Doctoral Dissertation

博士論文

Comparisons of larval and adult shell proteomes in molluscs

（軟体動物の成体殻と幼生殻における貝殻プロテオームの比較）

A Dissertation Submitted for Degree of Doctor of Philosophy

December 2018

平成 30 年 12 月博士（理学）申請

Department of Earth and Planetary Science, Graduate School of Science,

The University of Tokyo

東京大学大学院理学系研究科地球惑星科学専攻

ZHAO　RAN

趙　然

For my beloved mother, Li Shulan

献给我亲爱的老妈，李　淑蘭

Acknowledgments

First and foremost, I would like to show my deepest gratitude to my supervisor Prof. Kazuyoshi Endo, a pure, resourceful scholar and a gentle, responsible man, for his constant intellectual guidance and the kindness besides the science for an oversea student these years. Without the pushing and whiplashing from him, there would be no such a thesis. I also appreciate Associate Prof. T. Tsuihiji and Associate Prof. T. Sasaki of our group for their daily discussion and guidance. I would like to thank to the every member of Endo's laboratory for accepting me and the scientific discussion with them, especially, Dr. T. Takeuchi, who taught me experimental skills and shared experience with me, as well as Ishikawa-kun and Kobayashi-kun who taught me programming things for the bioinformatics works. I want to say thank you to Pro. N. Satoh of Okinawa Institute of Science and Technology, who provided supports for this project, as well as the member of his laboratory who did me a favor for generating the proteome data. Many thanks to Prof. S. Sasakura of Shimoda Marine Research Center, University of Tsukuba, who taught me electroporation technique. I also thank Prof. H. Wada and Dr. N. Hashimoto, of University of Tsukuba, who gave me a chance to learn microinjection skills.

Special appreciation to Dr. K. Nagai of Pearl Research Institute, Mikimoto, who taught us the skill of artificial fertilization of *Pintada fucata* and continuously supporting with samples. I appreciate Dr. S. Iwanaga of Nagasaki Prefectural Institute of Fisheries, Nagasaki, who provided us the larvae sample of *Crassostrea gigas.*

At last, I thank my mother for raising me up as well as her mental and financial support for my student career.

CONTENTS

Abstract

Although the functional importance of shell matrix proteins (SMPs) in shell formation, and their importance in understanding shell evolution have been generally recognized, potential importance of larval shells in those studies has been overlooked, and the SMPs of molluscan larval shells have been poorly characterized. Here, in chapter 2, I report comparisons of larval and adult shell matrix proteins within single species and between different species. By proteomic analysis combined with genomic and transcriptomic analyses, a total of 111 and 31 larval SMPs have been identified from the two bivalve species *Crassostrea gigas*, and *Pinctada fucata,* respectively, and they have been compared with their published adult counterparts. Comparisons between larval and adult SMPs in those species revealed that the larval SMPs are surprisingly different from the adult SMPs, exhibiting only four common SMPs shared between larval and adult SMPs in each species. Expression patterns of the genes encoding SMPs containing one or more functional domains, and forming a multigene family, such as Nacrein and Pif, clearly showed that, within those gene families, some members are highly and exclusively expressed at early developmental stages while others are adult stage specific, despite the fact that they contain the same domain(s), and expected to play similar or the same roles in shell formation. Pfam domain searches showed that VWA (von Willebrand factor type A), carbonic anhydrase, carbohydrate-binding module CBM_14, and EF-hand domains exist in both larval and adult SMPs in both species, indicating their indispensable roles in shell formation. The differences in the components between larval and adult SMPs suggest that the larval and adult shells originated and evolved independently. I could not determine whether the larval shells are more ancient in origin than the adult shells, or vice versa, because both are equally dissimilar between the two species.

In chapter 3, phylogenetic analyses have been performed on the three shell matrix

protein (SMP) families VWA and chitin-binding domain-containing protein (VWA-CB dcp), carbonic anhydrase (CA), and chitobiase, which exist in both the larval and adult shell proteomes in the bivalves *Crassostrea gigas* and *Pinctada fucata*. The results demonstrated different timing of deployment of larval and adult SMPs among those SMPs. Phylogenetic analysis performed on VWA, CB and Laminin G domains separately revealed that BMSPs, a VWA-CB dcp characterized by possession of multiple VWA domains, of bivalves and gastropods derived from a protein with a single VWA and two CB domains just like other VWA-CB dcps found in living bivalves and gastropods. In VWA-CB dcp and chitobiase, the gene duplications that gave rise to the larval and adult SMPs are inferred to have occurred before the divergence of the two bivalve species *C. gigas* and *P. fucata*. On the other hand, phylogenetic analyses of the SMPs containing a CA domain (CA-SMP), proteins which are remarkably expanded in molluscs and thought to play important roles in biomineralization, indicated that the duplications of the larval and adult CA-SMPs occurred after the divergence of the two bivalve species (possibly in Triassic), and in each lineage independently. Thus, the deployment of those CA-SMPs for larval or adult shell formation is considerably more recent than hitherto thought.

In chapter 4, in order to generate the first transgenic lineage of the pearl oyster *Pinctada fucata*, an endogenous β-tubulin promoter candidate has been cloned into *a green fluorescent protein* (*GFP*) expression plasmid. Electroporation and chemical mediated introduction of DNA were tested as a mean of delivering the DNA constructs into fertilized eggs. Successful deliveries of the foreign DNA via both methods have been confirmed by PCR amplification. However, fluorescent individuals could not be detected under microscope, which indicates some failures in the promoter sequence or in the *GFP*-coding sequence. On the other hand, the sequence of a SPARC gene of the limpet, *Nipponacmea fuscoviridus*, has been characterized. DNA and RNA constructs for CRISPR/Cas9 knock out of the SPARC gene have also been generated to perform *in vivo* functional analyses on this class of shell matrix proteins (SMP).

List of Tables

List of Figures

Chapter 1 General introduction

In living organisms, biomineralized tissues provide multiple functions: tissue support, storage of mineral ions, protecting the soft body from predators and from other environmental factors such as UV light (Lowenstam and Weiner 1989; Simkiss and Wilbur 1989). The overwhelming appearance of mineralized skeletons, including calcium phosphate, calcium carbonate and silica, opens the grand history of the metazoan taxa at the dawn of the Cambrian (Knoll and Carroll 1999; Morris 1998; Shubin and Marshall 2000). In the metazoan world, one of the phyla appeared abruptly in the Ediacaran. In contrast to many other groups or organisms, molluscs have an excellent fossil record extending some 550 million years with a substantial standing diversity of fossils so far discovered. As all but one of the eight main classes of molluscs had representatives in the Cambrian, including polyplacophorans, monoplacophorans, cephalopods, gastropods, bivalves (Lecointre and Le Guyader 2001; Runnegar 1996), and with the complex shell microstructures of early Cambrian molluscs (Feng and Sun 2003; Kouchinsky 2000), the history of molluscs is the key to understand the explosive Cambrian radiation of diverse animals accompanied with physiologically controlled mineralization.

Mollusc shells are generally composed of an outer organic layer-periostracum and inner calcified layers, over 95 weight % of which is calcium carbonate in the form of aragonite and/or calcite. The layers underneath the periostracum exhibit a diversity of microstructures, including “prismatic”, “nacreous”, “foliated”, “cross-lamellar”, “granular”, “composite-prismatic” and “homogeneous” (Boggild 1930; Carter and Clark 1985; Chateigner, et al. 2000). It is generally accepted that each shell structure develops in a specific part of shells. Also, combinations of several shell structures characteristically occur at certain taxonomic levels (Uozumi 1981), potentially reflecting phylogeny. In addition to the formation of those complex microstructures, various other aspects of shell biomineralization, including synthesis of transient

amorphous minerals, choice of calcium polymorphs, organization of crystals into complex shell textures are believed to be controlled by the organic molecules, collectively known as the shell matrix, which is a quantitatively minor constituent of the shell (0.1%~5% w/w according to different species and microstructures) (Mann 1988), and is secreted, at least for the formation of the adult shells, by an evolutionarily homologous organ known as the mantle (Lowenstam and Weiner 1989; Simkiss and Wilbur 1989).

Biomineralization in juvenile and adult molluscs

In molluscs, an indirect mode of development characterized by a transition from a ciliated trochophore to a veliger larva, followed by metamorphosis from the veliger to a juvenile, is shared by the majority of molluscan classes, in particular bivalves and gastropods (Martoja 1995). When the swimming veliger larva settles down onto a benthic existence, the disappearance of the velum is accompanied by the development of the foot, digestive gland and the reproductive organs (Bonar 1976). The resulting juvenile mollusc calcifies rapidly, and the growth of the shell approximately follows the von Bertalanffy law (Seed 1980). Classically, the physiology of molluscan shell calcification can be described as a succession of compartments (WILBUR and Saleuddin 1983), the central of which is the mantle, a thin organ, which coats the inner surface of the shell. The mantle is a polarized tissue comprised of three components: (1) the inner epithelium, in contact with the ambient medium, for example, seawater, (2) the mantle interior, which is comprised of pallial muscles, nerves, connective tissues, and (3) the outer epithelium, which mineralizes the shell (Fig. 1.1). Secreted by the mantle outer epithelium, the periostracum, composed of organic matters including polysaccharides and proteins, functions in providing a support or foundation for the mineralization and in setting the boundary and sealing the space where the mineralization occurs (Marin, et al. 2007).

Transported in the connective tissues via the hemolymph, precursor ions, i. e.,

calcium and bicarbonate ions, that are taken up from the ambient waters or from the metabolism (food and fluids) can finally reach the second compartment, the extrapallial space (Saleuddin and PETIT 1983) (Fig. 1.1). The extrapallial space is the place where "mysterious" transition from the liquid precursors to the solid crystals occurs. Recently, the concept of self-assemblage is put forward to depict how the amorphous ions and shell matrix interact in a controlled manner to produce organo-mineral material with fine structures (Marin, et al. 2007).

The matrix is a mixture of proteins, glycoproteins, acidic polysaccharides, chitin, and presumably lipids secreted by the mantle. It is consumed when interacting with the mineral ions and controlling the shape of the produced crystals (Marin *et al*., 2007). Previous studies have demonstrated that, at both transcriptional and protein levels, the secretory regime show differences according to the position of the cells involved in the process on the outer epithelium, in particular for the species which exhibit a bi-textured shell (e.g., an outer prismatic layer and an inner nacreous layer) (Fig. 1.1). The formation of the prisms is controlled by the matrix secreted from the outer epithelial cells in a more distal position than the epithelial cells involved in the nacre formation (Jolly, et al. 2004; Sudo, et al. 1997; Takeuchi and Endo 2006). Three types of SMPs, i.e., specific to the prism, specific to the nacre, and contained in both, have been discriminated in the pearl oyster *Pinctada fucata*, and their transcripts indicated congruent expression patterns in the outer epithelium (outer, middle, and inner pallial regions) of the mantle (Zhao, et al. 2018).

Molluscan shell formation is often cited as resulting in large measure from extracellular events mediated by the organic matrix secreted from the mantle epithelium (Lowenstam and Weiner 1989; Simkiss and Wilbur 1989). But, an hypothesis alternative to this matrix-mediated hypothesis was proposed on the oyster *Crassostrea virginica* (Mount, et al. 2004), positing that the crystal nucleation takes place in the crystallogenic cell (hemocyte), which supply nascent crystals to the mineralization front. So far, it is still under debate if this is a phenomenon particular

to this species and closely related oysters or a more general mechanism by which the molluscan shells are produced.

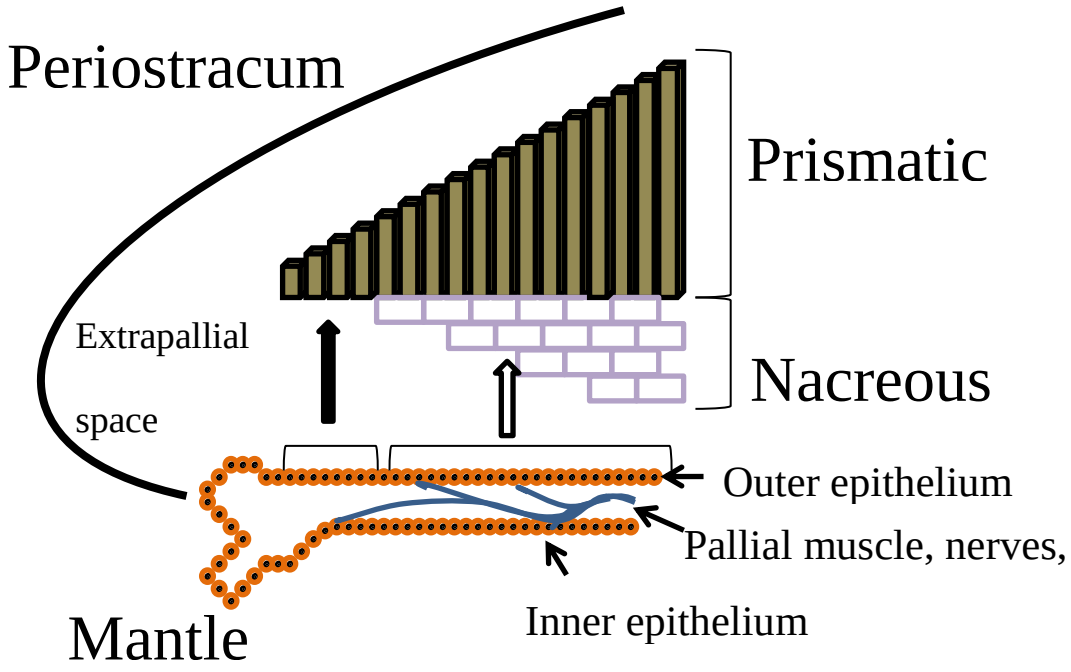


Fig. 1.1 Schematic of physiology of the shell formation in nacro-prismatic bivalves, redrawn from (Saleuddin and PETIT 1983).

Molluscan larval shell formation

In molluscs, the first step in shell biomineralization occurs during the trochophore larval stage. Most of the shell-secreting cells are descendants of the 2d blastomere, which develop into an ectodermal region of the embryo known as the "shell field" (Cather 1967; Cather, et al. 1976). The invagination of the cells in the center of the shell field produces the transitory structure called shell gland (KNIPRATH 1981). The invagination allows the cells at the periphery of the shell gland to produce the early organic membrane, which will be the first template for the deposition of calcium carbonate minerals. This organic lamella is the future periostracum. Between the periostracum and the cells of the shell field, the early mineralization of the embryo occurs (Kniprath 1980, 1977). In bivalves, secreted during the trochophore stage, the first shell is called prodissoconch I, which is characterized by granular microstructure (Lydie, et al. 2001). During the transition from the trochophore stage to the veliger stage, the prodissoconch II is secreted, which is characterized by the appearance of the growth lines (Lydie, et al. 2001). Then, following the metamorphosis of the veliger larva, the dissoconch shell or the teleoconch shell is formed. In gastropods, the terminology is different: the first shell formed during trochophore stage is protoconch I, the second shell formed during the veliger stage is protoconch II, and the postmetamorphic shell is called teleoconch (Jablonski 1980).

In spite of the active roles inferred for the shell gland or shell field invagination (Eyster and Morse 1984; Kniprath 1980, 1981, 1977) and the processes of the early shell formation have been depicted (Lydie, et al. 2001), in comparison to the other phyla, especially echinoderms and arthropods, the connection between the physiology of the larval shell development and the underlying genetic mechanisms, is poorly understood. Direct or indirect roles of homeobox gene *engrailed*, in marking the skeletal boundaries during molluscan embryo development were claimed by (Jacobs, et al. 2000; Moshel, et al. 1998; Nederbragt, et al. 2002; Wanninger and Haszprunar 2001). The role of *Hox* genes in patterning shell were suggested by the expression of

Hox1 in a ring of cells corresponding to the outer mantle edge in the trochophore larva and the expression of *Hox4* at the mantle in the later stage after larval shell is fully formed (Hinman, et al. 2003). Furthermore, an important but has long been underestimated aspect of larval shell formation corresponds to the intense enzymatic activities during the whole embryonic process. In the freshwater snail, *Lymanea*, alkaline phosphatase is highly expressed during the evagination, while the expression of DOPA-oxidase (tyrosinase) and peroxidase reached the highest level at the boundaries of shell gland and the surrounding cells (Timmermans 1968). In the larval stage of the mussel, *Mytilus*, high levels of expression of carbonic anhydrase (CA) were detected before the formation of the shell field in the gastrula embryo, the formation of shell gland and periostracum in the trochophore stage, and the deposition of mineral in the prodissoconch I and prodissoconch II stages (Medaković 2000). The importance of chitin synthase, a transmembrane glycosyltransferase that catalyzes the synthesis of chitin, was inferred by the presentence of the chitin synthase transcript in the cells contact with larval shell in early and late veliger stages of the mussel, *Mytilus galloprovincialis* (Weiss, et al. 2006).

Compared with those of their adult counterparts, molluscan larval shells are simpler in structure and mineralogies: the shell microstructures of larval shells are generally comprised of three mineralized layers, an outer prismatic layer (OP) below the periostracum, an inner prismatic layer next to the mantle epithelium (IP), and a granular homogeneous layer (G) between the two prismatic layers (Waller 1981; Weiss, et al. 2002), and are much simpler than those the adult shells, and the mineral component of larval shells is almost exclusively aragonite (Table 1.1) regardless of the mineralogies in the postlarval and adult forms (Carriker 1979; Cather 1967; Eyster 1982; Ivester 1972; Iwata and Akamatsu 1975; LaBarbera 1974; Stenzel 1964). By using focused ion beam (FIB) sample preparation technique and transmission electron microscopy (TEM), detailed processes of acquisition and the microstructure of the larval shell were depicted (Kudo et al. 2019; Yokoo et al. 2011). In *Crassostrea nippona* at 14 h after fertilization (in the trochophore stage), an outer layer of

aragonite started to form. At 72 h after fertilization (in the veliger stage), an additional granular layer was produced between the preexisting layer and the embryo to form a distinctive two-layered structure (Kudo, et al. 2010). In *Pinctada fucata*, at 18 h after fertilization, the larva was covered by the first shell made of aragonite. At 48 h after fertilization, a new homogenous aragonite layer with globular contrast formed under the initial layer. At this stage, inner layer consisting of calcite was found in a few specimens. One to three weeks after fertilization, a new aragonite layer with a prismatic contrast formed under the homogeneous layer (Yokoo, et al. 2011). However, no evidence of ACC was found in both studies. Despite it was suggested that amorphous calcium carbonate (ACC) is the precursor phase in the larval shell which transforms to aragonite during shell growth in *Mercenaria mercenaria* and *Crassostrea gigas* (Weiss, et al. 2002), in the two studies performed using FIB and TEM, no evidence of ACC was found. Taken together, these pieces of evidence agreed with the opinion pointed out by Taylor (1973), that larval shells are evolutionarily conserved.

It was assumed that the ancestor of extant invertebrates, with the exceptions of insects and cephalopod molluscs, use a similar embryogenesis mechanism for their development, based on which undifferentiated cells that are set aside from the embryogenesis *per se* and the subsequent utilization of genetic regulatory mechanisms that allowed the evolution of adult plans (Davidson 1990; Davidson, et al. 1995; Peterson, et al. 1997). Thus, larvae of living animals are supposed to show more ancestral common features than adults in this theory. Another hypothesis given by (Peterson, et al. 1997) suggested the larva forms are homologous and the genome of ancestors must have already included regulatory programs for embryo/larva development that are still utilized by their indirect developing modern descendants. If these hypotheses are correct, gene repertoires involved in the larval shell formation molluscs may show more similarities to that of their common ancestor than their adult counterparts.

Table 1.1 Crystal form of pre-metamorphic molluscan shells.

Species	Stage	Crystal phase	Reference
Aeolidia palillosa	Larva	Aragonite	Eyster, 1982
llyanassa obsoleta	Larva	Entirely aragonite	Cather, 1967
llyanassa obsoleta	Trochophore	Arag., trace calcite	Ivester, 1972
llyanassa obsoleta	Veliger	Aragonite	Ivester, 1972
Crassostrea virginica	PI, PII	Aragonite	Carriker, 1979
Crassostrea virginica	Veliger	Aragonite	Stenzel, 1964
Crassostrea gigas	Veliger	ACC, aragonite	Weiss, et al. 2002
Mervenaria merecenaria	Veliger	ACC, aragonite	Weiss, et al. 2002
Crassostrea nippona	PI, PII	Aragonite	Kudo, et al. 2010
Pinctada fucata	PI, PII	Aragonite, trace calcite	Yokoo, et al. 2011
Tndacna squamosa	PI, PII, juv (mix)	Aragonite	LaBarbera, 1974
Mytilaceans, 12 species	Larva	Aragonite	Carter, 1980
Patinopecten yessoensis	Larva	Aragonite	Iwata and Akamatsu, 1975

Shell matrix protein (SMP) and functional analyses of SMP

Since 1990s, accompanied by the discovery of various shell matrix proteins (SMPs), which cover a broad spectrum of p*I*, from the extremely acidic to the very basic, new concepts, including modularity, were utilized to interpret the observations as more complex realities. The different modules comprise the primary structure of these SMPs indicate their different roles in the calcification processes. For example, the existence of highly acidic modules in shell proteins has been predicted since the pioneering work of Weiner and Hood (1975). Today, this group includes MSI31 (Sudo, et al. 1997), MSP-1 (Sarashina and Endo 2001), Aspein (Tsukamoto, et al. 2004), Prismalin-14 (Suzuki, et al. 2004), MSP-2/SP-S (Hasegawa and Uchiyama 2005), Asp-rich protein families (Gotliv, et al. 2005) in Bivalvia, as well as LUSP-9, LUSP-10 and LUSP-23 in gastropoda (Marie, et al. 2013) (Table 1.2). One striking feature of these calcite-specific proteins is that they are singularly enriched in Asp residues, though the reason is unclear. Because of the enrichment of Asp residues which will be negatively charged under physiological conditions, these proteins have been considered to show low-affinity and high-capacity Ca^{2+} binding ability, which is compatible with a reversible binding of Ca^{2+} (Maurer, et al. 1996). The roles that Prismalin-14 and Aspein play in the regulation of the calcification were studied via *in vitro* carbonate precipitation experiments by Suzuki *et al*., (2004) and Takeuchi *et al*. (2008), respectively. Later, *RNAi* gene knockdown experiments confirmed the key roles of another acidic matrix protein *Pif* in the formation of aragonitic nacreous structure in *Pinctada fucata* (Suzuki, et al. 2009). Today, as the genomic and transcriptomic analyses using the next-generation DNA sequencing techniques are widely applied to shell proteomic projects, the number of shell proteins added to our knowledge is increasing fast, and the traditional theory that the shell matrix is where mineralization occurs is being challenged by a modified cellular model based on the identification of shell proteins in multiple organs in oyster, though the mantle is still the most important organ for the formation of adult shells (Wang, et al. 2013b).

Table 1.2 Unusually acidic molluscan shell proteins (with a theoretical isoelectric point below 4.5).

	Protein Name	Species	MW(kDa)	pI (Asp%, Glu%)	Swiss‐Prot/TrEMBL /Genome Accession Number	References
BIVALVIA	Aspein	*Pinctada fucata*	39.3/41.2	1.67 (57.6, 15.5)	Q76K52	Tsukamoto *et al*., 2004
	MSI31	*Pinctada fucata*	32.85/31	3.8 (5.1, 23.6)	O02401	Sudo *et al*., 1997
	Prismalin‐14	*Pinctada fucata*	11.9/13.5	4.24 (9.1, 24)	Q6F4C6	Suzuki *et al*., 2004
	MSP‐1	*Patinopecten yessoensis*	74.6/76.4	3.34 (19.6, 24.5)	Q95YF6	Sarashina and Endo, 1998, 2001
	MSP‐2/SP‐S	*Patinopecten yessoensis*	27.9/29.8	3.48 (20, 21.4)	Q6BC34	Hasegawa and Uchiyama, 2005
	Asp‐rich protein 1	*Atrina rigida*	6.6/8.5	3.34 (32.5, 5)	Q5Y821	Gotliv *et al*., 2005
	Asp‐rich protein 2	*Atrina rigida*	15/17	2.89 (41, 7.5)	Q5Y822	Gotliv *et al*., 2005
	Asp‐rich protein 3	*Atrina rigida*	16.5/18.4	2.75 (42, 6.9)	Q5Y823	Gotliv *et al*., 2005
	Asp‐rich protein 4	*Atrina rigida*	18/19.9	2.73 (40.3, 6.8)	Q5Y824	Gotliv *et al*., 2005
	Asp‐rich protein 5	*Atrina rigida*	17.4/19.3	2.76 (38.7, 7)	Q5Y825	Gotliv *et al*., 2005
	Asp‐rich protein 6	*Atrina rigida*	18.2/20	2.72 (40.9, 6.7)	Q5Y826	Gotliv *et al*., 2005
	Asp‐rich protein 7	*Atrina rigida*	25.8/23.9	2.54 (50.8, 5.7)	Q5Y827	Gotliv *et al*., 2005
	Asp‐rich protein 8	*Atrina rigida*	25.3/27.2	2.53 (48.6, 5.4)	Q5Y828	Gotliv *et al*., 2005
	Asp‐rich protein 9	*Atrina rigida*	18.2/20	2.72 (40.9, 6.7)	Q5Y829	Gotliv *et al*., 2005
	Asp‐rich protein 10	*Atrina rigida*	20/21.8	2.68 (41.6, 5.7)	Q5Y830	Gotliv *et al*., 2005
GASTROPODA	LUSP-9	*Lottia gigantea*	29	3.7 (11.6, 10,9)	B3A0Q7	Marie *et al*., 2013
	LUSP-10	*Lottia gigantea*	68	3.8 (16.9, 3.7)	Lgi_163637	Marie *et al*., 2013
	LUSP-23	*Lottia gigantea*	22	3.6 (24.3, 6.5)	B3A0S2	Marie *et al*., 2013

Problems with the current research on SMPs and prospects for transgenic molluscs

In the last decade, despite literally thousands of SMPs have been identified from many molluscan species and the functions of some of those SMPs in shell formation processes have been investigated, the functions have been mainly inferred from their chemical characteristics, and it is not very clear how well *in vitro* crystallization experiments mimicked the real physiological condition or how reliable the result of *RNAi* experiments are when the extent of gene knock down is incomplete and the temporal effects on phenotypes are lost in several days. This situation can be mainly attributed to the lack of a transgenic molluscan lineage to perform in vivo functional analyses via gene knock-out experiments, like those carried out on vertebrates and arthropods. Transmission of foreign DNA into embryos has been reported in a limited number of molluscs including abalones (Powers, et al. 1995; Tsai, et al. 1997), oysters (Buchanan, et al. 2001; Cadoret, et al. 1997a; Cadoret, et al. 1997b), limpets (Hashimoto, et al. 2012), and the slipper shell *Crepidula fornicata* (Perry and Henry 2015). However, none of these studies aimed at analyzing the functions of SMPs, and the methods for introduction of DNA reported for those animals are difficult to be repeated or to be applied to other molluscs, mainly because of the tiny size of molluscan eggs and the intrinsic differences among species, such as the structure and nature of the egg chorion.

Another problem of the current studies on SMPs, especially the shell proteome projects, which hinders our understanding of the shell formation processes, is that they mainly focus on the SMPs of adult animals, which can probably be explained by the tiny size of mollusc larvae, making the collection of the larval shells and extraction of SMPs from them difficult.

Novel gene editing technology CRISPR/Cas9

Despite the difficulties and the limited number of previous studies, the innovation of CRISPR/Cas9, a novel gene editing technology shed light on performing functional analyses of SMPs via transgenic molluscan lineages.

The CRISPR/Cas9 system is derived from *Streptococcus pyogenes* SF370 and is comprised of a riboprotein complex with a "guide" RNA (gRNA) and the endonuclease Cas9 (Fig. 1.2). The target DNA sequence can be any 20 nt sequence associated with a protospacer adjacent motif (PAM) whose recognition is specific to individual CRISPR systems (Garneau, et al. 2010; Gasiunas, et al. 2012; Jinek, et al. 2012). The binding between gRNA and the targeted DNA directs *Cas9* to mediate the cleavage of the target DNA upstream of PAM to create a double-strand break (DSB) within the protospacer. Then, The Non-Homologus End Joining (NHEJ) mechanism after cleavage will generate loss-of-function mutations within the targeted sequence (Fig. 1.2).

Until now, CRISPR/Cas9 has been successfully introduced into various metazoan animals, including mouse (Wang, et al. 2013a), human (Cong, et al. 2013), zebrafish (Hwang, et al. 2013), *Xenopus* (Blitz, et al. 2013), nematodes (Friedland, et al. 2013), *Drosophila* (Gratz, et al. 2013) and the mollusc *Crepidula fornicata* (Perry and Henry 2015). This technology has liberated researchers from the hard and tedious work of traditional gene-targeting methods. Compared with the previous designer nucleases, ZEN and TALEN, which have also been applied to various systems for generating double-strand breaks in the target region of the coding sequences (Beumer, et al. 2013; Sakuma, et al. 2013; Watanabe, et al. 2014), targeting of different target sites via CRISPR/Cas9 can be realized by simply modifying the gRNA sequences instead of revising the cutting nucleases, a property which is especially apposite when a number of target sites are to be tested.

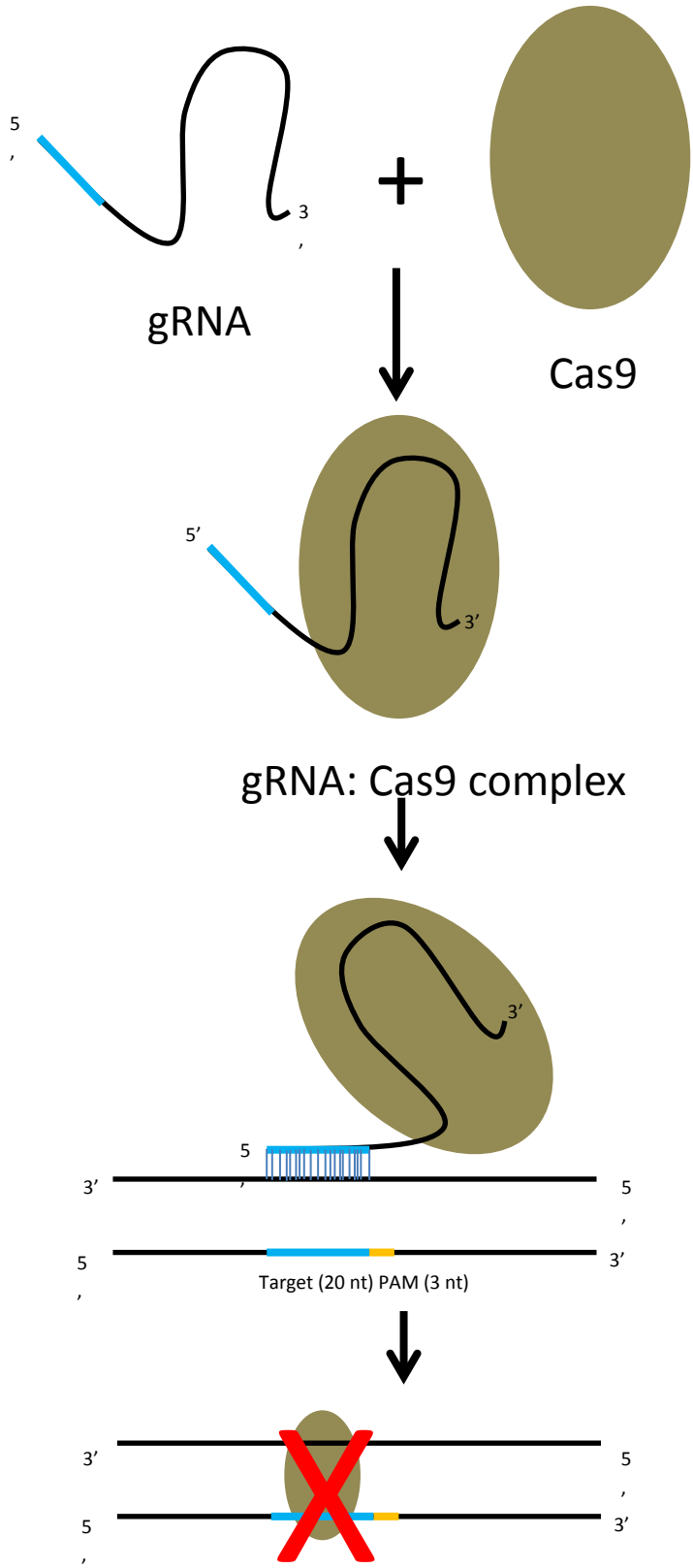


Fig. 1.2 Schematic of the type II CRISPR-mediated DNA double-strand break.

The aims of the present study

In the past, transcriptomics combined with shell proteomics revealed distinct repertoires of the SMPs of the adult shells among molluscan species (Marie, et al. 2013; Marie, et al. 2012; Marie, et al. 2011a; Zhang, et al. 2012). However, as for the larval shells, which could be evolutionarily more conserved than the adult shells, our knowledge is limited, and the components of the SMPs and whether the repertoires of larval SMPs are more similar to each other than those of their adult counterparts when compared between different species await detailed investigations. Thus, in chapter 2 of this thesis, as a first trial, the SMPs of the larval shells of the two pteriomorph bivalve species, *Crassostrea gigas* and *Pinctada fucata* have been characterized and compared, and further compared with the sequences of the adult counterparts already published for those species (Liu, et al. 2015; Takeuchi, et al. 2016; Zhang, et al. 2012).

