

# Genome-Wide Analysis of the Cyclin-Dependent Kinases (CDK) and Cyclin Family in Molluscs

YANG Qiong<sup>1)</sup>, YU Hong<sup>1), 2), \*</sup>, and LI Qi<sup>1), 2)</sup>

1) Key Laboratory of Mariculture of Ministry of Education, Ocean University of China, Qingdao 266003, China

2) Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

(Received September 7, 2020; revised November 16, 2020; accepted February 9, 2021)  
© Ocean University of China, Science Press and Springer-Verlag GmbH Germany 2021

**Abstract** Cell cycle regulation that plays a pivotal role during organism growth and development is primarily driven by cyclin-dependent kinases (CDKs) and Cyclins. Although CDK and Cyclin genes have been characterized in some animals, the studies of CDK and Cyclin families in molluscs, the ancient bilaterian groups with high morphological diversity, is still in its infancy. In this study, we identified and characterized 95 CDK genes and 114 Cyclin genes in seven representative species of molluscs, including *Octopus bimaculoides*, *Pomacea canaliculata*, *Biomphalaria glabrata*, *Lottia gigantea*, *Mizuhopecten yessoensis*, *Crassostrea gigas* and *Aplysia californica*. Genes in CDK and Cyclin families were grouped into eight and 15 subfamilies by phylogenetic analysis, respectively. It should be noted that duplication of *CDK9* gene was detected in *P. canaliculata*, *L. gigantea* and *M. yessoensis* genomes, which has never been recorded in animals. It is speculated that duplication may be the main course of expansion of the *CDK9* subfamily in the three molluscs, which also sheds new light on the function of *CDK9*. In addition, Cyclin B is the largest subfamily among the Cyclin family in the seven molluscs, with the average of three genes. Our findings are helpful in better understanding CDK and Cyclin function and evolution in molluscs.

**Key words** molluscs; CDK; Cyclin; genome-wide analysis

## 1 Introduction

Cyclin-dependent kinase (CDK) is a large family of serine/ threonine-specific kinases that control the eukaryotic cell cycle progression by binding cyclin partners (Pines, 1995; Johnson and Walker, 1999; Murray, 2004). During the cell cycle, cyclins accumulate and degrade periodically, and CDK/cyclin complexes are activated at specific cycle phases (Morgan, 1997). In addition to cell cycle regulation, CDKs and Cyclins are also involved in transcription, RNA processing, apoptosis and neurogenesis (Pines, 1995; Malumbres, 2011; Hydbring *et al.*, 2016).

The original member of CDK family was found in genetic screens for feast (Beach *et al.*, 1982), and the first Cyclin gene was identified in sea urchin eggs (Evans *et al.*, 1983). Subsequently, many members of the two families were identified in other species based on the conserved protein kinase domains. As the results, 6 to 8 CDKs and 9 to 15 Cyclins were identified in Fungi; 8 to 52 Cyclin-like protein and 25 to 29 CDK proteins were detected in plants (Wang *et al.*, 2004; La *et al.*, 2006; Ma *et al.*, 2013); and 11 to 28 Cyclin and 14 to 20 CDK genes were found in ani-

mals (Cao *et al.*, 2014; Malumbres, 2014). The increased complexity of cell cycle regulation in animals and plants might account for the more members of CDK and Cyclin families than in fungi, while the increased gene duplication events in plant genomes might explain the higher number of CDKs and Cyclins than in animals (Wang *et al.*, 2004). Researches on invertebrates have focused on some organisms, such as nematodes and sea urchins. Six CDK and 11 Cyclin genes are present in *Caenorhabditis elegans* (Boxem, 2006), 11 CDK and 14 Cyclin genes have been identified in *Strongylocentrotus purpuratus* (Cao *et al.*, 2014). Despite their functions in eukaryotic cell cycle regulation, CDKs and Cyclins have undergone an extraordinary degree of evolutionary divergence and specialization. Therefore, investigation of the evolutionary history of CDKs and Cyclins will enhance our understanding of animals and plants evolution and organism development.

Several studies have been conducted to disclose the evolutionary features of CDK and Cyclin families. *Cyclin A*, *Cyclin B3* and *Cyclin B* were conserved in animals and the number of *Cyclin A* and *B* varied among different organisms (Nieduszynski *et al.*, 2002). It was indicated that cell-cycle related CDKs became more evolutionarily and functionally diverse with transcription complexity increasing (Guo and Stiller, 2004). A comparative phylogenetic analysis of Cyclins from protists to plants, fungi and ani-

\* Corresponding author. Tel: 0086-532-82032773  
E-mail: hongyu@ouc.edu.cn

mals suggests that Cyclins can be divided into three groups (group I, groups II and III) (Ma *et al.*, 2013). Phylogenetic analysis of CDK and Cyclin proteins in 18 premetazoan lineages indicate that CDK4/6 subfamily and eumetazoans emerged simultaneously, while the evolutionary conservation of the Cyclin-D subfamily also tightly linked with eumetazoan appearance (Cao *et al.*, 2014).

Mollusca, the most speciose phylum in the marine realm with highly diverse body forms and lifestyles, plays an important role in evolution and ecosystem. Many molluscs especially the bivalve can attain old ages (over 100 years old), such as the Geoduck clam (*Panopea abrupta*), the freshwater pearl mussel (*Margaritifera margaritifera*), and the ocean quahog (*Arctica islandica*) (Ziuganov *et al.*, 2000; Strom *et al.*, 2004; Wanamaker *et al.*, 2008) which are increasingly regarded as longevity models (Abele *et al.*, 2009; Bodnar, 2009). Despite remarkable evolutionary and biological significance, knowledge of CDK and Cyclin families in Mollusca is still in its infancy. Recently along with the rapid development of high-throughput sequencing technologies, the number of sequenced molluscan genomes has been increased rapidly (Simakov *et al.*, 2013; Albertin *et al.*, 2015; Wang *et al.*, 2017), which provides a unique opportunity to enhance our understanding of CDK and Cyclin families in molluscs.

In this study, we identified and characterized CDK and Cyclin genes in seven species of mollusc based on the genome-wide data, including *Octopus bimaculoides*, *Pomacea canaliculata*, *Biomphalaria glabrata*, *Lottia gigantea*, *Mizuhopecten yessoensis*, *Crassostrea gigas* and *Aplysia californica*. Phylogenetic analysis and gene structure comparison of these proteins are conducted. The results reveal detailed evolutionary information of CDKs and Cyclin partners, providing insights into potential function of CDK and Cyclin genes in molluscs.

## 2 Materials and Methods

### 2.1 Database Searching and Identification of CDK and Cyclin Genes

To identify CDK and Cyclin genes in seven representative species of mollusc, including *O. bimaculoides*, *P. canaliculata*, *B. glabrata*, *L. gigantea*, *M. yessoensis*, *C. gigas*, and *A. californica*, their whole genomic sequence database from NCBI were searched using the query sequence generated from whole CDK and Cyclin family members in humans (*Homo sapiens*), vase tunicate (*Ciona intestinalis*), fruit fly (*Drosophila melanogaster*), purple sea urchin (*Strongylocentrotus purpuratus*), and starlet sea anemone (*Nematostella vectensis*) (Cao *et al.*, 2014). Tblastn was used to get the initial pool of CDK and Cyclin with a minimum e-value of 1e-5. After deleting the repeated entries, a unique set of sequences were kept for further analysis. All putative CDK and Cyclin family proteins collected by Blast searching were carried out a preliminary phylogenetic analysis. We verified the putatively identified Cyclin proteins by searching against SMART databases (<http://smart.emblheidelberg.de/>).

### 2.2 Phylogenetic Analysis and Classification of the CDK and Cyclin Gene Families

To investigate the evolutionary relationship of the CDK and Cyclin families, the CDK and Cyclin amino acid sequences of seven mollusc species and several representative metazoans, including *H. sapiens*, *C. intestinalis*, *D. melanogaster*, *N. vectensis*, *S. purpuratus* and *Danio rerio*, were used to perform the phylogenetic analysis. All the sequences of CDK genes were aligned using ClustalW (<http://www.ebi.ac.uk/clustalw/>) with the default parameters. Gblocks (<http://molevol.cmima.csic.es/castresana/index.html>) was used to eliminate poorly aligned positions and divergent regions. The phylogenetic trees were built with MEGA 7.0 using the Neighbor-joining (NJ) method with 1000 repetitions for the bootstrap test. Since the Cyclin family is not conserved enough like CDK family, the genes of Cyclin family were aligned with MAFFT v7.402 (Katoh and Standley, 2013) and poorly aligned sequences were removed. Only the conserved region (Cyclin-N and -C domains) were used for further phylogenetic analyses to identify the maximum likelihood (ML). ML constructed using RAxML v8.2.12 (Stamatakis, 2014) as implemented in the CIPRES Science Gateway v. 3.3 (<http://www.phylo.org/index.php>) with 1000 bootstrap and LG model. The tree was displayed with Interactive Tree of Life (ITOL, <https://itol.embl.de/>).

### 2.3 Sequence Analysis and Structural Characterization

The Compute PI/MW tool at Expert Protein Analysis System (ExPASy) site ([https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)) was used to calculation coding sequence (CDS), length molecular weight (MW) and isoelectric point (pI). According to the online software BUSCA (<http://busca.biocomp.unibo.it/>), all the subcellular localizations could be predicted. We used the MEME (<http://meme-suite.org/tools/meme>) to analyze the motifs of CDK and Cyclin proteins with the following parameter: minimum width of motif, six; maximum width of motif, 50; and number of motifs, 10. The ITOL was used to visualize the results. According to the result of the GSDS software (Hu *et al.*, 2015), the exon-intron structures of CDKs and Cyclins were shown in ITOL. To further analyze CDK9 genes, we use CLUSTAL W (<https://www.genome.jp/tools-bin/clustalw>) to get sequence alignment of the CDK9 genes from seven mollusc species, *H. sapiens* (Has-CDK9) and *S. purpuratus* (Spu-CDK9). Sequence alignment was exported into ESPRIPT 3.0 (<http://espript.ibcp.fr/ESPrIPT/ESPrIPT/>). Structural features were described with CDK9 in human (extracted from the CDK9-CyclinT complex, PDB code: 3TN9).

## 3 Results

### 3.1 Identification of CDK and Cyclin Genes

As summarized in Table 1, a total of 209 genes were identified in the seven molluscs, including 95 genes of CDK family and 114 genes of Cyclin family. The number of CDK genes was from 13 to 15, while the number of Cyclin genes

varied from 13 to 21. The component of CDK and Cyclin family members of each species was showed in Figs.1 and 2.

