Fish and Shellfish Immunology 89 (2019) 237–247

Contents lists available at ScienceDirect

Fish and Shellfish Immunology

journal homepage: www.elsevier.com/locate/fsi

Full length article

The lectin domain containing proteins with mucosal immunity and digestive functions in oyster *Crassostrea gigas*

Weilin Wang[a,b,c], Changhao Gong[a,c], Zirong Han[a,c], Xiaojing Lv[a,c], Shujing Liu[a,c], Lingling Wang[a,b,c], Linsheng Song[a,b,c,*]

[a] Liaoning Key Laboratory of Marine Animal Immunology, Dalian Ocean University, Dalian, 116023, China
[b] Functional Laboratory of Marine Fisheries Science and Food Production Process, Qingdao National Laboratory for Marine Science and Technology, Qingdao, 266200, China
[c] Liaoning Key Laboratory of Marine Animal Immunology and Disease Control, Dalian Ocean University, Dalian, 116023, China

ARTICLE INFO

Keywords:
Crassostrea gigas
Lectin domain
C-type lectin
Mucosal immunity
Digestive functions

ABSTRACT

Lectins are carbohydrate-binding proteins with lectin domains, which are extensively studied for their numerous roles in biological recognition. However, the lectin domain containing proteins (LDCPs) chimerized with other non-lectin domains have not received sufficient attention. In the present study, a genome-wide survey of LDCPs in oyster *Crassostrea gigas* was conducted, and an expansive 640 LDCPs derived from ten lectin domains were identified and functionally explored. In these LDCPs, a total of 282 kinds of domains were predicted, and 90% of the LDCPs contained more than one kind of domain. The lectin domains were frequently fused with non-lectin domains, such as epidermal growth factor domain and peptidase related domains, which supplied LDCPs with more diversity in structures and functions. The C-type lectin domains were the most abundant domains in LDCPs, and they were largely co-existed with non-lectin domains of complement activation-related domains (such as CUB domain and PAN-1 domain) but relative independence with other lectin domains. Furthermore, the C-type lectin domain containing proteins (CTLPs) found to mainly act as pattern immune recognition receptors and were highly expressed in mucosal tissues (digestive gland, male gonad and labial palp) to provide mucosal immune protections. The Concanavalin A-like lectin domains were the second richest domains in LDCPs, and they were mostly constructed into chimeric proteins with epidermal growth factor domain and peptidase related domains. The Concanavalin A-like lectin domain containing proteins (CALPs) were significantly enriched with peptidase activities and mainly expressed in digestive tissues. All the results suggested the mucosal immunity and digestive functions of oyster LDCPs, which provided a fresh idea about the functions of invertebrate lectin family.

1. Introduction

Lectins are carbohydrate-binding proteins that are highly specific for sugar moieties of other molecules [1]. The term “lectin” was originally introduced to refer to the fact that lectins could agglutinate red blood cells [2,3]. The presence of two or more carbohydrate-binding sites was thought to be the prerequisite for lectins, which allowing them to agglutinate cells [2,3]. For a long time, the well-studied lectins were composed solely of carbohydrate-binding domains, which actually should be referred to as sololectins.

With the expansion of knowledge on lectins, it becomes clear that some lectins are chimeric molecules consisting of multiple protein domains, and only the lectin domain exhibits lectin activity. These chimeric lectins together with the former mentioned sololectins compose the lectin domain containing proteins (LDCPs) family. There are plenty of evidences for the occurrence of lectin domains linked to unrelated domains, and these chimerolectins have been reported to be more widespread than the sololectins in plants [4]. The function of these chimerolectins can be predicted by their harbored domains because a domain represents a sequence with conserved functions. For example, the *Marasmius oreades* mushroom lectin MOA, once well known for its binding specificity for blood group B antigens with a N-terminal carbohydrate-binding domain and a C-terminal domain structurally resembling hydrolytic enzymes, also exhibits calcium-dependent cysteine protease activity [5]. One of the most well-known chimerolectins is mannose binding lectins (MBLs) functioning in

* Corresponding author. Liaoning Key Laboratory of Marine Animal Immunology, Dalian Ocean University, Dalian, 116023, China.
E-mail address: lshsong@dlou.edu.cn (L. Song).

https://doi.org/10.1016/j.fsi.2019.03.067
Received 25 January 2019; Received in revised form 19 March 2019; Accepted 26 March 2019
Available online 29 March 2019

complement lectin-activation pathway. MBL can form oligomerization by its collagen domain and bind to invaders by its lectin domain [6]. It is more appropriate to study these chimeric LDCPs in terms of their domain architectures rather than their "simple" carbohydrate-binding functions.

Lectins exist in almost all the organisms, including plants, animals as well as microorganisms, and the significantly varied numbers of lectin genes have been reported in different species [6–8]. Though the fundamental and diverse activities of lectins have been well studied in vertebrates, the expanded genes encoding lectin domains in invertebrate will provide much more possibilities to find novel functions of lectins [6]. For example, C type lectin genes are highly abundant in many metazoan genomes, with 283 members in the nematode *Caenorhabditis elegans* [9] and 266 members in the bivalve oyster *Crassostrea gigas* [10]. In the present study, the expansive LDCPs in oyster *Crassostrea gigas* genome were screened and analyzed with the purposes to (1) identify all the lectin domains and LDCPs in oyster, (2) analyze the distribution of lectin domain among the LDCPs, (3) explore the functions of different clusters of lectin domains, and provide a better understanding of invertebrate lectins.

2. Materials and methods

2.1. Data collection and bioinformatics analysis

The genome information of Pacific oyster *Crassostrea gigas* was acquired from the previous report [11] and its supplementary database (Table S14 of Supplementary Tables and Supplementary Information) [11]. The annotation items of genes, including InterPro (IPR), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), were collected for further functional predictions and enrichments. The protein domains of each gene were predicted depending on InterPro annotation (IPR items). For LDCPs screening, the InterPro annotation of each gene was analyzed, and the genes with IPR items containing lectin were all retrieved. These target gene sets were studied in depth for GO and KEGG enrichment by a free online bioinformatics tools OmicShare (http://www.omicshare.com/tools). The genome GO items and KEGG items were used as background for GO and KEGG enrichment, respectively, and all the bioinformatics analyses were conducted under default parameters. The transcriptomic representation of target genes (RPKM) in different adult organs (Table S14 from the report of Zhang et al., 2012) [11] was also analyzed. And the expression data of target gene sets were extracted directly without any modification. The expression heatmaps were generated by tool in OmicShare under default parameters.

2.2. Oysters and sample collection

The adult oysters *C. gigas*, with about 100–150 g weight, were collected from local farm (Dalian, China) and acclimated in aerated seawater at 15–20 °C for a week before experiments. Six tissues including digestive gland, gills, labial palp, adductor muscle, mantle and hemocytes were collected from six normal adult oysters as parallel samples to investigate the gene expression level. All the samples were stored at −80 °C after addition of 1 mL Trizol reagent (Invitrogen) for subsequent RNA extraction. The total RNA isolation and cDNA synthesis was conducted following the previous report [12]. The extracted RNA were qualitatively checked by agarose-gel electrophoresis and quantitatively evaluated by Nano-Drop 2100 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) before cDNA synthesis. And the cDNA synthesis was conducted by using the PrimeScript™ real-time PCR kit (Takara, Japan) under its instructions.

2.3. Real-time PCR analysis of mRNA expression levels

The mRNA expressions of ten candidate genes in different tissues were checked by SYBR Green fluorescent quantitative real-time PCR (RT-PCR) in an ABI Quantstudio sequence detection system according to the manual (Applied Biosystems). Specific primers (Table 1) were designed to amplify the fragments representing the corresponding genes, including macrophage mannose receptor 1 (CGI_10025847, CGI_10014755), C-type mannose receptor 2 (CGI_10020693, CGI_10001434), salivary C-type lectin (CGI_10012515), C-type lectin 4 (CGI_10018363), meprin (CGI_10007629, CGI_10007630), Fc receptor (CGI_10006368) and hemolymph lipopolysaccharide-binding protein (CGI_10008323) with the oyster EF (Elongation Factor) fragment as an internal control to calibrate the cDNA template. The RT-PCR reaction system and procedure were conducted according to the previous report [12]. The data were analyzed automatically with $2^{-\Delta\Delta CT}$ method using ABI 7500 SDS software V2.0, and all results were shown as mean ± S.E. (N = 6).