Two hypotheses have been given in the past to explain the calcification in metazoans. On one hand, calcification was the result of recruitment and orchestration of ancestral Precambrian functions which is not related to the mineralization (Marin, et al. 2003; Marin, et al. 2007). On the other hand, calcification was acquired independently by different metazoan lineages, and the similarities are the result of adaptive convergence (Marin, et al. 2007). In order to answer those puzzling questions about the origin of shell proteins which are deeply involved in the calcification process of molluscs, for example, where they came from, and how they were recruited by the shell, in chapter 3, phylogenetic analyses are performed on the three SMP gene families which are contained in both the larval and adult SMP repertoires of the two bivalve species *C. gigas* and *P. fucata* studied in chapter 2. The evolutionary relationships among each of those SMPs, possible evolutionary scenarios of domain structures, and the origin of their recruitment by the shells, particularly the timing of deployment for larval and adult shell formation relative to the divergence of the two bivalve species, are discussed.

In chapter 4, in order to generate a model organism in molluscs, which provides a platform for *in vivo* functional analyses of SMPs, including those from the nacro-prismatic molluscs in the future, introduction of foreign DNA by electroporation and chemical transfection methods have been tested on the fertilized eggs of *P. fucata*, and the DNA constructs based on the novel gene editing technology CRISPR/Cas9 have been generated based on the gene sequences of the limpet, *Nipponacmea fuscoviridus*. The results are summarized, and possible sources of problems are discussed.

Finally in chapter 5, based on the outcomes collectively obtained from the above studies in this thesis, general discussion and possible directions of further studies are given.

Chapter 2 Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation

Key words: biomineralization, mollusca, metamorphosis, proteome, shell matrix protein (SMP)

2.1 Introduction

In living organisms, biomineralized tissues provide multiple functions: tissue support, storage of mineral ions, protecting the soft body from predators and from other environmental factors such as UV light (Lowenstam and Weiner 1989; Simkiss and Wilbur 1989). Like in other metazoan lineages, acquisition of diverse mineralized exoskeletons is one of the reasons to explain the rapid establishment of shell-bearing molluscs at the dawn of the Cambrian times (Kawasaki, et al. 2004; Killian and Wilt 2008).

Despite being a minor fraction in the shells, the organic matrices, composed mainly of proteins, glycoproteins, chitin, and acidic polysaccharides, have pivotal roles in numerous aspects of shell formation, such as calcium carbonate nucleation, crystal growth, and choice of calcium carbonate polymorphs (Addadi, et al. 2006; Marin, et al. 2007). In recent years, with the help of high-throughput sequencing techniques, comparisons of molluscan shell matrix proteins through proteomic and transcriptomic studies revealed species-specific repertoires of shell matrix proteins among molluscs, including shell polymorph (i. e. calcite and aragonite) specific molecular toolkits within the same species (Liu, et al. 2015; Mann 1988; Marie, et al. 2013; Marie, et al. 2012; Marie, et al. 2011b; Zhang, et al. 2012).

The adult molluscan shells are mainly comprised of calcite or aragonite, or both, which are assembled in highly variable microstructures (Carter 1990; Kobayashi 1969; Taylor 1973). Molluscan larval shells, in contrast, are always composed only of

aragonite, showing similar microstructures to each other (Carriker 1979; Eyster 1982; Eyster 1986; LaBarbera 1974; Waller 1981; Weiss, et al. 2002), a fact which implies that the larval shells are evolutionarily highly conserved (Taylor 1973). With this simplicity in mineralogy and microstructures, and with the shorter period of time required to see the effects of potential *in vivo* manipulation experiments on the shell formation genes, the larval shells would arguably be a suitable model for understanding shell formation processes to answer such a simple but still unanswered question as how many proteins are required to build a shell. Studies of larval shell proteins would also be important in inferring the evolutionary antiquity of the larva, to test, for instance, the hypothesis that "set-aside cells" in the molluscan larvae provide the evolutionary and developmental ground on which adult structures are built (Peterson, et al. 1997). However, due perhaps to some technical difficulties, for example, in isolation of a certain amount of the tiny larval shells, a proteomic analysis on larval molluscam shells has not been performed.

The first shell of bivalves, prodissoconch I, is secreted during the trochophore stage, usually about 20 hours after fertilization (LaBarbera 1974). Prodissoconch I is then enlarged to form prodissoconch II during the transition stage to become the veliger (D-shape) larva. After the settlement of the veliger larva and metamorphosis to become juvenile, the shell is called the dissoconch (Lydie, et al. 2001). In this study, as the first trial, we applied acetic acid to extract the shell matrices from 24 h D-shape larval shells of the two pteriomorph bivalve species, *Crassostrea gigas* (*C. gigas*) and *Pinctada fucata* (*P. fucata*), whose adult shell protein repertoires have been characterized and proven to be rather different to each other (Liu, et al. 2015; Zhang, et al. 2012). Extracted shell matrix proteins were digested by trypsin, and then applied to LC-/MS/MS analysis. Shell-derived amino acid datasets were interrogated against the genome databases of *C. gigas* (Zhang, et al. 2012) and *P. fucata* (Takeuchi, et al. 2016).

2.2 Materials and methods

2.2.1 Protein extraction

24 h D-shape larvae of *Crassostrea gigas* and *Pinctada fucata* are gifts from Nagasaki Prefectural Institute of Fisheries and Mikomoto Pearl Research Institute, respectively (Fig. 2.1). The following shell cleaning and protein extraction method are same for the two species. Three mL of larvae were collected in 50 mL tube and incubated with 1 M sodium NaOH for overnight. The shells were washed with MilliQ water five times and were observed with stereomicroscope. The cleaning process was repeated until the soft tissue and the contaminant were completely removed. Then, cleaned shells were decalcified in 1M acetic acid. The solution was centrifuged at 4000g for 30 min to separate the supernatant and the pellet. The insoluble pellet was rinsed with MilliQ water three times and was lyophilized as acid-insoluble matrix (AIM). The acid soluble matrix (ASM) dissolved in the supernatant was recovered by methanol/chloroform precipitation method as previously described (Takeuchi, et al. 2016).

A B

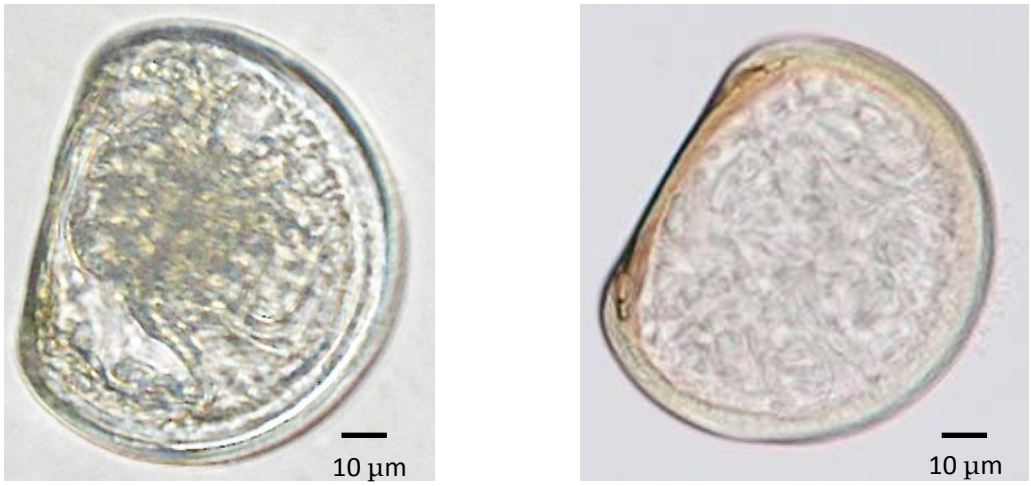


Fig. 2.1 Larval shells of *C. gigas* (A) and *P. fucata* (B) used in this study before NaOH treatment (Zhao, et al. 2018).

Adult shells of *Pinctada fucata* were provided by Mikimoto Pearl Research Institute. Shells were incubated in 1% NaOCl for 24 hours and mechanically washed

to remove superficial epibionts and periostracum. Two layers (outer prismatic and inner nacreous layers) were separated and crushed into tiny pieces. Ten grams of each shell layer was decalcified in 1M acetic acid overnight. Afterwards, acid soluble and insoluble matrices were obtained as mentioned above.

2.2.2 LC-MS/MS

AIM and ASM were suspended in solubilization buffer (1% SDS, 10mM DTT, 50mM Tris-HCl (pH 8.0)) and used for SDS-PAGE in 10%-20% gradient gel. Whole gel the proteins went through were excised and used for further analysis.

For the *Crassostrea gigas* shells, the excised gel bands were subjected to reduction/alkylation with dithiothreitol and iodoacetamide respectively followed by trypsin digestion overnight at 37℃. Peptides were extracted from gel using 5% formic Acid and 50% acetonitrile in water. After extraction, peptides were concentrated in Genevac EZ-2 Elite speed vacuum concentrator, and then resuspended in 0.1% formic acid in water for LC/MS analysis. An injection volume of 5 μL for each sample was injected into a Dionex Ultimate 3000 nano-UPLC system in tandem with a Thermo Q-Exactive Plus Mass Spectrometer, acquiring MS1 and MS2 of top 10 most intense peaks. Settings for the mass spectrometer are listed as follows: all ions with charge =< +2, and ions with charge >+8 were selected for MS/MS; Mass range = 350 to 1500 m/z; Mass tolerance for exclusion list, precursor high/low 10 ppm, using a time window of 20 sec. Peptides were separated in a Zorbax 300SB-C18 (0.3x150 mm; Agilent) column at 40 ℃, with 3 μL/min flow rate 90 min gradient. Acquired MS / MS spectra were subjected to database search against protein sequences of *C. gigas* gene models (Zhang, et al. 2012) and transcriptome assembly of *C. gigas* D-shaped larva, which is described below, complemented with the common Repository of Adventitious Proteins (cRAP; http://www.thegpm.org/crap/) database, using Proteome Discoverer software v1.4 (Thermo Fisher Scientific) - SEQUEST HT algorithm. Parameters set for the identification software are listed as follows: Select all MS/MS spectrum, if precursor ion charge is unknown, then +2 is assigned. No limit on mass

range; Mass tolerance for identification: precursor ion 20 ppm, fragment ion 0.1 Da. The results were filtered with a cut off 0.1 % false discovery rate at protein level. Not assigned spectra by SEQUEST HT were analyzed using PEAKs Studio 7 using de-novo sequencing first and then refined with database assisted search for alignment. Protein sequences identified by at least two unique peptides were accepted.

The *P. fucata* shell proteome data was provided by courtesy of Dr. Takeuchi, Dr. Yamada, and Prof. Sawada and retrieved from PRIDE DATABASE (https://www.ebi.ac.uk/pride/archive/).

2.2.3 Sequence analysis of SMPs

Blastp search against the UniProtKB/Swiss-Prot database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed using the default settings. SMART online service (http://smart.embl-heidelberg.de) and Pfam domain search (http://pfam.xfam.org/search) domain with default settings were applied to predict the functional domains, signal peptide, transmembrane domains and RLCDs. Molecular masses and isoelectric points (p*I*) of sequences were predicted using the Expasy ProtParam tool (www.expasy.org/tools/protparam.html). Sequence alignments performed on ClustalX programe embedded in Genetyx version 6 (Genetyx, Tokyo, Japan). Venn diagrams were drawn with VENNY 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/).

2.2.4 Transcriptome analysis

Total RNA of *Pinctada fucata* was extracted from the adult mantle tissues and 12 developmental stages including egg, 2 cells, 4 cells, early morula, blastula, 11 hour after fertilization trochophore, 13 h trochophore, 21 h D-shaped larva, 24 h D-shaped larva, 55 h D-shaped larva, 69 h D-shaped larva, and 96 h D-shaped larva with TRizol reagent (Chomczynski and Sacchi 1987). RNA-Seq libraries were prepared using a TruSeq RNA sample Prep Kit v2 (Illumina) and sequenced with Illumina GAIIx platform. Raw sequences were quality filtered and trimmed with Trimmomatic 0.32

(Bolger, et al. 2014), then mapped to the *P. fucata* gene model (Takeuchi, et al. 2016)using Bowtie2 (Langmead and Salzberg 2012) with default parameters. For gene expression analysis, TPM (transcripts per kilobase million) was calculated using eXpress 1.5.1 (Roberts and Pachter 2013) In the same way, RNA-Seq library of 24 h D-shaped larva of *Crassostrea gigas* was also prepared and sequenced with Illumina MiSeq. Transcriptome was assembled with Trinity 2.2.1 (Grabherr, et al. 2011).

2.3 Results and discussion

2.3.1 Shell matrix proteins (SMPs) of larval and adult shells

In order to acquire authentic shell matrix proteomes, only the proteins buttressed by at least two unambiguously identified peptide fragments were recorded in this study. The same standard was also adopted to reappraise the previously published adult shell matrix protein databases (Takeuchi, et al. 2016; Zhang, et al. 2012). With an overlap of 4 proteins, a shell proteome of *Crassostrea gigas* containing a total of 178 proteins, consisting of 111 larval shell proteins and 71 adult ones, was obtained (Fig. 2.2A, Table 2.1 and 2.3). In *Pinctada fucata*, a total of 31, 89, and 111 proteins were identified in the matrices of larval shell, adult nacre and adult prism, respectively. Taken together, a shell proteome of a total of 185 proteins, with 4 overlaps between larval and adult SMPs was established for *P. fucata* (Fig. 2.2b, Table 2.2 and 2.5).

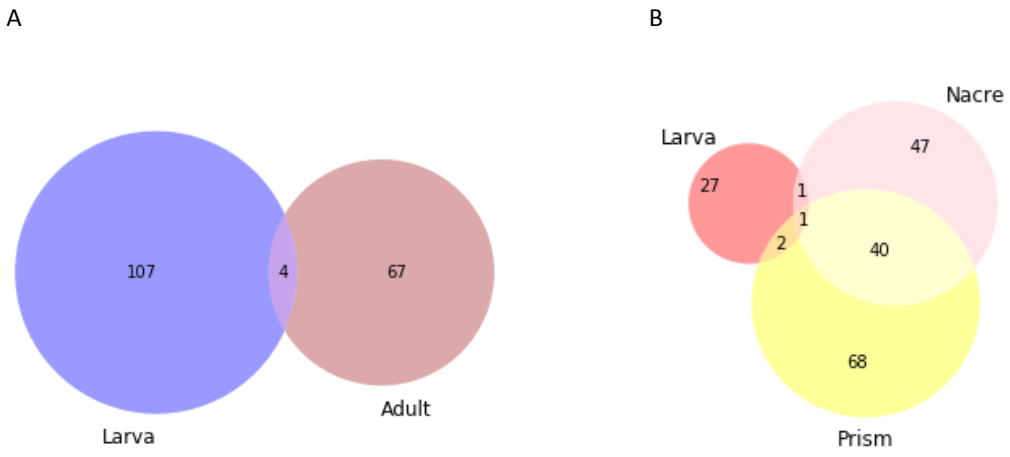
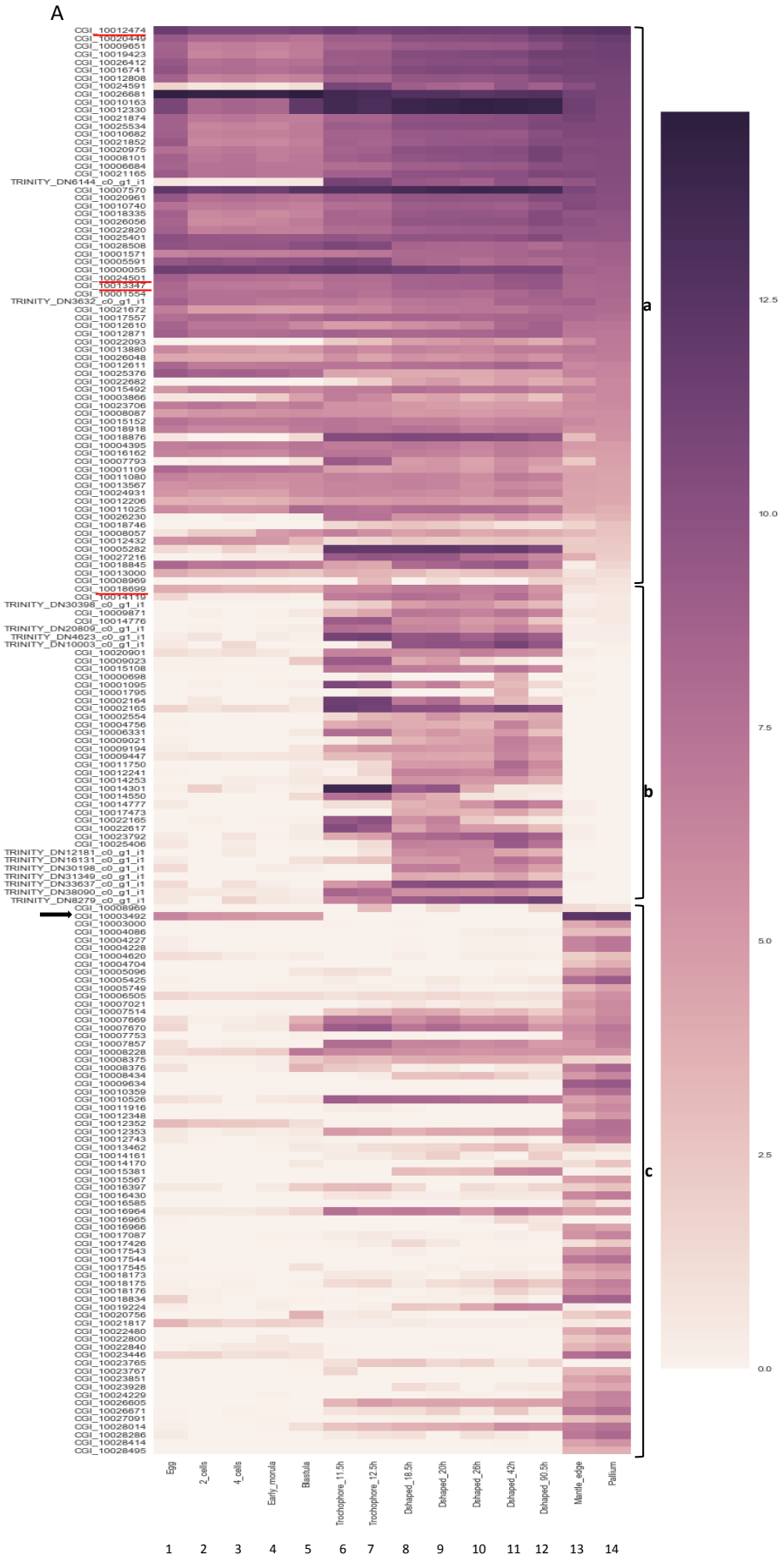


Fig. 2.2 Venn diagrams showing the number of shared and unshared shell matrix proteins between the larval and adult shells in *C. gigas* (A), and among larval, adult nacreous, and adult prismatic shells in *P. fucata* (B) (Zhao, et al. 2018).

2.3.2 Expression patterns of shell matrix protein genes in different developmental stages

Searched against the transcriptome database of *C. gigas* (Zhang, et al. 2012) and *P. fucata* (Takeuchi, et al. 2016), heatmaps showing the expression levels of the identified SMPs at each developmental stage in two species were drawn (Fig. 2.3). In Fig. 2.3A, SMPs of *C. gigas* which are highly expressed during both larval stages and adult stages comprise zone a, mainly by house-keeping genes and enzymatic proteins encoded genes, while other SMPs expressed at only larval stages and at only adult stages are located in zones b and c, respectively (Table 2.1). It is clear that expression patterns of larval SMPs are distinct from those of adult ones by gathering in zone a and zone b, with only one exception CGI_10003492 (Black arrowhead), an actin gene which peculiarly steadily expressed in the pre-adult stages until trochophore stage and proved to be the strongest expressed gene among all identified 178 SMPs of *C. gigas*. On the other hand, except the CGI_10008969, CGI_10012474, CGI_10013347 and CGI_10024501, 4 SMPs identified in both larval and adult shells of *C. gigas*, all the other 67 adult SMPs are located in zone c (A). Similarly, in Fig. 2.3B, except the 4 common proteins for both larval and adult shells of *P. fucata* which intensely expressed at pre-adult stages and adult stages in zone a, expression patterns of larval SMPs of (Table 2.2) in zone b explicitly discriminate to those of adult ones in zone c.


A
Egg
2_cells
4_cells
Early_morula
Blastula
Trochophore_11.5h
Trochophore_12.5h
Dshaped_18.5h
Dshaped_20h
Dshaped_26h
Dshaped_42h
Dshaped_90.5h
Mantle_edge
Pallium
1
2
3
4
5
6
7
8
9
10
11
12
13
14
a
b
c
12.5
10.0
7.5
5.0
2.5
0.0

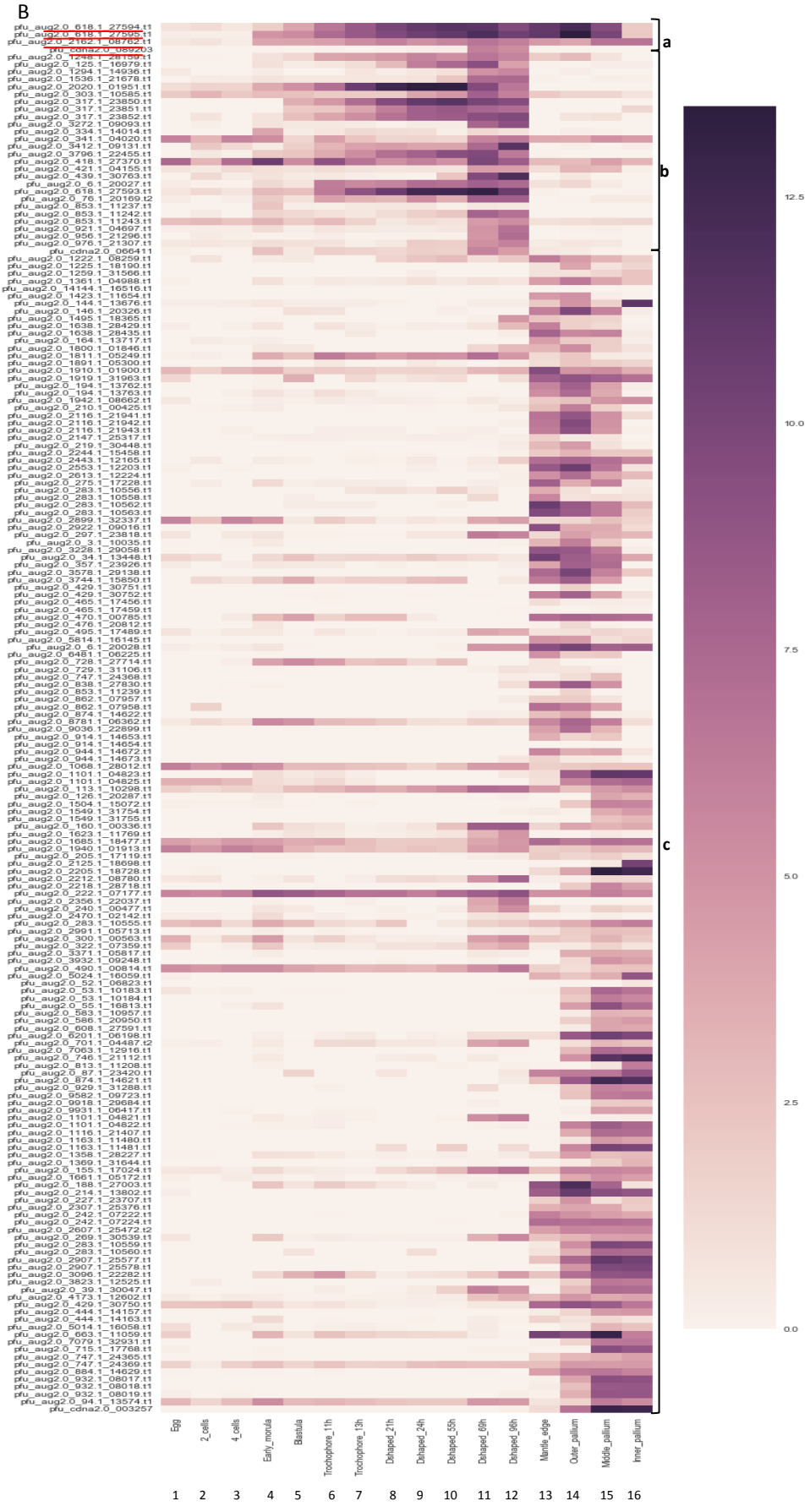

B
a
b
c
Egg
2_cells
4_cells
Early_morula
Blastula
Trochophore_11h
Trochophore_13h
Dshaped_21h
Dshaped_24h
Dshaped_55h
Dshaped_69h
Dshaped_96h
Mantle_edge
Outer_pallium
Middle_pallium
Inner_pallium
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
12.5
10.0
7.5
5.0
2.5
0.0

Fig2.3 Heatmaps showing the expression levels of the gene models for the identified SMPs in different developmental stages of *C. gigas* (A) and *P. fucata* (B). Gene models identified in both larval and adult shell proteomes, only in larval proteome, and only in adult proteome are classified into categories a, b, and c, respectively. (A) 1: egg; 2: 2 cells; 3: 4 cells; 4: early morula; 5: blastula; 6: trochophore 11.5 h; 7: trochophore 12.5 h; 8: D-shaped 18.5 h; 9: D-shaped 20 h; 10: D-shaped 26 h; 11: D-shaped 42 h; 12: D-shaped 90.5 h; 13: mantle edge; 14: pallium. (B) 1: egg; 2: 2 cells; 3: 4 cells; 4: early morula; 5: blastula; 6: trochophore 11 h; 7: trochophore 13 h; 8: D-shaped 21 h; 9: D-shaped 24 h; 10: D-shaped 55 h; 11: D-shaped 69 h; 12: D-shaped 96 h; 13: mantle edge; 14: outer pallium; 15: middle pallium; 16: inner pallium. Note that in *C. gigas*, a number of SMPs are highly expressed in both larval and adult stages. Common SMPs for larval and adult are underlined. Black arrow points an actin gene identified in larval shell of *C. gigas* (Zhao, et al. 2018).

2.3.3 Comparisons between larval and adult SMPs in each of the two species

To identify the contents of larval shell proteins and their possible roles in biomineralization, BLAST searches on the protein sequences obtained were performed. BLAST results show that, for *C. gigas* and *P. fucata*, 76 of 111 and 16 of 31 of identified larval SMPs are homologous to previously characterized proteins, respectively, including some well-known SMPs, such as Nacrein and PIF. In contrast to those of *P. fucata*, the larval SMPs of *C. gigas* are characterized by including a number of enzymes involved in diverse metabolic pathways, house-keeping gene products such as elongation factor 1 alpha and ribosomal proteins, as well as some possible structural proteins such as myosin, actin and tubulin, though this condition is consistent with the fact that the adult counterparts are also represented by those house-keeping proteins (Zhang, et al. 2012). This observation suggests that not only adult shells (Wang, et al. 2013b) but also larval shells may be formed under the cell-mediated processes in *C. gigas*. Comparisons between the larval and adult SMPs indicate that they are strikingly different in both species (Fig. 2.4 and Table 2.3). Within species, 4 entries retrieved from both larval and adult shells (Fig. 2.1) are ATP synthase subunit alpha, ATP synthase subunit beta, Elongation factor 1 alpha, L-ascorbate oxidase and Collagen alpha-5(VI) chain, Valine-rich protein, Mantle protein, pfu_aug2.0_2162.1_08762.t1 (unknown) for *C.gigas* and *P. fucata*, respectively (Fig. 2.4, Table 2.3). Although in the blast search results, common genes between larval and adult shells also include Nacrein, PIF, Peptidyl-prolyl cis-trans isomerase B and Nacrein, PIF, Beta-hexosaminidase in *C. gigas* and *P. fucata*,

respectively (Fig. 2.4), gene IDs and sequences of these proteins are distinct between the larva and the adult (Table 2.3, Figs. 2.8, 2.9), which indicates different homologs of these gene families are recruited by larval and adult shells. On the other hand, across species, larval shells are amazingly as dissimilar as their adult counterparts. Common SMPs take up 4.2% (6/142) between larval shells and 4.8% (11/230) between adult shells, despite the number is bargained by the fact that more than five times of SMPs have been identified in the adult shell of *P. fucata* compared with in its larval shell.

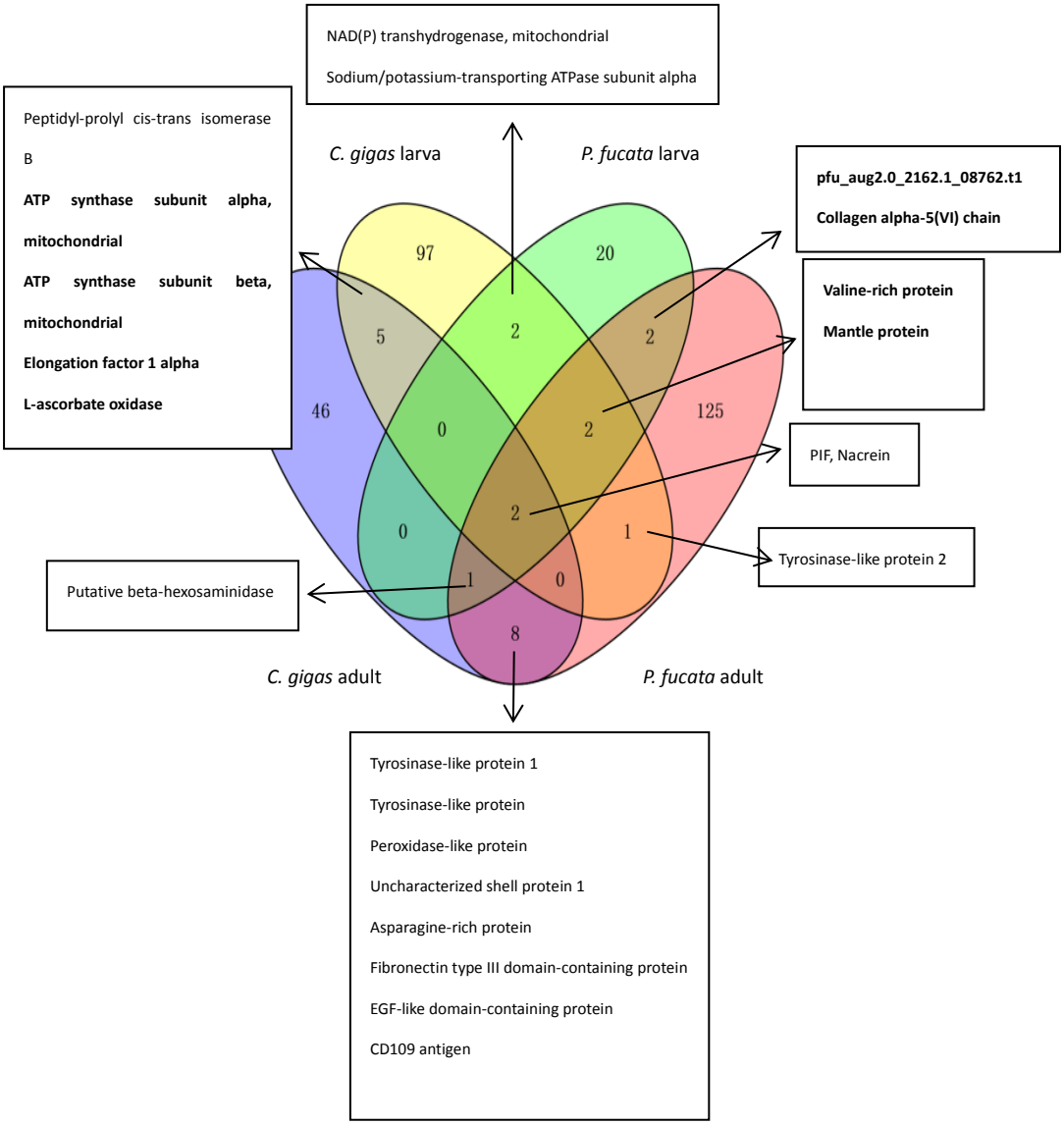


Figure 2.4 Comparisons of larval and adult SMPs of *C. gigas* and *P. fucata*. Proteins shown in bold characters represent SMPs identified in both larval and adult shells. Nacrein, PIF, Peptidyl-prolyl cis-trans isomerase B and Nacrein, PIF, Beta-hexosaminidase were identified in both larval and adult shell of *C.gigas* and *P. fucata*, respectively (Fig. 3), while gene IDs and sequences encode these proteins are different between larval and adult shells (Table 3). Venn diagram was drew using *VENNY 2.1* (http://bioinfogp.cnb.csic.es/tools/venny/).