Table 1 Distribution of CDK and Cyclin family proteins in representative species of Mollusca

Species	Class	CDK	Cyclin
<i>Octopus bimaculoides</i>	Cephalopoda	13	16
<i>Pomacea canaliculata</i>	Gastropoda	14	15
<i>Biomphalaria glabrata</i>	Gastropoda	13	13
<i>Lottia gigantea</i>	Gastropoda	14	18
<i>Aplysia californica</i>	Gastropoda	14	15
<i>Mizuhopecten yessoensis</i>	Bivalvia	14	18
<i>Crassostrea gigas</i>	Bivalvia	13	19
Total		95	114

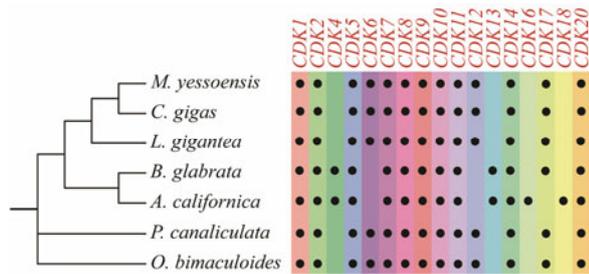


Fig.1 Schematic representation of the distribution of different CDK family members in mollusc species. A black dot indicates the presence of clear homologs of CDK family members. Phylogenetic relationships of these organisms are derived from COI genes using MEGA 7.0 by the neighbor joining.

The bioinformation on the CDK and Cyclin family genes of seven species is provided in Table 2 and Table 3, including name, identifier (ID), number of amino acid (aa), iso-

Table 2 The bioinformation of the CDK family genes in seven mollusc species

Species	Gene name	Gene ID	aa	MV(Da)	pI	SL
<i>Aplysia californica</i>	<i>Cyclin-dependent kinase 1 like</i>	LOC101853987	349	40324.45	5.69	Cytoplasm
	<i>Cyclin-dependent kinase 2 like</i>	LOC101846074	316	35923.41	7.11	Cytoplasm
	<i>Cyclin-dependent kinase 4 like</i>	LOC101849290	418	47098.35	6.80	Mitochondrion
	<i>Cyclin-dependent-like kinase 5</i>	LOC101853437	294	33562.51	7.03	Cytoplasm
	<i>Cyclin-dependent kinase 7 like</i>	LOC101863685	341	38453.22	7.71	Cytoplasm
	<i>Cyclin-dependent kinase 8 like</i>	LOC101862428	473	54112.66	8.57	Cytoplasm
	<i>Cyclin-dependent kinase 9 like</i>	LOC101857999	369	42480.35	9.23	Cytoplasm
	<i>Cyclin-dependent kinase 10 like</i>	LOC101858693	402	45334.39	7.13	Cytoplasm
	<i>Cyclin-dependent kinase 11B like</i>	LOC101864621	779	89745.12	5.75	Nucleus
	<i>Cyclin-dependent kinase 13 like</i>	LOC101847916	1212	135440.16	9.46	Nucleus
	<i>Cyclin-dependent kinase 14 like</i>	LOC101862276	539	60331.54	9.28	Nucleus
	<i>Cyclin-dependent kinase 16 like</i>	LOC101854558	272	31371.04	9.34	Nucleus
	<i>Cyclin-dependent kinase 18 like</i>	LOC101848503	799	89199.05	9.14	Nucleus
	<i>Cyclin-dependent kinase 20 like</i>	LOC101850770	347	39383.58	6.54	Cytoplasm
<i>Biomphalaria glabrata</i>	<i>Cyclin-dependent kinase 1 like</i>	LOC106073179	308	35667.42	6.47	Cytoplasm
	<i>Cyclin-dependent kinase 2 like</i>	LOC106065664	306	34857.50	7.66	Cytoplasm
	<i>Cyclin-dependent kinase 4 like</i>	LOC106060252	402	45768.19	8.14	Cytoplasm
	<i>Cyclin-dependent-like kinase 5</i>	LOC106076847	281	32122.44	8.86	Cytoplasm
	<i>Cyclin-dependent kinase 7 like</i>	LOC106074176	342	38653.85	8.69	Cytoplasm
	<i>Cyclin-dependent kinase 8 like</i>	LOC106060752	493	56610.84	8.48	Cytoplasm
	<i>Cyclin-dependent kinase 9 like</i>	LOC106065241	323	37291.44	9.30	Cytoplasm
	<i>Cyclin-dependent kinase 10 like</i>	LOC106060771	371	42633.41	8.65	Cytoplasm
	<i>Cyclin-dependent kinase 11B like</i>	LOC106053129	800	92309.28	5.2	Nucleus
	<i>Cyclin-dependent kinase 13 like</i>	LOC106051172	1198	134986.17	9.49	Nucleus
<i>Cyclin-dependent kinase 14 like</i>	LOC106070902	518	57886.68	9.26	Nucleus	
<i>Cyclin-dependent kinase 17 like</i>	LOC106056030	526	59768.53	8.89	Cytoplasm	

(to be continued)

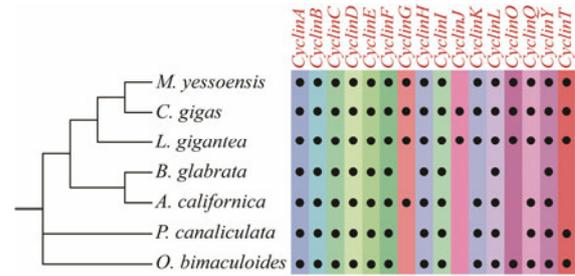


Fig.2 Schematic representation of the distribution of different Cyclin family members in mollusc species. A black dot indicates the presence of clear homologs of Cyclin family members. Phylogenetic relationships of these organisms are derived from COI genes using MEGA 7.0 by the neighbor joining.

electric points (pIs), molecular weight (MW), subcellular localization and protein length. The lengths of the proteins encoded by the CDK genes is from 207 to 1538 aa, with the predicted MW ranging from 23.57 to 172.26 kD. The lengths of the Cyclin proteins varies from 119 to 1162 aa, with the predicted MW varying from 13.48 to 128.23 kD. Among the seven mollusc species, the average lengths of CDK and Cyclin proteins in *L. gigantea* are far shorter than those in the other species. For CDK family, the pI was between 5.2 and 9.65, with an average pI of 7.94. Overall, 72% of the CDK family proteins had a pI more than 7, suggesting that the proteins are rich in acidic amino acids. For Cyclin family, the pI was between 4.56 and 10.21, with an average pI of 7.03. Among the seven molluscs, Cyclin proteins in *P. canaliculata*, *L. gigantea* and *B. glabrata* are rich in acidic amino acids, while Cyclin proteins in *A. californica*, *C. gigas* and *O. bimaculoides* are rich in alkaline amino acids.

(continued)

Species	Gene name	Gene ID	aa	MV(Da)	pI	SL
<i>Biomphalaria glabrata</i>	<i>Cyclin-dependent kinase 20 like</i>	LOC106077158	345	39103.31	6.40	Cytoplasm
	<i>Cyclin-dependent kinase 1</i>	LOC105346130	302	34767.45	8.77	Cytoplasm
	<i>Cyclin-dependent kinase 2</i>	LOC105325500	273	31119.97	6.39	Cytoplasm
	<i>Cyclin-dependent-like kinase 5</i>	LOC105344147	307	35163.58	7.08	Cytoplasm
	<i>Cyclin-dependent kinase 6</i>	LOC105327718	331	37997.29	5.56	Cytoplasm
	<i>Cyclin-dependent kinase 7</i>	LOC105317788	341	38673.04	8.90	Cytoplasm
	<i>Cyclin-dependent kinase 8</i>	LOC105330051	439	50809.92	8.87	Cytoplasm
	<i>Cyclin-dependent kinase 9</i>	LOC105331277	401	45542.43	9.29	Cytoplasm
<i>Crassostrea gigas</i>	<i>Cyclin-dependent kinase 10</i>	LOC105319320	379	43230.11	8.34	Cytoplasm
	<i>Cyclin-dependent kinase 11B</i>	LOC105335549	805	93715.25	5.66	Nucleus
	<i>Cyclin-dependent kinase 12</i>	LOC105347499	1254	140359.71	9.65	Nucleus
	<i>Cyclin-dependent kinase 14</i>	LOC105345433	563	63876.56	9.10	Cytoplasm
	<i>Cyclin-dependent kinase 17</i>	LOC105323040	487	55235.06	8.57	Nucleus
	<i>Cyclin-dependent kinase 20</i>	LOC105329436	343	38937.04	6.54	Cytoplasm
	Hypothetical protein	GeneID20242291	255	29038.46	6.08	Cytoplasm
	Hypothetical protein	GeneID20241075	345	39290.22	6.01	Cytoplasm
	Hypothetical protein	GeneID20244259	469	53223.91	8.75	Cytoplasm
	Hypothetical protein	GeneID20234475	315	35823.18	8.30	Cytoplasm
<i>Lottia gigantea</i>	Hypothetical protein	GeneID20234150	389	45244.09	8.63	Cytoplasm
	Hypothetical protein	GeneID20239796	329	37878.99	8.89	Cytoplasm
	Hypothetical protein	GeneID20247146	385	44433.34	8.05	Cytoplasm
	Hypothetical protein	GeneID20245965	355	40991.65	9.32	Cytoplasm
	Hypothetical protein	GeneID20244620	370	42831.72	9.10	Cytoplasm
	Hypothetical protein	GeneID20246230	421	48908.21	8.82	Cytoplasm
	Hypothetical protein	GeneID20246046	338	38491.45	7.01	Cytoplasm
	Hypothetical protein	GeneID20246924	357	40843.33	6.65	Cytoplasm
	Hypothetical protein	GeneID20245758	300	40843.33	7.73	Cytoplasm
	Hypothetical protein	GeneID20248236	302	34804.35	8.26	Cytoplasm
	<i>Cyclin-dependent kinase 1 like</i>	LOC110448550	304	35107.79	8.58	Cytoplasm
	<i>Cyclin-dependent kinase 2 like</i>	LOC110446820	209	23567.31	8.88	Cytoplasm
	<i>Cyclin-dependent-like kinase 5</i>	LOC110443310	306	35121.21	6.72	Cytoplasm
	<i>Cyclin-dependent kinase 6 like</i>	LOC110452126	334	37547.70	5.25	Cytoplasm
	<i>Cyclin-dependent kinase 7 like</i>	LOC110443310	340	38786.85	7.16	Cytoplasm
	<i>Cyclin-dependent kinase 8 like</i>	LOC110456883	462	53195.86	8.87	Cytoplasm
	<i>Cyclin-dependent kinase 9 like</i>	LOC110455838	396	45303.21	9.19	Cytoplasm
	<i>Mizuhopecten yessoensis</i>	<i>Cyclin-dependent kinase 9 like</i>	LOC110441367	406	46812.90	9.22
<i>Cyclin-dependent kinase 10 like</i>		LOC110443388	381	43332.29	8.42	Cytoplasm
<i>Cyclin-dependent kinase 11B like</i>		LOC110458839	786	91502.18	5.28	Nucleus
<i>Cyclin-dependent kinase 12 like</i>		LOC110465128	1509	168036.26	9.43	Nucleus
<i>Cyclin-dependent kinase 14 like</i>		LOC110460375	388	43443.78	6.90	Cytoplasm
<i>Cyclin-dependent kinase 17 like</i>		LOC110448834	462	52254.92	8.03	Nucleus
<i>Cyclin-dependent kinase 20 like</i>		LOC110458038	343	38809.74	6.13	Cytoplasm
<i>Cyclin-dependent kinase 1 like</i>		LOC106880228	304	35059.48	7.09	Cytoplasm
<i>Cyclin-dependent kinase 2 like</i>		LOC106876684	277	31959.04	6.76	Cytoplasm
<i>Cyclin-dependent-like kinase 5</i>		LOC106880362	296	33712.73	7.59	Cytoplasm
<i>Cyclin-dependent kinase 6 like</i>		LOC106881146	281	32143.16	7.63	Cytoplasm
<i>Cyclin-dependent kinase 7 like</i>		LOC106883163	284	32319.45	6.01	Cytoplasm
<i>Cyclin-dependent kinase 8 like</i>		LOC106878348	461	52741.77	9.05	Cytoplasm
<i>Octopus bimaculoides</i>		<i>Cyclin-dependent kinase 9 like</i>	LOC106871831	366	42452.25	9.17
	<i>Cyclin-dependent kinase 10 like</i>	LOC106884424	378	43703.84	8.87	Cytoplasm
	<i>Cyclin-dependent kinase 11B like</i>	LOC106872191	819	95046.49	6.18	Nucleus
	<i>Cyclin-dependent kinase 12 like</i>	LOC106872449	1538	172264.59	9.43	Nucleus
	<i>Cyclin-dependent kinase 14 like</i>	LOC106877593	511	57441.61	8.73	Nucleus
	<i>Cyclin-dependent kinase 17 like</i>	LOC106879770	485	55107.16	8.54	Nucleus
	<i>Cyclin-dependent kinase 20 like</i>	LOC106883147	277	31332.10	7.66	Cytoplasm
	<i>Cyclin-dependent kinase 1 like</i>	LOC112558173	302	34856.67	8.90	Cytoplasm
<i>Pomacea canaliculata</i>	<i>Cyclin-dependent kinase 2 like</i>	LOC112556656	299	34252.60	8.34	Cytoplasm
	<i>Cyclin-dependent-like kinase 5</i>	LOC112564382	296	34090.17	6.46	Cytoplasm
	<i>Cyclin-dependent kinase 6 like</i>	LOC112556232	352	40137.83	6.35	Cytoplasm
	<i>Cyclin-dependent kinase 7 like</i>	LOC112553685	347	39426.77	9.08	Cytoplasm
	<i>Cyclin-dependent kinase 8 like</i>	LOC112566334	461	53007.09	9.34	Cytoplasm