Table 1
Primers used for Real-time PCR in this study.

Gene ID	Primer Sequence (5′-3′)
CGI_10025847	Forward: ACTTACAAAGAGCAACCGCAAC
	Reverse: CTCAGATAGGGAGAGTGTGGGA
CGI_10006368	Forward: CTTTTTTTACAGATGTTCCAGCCAC
	Reverse: TCCATAACCATTCTCCGTCTACCA
CGI_10008323	Forward: GGTGTCCTAACGGGTCTGTGATA
	Reverse: CCACTCGTAGTCATACCTCGGATA
CGI_10020693	Forward: GTATCCCGACAAGACTGACAACC
	Reverse: TTCCTCCGCTACCGAAACCT
CGI_10018363	Forward: ATCAGCAGAGACAAAGTATGGGGAGA
	Reverse: GGGTTCGCCAGTGCCAAATG
CGI_10012515	Forward: CCGTAGATTTAGGAAAAGAAGGAC
	Reverse: GCTGTAGTGACATTTTGTATCGC
CGI_10001434	Forward: TTCTCTTTCATCATCGCTGTCATC
	Reverse: ACCATCTTCTTCCGTTTTTCTTTT
CGI_10014755	Forward: GACTGTGAAGCCCGAGGAAA
	Reverse: GCTGTTTGCTCACCGTATTGTATT
CGI_10007629	Forward: TGGTGTGGTTCACTGGACTATT
	Reverse: CCTTCCTCAGGCAACCAGTA
CGI_10007630	Forward: TTGACTCTGTCCGATGGTTGT
	Reverse: CGAAGGAGTGAGCCGAGTAG
*Cg*EF	Forward: AGTCACCAAGGCTGCACAGAAAG
	Reverse: TCCGACGTATTTCTTTGCGATGT

2.4. Statistical analysis

All data were subjected to one-way analysis of variance (one-way ANOVA) followed by a post hoc multiple-comparisons (Tukey's) test. Differences were considered significant at $p < 0.05$.

3. Results

3.1. The lectin domain containing proteins (LDCPs) in oyster genome

Ten kinds of IPRs related to lectin domains were identified (Table 2) from all the retrieved IPR items after the annotation of oyster genome. A total of 640 LDCPs containing at least one of the ten lectin IPRs were obtained (Supplementary Table S1). In these LDCPs, there were 282 IPRs (domains) identified, and about 90% (574) of LDCPs contained more than one IPR, 66 LDCPs contained only one IPR, 254 LDCPs contained two IPRs, and 123 LDCPs contained three IPRs (Fig. 1). There were three richest lectin IPRs, C-type lectin fold (IPR016187), C-type lectin (IPR001304) and Concanavalin A-like lectin/glucanase (IPR008985), with 352, 273 and 189 containing proteins predicted, respectively.

3.2. The distribution of lectin domains among the LDCPs

To explore the possible coexistence of lectin domains in LDCPs, the proteins containing the four most abundant lectin domains (C-type

Table 2
IPRs related to lectin domains identified in oyster *C. gigas* genome.

IPR ID	IPR description	Gene number
IPR016187	C-type lectin fold	352
IPR001304	C-type lectin	273
IPR008985	Concanavalin A-like lectin/glucanase	189
IPR006585	Fucolectin tachylectin-4 pentraxin-1	50
IPR000922	D-galactoside/L-rhamnose binding SUEL lectin domain	30
IPR000772	Ricin B lectin	18
IPR008997	Ricin B-related lectin	18
IPR001079	Galectin, carbohydrate recognition domain	11
IPR019019	H-type lectin domain	8
IPR005052	Legume-like lectin	3

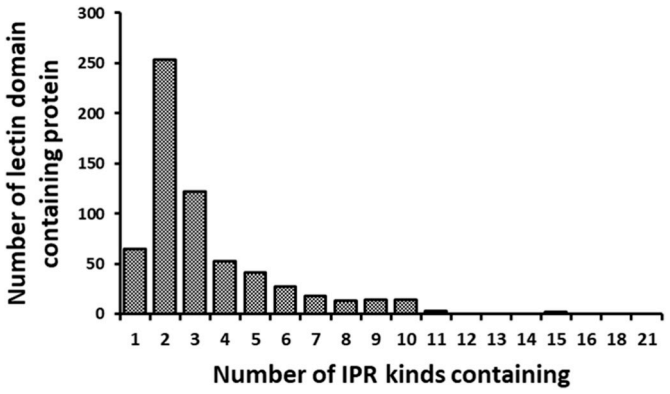

Fig. 1. The statistic numbers of lectin domain containing proteins (LDCPs) with varied IPR kinds in oyster *C. gigas*.

lectin fold, C-type lectin, Concanavalin A-like lectin and Fucolectin) were analyzed by Venn diagram. The result showed that almost all the C-type lectin domain containing proteins (CTLPs) (272 of the 273 proteins) also harbored the C-type lectin fold domain (Fig. 2A). However, the coexistence of C-type lectin domain with the other two lectin domains was rarely observed in the LDCPs. Only four molecules simultaneously contained C-type lectin domain and Concanavalin A-like lectin domain, and one contained C-type lectin domain and Fucolectin domain. There was no LDCP identified simultaneously containing Concanavalin A-like lectin and Fucolectin domain (Fig. 2A). The coexistence of C type lectin domain with the rest lectin domains was similar rare. For instance, only two CTLPs contained D-galactoside/L-rhamnose binding SUEL lectin domain, and no CTLP contained Ricin B lectin and Ricin B-related lectin domains (data not shown).

In oyster *C. gigas,* about 90% of the LDCPs were found to contain multiple domains, but the coexistence of lectin domains was rarely observed except the coexistence of C-type lectin with C-type lectin fold. The other 272 non-lectin domains in LDCPs were further analyzed. Among the 640 LDCPs, 258 LDCPs contained only lectin domains (solectin), and 382 LDCPs contained non-lectin domains (chimerolectin). Most of the solectins (82.6%) were C-type lectin (176) or C-type lectin fold (213) domain containing proteins. Most of the chimerolectins contained Concanavalin A-like lectin (168) and C-type lectin fold domains (139) (Fig. 2B).

The non-lectin domains of top 20 abundance in LDCPs were identified (Table 3). There were 67 proteins containing Epidermal growth factor-like type 3 (IPR000742) domain, 67 proteins containing MAM domains (IPR000998), 64 proteins containing Epidermal growth factor-like (IPR006210) domains, and 61 proteins containing Galactose-binding domain-like (IPR008979) domains. Some other common domains were also identified to co-exist with lectin domains, such as 13 Chitin binding domain (IPR002557), 19 Carbohydrate-binding WSC (IPR002889, IPR013994) domains, and 36 Immunoglobulin related (IPR007110, IPR003598, IPR003599, IPR013098, IPR013151, IPR014756, IPR003596) domains.

3.3. Functional prediction of the LDCPs

The GO items and KEGG items of these 640 LDCPs were retrieved to predict their potential functions with the genome GO and KEGG items as background information, respectively. There were eight GO items significantly enriched at level two (Table 4), with most abundant in binding (GO:0005488), carbohydrate binding (GO:0030246), and calcium ion binding (GO:0005509) activity. Unexpectedly, the peptidase activity was also significantly enriched, such as endopeptidase activity (GO:0004175), metalloendopeptidase activity (GO:0004222), and metallopeptidase activity (GO:0008237).

KEGG pathway enrichments were analyzed to find the possible molecular pathways related to LDCPs. Six KEGG items were significantly enriched, and the LDCPs mainly participated in the phagosome (ko04145) pathway of transport and catabolism, the protein digestion and absorption (ko04974) pathway of digestive system, the ECM-receptor interaction (ko04512) pathway and Notch signaling (ko04330) pathway of signaling molecules and interaction, the hematopoietic cell lineage (ko04640) pathway, and complement and coagulation cascades (ko04610) pathway of immune system (Fig. 3, Table 5).