2.3.4 Common domains between larval and adult SMPs

In order to infer possible functions of identified SMPs, domain searches were performed. Similar to the results of BLAST searches, the predicted domains of larval and adult shell matrix proteins are rather diverse by showing only 9 and 5 common domains between larval and adult SMPs for each species, respectively (Figs. 2.5a, b). Essential roles in the shell formation processes are suggested for three domains, carbohydrate-binding domain 14 (CBM_14), von Willebrand factor type A (VWA), carbonic anhydrase (CA), by their presence in the larval and adult shells of both species. Chitin provides the framework for $CaCO_3$ crystallization. CBM_14 is known as the Chitin binding Peritrophin-A domain. VWA-containing proteins are extracellular and characterized to be involved in forming multiprotein complexes. CA is a metalloenzyme that is critical for the hydration of carbon dioxide, $CO_2 + H_2O \rightarrow HCO_3^- + H^+$ (equation 1), which provides bicarbonate ions that directly react with Ca^{2+} to form $CaCO_3$. These three domains and the proteins containing them should be indispensable in the formation of both larval and adult shells (Figs. 2.4, 2.5d). In addition, although not shared by both larval and adult shells in both species, there are some other domains which might play important roles in the formation of biominerals, including Glycoside hydrolase family 20 (glycol_hydro_20), EF-hand, Lytic polysaccharide mono-oxygenese (LPMO_10), and Tyrosinase (Fig. 2.5). Glycoside hydrolases assist in the hydrolysis of glycosidic bonds in complex sugars. Glycol_hydro_20 might thus be in charge of the degradation of chitin, one of the most essential polysaccharides in shell formation. EF-hand is another domain common to larval and adult SMPs, although EF-hand_5 is employed for the larval shells of *C. gigas* and EF-hand_1 for others. With a Ca^{2+} binding short loop region, EF-hand domains are found in a large family of calcium-binding proteins, including extracellular signaling proteins, such as calmodulin, troponin C and S100B. LPMO_10 degrades cellulose, which resemble chitin in structure. Tyroninase is thought to function in the periostracum formation in *P. fucata* (Zhang and Zhang 2006).

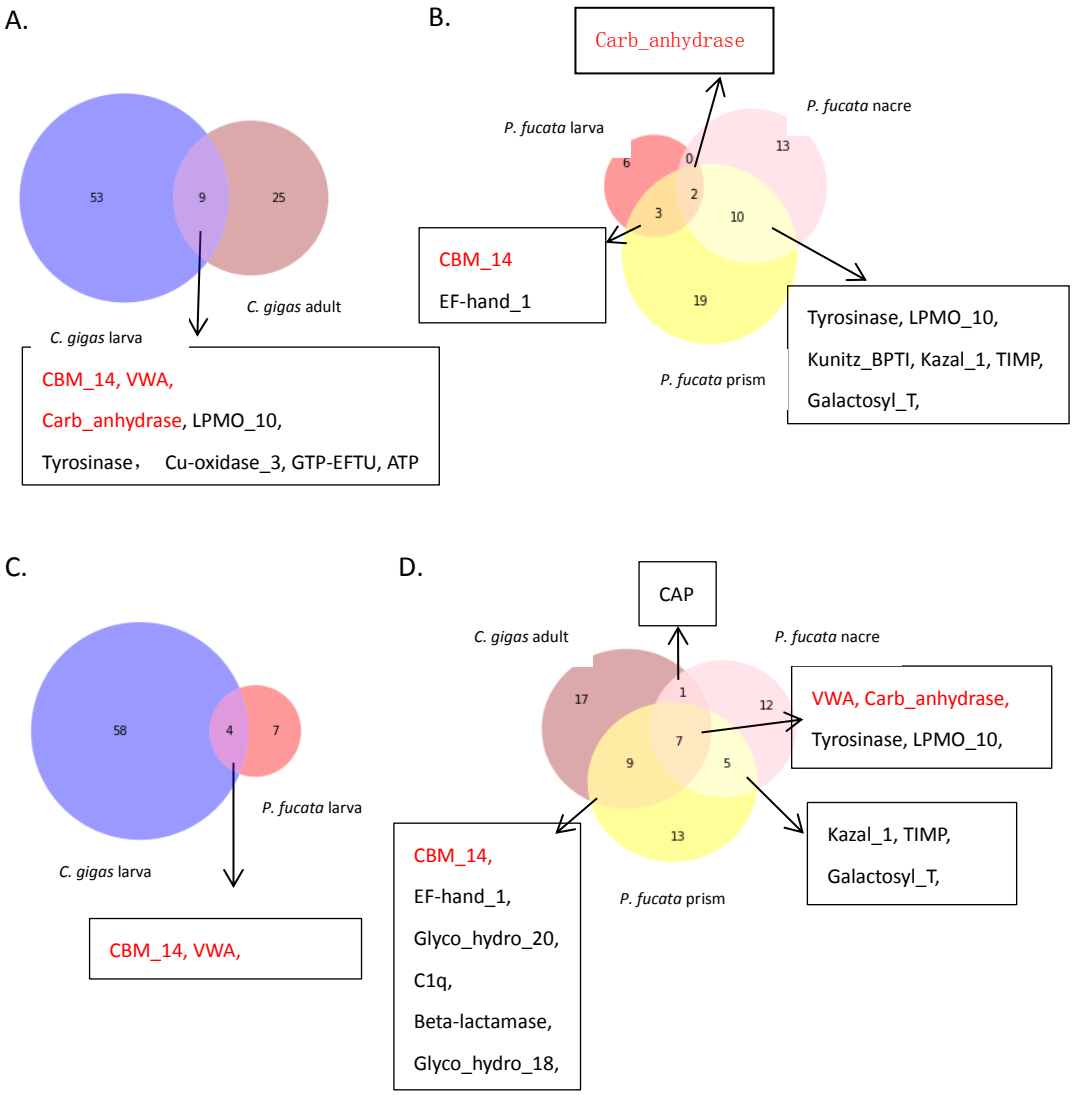


Fig. 2.5 Comparisons of domains included in the larval and adult SMPs of *C. gigas* and *P. fucata*. A, Nine domains are shared between larval and adult shells of *C. gigas*. B, Five domains are shared between larval and adult shells of *P. fucata*. C, Four domains are shared between the larval shells of *C. gigas* and *P. fucata*. D, Seventeen domains are shared between the adult shells of *C. gigas* and *P. fucata*. Common domains shared by both larva and adult of two species are colored in red.

The domains common to larval shells of both species include NAD(P) transhydrogenase beta subunit (PNTB), which is supposed to play important roles for larval shell formation in addition to CBM_14, VWA and CA (Fig. 2.5c). Considering the transition of H^+ in the reaction it catalyzes, despite its original role in anabolic pathways, NAD(P) transhydrogenase might help remove the H^+ produced by hydration of CO_2 and precipitation of $CaCO_3$ ($Ca^{2+} + HCO_3^- \rightarrow CaCO_3 + H^+$; equation 2), and therefore would assist both the equations 1 and 2 to proceed toward right. A shell protein with NAD(P) binding domain has also been identified in the brachiopod *Laqueus rubellus* (Isowa, et al. 2015), whose shell is also comprised of calcium carbonate.

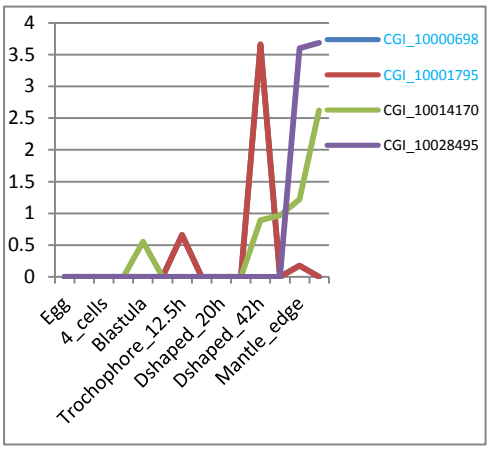
Besides those domains shared between larval shells, adult SMPs show some domains common to both species, domains which are already suggested to have diverse possible functions among shell proteins, such as Glyco_hydro_18, Fibronectin type III domain (fn3), CAP, Trypsin and Trypsin inhibitor Kunitz_BPTI (Fig. 2.5d). Also, some domains, which were not identified in the previous published proteome of the adult shell of *P. fucata* (Liu, et al. 2015) were identified in our database, such as glycol_hydro_20, EF-hand and Kunitz. Above all, although domains identified from larval SMPs are different from those of the adult counterparts, comparable members exist, and appear to have been recruited to fulfill primary biomineralization processes such as framework organization, ion concentration control, and crystal growth.

2.3.5 Larval and adult SMPs show different expression patterns during development

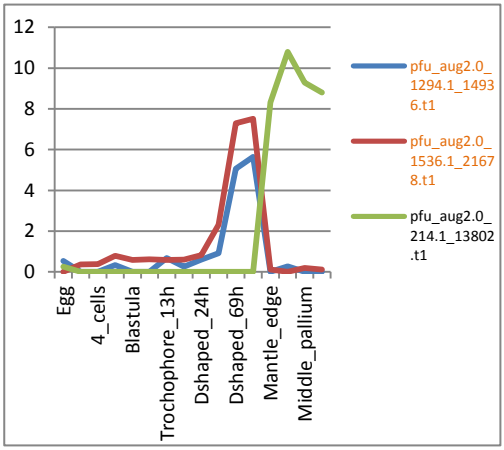
Considering the very limited concurrence of common SMPs between larval and adult shells in both species, it seems that each species possesses two "versions" of SMP repertoires, and this discrepancy appears to be extended to the realm of gene families whose members encode similar domains. For instance, we noticed that in the carbonic anhydrase gene families, multiple homologs occur in larval and adult shells, including the larval SMPs, CGI_10000698 and CGI_10001795, which are annotated to be Nacrein-like proteins, as well as the adult SMPs, CGI_10014170 and CGI_10028495. Their amino acid sequences are rather different as the alignment illustrates (Fig. 2.8). In the region of CA domain, sequences of the larval Nacrein are highly similar to each other and are distinguishable from those of adult ones, indicating that gene duplications may have occurred in each of the larval and adult CA-containing SMP genes. A similar condition is also observed in other genes, including those encoding PIF, Tyrosinase and Beta-hexosaminidase (Table 2.3), which contain the supposedly essential shell formation related domains, i. e., Chitin-binding, VWA, and Tyrosinase, respectively. SMPs identified in larval shells and in adult shells show clearly distinguishable expression patterns (Figs 2.6 A-I). The expression levels of larval SMPs show the peak at the D-shape larval stage or before that, and are apparently reduced or disappeared in the adult stage at the mantle edge and the mantle pallium, which are the main tissues in charge of the adult shell formation. In contrast, the gene expressions of most adult SMPs are barely detected in the larval stages, but are steeply increased in the adult shell forming tissues. As a few exceptions, the gene models encoding PIF, namely, CGI_10012352 and CGI_10028014, are relatively strongly expressed from the trochophore stage to D-shape stage, and their expression levels are indeed further increased in the mantle edge and pallium (Fig. 2.6C). An exception was also observed from CGI_10007857, a Beta-hexosaminidase/Chitobiase identified from the adult shell of *C. gigas*, which exhibits even higher level of expression in the larval stages than that of the adult stage (Fig. 2.6 E). Taken together, larval and adult SMPs repertoires are unambiguously

distinguishable, and within the same gene families, larval and adult homologs follow different and stage-specific expression patterns.

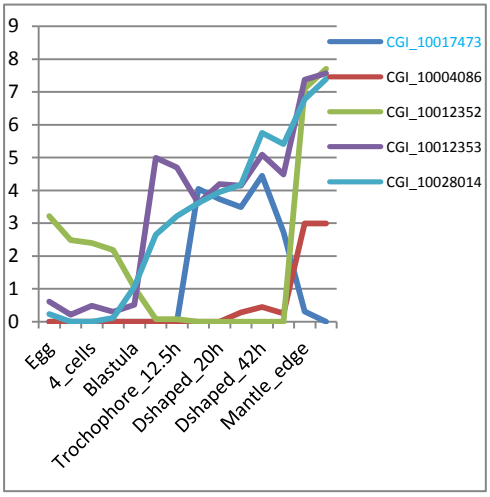
A. Nacrein-like (*C. gigas*)



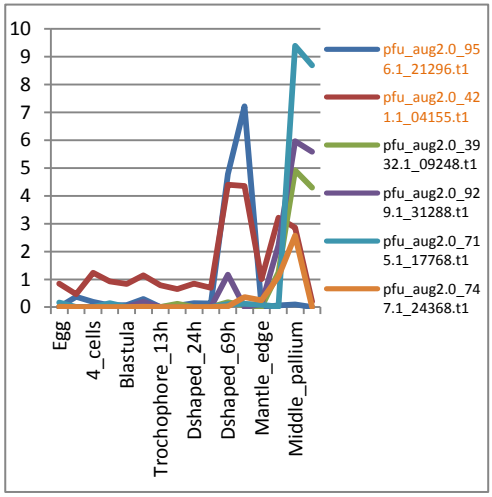
B. Nacrein-like (*P. fucata*)



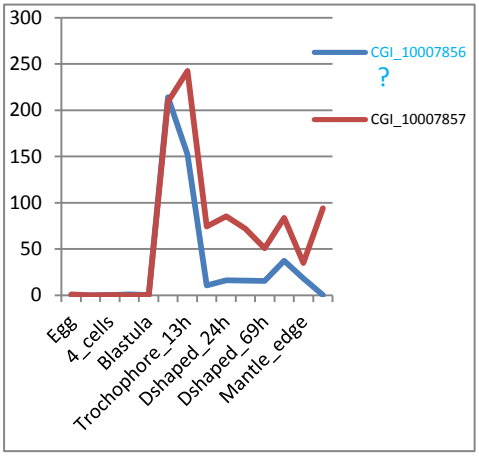
C. PIF (*C. gigas*)



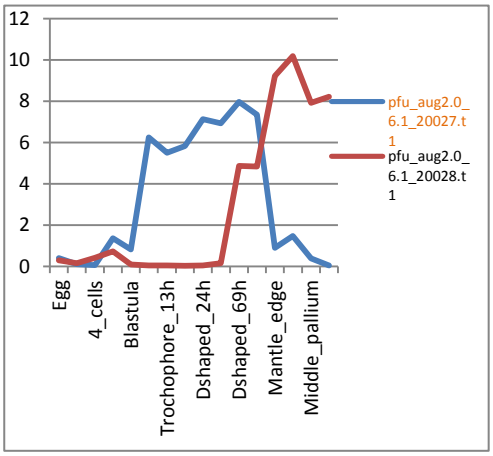
D. PIF (*P. fucata*)



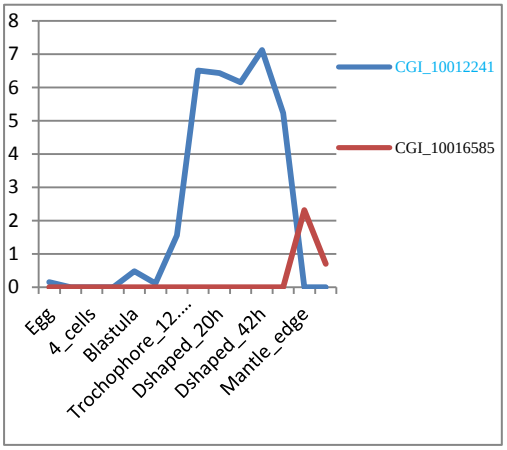
E. Beta-hexosaminidase (*C. gigas*)



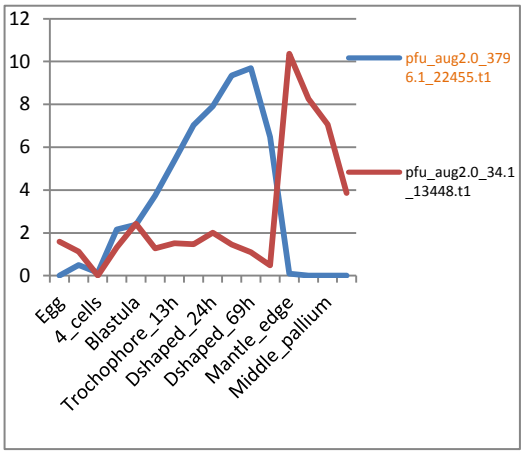
F. Beta-hexosaminidase (*P. fucata*)



G. EF-hand domain-containing proteins (*C. gigas*)



H. EF-hand domain-containing proteins (*C. gigas*)



I. Tyrosinase (*C. gigas*)

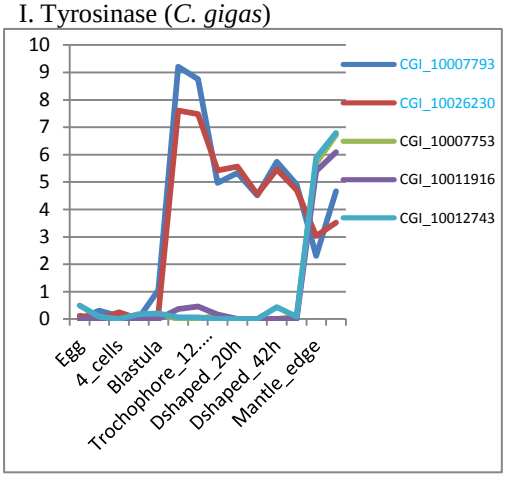


Fig. 2.6 Comparisons of the expression patterns of larval and adult SMPs within gene families. A and B, expression levels of gene models encoding Nacrein-like or Nacrein in *C. gigas* and *P. fucata*, respectively. C and C, expression levels of gene models encoding PIF in *C. gigas* and *P. fucata*, respectively. E, expression of Beta-hexosaminidases in *C. gigas*, Question mark indicates the gene was picked out by BLAST search against the genome of *C. gigas*, and if the gene is a SMP is questionable. F, expression levels of gene models encoding Beta-hexosaminidase in *P. fucata*. G, EF-hand domain-containing proteins in *C. gigas*. H, expression of EF-hand domain-containing proteins in *P. fucata*. I, expression levels of gene models encoding Tyrosinase in *C. gigas*. Gene models shown in blue characters and in orange characters represent larval SMPs of *C. gigas* and of *P. fucata*, respectively (Takeuchi, et al. 2012; Takeuchi, et al. 2016; Zhang, et al. 2012; Zhao, et al. 2018).

2.3.6 Larval shell matrix proteins and their putative functions

Based on the supposed functions inferred from the known homologous proteins and from the domains they contain, the identified larval SMPs can be roughly sorted into several groups: extracellular matrix (ECM) related, enzymes, acidic (low p*I* and/or D-rich) calcium binding, house-keeping, and others, which are not homologous to any previously characterized proteins or do not contain any identifiable domains (Tables 2.1, 2.2). In this section, SMPs from the two species will be discussed separately and previously discussed SMPs will only be mentioned briefly.

2.3.6.1 Larval SMPs of *Crassostrea gigas*

Extracellular shell matrix (ECM) related proteins: chitin-binding proteins, proteins involved in Wnt signaling and other proteins with extracellular domains. Four proteins contain chitin-binding domain, including PIF, Collagen, and two unknown proteins. Pif, the precursor of Pif97 and Pif80, contains VWA and chitin-binding domains, whose importance in the nacre formation has been confirmed (Suzuki, et al. 2009). CGI_10017473 is 26% identical to *Pinctada margaritifera* Pif97 and contains a VWA domain and two chitin-binding domains. Its chitin-binding domains are preceded by a T-rich motif (aa148-181; 65% T), which is not found in any of the four adult Pif homologs. However, similar arrangement of chitin-binding domain and T-rich motif has been reported in three Pif-like proteins of *Lottia gigantea* (Mann, et al. 2012). CGI_10009194 shows similarity to Collagen alpha-4(VI) with 22% identity, whose four VWA domains are followed by a chitin-binding domain (Fig. 2.7A). The triple helical structure of collagen prevents it from being broken down by enzymes, it enables adhesiveness of cells and it is important for the proper assembly of the extracellular matrix (Cunniffe and O'Brien 2011). Similar functions can be expected in molluscan shells. CGI_10014550 and CGI_10004756 are not homologous to any characterized protein. The presence of the signal peptide indicates that they are secreted into the extracellular shell matrix to take part in the formation of chitin framework. Besides chitin-binding domain, the former also contains a Laminin G (LamG) domain. As the important active part of lamina,

laminin glycoproteins are secreted and incorporated into cell-associated extracellular matrices, influencing cell differentiation, migration and adhesion by providing a network foundation for cells and organs (Timpl, et al. 1979). This protein might play an essential role in the shell matrix in a similar way. Embryonic processes controlled by Wnt signaling are necessary for proper formation of important tissues including bones, heart and muscles. CGI_10011025 shows high overall sequence similarity (88%) to Sperm-associated antigen 6 with an Armadillo domain. Armadillo/beta-catenin-like repeat mediates interaction of beta-catenin with its ligands, involved in transducing the Wnt signal during embryonic development (Peifer, et al. 1994; Peifer, et al. 1991). Another putative Wnt signaling related protein is CGI_10012432 that contains a signal peptide and two transmembrane domains, which display a 50% identity to a human Smoothened homolog, a Frizzled protein that serves as cell-surface receptor for Wnts (Malbon 2004). CGI_10018845 is 34% identical to human Periostin, which contains two Fasciclin (FAS1) domains. This gene is on average 29 folds and 5 folds highly expressed during early developmental stages (from Trochophore 11.5 h to D-shape 90.5 h) compared with mantle edge and pallium, respectively. FAS1 is an extracellular domain present in many secreted and membrane-anchored proteins (Huber and Sumper 1994). Similar to the fibronectin-like proteins with type III domain that are highly expressed in the mantle, FAS1 domain containing proteins mediate cell adhesion through an interaction with integrin (Kim, et al. 2000).

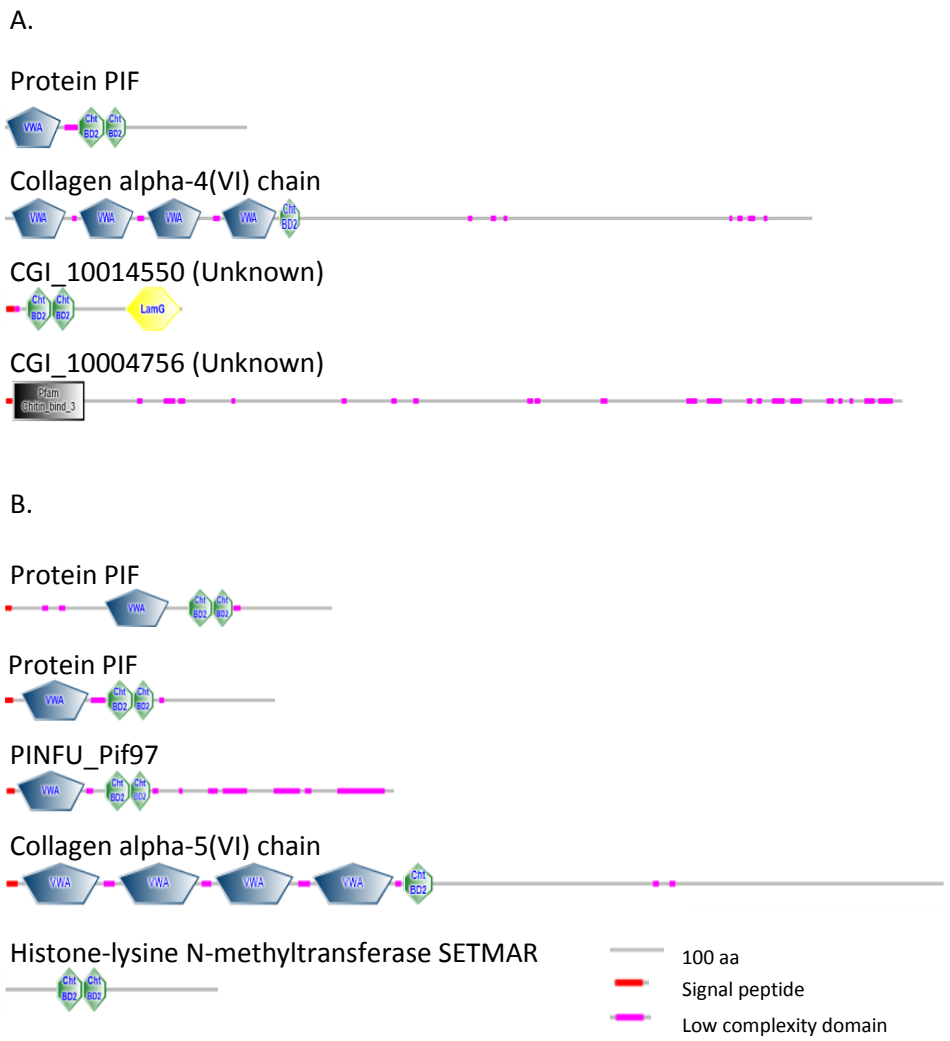


Fig. 2.7 Schematic representations of SMPs containing chitin-binding domains identified by SMART searches. A, SMPs with chitin-binding domain in *C. gigas*. B, SMPs with chitin-binding domain in *P. fucata*. PINFU_Pif97 is a PIF homolog found in adult *P. fucata*. VWA: von Willebrand factor type A domain; ChtBD2: carbohydrate-binding domain 2 domain; LamG: Laminin G domain.

Enzymatic proteins: SMPs grouped in this category include CA, NADP, NAD(P) transhydrogenase, redox enzymes, tyrosinases, isomerases, and others. Carbonic anhydrase is considered to be one of the most important enzymes in $CaCO_3$ biomineralization for its ability to catalyze hydration of carbon dioxide in the water to form hydrogen carbonate. In molluscs, Nacrein is the first isolated and characterized SMP containing carbonic anhydrase domains, a protein which has been identified in both nacreous and prismatic layers of pearl oysters (Miyamoto, et al. 1996; Miyashita 2000) and in various other molluscs (Mann, et al. 2012; Marie, et al. 2011a; Miyamoto, et al. 1996; Miyamoto, et al. 2003; Norizuki and Samata 2008). The gene models CGI_10000698 and CGI_10001795 containing a CA domain are homologous to *Mytilus californianus* Nacrein-like protein gene (Fig. 2.8). NAD(P) transhydrogenase and Sodium/potassium-transporting ATPase subunit alpha are the only two SMPs found in the larval shell of both species while not in their adult counterparts (Fig. 2.4, Table 2.3). In previous section, we have discussed the potential role of NAD(P) transhydrogenase in regulating hydrogen ion concentration in the formation of $CaCO_3$, although further experiments are required to corroborate it. In addition to the classical function in ion transporting, Sodium/potassium-transporting ATPase has also been shown to mediate the extracellular signal transduction through protein-protein interaction (Tian, et al. 2006) and to control neuron activity states (Zylbertal, et al. 2017); consequently, it may not simply be a "house-keeping" protein for ion gradient. Some other redox enzymes are also found in larval shell, including Ferric-chelate reductase 1 and L-ascorbate oxidase, one of the 4 common SMPs identified in both larval and adult shells, as well as two tyrosinases. Ferric-chelate reductase 1 contains a reeler domain, which is found in association with fibronectin type III domain and the Kunitz trypsin protease inhibitor domain. L-ascorbate oxidase acid, also known as Vitamin C, performs numerous physiological functions, such as synthesis of collagen, catabolism of tyrosine, and a prominent role in immune system (Gropper and Smith 2012; Wintergerst, et al. 2006). As the oxidase that controls the production of melanin, tyrosinases have been reported in several molluscan shells (Mann, et al. 2012; Marie, et al. 2012; Nagai, et al. 2007) and non-pigmented pallial

mantle (Zhang, et al. 2012), indicating that its function should not be restricted to the oxidation of tyrosine, and it might be necessary for biomeralization too. Indeed, tyrosinase was suggested to take part in the formation of periostracum (Zhang and Zhang 2006) which seals the extrapallial space from the environment, consequently allowing the accumulation of ions in sufficient concentration for crystallization to occur. In summary, despite further pieces of evidence are required, plenty of redox enzymes are supposed to join in the process of shell generation besides whose original parts in the embryonic generation of nervous system or immune system. Several isomerases and chaperonins have been found as well. Isomerases convert a molecule from one isomer to another, which facilitate the intracellular rearrangement in which bonds are broken or formed. Chaperonins are believed to display a sort of enzymatic function to add protein folding correctly, thus prevent aggregation. Although functions of these proteins here are unknown, by recalling that cyclophilin-type peptidyl-prolyl isomerases were detected in different shells (Jackson, et al. 2009; Marie, et al. 2013), as well as the severe osteogenesis imperfecta induced by cyclophilin deficiency in mice (Choi, et al. 2009), similar functions of these proteins in the proper shell calcification activity can be deduced. Other enzymatic proteins are also included in the larval shell, such as ATPase, ATP synthase and glycosyltransferase etc. The presence of signal peptide and the D-rich domain of CGI_10003866, a homolog of Zinc transporter ZIP12, a membrane transport protein that regulates zinc intracellular and cytoplasmic concentrations, indicates its extracellular destination and potential roles related to metalloenzymes, such as carbonic anhydrase, the active site of which contains a zinc ion.

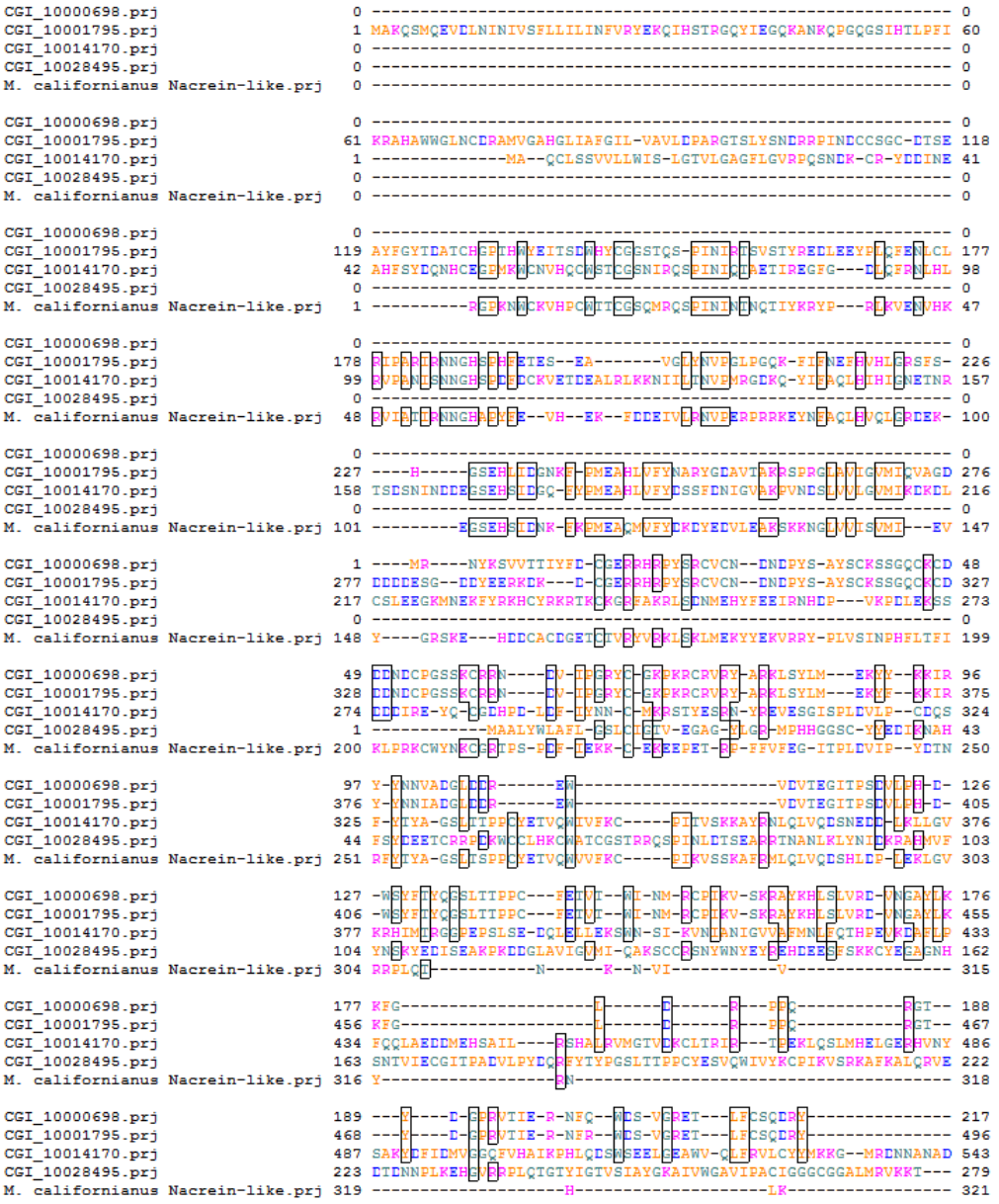


Fig. 2.8 Alignment of Nacrein-like proteins in *C. gigas*. CGI_10000698 and CGI_10001795 are larval SMPs. CGI_10014170 and CGI_10028495 are adult SMPs.

Acidic proteins: the existence of unusually acidic proteins has been known and confirmed by the pioneering work on the soluble proteins of the organic matrix of molluscan shells (Maurer, et al. 1996; Weiner 1979, 1983; Weiner and Hood 1975). However, because of the technical difficulties to isolate these proteins, until now only a few of them have been reported (Gotliv, et al. 2005; Marie, et al. 2013; Sarashina and Endo 1998; Suzuki, et al. 2009; Tsukamoto, et al. 2004) Although the function of which are poorly understood with only limited reports (Suzuki *et al*., 2009; Takeuchi *et al*., 2008), because their side chains are negatively charged under physiological environment, extreme acidic proteins (p*I* < 4.5), usually with D-rich domains, are supposed to bind calcium ions easily (Jackson, et al. 2009). Here, we report several proteins with p*I*s between 4.3 and 5.0, among which CGI_10012241 is the most acidic in the larval SMPs of *C. gigas*, with a predicted p*I* of 4.3, containing the typical canonical calcium-binding EF-hand domain (Kretsinger 1976).