(to be continued)

(continued)

Species	Gene name	Gene ID	aa	MV(Da)	pI	SL
<i>Pomacea canaliculata</i>	<i>Cyclin-dependent kinase 9 like</i>	LOC112574862	374	43554.36	9.09	Cytoplasm
	<i>Cyclin-dependent kinase 9 like</i>	LOC112575238	365	42398.45	9.17	Cytoplasm
	<i>Cyclin-dependent kinase 10 like</i>	LOC112576905	385	44234.19	9.07	Cytoplasm
	<i>Cyclin-dependent kinase 11B like</i>	LOC112556334	876	102025.48	6.15	Endomembrane system
	<i>Cyclin-dependent kinase 12 like</i>	LOC112563814	1366	151859.73	9.37	Nucleus
	<i>Cyclin-dependent kinase 14 like</i>	LOC112569616	518	58116.28	9.22	Nucleus
	<i>Cyclin-dependent kinase 17 like</i>	LOC112571428	474	53459.24	8.76	Nucleus
	<i>Cyclin-dependent kinase 20 like</i>	LOC112571983	348	39541.86	6.79	Cytoplasm

Table 3 The bioinformation of the Cyclin family genes in seven mollusc species

Species	Gene name	Gene ID	aa	MV(Da)	pI	SL
<i>Aplysia californica</i>	<i>Cyclin A like</i>	LOC101862222	438	48632.88	5.98	Organelle membrane
	<i>Cyclin B2 like</i>	LOC106012287	160	18075.21	6.07	Nucleus
	<i>Cyclin B3 like</i>	LOC101852415	442	49746.35	6.25	Organelle membrane
	<i>Cyclin B like</i>	LOC101850386	442	49577.98	7.67	Plasma membrane
	<i>Cyclin C like</i>	LOC101860286	293	34150.51	6.53	Endomembrane system
	<i>Cyclin D2 like</i>	LOC101847634	307	34944.62	5.64	Endomembrane system
	<i>Cyclin E1 like</i>	LOC101845141	444	50530.48	5.72	Nucleus
	<i>Cyclin F like</i>	LOC101850094	1162	128227.24	6.23	Endomembrane system
	<i>Cyclin H like</i>	LOC101847105	344	39935.15	8.59	Organelle membrane
	<i>Cyclin I like</i>	LOC101848393	290	32757.42	6.45	Plasma membrane
	<i>Cyclin I like</i>	LOC101847777	347	39070.38	7.99	Endomembrane system
	<i>Cyclin K like</i>	LOC101863951	593	65646.9	8.75	Plasma membrane
	<i>Cyclin L1 like</i>	LOC101855525	535	61983.32	9.99	Organelle membrane
	<i>Cyclin Y like</i>	LOC101863542	372	42968.88	6.98	Nucleus
	<i>FAM58A like</i>	LOC101847523	235	27387.89	8.42	Endomembrane system
<i>Biomphalaria glabrata</i>	<i>Cyclin A like</i>	LOC106061643	435	49344.56	6.17	Endomembrane system
	<i>Cyclin B3 like</i>	LOC106057744	442	50240.11	8.18	Plasma membrane
	<i>Cyclin B like</i>	LOC106074749	432	48862.72	8.73	Organelle membrane
	<i>Cyclin B like</i>	LOC106071160	450	51562.36	8.05	Endomembrane system
	<i>Cyclin B like</i>	LOC106062617	307	35406.07	5.31	Endomembrane system
	<i>Cyclin C like</i>	LOC106071702	290	33539.92	6.32	Organelle membrane
	<i>Cyclin D2 like</i>	LOC106059492	302	34278.59	5.30	Plasma membrane
	<i>Cyclin E like</i>	LOC106065422	278	31527.09	5.09	Endomembrane system
	<i>Cyclin F like</i>	LOC106067840	1132	125691.36	8.35	Plasma membrane
	<i>Cyclin H like</i>	LOC106054066	326	38195.82	6.63	Organelle membrane
	<i>Cyclin I like</i>	LOC106075664	338	38546.8	6.63	Endomembrane system
	<i>Cyclin L1 like</i>	LOC106078715	516	60730.04	9.97	Organelle membrane
	<i>Cyclin Y like</i>	LOC106056699	358	41199.55	5.48	Nucleus
	<i>Cyclin A</i>	LOC105348733	476	53201.43	5.40	Plasma membrane
	<i>Cyclin B</i>	LOC105329592	496	56623.04	9.08	Organelle membrane
<i>Cyclin B</i>	LOC105317174	432	49017.76	7.16	Organelle membrane	
<i>Cyclin B3</i>	LOC105347974	459	51961.69	8.24	Organelle membrane	
<i>Cyclin C</i>	LOC105328264	283	33192.49	6.00	Organelle membrane	
<i>Cyclin D2</i>	LOC105330764	291	33344.44	5.14	Endomembrane system	
<i>Cyclin E</i>	LOC105332299	481	55701.39	5.32	Mitochondrion	
<i>Cyclin F</i>	LOC105347344	757	85072.35	5.52	Endomembrane system	
<i>Cyclin G1</i>	LOC105343927	350	39638.57	6.01	Endomembrane system	
<i>Cyclin H like</i>	LOC105333009	283	33095.58	8.12	Endomembrane system	
<i>Cyclin I</i>	LOC105343928	326	36977.11	8.04	Nucleus	
<i>Cyclin J</i>	LOC105319735	308	35644.15	6.07	Plasma membrane	
<i>Cyclin K</i>	LOC105343893	579	64506.39	8.89	Cytoplasm	
<i>Cyclin L1</i>	LOC105343487	465	54069.95	10.16	Organelle membrane	
<i>Cyclin O</i>	LOC105335536	371	42229.66	4.67	Plasma membrane	
<i>Cyclin S13-7 like</i>	LOC105326680	392	44702.16	8.32	Organelle membrane	
<i>Cyclin T2</i>	LOC105330254	791	88394.16	9.19	Plasma membrane	
<i>Cyclin Y like</i>	LOC105323010	353	40471.73	6.17	Nucleus	
<i>FAM58A like</i>	LOC105347660	228	26662.03	6.34	Endomembrane system	
<i>Lottia gigantea</i>	Hypothetical protein	Gene ID 20234850	252	29161.5	8.56	Nucleus
	Hypothetical protein	Gene ID20231625	258	29927.73	8.28	Endomembrane system
	Hypothetical protein	Gene ID 20250518	240	28025.58	6.41	Endomembrane system
	Hypothetical protein	Gene ID20251317	190	22095.74	5.99	Endomembrane system

(to be continued)

(continued)

