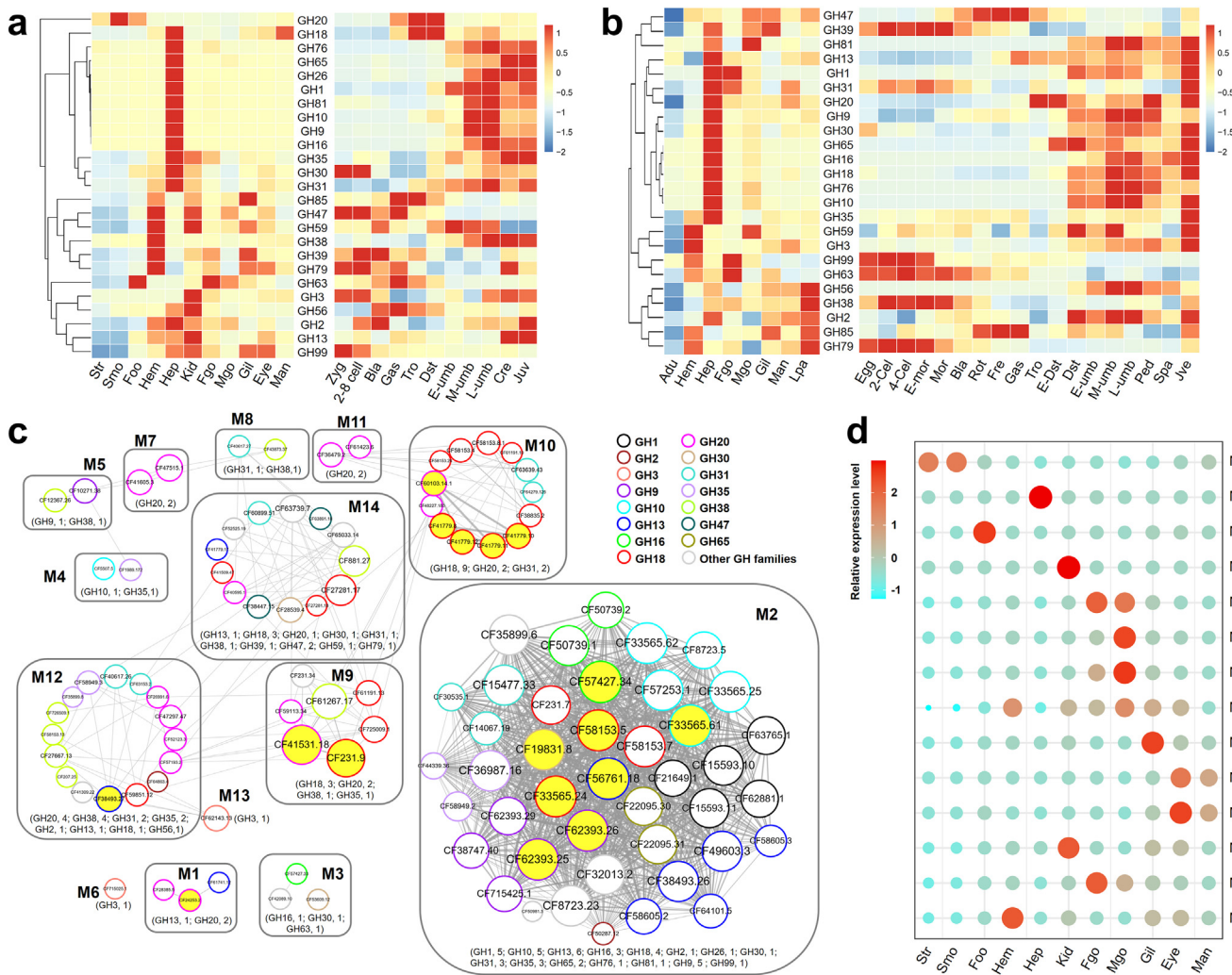


**Fig. 2.** The expansion characteristics of molluscan GH18 family genes. (a) Overview of tandemly arranged GH18 family genes in 7 molluscan genomes. Boxes with different colors represent genes on genomic scaffolds (yellow: GH18 genes; blue: non-GH genes). The solid line between genes indicates gene adjacency. Only the longest GH18 tandem array of each species is shown, and the full view can be found in Fig. S4. The numbers below the GH genes correspond to the gene IDs. (b) Phylogenetic analysis of GH18 family genes from seven molluscan species, *C. farreri* (CF), *C. gigas* (CG), *P. fucata* (PF), *B. platifrons* (BP), *M. philippinarum* (MP), *L. gigantea* (LG), and *O. bimaculoides* (OB). A bootstrap analysis of 2000 replications was carried out on the tree inferred from the Neighbor-joining method. Bootstrap values larger than 40 are shown at each branch of the tree. Molluscan GH18 genes are grouped into nine different groups (Group A-I). The conserved domains of each gene are displayed in the outer layer. (c) The sequence logos of the active centre residues in the catalytic domain of each GH18 group. The arrows indicate the positions of residues (D and E) that are essential for the catalytic activity of chitinase. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**4. Discussion**

Carbohydrates are the most abundant and widely distributed organic compounds on Earth and play a huge diversity of roles, such as energy conversion, cell structural constitution and cell-

cell recognition processes, in living organisms [60]. GHs play essential roles in these important biological processes, primarily by polysaccharide depolymerization [6,7]. As the largest marine phylum in the animal kingdom, molluscs are mainly herbivorous, with their main nutrition source being algae (e.g., diatoms, brown