House-keeping proteins: in this group, a total of 32 out of the 43 gene models grouped in this category lack a signal peptide, transmembrane domains or RLCDs, thus are not typical secreted or extracellular proteins that theoretically function in biomineralization. Although the extent of their contribution to the shell formation is argued, some of them including Actin, Tubulin, Elongation Factor 1a, and ribosomal proteins etc. have been observed in diverse metazoan calcified skeletons (Drake, et al. 2013; Isowa, et al. 2015; Jackson, et al. 2015; Mass, et al. 2014; Rahman, et al. 2013; Zhang, et al. 2012). Also, a speculation about a genuine function of actin that help cells move along their secretory tracks during the growth of brachiopod shell has been pointed out (Jackson, et al. 2015). Interestingly, except CGI_10003492, all members of this group belong to zone a in Fig. 2.3 A, in other words, they show the typical expression pattern of house-keeping genes. As an actin, the expression pattern of CGI_10003492 is rather curious. It is strongly expressed since the first cleavage till blastula stage, while the expression is paused when the first shell is formed at the trochophore stage, then the expression is restarted and even strengthened at the adult stage. However, this actin was identified in the larval shell, but not in the adult shell.

One explanation to this conflict is that after being translated in the very early stage, it is transported into the larval shell, while its destination at the adult stage is not the shell any more, but some other tissues.

Other proteins: this last group is comprised of 33 proteins, including homologs to Mantle protein, Valine-rich protein, Salivary glue protein Sgs-4, ER membrane protein complex subunit 4. A total of 29 proteins do not show a sequence similarity to any characterized proteins. In contrast to the identified house-keeping proteins, most of the these proteins contain a signal peptide, transmembrane domains, or RLCDs, which vindicates their connection to the shell formation, However, none of them is predicted to hold a function known domain, thus their roles in the biomineralization processes still remain obscure.

2.3.6.2 Larval SMPs of *Pinctada fucata*

Detailed information of SMPs identified in larval shell of *P. fucata* is listed in Table 2.2. Proteins that have been discussed previously will only be mentioned briefly in this section.

Extracellular shell matrix (ECM) related proteins: Three proteins have been predicted to contain the VWA and chitin-binding domains, including PIF and Collagen alpha-5(VI) chain protein, which was found in both larval and adult shells of *P. fucata*. As the larval version of PIF, pfu_aug2.0_421.1_04155.t1 and pfu_aug2.0_956.1_21296.t1 are highly similar to each other, and show lower sequence similarities, 27% and 25%, to the adult Pif97, respectively (Fig. 2.9). pfu_aug2.0_303.1_10585.t1 is another protein containing two consecutive chitin-binding domains, a sequence which is 31% identical (aa26-aa115) to human Histone-lysine N-methyltransferase SETMAR (Fig. 2.7 B). The SETMAR gene knockout line of mice exhibited abnormal retinal pigmentation or atypical peripheral blood lymphocyte parameters (Gardin and White 2011), however, the full length of human SETMAR gene does not contain a chitin-binding domain but SET domain and a transposase domain, which indicates the function of pfu_aug2.0_303.1_10585.t1 may have no relationship to SETMAR anymore, and the sequence similarity might have come from a gene recombination event. pfu_aug2.0_921.1_04697.t1 contains tandem Kazal domain repeats and a Membrane attack complex component/perforin (MACPF) domain. The former is known as the serine protease inhibitor and the latter is common to the membrane attack complex (MAC) proteins of the complement system (C6, C7, C8α, C8β and C9) and perforin (PF), which play roles in immunity system, such as lysing virally infected and transformed cells and delivering of cytotoxic proteases that cause cell death (Rosado, et al. 2008; Voskoboinik, et al. 2006). In the wild, the shell protect the animal against the bacterial degradation and infection, these functions are supposed to be essential to survive in the complex seawater containing various infectious microbes. Pfu_aug2.0_853.1_11237.t1 shows a 51% sequence similarity to *Acropora millepora* Mucin-like protein, and contains a

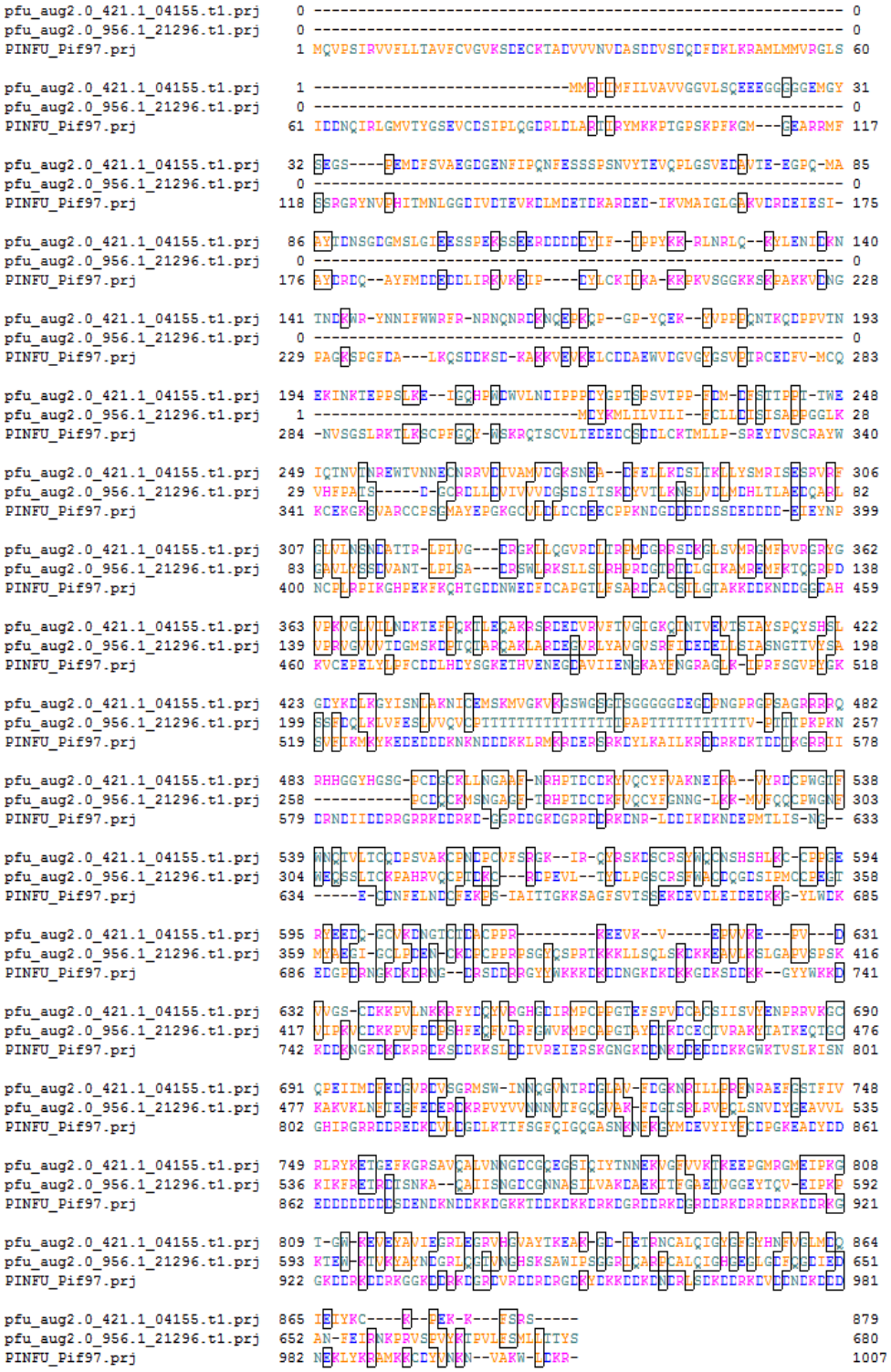


Fig. 2.9 Alignment of PIFs in *P. fucata*. pfu_aug2.0_421.1_04155.t1 and pfu_aug2.0_956.1_21296.t1 are homologous to the adult Pif97 of *P. fucata*.

thrombospondin type-1 (TSP1) repeat. TSP1 is a matricellular glycoprotein that inhibits the proliferation and migration of endothelial cells by triggering a series of gene expression that leads to the activation of caspases and apoptosis of the cell (Haviv, et al. 2005; Morris and Kyriakides 2014). This antiangiogenic function strongly suggests that the protein may directly take part in controlling the shell synthesis by deciding the fate of cells elaborated. Pfu_aug2.0_976.1_21307.t1 is 51% identical to *Sus scrofa* negative elongation factor D, containing a typical extracellular zona pellucida domain, which suggests its destination is the shell matrix. Homologs of *Bos taurus* Thyroglobulin are also identified. Thyroglobulin is the precursor of the thyroid hormones, which act on nearly every cell in the body, to increase the basal metabolic rate, to regulate the protein synthesis and the bone growth etc. Thyroid hormones have also been reported to induce the larval abalone metamorphosis (Fukazawa, et al. 2001).

Enzymatic proteins: Two homologs of Nacrein-like protein are identified. Alignment demonstrates that pfu_aug2.0_1536.1_21678.t1 shows higher sequence similarity to the adult Nacrein and the *Pinctada maxima* Nacrein-like protein M than to its larval companion pfu_aug2.0_1294.1_14936.t, indicating different ancestries for these proteins (Fig. 2.10). The gene model pfu_aug2.0_6.1_20027.t1 is identical to beta-hexosaminidases of *Pinctada margaritifera* and contains the N-terminal domain of chitobiases/beta-hexosaminidases (Fig. 2.11 and Table 2.2). Chitobiases degrade chitin (Tews, et al. 1996), thus this protein might be recruited for remodeling chitin framework.

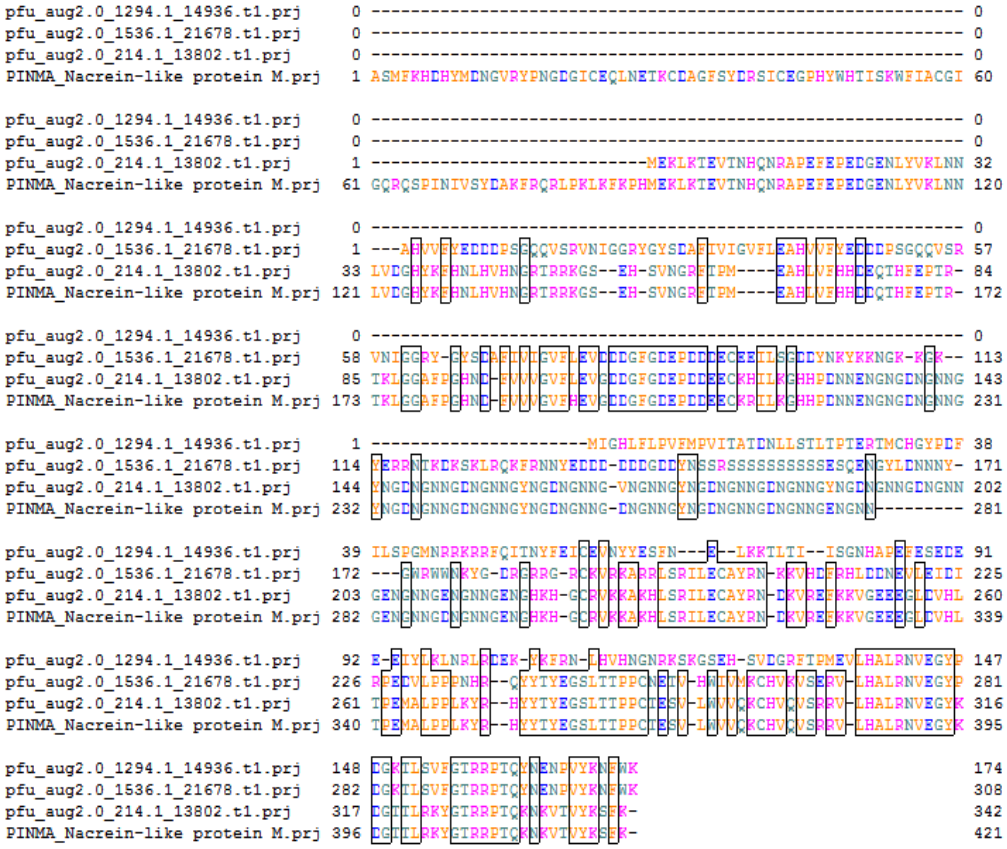


Fig. 2.10 Alignment of Nacreins in *P. fucata*. pfu_aug2.0_1294.1_14936.t1 and pfu_aug2.0_1536.1_21678.t1 are larval SMPs, while pfu_aug2.0_214.1_13802.t1 is adult SMP. Either of which is homologous to the Nacrein-like protein of *Pincdata maxima*.

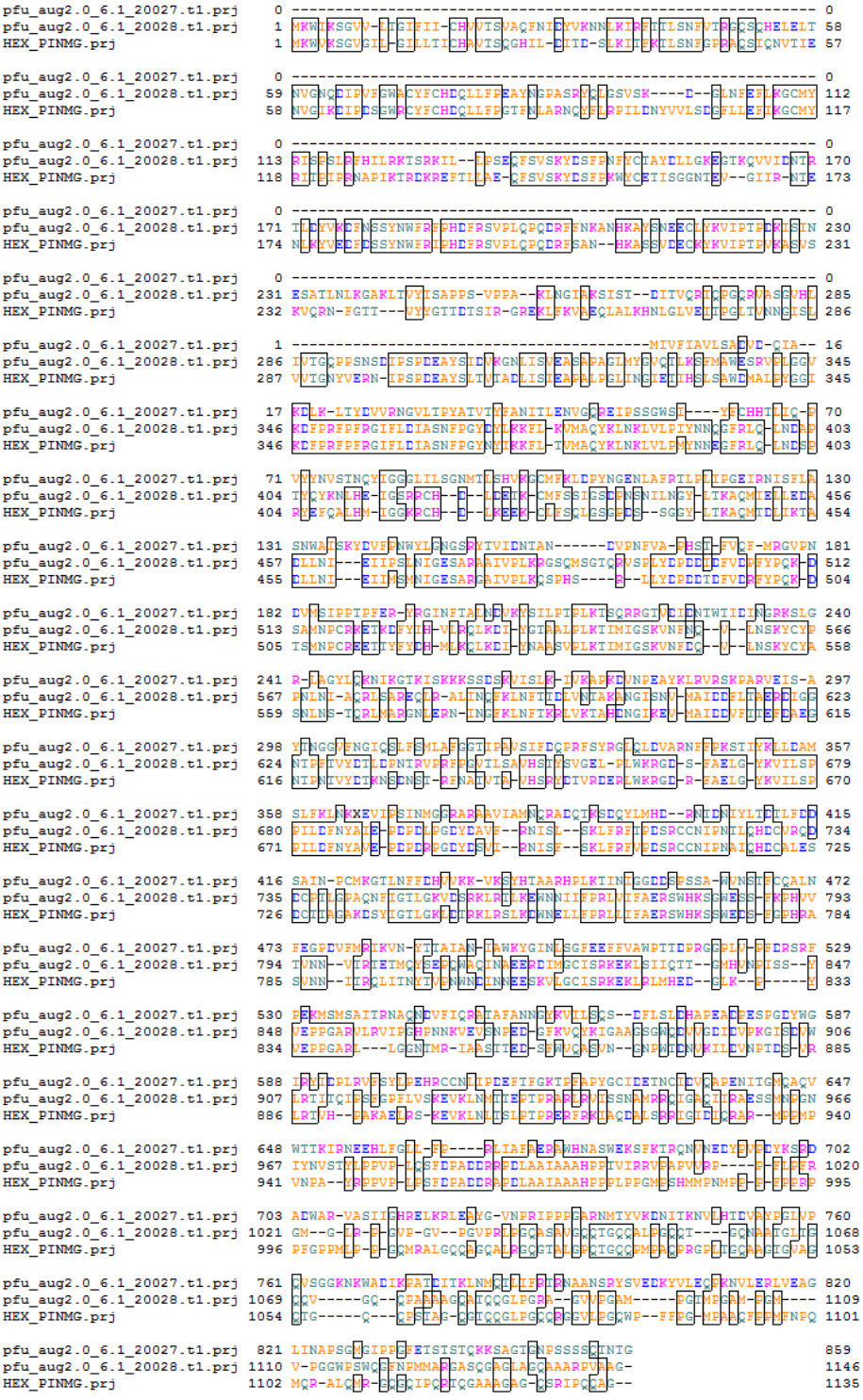


Fig. 2.11 Alignment of beta-hexosaminidase in *P. fucata*. Larval SMP pfu_aug2.0_6.1_20027.t1 and adult SMP pfu_aug2.0_6.1_20028.t1 both show sequence similarities to the putative beta-hexosaminidase of *Pinctada margaritifera*.

Acidic proteins: Three extremely acidic proteins have been found in the larval shell of *P. fucata*. pfu_aug2.0_3272.1_09093.t1 and pfu_aug2.0_2162.1_08762.t1 are secreted D-rich proteins with p*I*s of 4.2 and 3.8, respectively. Instead of the acidic D- or E-rich domains, pfu_aug2.0_3796.1_22455.t1, encoding a protein with a p*I* of 4.4, contains an S-rich domain (aa 95-119, 32%) and two calcium-binding EF-hand domains.

Other proteins: This group is comprised of 13 members including 3 entries, which are identical to melanoma-associated antigen (MAGE)-like protein 2, Valine-rich protein, and Mantle protein respectively, and the 10 remaining entries are not homologous to any known proteins. pfu_aug2.0_3412.1_09131.t1 shows a rather high similarity (52%) to mouse melanoma-associated antigen (MAGE)-like protein 2. However, the sequence similarity mainly comes from the matching of three repetitive amino acids, P, V and Q. Moreover, in mammals, members of MAGE gene family are described as completely silent in normal tissues, with exception of male germ cells, and, for some of them, placenta. By contrast, these genes are expressed in various kinds of tumors (Chomez, et al. 2001). Therefore, it is insufficient to claim that this protein is a molluscan MAGE-like protein. However, such a protein reminds us a hypothesis that functional motifs are rescued as "building blocks" by different "mosaic" proteins (Marin, et al. 2007).

2.4 Conclusions

Unlike adult shells containing diverse polymorphs of $CaCO_3$, molluscan larval shells all contain aragonite and similar microstructures (Carriker 1979; Eyster 1982; Eyster 1986; LaBarbera 1974; Waller 1981; Weiss, et al. 2002). By comparing the larval and adult SMP repertoires of two revolutionarily closely related species, we found some common genes that might play important roles in the larval shell formation, including Nacrein-like proteins, PIF, and NAD(P) transhydrogenase etc. (Fig. 2.4). Nevertheless, we cannot simply conclude larval shells share more intrinsic similarities than their adult counterparts, because the majority of larval SMPs are still diverse between two species and most of those larval common genes have also been found in their adult shells (Fig. 2.4 and adult shells of other molluscan species (LaBarbera 1974; Marie, et al. 2013). Therefore, aragonite precipitation in molluscan larva could be unrelated to the contents of SMPs. On the other hand, for the same species, SMPs repertoires of larval and adult shell are amazingly different to each other (Figs. 2.4, 2.5), even within gene families containing the same domains, different members with definitely distinct expression patterns are recruited by larval and adult shells (Fig. 2.5, Table 2.3), which indicates the larval and the adult shell have evolved independently from each other.

This first attempt for extracting molluscan larval SMPs seems effective through identifying a considerable sum of proteins possibly involved in the biomineralization process directly or indirectly, nevertheless, we do not consider the ones recorded here represent all proteins contained in the shells. Some of them, including such proteins as highly acidic proteins, may not be detected due to a lack of the trypsin cleavage sites, while proteins which have been identified might be occluded from the datasets because of the relatively high criteria applied in this study compared with those in some of previous researches (Liu, et al. 2015; Zhang, et al. 2012). By demonstrating the different SMPs repertoires between larval and adult shells, and identifying genes expressed in larval shells uniquely, we hope to provide some new clues to the establishment of molluscan shell formation models and the functional analyses of

various SMPs, e.g. *in vivo* gene manipulation experiments, in the future.

Several shell formation models have been proposed until now: the classical model of molluscan shell formation postulates that the mineralization occurs in a mantle-secreted matrix, where the acidic proteins-controlled nucleation and $CaCO_3$ crystallization occur in the framework of chitin and silk proteins (Furuhashi, et al. 2009; Weiner and Traub 1984). The cell-mediated models suggest calcium carbonate crystals and shell proteins produced by other organs are transported to the biomineralization front via the assistance of haemocytes, though the shell matrix is still the most important place for shell formation (Mount, et al. 2004; Wang, et al. 2013b). The involvement of cells is supported by the presence of many house-keeping genes, which are distributed to diverse metabolic pathways in the shell proteins datasets (Liu, et al. 2015; Wang, et al. 2013b; Zhang, et al. 2012).

Many identified larval SMPs of *C. gigas* are classified as house-keeping proteins (Table 2.1), a condition which is similar to the one previously reported for the proteome of the adult shell (Zhang, et al. 2012). Deposition of numerous house-keeping proteins distributed to diverse metabolic pathways in the shell has been taken as one of the critical evidence supporting the involvement of cells in biomineralization (Wang, et al. 2013b; Zhang, et al. 2012). However, we cannot completely exclude the possibility that the sample of larval shells was contaminated by the soft tissues left over after the treatment with NaOH before decalcification, although the eradication of the soft tissues was confirmed under microscope.

Table 2.1 Shell matrix proteins identified in the larva of *C. gigas*.

ID	E-vaule	Organism	UniProt best hit	SP	RLCDs	Pfam domain	Molecular mass (kDa)	p*I*	Location in heatmap
CGI_10000698	2E-23	Mytilus californianus	Nacrein-like protein (Fragment)	NO	NO	PF00194:Alpha carbonic anhydrase	2.5	9.1	b
CGI_10001795	4E-52	Mytilus californianus	Nacrein-like protein (Fragment)	NO	NO	PF00194:Alpha carbonic anhydrase	5.8	7.5	b
CGI_10012610	0	Sus scrofa	Isocitrate dehydrogenase [NADP], mitochondrial (Fragment)	NO	NO	PF00180:Isopropylmalate dehydrogenase-like domain	50	7.6	a
CGI_10012611	0	Bos taurus	NAD(P) transhydrogenase, mitochondrial	NO	YES	PF12769:NAD(P) transhydrogenase, alpha subunit, C-terminal PF01262:Alanine dehydrogenase/pyridine nucleotide transhydrogenase, NAD(H)-binding domain PF05222:Alanine dehydrogenase/pyridine nucleotide transhydrogenase, N-terminal PF02233:NADP transhydrogenase, beta subunit	134	8.4	a
CGI_10005591	0	Taenia solium	Sodium/potassium-transporting ATPase subunit alpha	NO	NO	PF00702:HAD-like domain PF00689:Cation-transporting P-type ATPase, C-terminal PF00122:P-type ATPase, A domain PF00690:Cation-transporting P-type ATPase, N-terminal	116	5.5	a
CGI_10028508	2E-69	Caenorhabditis elegans	Probable sodium/potassium-transporting ATPase subunit beta-3	NO	NO	PF00287:Sodium/potassium-transporting ATPase subunit beta	35	5.4	a
CGI_1001	3E-18	Mus	Ferric-chela	YE	YES	PF02014:Reeler domain	30	9.5	b

4119		musculus	te reductase 1	S					
CGI_10008969	1E-29	Cucumis sativus	L-ascorbate oxidase	NO	YES	PF07732:Multicopper oxidase, type 3 PF00394:Multicopper oxidase, type 1 PF07731:Multicopper oxidase, type 2	139	8.5	b
CGI_10026230	1E-44	Caenorhabditis elegans	Putative tyrosinase-like protein tyr-1	NO	YES	PF00271:Helicase, C-terminal PF00264:Tyrosinase copper-binding domain	64	9.4	a
CGI_10007793	4E-40	Pinctada margaritifera	Tyrosinase-like protein 2	YES	NO	PF00264:Tyrosinase copper-binding domain	74	9.3	a
CGI_10018746	3E-15	Ptilota filicina	Polyenoic fatty acid isomerase	YES	NO	PF01593:Amine oxidase	61	8.5	a
TRINITY_DN20809_c0_g1_i1	2E-13	Ptilota filicina	Polyenoic fatty acid isomerase	YES	NO	PF01593:Amine oxidase	59	7.3	b
CGI_10013880	7E-89	Gallus gallus	Peptidyl-prolyl cis-trans isomerase B	NO	NO	PF00160:Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain	22	8.8	a
CGI_10026048	0	Rattus norvegicus	Protein disulfide-isomerase	YES	YES	PF00085:Thioredoxin domain PF13848	55	4.6	a
CGI_10004395	0	Oryctolagus cuniculus	T-complex protein 1 subunit zeta	NO	NO	PF00118:Chaperonin Cpn60/TCP-1 family	58	6.2	a
CGI_10011080	2E-43	Oryzias latipes	10 kDa heat shock protein, mitochondrial	NO	NO	PF00166:Chaperonin Cpn10	18	5.6	a
CGI_10013567	0	Gallus gallus	T-complex protein 1 subunit eta	NO	NO	PF00118:Chaperonin Cpn60/TCP-1 family	47	6.7	a
CGI_10018918	0	Drosophila melanogaster	T-complex protein 1 subunit alpha	NO	NO	PF00118:Chaperonin Cpn60/TCP-1 family	75	6.8	a

CGI_1001 5152	0	Xenopus laevis	Transitional endoplasmic reticulum ATPase	NO	YES	PF02933:CDC48, domain 2 PF02359:CDC48, N-terminal subdomain PF00004:ATPase, AAA-type, core	87	5.2	a
CGI_1001 3347	0	Drosophila melanogaster	ATP synthase subunit beta, mitochondrial	NO	NO	PF00306:ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal PF00006:ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain	45	5	a
CGI_1002 4501	0	Pongo abelii	ATP synthase subunit alpha, mitochondrial	NO	NO	PF02874:ATPase, F1 complex alpha/beta subunit, N-terminal domain PF00006:ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain PF00306:ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal	60	8.5	a
CGI_1000 3866	2E-78	Bos taurus	Zinc transporter ZIP12	YES	YES	PF02535:Zinc/iron permease	71	5.2	a
CGI_1001 2206	0	Homo sapiens	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	NO	YES	PF02516:Oligosaccharyl transferase, STT3 subunit	88	9.2	a
CGI_1001 3000	8E-60	Mus musculus	Deoxyribonuclease gamma	YES	NO	PF03372:Endonuclease/exonuclease/phosphatase	50	4.9	b
CGI_1001 7473	3E-48	Pinctada margaritifera	Protein PIF	NO	YES	PF00092:von Willebrand factor, type A PF01607:Chitin binding domain	66	8.7	b
CGI_1000 9194	3E-39	Mus musculus	Collagen alpha-4(VI) chain	NO	YES	PF00092:von Willebrand factor, type A PF01607:Chitin binding domain	94	11	b
CGI_1000 4756	-	-	-	YES	YES	PF03067:Chitin-binding, domain 3	248	5.6	b

CGI_1001	-	-	-	YE	YES	PF01607:Chitin binding	46	8.4	b
4550				S		domain			
CGI_1001	0	Mus	Sperm-asso	NO	NO	PF00514:Armadillo	55	6.3	a
1025		musculus	ciated						
			antigen 6						
CGI_1001	0	Homo	Smoothene	YE	YES	PF01392:Frizzled domain	115	9.4	a
2432		sapiens	d homolog	S		PF01534:Frizzled protein			
CGI_1001	3E-38	Homo	Periostin	YE	NO	PF02469:FAS1 domain	30	6.9	b
8845		sapiens		S					
CGI_1001	-	-	-	YE	YES	PF13202:EF-hand domain	54	4.3	b
2241				S					
CGI_1000	-	-	-	YE	YES	-	19	4.4	b
5282				S					
CGI_1002	-	-	-	NO	YES	-	43	4.5	b
2165									
CGI_1001	0	Aplysia	78 kDa	YE	YES	PF00012:Heat shock protein	73	5	a
5492		californica	glucose-reg	S		70 family			
			ulated						
			protein						
CGI_1002	8E-23	Pinctada	Mantle	YE	YES	-	28	10	a
4591		maxima	protein	S					
TRINITY_	3E-38	Pinctada	Valine-rich	YE	YES	-	30	10	a
DN6144_c		maxima	protein	S					
0_g1_i1									
TRINITY_	0.0000	Drosophila	Salivary	YE	NO	-	11	11	b
DN31349_	03	melanogaster	glue protein	S					
c0_g1_i1			Sgs-4						
CGI_1001	5E-19	Danio rerio	ER	NO	YES	PF06417:Protein of unknown	38	9.3	b
1750			membrane			function DUF1077, TMEM85			
			protein						
			complex						
			subunit 4						
CGI_1000	-	-	-	NO	YES	-	2.3	12	b
1095									
CGI_1000	-	-	-	NO	YES	-	1.6	8.6	b
2164									
CGI_1000	-	-	-	YE	NO	-	1.5	9.4	b
2165				S					
CGI_1000	-	-	-	YE	NO	-	0.9	9.8	b
2554				S					
CGI_1000	-	-	-	NO	YES	-	38	5.4	b
6331									
CGI_1000	-	-	-	YE	YES	-	53	11	b
9021				S					

CGI_10009023	-	-	-	NO	YES	-	94	11	b
CGI_10009447	-	-	-	NO	YES	-	181	10	b
CGI_10009871	-	-	-	NO	YES	-	93	6.2	b
CGI_10014253	-	-	-	NO	YES	-	23	5	b
CGI_10014301	-	-	-	NO	YES	-	46	8.8	b
CGI_10014776	-	-	-	YES	YES	-	28	6.7	b
CGI_10014777	-	-	-	NO	YES	-	21	4.9	b
CGI_10015108	-	-	-	NO	YES	-	85	8.6	b
CGI_10020901	-	-	-	YES	NO	-	34	6.4	b
CGI_10022617	-	-	-	YES	YES	-	12	9.4	b
CGI_10022682	-	-	-	NO	YES	-	25	9.4	a
CGI_10023792	-	-	-	YES	YES	-	31	7.8	b
CGI_10025406	-	-	-	NO	YES	-	137	8.9	b
CGI_10027216	-	-	-	YES	YES	-	47	6.5	b
TRINITY_DN10003_c0_g1_i1	-	-	-	YES	YES	-	18	9.4	b
TRINITY_DN12181_c0_g1_i1	-	-	-	NO	YES	-	10	11	b
TRINITY_DN16131_c0_g1_i1	-	-	-	NO	YES	-	33	13	b
TRINITY_DN30198_c0_g1_i1	-	-	-	NO	NO	-	18	5.8	b
TRINITY_DN30398_c0_g1_i1	-	-	-	YES	NO	-	10	8.9	b
TRINITY_	-	-	-	NO	YES	-	10	10	b

DN33637_ c0_g1_i1									
TRINITY_ DN4623_c 0_g1_i1	-	-	-	NO	YES	-	13	6.6	b
TRINITY_ DN8279_c 0_g1_i1	-	-	-	NO	NO	-	40	5.2	b
TRINITY_ DN38090_ c0_g1_i1	-	-	-	NO	YES	-	13	12	b
CGI_1002 5376	1E-76	Asterias rubens	Histone H2A	NO	NO	PF16211 PF00125:Histone H2A/H2B/H3	14	11	a
CGI_1000 0055	1E-93	Xenopus tropicalis	Histone H3.3	NO	NO	PF00125:Histone H2A/H2B/H3	1.5	11	a
CGI_1000 8057	4E-68	Tigriopus californicus	Histone H2B.3	NO	YES	PF00125:Histone H2A/H2B/H3	13	11	a
CGI_1000 8087	8E-66	Xenopus tropicalis	Histone H4	NO	NO	PF15511	11	11	a
CGI_1000 1554	0	Rattus norvegicus	Eukaryotic initiation factor 4A-II	NO	NO	PF00271:Helicase, C-terminal PF00270:DEAD/DEAH box helicase domain	3.9	5.8	a
CGI_1000 1571	2E-102	Artemia salina	Elongation factor 1-gamma	NO	NO	PF02798:Glutathione S-transferase, N-terminal PF00043:Glutathione S-transferase, C-terminal	2.7	9.3	a
CGI_1001 2474	0	Danio rerio	Elongation factor 1-alpha	NO	NO	PF03143:Translation elongation factor EFTu/EF1A, C-terminal PF03144:Translation elongation factor EFTu/EF1A, domain 2 PF00009	50	9.1	a
CGI_1001 8876	0	Placopecten magellanicus	Actin, adductor muscle	NO	NO	PF00022:Actin family	42	5.3	a
CGI_1000 3492	0	Crassostrea gigas	Actin	NO	NO	PF00022:Actin family	42	5.3	c
CGI_1000 7570	0	Rattus norvegicus	Tubulin alpha-1A chain	NO	NO	PF03953:Tubulin/FtsZ, 2-layer sandwich domain PF00091:Tubulin/FtsZ, GTPase domain	53	5	a
CGI_1001	0	Paracentrotus	Tubulin	NO	NO	PF00091:Tubulin/FtsZ,	50	4.8	a