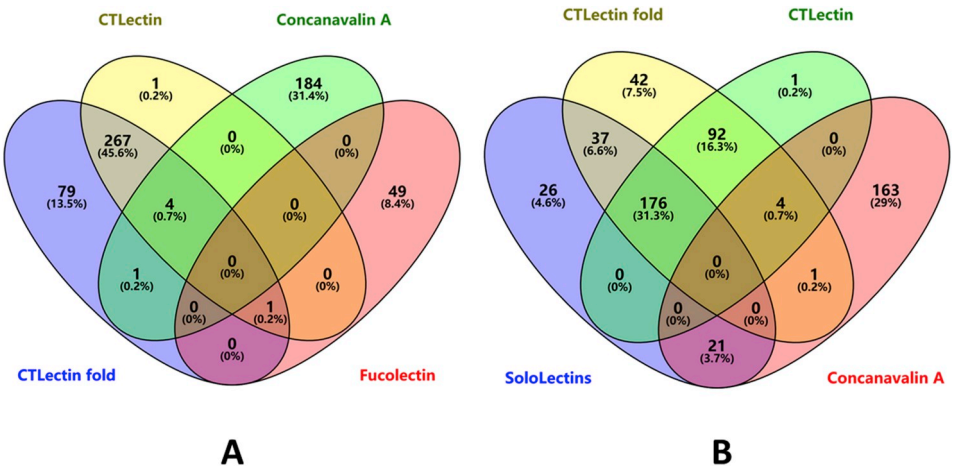

Fig. 2. The distributions of four most abundant lectin domains among the lectin domain containing proteins (LDCPs) (A) and lectin domain only proteins (Sololectin) (B) in oyster *C. gigas*.

Table 3
The non-lectin domains of top 20 abundance in LDCPs of oyster.

IPR	IPR description	Gene number	IPR	IPR description	Gene number
IPR000742	Epidermal growth factor-like, type 3	67	IPR016060	Complement control module	28
IPR000998	MAM domain	67	IPR002353	Type II antifreeze protein	27
IPR006210	Epidermal growth factor-like	64	IPR012680	Laminin G, subdomain 2	26
IPR008979	Galactose-binding domain-like	61	IPR001870	B302/SPRY domain	25
IPR000421	Coagulation factor 5/8 type, C-terminal	55	IPR001791	Laminin G domain	24
IPR006209	EGF	34	IPR001506	Peptidase M12A, astacin	23
IPR000859	CUB	33	IPR002172	LDLR class A repeat	22
IPR001881	EGF-like calcium-binding	32	IPR003609	Apple-like	21
IPR003014	PAN-1 domain	29	IPR003877	SPla/RYanodine receptor SPRY	21
IPR000436	Sushi/SCR/CCP	28	IPR006026	Peptidase, metallopeptidase	20

Table 4
The significantly enriched GO items of the 640 LDCPs.

GO ID	Description	GeneRatio	BgRatio (13029)	Pvalue	FDR
GO:0030246	carbohydrate binding	77 (14.29%)	144 (1.11%)	0	0
GO:0005488	binding	526 (97.59%)	9616 (73.8%)	0	0
GO:0005201	extracellular matrix structural constituent	8 (1.48%)	11 (0.08%)	0	0
GO:0004222	metalloendopeptidase activity	23 (4.27%)	149 (1.14%)	0	0.000002
GO:0008237	metallopeptidase activity	23 (4.27%)	200 (1.54%)	0.000009	0.000281
GO:0005509	calcium ion binding	42 (7.79%)	543 (4.17%)	0.000069	0.001765
GO:0004175	endopeptidase activity	27 (5.01%)	292 (2.24%)	0.00008	0.001765
GO:0008061	chitin binding	13 (2.41%)	91 (0.7%)	0.000089	0.001765

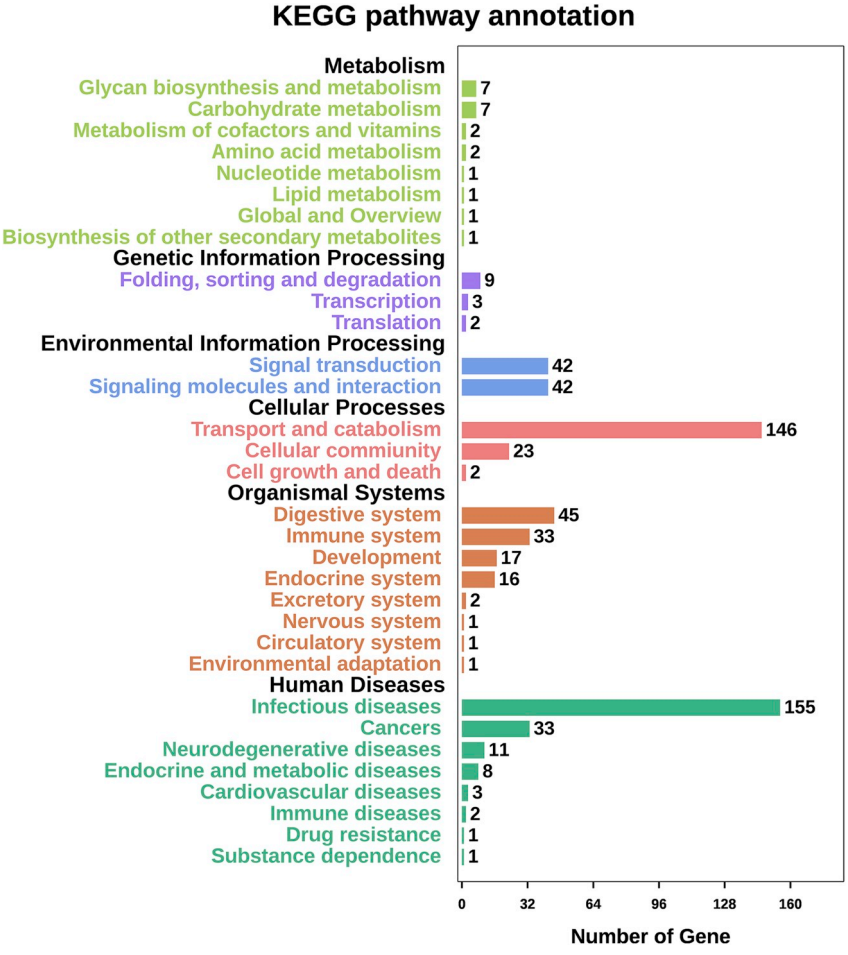

Fig. 3. The KEGG pathway enrichment of all the 640 lectin domain containing proteins (LDCPs).

Table 5
The significantly enriched KEGG pathways of the 640 LDCPs.

Pathway ID	Pathway	DEGs genes with pathway annotation	All genes with pathway annotation	Pvalue	Qvalue
ko04145	Phagosome	141	303	9.1E-127	9E-125
ko04974	Protein digestion and absorption	40	164	3.05E-21	1.1E-19
ko04640	Hematopoietic cell lineage	17	53	9.57E-12	2.5E-10
ko04512	ECM-receptor interaction	25	270	7.18E-05	0.00149
ko04610	Complement and coagulation cascades	13	104	0.000238	0.00354
ko04330	Notch signaling pathway	13	142	0.004366	0.04541