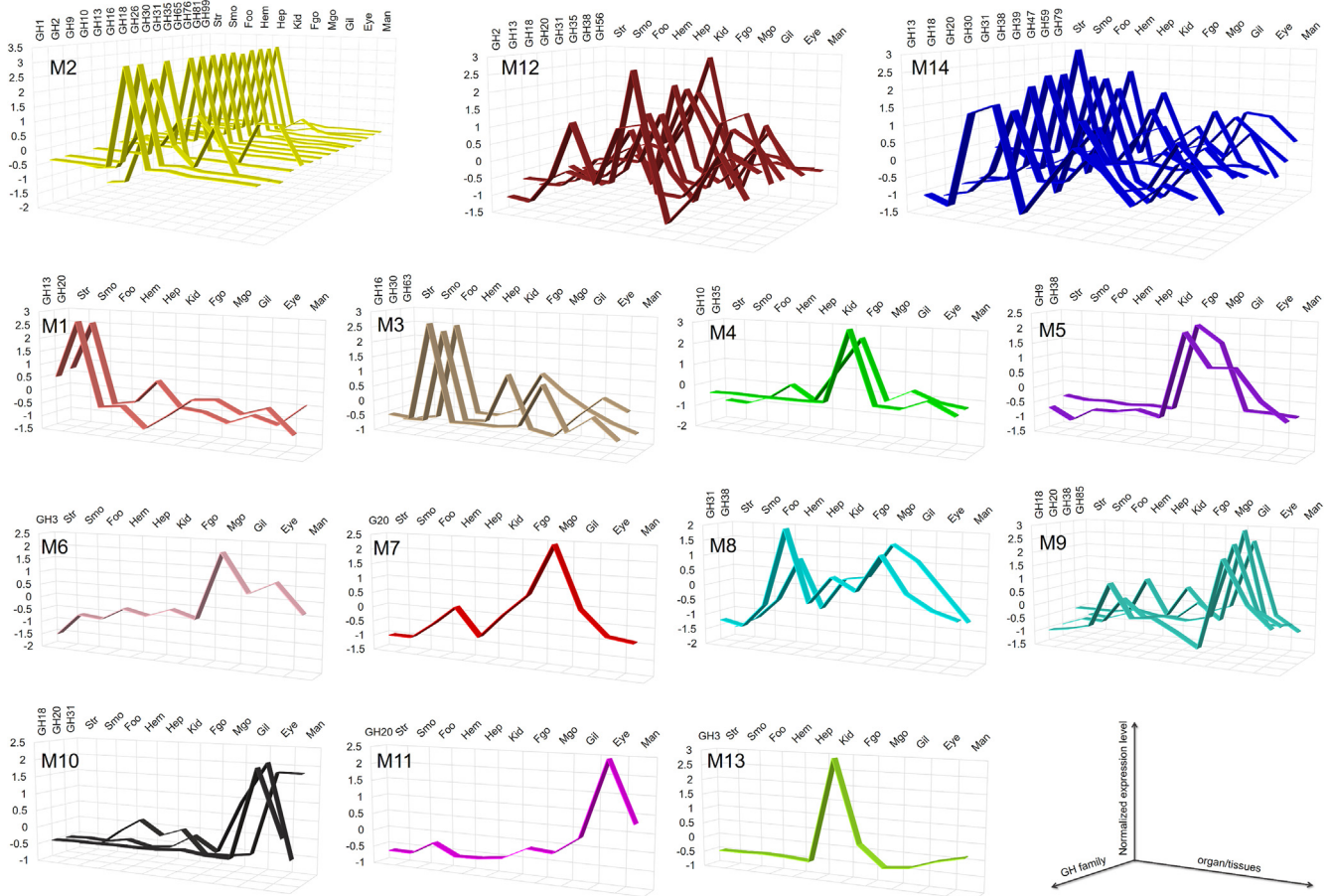
**Fig. 3.** Overview of the expression profile of GH family genes. (a–b) The expression patterns of all GH families in various tissues/organs and developmental stages of *C. farreri* (a) and *C. gigas* (b). For each species, the left heatmap shows the expression patterns of GH families in various tissues/organs (GH families clustered according to their expression patterns), and the right heatmap shows the expression pattern of GH families during development (the same order of GH families as in the left heatmap). The abbreviations and the corresponding full description of developmental stages/adult organs are shown in Table S3. (c) Gene coexpression network of GH family genes in the main tissues/organs of *C. farreri*. For network visualization, each node represents a gene, and the yellow polygonal nodes represent hub genes. The border color of a node indicates the GH family and the node size represents the intramodular connectivity. The GH numbers of each family in the module are listed at the bottom. (d) The dot plot shows the modular expression patterns across adult organs. The dot color and dot size both represent the relative expression level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

algae and green algae), which provide different types of polysaccharides, including cellulose, hemicellulose, chitin and laminarin [1,11,17–24]. Understanding the genomic distribution, expression and evolutionary characteristics of the molluscan GH family is of importance for understanding their functional divergence and adaptive roles in molluscs.

In this study, the full sets of GH family genes from seven molluscan genomes were identified and grouped into 26 different GH families based on their sequence similarity and catalytic domains supported by the CAZy database [3,4]. According to the global view of GH family distribution among 21 species representing the main metazoan groups, the GH9 and GH10 families are prominently expanded in gastropods and bivalves whereas they are contracted in cephalopod, which is possibly a reflection of adaptive evolution for the herbivory in molluscs and carnivory in cephalopod [12,13,61]. This finding is very interesting, as GH9 (cellulases) and GH10 (endo-1,4- $\beta$ -xylanase) are mainly responsible for the hydrolysis of cellulose and hemicellulose, which are important components of the algal cell wall [16,17]. Moreover,

molluscs are among very few animals that have retained cellulose digestion abilities [16,17,62–64]. The expansion of these two families may indicate adaptive evolution for the degradation and utilization of cellulose and hemicellulose components in algal cell walls. Compared with the shallow-water mussel *M. philippinarum*, the deep-sea mussel *B. platifrons* shows much reduced numbers of GH9 and GH10 family genes, possibly as an adaptation to its unusual nutritional supply via chemoautotrophic endosymbionts [46]. The GH18 (chitinase) and GH20 (beta-N-acetylhexosaminidase) families were also expanded in molluscs. The beta-N-acetylhexosaminidase is an exo-type hydrolytic enzyme involved in chitin degradation by cleaving chitin oligosaccharides into the N-acetylglucosamine monomers, which has synergistic effect with GH18 family on chitin hydrolysis [65–67]. Chitin is not only the main polysaccharide from molluscan herbivore diets but also the main component of molluscan shell organic scaffolds [12,68,69], which may suggest the adaptive evolution of the GH18 family for the digestion of food-source chitin and the calcification of shells. Although diverse enzyme functions have been identified in GH9,