Species	Gene name	Gene ID	aa	MV(Da)	pI	SL
	Hypothetical protein	Gene ID 20246422	268	30977.27	6.39	Endomembrane system
	Hypothetical protein	Gene ID20246595	327	37778.55	8.30	Endomembrane system
	Hypothetical protein	Gene ID20231798	259	30480.69	6.03	Plasma membrane
	Hypothetical protein	GeneID20244950	283	32262.58	7.68	Nucleus
	Hypothetical protein	Gene ID20240526	246	28417.69	8.23	Organelle membrane
	Hypothetical protein	Gene ID 20251369	119	13482.71	7.11	Endomembrane system
<i>Lottia gigantea</i>	Hypothetical protein	Gene ID20230294	597	67949.11	6.74	Endomembrane system
	Hypothetical protein	Gene ID20245912	416	47791.65	5.57	Nucleus
	Hypothetical protein	Gene ID20244919	284	32306.59	5.20	Endomembrane system
	Hypothetical protein	Gene ID 20246312	293	34071.35	8.03	Endomembrane system
	Hypothetical protein	Gene ID20247052	280	32927.36	5.85	Endomembrane system
	Hypothetical protein	Gene ID20239855	373	42369.17	6.72	Organelle membrane
	Hypothetical protein	Gene ID20244560	246	27994.84	5.68	Endomembrane system
	Hypothetical protein	Gene ID 20247260	416	47048.34	5.34	Plasma membrane
	<i>Cyclin A like</i>	LOC110457210	426	48319.24	5.73	Endomembrane system
	<i>Cyclin B2 like</i>	LOC110459617	120	13902.52	4.70	Nucleus
	<i>Cyclin B3 like</i>	LOC110441995	470	53347.17	8.75	Endomembrane system
	<i>Cyclin B like</i>	LOC110443637	460	51643.46	7.64	Organelle membrane
	<i>Cyclin C like</i>	LOC110462430	207	24651.43	6.21	Endomembrane system
	<i>Cyclin D2 like</i>	LOC110454764	293	33657.09	5.15	Plasma membrane
	<i>Cyclin D2 like</i>	LOC110444575	143	16823.79	4.78	Nucleus
	<i>Cyclin E like</i>	LOC110443792	438	49996.98	5.71	Nucleus
<i>Mizuhopecten yessoensis</i>	<i>Cyclin F like</i>	LOC110461504	794	89574.03	6.31	Plasma membrane
	<i>Cyclin G1 like</i>	LOC110441487	370	42327.8	5.36	Endomembrane system
	<i>Cyclin H like</i>	LOC110451027	327	37828.79	8.57	Organelle membrane
	<i>Cyclin I like</i>	LOC110441488	336	38617.73	6.40	Nucleus
	<i>Cyclin K like</i>	LOC110446434	563	61970.42	9.06	Plasma membrane
	<i>Cyclin L1 like</i>	LOC110458915	472	55269.27	10.2	Organelle membrane
	<i>Cyclin O like</i>	LOC110455988	462	51200.11	4.56	Plasma membrane
	<i>Cyclin T2 like</i>	LOC110443050	821	91190.45	9.19	Plasma membrane
	<i>Cyclin Y like</i>	LOC110452085	354	40390.85	6.42	Nucleus
	<i>FAM58A like</i>	LOC110464947	232	27445.67	6.55	Endomembrane system
	<i>Cyclin A3-1</i>	LOC106867971	401	45996.09	7.47	Plasma membrane
	<i>Cyclin A like</i>	LOC106879537	457	52092.23	5.72	Organelle membrane
	<i>Cyclin B3 like</i>	LOC106882904	420	48814.73	7.94	Organelle membrane
	<i>Cyclin B like</i>	LOC106879992	384	43863.63	9.03	Organelle membrane
	<i>Cyclin C like</i>	LOC106867551	290	33916.23	6.45	Endomembrane system
	<i>Cyclin D2 like</i>	LOC106878898	286	32866.25	5.15	Endomembrane system
	<i>Cyclin E like</i>	LOC106874854	436	49233.78	6.47	Nucleus
<i>Octopus bimaculoides</i>	<i>Cyclin F like</i>	LOC106883185	780	87353.83	6.76	Plasma membrane
	<i>Cyclin H like</i>	LOC106879470	313	36671.25	5.83	Endomembrane system
	<i>Cyclin I like</i>	LOC106881559	328	38109.24	8.18	Endomembrane system
	<i>Cyclin K like</i>	LOC106871018	674	74331.63	9.06	Plasma membrane
	<i>Cyclin L1 like</i>	LOC106882358	478	56177.31	10.07	Organelle membrane
	<i>Cyclin O</i>	LOC106872739	360	42281.57	7.93	Plasma membrane
	<i>Cyclin T1 like</i>	LOC106879624	604	66781.68	9.33	Plasma membrane
	<i>Cyclin Y like</i>	LOC106884459	370	42075.54	8.33	Nucleus
	<i>FAM58A like</i>	LOC106879276	230	27006.26	8.92	Endomembrane system
	<i>Cyclin A like</i>	LOC112560487	433	48959.77	5.17	Organelle membrane
	<i>Cyclin B3 like</i>	LOC112555046	332	38130	6.48	Endomembrane system
	<i>Cyclin B like</i>	LOC112555595	314	35548.19	5.29	Endomembrane system
	<i>Cyclin B like</i>	LOC112553650	461	51222.91	8.68	Organelle membrane
	<i>Cyclin C like</i>	LOC112554859	299	34726.07	6.34	Plasma membrane
	<i>Cyclin D2 like</i>	LOC112566123	297	33617.77	5.05	Endomembrane system
	<i>Cyclin E2 like</i>	LOC112577049	467	53680.42	5.92	Endomembrane system
<i>Pomacea canaliculata</i>	<i>Cyclin F like</i>	LOC112564092	763	84712.55	6.36	Plasma membrane
	<i>Cyclin H like</i>	LOC112557535	324	38457.24	6.07	Organelle membrane
	<i>Cyclin I like</i>	LOC112559393	328	37207.27	6.64	Endomembrane system
	<i>Cyclin K like</i>	LOC112570940	570	62518.29	9.09	Plasma membrane
	<i>Cyclin L1 like</i>	LOC112556543	490	56800.93	10.21	Endomembrane system
	<i>Cyclin Q like</i>	LOC112572637	232	26812.06	6.86	Endomembrane system
	<i>Cyclin T1 like</i>	LOC112558079	826	92553.59	8.68	Endomembrane system
	<i>Cyclin Y like</i>	LOC112556592	375	43055.82	6.81	Cytoplasm

### 3.2 Phylogenetic Analysis and Classification of the CDK and Cyclin Genes

To analyze the characteristics of molluscan CDK and Cyclin proteins, and to examine the CDK and Cyclin genes in molluscs and other representative animals in evolutionary terms, phylogenetic tree was constructed using the CDK proteins from the seven molluscs, human, vase tunicate, fruit fly, starlet sea anemone, purple sea urchin and zebrafish. As shown in Fig.3, the CDK family was clustered into eight groups: CDK1 (including *CDK1*, *CDK2* and *CDK3*), CDK4/6 (including *CDK4* and *CDK6*), CDK5 (including *CDK5*, *CDK16*, *CDK17*, *CDK18* and *CDK14*), CDK7, CDK20, CDK8, CDK9 (including *CDK9*, *CDK12* and *CDK13*), and CDK10/11 (including *CDK10* and *CDK11*). Because Cyclin sequences diverged greatly, a reliable Cyclin phylogenetic tree failed to be obtained between molluscs and other animals like CDKs. We constructed ML trees from the seven organisms in Mollusca, human, purple sea urchin, part genes of vase tunicate and fruit fly (Fig.4). According to the phylogenetic tree, the Cyclin family in molluscs could be divided into 15 groups: Cyclin A, Cyclin B, Cyclin C, Cyclin D, Cyclin E, Cyclin F, Cyclin G/I, Cyclin H, Cyclin J, Cyclin K, Cyclin L, Cyclin O, Cyclin Q, Cyclin T and Cyclin Y (Fig.4). Cyclin B subfamily has the largest number of members (2 Cyclin B-like and a Cyclin B3) in Mollusca.

netic tree failed to be obtained between molluscs and other animals like CDKs. We constructed ML trees from the seven organisms in Mollusca, human, purple sea urchin, part genes of vase tunicate and fruit fly (Fig.4). According to the phylogenetic tree, the Cyclin family in molluscs could be divided into 15 groups: Cyclin A, Cyclin B, Cyclin C, Cyclin D, Cyclin E, Cyclin F, Cyclin G/I, Cyclin H, Cyclin J, Cyclin K, Cyclin L, Cyclin O, Cyclin Q, Cyclin T and Cyclin Y (Fig.4). Cyclin B subfamily has the largest number of members (2 Cyclin B-like and a Cyclin B3) in Mollusca.

### 3.3 CDK and Cyclin Gene Structure and Conserved Motifs

To further interpret the structural diversity of CDK and Cyclin proteins, the gene structure and conserved motifs