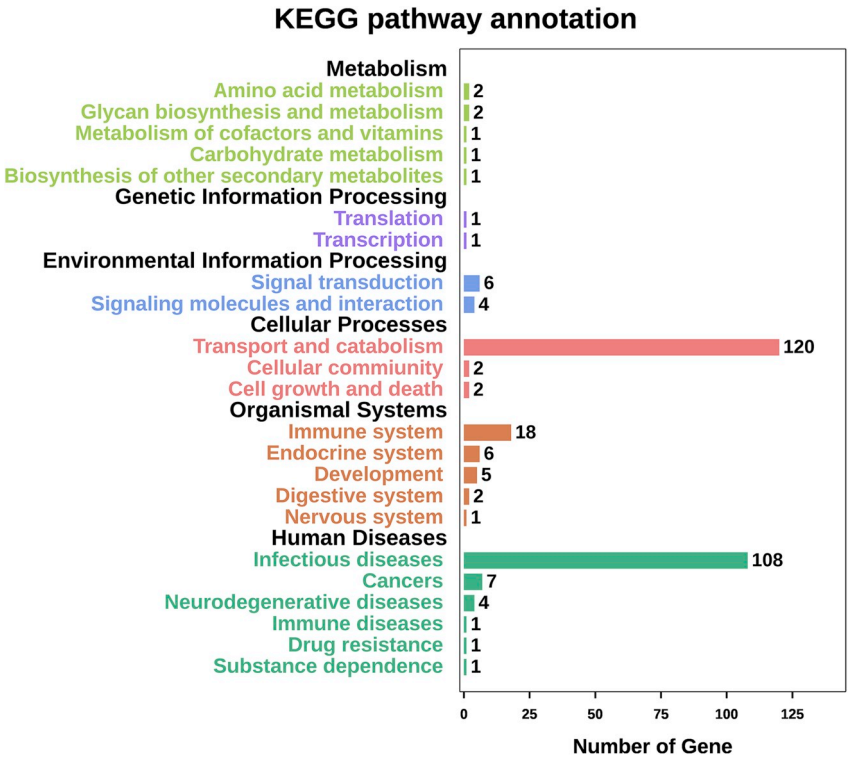

Fig. 4. The KEGG pathway enrichment of C-type lectin domain containing proteins (CTLPs).

3.4. The C-type lectin domain containing proteins (CTLPs) function in mucosal immunity

The C-type lectin domain containing protein (CTLP) with 273 members was the largest lectin cluster in oyster. Several classical C-type lectin families were identified, such as 27 C-type lectins, 33 C-type mannose receptors, 20 macrophage mannose receptors, 15 low affinity immunoglobulin epsilon Fc receptors, 4 galactose-specific C-type lectins, and 4 salivary c-type lectins. About 176 CTLPs (sololectins) contained just lectin domains and 97 CTLPs (chimerolectins) contained other non-lectin domains. Among the 97 chimeric CTLPs, the most abundant non-lectin domains fused with C-type lectin domain were CUB (IPR000859) domain, PAN-1 domain (IPR003014), Sushi/SCR/CCP (IPR000436), Complement control module (IPR016060), and Type II antifreeze protein (IPR002353).

KEGG pathway enrichment was conducted to further understand the functions of the CTLPs. These genes were significantly enriched in immune related pathways, such as the phagosome (ko04145) pathway of transport and catabolism, and hematopoietic cell lineage pathway (ko04640) (Fig. 4). In the phagosome pathway, the 119 genes functioned as phagocytosis-promoting receptors, such as complement receptors, integrins, and C-lectin receptors (Fig. 5). In hematopoietic cell lineage pathway, the 15 genes functioned as CD23 and low affinity immunoglobulin epsilon Fc receptor.

The expression patterns of CTLPs in different adult tissues were further examined with the data from previous report [11]. More CTLPs with a higher expression level in mucosal tissues were witnessed in the expression heatmap, such as 70 CTLPs in digestive gland (Dgl), 40 CTLPs in male gonad (Mgo), and 35 in labial palp (Lpa) (Fig. 6). Meanwhile, ten CTLPs were also found to be highly expressed in circulatory hemocytes.

3.5. The Concanavalin A-like lectin domain containing proteins (CALPs) function in digestive system

The Concanavalin A-like lectin domain containing proteins (CALPs) were the second largest lectin cluster, and most of the CALPs were chimeric lectin containing proteins. The most abundant non-lectin domains identified in CALPs were the Epidermal growth factor domains (IPR000742, IPR006210, IPR006209, IPR001881), MAM domain (IPR000998), and Peptidase domains (IPR001506, IPR006026). A plenty of proteins with protease activity were identified, such as 20 blastula proteases, four endo-beta-*d*-glucanases, and four meprin metallopeptidases. The KEGG pathway enrichment was conducted for more detailed functions of CALPs. These genes were significantly enriched in the protein digestion and absorption (ko04974) pathway of digestive system, the ECM-receptor interaction (ko04512) pathway of signaling molecules and interaction (Fig. 7). For example, 31 genes functioned as peptidase and collagen of the protein digestion and absorption (ko04974) pathway in digestive system (Fig. 8), including

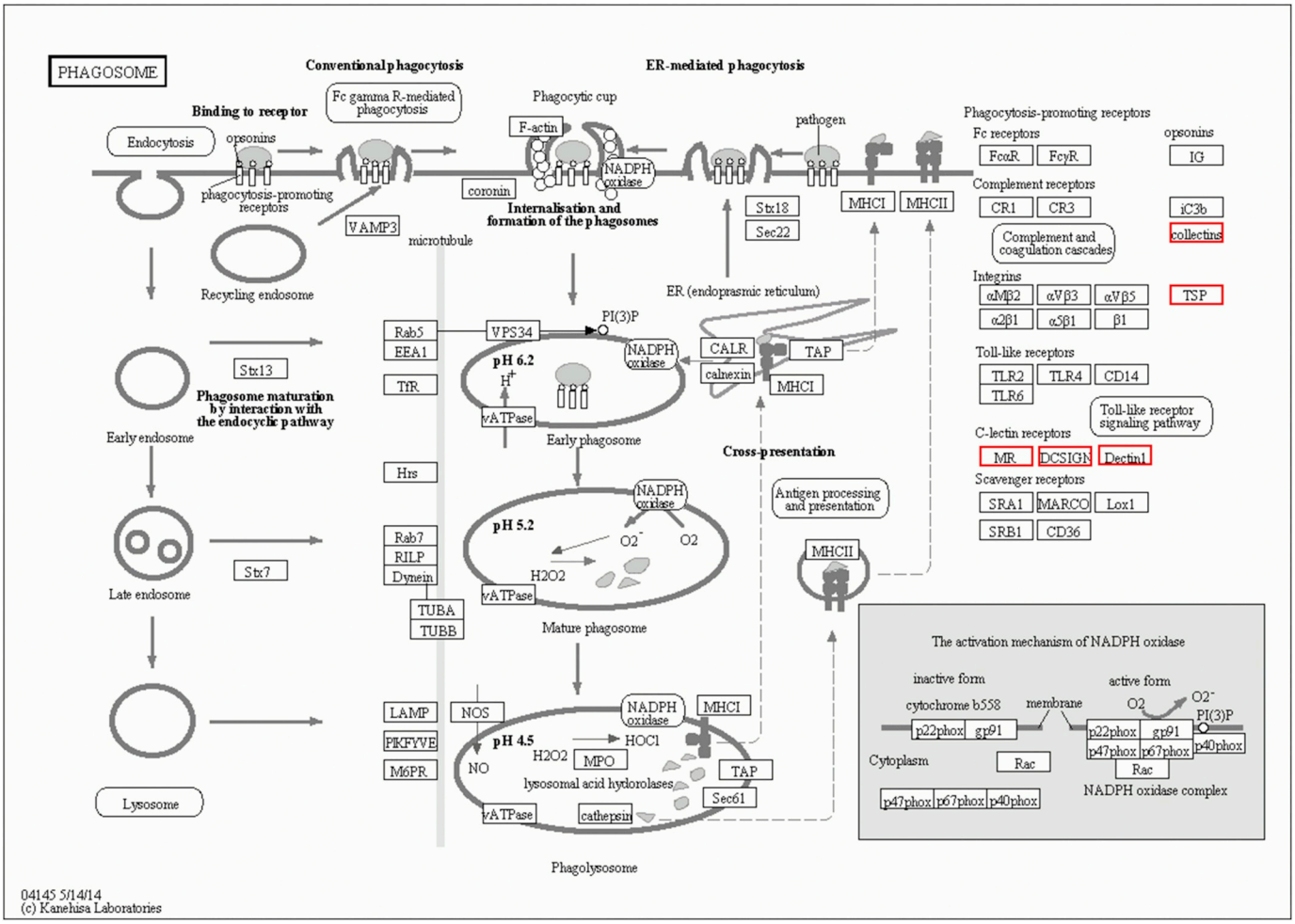

Fig. 5. The C-type lectin domain containing proteins (CTLPs) highly enriched in phagosome pathway.

blastula protease, meprin subunits A alpha and beta. The expression patterns of 25 peptidases in different adult organs were further explored, and most of the peptidases were observed to be higher expressed in digestive gland (Dgl) (Fig. 9).