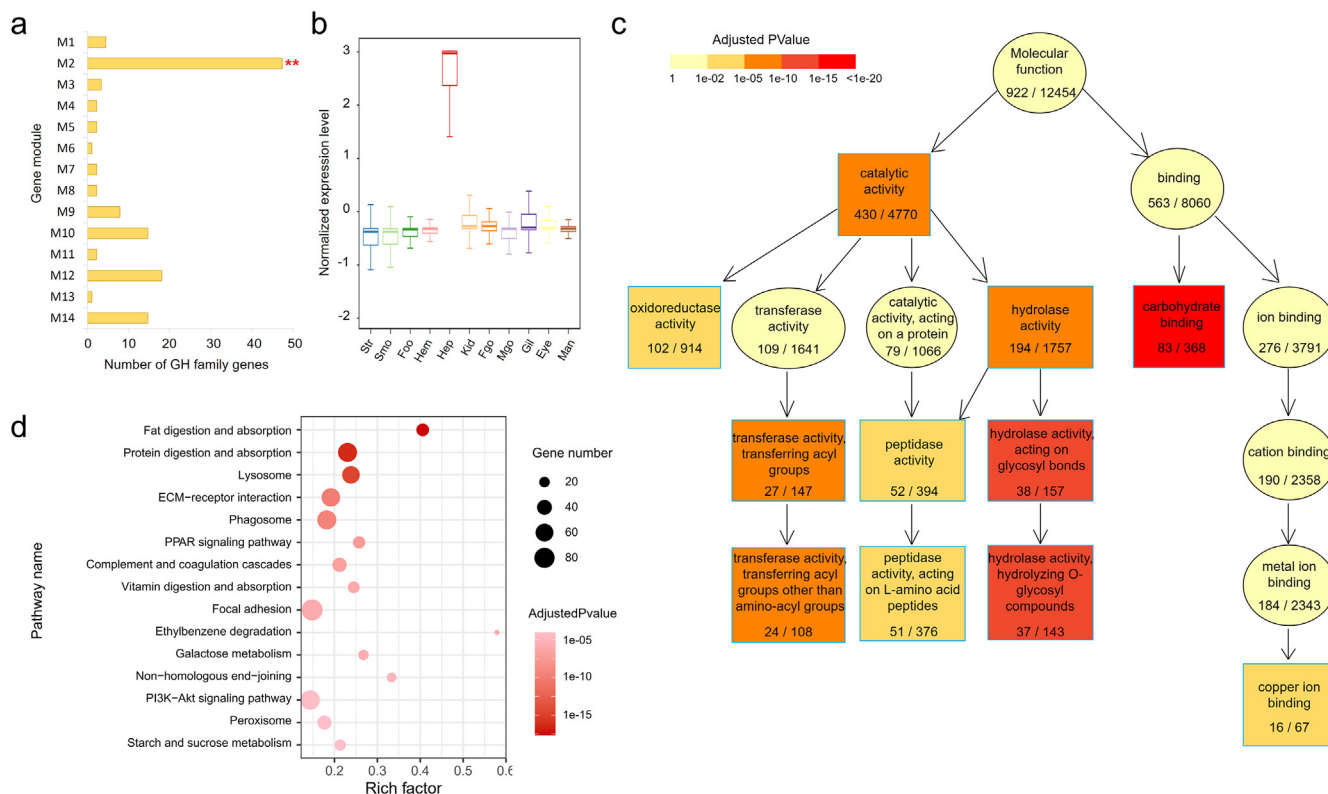
**Fig. 4.** The relative expression pattern of GH genes in each module across major tissues/organs of *C. farreri*. The x-axis, y-axis and z-axis of the 3D line graph represent different tissues/organs, the GH family and relative expression levels, respectively.

GH10, GH18 and GH20 according to the CAZy database, only one kind of enzyme genes that closely related to the hydrolysis of algal polysaccharides have been identified and expanded in each of the expanded GH family, which suggested the adaptive evolution of molluscs for algae feeding.

The characteristic gene structure of the GH family is its catalytic domain, which is often a  $(\beta/a)_8$  TIM barrel structure with conserved catalytic residues that shape the active pocket. The non-catalytic subdomains with  $\beta$ -sandwich fold, zincin-like fold or jelly-roll fold structures beside the catalytic units are involved in the modification of the pocket-shaped catalytic site and are supposed to be critical for the stabilization of enzyme-carbohydrate interactions [70–72]. The catalytic domains are responsible for the hydrolysis reaction of polysaccharides [1,26]. However, the insoluble feature of polysaccharides makes the GHs inefficient when targeting glycosidic bonds. To overcome these problems, some GH family genes often possess one or more CBMs, which can specifically bind to polysaccharide substrates and efficiently promote adsorption of the enzyme onto insoluble polysaccharides [73]. Approximately 20% members of all molluscan GH family genes contain one or more CBMs, which is higher than the rough ratio (7%) of the CAZy database entries containing one or more CBMs [4]. It is interesting that most of the CBM-containing families (GH9, GH10, GH18, GH20) that participate in cellulose, hemicellulose and chitin digestion are expanded in molluscs [30,74–76]. Especially in the GH18 family, 35% of copies possessed up to five CBMs located at the N/C-terminus or interspersed among catalytic domains, which can promote the association of the enzyme with the substrate. The location of CBMs has no effect on their carbohydrate-binding activ-