0163		lividus	beta chain			GTPase domain PF03953:Tubulin/FtsZ, 2-layer sandwich domain			
CGI_1001 2330	0	Paracentrotus lividus	Tubulin beta chain	NO	NO	PF00091:Tubulin/FtsZ, GTPase domain PF03953:Tubulin/FtsZ, 2-layer sandwich domain	50	4.8	a
CGI_1000 8101	1E-101	Argopecten irradians	40S ribosomal protein S18	NO	NO	PF00416:Ribosomal protein S13	18	11	a
CGI_1000 9651	2E-86	Homo sapiens	40S ribosomal protein S5	NO	NO	PF00177:Ribosomal protein S7 domain	14	11	a
CGI_1001 0740	7E-97	Rattus norvegicus	40S ribosomal protein S13	NO	NO	PF08069:Ribosomal protein S13/S15, N-terminal PF00312:Ribosomal protein S15	17	11	a
CGI_1001 6741	7E-143	Danio rerio	40S ribosomal protein S4, X isoform	NO	NO	PF16121 PF00900:Ribosomal protein S4e, central region PF00467:KOW PF08071:Ribosomal protein S4e, N-terminal	28	10	a
CGI_1001 9423	3E-142	Aplysia californica	40S ribosomal protein S6	NO	YES	PF01092:Ribosomal protein S6e	32	10	a
CGI_1002 0449	6E-114	Xenopus laevis	40S ribosomal protein S7	NO	NO	PF01251:Ribosomal protein S7e	22	10	a
CGI_1002 0961	6E-91	Heteropneustes fossilis	40S ribosomal protein S16	NO	NO	PF00380:Ribosomal protein S9	17	10	a
CGI_1002 1672	3E-150	Ictalurus punctatus	40S ribosomal protein S2	NO	YES	PF00333:Ribosomal protein S5, N-terminal PF03719:Ribosomal protein S5, C-terminal	30	10	a
CGI_1002 1852	6E-165	Pinctada fucata	40S ribosomal protein SA	NO	NO	PF00318:Ribosomal protein S2 PF16122	33	4.8	a
CGI_1002 5534	3E-141	Rattus norvegicus	40S ribosomal protein S3	NO	NO	PF07650:K Homology domain, type 2 PF00189	26	9.4	a
CGI_1002 6056	1E-105	Drosophila melanogaster	40S ribosomal	NO	NO	PF01479:RNA-binding S4 domain	64	9.9	a

			protein S9			PF00163:Ribosomal protein S4/S9, N-terminal			
CGI_1001 0682	7E-146	Rana sylvatica	60S acidic ribosomal protein P0	NO	NO	PF00428 PF00466:Ribosomal protein L10P	34	8.7	a
CGI_1002 1165	3E-92	Rattus norvegicus	60S ribosomal protein L18	NO	YES	-	21	12	a
CGI_1001 2808	9E-59	Lumbricus rubellus	60S ribosomal protein L14	NO	NO	PF01929:Ribosomal protein L14	18	11	a
CGI_1000 6684	0	Urechis caupo	60S ribosomal protein L4	NO	YES	PF00573:Ribosomal protein L4/L1e PF14374:60S ribosomal protein L4, C-terminal domain	43	11	a
CGI_1002 0975	2E-110	Pongo abelii	60S ribosomal protein L7	NO	NO	PF00327:Ribosomal protein L30, ferredoxin-like fold domain PF08079:Ribosomal protein L30, N-terminal	29	11	a
CGI_1002 1874	2E-89	Bos taurus	60S ribosomal protein L13a	NO	YES	PF00572:Ribosomal protein L13	35	10	a
CGI_1002 2820	5E-130	Drosophila melanogaster	60S ribosomal protein L10	NO	NO	PF00252:Ribosomal protein L10e/L16	25	10	a
CGI_1001 8335	8E-42	Xenopus tropicalis	60S ribosomal protein L8	NO	NO	PF03947:Ribosomal protein L2, C-terminal PF00181:Ribosomal Proteins L2, RNA binding domain	18	11	a
CGI_1002 6412	3E-90	Homo sapiens	60S ribosomal protein L18a	NO	NO	PF01775:Ribosomal protein L18a/LX	21	11	a
CGI_1001 2871	1E-172	Mus musculus	Phosphate carrier protein, mitochondrial	YES	YES	PF00153:Mitochondrial substrate/solute carrier PF07686:Immunoglobulin V-set domain	56	5.8	a
CGI_1001 6162	0	Bos taurus	Stress-70 protein, mitochondrial	NO	YES	PF00012:Heat shock protein 70 family	76	5.7	a

CGI_1001 7557	9E-123	Rattus norvegicus	ADP-ribosylation factor 2	NO	NO	PF00025:Small GTPase superfamily, ARF/SAR type	21	6.2	a
CGI_1002 5401	4E-139	Chlamydomonas reinhardtii	ADP,ATP carrier protein	NO	NO	PF00153:Mitochondrial substrate/solute carrier	33	9.6	a
CGI_1001 8699	5E-42	Homo sapiens	Cytochrome P450 2J2	NO	NO	PF00067:Cytochrome P450	27	6.4	a
CGI_1002 3706	4E-99	Gallus gallus	Cytochrome P450 1A5	NO	YES	PF07678:Alpha-macroglobulin complement component PF07703:Alpha-2-macroglobulin, N-terminal 2 PF01835:Alpha-2-macroglobulin, N-terminal PF00207:Alpha-2-macroglobulin PF00067:Cytochrome P450 PF07677:Alpha-macroglobulin, receptor-binding	229	6	a
CGI_1002 6681	1E-155	Pongo pygmaeus	Polyubiquitin-B	NO	NO	PF00240:Ubiquitin domain	25	6.4	a
CGI_1000 1109	2E-32	Homo sapiens	Heme-binding protein 2	NO	NO	PF04832:SOUL haem-binding protein	2.4	9.4	a
CGI_1002 2093	0	Argopecten irradians	Myosin heavy chain, striated muscle	NO	NO	PF00063:Myosin head, motor domain PF01576:Myosin tail PF02736:Myosin, N-terminal, SH3-like	230	5.5	a
TRINITY_ DN3632_c 0_g1_i1	-	-	-	NO	NO	PF02937：COX6C domain	12	9.4	a
CGI_1002 4931	1E-106	Drosophila melanogaster	Protein l(2)37Cc	NO	NO	PF01145:Band 7 protein	23	7.9	a

Bold text: common SMPs for larva and adult; '-' means not applicable; SP: signal peptide; RLCD: repeats of low complexity domains.

Table 2.2 Shell matrix proteins identified in the larva of *P. fucata*.

ID	E-vaule	Organism	UniProt best hit	SP	RLCDs	Pfam domain	Molecular mass (kDa)	p*I*	Location in heatmap
pfu_aug2.0_1294.1_14936.t1	3E-18	Pinctada maxima	Nacrein-like protein M	YES	NO	PF00194:Alpha carbonic anhydrase	21	8.5	a
pfu_aug2.0_1536.1_21678.t1	2E-46	Pinctada maxima	Nacrein-like protein M	NO	NO	PF00194:Alpha carbonic anhydrase	36	5.5	a
pfu_aug2.0_6.1_20027.t1	4E-72	Pinctada margaritifera	Putative beta-hexosaminidase	NO	NO	PF03173:Chitobiase/beta-hexosaminidases, N-terminal domain PF02838:Beta-hexosaminidase, bacteial type, N-terminal PF00728:Glycoside hydrolase family 20, catalytic domain	96	9.2	a
pfu_aug2.0_341.1_04020.t1	0	Bos taurus	NAD(P) transhydrogenase, mitochondrial	NO	YES	PF05222:Alanine dehydrogenase/pyridine nucleotide transhydrogenase, N-terminal PF02233:NADP transhydrogenase, beta subunit PF12769:NAD(P) transhydrogenase, alpha subunit, C-terminal PF01262:Alanine dehydrogenase/pyridine nucleotide transhydrogenase, NAD(H)-binding domain	122	6.7	a
pfu_aug2.0_418.1_27370.t1	1E-56	Taenia solium	Sodium/potassium-transporting ATPase subunit alpha	NO	NO	-	16	4.9	a
pfu_aug2.0_421.1_04155.t1	3E-55	Pinctada fucata	Protein PIF	YES	YES	PF00092:von Willebrand factor, type A	99	6.9	a

						PF01607:Chitin binding domain			
pfu_aug2.0_956.1_21296.t1	4E-55	Pinctada fucata	Protein PIF	YES	YES	PF00092:von Willebrand factor, type A PF01607:Chitin binding domain	75	8.3	a
pfu_cdna2.0_089203	8E-44	Mus musculus	Collagen alpha-5(VI) chain	YES	YES	PF01607:Chitin binding domain PF00092:von Willebrand factor, type A	262	7.4	b
pfu_aug2.0_303.1_10585.t1	6E-07	Homo sapiens	Histone-lysine N-methyltransferase SETMAR	NO	NO	PF01607:Chitin binding domain	55	8.8	a
pfu_aug2.0_921.1_04697.t1	-	-	-	NO	YES	PF07648:Kazal domain PF01823:Membrane attack complex component/perforin (MACPF) domain	136	9.5	a
pfu_aug2.0_853.1_11237.t1	2E-45	Acropora millepora	Mucin-like protein (Fragment)	NO	YES	PF00090:Thrombospondin type-1 (TSP1) repeat	69	9.2	a
pfu_aug2.0_976.1_21307.t1	1E-28	Sus scrofa	Negative elongation factor D	NO	YES	PF00100:Zona pellucida domain PF04858:TH1 protein	105	8.3	a
pfu_aug2.0_317.1_23850.t1	-	-	-	NO	NO	PF00086:Thyroglobulin type-1	8	7.6	a
pfu_aug2.0_317.1_23851.t1	1E-08	Bos taurus	Thyroglobulin	YES	NO	PF00086:Thyroglobulin type-1	16	6.3	a
pfu_aug2.0_317.1_23852.t1	1E-08	Bos taurus	Thyroglobulin	NO	NO	PF00086:Thyroglobulin type-1	19	8	a
pfu_aug2.0_3272.1_09093.t1	-	-	-	YES	YES	-	51	4.2	a
pfu_aug2.0_2162.1_08762.t1	-	-	-	YES	YES	-	43	3.8	b
pfu_aug2.0_3796.1_22455.t1	-	-	-	NO	YES	-	39	4.4	a
pfu_aug2.0_3	7E-07	Mus	MAGE-like protein 2	NO	YES	-	17	5	a

412.1_09131.t1		musculus							
pfu_aug2.0_618.1_27594.t1	3E-87	Pinctada maxima	Valine-rich protein	NO	YES	-	61	9.9	b
pfu_aug2.0_618.1_27595.t1	5E-65	Pinctada maxima	Mantle protein	YES	YES	-	30	10	b
pfu_aug2.0_334.1_14014.t1	-	-	-	NO	YES	-	17	5	a
pfu_aug2.0_1248.1_28159.t1	-	-	-	YES	YES	-	91	10	a
pfu_aug2.0_125.1_16979.t1	-	-	-	YES	YES	-	50	11	a
pfu_aug2.0_2020.1_01951.t1	-	-	-	NO	YES	-	22	12	a
pfu_aug2.0_439.1_30763.t1	-	-	-	YES	YES	-	27	9.4	a
pfu_aug2.0_618.1_27593.t1	-	-	-	YES	YES	-	54	6.5	a
pfu_aug2.0_76.1_20169.t2	-	-	-	NO	YES	-	86	11	a
pfu_aug2.0_853.1_11242.t1	-	-	-	YES	NO	-	26	12	a
pfu_aug2.0_853.1_11243.t1	-	-	-	NO	YES	-	101	11	a
pfu_cdna2.0_066411	-	-	-	YES	YES	-	181	11	a

Bold text: common SMPs for larva and adult; '-' means not applicable; SP: signal peptide; RLCD: repeats of low complexity domains.

Table 2.3 List of the common proteins (shown by gene model IDs) between larval and adult shells of *C. gigas* and *P. fucata* revealed by Blastp search.

	C. gigas larval SMPs	*C. gigas* adult SMPs	*P. fucata* larval SMPs	*P. fucata* adult SMPs
Nacrein	CGI_10000698 CGI_10001795	CGI_10014170 CGI_10028495	pfu_aug2.0_1294.1_14936.t1 pfu_aug2.0_1536.1_21678.t1	pfu_aug2.0_214.1_13802.t1
Pif	CGI_10017473	CGI_10004086 CGI_10012352 CGI_10012353 CGI_10028014	pfu_aug2.0_956.1_21296.t1 pfu_aug2.0_421.1_04155.t1	pfu_aug2.0_3932.1_09248.t1 pfu_aug2.0_929.1_31288.t1 pfu_aug2.0_715.1_17768.t1 pfu_aug2.0_747.1_24368.t1
Tyrosinase	CGI_10007793 (2) CGI_10026230 (tyr-1)	CGI_10007753 (1) CGI_10011916 (tyr-3) CGI_10012743		pfu_aug2.0_2553.1_12203.t1 pfu_aug2.0_6481.1_06225.t1 (1) pfu_aug2.0_9036.1_22899.t1 (2) pfu_aug2.0_914.1_14653.t1 (1) pfu_aug2.0_914.1_14654.t1 (1) pfu_aug2.0_242.1_07222.t1 (1) pfu_aug2.0_242.1_07224.t1 (1)
Peptidyl-prolyl cis-trans isomerase B	CGI_10013880	CGI_10023851		
L-ascorbate oxidase	CGI_10008969			
Elongation factor 1-alpha	CGI_10001571 (gamma) CGI_10012474 (alpha)			
ATP synthase subunit beta, mitochondrial	CGI_10013347			
ATP synthase subunit alpha, mitochondrial	CGI_10024501			
Valine-rich protein	TRINITY_DN6144_c0_g1_i1		pfu_aug2.0_618.1_27594.t1	
Mantle protein	CGI_10024591		pfu_aug2.0_618.1_27595.t1	
Collagen	CGI_10009194 Collagen alpha-4(VI) chain		pfu_cdna2.0_089203 Collagen alpha-5(VI) chain	pfu_cdna2.0_089203 Collagen alpha-5(VI) chain pfu_aug2.0_2218.1_28718.t1 Collagen alpha-1(I) chain
Putative beta-hexosaminidase		CGI_10007857	pfu_aug2.0_6.1_20027.t1	pfu_aug2.0_6.1_20028.t1
Peroxidase-like protein		CGI_10017426		pfu_aug2.0_2147.1_25317.t1
Uncharacterized shell protein 1		CGI_10016430		pfu_aug2.0_275.1_17228.t1 pfu_aug2.0_2607.1_25472.t2
Asparagine-rich protein		CGI_10010359		pfu_aug2.0_1358.1_28227.t1
CD109 antigen		CGI_10023765		pfu_aug2.0_144.1_13676.t1

Fibronectin		CGI_10016964 (2) CGI_10016965 (2) CGI_10016966 (2)		pfu_aug2.0_429.1_30751.t1 (1) pfu_aug2.0_429.1_30752.t1 (1) pfu_aug2.0_429.1_30750.t1 (2)
EGF-like domain-containing protein		CGI_10017544 (2) CGI_10017545 (2)		pfu_aug2.0_3578.1_29138.t1 (1) pfu_aug2.0_2116.1_21941.t1 (1) pfu_aug2.0_2116.1_21942.t1 (2) pfu_aug2.0_2116.1_21943.t1 (2) pfu_aug2.0_838.1_27830.t1 (2)
NAD(P) transhydrogenase, mitochondrial		CGI_10012611		pfu_aug2.0_341.1_04020.t1
Sodium/potassium-transporting ATPase subunit alpha		CGI_10005591		pfu_aug2.0_418.1_27370.t1
pfu_aug2.0_2162.1_08762.t1			**pfu_aug2.0_2162.1_08762.t1**	

Characters in bold are proteins identified in both larval and adult shells of *C. gigas* or *P. fucata*.

Table 2.4 Shell matrix proteins identified in the adult of *C. gigas*.

ID	E-value	Organism	UniProt best hit	SP	RLCDs	Pfam domain	Molecular mass (kDa)	p*I*	Location in heatmap
CGI_10014170	9E-77	Mytilus californianus	Nacrein-like protein (Fragment)	YES	NO	PF00194:Alpha carbonic anhydrase PF00042:Globin	62	5.7	c
CGI_10028495	5E-27	Mytilus californianus	Nacrein-like protein (Fragment)	YES	NO	PF00194:Alpha carbonic anhydrase	31	8.5	c
CGI_10023928	2E-97	Gluconobacter oxydans	L-sorbose 1-dehydrogenase	NO	NO	PF05199:Glucose-methanol-choline oxidoreductase, C-terminal PF00732:Glucose-methanol-choline oxidoreductase, N-terminal	65	9.1	c
CGI_10008969	1E-29	Cucumis sativus	L-ascorbate oxidase	NO	YES	PF07732:Multicopper oxidase, type 3 PF00394:Multicopper oxidase, type 1 PF07731:Multicopper oxidase, type 2	139	8.5	b
CGI_10004620	2E-09	Bacillus subtilis	Spore cortex-lytic enzyme	NO	NO	PF07486:Cell wall hydrolase, SleB	17	5	c
CGI_10007857	9E-113	Pinctada margaritifera	Putative beta-hexosaminidase	NO	YES	PF02838:Beta-hexosaminidase, bacteial type, N-terminal PF03173:Chitobiase/beta-hexosaminidases, N-terminal domain PF00728:Glycoside hydrolase family 20, catalytic domain	112	6.6	c
CGI_10014161	2E-54	Haliotis rufescens	Chymotrypsin-like serine proteinase	NO	NO	PF00089:Serine proteases, trypsin domain	30	5.9	c
CGI_10015381	5E-53	Gadus morhua	Chymotrypsin B	YES	NO	PF00089:Serine proteases, trypsin domain	30	6.4	c
CGI_10017426	5E-124	Pinctada margaritifera	Peroxidase-like protein	NO	YES	PF03098:Haem peroxidase, animal	87	8.4	c

CGI_1002667 1	1E-15 8	Drosophila melanogaster	Probable sulfite oxidase, mitochondrial	NO	YES	PF00173:Cytochrome b5-like heme/steroid binding domain PF02931:Neurotransmitter-gated ion-channel ligand-binding domain PF02932:Neurotransmitter-gated ion-channel transmembrane domain PF00092:von Willebrand factor, type A PF03404:Moybdenum cofactor oxidoreductase, dimerisation PF00174:Oxidoreductase, molybdopterin-binding domain	196	5.6	c
CGI_1001883 4	8E-09	Rattus norvegicus	Extracellular superoxide dismutase [Cu-Zn]	YES	NO	PF00080:Superoxide dismutase, copper/zinc binding domain	48	5.6	c
CGI_1000775 3	2E-31	Pinctada margaritifera	Tyrosinase-like protein 1	NO	YES	PF00264:Tyrosinase copper-binding domain	80	9.1	c
CGI_1001191 6	2E-37	Caenorhabditis elegans	Putative tyrosinase-like protein tyr-3	NO	YES	PF00264:Tyrosinase copper-binding domain	67	8.7	c
CGI_1001274 3	6E-44	Pinctada maxima	Tyrosinase-like protein	NO	YES	PF00264:Tyrosinase copper-binding domain	116	8.1	c
CGI_1002385 1	5E-62	Gallus gallus	Peptidyl-prolyl cis-trans isomerase B	NO	YES	PF00160:Cyclophilin-type peptidyl-Sheet3!A113+Sheet3!A113+Sheet1!prolyl cis-trans isomerase	28	7.7	c

						domain			
CGI_1001639	7E-12	Mus	Histone-lysine	NO	YES	PF13771	549	6.4	c
7	5	musculus	N-methyltransferas			PF00759:Glycoside			
			e 2C			hydrolase, family 9			
						PF00628:Zinc			
						finger, PHD-finger			
						PF00264:Tyrosinas			
						e copper-binding			
						domain			
CGI_1000542	0	Crassostrea	Gigasin-6	YES	YES	PF00144:Beta-lacta	62	6.4	c
5		gigas				mase-related			
CGI_1001235	2E-09	Pinctada	Protein PIF	NO	YES	PF07679:Immunogl	123	4.9	c
2		margaritifer				obulin I-set			
		a							
CGI_1001235	7E-07	Pinctada	Protein PIF	NO	YES	PF01607:Chitin	61	4.9	c
3		margaritifer				binding domain			
		a							
CGI_1002801	3E-42	Pinctada	Protein PIF	YES	YES	PF00092:von	129	5.1	c
4		fucata				Willebrand factor,			
						type A			
CGI_1000408	1E-35	Pinctada	Protein PIF	NO	NO	PF00092:von	165	5.6	c
6		fucata				Willebrand factor,			
						type A			
CGI_1002841	6E-30	Xenopus	Kielin/chordin-like	NO	NO	PF00093:VWFC	80	7.8	c
4		laevis	protein			domain			
CGI_1001346	-	-	-	YES	YES	PF03067:Chitin-bin	61	9.6	c
2						ding, domain 3			
CGI_1001708	-	-	-	NO	YES	PF01607:Chitin	44	9.1	c
7						binding domain			
CGI_1001817	-	-	-	YES	YES	PF03067:Chitin-bin	89	9.7	c
6						ding, domain 3			
CGI_1002660	1E-14	Homo	Chitotriosidase-1	NO	YES	PF01607:Chitin	79	8.5	c
5	2	sapiens				binding domain			
						PF00704:Glycoside			
						hydrolase family			
						18, catalytic domain			
CGI_1001696	2E-16	Pinctada	Fibronectin type III	YES	YES	PF00041:Fibronecti	89	6.4	c
4	2	margaritifer	domain-containing			n type III			
		a	protein 2						
CGI_1001696	2E-13	Pinctada	Fibronectin type III	NO	YES	PF00041:Fibronecti	70	6	c
5	6	margaritifer	domain-containing			n type III			
		a	protein 2						
CGI_1001696	2E-50	Pinctada	Fibronectin type III	NO	NO	PF00041:Fibronecti	29	8.9	c
6		margaritifer	domain-containing			n type III			

		a	protein 2						
CGI_1000300 0	3E-14	Homo sapiens	Complement C1q-like protein 2	YES	NO	PF00386:C1q domain	82	5	c
CGI_1002828 6	-	-	-	NO	NO	PF14625:Lustrin, cysteine-rich repeated domain PF05375:Pacifastin domain	25	8.7	c
CGI_1001754 3	0	Crassostrea gigas	Gigasin-2	YES	NO	-	41	7.4	c
CGI_1001754 4	4E-83	Pinctada maxima	EGF-like domain-containing protein 2	YES	NO	-	40	5.2	c
CGI_1001754 5	8E-74	Pinctada maxima	EGF-like domain-containing protein 2	YES	NO	-	40	6	c
CGI_1002075 6	4E-12	Caretta caretta	Chelonianin	NO	YES	PF00014:Pancreatic trypsin inhibitor Kunitz domain	83	10	c
CGI_1001556 7	9E-19	Anemonia sulcata	KappaPI-actitoxin-Avd3a	NO	NO	PF00014:Pancreatic trypsin inhibitor Kunitz domain	11	6.1	c
CGI_1000470 4	2E-17	Clostridium cellulolyticum	Translation initiation factor IF-2	NO	YES	PF00095:WAP-type 'four-disulfide core' domain	184	9	c
CGI_1000509 6	-	-	-	NO	YES	PF00095:WAP-type 'four-disulfide core' domain	126	9.7	c
CGI_1002248 0	5E-25	Mus musculus	Glioma pathogenesis-related protein 1	NO	NO	PF00188:CAP domain	14	8.7	c
CGI_1001052 6	5E-15	Aplysia californica	Temptin	YES	NO	-	16	7.6	c
CGI_1002181 7	8E-53	Caenorhabditis elegans	Vitellogenin-6	YES	YES	PF09172:Vitellinogen, open beta-sheet PF00094:von Willebrand factor, type D domain PF01347:Lipid transport protein, N-terminal	273	9.1	c
CGI_1000751 4	6E-06	Oncorhynchus mykiss	Retinol-binding protein 4-B	YES	NO	PF00061:Lipocalin/ cytosolic fatty-acid binding domain	29	6.5	c

CGI_1001658 5	1E-15	Arabidopsis thaliana	Calmodulin-like protein 11	NO	YES	PF13499:EF-hand domain pair	66	4.6	c
CGI_1000650 5	-	-	-	YES	YES	-	115	4.5	c
CGI_1001035 9	1E-29	Pinctada margaritifer a	Asparagine-rich protein	NO	YES	-	61	4.8	c
CGI_1000574 9	8E-14 8	Crassostrea gigas	Gigasin-3a (Fragment)	YES	YES	-	72	8.5	c
CGI_1000837 6	2E-36	Crassostrea gigas	Gigasin-4 (Fragment)	YES	YES	-	23	9.3	c
CGI_1000837 5	4E-10	Crassostrea gigas	Gigasin-4 (Fragment)	NO	YES	-	92	8.3	c
CGI_1001643 0	1E-20	Pinctada maxima	Uncharacterized shell protein 1	NO	YES	-	16	6.5	c
CGI_1002709 1	9E-06	Arabidopsis thaliana	Proline-rich protein 2	NO	YES	-	180	10	c
CGI_1002280 0	2E-38	Ruditapes philippinar um	Insoluble matrix shell protein 1 (Fragment)	NO	NO	-	29	5.6	c
CGI_1000822 8	3E-13	Mus musculus	Transmembrane protein 145	NO	YES	-	58	8.8	c
CGI_1000422 7	-	-	-	NO	YES	-	85	11	c
CGI_1000422 8	-	-	-	YES	NO	-	49	9.9	c
CGI_1000766 9	-	-	-	NO	YES	-	36	9.3	c
CGI_1000767 0	-	-	-	YES	NO	-	32	8.6	c
CGI_1000843 4	-	-	-	NO	YES	-	70	9.9	c
CGI_1000963 4	-	-	-	NO	YES	-	73	9.2	c
CGI_1001817 3	-	-	-	NO	YES	-	105	9.6	c
CGI_1001817 5	-	-	-	NO	YES	-	88	9	c
CGI_1001922 4	-	-	-	YES	YES	-	26	7.2	c
CGI_1002284 0	-	-	-	NO	YES	-	137	9.2	c
CGI_1002344 6	-	-	-	NO	YES	-	35	10	c

CGI_10024229	-	-	-	NO	YES	-	39	9.6	c
CGI_10012474	0	Danio rerio	Elongation factor 1-alpha	NO	NO	PF03143:Translation elongation factor EFTu/EF1A, C-terminal PF03144:Translation elongation factor EFTu/EF1A, domain 2 PF00009	50	9.1	a
CGI_10013347	0	Drosophila melanogaster	ATP synthase subunit beta, mitochondrial	NO	NO	PF00306:ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal PF00006:ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain	45	5	a
CGI_10024501	0	Pongo abelii	ATP synthase subunit alpha, mitochondrial	NO	NO	PF02874:ATPase, F1 complex alpha/beta subunit, N-terminal domain PF00006:ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain PF00306:ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal	60	8.5	a
CGI_10007021	6E-77	Rattus norvegicus	Sodium-dependent multivitamin transporter	NO	NO	PF00474:Sodium/solute symporter	86	6.8	c
CGI_10012348	1E-83	Homo sapiens	Hemicentin-1	NO	YES	PF13927 PF13895 PF07679:Immunoglobulin I-set	643	5.3	c
CGI_10023765	1E-164	Homo sapiens	CD109 antigen	NO	YES	PF00207:Alpha-2-macroglobulin PF07678:Alpha-macroglobulin	94	5.3	c

						complement component PF10569:Alpha-2-macroglobulin, thiol-ester bond-forming PF07677:Alpha-macroglobulin, receptor-binding			
CGI_10023767	6E-64	Homo sapiens	CD109 antigen	NO	YES	PF07703:Alpha-2-macroglobulin, N-terminal 2 PF01835:Alpha-2-macroglobulin, N-terminal	95	5.6	c

Bold text: common SMPs for larva and adult; '-' means not applicable; SP: signal peptide; RLCD: repeats of low complexity domains.

Table 2.5 Shell matrix proteins identified in the adult of *P. fucata*.

ID	E-value	Organism	UniProt best hit	SP	RLCDs	Pfam domain	Molecular mass	p*I*	Location in heatmap
pfu_aug2.0_214.1_13802.t1	0	Pinctada fucata	Nacrein	NO	NO	PF01391:Collagen triple helix repeat PF00194:Alpha carbonic anhydrase	38	6.2	c
pfu_aug2.0_6.1_20028.t1	0	Pinctada margaritifera	Putative beta-hexosaminidase	NO	YES	PF03173:Chitobiase/beta-hexosaminidases, N-terminal domain PF02838:Beta-hexosaminidase, bacteial type, N-terminal PF00728:Glycoside hydrolase family 20, catalytic domain	127	9.1	c
pfu_aug2.0_194.1_13762.t1	0	Pinctada margaritifera	Putative chitinase 1	NO	NO	PF00704:Glycoside hydrolase family 18, catalytic domain	54	5.3	c
pfu_aug2.0_194.1_13763.t1	0	Pinctada maxima	Putative chitinase	YES	NO	PF00704:Glycoside hydrolase family 18, catalytic domain PF01607:Chitin binding domain	64	8.6	c
pfu_aug2.0_2244.1_15458.t1	6E-12	Arthrobacter globiformis	Dextranase	NO	YES	PF03718:Glycoside hydrolase, family 49	55	9.7	c
pfu_aug2.0_164.1_13717.t1	7E-23	Bos taurus	Tissue-type plasminogen activator	NO	NO	PF00089:Serine proteases, trypsin domain	38	9.2	c
pfu_aug2.0_14144.1_16516.t1	1E-12	Arabidopsis thaliana	Proline-rich protein 1	NO	NO	PF03098:Haem peroxidase, animal	25	7.8	c
pfu_aug2.0_2147.1_25317.t1	0	Pinctada margaritifera	Peroxidase-like protein	YES	YES	PF03098:Haem peroxidase, animal	84	8.8	c
pfu_aug2.0_2613.1_12224.t1	1E-53	Pinctada margaritifera	Peroxidase-like protein	NO	YES	PF03098:Haem peroxidase, animal	38	9.2	c
pfu_aug2.0_465.1_17456.t1	0	Pinctada margaritifera	Peroxidase-like protein	NO	NO	PF03098:Haem peroxidase, animal	76	9.3	c
pfu_aug2.0_465.1_17459.t1	0	Pinctada margaritifera	Peroxidase-like protein	NO	YES	PF03098:Haem peroxidase, animal	76	8.9	c

pfu_aug2.0_1225.1_18190.t1	-	-	-	YES	NO	-	20	8.5	c
pfu_aug2.0_608.1_27591.t1	2E-60	Aplysia kurodai	Aplysianin-A	NO	YES	PF01593:Amine oxidase	98	9.5	c
pfu_aug2.0_1361.1_04988.t1	5E-24	Gallus gallus	DBH-like monooxygenase protein 1	NO	NO	PF01082:Copper type II, ascorbate-dependent monooxygenase, N-terminal PF03712:Copper type II ascorbate-dependent monooxygenase, C-terminal	62	8.6	c
pfu_aug2.0_94.1_13574.t1	0	Bos taurus	Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial	NO	YES	PF05187:Electron transfer flavoprotein-ubiquinone oxidoreductase PF13450	265	5.1	c
pfu_aug2.0_2553.1_12203.t1	0	Pinctada maxima	Tyrosinase-like protein	NO	NO	PF00264:Tyrosinase copper-binding domain	50	6	c
pfu_aug2.0_6481.1_06225.t1	3E-50	Pinctada margaritifera	Tyrosinase-like protein 1	YES	YES	PF00264:Tyrosinase copper-binding domain	44	9.3	c
pfu_aug2.0_9036.1_22899.t1	2E-62	Pinctada margaritifera	Tyrosinase-like protein 2	YES	YES	-	16	9.5	c
pfu_aug2.0_914.1_14653.t1	0	Pinctada margaritifera	Tyrosinase-like protein 1	NO	YES	PF00264:Tyrosinase copper-binding domain	53	8.9	c
pfu_aug2.0_914.1_14654.t1	9E-40	Pinctada margaritifera	Tyrosinase-like protein 1	NO	NO	PF00264:Tyrosinase copper-binding domain	15	6.3	c
pfu_aug2.0_242.1_07222.t1	0	Pinctada margaritifera	Tyrosinase-like protein 1	YES	YES	PF00264:Tyrosinase copper-binding domain	60	6.4	c
pfu_aug2.0_242.1_07224.t1	0	Pinctada margaritifera	Tyrosinase-like protein 1	YES	YES	PF00264:Tyrosinase copper-binding domain	58	6.7	c
pfu_aug2.0_1358.1_28227.t1	0	Pinctada margaritifera	Asparagine-rich protein	NO	YES	PF01762:Glycosyl transferase, family 31	114	9.4	c
pfu_aug2.0_444.1_14157.t1	3E-24	Homo sapiens	Sulfotransferase 1C2	YES	NO	PF00685:Sulfotransferase domain	85	7.3	c
pfu_aug2.0_8	-	-	-	NO	NO	PF00144:Beta-lactamase-relat	40	5.9	c