3.6. Validation of the mRNA expression levels of candidate genes

The mRNA expressions of ten candidate genes involved in either mucosal immunity or digestive function were examined in different tissues by RT-PCR. The expression patterns of CTLPs were found to be of tissue-specific. For instance, the macrophage mannose receptor 1 (CGI_10014755) and C-type mannose receptor 2 (CGI_10001434) were highest expressed in labial palp (Fig. 10A). The immunoglobulin Fc receptor (CGI_10006368), hemolymph lipopolysaccharide binding protein (CGI_10008323), C-type mannose receptor 2 (CGI_10020693) and macrophage mannose receptor 1 (CGI_10025847) were highest expressed in digestive gland (Fig. 10B). The C-type lectin 4 (CGI_10018363) and salivary c-type lectin (CGI_10012515) were highest expressed in gills (Fig. 10C). The two digest related metalloprotease, meprin A subunit alpha (CGI_10004774) and beta (CGI_10007629), were both highest expressed in digestive gland (Fig. 10D).

4. Discussion

Lectins are common molecules among almost all organisms and have been well-studied in prokaryotes, plants and animals with various physical functions. In microorganisms, such as amoeba and bacteria, lectins serve as important infective means for attachment to the target cells via surface carbohydrates [13]. In plants, lectins are involved in the protection against pathogenic microorganisms and also symbiosis with nitrogen-fixing bacteria [14,15]. More functions have been reported for animal lectins, such as self/non-self recognition, cell-cell interactions, mediation of endocytosis and bactericide [16–18]. The diverse roles of lectins are supported by their abundant kinds of structures, and they have been divided into several protein families based on the structural diversity of lectin domains [9]. It has been reported that there are at least 12 structural families in animals, such as C-type (Ca^{2+}-dependent), S-type (thiol-dependent, known as galectins), and F-type (fucose recognition domain) families [19]. Moreover, large amount of non-lectin domains have also been predicted in lectins, which endow more functions for lectin groups. For a long time, lectins have been extensively studied for their classical carbohydrate-binding domains, but not for the non-lectin domains. So it is quite valuable to investigate lectin domain containing proteins (LDCPs) from the whole structure view and find novel functions of these traditional proteins. Fortunately, the great number of lectin genes in invertebrate genome, such as 266 C-type lectins in the bivalve oyster *Crassostrea gigas* [10], provides a good opportunity and possibility to better understand the functions of lectins.

LDCPs are a large protein family, and they can be diversified by various lectin domains. In the present study, about 57 lectin related domains were identified after screening the online Interpro database, and most of them were plant-nominated, such as jacalin-like lectin domain (IPR001229), Ricin B lectin domain (IPR000772), and Legume lectin domain (IPR001220). At least 12 kinds of lectin domains have

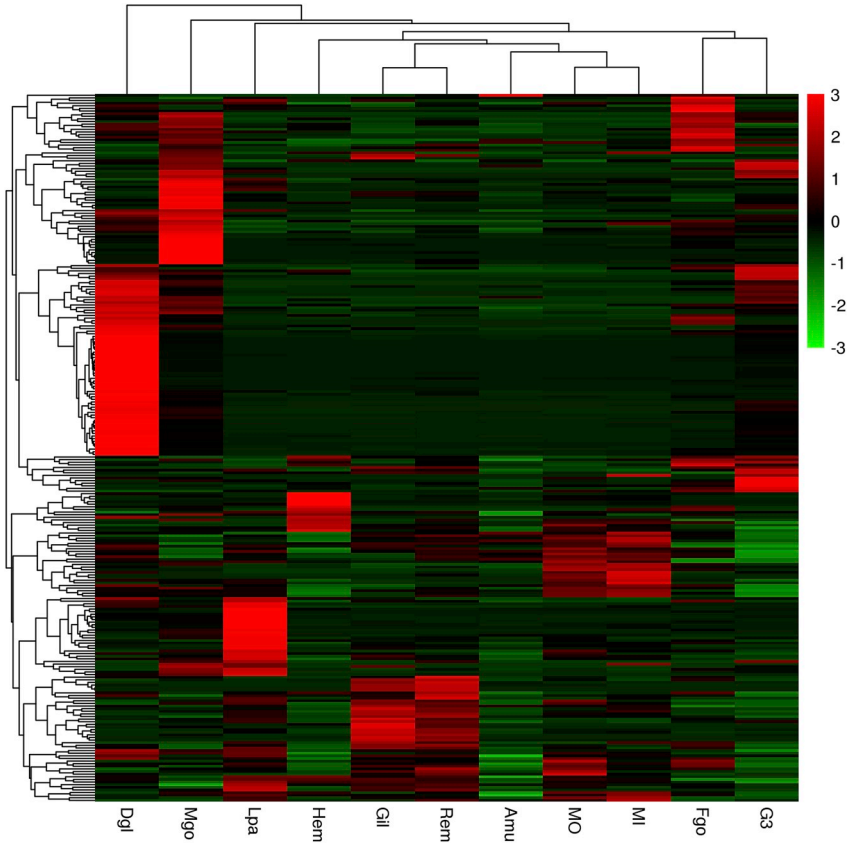

Fig. 6. The mRNA expression level of C-type lectin domain containing proteins (CTLPs) in different oyster tissues. (Dgl, digestive gland; Mgo, male gonad; Lpa, labial palp; Hem, hemocyte; Gil, gill; Rem, mixture of remaining tissues; Amu, adductor muscle; Fgo, female gonad; MO, outer edge mantle; MI, inner pallial mantle; G3, female gonad from F1 offspring of family "G3").

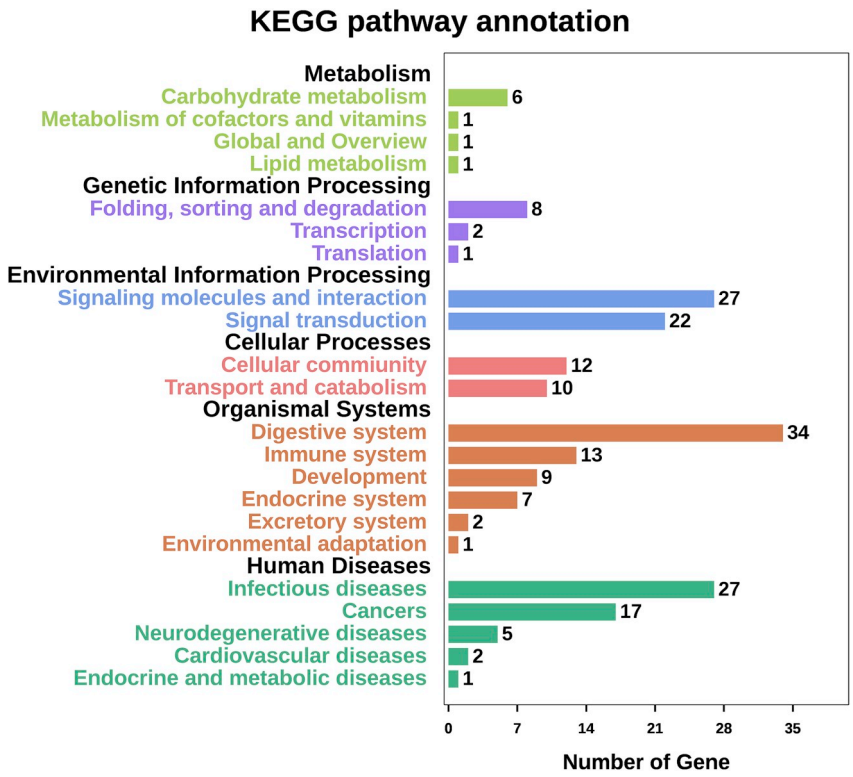

Fig. 7. The KEGG pathway enrichment of Concanavalin A-like lectin domain containing proteins (CALPs).