ity [77,78]. Approximately 8% of GH9 members in molluscs possessed CBMs, while no CBM domain was identified in cellulases of the termite *C. secundus*. Termites are one of the insects that can degrade lignocellulose efficiently with the aid of intestinal microbiota (e.g. bacteria, fungi and flagellates) [79,80]. The cellulose-acting enzymes from flagellates usually have no CBMs, which show activity towards water-soluble cellulose and hardly any activity towards insoluble cellulose. In contrast, cellulases from the symbiotic bacteria of termites usually contain CBMs and show remarkable activities towards insoluble cellulose [81–83]. The wide adoption of CBMs in molluscan-expanded GH families is therefore evolutionarily advantageous and facilitates the functional achievement of GH families in the digestion of diverse marine algal polysaccharides. In view of the important roles of intestinal microbiota in the energy metabolism, the symbiotic microbiota in molluscs is probably involved in the hydrolyzation of polysaccharides from marine algae. However, the exact roles of symbiotic microbiota in the complex metabolic process in Mollusca are still poorly understood, which deserves in-depth analysis for more comprehensive understanding on the evolution of molluscan feeding habits.

Moreover, some GH genes also had an auxiliary domain, such as the Fn3-like domain in the GH3 family, the trefoil domain in GH31 family, the Alpha-mann\_mid domain in the GH38 family and the SLC3A2\_N domain in the GH13 family. The Fn3-like domain in GH3 has been suggested to stabilize the incomplete catalytic domain and can aid the hydrolysis of cellulose by modifying its surface [84,85]. The SLC3A2\_N domain in GH13 can mediate the transmembrane transport of all essential amino acids as a compo-



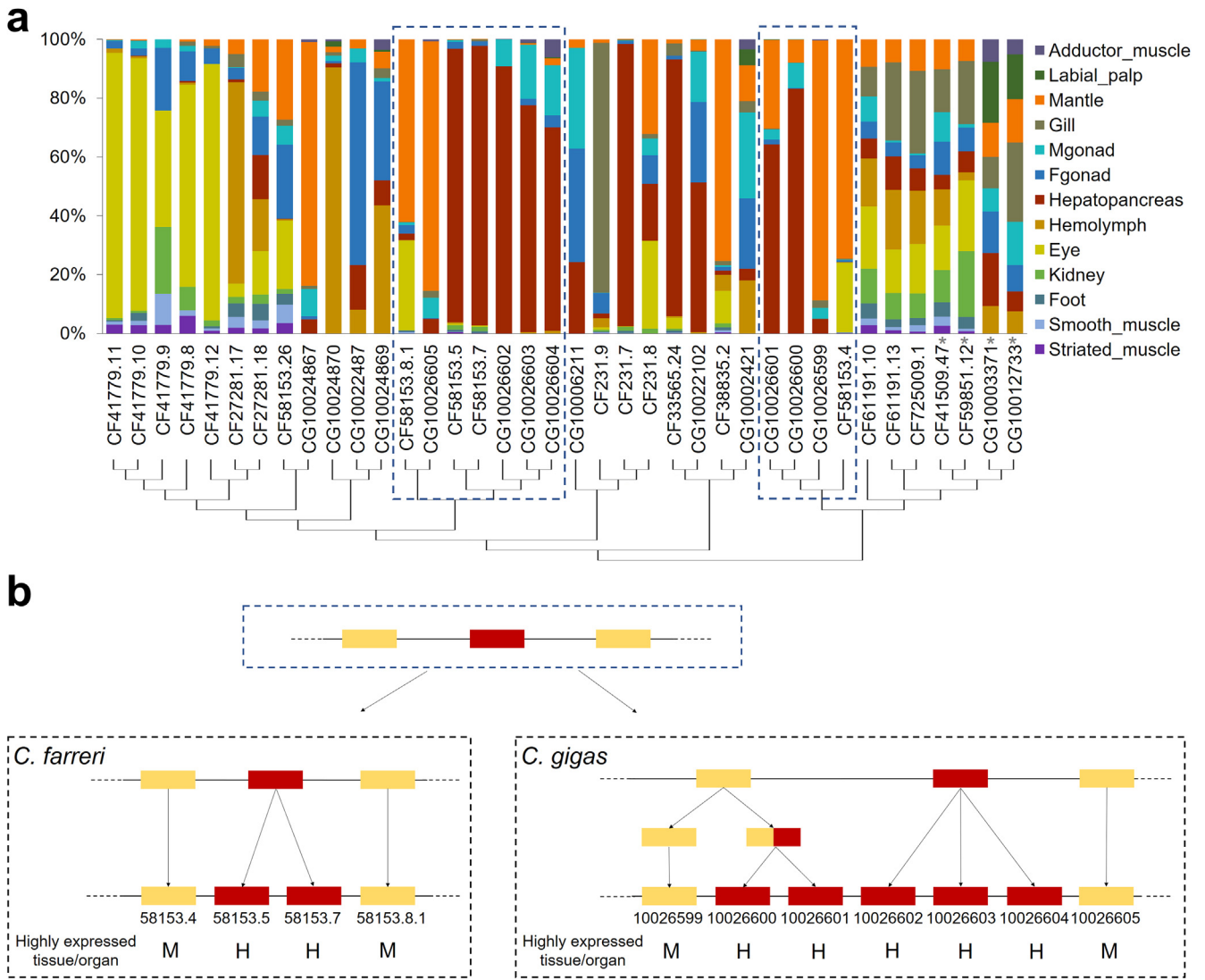
**Fig. 5.** Functional analysis of the GH gene-enriched hepatopancreas-related module M2. (a) Distribution of 110 GH genes in 14 gene coexpression modules. The modules that were significantly enriched with GH family genes (FDR = 5.20e−15) are marked with asterisks\*\*. (b) Boxplots show the relative expression pattern of genes in M2 across major tissues/organs of *C. farreri*. The box colors represent different tissues/organs and the inside lines indicate the median values. (c) Directed acyclic graph (DAG) of GO terms corresponding to molecular function by enrichment analysis of genes in M2. GO terms with FDR < 0.05 are colored from yellow to red with increasing significance. (d) Top 15 KEGG pathways with adjusted enrichment p-values < 0.05 reveal the functional roles of module M2 genes (see Table S6 for the full result). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ment of heteromeric amino acid transport (HAT) [86]. The duplicated trefoil domain in the GH31 family is probably involved in protein–protein or carbohydrate moieties interactions [87]. The alpha-mann\_mid domains are suggested to be involved in enzyme folding and substrate recognition [88]. Above all, different kinds of auxiliary domains are suggested to have important contributions to substrate binding and hydrolysis efficiency.