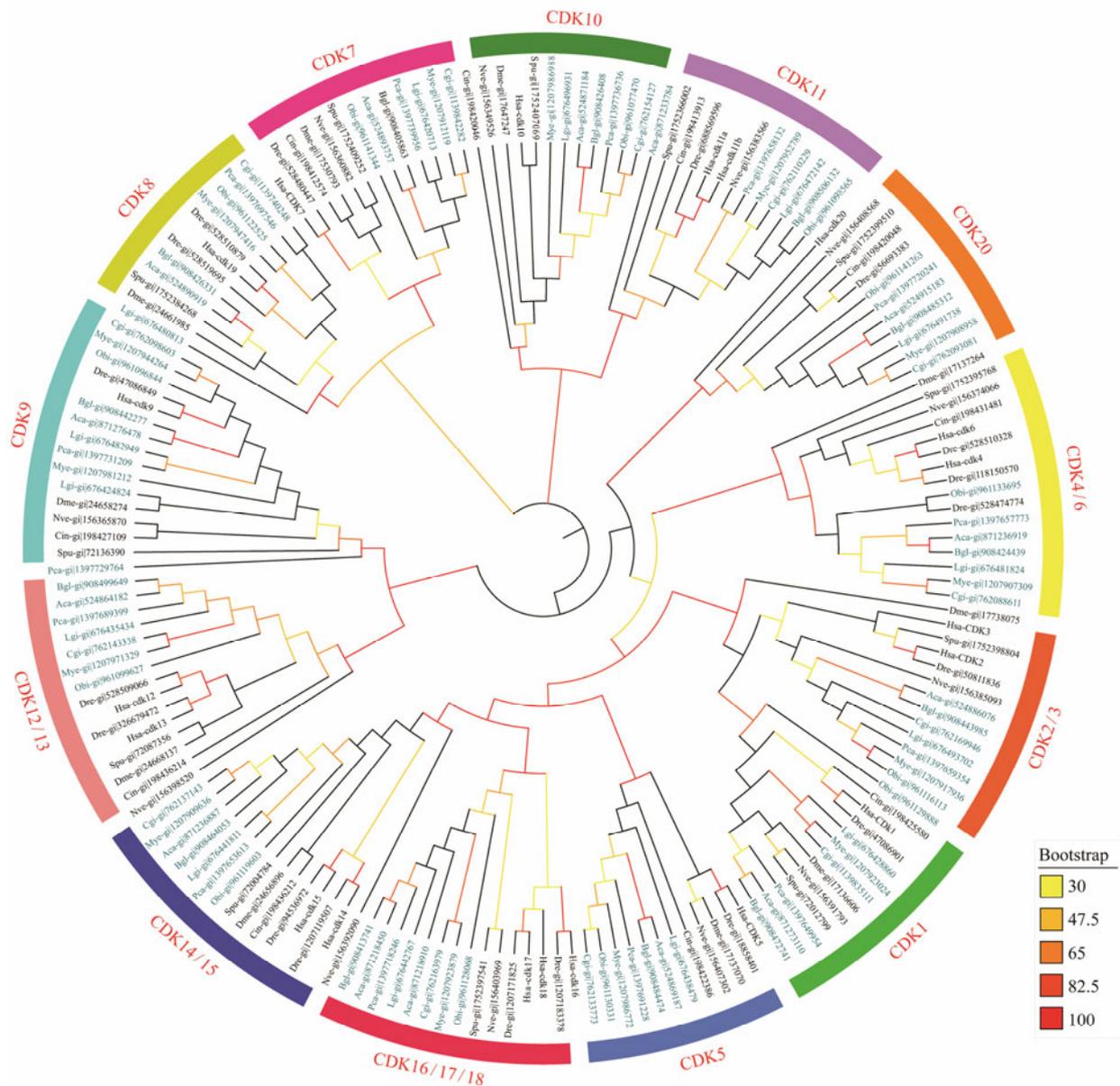


Fig.3 Phylogenetic relationships of CDK genes in Mollusca and several representative metazoans, including human, vase tunicate, fruit fly, starlet sea anemone, purple sea urchin and zebrafish. The phylogenetic tree is constructed using the neighbor-joining (NJ) method with 1000 bootstrap. The bootstrap values are represented by various colors. The CDK genes of Mollusca are marked with blue and other organisms' CDK genes are with black. All proteins are labeled with species names followed by accession numbers.

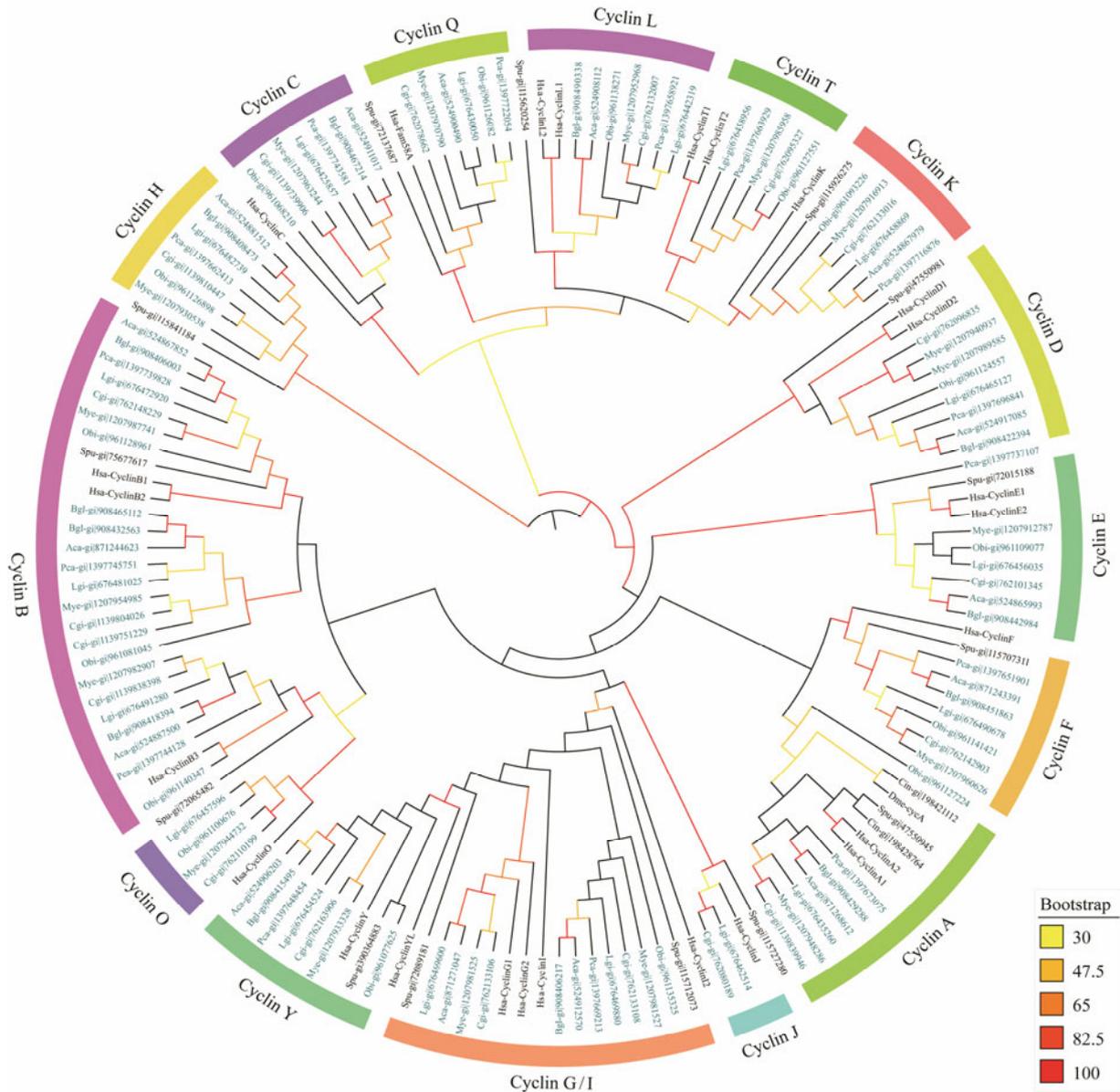


Fig.4 Phylogenetic relationships of Cyclin genes in Mollusca and several representative metazoans, including human, purple sea urchin, part genes of vase tunicate and fruit fly. The phylogenetic tree is constructed using the maximum likelihood (ML) method with 1000 bootstrap. The bootstrap values are represented by various colors. The Cyclin genes of Mollusca are marked with blue and other organisms' Cyclin genes are with black. All proteins are labeled with species names followed by accession numbers.

were analyzed (Fig.5 and Fig.6). The structures of the CDK and Cyclin genes were found to be moderately conserved among the various subfamilies, and the number and location of exons were similar in each subfamily, indicating similar function. The highest intron disruption was noted in the members of the *CDK11*, with intron disruption from 19 to 20, except for *L. gigantea* with seven intron disruption. The lowest intron disruption was noted in the members of the *CDK9*, with intron disruption varying from six to nine. For Cyclin family, Cyclin F subfamily showed the highest intron disruption ranging between 14 and 17, except for *L. gigantea* with eight intron disruption; Cyclin D subfamily showed the lowest intron disruption ranging from four to eight. In general, these results indicate that the CDKs and Cyclins in each group possess a similar number of exons, which further supports the evolutionary classification. Align-

ments of all CDK9 proteins from the seven molluscs was shown in Fig.7. Multiple sequence alignment of CDK9 proteins revealed a highly conserved CDK domain.

As a consensus or a conserved region in the protein or nucleotide sequences, motifs were analyzed in this study (Fig.5 and Fig.6). Totally, 10 conserved motifs of molluscan CDKs and Cyclins were identified using MEME. The length of these motifs varied from 15 to 29 aa in CDK family, and ranged between 15 and 41 aa in Cyclin family. For Cyclin genes, motifs 1, 3 and 4 were identified as N-terminal domains of Cyclin, while motifs 5 and 9 were identified as C-terminal domains of Cyclin. As CDK family, motifs 1, 2, 3, 6 and 7 were identified as protein kinase domains. Motif 2 (QLLRGJAYCHSNRILHRDLKPNJLI) and motif 5 (DQLDRIFKVLGTPTEETWPGV) were common in all the seven genomes, except CDK16 in *A. cali-*

*formica*. Taken together, the finding of similar gene structures and conserved motifs within the same subfamily further supports the accuracy of the phylogenetic tree. On the other hand, the structural differences between different subfamilies also indicate functional diversity of the CDK and Cyclin genes in Mollusca.

### 4 Discussion

Cell cycle is controlled by the regulatory units, cyclins, with the catalytic units, cyclin-dependent kinases. With the evolution of eukaryotes, the number of CDKs and Cyclins

increased (Gunbin *et al.*, 2011; Cao *et al.*, 2014). For example, *Saccharomyces cerevisiae* contains 6 CDKs and 15 Cyclins, whereas in human, the gene number is up to 20 (CDKs) and 29 (Cyclins) (Malumbres and Barbacid, 2005; Malumbres, 2014). However, investigation of cycle regulation in eukaryotes is limited, especially in Mollusca. Our work represents the first genome-wide identification of CDK and Cyclin family members in molluscs and provides insights into molecular evolution. In our study, we identified 95 genes in CDK family and 114 genes in Cyclin family from the seven molluscs. The number of CDK genes ranged from 13 to 15, and the Cyclin genes' number varied from 13 to 21,