been identified in animal lectins [3,20]. In *C. gigas*, ten lectin domains were identified, and the most abundant domains were the typical C-type lectin fold, C-type lectin, and Concanavalin A-like lectin domains. C-type lectin domain is one of the most studied domains with the unique structural fold and the requirement of Ca^{2+} for ligand binding [21]. C-type lectin domain is also the most ancient structure during

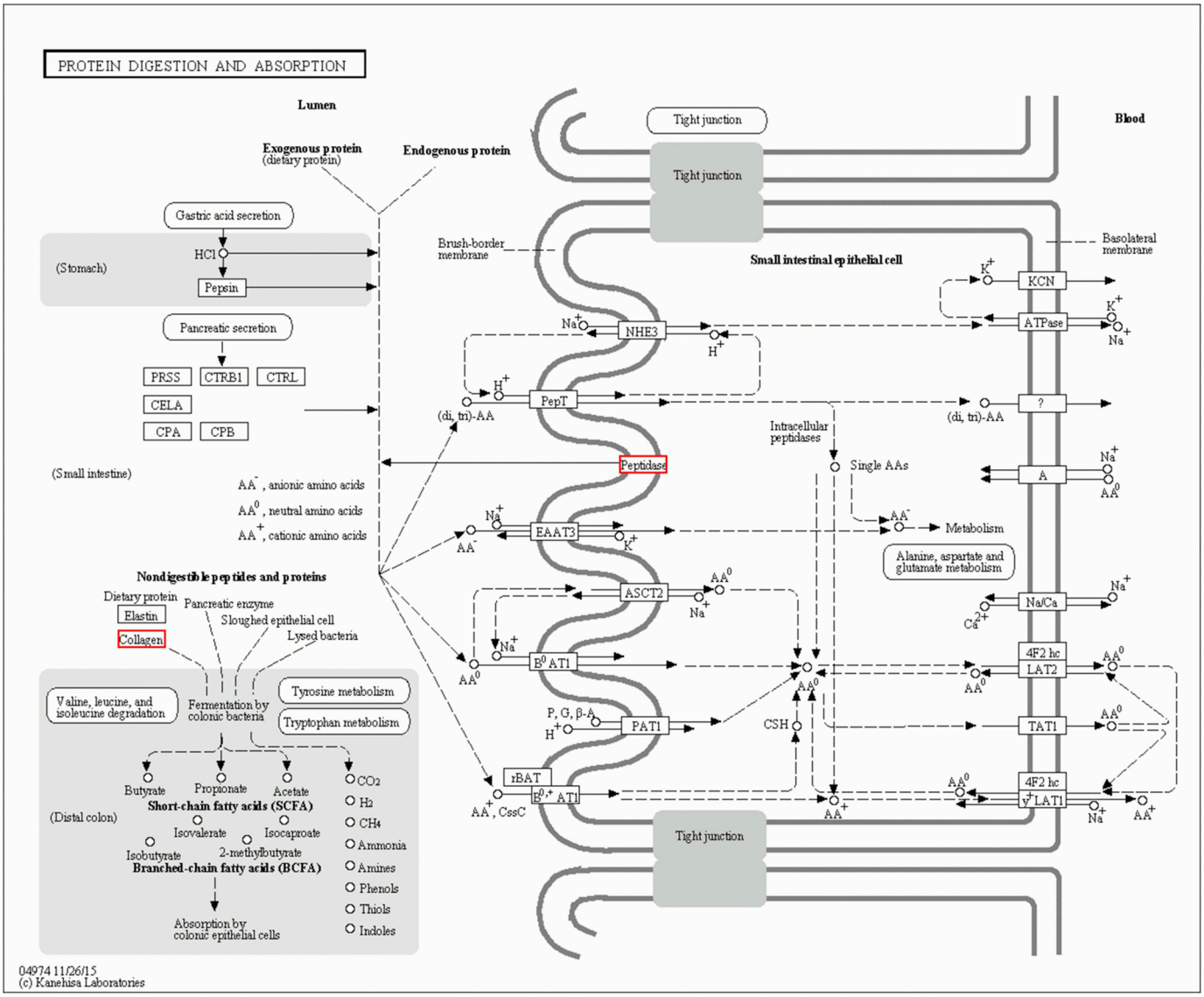

Fig. 8. The Concanavalin A-like lectin domain containing proteins (CALPs) highly enriched in protein digestion and absorption pathway.

evolution, which has presented in the simple multicellular *Trichoplax adhaerens* with large amount (45) [8,22]. In oysters, there is an expansive number (273) of C-type lectin domains, which is larger than that (100) in mammalia *Homo sapiens* [8]. The LDCPs (640 genes) in oyster was also more diversified than that in mammalian. And the abundance of LDCPs in oyster might partially attribute to the various lectin domains encoded by oyster genome.

The lectin domains of oyster LDCPs are found to be frequently fused with other non-lectin domains, which might be another mechanism for LDCPs diversification. Though there are only ten lectin domains identified in oyster, the LDCPs are structurally diversified by associating with a variety of non-lectin domains to constitute chimeric proteins. About 90% of oyster LDCPs contain more than one domain, and 60% of them are chimeric proteins associated with non-lectin domains, which might endow the LDCPs with multiple structural and functional properties. It seems that the lectin domains are more likely to construct chimeric proteins in lower organisms. In animals, there are few reports about the proportions of chimeric lectins, but there are countless chimerolectins in plants, which are even more than the sololectins [23–25]. It might be a universal mechanism for lectins diversification by fusing lectin domain with others, especially in lectin-expansive invertebrates.

The various lectin and non-lectin domains endow LDCPs with more roles other than just as pattern recognition receptors in innate immune system. The detailed knowledge about the domain component of LDCPs is necessary for the better understanding of their functions. In oyster LDCPs, two kinds of lectin domains (the C-type lectin, Concanavalin A-like lectin domains) and two kinds of non-lectin domains (EGF-like domain, MAM domain) were rather abundant in the total 282 identified domains. The EGF-like domain is commonly found in the extracellular domain of membrane-bound proteins, which is important for cell-cell interaction [26,27]. The MAM domain, an overlapping homologous of Concanavalin A-like lectin domains, plays a role in homodimerization of protein-tyrosine phosphatase and is critical for enzyme function [28]. In addition, many peptidase related domains were also identified in oyster, such as peptidase M12A astacin and peptidase metallopeptidase. These non-lectin domains might supply LDCPs with functions more than just carbohydrate-binding activities. As expected, the obvious endopeptidase and metallopeptidase activities were also significantly enriched in LDCPs by GO enrichment. Furthermore, KEGG pathway analysis revealed that the LDCPs were not only significantly enriched in phagosome pathway functioning as pattern recognition receptors in innate immune system, but also enriched in the protein digestion and absorption pathway functioning as peptidases. And it