To understand molluscan GH family expansion for functional divergence, we focused on the GH18 family (chitinase), which is known to have complex and diverse biological functions and participate in molluscan shell formation, algal digestion, immunity, and early embryonic development [68,69,89–91]. GH18 family genes are widely distributed in representatives of all kingdoms. Extensive studies on the domain architecture, functional divergence and structure-function relationship have been carried out in mammals, insects, plants, fungi, bacteria, archaea, nematodes and viruses [92–94]. However, little is known about the characteristics and evolution of the molluscan GH18 family. The expansion of GH18 genes was previously found in insects [95,96]. In this study, molluscan GH18 family genes showed more prominent expansions than those of insects. Most of the expanded GH18 genes reside in tandem arrays, possibly resulting from recent gene duplications. Based on the phylogenetic analysis of molluscan GH18 family genes, species-specific members are dominant in several molluscan species, possibly indicating their recent intraspecific expansion. Functional analysis of active residues in molluscan GH18 catalytic domains showed their good preservation of key residues [e.g., glutamate residue (E) and aspartic acid (D)] that are crucial for the retention of enzymatic activity [97,98], indicat-

ing that the functionality of the expanded GH18 copies is possibly ensured by purifying selection.

Gene expression profiling showed that the molluscan GH family genes are generally widely expressed in a variety of adult tissues/organs as well as at different developmental stages, which may be related to the extensive involvement of GH in various biological processes [6,7]. In particular, approximately half of GH families showed highly restricted expression in the hepatopancreas, which was related to the important roles of GH families in the digestion of polysaccharides. During the developmental process, the highly expressed GH families in the hepatopancreas started to be expressed after D-larval period and reached peak expression levels in the later umbo larval period. This corresponds well to the developmental formation of the molluscan hepatopancreas and the nutritional supply change as molluscs consumed their egg-stored energy and then started to obtain energy by ingesting algae in the surrounding aquatic environment [99]. According to the co-expression network, GH genes from 25 GH families were widely distributed in 14 gene modules and 15% of them were hub genes, which suggested their key roles in multiple biological processes. The GH gene-enriched module was the hepatopancreas-specific module. The wide-presence of diverse highly expressed GH families (14 GH families) in the hepatopancreas suggests a complex polysaccharide-hydrolase system in scallop, which coincides with the largely unique substrate specificity of polysaccharide-degrading enzymes and the diverse marine algal polysaccharides consumed by scallop [1,9,14]. Moreover, genes from the same GH family are widely distributed in different gene modules, which suggests their potential functional diversity. Taking GH18 genes



**Fig. 6.** The within-cluster functional diversity of GH18 family genes. (a) Phylogenetic relationships of GH18 family genes and their relative expression levels across major tissues/organs of *C. farreri* and *C. gigas*. The grey asterisks tagged the di-N-acetylchitinase genes. (b) The putative evolution of hepatopancreas-related and mantle-related GH18 genes within the same cluster of *C. farreri* and *C. gigas*. The yellow boxes indicate the GH18 genes that are highly expressed in the mantle (M), while the red boxes indicate the GH18 genes that are highly expressed in the hepatopancreas (H). The putative expansion histories of GH18 genes were inferred based on their phylogenetic relationships and expressional preference. The numbers below the GH genes correspond to the gene IDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as an example, in addition to the highly expressed GH18 genes in the hepatopancreas, which are related to the digestion of chitin-containing food, four tandem-arranged genes are specifically expressed in the eyes of scallop (also the hub genes of eye-related module M10) and are therefore potentially involved in eye function. Chitin has been identified in the compound eyes of arthropods, where it is considered as a part of the visual system [100]. This may suggest that the expansion of eye-related GH18 in the scallop could be responsible for maintaining eye function in the scallop by chitin hydrolysis. The within-cluster expressional divergence of the GH18 family was found both in scallop and oyster. The expression features and gene location of cluster members suggested that both hepatopancreas-related and mantle-related GH18 copies are already present in the common ancestor of scallop and oyster, and recent expansion of hepatopancreas-related copies independently occurred after the divergence between scallop and oyster. The higher expansion of hepatopancreas-related copies in oyster than scallop is likely a reflection of their differential efficiency of algae digestion based on the previous observation of

the higher filtration efficiency of oyster over scallop especially at low algal concentrations [101].

### 5. Conclusions

We conducted the most comprehensive genome-wide identification, expressional and evolutionary analysis of GH families to date for the largest marine phylum Mollusca. We revealed that compared with other animal groups, the remarkable expansion of GH9, GH10, GH18 and GH20 families in molluscs plays prominent roles in consolidating their algae-feeding habits. The wide adoption of CBMs in molluscan-expanded GH families likely contributes to their efficient hydrolysis of marine algal polysaccharides. Gene expression and network analysis revealed that the hepatopancreas was the main organ for the prominent expression of approximately half of GH families, which corresponds well to the herbivorous diets of molluscan species and the digestive roles of the hepatopancreas. We also revealed the evolutionary signs of differential expansion and functional divergence of the GH18 family for possi-

bly contributing to lineage-specific adaptation. Taken together, through systematic analysis of GH families at both genomic and transcriptomic levels, our findings provide important clues for understanding the functional diversity and evolution of molluscan GH gene families and their association with molluscan algae-feeding biology.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

We acknowledge the grant support from National Key Research and Development Program of China (2018YFC0310802), National Natural Science Foundation of China (U1706203 and 31702330), Fundamental Research Funds for the Central Universities (201762001 and 201841001), and Taishan Scholar Project Fund of Shandong Province of China.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.09.021>.