781.1_06362.t1						ed			
pfu_aug2.0_205.1_17119.t1	3E-06	Homo sapiens	E3 ubiquitin-protein ligase TRIM33	NO	YES	PF00643:B-box-type zinc finger	149	6.9	c
pfu_aug2.0_275.1_17228.t1	3E-44	Pinctada maxima	Uncharacterized shell protein 1	NO	YES	PF01359:Transposase, type 1	31	9.2	c
pfu_aug2.0_2443.1_12165.t1	3E-67	Paralichthys olivaceus	Poly(U)-specific endoribonuclease	YES	YES	PF01033:Somatomedin B domain PF09412:Endoribonuclease XendoU	37	5.1	c
pfu_aug2.0_3932.1_09248.t1	4E-15	Pinctada margaritifera	Protein PIF	YES	YES	PF00092:von Willebrand factor, type A	68	8.2	c
pfu_aug2.0_929.1_31288.t1	1E-21	Pinctada margaritifera	Protein PIF	NO	NO	-	35	8	c
pfu_aug2.0_715.1_17768.t1	0	Pinctada fucata	Protein PIF	YES	YES	PF00092:von Willebrand factor, type A	108	4.9	c
pfu_aug2.0_747.1_24368.t1	4E-10	Pinctada fucata	Protein PIF	NO	YES	-	366	5.2	c
pfu_aug2.0_269.1_30539.t1	2E-91	Locusta migratoria	Apolipophorins	NO	YES	PF00094:von Willebrand factor, type D domain PF09172:Vitellinogen, open beta-sheet PF01347:Lipid transport protein, N-terminal	441	8.8	c
pfu_cdna2.0_089203	8E-44	Mus musculus	Collagen alpha-5(VI) chain	YES	YES	PF01607:Chitin binding domain PF00092:von Willebrand factor, type A	262	7.4	b
pfu_aug2.0_2218.1_28718.t1	2E-13	Gallus gallus	Collagen alpha-1(I) chain	NO	YES	PF01391:Collagen triple helix repeat	236	9.9	c
pfu_aug2.0_219.1_30448.t1	3E-148	Rattus norvegicus	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	YES	NO	PF00084:Sushi/SCR/CCP domain PF00092:von Willebrand factor, type A PF01607:Chitin binding domain	462	4.9	c
pfu_aug2.0_2	-	-	-	YES	YES	PF01607:Chitin binding	202	6.5	c

97.1_23818.t1				S		domain			
pfu_aug2.0_39.1_30047.t1	-	-	-	NO	YES	PF03067:Chitin-binding, domain 3	43	7.1	c
pfu_aug2.0_747.1_24365.t1	-	-	-	NO	YES	-	94	6	c
pfu_aug2.0_210.1_00425.t1	-	-	-	NO	YES	PF03567:Sulfotransferase PF01607:Chitin binding domain	147	9.9	c
pfu_aug2.0_7063.1_12916.t1	1E-39	Mytilus californianus	Shell matrix protein (Fragment)	NO	YES	PF13385	74	6.3	c
pfu_aug2.0_53.1_10184.t1	3E-47	Mytilus californianus	Shell matrix protein (Fragment)	NO	YES	PF13385	92	5	c
pfu_aug2.0_1919.1_31963.t1	2E-09	Danio rerio	Caprin-2	YES	NO	PF00386:C1q domain	18	9.2	c
pfu_aug2.0_470.1_00785.t1	4E-08	Danio rerio	Caprin-2	YES	NO	PF00386:C1q domain	18	9.2	c
pfu_aug2.0_2116.1_21941.t1	1E-93	Pinctada maxima	EGF-like domain-containing protein 1 (Fragment)	YES	YES	-	30	6.1	c
pfu_aug2.0_2116.1_21942.t1	5E-176	Pinctada margaritifera	EGF-like domain containing protein 2	YES	YES	-	41	8.4	c
pfu_aug2.0_2116.1_21943.t1	1E-108	Pinctada maxima	EGF-like domain-containing protein 2	NO	YES	-	31	5.8	c
pfu_aug2.0_1222.1_08259.t1	4E-09	Dipetalogaster maximus	Serine protease inhibitor dipetalogastin (Fragment)	NO	NO	PF07648:Kazal domain	19	8	c
pfu_aug2.0_283.1_10556.t1	4E-18	Dipetalogaster maximus	Serine protease inhibitor dipetalogastin (Fragment)	YES	YES	PF07648:Kazal domain PF00050:Kazal domain	28	8.9	c
pfu_aug2.0_283.1_10558.t1	7E-16	Dipetalogaster maximus	Serine protease inhibitor dipetalogastin (Fragment)	YES	YES	PF00050:Kazal domain	26	8.7	c
pfu_aug2.0_2	1E-18	Dipetalo	Serine protease inhibitor	YE	NO	PF00050:Kazal domain	22	8.5	c

83.1_10562.t1		gaster maximus	dipetalogastin (Fragment)	S					
pfu_aug2.0_283.1_10563.t1	2E-24	Dipetalogaster maximus	Serine protease inhibitor dipetalogastin (Fragment)	YES	NO	PF07648:Kazal domain PF00050:Kazal domain	27	8.5	c
pfu_aug2.0_283.1_10555.t1	3E-11	Dipetalogaster maximus	Serine protease inhibitor dipetalogastin (Fragment)	YES	NO	PF00050:Kazal domain	19	8.5	c
pfu_aug2.0_283.1_10559.t1	2E-12	Melithaea caledonica	Four-domain proteases inhibitor	YES	NO	PF07648:Kazal domain PF00050:Kazal domain	18	8.3	c
pfu_aug2.0_283.1_10560.t1	2E-13	Dipetalogaster maximus	Serine protease inhibitor dipetalogastin (Fragment)	YES	YES	PF07648:Kazal domain PF00050:Kazal domain	18	8.2	c
pfu_aug2.0_3578.1_29138.t1	2E-176	Pinctada maxima	EGF-like domain-containing protein 1 (Fragment)	NO	YES	-	36	5.8	c
pfu_aug2.0_429.1_30751.t1	1E-81	Pinctada margaritifera	Fibronectin type III domain-containing protein 1	NO	YES	-	83	9.4	c
pfu_aug2.0_429.1_30752.t1	0	Pinctada margaritifera	Fibronectin type III domain-containing protein 1	YES	YES	PF00041:Fibronectin type III	87	5.8	c
pfu_aug2.0_429.1_30750.t1	0	Pinctada margaritifera	Fibronectin type III domain-containing protein 2	NO	NO	PF00041:Fibronectin type III	59	5.7	c
pfu_aug2.0_838.1_27830.t1	1E-69	Pinctada margaritifera	EGF-like domain containing protein 2	YES	YES	-	44	8.5	c
pfu_aug2.0_1638.1_28429.t1	2E-49	Drosophila melanogaster	Papilin	NO	NO	PF00014:Pancreatic trypsin inhibitor Kunitz domain	36	9.4	c
pfu_aug2.0_1638.1_28435.t1	2E-35	Pinctada margaritifera	BPTI/Kunitz domain-containing protein 2	NO	NO	PF00014:Pancreatic trypsin inhibitor Kunitz domain	15	8	c
pfu_aug2.0_5814.1_16145.t1	6E-37	Pinctada margaritifera	BPTI/Kunitz domain-containing protein 1	YES	YES	PF00014:Pancreatic trypsin inhibitor Kunitz domain	16	9.3	c

pfu_aug2.0_729.1_31106.t1	2E-29	Bitis gabonica	Kunitz-type serine protease inhibitor bitisilin-3 (Fragment)	NO	YES	PF00014:Pancreatic trypsin inhibitor Kunitz domain	166	9.6	c
pfu_aug2.0_1101.1_04823.t1	1E-104	Pinctada margaritifera	BPTI/Kunitz domain-containing protein 5	YES	YES	PF00014:Pancreatic trypsin inhibitor Kunitz domain	21	9.9	c
pfu_aug2.0_1101.1_04825.t1	5E-100	Pinctada margaritifera	BPTI/Kunitz domain-containing protein 3	YES	NO	PF00014:Pancreatic trypsin inhibitor Kunitz domain	18	9.1	c
pfu_aug2.0_2907.1_25577.t1	7E-61	Pinctada margaritifera	BPTI/Kunitz domain-containing protein 4	YES	YES	PF00014:Pancreatic trypsin inhibitor Kunitz domain PF02822:Antistasin-like domain	23	8.9	c
pfu_aug2.0_2907.1_25578.t1	4E-46	Pinctada margaritifera	BPTI/Kunitz domain-containing protein 4	NO	YES	PF02822:Antistasin-like domain PF00014:Pancreatic trypsin inhibitor Kunitz domain	15	9.2	c
pfu_aug2.0_1101.1_04821.t1	1E-26	Sabellastarte magnifica	Carboxypeptidase inhibitor SmCI	NO	YES	PF00014:Pancreatic trypsin inhibitor Kunitz domain	50	9.6	c
pfu_aug2.0_1101.1_04822.t1	1E-35	Pinctada maxima	BPTI/Kunitz domain-containing protein 2	YES	YES	PF00125:Histone H2A/H2B/H3 PF00014:Pancreatic trypsin inhibitor Kunitz domain	57	9.9	c
pfu_aug2.0_6201.1_06198.t1	9E-79	Pinctada maxima	BPTI/Kunitz domain-containing protein 2	NO	YES	PF00014:Pancreatic trypsin inhibitor Kunitz domain	32	8.4	c
pfu_aug2.0_701.1_04487.t2	2E-21	Mus musculus	Peptidase inhibitor 16	NO	YES	PF00188:CAP domain PF01549:ShKT domain	209	9.6	c
pfu_aug2.0_3228.1_29058.t1	3E-40	Pinctada margaritifera	NTR domain-containing protein	YES	NO	PF00965:Protease inhibitor I35 (TIMP)	17	9	c
pfu_aug2.0_944.1_14673.t1	5E-23	Pinctada margaritifera	NTR domain-containing protein	NO	YES	PF13855:Leucine-rich repeat PF00057:Low-density lipoprotein (LDL) receptor class A repeat	145	5.4	c
pfu_aug2.0_586.1_20950.t1	-	-	-	NO	YES	PF00965:Protease inhibitor I35 (TIMP)	33	8.1	c
pfu_aug2.0_1661.1_05172.t1	-	-	-	NO	YES	PF00965:Protease inhibitor I35 (TIMP)	76	9.1	c

pfu_aug2.0_2607.1_25472.t2	4E-34	Pinctada margaritifera	Uncharacterized shell protein 1	YES	YES	PF00965:Protease inhibitor I35 (TIMP)	35	9.2	c
pfu_aug2.0_5014.1_16058.t1	-	-	-	YES	YES	PF00965:Protease inhibitor I35 (TIMP)	74	8.8	c
pfu_aug2.0_1068.1_28012.t1	7E-112	Homo sapiens	Zinc finger CCCH domain-containing protein 15	NO	YES	PF00642:Zinc finger, CCCH-type PF16543	36	8.8	c
pfu_aug2.0_1940.1_01913.t1	8E-114	Homo sapiens	Zinc finger CCCH domain-containing protein 15	NO	YES	PF16543 PF00642:Zinc finger, CCCH-type	41	8.6	c
pfu_aug2.0_357.1_23926.t1	1E-07	Rhodobacter blasticus	ATP synthase subunits region ORF 7	YES	NO	PF09458:H-type lectin domain	25	7	c
pfu_aug2.0_495.1_17489.t1	2E-126	Homo sapiens	Adhesion G protein-coupled receptor L3	YES	YES	PF01825:GPS motif PF00059:C-type lectin PF16489 PF00002:GPCR, family 2, secretin-like	104	8.5	c
pfu_aug2.0_813.1_11208.t1	1E-14	Danio rerio	Natterin-like protein	NO	YES	PF01419:Jacalin-like lectin domain PF03318:Clostridium epsilon toxin ETX/Bacillus mosquitocidal toxin MTX2	45	8.4	c
pfu_aug2.0_5024.1_16059.t1	2E-18	Danio rerio	Natterin-like protein	NO	YES	-	39	5.7	c
pfu_aug2.0_663.1_11059.t1	4E-13	Pinctada maxima	Shematrin-like protein 3	NO	YES	-	45	9.7	c
pfu_cdna2.0_003257	2E-91	Pinctada fucata	N16.3 matrix protein	YES	YES	-	16	5.7	c
pfu_aug2.0_1259.1_31566.t1	-	-	-	NO	YES	PF00024:PAN/Apple domain	26	5.3	c
pfu_aug2.0_1910.1_01900.t1	8E-52	Mus musculus	Bromodomain-containing protein 8	NO	YES	PF00439:Bromodomain	189	5.1	c
pfu_aug2.0_2899.1_32337.t1	1E-62	Mus musculus	Leucine-rich repeat and calponin homology domain-containing protein 1	NO	YES	PF13855:Leucine-rich repeat PF00307:Calponin homology domain	196	9.8	c
pfu_aug2.0_	-	-	-	YES	YES	-	43	3.8	b

2162.1_08762.t1				S					
pfu_aug2.0_126.1_20287.t1	1E-06	Dictyostelium discoideum	Midasin	NO	YES	-	192	3.9	c
pfu_aug2.0_444.1_14163.t1	-	-	-	NO	YES	-	109	3.7	c
pfu_aug2.0_618.1_27594.t1	3E-87	Pinctada maxima	Valine-rich protein	NO	YES	-	61	9.9	b
pfu_aug2.0_618.1_27595.t1	5E-65	Pinctada maxima	Mantle protein	YES	YES	-	30	10	b
pfu_aug2.0_490.1_00814.t1	2E-166	Rattus norvegicus	DnaJ homolog subfamily B member 11	YES	YES	PF00226:DnaJ domain PF01556:Chaperone DnaJ, C-terminal	41	5.7	c
pfu_aug2.0_3.1_10035.t1	-	-	-	YES	YES	-	170	9.8	c
pfu_aug2.0_34.1_13448.t1	-	-	-	YES	YES	-	20	8.2	c
pfu_aug2.0_874.1_14622.t1	8E-133	Pinctada margaritifera	Uncharacterized shell protein 2	NO	YES	-	23	9.7	c
pfu_aug2.0_160.1_00336.t1	5E-06	Lottia gigantea	Uncharacterized shell protein 26 (Fragment)	YES	NO	-	37	8.5	c
pfu_aug2.0_746.1_21112.t1	7E-62	Pinctada fucata		YES	YES	-	16	4.8	c
pfu_aug2.0_9582.1_09723.t1	5E-45	Pinctada fucata		NO	YES	-	14	4.5	c
pfu_aug2.0_240.1_00477.t1	5E-09	Aplysia californica	Buccalin	NO	NO	-	33	11	c
pfu_aug2.0_1116.1_21407.t1	2E-82	Pinctada maxima	Methionine-rich nacre protein	YES	YES	-	50	10	c
pfu_aug2.0_1163.1_11481.t1	3E-47	Pinctada maxima	Uncharacterized protein 3	YES	YES	-	10	11	c
pfu_aug2.0_1	1E-44	Pinctada	Serine, glycine and	YE	YES	-	33	9.5	c

88.1_27003.t1		maxima	glutamine-rich protein	S					
pfu_aug2.0_3096.1_22282.t1	7E-06	Mytilus californianus	Mytilin-3	NO	YES	-	15	10	c
pfu_aug2.0_944.1_14672.t1	1E-34	Pinctada margaritifera	Uncharacterized shell protein 6	YES	YES	-	17	9.2	c
pfu_aug2.0_1423.1_11654.t1	-	-	-	NO	YES	-	20	10	c
pfu_aug2.0_146.1_20326.t1	-	-	-	NO	YES	-	15	9.6	c
pfu_aug2.0_1800.1_01846.t1	-	-	-	NO	NO	-	15	5.2	c
pfu_aug2.0_1811.1_05249.t1	-	-	-	YES	YES	-	69	11	c
pfu_aug2.0_1942.1_08662.t1	-	-	-	YES	NO	-	50	5.1	c
pfu_aug2.0_3744.1_15850.t1	-	-	-	YES	NO	-	12	8	c
pfu_aug2.0_476.1_20812.t1	-	-	-	YES	YES	-	145	4.7	c
pfu_aug2.0_728.1_27714.t1	-	-	-	YES	YES	-	16	9.5	c
pfu_aug2.0_853.1_11239.t1	-	-	-	YES	YES	-	197	11	c
pfu_aug2.0_862.1_07957.t1	-	-	-	NO	NO	-	18	7.5	c
pfu_aug2.0_862.1_07958.t1	-	-	-	YES	YES	-	9	8.1	c
pfu_aug2.0_113.1_10298.t1	-	-	-	YES	NO	-	23	9	c

pfu_aug2.0_1504.1_15072.t1	-	-	-	NO	YES	-	76	9.5	c
pfu_aug2.0_1549.1_31754.t1	-	-	-	YES	YES	-	25	5.3	c
pfu_aug2.0_1549.1_31755.t1	-	-	-	NO	YES	-	25	5.4	c
pfu_aug2.0_1685.1_18477.t1	-	-	-	YES	NO	-	16	9.1	c
pfu_aug2.0_2125.1_18698.t1	-	-	-	YES	YES	-	57	8.9	c
pfu_aug2.0_2205.1_18728.t1	-	-	-	NO	YES	-	58	4.7	c
pfu_aug2.0_2212.1_08780.t1	-	-	-	YES	NO	-	17	11	c
pfu_aug2.0_2356.1_22037.t1	-	-	-	YES	NO	-	17	11	c
pfu_aug2.0_2991.1_05713.t1	-	-	-	YES	YES	-	105	8.3	c
pfu_aug2.0_3371.1_05817.t1	-	-	-	YES	YES	-	87	9.7	c
pfu_aug2.0_52.1_06823.t1	-	-	-	YES	YES	-	57	9.3	c
pfu_aug2.0_53.1_10183.t1	-	-	-	YES	YES	-	60	9.8	c
pfu_aug2.0_55.1_16813.t1	-	-	-	YES	NO	-	32	11	c
pfu_aug2.0_583.1_10957.t1	-	-	-	NO	YES	-	72	8.7	c
pfu_aug2.0_87.1_23420.t1	-	-	-	YES	YES	-	10	5.2	c
pfu_aug2.0_874.1_14621.t1	-	-	-	YES	YES	-	14	9.2	c

pfu_aug2.0_9	-	-	-	NO	YES	-	27	9.1	c
918.1_29684.									
t1									
pfu_aug2.0_9	-	-	-	NO	YES	-	39	8.4	c
931.1_06417.									
t1									
pfu_aug2.0_1	-	-	-	YE	YES	-	180	9.8	c
163.1_11480.				S					
t1									
pfu_aug2.0_1	-	-	-	YE	NO	-	49	5.2	c
369.1_31644.				S					
t1									
pfu_aug2.0_1	-	-	-	YE	NO	-	49	8.4	c
55.1_17024.t				S					
1									
pfu_aug2.0_2	-	-	-	NO	NO	-	52	9.3	c
27.1_23707.t									
1									
pfu_aug2.0_2	-	-	-	NO	NO	-	21	5.3	c
307.1_25376.									
t1									
pfu_aug2.0_3	-	-	-	YE	YES	-	36	6.5	c
823.1_12525.				S					
t1									
pfu_aug2.0_4	-	-	-	NO	YES	-	132	9.5	c
173.1_12602.									
t1									
pfu_aug2.0_7	-	-	-	YE	NO	-	23	9.4	c
079.1_32931.				S					
t1									
pfu_aug2.0_7	-	-	-	YE	NO	-	46	5.7	c
47.1_24369.t				S					
1									
pfu_aug2.0_8	-	-	-	NO	YES	-	110	9.8	c
84.1_14629.t									
1									
pfu_aug2.0_9	-	-	-	NO	NO	-	28	6.9	c
32.1_08017.t									
1									
pfu_aug2.0_9	-	-	-	YE	NO	-	23	8.7	c
32.1_08018.t				S					
1									
pfu_aug2.0_9	-	-	-	NO	NO	-	43	9.3	c
32.1_08019.t									

1									
pfu_aug2.0_1623.1_11769.t1	-	-	-	NO	YES	PF16026	49	9.4	c
pfu_aug2.0_222.1_07177.t1	4E-158	Strongylocentrotus purpuratus	33 kDa inner dynein arm light chain, axonemal	NO	NO	PF10211:Axonemal dynein light chain	29	9.2	c
pfu_aug2.0_322.1_07359.t1	9E-125	Chlamydomonas reinhardtii	Dynein beta chain, flagellar outer arm	YES	YES	PF03028:Dynein heavy chain domain PF12774 PF12781 PF12775 PF08385:Dynein heavy chain, domain-1 PF12777:Dynein heavy chain, coiled coil stalk PF08393:Dynein heavy chain, domain-2 PF13306:Leucine rich repeat 5	626	5.6	c
pfu_aug2.0_2470.1_02142.t1	0	Mus musculus	Dynein heavy chain 8, axonemal	NO	NO	PF08385:Dynein heavy chain, domain-1 PF08393:Dynein heavy chain, domain-2	205	5.8	c
pfu_aug2.0_300.1_00563.t1	0	Homo sapiens	Dynein heavy chain 10, axonemal	NO	YES	PF12781 PF08393:Dynein heavy chain, domain-2 PF07728:ATPase, dynein-related, AAA domain PF12775 PF12774 PF12780:Dynein heavy chain, P-loop containing D4 domain PF03028:Dynein heavy chain domain PF12777:Dynein heavy chain, coiled coil stalk	294	5.7	c
pfu_aug2.0_1891.1_05300.t1	4E-08	Rhodobacter blasticus	ATP synthase subunits region ORF 7	NO	NO	PF09458:H-type lectin domain	23	5.8	c
pfu_aug2.0_144.1_13676.t1	1E-100	Homo sapiens	CD109 antigen	YES	YES	PF07678:Alpha-macroglobulin complement component PF10569:Alpha-2-macroglobulin, thiol-ester bond-forming	295	9.1	c

						PF01835:Alpha-2-macroglobulin, N-terminal PF00207:Alpha-2-macroglobulin PF07677:Alpha-macroglobulin, receptor-binding			
pfu_aug2.0_1495.1_18365.t1	3E-15	Homo sapiens	Ribosome-binding protein 1	NO	YES	-	477	9.6	c
pfu_aug2.0_2922.1_09016.t1	-	-	-	NO	NO	PF00241:Actin-depolymerising factor homology domain	14	4.8	c

Bold text: common SMPs for larva and adult; '-' means not applicable; SP: signal peptide; RLCD: repeats of low complexity domains.

Chapter 3 Phylogenetic comparisons revealed mosaic histories of deployment among larval and adult shell matrix proteins in pteriomorph bivalves

Key words: carbonic anhydrase, chitobiase, molecular evolution, molluscs, shell matrix protein (SMP), VWA and chitin-binding domain-containing protein

3.1 Introduction

The overwhelming appearance of mineralized skeletons, including calcium phosphate, calcium carbonate and silica, opens the grand history of the metazoan taxa at the dawn of the Cambrian (Knoll and Carroll 1999; Morris 1998; Shubin and Marshall 2000). Among the metazoan biominerals, calcium carbonate skeletons of the Mollusca provide exceptional resources for studying the processes of biomineralization by exhibiting tremendously abundant morphologies (Lowenstam and Weiner 1989), as well as a huge diversity of special microstructures identifiable and characteristic for each species, publications on the classification of which are available (Boggild 1930; Carter and Clark 1985; Chateigner, et al. 2000). Despite this complexity, the adult molluscan shells are secreted by an evolutionarily homologous organ known as the mantle (Jolly, et al. 2004; Sudo, et al. 1997), and different aspects of the shell formation processes are controlled by the organic molecules, collectively known as the shell matrix, which is a quantitatively minor constituent of the shell (0.1%~5% w/w according to different species and microstructures) (Mann 1988). To date, a large number of shell matrix proteins (SMPs) have been found in molluscs, from the acidic ones, MSI31 (Sudo, et al. 1997), MSP-1 (Sarashina and Endo 2001), Aspein (Tsukamoto, et al. 2004), Prismalin-14 (Suzuki, et al. 2004) and Asp-rich protein families (Gotliv, et al. 2005), to the more basic ones such as N66 (Kono, et al. 2000), Shematrins (Yano, et al. 2006) and KRMPs (Zhang, et al. 2006). As the number of identified SMPs increases from different molluscan species, lineage-specific repertoires of SMPs (Marie, et al. 2013; Marie, et al. 2011a),

respective molecular toolkits that control the formation of prism and nacre within one species (Marie, et al. 2012; Takeuchi and Endo 2006), as well as the dual gene repertoires in charge of larval and the adult shell formation (Zhao, et al. 2018) have been noticed. Meanwhile, only a few homologous SMPs have been reported to be shared between different species, and the rare examples that are shared at the level between gastropods and bivalves include Carbonic anhydrase, Pif/BMSP-like protein, Perlucin and Perlwapin (Marie, et al. 2013; Marie, et al. 2011a; Zhao, et al. 2018).

In order to explain the evolutionary relationships among the various SMP repertories and how the components of them were recruited by the shell, two extreme scenarios have been given (Marin, et al. 2007). On one hand, the "ancient heritage" scenario was generally favored by the fossil record of skeletal elements suddenly shown up in Tommotian rocks, and the fossil record suggests that the main classes of molluscs had representatives in the Cambrian, including polyplacophores, monoplacophores, cephalopods, gastropods, bivalves (Lecointre and Le Guyader 2001; Runnegar 1996), as well as the complex shell microstructures of early Cambrian molluscs (Feng and Sun 2003; Kouchinsky 2000). A single and ancient origin of common SMPs is suggested by the fast exploitation of all textural combinations and most of the design possibilities for building their skeleton (Thomas, et al. 2000) resulted from recruitment of Precambrian functions, which were not related to mineralization (Marin, et al. 2003). In addition, the primary structure of SMPs may also support the speculation of the antiquity of some SMPs. For instance, the high similarities between the functional domains of the molluscan carbonic anhydrases (CAs), including the SMPs of Nacrein and N66, and CAs of other metazoans have been noticed. Because the conversion of carbon dioxide into bicarbonate is an ancient function, and this function could be primordial in calcium carbonate biomineralization, and because carbonic anhydrase domains have been found in both bivalves and gastropods, it is hard to think of such a key function results from a recent recruitment (Marin, et al. 2007).

On the other hand, the “recent heritage and fast evolution” scenario was supported by the unique origins of different shell matrices indicated by the “independent inventions” based on the phylogenetic comparisons of homologous genes (Marin, et al. 2007). The transcriptomic data indicated that 85% of the secreted proteins of the abalone *Haliotis asinina* are unknown and only 19% of the secreted proteins of *H. asinina* are homologous to those of the patellogastropod *Lottia scutum* (Jackson, et al. 2006), suggesting that the shell is constructed from a rapidly evolving secretome (Marin, et al. 2007). This scenario was also supported by the phylogenetic tree of dermatopontins of eight gastropod species (Sarashina, et al. 2006). Dermatopontin is an ancient protein, because it is found in various metazoans including sponges and human (Fujii, et al. 1992; Neame, et al. 1989; Schütze, et al. 2001) with general functions in cell-matrix interactions and matrix assembly (Marxen and Becker 1997; Marxen, et al. 2003). However, in the two gastropod lineages, Basommatophora (pond snails) and Stylommatophora (land snails), the recruitment of dermatopontin to the shell occurred twice independently (Sarashina, et al. 2006).

Although the mollucan adult shells show complex micro-textures and different mineralogies, molluscan larval shells have microstructures similar to each other, and are composed only of aragonite (Eyster 1986; Eyster 1983; Iwata 1980; LaBarbera 1974; Weiss, et al. 2002), implying that larval shells are evolutionarily highly conserved (Taylor 1973). If we take the commonality in mineralogy and microstructures as a hallmark of “primitive” states of shells, studies of the larval SMPs could shed light on the reconstruction of the ancestral features of larval shells as well as the origin of SMPs among different lineages.

With the help of advanced next-generation sequencing and proteomic techniques, the first larval shell proteomes of two pteriomorph bivalve species, the pacific oyster, *Crassostrea gigas* and the pearl oyster, *Pinctada fucata* were published (Zhao, et al. 2018). In this study, we performed phylogenetic analyses on the SMPs which were shown to be shared by the larval and adult shells in both species, and discuss the time

points when they were recruited as SMPs.

3.2 Materials and methods

3.2.1 Data resources

The details of the shell matrix proteins (SMPs) of *Crassostrea gigas* and *Pinctada fucata*, including amino acid sequences, results of annotation, and gene expression patterns indicated by RNA-Seq can be found in our previously published report (Zhao, et al. 2018). The amino acid sequences and other details of the shell and other proteins of the other species (*Lottia gigantea*, *Mitulus galloprovincialis*, *Octopus bimaculoides*, *Homo sapiens* and *Arabidopsis thaliana*) were obtained from the public databases including InterPro protein analysis and classification (http://www.ebi.ac.uk/interpro/), Genebank (https://www.ncbi.nlm.nih.gov/genbank/), *Pinctada fucata* genome project web site (http://marinegenomics.oist.jp/pearl/viewer/info?project_id=36) (Tables 3.1, 3.2, 3.3 and 3.4).

3.2.2 Sequence analysis

Blastp searches against the UniProtKB/Swiss-Prot database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were performed on the SuperComputer facilities of International Institute of Genetics using the default settings. SMART online service (http://smart.embl-heidelberg.de) and BLAST were applied to predict the presence of functional domains and the signal peptide. Sequence alignments were performed on ClustalX program embedded in Genetyx version 6 (Genetyx, Tokyo, Japan).

3.2.3 Alignments and setting for phylogenetic trees

Multiple sequences were aligned using ClustalW embedded in the phylogenetic analysis tool MEGA 7.0 (Kumar, et al. 2016) with default settings. Alignment results were submitted to the online Gblocks server (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) to remove poorly aligned regions and divergent regions of protein alignment with default settings (stringent). Phylogenetic analysis was performed using maximum-likelihood (ML) method on MEGA 7.0. Both LG or Poisson models were applied. Reliability was

examined by bootstrap analysis based on 1000 replicates. Polychotomies were generated by collapsing the nodes with a bootstrap value lower than 50.

3.2.4 Phylogenetic analyses of CB, VWA and Laminin G domains in molluscan shell matrix proteins

According to a genome-wide survey based on the InterProScan online database as well as our proteomic work, an overwhelming expansion of VWA (IPR002035) and chitin-binding domain (IPR002557)-containing protein (VWA-CB dcp) in molluscs has been reported compared with other taxa (Zhao, et al. 2018). Pif is a representative of this class of proteins, and its functions in the formation of nacre in *Pinctada fucata* have been documented (Suzuki, et al. 2009). As molluscan SMPs, VWA-CB dcps or Pif/BMSP-like proteins were reported in *Mytilus galloprovincialis* (Suzuki, et al. 2011), *Lottia gigantea* (Mann, et al. 2012; Marie, et al. 2013) and in both the larval and adult shells of *Crassostrea. gigas* and *P. fucata* (Suzuki, et al. 2009; Zhang, et al. 2012; Zhao, et al. 2018). BMSPs (blue mussel shell proteins) contains more than one VWA domains, while other VWA-CB dcps of molluscan SMPs possess only one VWA domain. In order to reconstruct the evolutionary relationships among these shell-specific VWA-CB dcps, phylogenetic analyses were performed on 37 (Fig. 3.16) and 127 (Fig. 3.17) alignable amino acid residues of CB and VWA domains, respectively, through maximum likelihood method with LG (Figs. 3.1 A and B; 3.2 A and B) and Poisson (Figs. 3.7 A and B; 3.8 A and B) models. Domains of the VWA-CB dcp of the tunicate *Ciona intestinalis* were used as the outgroup. Predicted domain structures of the proteins and the place where the SMP were identified are illustrated in Fig. 3.1 C and Fig. 3.2 C. Phylogenetic analyses were also performed on the Laminin G domain, which was identified by SMART and BLAST domain searches, for 57 amino acid residues (Fig. 3.19 A) and on the concatenated CB and Laminin G domains for 86 amino acid residues (Fig. 3.19 B), respectively, with LG (Figs. 3.3 A, B, C and D) and Poisson (Figs. 3.10 A, B, C and D) models.

3.2.5 Phylogenetic analyses of carbonic anhydrase (CA) in molluscs

The CA domain is another common domain identified in both larval and adult shell proteomes of *C. gigas* and *P. fucata* in our previous study (Zhao, et al. 2018). CA genes are highly expanded in molluscs compared with other protostomian animals (Zhao, et al. 2018), though reports of shell-specific CAs are still poor and none has been reported from cephalopods. In order to estimate the origin of shell-specific CAs in molluscs, phylogenetic analyses were performed on the CA domains of the four molluscan species, *Crassostrea gigas*, *Pinctada fucata*, *Lottia gigantea* and *Octopus bimaculoides*, the genomic and/or proteomic data of which are deposited in the public databases. Trees were generated based on 82 amino acid residues (ClustalW alignment result was modified by GBlocks) (Fig. 3.20), respectively, via LG and Poisson models. A CA domain of the plant *Arabidopsis thaliana* was used as the outgroup. Phylogenetic trees were generated through LG (Figs. 3.4 A and B) and Poisson (Figs. 3.11 A and B) models, respectively. Phylogenetic analyses were also performed on the 55 unambiguously aligned amino acid residues of CAs of molluscan shells and those of human (Fig. 3.21) via LG (Figs. 3.12 A and B) and Poisson (Figs. 3.12 C and D) models in order to infer the relationships among them.