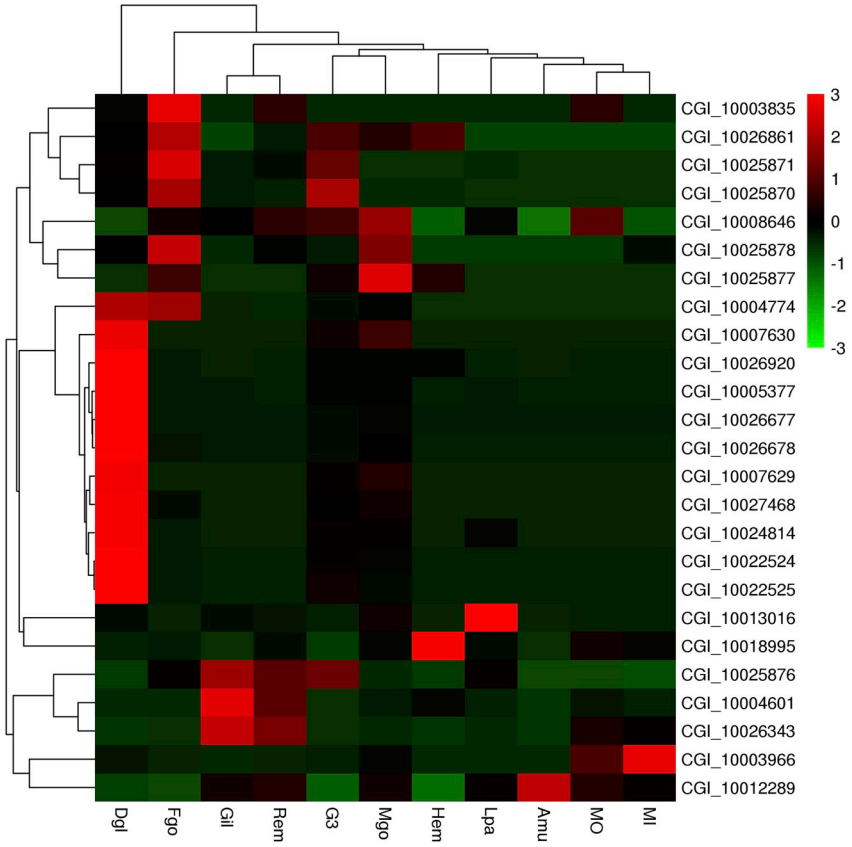

Fig. 9. The heatmap of mRNA expression of Concanavalin A-like lectin domain containing peptidases in different oyster tissues. (Dgl, digestive gland; Mgo, male gonad; Lpa, labial palp; Hem, hemocyte; Gil, gill; Rem, mixture of remaining tissues; Amu, adductor muscle; Fgo, female gonad; MO, outer edge mantle; MI, inner pallial mantle; G3, female gonad from F1 offspring of family "G3").

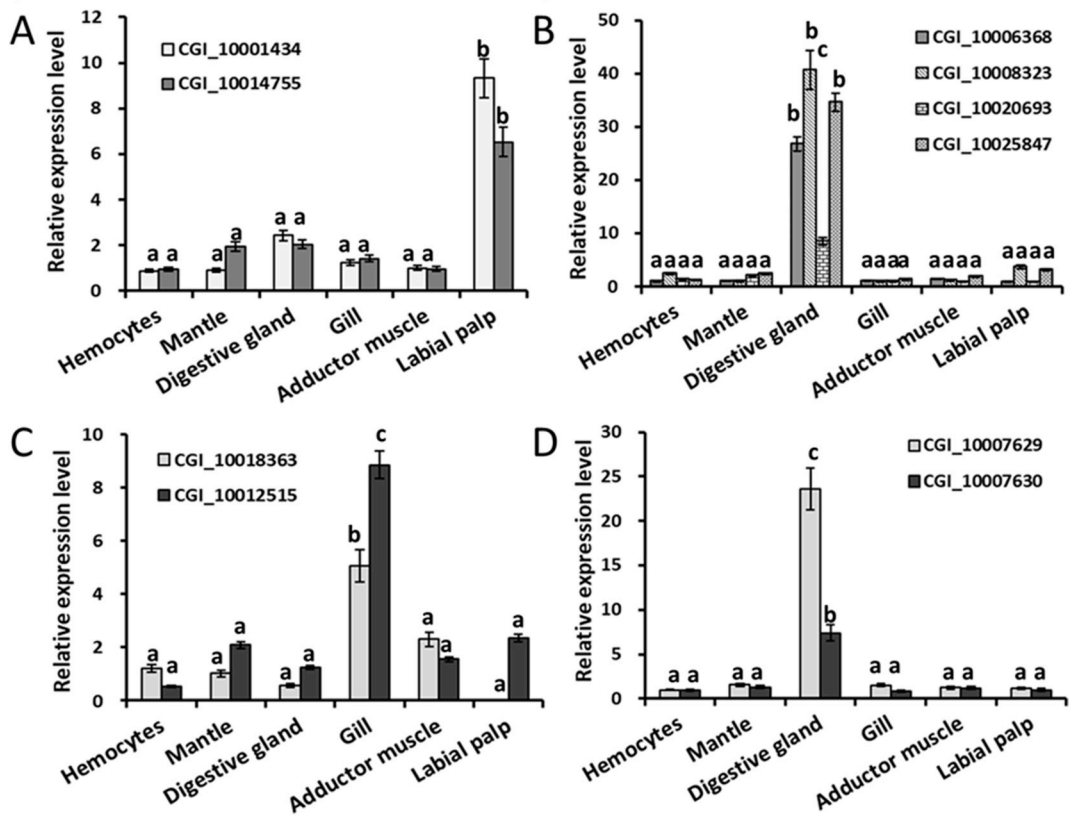

Fig. 10. Relative expression level of ten candidate genes involved in either mucosal immunity or digestive function. Each value is shown as mean ± S.D. (N = 6, $p < 0.05$).

could be speculated that these LDCPs might be classified into different sub-clusters based on their significantly varied functions.

The classical C-type lectin domains in oyster were co-existed mostly with the complement activation-related non-lectin domains, and these proteins were highly expressed in mucosal tissues for mucosal immune protection. C-type lectin fold domain is always overlapping with C-type lectin domain and also presents in non-lectin proteins. In the present study, only the classical C-type lectin domain was further analyzed. The C-type lectins are most well-known as pattern recognition receptors or hemolymph opsonin in lectin pathway of complement in innate immunity [29,30]. In oyster, about two thirds of the C-type lectin domain-containing proteins (CTLPs) were sololectins, which were revealed by KEGG enrichment to function as pattern recognition receptors or phagocytosis promoting receptors, such as the macrophage mannose receptors. The other chimeric CTLPs were mainly fused with complement activation-related domains, such as CUB domain (for complement C1r/C1s, Uegf, Bmp1) and Sushi/SCR/CCP domain. The CUB domain is a component of mammalian complement C1s/C1r and a protease cleaving C4, which is involved in complement activation [31,32]. The Sushi domain is known to be involved in the binding of several complement factors to fragments C3b and C4b [33]. The chimeric lectins fused with these two domains might facilitate the complement activation of lectin pathway in a shortcut manner compared with that in vertebrates. It was suggested that the CTLPs in oysters might be mainly functioned in immune defense and play important roles in innate immunity. Correspondingly, most of the CTLPs were found to be highly expressed in mucosal tissues, such as digestive gland, labial palp and gills. And these genomic expression data were further confirmed by RT-PCR detection. As filter-feeding marine animals, oysters exchange oxygen and filter foods in water environment all the times with the gills, labial palp and digestive gland, so it is quit available at these mucosal tissues for pathogenic invaders [34,35]. Reasonably, the highly expressed CTLPs at these tissues could provide necessary mucosal immune protections for hosts.

By comparison, the Concanavalin A-like lectin domain was mostly constructed into chimeric proteins with various functions, and significantly enriched with peptidase activities functioning in digestive tissues of oyster. About 89% of the Concanavalin A-like lectin domain containing proteins (CALPs) were chimerolectins, and most of them were fused with Peptidase domain and EGF domain. As expected, the CALPs were mainly enriched in the protein digestion and absorption pathway, and the ECM-receptor interaction pathway. Most peptidase domain containing chimeric CALPs were found to be highly expressed in digestive gland, which was validated by RT-PCR results. It could be inferred that the CTLPs were most responsible for immune functions, while the CALPs for digestive functions in oyster *C. gigas*, which had been rarely reported in other animals.

In conclusion, ten lectin domains and an expansive 640 lectin domain containing proteins were identified and functionally analyzed in oyster *C. gigas*. The C-type lectin fold and C-type lectin domains were the most abundant domains. The lectin domains kept relative independence with each other, but frequently fused with non-lectin domains, which supplied LDCPs with more diversification of structure and function. The classical C-type lectin domains co-existed mostly with complement activation-related non-lectin domains, and these proteins were highly expressed in mucosal tissues. The Concanavalin A-like lectin domain were mostly constructed into chimeric proteins with various function, and significantly enriched with peptidase activities functioning in digestive tissues. All results revealed the involvement in mucosal immunity and digestive functions of oyster lectin domain containing proteins, which provided a fresh idea about the functions of invertebrate lectin family.