### References

- Henrissat B, Davies GJ. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol* 2000;124:1515–9.
- Stutz AE, Wrodnigg TM. Imino sugars and glycosyl hydrolases: historical context, current aspects, emerging trends. *Adv Carbohydr Chem Biochem* 2011;66:187–298.
- Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 1991;280:309–16.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucl Acids Res* 2014;42:D490–5.
- Naumoff DG. Hierarchical classification of glycoside hydrolases. *Biochemistry* 2011;76:622–35.
- Coutinho PM, Henrissat B. Life with no sugars?. *J Mol Microbiol Biotechnol* 1999;1:307–8.
- Davies GJ, Henrissat B. Structures and mechanisms of glycosyl hydrolases. *Structure* 1995;3:853–9.
- Berlemont R, Martiny AC. Glycoside hydrolases across environmental microbial communities. *PLoS Comput Biol* 2016;12:e1005300.
- Coines J, Raich L, Rovira C. Modeling catalytic reaction mechanisms in glycoside hydrolases. *Curr Opin Chem Biol* 2019;53:183–91.
- Nakamura T, Fahmi M, Tanaka J, Seki K, Kubota Y, et al. Genome-wide analysis of whole human glycoside hydrolases by data-driven analysis in Silico. *Int J Mol Sci* 2019;20:6290.
- Ojima T. Marine enzymes for biocatalysis: polysaccharide-degrading enzymes from herbivorous marine invertebrates. Philadelphia: Woodhead Publishing; 2013. p. 333–71.
- Steneck RS, Watling L. Feeding capabilities and limitation of herbivorous molluscs: a functional group approach. *Mar Biol* 1982;68:299–319.
- Haszprunar G, Wanninger A. Molluscs. *Curr Biol* 2012;22:510–4.
- Ahmadi A, Moghadamtousi SZ, Abubakar S, Zandi K. Antiviral potential of algae polysaccharides isolated from marine sources: a review. *BioMed Res Int* 2015;2015:825203.
- Alves A, Caridade SG, Mano JF, Sousa RA, Reis RL. Extraction and physico-chemical characterization of a versatile biodegradable polysaccharide obtained from green algae. *Carbohydr Res* 2010;345:2194–200.
- Michel G, Tonon T, Scornet D, Cock JM, Kloareg B. The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytol* 2010;188:82–97.
- Domozych DS, Giancia M, Fangel JU, Mikkelsen MD, Ulvskov P, et al. The cell walls of green algae: a journey through evolution and diversity. *Front Plant Sci* 2012;3:82.
- Synetsya A, Čopíková J, Kim WJ, Park YI. Cell wall polysaccharides of marine algae. In: Kim SK, editor. Springer handbook of marine biotechnology. Berlin, Heidelberg: Springer Handbooks. Springer; 2015.
- Durkin CA, Mock T, Armbrust EV. Chitin in diatoms and its association with the cell wall. *Eukaryot Cell* 2009;8:1038–50.
- Pearlmutter NL, Lembi CA. Localization of chitin in algal and fungal cell walls by light and electron microscopy. *J Histochem Cytochem* 1978;26:782–91.
- Synowiecki J, Al-Khateeb NA. Production, properties, and some new applications of chitin and its derivatives. *Crit Rev Food Sci Nutr* 2003;43:145–71.
- Chiovitti A, Molino P, Crawford SA, Teng R, Spurck T, et al. The glucans extracted with warm water from diatoms are mainly derived from intracellular chrysolaminarin and not extracellular polysaccharides. *Eur J Phycol* 2004;39:117–28.
- Maeda M, Nishizawa K. Fine structure of laminarin of *Eisenia bicyclis*. *J Biochem* 1968;63:199–206.
- Pang Z, Otaka K, Maoka T, Hidaka K, Ishijima S, et al. Structure of beta-glucan oligomer from laminarin and its effect on human monocytes to inhibit the proliferation of U937 cells. *Biosci Biotechnol Biochem* 2005;69:553–8.
- Guo R, Ding M, Zhang S, Xu G, Zhao F. Molecular cloning and characterization of two novel cellulase genes from the mollusc *Ampullaria crosseana*. *J Comp Physiol B* 2008;178:209–15.
- Suzuki K, Ojima T, Nishita K. Purification and cDNA cloning of a cellulase from abalone *Haliotis discus hannai*. *Eur J Biochem* 2003;270:771–8.
- Davison A, Blaxter M. Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol* 2005;22:1273–84.
- Watanabe H, Tokuda G. Cellulolytic systems in insects. *Annu Rev Entomol* 2010;55:609–32.
- Furuhashi T, Schwarzinger C, Miksik I, Smrz M, Beran A. Molluscan shell evolution with review of shell calcification hypothesis. *Comp Biochem Physiol B Biochem Mol Biol* 2009;154:351–71.
- Huang Q, Xie X, Liang G, Gong F, Wang Y, et al. The GH18 family of chitinases: their domain architectures, functions and evolutions. *Glycobiology* 2012;22:23–34.
- Lyu Q, Jiao W, Zhang K, Bao Z, Wang S, et al. Proteomic analysis of scallop hepatopancreatic extract provides insights into marine polysaccharide digestion. *Sci Rep* 2016;6:34866.
- Ojima T, Rahman MM, Kumagai Y, Nishiyama R, Narsico J, et al. Polysaccharide-degrading enzymes from marine gastropods. *Method Enzymol* 2018;605:457–97.
- Simakov O, Marletaz F, Cho S, Edsinger-Gonzales E, Havlak P, et al. Insights into bilaterian evolution from three spiral genomes. *Nature* 2013;493:526–31.
- Albertin CB, Simakov O, Mitros T, Wang Z, Pungor JR, et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 2015;524:220–4.
- Zhang G, Fang X, Guo X, Li L, Luo R, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012;490:49–54.
- Wang S, Zhang J, Jiao W, Li J, Xun X, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol* 2017;1:0120.
- Li Y, Sun X, Hu X, Xun X, Zhang J, et al. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nat Commun* 2017;8:1721.
- Takeuchi T. Molluscan genomics: implications for biology and aquaculture. *Curr Mol Bio Rep* 2017;3:297–305.
- Yang Z, Zhang L, Hu J, Wang J, Bao Z, et al. The evo-devo of molluscs: insights from a genomic perspective. *Evol Dev* 2020. e12336.
- Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 1997;7:637–44.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, et al. The Pfam protein families database. *Nucl Acids Res* 2012;40:290–301.
- The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47: D506–D515.
- O'Leary N, Wright M, Brister J, Ciuffo S, Haddad D, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucl Acids Res* 2016;44:D733–45.
- The Gene Ontology Consortium, Ashburner M, Ball C, Blake J, Botstein D, et al. Gene ontology: tool for the unification of biology. *Nat Genet* (2000) 25: 25–29.
- The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* (2019) 47: D330–D338.
- Sun J, Zhang Y, Xu T, Zhang Y, Mu H, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat Ecol Evol* 2017;1:0121.
- Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, et al. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zool Lett* 2016;2:3.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;30:2725–9.
- Liu S, Li H, Yang Y, Yang R, Yang W, et al. Genome-wide identification of chitinase and chitin deacetylase gene families in the oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Comp Biochem Physiol Part D* 2018;27:13–22.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–90.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis, statistical applications in genetics and molecular biology. *Stat Appl Genet Mol Biol* 2005;4:e17.