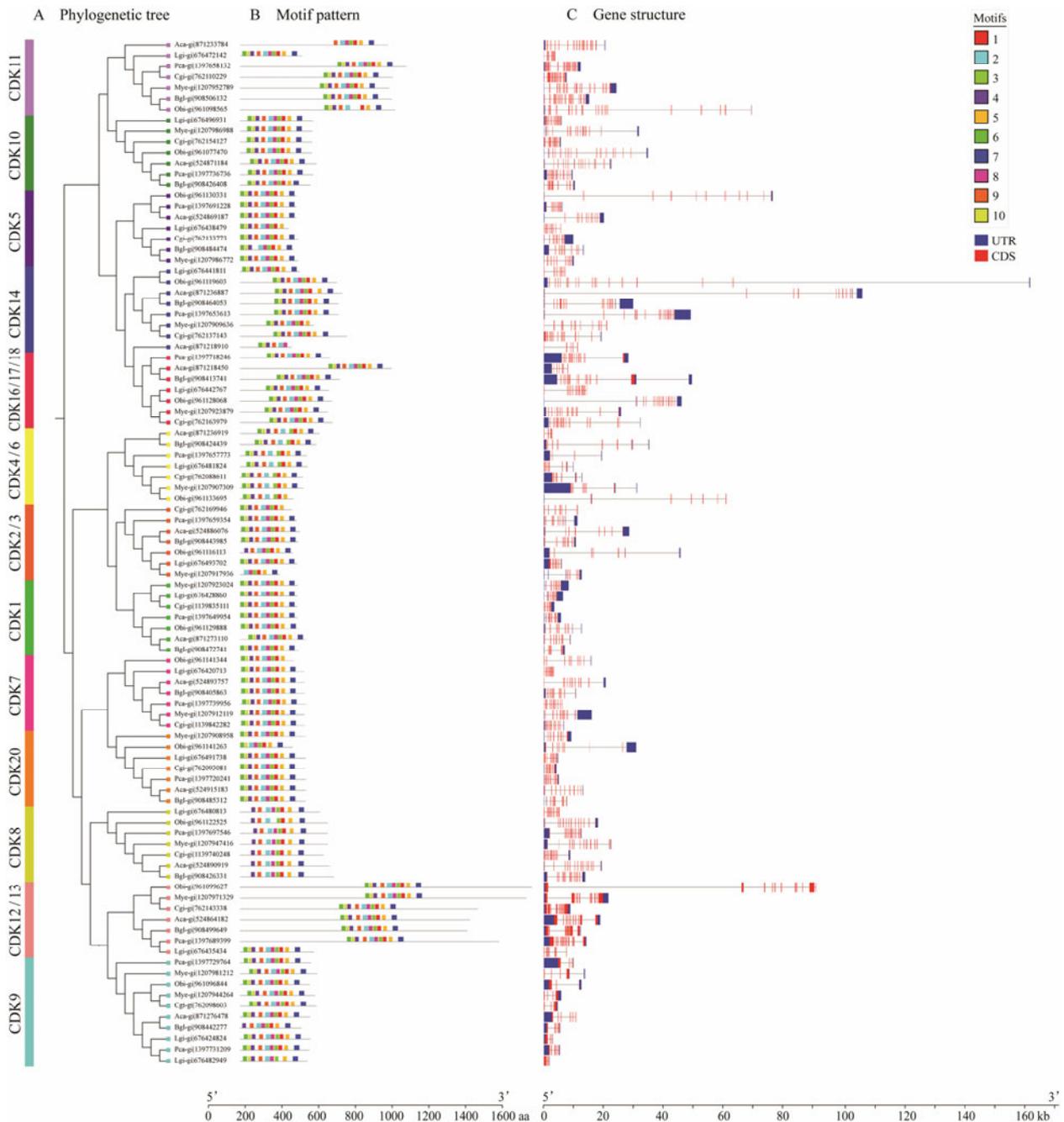


Fig.5 Phylogenetic relationships, gene structure and architecture of conserved protein motifs in CDK genes from seven mollusc species. (A) The phylogenetic tree is constructed based on the conserved structure of seven mollusc species CDK proteins using MEGA 7.0 software. (B) Exons and introns of CDK genes. Blue boxes indicate untranslated regions; red boxes indicate exons; and black lines indicate introns. (C) The motif composition of CDK proteins. The motifs are displayed in boxes with different colors. The length of the protein can be estimated using the scale at the bottom.

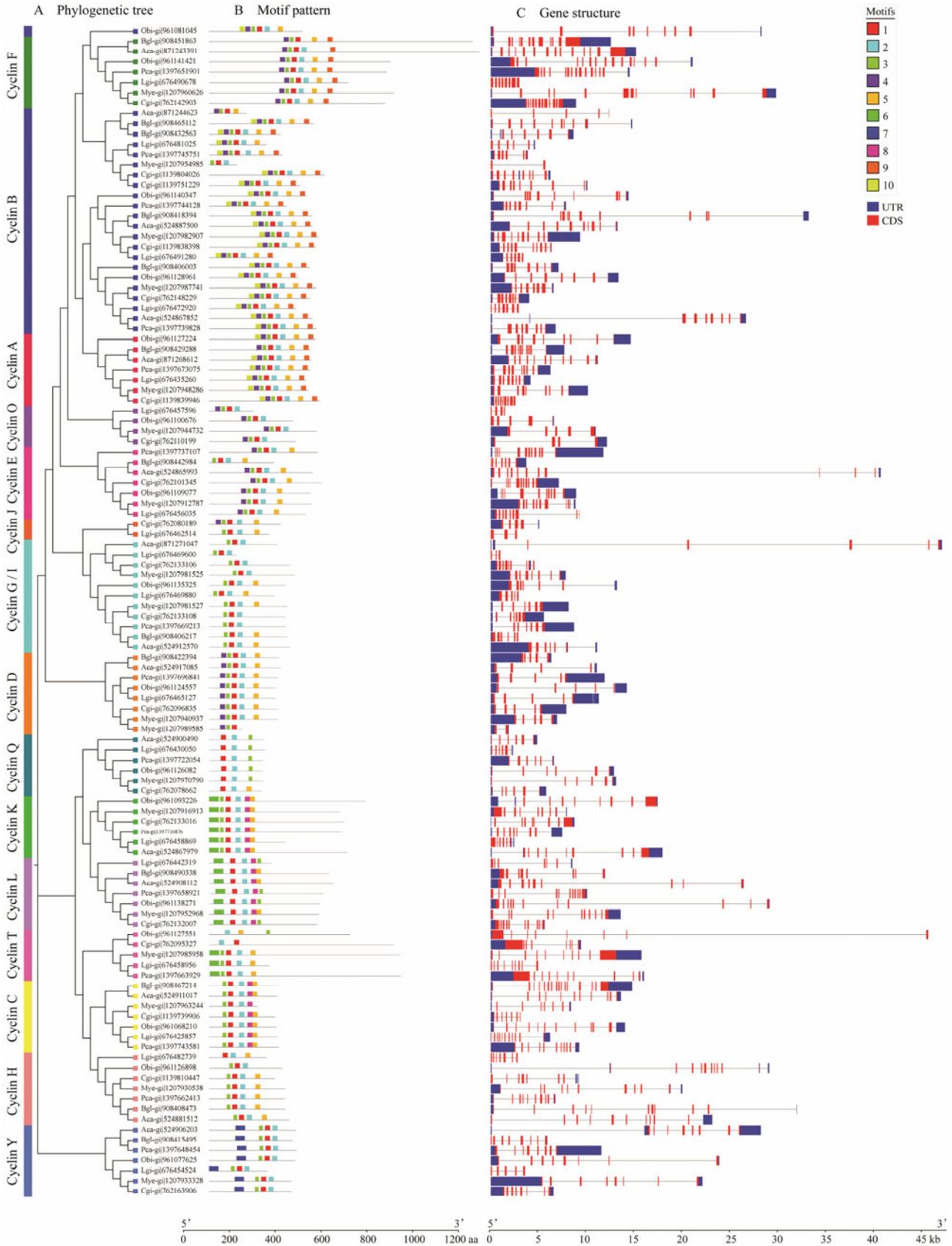


Fig.6 Phylogenetic relationships, gene structure and architecture of conserved protein motifs in Cyclin genes from seven mollusc species. (A) The phylogenetic tree is constructed based on the conserved structure of Cyclin proteins from seven mollusca species using MEGA 7.0 software. (B) Exon and intron of Cyclin genes. Blue boxes indicate untranslated regions; red boxes indicate exons; and black lines indicate introns. (C) The motif composition of Cyclin proteins. The motifs are displayed in boxes with different colors. The length of the protein can be estimated using the scale at the bottom.

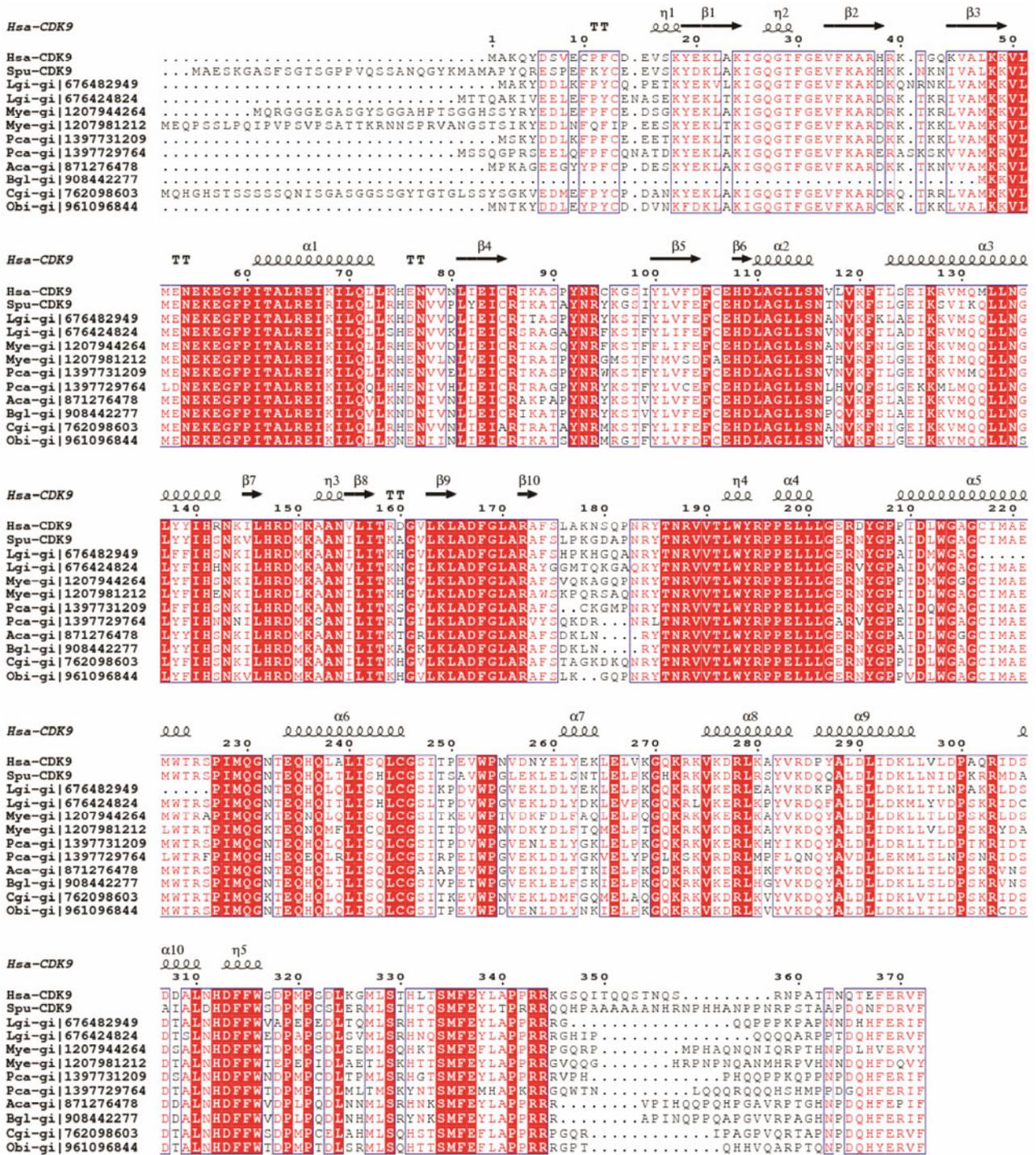


Fig.7 Sequence alignment of the *CDK9* genes from seven mollusc species, *H. sapiens* (*Has-CDK9*), and *S. purpuratus* (*Spu-CDK9*). Red shading denotes the extent of sequence conservation.

which is consistent with evolution trend of CDK and Cyclin in premetazoan lineages. Additionally, number and composition of the CDK genes were more stable than genes of the Cyclin families in the seven mollusc species, suggesting that CDK family was more conserved than Cyclin family and members of the Cyclin family might function in a species-specific manner.