3.2.6 Phylogenetic analyses on chitobiases

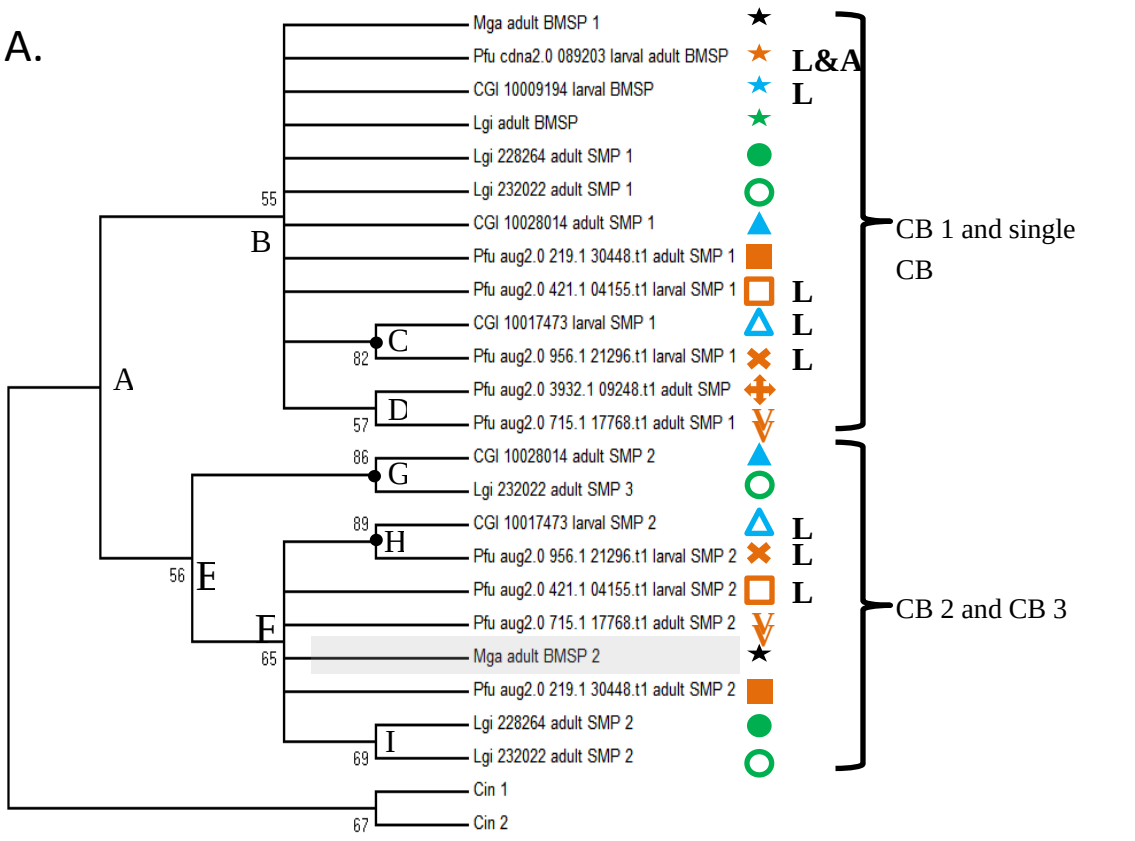
A combined domain search with SMART and BLAST (E-value: 10^{-5}) was performed on CGI_10007856, Pfu_aug2.0_6.1_20027.t1 and two adult shell-chitobiase CGI_10007857 and Pfu aug2.0_6.1_20028.t1 (Zhang, et al. 2012; Zhao, et al. 2018), revealing that they all possess conserved sequences of the four domains, CHB_HEX domain (IPR004866), Glyco_hydro_20b domain (IPR015882), Glyco_hydro_20 domain (IPR015883) and CHB_HEX_C domain (IPR004867) (Fig. 3.5 A). Phylogenetic analyses on the shell-chitobiase (Figs. 3.6 A and B, 3.15 A and B) were performed using the genes possessing the all four domains via the combined domain searches in the four molluscan species, *C. gigas*, *P. fucata*, *L. gigantea* and *O. bimaculoides* (Table 3.3). A sequence of the brachiopod *Lingula anatina*, Lan_1530 (Luo, et al. 2015), containing the same four domains was taken as the outgroup. Trees were generated based on 523 aa (ClustalW alignment of the concatenated sequences

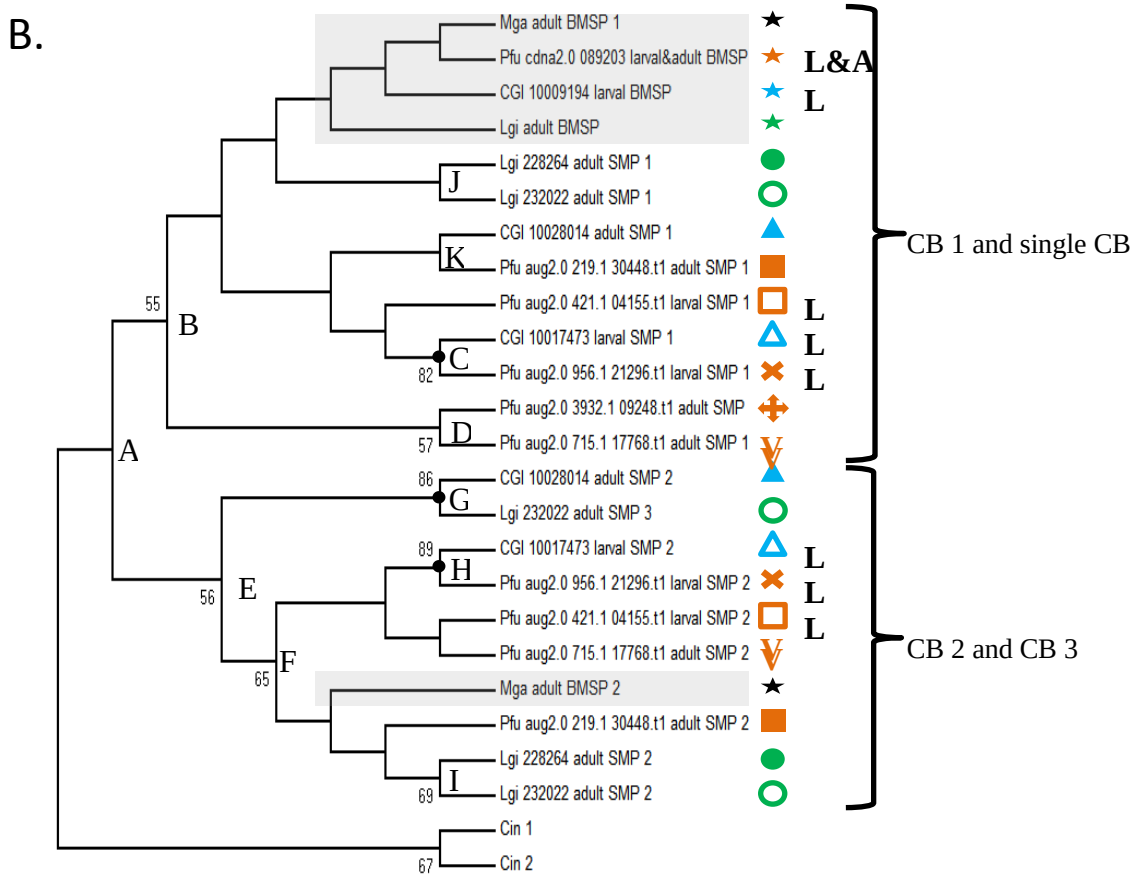
of the four domains modified by GBlocks) (Fig. 3.22) as well as on 106 aa (gaps were included in the former alignment) (Fig. 3.23), respectively, via LG (Figs. 3.6 A and B; Figs. 3.15 A and B) and Poisson (Figs. 3.14 A and B; 3.15 C and D) models. Also, the BLAST searches (E-value cut-off: 10^{-5}) indicated that the family 18 chitinases and the family 20 chitobiases (Table 3.4) are not homologous in *C. gigas* and *P. fucata*, therefore, phylogenetic analyses were performed only on the family 20 chitobiases-like proteins of both species.

3.3 Results and Discussion

3.3.1 Evolutionary history of molluscan VWA-CB dcps

In the phylogenetic trees of CB domains (Figs. 3.1 A and B; 3.7 A and B), CB 1 and the single CB domains form a monophyletic group separated from that of CB 2 and CB 3, indicating that all extant shell-specific VWA-CB dcps are originated from an ancestor with two CB domains at node A (Figs. 3.1 A and B; 3.7 A and B). The clustering of a single group comprised of all BMSPs of both larval and adult shells of the bivalve and the gastropod (Figs. 3.1 B and 3.8 B) suggests that a possible duplication event of the single VWA domain-containing VWA-CB dcp in the common ancestor of the bivalves and the gastropods gave rise to the first BMSP. BMSPs are nested within in the group comprised of the CB 1 or single CB domains of the other VWA-dcps (Figs. 3.1 A and B; 3.7 A and B), indicating that CB 1 or single CB domain of the first BMSP and those of other VWA-dcps are orthologues inherited from the CB 1 domain of their common ancestor at node A. The BMSP of *M. galloprovincialis* possesses two CB domains, and the second CB domain was clustered with the monophyletic group of CB 2 domains of the other VWA-CB dcps. This result indicates that BMSP originally also had two CB domains like other VWA-CB dcps in the ancestral protein at node A. It is inferred that the CB 2 domain has been lost in the BMSPs of *C. gigas*, *P. fucata* and *L. gigantea*, rather than gained only in *M. galloprovincialis.*


A.
Mga adult BMSP 1
Pfu cdna2.0 089203 larval adult BMSP
CGI 10009194 larval BMSP
Lgi adult BMSP
Lgi 228264 adult SMP 1
Lgi 232022 adult SMP 1
CGI 10028014 adult SMP 1
Pfu aug2.0 219.1 30448.t1 adult SMP 1
Pfu aug2.0 421.1 04155.t1 larval SMP 1
CGI 10017473 larval SMP 1
Pfu aug2.0 956.1 21296.t1 larval SMP 1
Pfu aug2.0 3932.1 09248.t1 adult SMP
Pfu aug2.0 715.1 17768.t1 adult SMP 1
CGI 10028014 adult SMP 2
Lgi 232022 adult SMP 3
CGI 10017473 larval SMP 2
Pfu aug2.0 956.1 21296.t1 larval SMP 2
Pfu aug2.0 421.1 04155.t1 larval SMP 2
Pfu aug2.0 715.1 17768.t1 adult SMP 2
Mga adult BMSP 2
Pfu aug2.0 219.1 30448.t1 adult SMP 2
Lgi 228264 adult SMP 2
Lgi 232022 adult SMP 2
Cin 1
Cin 2
L&A
L
L
L
L
L
L
L
L
CB 1 and single CB
CB 2 and CB 3
A
B
C
D
E
F
G
H
I
55
82
57
86
89
56
65
69
67


B.
Mga adult BMSP 1
Pfu cdna2.0 089203 larval&adult BMSP
CGI 10009194 larval BMSP
Lgi adult BMSP
Lgi 228264 adult SMP 1
Lgi 232022 adult SMP 1
CGI 10028014 adult SMP 1
Pfu aug2.0 219.1 30448.t1 adult SMP 1
Pfu aug2.0 421.1 04155.t1 larval SMP 1
CGI 10017473 larval SMP 1
Pfu aug2.0 956.1 21296.t1 larval SMP 1
Pfu aug2.0 3932.1 09248.t1 adult SMP
Pfu aug2.0 715.1 17768.t1 adult SMP 1
CGI 10028014 adult SMP 2
Lgi 232022 adult SMP 3
CGI 10017473 larval SMP 2
Pfu aug2.0 956.1 21296.t1 larval SMP 2
Pfu aug2.0 421.1 04155.t1 larval SMP 2
Pfu aug2.0 715.1 17768.t1 adult SMP 2
Mga adult BMSP 2
Pfu aug2.0 219.1 30448.t1 adult SMP 2
Lgi 228264 adult SMP 2
Lgi 232022 adult SMP 2
Cin 1
Cin 2
L&A
L
L
L
L
L
L
L
L
CB 1 and single CB
CB 2 and CB 3
A
B
C
D
E
F
G
H
I
J
K
55
82
57
86
89
56
65
69
67

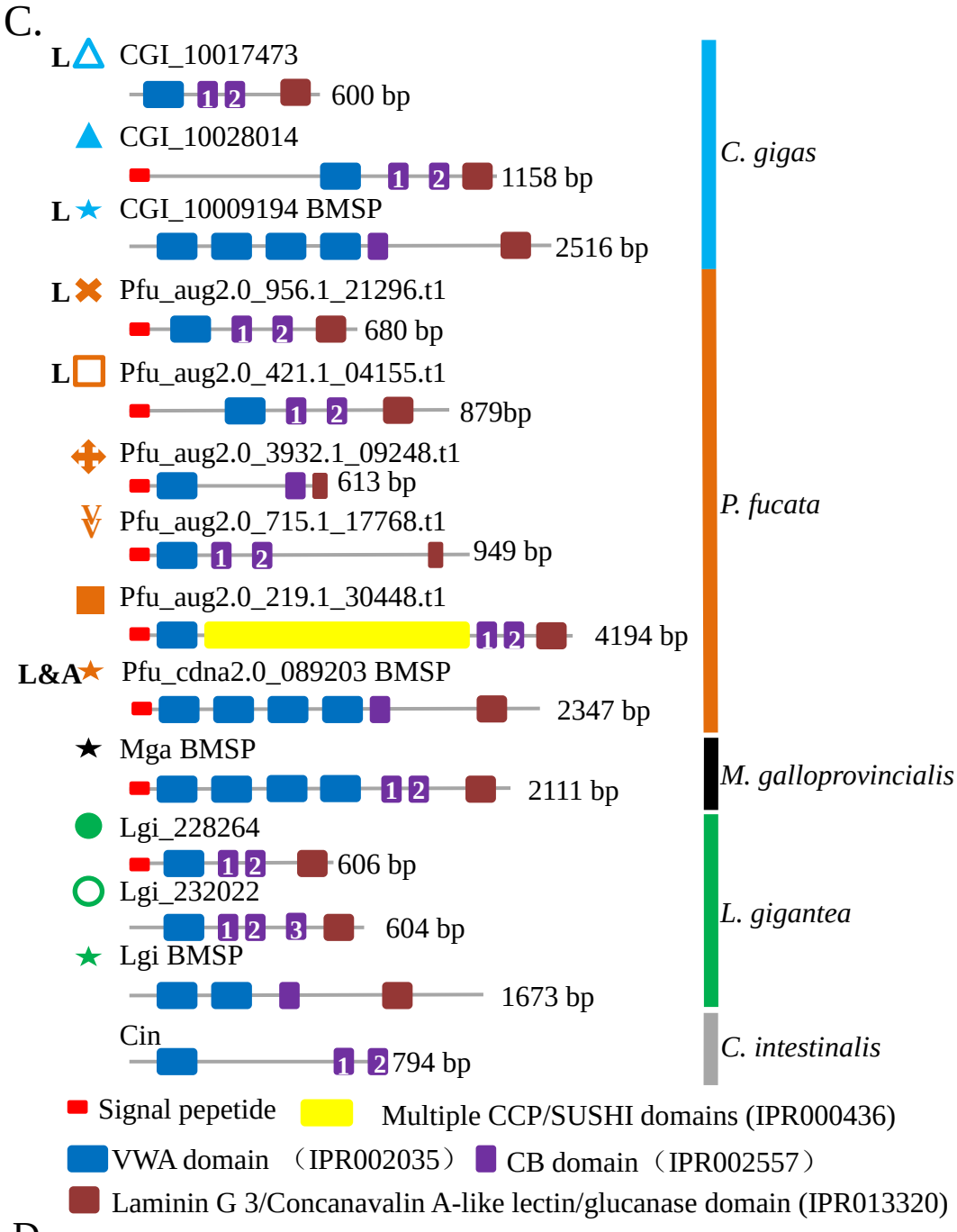

C.
L
CGI_10017473
600 bp
CGI_10028014
1158 bp
L
CGI_10009194 BMSP
2516 bp
L
Pfu_aug2.0_956.1_21296.t1
680 bp
L
Pfu_aug2.0_421.1_04155.t1
879bp
Pfu_aug2.0_3932.1_09248.t1
613 bp
Pfu_aug2.0_715.1_17768.t1
949 bp
Pfu_aug2.0_219.1_30448.t1
4194 bp
L&A
Pfu_cdna2.0_089203 BMSP
2347 bp
Mga BMSP
2111 bp
Lgi_228264
606 bp
Lgi_232022
604 bp
Lgi BMSP
1673 bp
Cin
794 bp
C. gigas
P. fucata
M. galloprovincialis
L. gigantea
C. intestinalis
Signal pepetide
Multiple CCP/SUSHI domains (IPR000436)
VWA domain （IPR002035）
CB domain（IPR002557）
Laminin G 3/Concanavalin A-like lectin/glucanase domain (IPR013320)
D.
Other single VWA dcps
Lgi_232022
Duplication
Lgi_Dp
Lgi_228264

Fig. 3.1. Phylogenetic analyses of CB domains of VWA-CB dcps of molluscan shells through LG model on 37 amino acid residues. A, Polychotomes are generated if the bootstrap value of a node is lower than 50. Nodes A-I were indicated. B, All dichotomes are preserved. Bootstrap values are shown if over 50, and marked with black dots if over 80. Nodes A-J were indicated. C, Schematics of domain structure of VWA-CB dcps. D, History of CB domain duplication in VWA-CB dcps of *Lottia gigantea*. In the tree region, same proteins were marked by same marks. BMSPs were marked by stars and grep squares. A SMP was marked by 'L' if it is a larval SMP. Pfu_cdna2.0_089203 was identified from both larval and adult shells of *P. fucata*. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Cin, *Ciona intestinalis*.

The group of the larval VWA-CB dcps of *C. gigas* and *P. fucata* at node C (Figs. 3.1 A and B) and at node D (Figs. 3.7 A and B) indicate a recruitment of the VWA-CB dcp to the larval shell of the common ancestor of *C. gigas* and *P. fucata* before the divergence of the two species. A similar situation was suggested to have occurred to the adult shell of the ancestor of the two species at the node K (Fig. 3.1 B) and the node J (Fig. 3.7 B), though both were not supported by a high bootstrap value.

Interestingly, the CB 3 of Lgi_232022 is orthologous to the CB 2 of CGI_10028014, forming a clade at node G (Figs. 3.1 A and B) and at node E (Figs. 3.7 A and B), while the CB 2 of Lgi_232022 is orthologous to the CB 2 of Lgi_228264, forming a clade at node I (Figs. 3.1 A and B) and at G (Figs. 3.7 A and B), both of which are supported by relatively high bootstrap values. The evolutionary history of these two VWA-CB dcps of *L. gigantea* can be deduced from these relationships as follows (Fig. 3.1 D): in *L. gigantea*, first, a VWA-CB dcp containing two CB domains was inherited from the common ancestor of the bivalve and the gastropod as those of other species. Then, a duplication of the second CB domain (in the notation of other single VWA dcps) produced another CB domain between the original two, which generated Lgi_232022. Next, duplication of Lgi_232022 produced another copy of this protein Lgi_Dp. Last, the third CB domain (in the notation of the Lgi_232022 of *L. gigantea*) occurred in the Lgi_Dp, finally generating Lgi_228264 (Fig. 3.1 D).

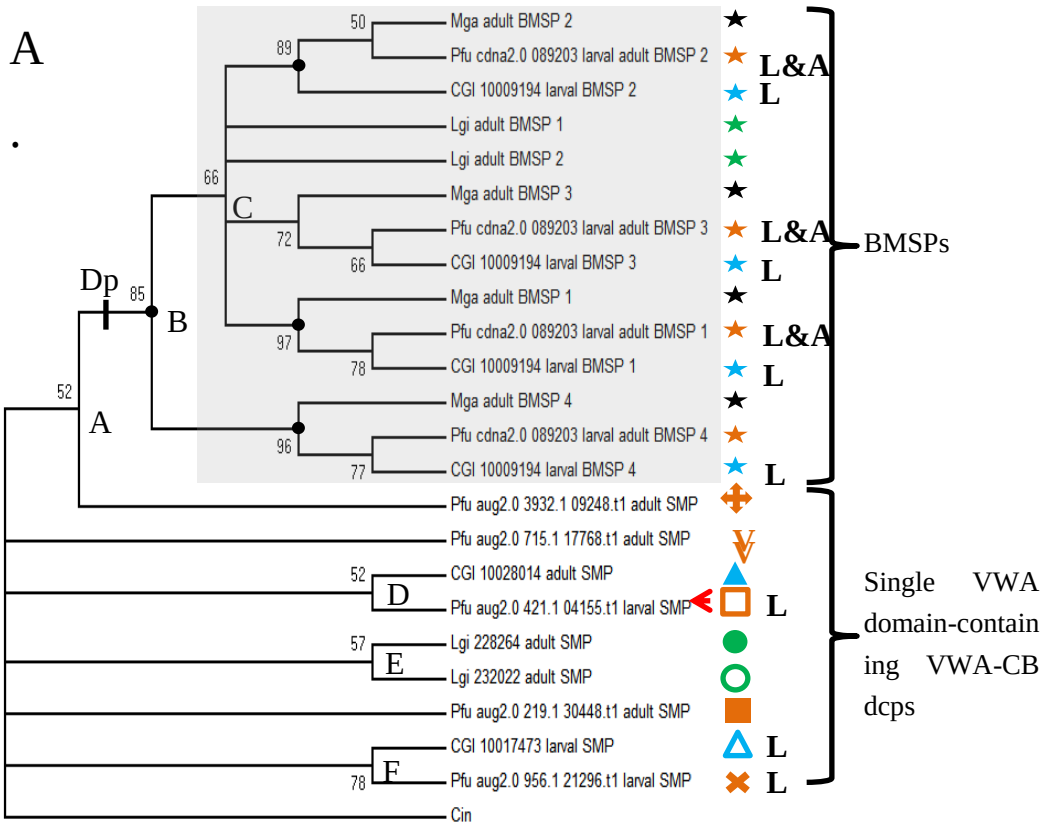
Although the low bootstrap values of some nodes suggest the possibility of other

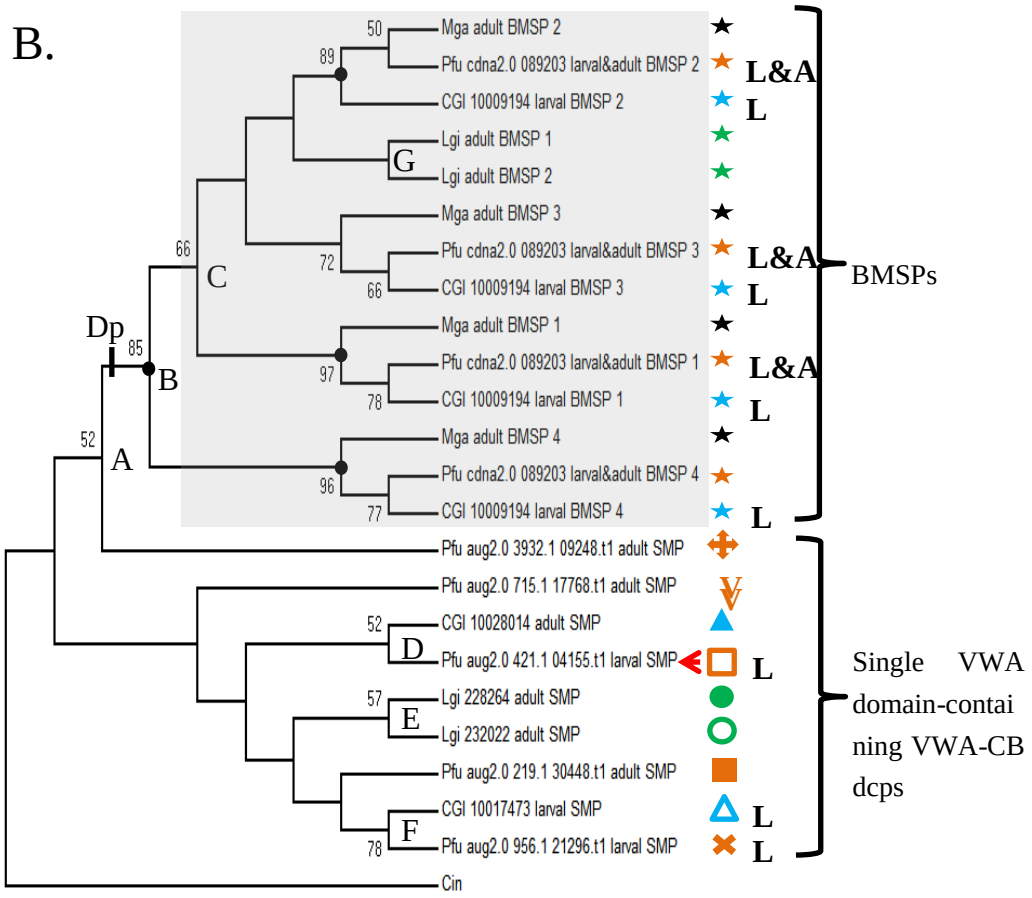
phylogenetic relationships, topologies of the trees constructed via LG model (Figs. 3.1 A and B) and Poisson model (Figs. 3.7 A and B) generally exhibited high identities.

The occurrence of multiple VWA domains distinguishes the BMSPs from other VWA-CB dcps (Figs. 3.1 C and 3.2 C). In the reconstructed phylogenetic tree (LG model) of VWA domains (Figs. 3.2 A and B; 3.8 A and B), except for those of the BMSP of *L. gigantea*, multiple VWA domains of BMSPs are separated into four monophyletic groups according to the cognate number of the VWA domains. Meanwhile, multiple VWA domains of BMSPs are clustered together with the VWA domain of a single VWA domain-containing VWA-CB dcp of *P. fucata*, Pfu aug2.0 3932.1 09248.t1 (Figs. 3.2 A and B; 3.8 A and B). Taken together, possession of a single VWA domain in the common ancestor of BMSPs and single VWA domain-containing VWA-CB dcps can be inferred. Duplication (Dp) events of the VWA domain generated multiple VWA domains of BMSPs in the common ancestor of the bivalves and the gastropods before their divergence (in the internal branch between nodes A and B; Figs. 3.2 A and B; 3.8 A and B). This scenario is different from that indicated by a previous study (Suzuki, et al. 2013), in which multiple VWA domains of BMSPs were suggested to have been produced by independent duplications in each species. The previous study was based on an alignment between the BMSP of *M. galloprovincialis*, an SMP identified using calcium carbonate-binding assay (Suzuki, et al. 2013; Suzuki, et al. 2011), and LG236719, a multiple VWA domain-containing protein deduced from the genome data of *Lottia gigantea* (v1.0, http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html). Indeed, phylogenetic analyses of the VWA domains of the VWA-CB dcps identified from the shells as well as those of this theoretical BMSP LG236719 (Fig. 3.18) based on 138 amino acid residues (Figs. 3.9 A, B, C and D) show that the VWA domains of LG236719 remained out of the cluster formed by the other BMSPs (including the Lgi BMSP which was identified from the shell proteome of *L. gigantea*), forming a single cluster with one or two single VWA domain-containing VWA-CB dcps of *L. gigantea* (Figs. 3.9 B and D). This observation indicates species-specific duplications of the

VWA domains for those VWA-CB dcps that is distinct from those of other BMSPs. However, it is dubious that LG236719 of *L. gigantea* is a shell protein, since its presence in the shell has not been confirmed by shell proteome analysis, and it is not found in the published shell proteome data of *L. gigantea* (Marie, et al. 2013). Therefore, it appears possible that the hypothesis proposed in a previous study (Suzuki, et al. 2013) was built on a comparison between SMPs and a non-SMP, which may have other physiological functions.

A recruitment of the VWA-CB dcps to the larval shell of the common ancestor of *C. gigas* and *P. fucata* before the speciation was also suggested by node F (Figs. 3.2 A and B) and node C (Figs. 3.8 A and B), where the larval SMPs of the two species form a single group.


A
Mga adult BMSP 2
Pfu cdna2.0 089203 larval adult BMSP 2
CGI 10009194 larval BMSP 2
Lgi adult BMSP 1
Lgi adult BMSP 2
Mga adult BMSP 3
Pfu cdna2.0 089203 larval adult BMSP 3
CGI 10009194 larval BMSP 3
Mga adult BMSP 1
Pfu cdna2.0 089203 larval adult BMSP 1
CGI 10009194 larval BMSP 1
Mga adult BMSP 4
Pfu cdna2.0 089203 larval adult BMSP 4
CGI 10009194 larval BMSP 4
Pfu aug2.0 3932.1 09248.t1 adult SMP
Pfu aug2.0 715.1 17768.t1 adult SMP
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
Lgi 228264 adult SMP
Lgi 232022 adult SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
Cin
Dp
A
B
C
D
E
F
L&A
L
BMSPs
Single VWA domain-containing VWA-CB dcps


B.
Mga adult BMSP 2
Pfu cdna2.0 089203 larval&adult BMSP 2
CGI 10009194 larval BMSP 2
Lgi adult BMSP 1
Lgi adult BMSP 2
Mga adult BMSP 3
Pfu cdna2.0 089203 larval&adult BMSP 3
CGI 10009194 larval BMSP 3
Mga adult BMSP 1
Pfu cdna2.0 089203 larval&adult BMSP 1
CGI 10009194 larval BMSP 1
Mga adult BMSP 4
Pfu cdna2.0 089203 larval&adult BMSP 4
CGI 10009194 larval BMSP 4
Pfu aug2.0 3932.1 09248.t1 adult SMP
Pfu aug2.0 715.1 17768.t1 adult SMP
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
Lgi 228264 adult SMP
Lgi 232022 adult SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
Cin
Dp
A
B
C
D
E
F
G
L&A
L
BMSPs
Single VWA domain-containing VWA-CB dcps

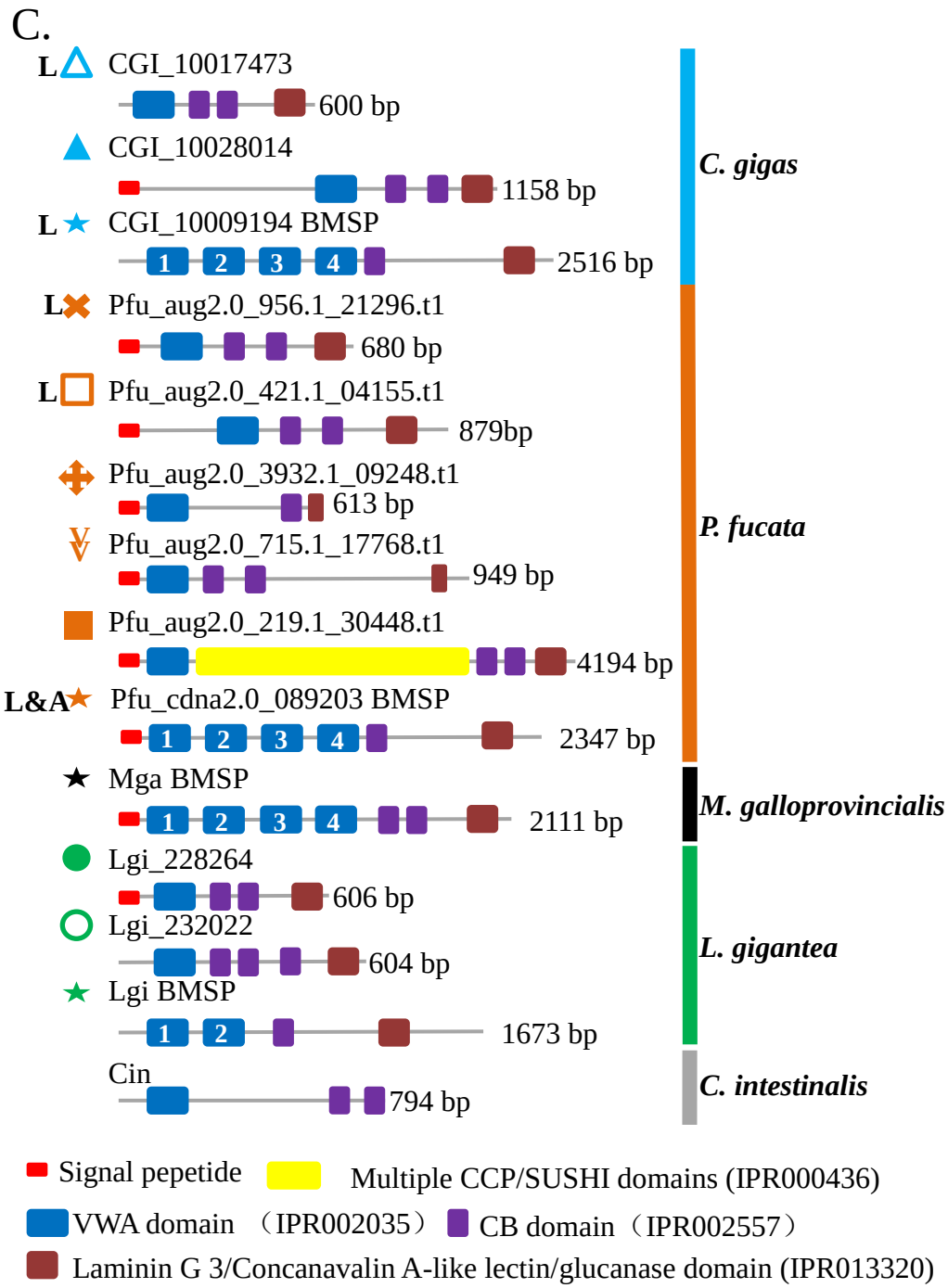


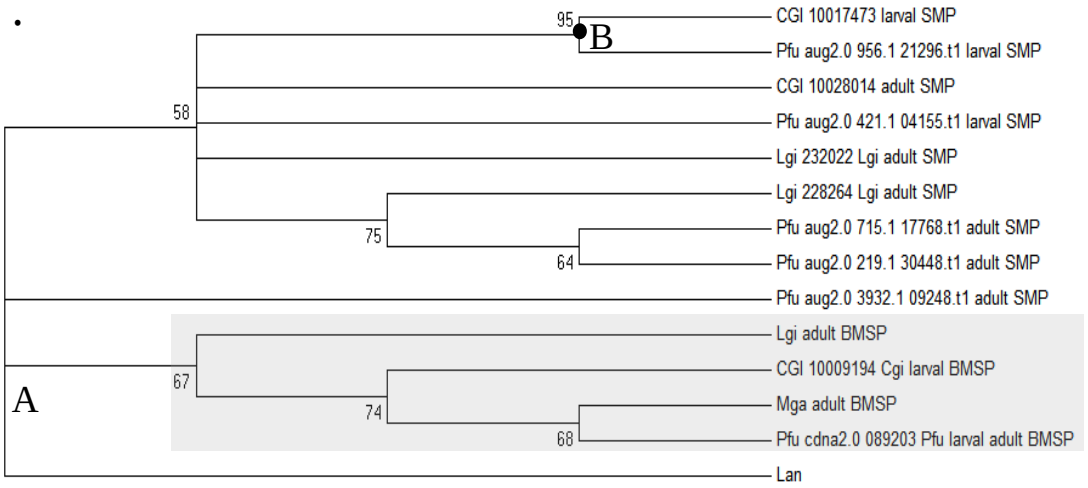
Fig. 3.2. Phylogenetic analyses of VWA domains of VWA-CB dcps of molluscan shells through LG model on 127 amino acid residues. A, Polychotomes are generated if the bootstrap value of a node is lower than 50. Nodes A-F were indicated. B, All dichotomes are preserved. Bootstrap values are shown if over 50, and marked with black dots if over 80. Nodes A-G were indicated. C, Schematics of domain structure of VWA-CB dcps. In the tree region, same proteins were marked by same marks. BMSPs were marked by stars and grep squares. A SMP was marked by 'L' if it is a larval SMP. Pfu_cdna2.0_089203 was identified from both larval and adult shells of *P. fucata*. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Cin, *Ciona intestinalis*.