- [54] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform* 2008;9:559.
- [55] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995;57:289–300.
- [56] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucl Acids Res* 2009;37:1–13.
- [57] Chen S, Yang P, Jiang F, Wei Y, Ma Z, et al. De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. *PLoS One* 2010;5:e15633.
- [58] Shannon P, Markiel A, Ozier O, Baliga NS, Wang J, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
- [59] Simakov O, Marletaz F, Cho S, Edsingerzales E, Havlak P, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature* 2013;493:526–31.
- [60] Khowala S, Verma D, Banik SP. Biomolecules: Carbohydrates. *National Science Digital Library* (2008). pp. 1–93.
- [61] Villanueva R, Perricone V, Fiorito G. Cephalopods as predators: a short journey among behavioral flexibilities, adaptations, and feeding habits. *Front Physiol* 2017;8:598.
- [62] Shelomi M, Heckel DG, Pauchet Y. Ancestral gene duplication enabled the evolution of multifunctional cellulases in stick insects (Phasmatodea). *Insect Biochem Mol Biol* 2016;71:1–11.
- [63] Tokuda G, Watanabe H, Matsumoto T, Noda H. Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis* (Shiraki): distribution of cellulases and properties of endo-beta-1,4-glucanase. *Zool Sci* 1997;14:83–93.
- [64] Watanabe H, Nakamura M, Tokuda G, Yamaoka I, Scrivener AM, et al. Site of secretion and properties of endogenous endo-beta-1,4-glucanase components from *Reticulitermes speratus* (Kolbe), a Japanese subterranean termite. *Insect Biochem Mol Biol* 1997;27:305–13.
- [65] Fukamizo T, Kramer KJ. Mechanism of chitin oligosaccharides hydrolysis by the binary chitinase system in insect molting fluid. *Insect Biochem* 1985;15(1):1–7.
- [66] Fukamizo T, Kramer KJ. Mechanism of chitin hydrolysis by the binary chitinase system in insect molting fluid. *Insect Biochem* 1985;15(2):141–5.
- [67] Koga D, Hoshika H, Matsushita M, Tanaka A, Ide A, et al. Purification and characterization of beta-N-acetylhexosaminidase from the liver of a prawn, *Penaeus japonicus*. *Biosci Biotechnol Biochem* 1996;60(2):194–9.
- [68] Badarotti F, Thuau R, Lelong C, Dubos M, Favrel P. Characterization of an atypical family 18 chitinase from the oyster *Crassostrea gigas*: evidence for a role in early development and immunity. *Dev Comp Immunol* 2007;31:559–70.
- [69] Okada Y, Yamaura K, Suzuki T, Itoh N, Osada M, et al. Molecular characterization and expression analysis of chitinase from the Pacific oyster *Crassostrea gigas*. *Comp Biochem Physiol B Biochem Mol Biol* 2013;165:83–9.
- [70] Talens-Perales D, Górska A, Huson DH, Polaina J, Marín-Navarro J. Analysis of domain architecture and phylogenetics of Family 2 Glycoside Hydrolases (GH2). *PLoS One* 2016;11(12):e0168035.
- [71] Val-Cid C, Biarnés X, Fajies M, Planas A. Structural-functional analysis reveals a specific domain organization in family GH20 hexosaminidases. *PLoS One* 2015;10(5):e0128075.
- [72] Deane JE, Graham SC, Kim NN, Stein PE, McNair R, et al. Insights into Krabbe disease from structures of galactocerebrosidase. *Proc Natl Acad Sci USA* 2011;108(37):15169–73.
- [73] Boraston AB, Bolam DN, Gilbert HJ, Davies CJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 2004;382:769–81.
- [74] Wang M, Zheng F, Wang T, Lyu Y, Alteen MG, et al. Characterization of *Stacckbrandtia nassauensis* GH 20 beta-hexosaminidase, a versatile biocatalyst for chitobiose degradation. *Int J Mol Sci* 2019;20:1243.
- [75] Naumoff DG. GH10 family of glycoside hydrolases: structure and evolutionary connections. *Mol Biol* 2016;50:132–40.
- [76] Kundu S. Insights into the mechanism(s) of digestion of crystalline cellulose by plant class C GH9 endoglucanases. *J Mol Model* 2019;25:240.
- [77] Mamo G, Hatti-Kaul R, Mattiasson B. Fusion of carbohydrate binding modules from *Thermotoga neopolitana* with a family 10 xylanase from *Bacillus halodurans* S7. *Extremophiles* 2007;11:169–77.
- [78] Shin ES, Yang MJ, Jung KH, Kwon EJ, Jung JS, et al. Influence of the transposition of the thermostabilizing domain of *Clostridium thermocellum* xylanase (XynX) on xylan binding and thermostabilization. *Appl Environ Microbiol* 2002;68:3496–501.