#### 4.1 The Features of CDK Family in Molluscs

Based on our results, CDK genes in the seven molluscs are significantly conserved. The number of CDK family

members are steady, ranging from 13 to 15, and can be divided into eight subfamilies, which is consistent with previous reports (Liu and Kipreos, 2000; Guo and Stiller, 2004). The eight subfamilies can be classified into three cell-cycle-related subfamilies (CDK1/2, CDK4/6 and CDK5) and five transcriptional subfamilies (CDK7, CDK8, CDK9, CDK10/11 and CDK20) according to their functional characteristics (Liu and Kipreos, 2000; Cao *et al.*, 2014). Cell-cycle-related subfamilies of CDK binding with Cyclin A, Cyclin B, Cyclin D and Cyclin E promote each phase of the cell cycle. The different functions and structures of these CDK

subfamilies have been described in an excellent review (Wood and Endicott, 2018).

According to the phylogenetic tree, CDK5 subfamily is the most multiple subfamily with three clades including CDK5, CDK16/17/18, and CDK14/15. CDK5 subfamily in metazoan differs greatly (Mikolcevic *et al.*, 2012), but in the seven species of mollusc, it is more conservative based on gene constitution, phylogenetic and motif analyses. *CDK14* and *CDK15* were both detected in vertebrates (Mikolcevic *et al.*, 2012), but in the seven mollusc species, only CDK14 was identified. CDK16/17/18 is also named as PCTAIRE 1/2/3 (PCTK1/2/3), which contains a PCTAIRE sequence in the C-helix characterized by a conserved catalytic domain. In mammals, *CDK16*, *CDK17* and *CDK18* are expressed in neurons, suggesting that they play an essential role in the nervous system (Hirose *et al.*, 1997; Herskovits and Davies, 2006; Shimizu *et al.*, 2014). At present, PCTAIRE is studied as a new potential cancer treatment target (Dixon-Clarke *et al.*, 2017; Wang *et al.*, 2018). In this study, only one gene of CDK16/17/18 group, *CDK17*, was found in six mollusc species, while two genes, *CDK16* and *CDK18*, were found in *A. californica*. *A. californica* has large neurons and is suited for neurobiology study, *CDK16* and *CDK18* might play important roles. *CDK17* was highly conserved among the vertebrates, and most eumetazoa contain only *CDK17* (Mikolcevic *et al.*, 2012). So, it is suggested that *CDK17* might originate earlier than CDK16/18. Our results seem to provide additional evidence to support this scenario.

Except for *CDK9*, the other seven subfamilies of CDK genes have only one duplication in the seven mollusc species. The gene structure and motif analyses of the seven CDK subfamilies failed to find big differences in the exon number, multiple copies and homologous genes, which further highlighted that the seven types of CDK appeared to be widely conserved among different mollusc species. The CDK9 subfamily consists of two clades, CDK9 and CDK12/13 (Liu and Kipreos, 2000). It was referred that the CDK9 subfamily split into two clades before the divergence of metazoans and fungi (Cao *et al.*, 2014). In our results, all the seven molluscs contain CDK9 and CDK12/13. For CDK12/13, they are all CTD kinases with similar function (Kondrashov, 2012; Zhang *et al.*, 2016). In this study, they were not detected in the same species, which indicates there are substitutions between them. Notably, for *CDK9* genes, there are two duplications in *P. canaliculate*, *L. gigantea* and *M. yessoensis*. So far, it is the first time to identify the duplications of *CDK9* genes in animal genomes. According to the result of phylogenetic analysis, the two duplications do not cluster tightly, but clustered with other species' *CDK9*. For example, *L. gigantea* clustering respectively with *D. melanogaster* and *P. canaliculate* (Fig.3). Gene duplication has been the main course of expansion of the various gene families, and is associated with the adaptation of animals to the changing environments (Kondrashov, 2012). As a subunit of the positive transcription elongation factor b complex, CDK9 regulates transcription elongation in cooperation with Cyclin T. It also forms a complex with Cyclin K to regulate DNA damage signaling in replicating cells and

recover from a transient replication arrest (Yu *et al.*, 2010). CDK9 is a multifunctional kinase involved in a broad range of physiological processes, including myogenesis, cell growth, cellular viability and apoptosis (De Falco and Giordano, 1998; Franco *et al.*, 2018). Actually, no study on CDK9 genes has been reported in molluscs by now, and our understanding of the CDK9 functions is limited. The results in our study suggested that *P. canaliculate*, *L. gigantea* and *M. yessoensis* might be good materials to explore the CDK9 functions and evolution.

## 4.2 The Features of Cyclin Family in Molluscs

114 Cyclin family genes identified in this study can be divided into three major groups (Group I, Group II, and Group III), which is consistent with the previous study (Ma *et al.*, 2013). According to our results, Group I includes *Cyclins A, B, D, E, F, G, I, J* and *O*, Group II includes Cyclin *Y*, and Group III includes *Cyclins C, H, L, K, T*, and *Q* (*Fam58*). All types of Cyclin genes discovered in metazoan were also detected in the seven mollusc species. Cyclin A, B, D and E cooperated with CDK1 and CDK4/6 regulate cell cycle directly (Malumbres, 2014). It should be noted that Cyclin B in the seven mollusc species is the biggest subfamily with the average of three genes in each species, containing two *Cyclin B-like* genes and *Cyclin B3* gene. They are divided into three clades respectively. It is well known that Cyclin B in partner with CDK1 drive G2-M transition in mitosis. Cyclin B is multiple in different species, for example, there are three Cyclin B genes in human, mouse and zebrafish; two in purple sea urchin, cattle and dog; one in Rhesus monkey, zebra finch and Florida lancelet (Gunbin *et al.*, 2011; Cao *et al.*, 2014). Cyclin B in invertebrates evolved both rapidly and at uneven rates (Gunbin *et al.*, 2011). In most animals, two conserved B-type cyclins are detected: Cyclin B-like protein and Cyclin B3 (Nieduszynski *et al.*, 2002). Cyclin B3 is more important for regulation of meiosis than mitosis and is relatively conserved in vertebrate and invertebrate (Nguyen *et al.*, 2002; van der Voet *et al.*, 2009; Miles *et al.*, 2010). For Cyclin B-like protein, some organisms like human have two genes including *Cyclin B1* and *Cyclin B2*, which can compensate each other in function (Chotiner *et al.*, 2019). In our study, two Cyclin B-like genes were detected. Based on the phylogenetic analysis, the two *Cyclin B-like* genes clustered into two distinct clades, one of which clustered together with *Cyclin B1* and *Cyclin B2* of human, suggesting that this Cyclin B-like protein has a high homology with Cyclin B1 and Cyclin B2 in human, while the other one may carry out specific functions in molluscs. Mollusca is the second largest group with more than 100000 species, which is widely distributed in lakes, marshes, oceans, mountains and other environments, so as to adapt to different habitats, the morphological structure and lifestyle of various groups are very different. A different *Cyclin B-like* gene perhaps is a distinct developmental strategy in the adaptation to their changeable living environment (Nieduszynski *et al.*, 2002; Gunbin *et al.*, 2011).

Except for the cell-cycle-regulate genes (*Cyclin A*, Cy-

*clin B, Cyclin D and Cyclin E), Cyclin C, Cyclin F, Cyclin L and Cyclin Y* in the seven mollusc species are relatively conserved without genes duplication and with similar gene structure and motif. There is little difference of Cyclin family composition between classes, except for *L. gigantea*. Compared with other gastropods, the Cyclin family of *L. gigantea* is more similar to bivalves, especially *C. gigas*. This may be because both *C. gigas* and *L. gigantea* live in the intertidal zone and face a complex and varied environment. Additionally, a recent research suggested *Cyclin F* should not be part of the Cyclin family as it has no characteristic Lys Glu pair (Quandt *et al.*, 2020). In this study, we still analyzed *Cyclin F* as it consists of two typical cyclin domains, Cyclin N and Cyclin C. In addition, *Cyclin K* and *Cyclin Q* were not discovered in *B. glabrata*. However, according to recent studies, *Cyclin K* and *Cyclin Q* are conserved in metazoan species and are specific to animals (Ma *et al.*, 2013; Cao *et al.*, 2014). We found partial sequence of *Cyclin Q* in *B. glabrata* genome. The lack of whole *Cyclin Q* and *Cyclin K* genes was possibly because of genome incompleteness (N50: 7298 bp; L50: 32153 bp).

## Acknowledgements

This study was supported by the grants from the National Natural Science Foundation of China (No. 31672649), and the National Key R&D Program of China (No. 2018YFD0900200).