Acknowledgements

We are grateful to all the laboratory members for their technical advice and helpful discussions. This research was supported by National Science Foundation of China (No. 31802330), National Key R&D Projects (2018YFD0900502), AoShan Talents Cultivation Program Supported by Qingdao National Laboratory for Marine Science and Technology (No. 2017ASTCP-OS13), and the Fund for Outstanding Talents and Innovative Team of Agricultural Scientific Research, the Distinguished Professor of Liaoning, Key R&D Program of Liaoning Province (201703165, 201703148), and Natural Science Foundation of Liaoning, China (20170520056), Dalian High Level Talent Innovation Support Program (2015R020), and Talented Scholars in Dalian Ocean University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fsi.2019.03.067.

References

[1] V. Lakhtin, M. Lakhtin, V. Alyoshkin, Lectins of living organisms. The overview, Anaerobe 17 (6) (2011) 452–455.
[2] N. Sharon, H. Lis, History of lectins: from hemagglutinins to biological recognition molecules, Glycobiology 14 (11) (2004) 53–62.
[3] D.C. Kilpatrick, Animal lectins: a historical introduction and overview, Biochim. Biophys. Acta Gen. Subj. 1572 (2–3) (2002) 187–197.
[4] L. Eggermont, B. Verstraeten, E.J.M. Van Damme, Genome-wide screening for lectin motifs in Arabidopsis thaliana, Plant Genome 10 (2) (2017) 1–17.
[5] G. Cordara, W. Egge-Jacobsen, H.T. Johansen, H.C. Winterf, I.J. Goldstein, K. Sandvig, et al., *Marasmius oreades* agglutinin (MOA) is a chimerolectin with proteolytic activity, Biochem. Biophys. Res. Commun. 408 (3) (2011) 405–410.
[6] B. Pees, W. Yang, A. Zárate-Potes, H. Schulenburg, K. Dierking, High innate immune specificity through diversified C-type lectin-like domain proteins in invertebrates, J. Innate Immun. 8 (2) (2016) 129–142.
[7] J. Unitt, D. Hornigold, Plant lectins are novel Toll-like receptor agonists, Biochem. Pharmacol. 81 (11) (2011) 1324–1328.
[8] M. Lakhtin, V. Lakhtin, V. Alyoshkin, S. Afanasyev, Lectins of beneficial microbes: system organisation, functioning and functional superfamily, Benef. Microbes 2 (2) (2011) 155–165.
[9] L. Bauters, D. Naalden, G. Gheysen, The distribution of lectins across the phylum Nematoda: a genome-wide search, Int. J. Mol. Sci. 18 (1) (2017) 91.
[10] L. Zhang, L. Li, X. Guo, G.W. Litman, L.J. Dishaw, G. Zhang, Massive expansion and functional divergence of innate immune genes in a protostome, Sci. Rep. 5 (2015) 8693.
[11] G. Zhang, X. Fang, X. Guo, L. Li, R. Luo, F. Xu, et al., The oyster genome reveals stress adaptation and complexity of shell formation, Nature 490 (7418) (2012) 49.
[12] W. Wang, M. Li, L. Wang, H. Chen, Z. Liu, Z. Jia, et al., The granulocytes are the main immunocompetent hemocytes in *Crassostrea gigas*, Dev. Comp. Immunol. 67 (2017) 221–228.
[13] R.S. Singh, A.K. Walia, Microbial lectins and their prospective mitogenic potential, Crit. Rev. Microbiol. 40 (4) (2014) 329–347.
[14] P. Singh, L.Z. Zimmerli, Lectin receptor kinases in plant innate immunity, Front. Plant Sci. 4 (2013) 124.
[15] H. Rüdiger, H.J. Gabius, Plant lectins: occurrence, biochemistry, functions and applications, Glycoconj. J. 18 (8) (2001) 589–613.
[16] W.I. Weis, M.E. Taylor, K. Drickamer, The C-type lectin superfamily in the immune system, Immunol. Rev. 163 (1) (1998) 19–34.
[17] K. Drickamer, M.E. Taylor, Biology of animal lectins, Annu. Rev. Cell Biol. 9 (1) (1993) 237–264.
[18] K. De Schutter, E.J.M. Van Damme, Protein-carbohydrate interactions as part of plant defense and animal immunity, Molecules 20 (5) (2015) 9029–9053.
[19] M.E. Taylor, K. Drickamer, Convergent and divergent mechanisms of sugar recognition across kingdoms, Curr. Opin. Struct. Biol. 28 (2014) 14–22.
[20] K. Drickamer, C-type lectin-like domains, Curr. Opin. Struct. Biol. 9 (5) (1999) 585–590.
[21] I.M. Dambuza, G.D. Brown, C-type lectins in immunity: recent developments, Curr. Opin. Immunol. 32 (2015) 21–27.
[22] H. Suga, Z. Chen, A. De Mendoza, A. Sebé-Pedrós, M.W. Brown, E. Kramer, et al., The Capsaspora genome reveals a complex unicellular prehistory of animals, Nat. Commun. 4 (2013) 3325.
[23] A.N. Zelensky, J.E. Gready, The C-type lectin like domain superfamily, FEBS J. 272 (24) (2005) 6179–6217.
[24] I.C. Michelow, M. Dong, B.A. Mungall, M.L. Yantosca, C. Lear, X. Ji, et al., A novel L-ficolin/mannose-binding lectin chimeric molecule with enhanced activity against Ebola virus, J. Biol. Chem. M110 (2010) 106260.
[25] K. Yamamoto, Y. Konami, T. Osawa, T. Irimura, Alteration of the carbohydrate-binding specificity of the *Bauhinia purpurea* lectin through the preparation of a chimeric lectin, J. Biochem. 111 (1) (1992) 87–90.
[26] I.D. Campbell, P. Bork, Epidermal growth factor-like modules, Curr. Opin. Struct. Biol. 3 (3) (1993) 385–392.
[27] J. Engel, EGF-like domains in extracellular matrix proteins: localized signals for

growth and differentiation? FEBS Lett. 251 (1–2) (1989) 1–7.
[28] V.B. Cismasiu, S.A. Denes, H. Reiländer, H. Michel, S.E. Szedlacsek, The MAM (meprin/A5-protein/PTPmu) domain is a homophilic binding site promoting the lateral dimerization of receptor-like protein tyrosine phosphatase mu, J. Biol. Chem. 279 (26) (2004) 26922–26931.
[29] T.B.H. Geijtenbeek, S.I. Gringhuis, Signalling through C-type lectin receptors: shaping immune responses, Nat. Rev. Immunol. 9 (7) (2009) 465.
[30] J.C. Hoving, G.J. Wilson, G.D. Brown, Signalling C-Type lectin receptors, microbial recognition and immunity, Cell Microbiol. 16 (2) (2014) 185–194.
[31] P. Bork, G. Beckmann, The CUB domain: a widespread module in developmentally regulated proteins, J. Mol. Biol. 231 (2) (1993) 539–545.
[32] S.V. Petersen, S. Thiel, J.C. Jensenius, The mannan-binding lectin pathway of complement activation: biology and disease association, Mol. Immunol. 38 (2–3) (2001) 133–149.
[33] T. Izumi, T. Hashiguchi, G. Castaman, A. Tosetto, F. Rodeghiero, A. Girolami, et al., Type I factor XIII deficiency is caused by a genetic defect of its b subunit: insertion of triplet AAC in exon III leads to premature termination in the second Sushi domain, Blood 87 (7) (1996) 2769–2774.
[34] E.C. Lavelle, G. Grant, A. Pusztai, U. Pfüller, D.T. O'Hagan, Mucosal immunogenicity of plant lectins in mice, Immunology 99 (1) (2000) 30–37.
[35] M.A. Clark, B.H. Hirst, M.A. Jepson, Lectin-mediated mucosal delivery of drugs and microparticles, Adv. Drug Deliv. Rev. 43 (2–3) (2000) 207–223.