- [79] Breznak JA, Brune A. Role of microorganisms in the digestion of lignocellulose by termites. *Annu Rev Entomol* 1994;39:453–87.
- [80] Tokuda G, Watanabe H. Hidden cellulases in termites: revision of an old hypothesis. *Biol Lett* 2007;3:336–9.
- [81] Nakatani F, Kawaguchi T, Takada G, Sumitani J, et al. Cloning and sequencing of an endoglucanase gene from *Scopulariopsis brevicaulis* TOF-1212, and its expression in *Saccharomyces cerevisiae*. *Biosci Biotechnol Biochem* 2000;64:1238–46.
- [82] Warnecke F, Luginbühl P, Ivanova N, Ghasseman M, Richardson TH, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-eating higher termite. *Nature* 2007;450:560–5.
- [83] König H, Li L, Fröhlich J. The cellulolytic system of the termite gut. *Appl Microbiol Biotechnol* 2013;97:7943–62.
- [84] Kataeva IA, Seidel III RD, Shah A, West LT, Li X, et al. The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase ChhA promotes hydrolysis of cellulose by modifying its surface. *Appl Environ Microbiol* 2002;68:4292–300.
- [85] Gudmundsson M, Hansson H, Karkehabadi S, Larsson A, Stals I, et al. Structural and functional studies of the glycoside hydrolase family 3 beta-glucosidase Cel3A from the moderately thermophilic fungus *Rasamsonia emersonii*. *Acta Crystallogr D Struct Biol* 2016;72(Pt 7):860–70.
- [86] Digomann D, Linge A, Dubrovskaya A. SLC3A2/CD98hc, autophagy and tumor radioresistance: a link confirmed. *Autophagy* 2019;15:1850–1.
- [87] Järvä MA, Lingford JP, John A, Soler NM, Scott NE, et al. Trefol factors share a lectin activity that defines their role in mucus. *Nat Commun* 2020;11:2265.
- [88] Heikinheimo P, Helland R, Leiros HK, Leiros I, Karlsen S, et al. The structure of bovine lysosomal alpha-mannosidase suggests a novel mechanism for low-pH activation. *J Mol Biol* 2003;327(3):631–44.
- [89] Yang B, Zhang M, Li L, Pu YW, et al. Molecular Analysis of atypical family 18 chitinase from Fujian oyster *Crassostrea angulata* and its physiological role in the digestive system. *PLoS One* 2015;10:e0129261.
- [90] Li H, Wang D, Deng Z, Huang G, Fan S, et al. Molecular characterization and expression analysis of chitinase from the pearl oyster *Pinctada fucata*. *Comp Biochem Physiol B Biochem Mol Biol* 2017;203:141–8.
- [91] Wang G, Xu B, Bai Z, Li J. Two chitin metabolic enzyme genes from *Hyriopsis cumingii*: cloning, characterization, and potential functions. *Genet Mol Res* 2012;11:4539–51.
- [92] Chen W, Jiang X, Yang Q. Glycoside hydrolase family 18 chitinases: the known and the unknown. *Biotechnol Adv* 2020;43:107553.
- [93] Karlsson M, Stenlid J. Evolution of family 18 glycoside hydrolases: diversity, domain structures and phylogenetic relationships. *J Mol Microbiol Biotechnol* 2009;16:208–23.
- [94] Funkhouser JD, Aronson Jr NN. Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC Evol Biol* 2007;7:96.
- [95] Zhu Q, Arakane Y, Banerjee D, Beeman RW, Kramer KJ, et al. Domain organization and phylogenetic analysis of the chitinase-like family of proteins in three species of insects. *Insect Biochem Mol Biol* 2008;38:452–66.
- [96] Zhu Q, Deng Y, Vanka P, Brown SJ, Muthukrishnan S, et al. Computational identification of novel chitinase-like proteins in the *Drosophila melanogaster* genome. *Bioinformatics* 2004;20:161–9.
- [97] Lu Y, Zen K, Muthukrishnan S, Kramer KJ. Site-directed mutagenesis and functional analysis of active site acidic amino acid residues D142, D144 and E146 in *Manduca sexta* (tobacco hornworm) chitinase. *Insect Biochem Mol Biol* 2002;32:1369–82.
- [98] Arakane Y, Muthukrishnan S. Insect chitinase and chitinase-like proteins. *Cell Mol Life Sci* 2010;67:201–16.
- [99] Cragg SM. Development, physiology, behaviour and ecology of scallop larvae. *Dev Aquac Fish Sci* 2006;35:45–122.
- [100] Kaya M, Sargin I, Al-jaf I, Erdogan S, Arslan G. Characteristics of corneal lens chitin in dragonfly compound eyes. *Int J Biol Macromol* 2016;89:54–61.
- [101] Palmer RE, Williams LG. Effect of particle concentration on filtration efficiency of the bay scallop *Argopecten irradians* and the oyster *Crassostrea virginica*. *Ophelia* 1980;19:163–74.