## References

- Abele, D., Brey, T., and Philipp, E., 2009. Bivalve models of aging and the determination of molluscan lifespans. *Experimental Gerontology*, **44** (5): 307-315, DOI: 10.1016/j.exger.2009.02.012.
- Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., *et al.*, 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, **524** (7564): 220-224, DOI: 10.1038/nature14668.
- Beach, D., Durkacz, B., and Nurse, P., 1982. Functionally homologous cell cycle control genes in budding and fission yeast. *Nature*, **300** (5894): 706-709, DOI: 10.1038/300706a0.
- Bodnar, A., 2009. Marine invertebrates as models for aging research. *Experimental Gerontology*, **44** (8): 477-484, DOI: 10.1016/j.exger.2009.05.001.
- Boxem, M., 2006. Cyclin-dependent kinases in *C. elegans*. *Cell Division*, **1**: 1-12, DOI: 10.1186/1747-1028-1-6.
- Cao, L., Chen, F., Yang, X., Xu, W., Xie, J., and Yu, L., 2014. Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. *BMC Evolutionary Biology*, **14** (1): 1-16, DOI: 10.1186/1471-2148-14-10.
- Chotiner, J. Y., Wolgemuth, D. J., and Wang, P. J., 2019. Functions of cyclins and CDKs in mammalian gametogenesis. *Biology of Reproduction*, **101** (3): 591-601, DOI: 10.1093/biolre/iox070.
- De Falco, G., and Giordano, A., 1998. CDK9 (PITALRE): A multifunctional cdc2-related kinase. *Journal of Cellular Physiology*, **177** (4): 501-506, DOI: 10.1002/(SICI)1097-4652(199812)177:4<501::AID-JCP1>3.0.CO;2-4.
- Dixon-Clarke, S. E., Shehata, S. N., Krojer, T., Sharpe, T. D., von Delft, F., Sakamoto, K., *et al.*, 2017. Structure and inhibitor specificity of the PCTAIRE-family kinase CDK16. *Biochemical Journal*, **474**: 699-713, DOI: 10.1042/BCJ20160941.
- Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D., and Hunt, T., 1983. Cyclin: A protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell*, **33** (2): 389-396, DOI: 10.1016/0092-8674(83)90420-8.
- Franco, L. C., Morales, F., Boffo, S., and Giordano, A., 2018. CDK9: A key player in cancer and other diseases. *Journal of Cellular Biochemistry*, **119** (2): 1273-1284, DOI: 10.1002/jcb.26293.
- Gunbin, K. V., Suslov, V. V., Turnaev, I. I., Afonnikov, D. A., and Kolchanov, N. A., 2011. Molecular evolution of cyclin proteins in animals and fungi. *BMC Evolutionary Biology*, **11** (1): 224, DOI: 10.1186/1471-2148-11-224.
- Guo, Z., and Stillier, J. W., 2004. Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. *BMC Genomics*, **5** (1): 69, DOI: 10.1186/1471-2164-5-69.
- Herskovits, A., and Davies, P., 2006. The regulation of tau phosphorylation by PCTAIRE 3: Implications for the pathogenesis of Alzheimer's disease. *Neurobiology of Disease*, **23** (2): 398-408, DOI: 10.1016/j.nbd.2006.04.004.
- Hirose, T., Tamaru, T., Okumura, N., Nagai, K., and Okada, M., 1997. PCTAIRE 2, a Cdc2-related serine/threonine kinase, is predominantly expressed in terminally differentiated neurons. *European Journal of Biochemistry*, **249** (2): 481-488, DOI: 10.1111/j.1432-1033.1997.t01-1-00481.x.
- Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G., 2015. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*, **31** (8): 1296-1297, DOI: 10.1093/bioinformatics/btu817.
- Hydbring, P., Malumbres, M., and Sicinski, P., 2016. Non-canonical functions of cell cycle cyclins and cyclin-dependent kinases. *Nature Reviews Molecular Cell Biology*, **17** (5): 280-292, DOI: 10.1038/nrm.2016.27.Epub2016Apr1.
- Johnson, D., and Walker, C., 1999. Cyclins and cell cycle checkpoints. *Annual Review of Pharmacology and Toxicology*, **39** (1): 295-312, DOI: 10.1146/annurev.pharmtox.39.1.295.
- Katoh, K., and Standley, D. M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, **30** (4): 772-780, DOI: 10.1093/molbev/mst010.
- Kondrashov, F. A., 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences*, **279** (1749): 5048-5057, DOI: 10.1098/rspb.2012.1108.
- La, H., Li, J., Ji, Z., Cheng, Y., Li, X., Jiang, S., *et al.*, 2006. Genome-wide analysis of cyclin family in rice (*Oryza sativa* L.). *Molecular Genetics and Genomics*, **275** (4): 374-386, DOI: 10.1007/s00438-005-0093-5.
- Liu, J., and Kipreos, E. T., 2000. Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): Differential conservation of CAKs in yeast and metazoa. *Molecular Biology and Evolution*, **17** (7): 1061-1074, DOI: 10.1093/oxfordjournals.molbev.a026387.
- Ma, Z., Wu, Y., Jin, J., Yan, J., Kuang, S., Zhou, M., *et al.*, 2013. Phylogenetic analysis reveals the evolution and diversification of cyclins in eukaryotes. *Molecular Phylogenetics and Evolution*, **66** (3): 1002-1010, DOI: 10.1016/j.ympev.2012.12.007.
- Malumbres, M., 2011. Physiological relevance of cell cycle kinases. *Physiological Reviews*, **91** (3): 973-1007, DOI: 10.1152/physrev.00025.2010.
- Malumbres, M., 2014. Cyclin-dependent kinases. *Genome Biology*, **15** (6): 1-10, DOI: 10.1186/gb4184.
- Malumbres, M., and Barbacid, M., 2005. Mammalian cyclin-dependent kinases. *Trends in Biochemical Sciences*, **30** (11): 630-641, DOI: 10.1016/j.tibs.2005.09.005.

- Mikolcevic, P., Rainer, J., and Geley, S., 2012. Orphan kinases turn eccentric: A new class of cyclin Y-activated, membrane-targeted CDKs. *Cell Cycle*, **11** (20): 3758-3768, DOI: 10.4161/cc.21592.
- Miles, D. C., van den Bergen, J. A., Sinclair, A. H., and Western, P. S., 2010. Regulation of the female mouse germ cell cycle during entry into meiosis. *Cell Cycle*, **9** (2): 408-418, DOI: 10.4161/cc.9.2.10691.
- Morgan, D. O., 1997. Cyclin-dependent kinases: Engines, clocks, and microprocessors. *Annual Review of Cell and Developmental Biology*, **13** (1): 261-291, DOI: 10.1146/annurev.cellbio.13.1.261.
- Murray, A. W., 2004. Recycling the cell cycle: Cyclins revisited. *Cell*, **116** (2): 221-234, DOI: 10.1016/s0092-8674(03)01080-8.
- Nguyen, T. B., Manova, K., Capodici, P., Lindon, C., Bottega, S., Wang, X. Y., *et al.*, 2002. Characterization and expression of mammalian cyclin B3, a prepachytene meiotic cyclin. *Journal of Biological Chemistry*, **277** (44): 41960-4169, DOI: 10.1074/jbc.M203951200.
- Nieduszynski, C. A., Murray, J., and Carrington, M., 2002. Whole-genome analysis of animal A- and B-type cyclins. *Genome Biology*, **3** (12): research0070.1-0070.8, DOI: 10.1186/gb-2002-3-12-research0070.
- Pines, J., 1995. Cyclins and cyclin-dependent kinases: A biochemical view. *Biochemical Journal*, **308**: 697-711, DOI: 10.1042/bj3080697.
- Quandt, E., Ribeiro, M. P., and Clotet, J., 2020. Atypical cyclins: The extended family portrait. *Cellular and Molecular Life Sciences*, **77** (2): 231-242, DOI: 10.1007/s00018-019-03262-7.
- Shimizu, K., Uematsu, A., Imai, Y., and Sawasaki, T., 2014. Pctaire1/Cdk16 promotes skeletal myogenesis by inducing myoblast migration and fusion. *FEBS Letters*, **588** (17): 3030-3037, DOI: 10.1016/j.febslet.2014.05.060.
- Simakov, O., Marletaz, F., Cho, S. J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., *et al.*, 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature*, **493** (7433): 526-531, DOI: 10.1038/nature11696.
- Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30** (9): 1312-1313, DOI: 10.1093/bioinformatics/btu033.
- Strom, A., Francis, R. C., Mantua, N. J., Miles, E. L., and Peterson, D. L., 2004. North Pacific climate recorded in growth rings of geoduck clams: A new tool for paleoenvironmental reconstruction. *Geophysical Research Letters*, **31** (6): L06206, DOI: 10.1029/2004GL019440.
- van der Voet, M., Lorson, M. A., Srinivasan, D. G., Bennett, K. L., and van den Heuvel, S., 2009. *C. elegans* mitotic cyclins have distinct as well as overlapping functions in chromosome segregation. *Cell Cycle*, **8** (24): 4091-4102, DOI: 10.4161/cc.8.24.10171.
- Wanamaker, A. D., Heinemeier, J., Scourse, J. D., Richardson, C. A., Butler, P. G., Eiriksson, J., *et al.*, 2008. Very long-lived mollusks confirm 17th century AD tephra-based radiocarbon reservoir ages for North Icelandic shelf waters. *Radiocarbon*, **50** (3): 399-412, DOI: 10.1017/S0033822200053510.
- Wang, G., Kong, H., Sun, Y., Zhang, X., Zhang, W., Altman, N., *et al.*, 2004. Genome-wide analysis of the cyclin family in Arabidopsis and comparative phylogenetic analysis of plant cyclin-like proteins. *Plant Physiology*, **135** (2): 1084-1099, DOI: 10.1104/pp.104.040436.
- Wang, H. T., Liu, H. L., Min, S. P., Shen, Y. B., Li, W., Chen, Y. Q., *et al.*, 2018. CDK16 overexpressed in non-small cell lung cancer and regulates cancer cell growth and apoptosis via a p27-dependent mechanism. *Biomedicine & Pharmacotherapy*, **103**: 399-405, DOI: 10.1016/j.biopha.2018.04.080.
- Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., *et al.*, 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution*, **1** (5): 1-12, DOI: 10.1038/s41559-017-0120.
- Wood, D. J., and Endicott, J. A., 2018. Structural insights into the functional diversity of the CDK-cyclin family. *Open Biology*, **8** (9): 180112, DOI: 10.1098/rsob.180112.
- Yu, D. S., Zhao, R. X., Hsu, E. L., Cayer, J., Ye, F., Guo, Y., *et al.*, 2010. Cyclin-dependent kinase 9-cyclin K functions in the replication stress response. *Embo Reports*, **11** (11): 876-882, DOI: 10.1038/embor.2010.153.
- Zhang, T., Kwiatkowski, N., Olson, C. M., Dixon-Clarke, S. E., Abraham, B. J., Greifenberg, A. K., *et al.*, 2016. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nature Chemical Biology*, **12** (10): 876-884, DOI: 10.1038/nchembio.2166.
- Ziuganov, V., San Miguel, E., Neves, R. J., Longa, A., Fernandez, C., Amaro, R., *et al.*, 2000. Life span variation of the freshwater pearl shell: A model species for testing longevity mechanisms in animals. *Ambio*, **29** (2): 102-105, DOI: 10.1639/0044-7447.

(Edited by Qiu Yantao)