Notably, unlike the other larval VWA-CB containing SMP of *P. fucata*, Pfu_aug2.0_956.1_21296.t1, which forms a clade with the larval SMP of *C. gigas* CGI_10017473 (Figs. 3.1 A and B, node H; Figs. 3.7 A and B, node F; Figs. 3.2 A and B, node F; Figs. 3.8 A and B, node C), the larval SMP, Pfu_aug2.0_421.1_04155.t1 forms a clade with an adult SMP of *C. gigas*, CGI_10028014 (Figs. 3.2 A and B, node D; Figs. 3.8 A and B, node D), suggesting that it derived from an ancestrally adult SMP or the ancestrally larval common ancestor gave rise to an adult SMP. An interesting observation concerning this is that, in contrast to other VWA-CB domain containing SMPs of *P. fucata*, it exhibits an expression pattern with double peaks, being highly expressed in both the larval and adult stages in the pearl oyster (Fig 3.8 C), suggesting that it retains a transitional phase in its history that an original adult shell protein was recruited by the larval SMP repertoire later for the formation of the larval shell, or vice versa. The genetic mechanism behind this hypothesis might be explained in terms of heterochronic gene expression.

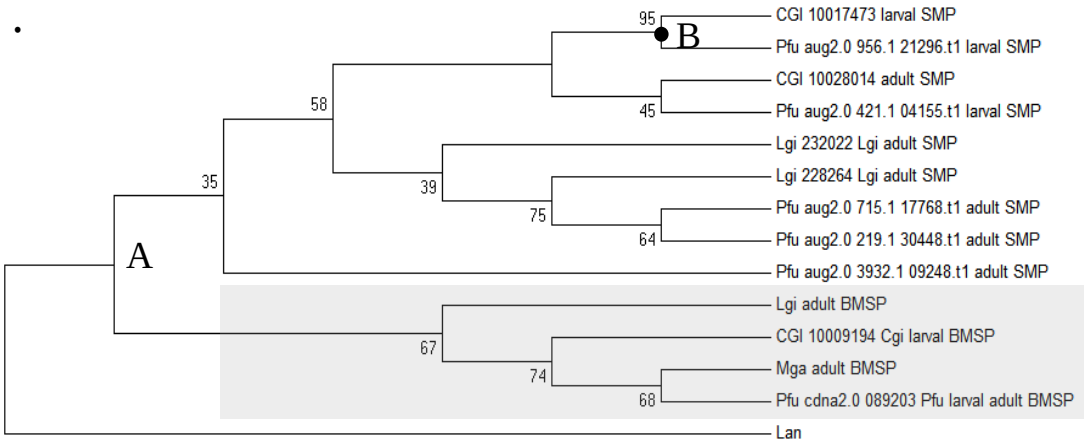
Laminin G domains have been reported from the downstream region of the chitin-binding domain in Pif/BMSP-like proteins (Suzuki, et al. 2013). In this study, a combined domain search using SMART and BLAST revealed conserved Laminin G-like region from the VWA-CB dcps (Figs. 3.1 C and 3.2 C), though only the N-terminal or the C-terminal sequence of the domain were identified from the complete sequences of the two adult SMPs of *P. fucata*, Pfu_aug2.0_715.1_17768.t1 and Pfu_aug2.0_3932.1_09248.t1 (Fig. 3.19 A). Phylogenetic analyses performed on the Laminin G domain (Figs. 3.3 A and B and 3.10 A and B) and concatenated sequences of the CB domain and Laminin G domain regions (Figs. 3.3 C and D and 3.10 C and D), because the CB1 and single CB domain of VWA-CB dcps are inferred to be orthologues (Figs. 3.1 A and B, 3.7 A and B). Again, the topology suggested the recruitment of the VWA-CB dcp to the larval shell of the common ancestor of *C. gigas* and *P. fucata* before the divergence of the two species (Figs. 3.3 A, B, C and D, node B; 3.10 A, B, C, D, E and F, node B), and the duplication of the VWA domain occurred in the common ancestor of the bivalves and the gastropods gave rise to

BMSPs in the shell, which can be revealed by the monophyletic group of BMSPs (Figs. 3.3 A, B, C and D; 3.10 A, B, C, D, E and F). The phylogram of Laminin G domain suggested that more amino acid substitutions occurred to this region of Pfu_aug2.0_3932.1_09248.t1 compared with other VWA-CB dcps analyzed in this study (Figs. 3.10 E and F), a fact which can also be seen from the alignment (Fig. 3.19 A).

A



B



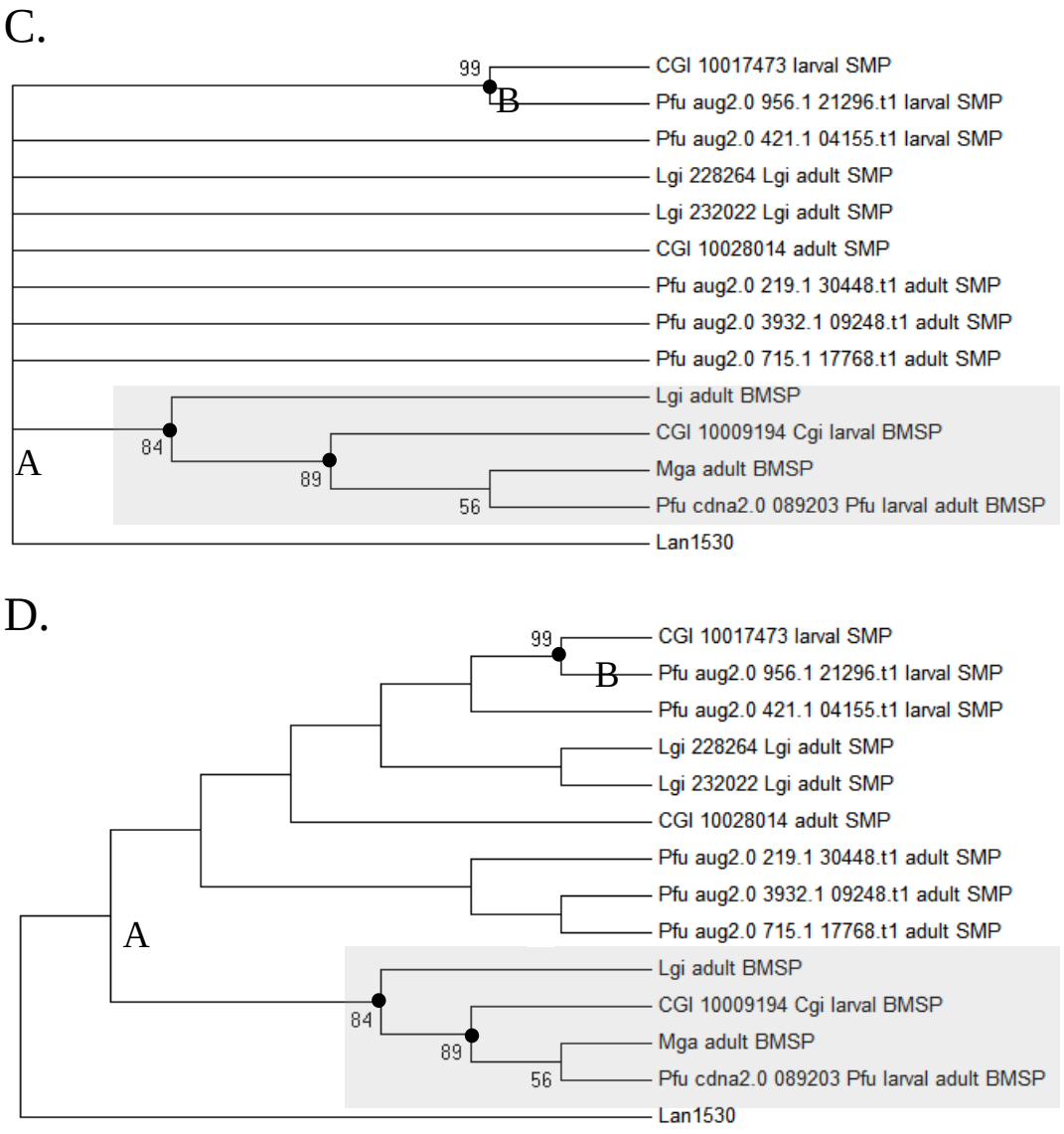


Fig. 3.3. Phylogenetic analyses of Laminin G domains on 57 aa and the concatenated sequence of a CB domain and the Laminin G domain on 86 aa of VWA-CB dcps of molluscan shells through LG model, respectively. A, Phlylogenetic tree performed on Laminin G domains. Polychotomes are generated if the bootstrap value of a node is lower than 50. B, Phlylogenetic tree performed on Laminin G domains. All dichotomes are preserved. C, Phlylogenetic tree performed on the concatenated sequence of a CB domain and the Laminin G domain. Polychotomes are generated if the bootstrap value of a node is lower than 50. D, Phlylogenetic tree performed on the concatenated sequence of a CB domain and the Laminin G domain. All dichotomes are preserved. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*. The group formed by BMSPs are

Taken together, the phylogenetic analyses of the chitin-binding domain, VWA domain and Laminin G domain performed on the VWA-CB dcps identified in the shell of the bivalves and the gastropods all suggested that VWA-CB dcps were possessed by the common ancestor of *C. gigas* and *P. fucata* and that the BMSPs originated from a single-VWA-domain containing VWA-CB domain containing SMP by the duplications of the VWA domains and existed in the shell of the common ancestor of the bivalves and the gastropods.

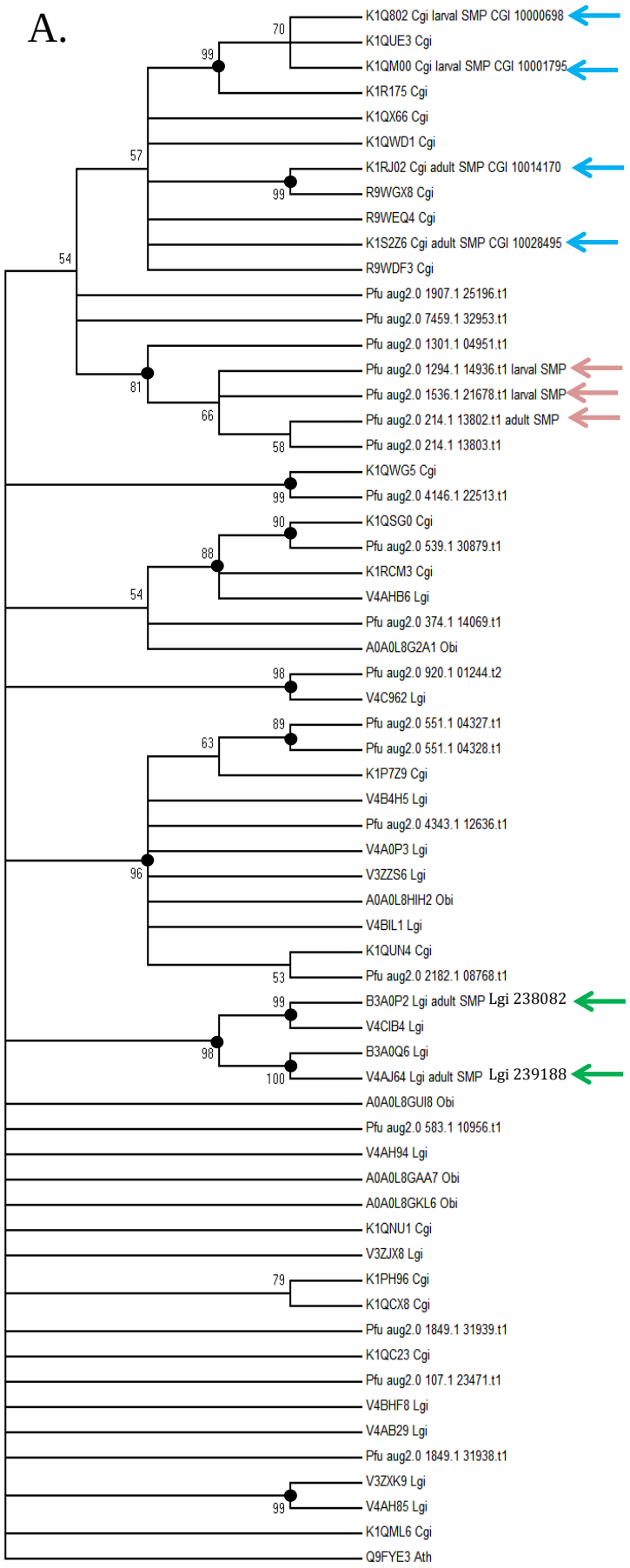
3.3.2 Relatively recent recruitment of CAs to the shells of bivalves

The conversion of carbon dioxide into bicarbonate was thought to be an ancestral function in carbonate biomineralization (Marin, et al. 2007). CAs were reported from a number of the adult shells of bivalves and gastropods (Mann, et al. 2012; Marie, et al. 2013; Marie, et al. 2011a; Zhang, et al. 2012; Zhao, et al. 2018), as well as the larval shells of bivalves (Zhao, et al. 2018). CA domains are highly expanded in molluscs including bivalves, gastropods and cepholopods compared to most other metazoan phyla (Zhao, et al. 2018). Taken together, it is tempting to conclude that CAs contained in the shells of extant molluscan species are inherited from the molluscan common ancestor in early Cambrian before the divergence of main classes (Lecointre and Le Guyader 2001; Sun, et al. 2017). However, surprisingly, CAs extracted from the shells of *C. gigas*, *P. fucata* and *L. gigantea* formed three separated clusters, each of which comprises of proteins of one species only (Figs. 3.4 A and B; 3.11 A and B; 3.12 A, B, C and D), which indicated the recruitment of CAs as SMPs occurred independently to each molluscan species. In bivalves, the recruitment is inferred to have occurred even after the divergence of *C. gigas* and *P. fucata* (Figs. 3.4 A and B; 3.11 A and B; 3.12 A, B, C and D), a divergence which is estimated to have occurred most probably in Triassic, or in Silurian at the earliest, depending on the phylogenetic interpretations of fossil taxa, based on the fossil record (Tracey, et al. 1993), or in the period from Carboniferous to Triassic, based on the molecular clock (Sun, et al. 2017). In either scenario, the functional diversifications of CA-SMPs in the larval and adult shells in those two bivalve species were rather recent events than

expected. Although the CA SMPs and the secreted CA of human form a cluster (Figs. 3.12 A and B), this relationship is not supported by a high bootstrap value and the topology was not repeated in the tree generated by the Poisson model (Figs. 3.12 C and D). Therefore, it is not certain to say if the CA of molluscan shells exhibit more similarities to any sort of the human CAs than to the other. Meanwhile, phylogenetic analyses performed on molluscan CAs indicate the larval or the adult shell CA can be derived from the duplication of each other.

In this study, independent recruitments of CAs to the shell of *C. gigas* and *P. fucata* after speciation were discovered. In the meantime, the topology of phylogenetic analyses agreed with two previous opinions (Zhao, et al. 2018): 1, the lineage-specific expansion of the CAs between bivalves and gastropods is supported by the clusters consisting entirely of CAs of bivalves or those of gastropods. 2 multiple homologs of CAs in the larval or adult molluscan shells were produced by independent duplications of CAs within each species (Figs. 3.4 A and B; 3.11 A and B; 3.12 A, B, C and D).

As a characteristics of Nacrein and N66 proteins of *Pinctada* species (Smith‐Keune and Jerry 2009), the NG-repeat region was demonstrated to have an inhibitory function *in vitro* in the rate of $CaCO_3$ precipitation (Miyamoto, et al. 2005). An important role of the NG-repeat domain for the inter-molecular interactions in the biomineralization processes (Kono, et al. 2000; Miyamoto, et al. 1996; Miyamoto, et al. 2005; Norizuki and Samata 2008; Smith‐Keune and Jerry 2009) and its calcium-binding ability have also been suggested, although the latter role is still under debate (Miyamoto, et al. 1996; Norizuki and Samata 2008). However, no NG-repeat domain was identified from the larval Nacrein protein of *P. fucata* nor from both the larval and adult Nacreins of *C. gigas* (Fig. 3.13). The absence of this domain suggests a possibility of functional divergence of Nacrein proteins between larval and adult SMPs in *P. fucata* and between *C. gigas* and *P. fucata*.


A.
K1Q802 Cgi larval SMP CGI 10000698
K1QUE3 Cgi
K1QM00 Cgi larval SMP CGI 10001795
K1R175 Cgi
K1QX66 Cgi
K1QWD1 Cgi
K1RJ02 Cgi adult SMP CGI 10014170
R9WGX8 Cgi
R9WEQ4 Cgi
K1S2Z6 Cgi adult SMP CGI 10028495
R9WDF3 Cgi
Pfu aug2.0 1907.1 25196.t1
Pfu aug2.0 7459.1 32953.t1
Pfu aug2.0 1301.1 04951.t1
Pfu aug2.0 1294.1 14936.t1 larval SMP
Pfu aug2.0 1536.1 21678.t1 larval SMP
Pfu aug2.0 214.1 13802.t1 adult SMP
Pfu aug2.0 214.1 13803.t1
K1QWG5 Cgi
Pfu aug2.0 4146.1 22513.t1
K1QSG0 Cgi
Pfu aug2.0 539.1 30879.t1
K1RCM3 Cgi
V4AHB6 Lgi
Pfu aug2.0 374.1 14069.t1
A0A0L8G2A1 Obi
Pfu aug2.0 920.1 01244.t2
V4C962 Lgi
Pfu aug2.0 551.1 04327.t1
Pfu aug2.0 551.1 04328.t1
K1P7Z9 Cgi
V4B4H5 Lgi
Pfu aug2.0 4343.1 12636.t1
V4A0P3 Lgi
V3ZZS6 Lgi
A0A0L8HIH2 Obi
V4BIL1 Lgi
K1QUN4 Cgi
Pfu aug2.0 2182.1 08768.t1
B3A0P2 Lgi adult SMP Lgi 238082
V4CIB4 Lgi
B3A0Q6 Lgi
V4AJ64 Lgi adult SMP Lgi 239188
A0A0L8GUI8 Obi
Pfu aug2.0 583.1 10956.t1
V4AH94 Lgi
A0A0L8GAA7 Obi
A0A0L8GKL6 Obi
K1QNU1 Cgi
V3ZJX8 Lgi
K1PH96 Cgi
K1QCX8 Cgi
Pfu aug2.0 1849.1 31939.t1
K1QC23 Cgi
Pfu aug2.0 107.1 23471.t1
V4BHF8 Lgi
V4AB29 Lgi
Pfu aug2.0 1849.1 31938.t1
V3ZXK9 Lgi
V4AH85 Lgi
K1QML6 Cgi
Q9FYE3 Ath

B.

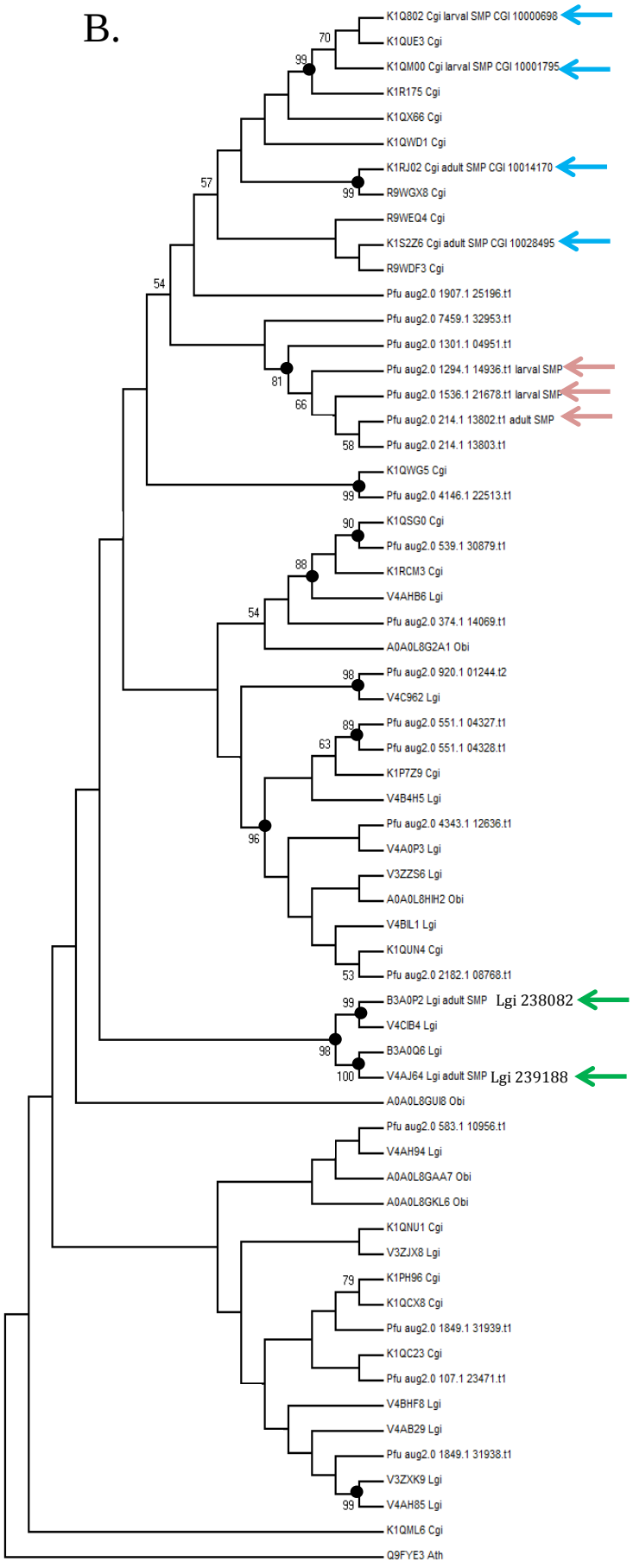

K1Q802 Cgi larval SMP CGI 10000698
K1QUE3 Cgi
K1QM00 Cgi larval SMP CGI 10001795
K1R175 Cgi
K1QX66 Cgi
K1QWD1 Cgi
K1RJ02 Cgi adult SMP CGI 10014170
R9WGX8 Cgi
R9WEQ4 Cgi
K1S2Z6 Cgi adult SMP CGI 10028495
R9WDF3 Cgi
Pfu aug2.0 1907.1 25196.t1
Pfu aug2.0 7459.1 32953.t1
Pfu aug2.0 1301.1 04951.t1
Pfu aug2.0 1294.1 14936.t1 larval SMP
Pfu aug2.0 1536.1 21678.t1 larval SMP
Pfu aug2.0 214.1 13802.t1 adult SMP
Pfu aug2.0 214.1 13803.t1
K1QWG5 Cgi
Pfu aug2.0 4146.1 22513.t1
K1QSG0 Cgi
Pfu aug2.0 539.1 30879.t1
K1RCM3 Cgi
V4AHB6 Lgi
Pfu aug2.0 374.1 14069.t1
A0A0L8G2A1 Obi
Pfu aug2.0 920.1 01244.t2
V4C962 Lgi
Pfu aug2.0 551.1 04327.t1
Pfu aug2.0 551.1 04328.t1
K1P7Z9 Cgi
V4B4H5 Lgi
Pfu aug2.0 4343.1 12636.t1
V4A0P3 Lgi
V3ZZS6 Lgi
A0A0L8HIH2 Obi
V4BIL1 Lgi
K1QUN4 Cgi
Pfu aug2.0 2182.1 08768.t1
B3A0P2 Lgi adult SMP
Lgi 238082
V4CIB4 Lgi
B3A0Q6 Lgi
V4AJ64 Lgi adult SMP
Lgi 239188
A0A0L8GUI8 Obi
Pfu aug2.0 583.1 10956.t1
V4AH94 Lgi
A0A0L8GAA7 Obi
A0A0L8GKL6 Obi
K1QNU1 Cgi
V3ZJX8 Lgi
K1PH96 Cgi
K1QCX8 Cgi
Pfu aug2.0 1849.1 31939.t1
K1QC23 Cgi
Pfu aug2.0 107.1 23471.t1
V4BHF8 Lgi
V4AB29 Lgi
Pfu aug2.0 1849.1 31938.t1
V3ZXK9 Lgi
V4AH85 Lgi
K1QML6 Cgi
Q9FYE3 Ath

Fig. 3.4. Phylogenetic trees of CA domains in molluscs via LG model. A, Polychotomes are generated if the bootstrap value of a node is lower than 50. B, All dichotomes are preserved. Trees were generated on 82 amino acid residues. SMPs were indicated by blue (*C. gigas*), Pink (*P. fucata*) and green (*L. gigantea*) arrowheads. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides.* Ath, *Arabidopsis thaliana*.

3.3.3 Chitobiases were possessed by larval and adult shells of the last common ancestor of *C. gigas* and *P. fucata*

Chitin and fibroin-like proteins are considered to be integral to the shell matrix to provide the framework for the nucleation and growth of crystals (Furuhashi, et al. 2009; Weiner and Traub 1984). The family 18 chitinase and the family 20 chitobiase are chitinolytic enzymes. The former are essential enzymes for chitin degradation and the latter hydrolyzes *N*,*N*'-diacetylchitobiose [$(GlcNAc)_2$] produced by the former to GlcNAc (Suzuki, et al. 2007). Chitinases and chitobiases have been reported from organic matrices of adult molluscan shells (Gotliv, et al. 2005; Kintsu, et al. 2017; Mann, et al. 2012; Zhang, et al. 2012; Zhao, et al. 2018). On the other hand, a family 20 chitobiase, Pfu_aug2.0_6.1_20027.t1, was identified as the only chitinolytic enzyme in the larval shell for the first time (Fig. 3.5 A, Table 3.4) (Zhao, et al. 2018). A BLASTP search of the *P. fucata* larval chitobiase against the gene models of the whole genome of *C. gigas* revealed that the protein predicted from CGI_10007856 is highly similar (Identity: 70%; E-value: 0) to the larval shell-chitobiase of *P. fucata* (Zhao, et al. 2018) (Fig. 3.5 A), a fact which suggested that this gene is a potential SMP escaped from the detection by the proteomic analysis. Transcriptomic data showed that the expression of CGI_10007856 reaches the peak at the trochophore stage, when the larval shell starts to form, while compared with that of the larval shell-chitobiase of *P. fucata*, a suddenly reduction with the appearance of D-shape larva was observed (Fig. 3.5 B), a fact which might explain why it was undetectable in the larval shell proteome. These evidences suggested that the chitobiase SMP exists in the larval shells of the oyster, and that it might be another common SMP shared by both larval and adult shells of *C. gigas* and *P. fucata*.

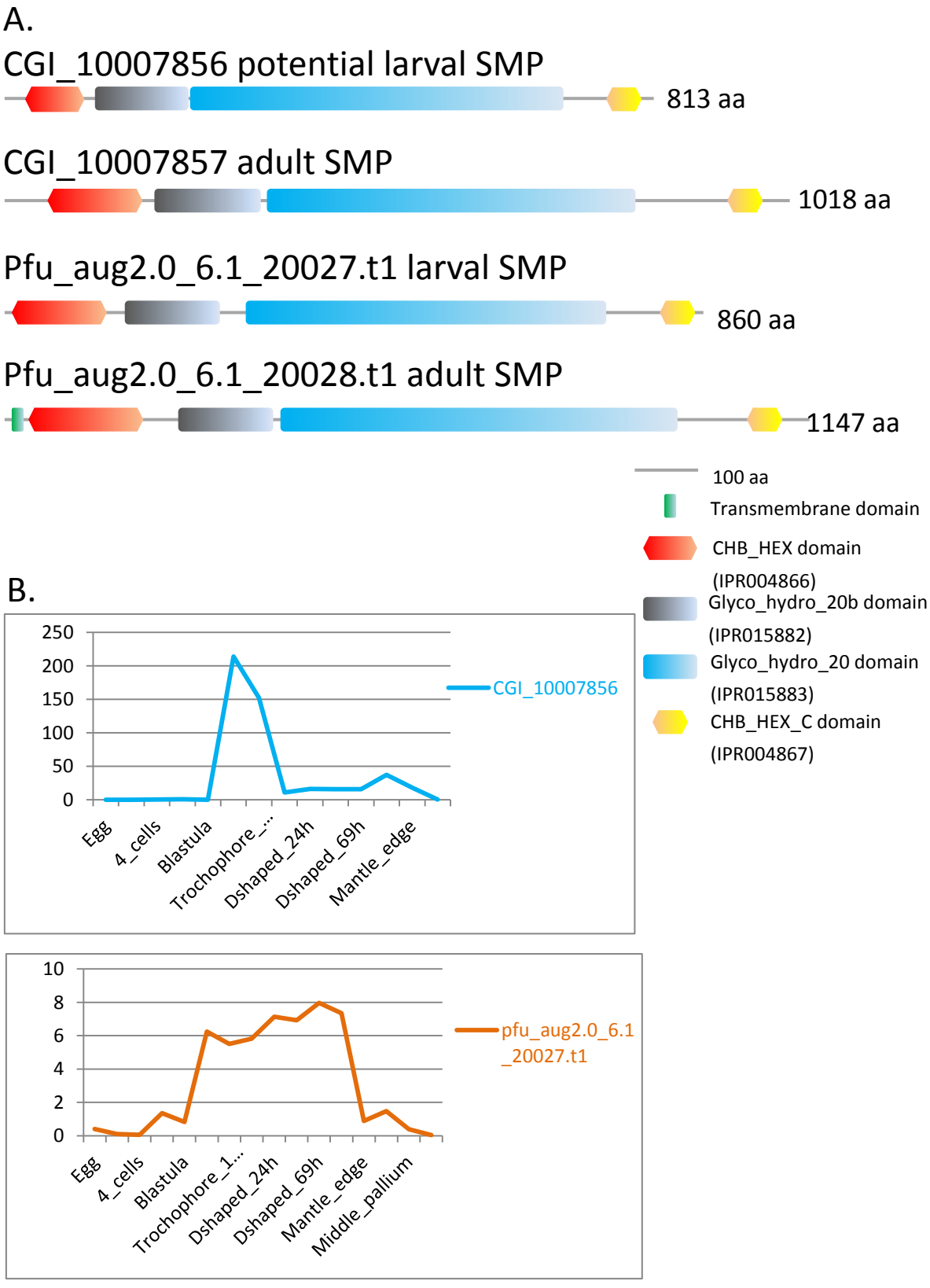


Fig. 3.5. Chitobiases identified in the shell of *C. gigas* and *P. fucata*. A, domain structures of chitobiases. B, Transcript expression patterns during development of the larval chitobiase of *C. gigas* (blue) and *P. fucata* (orange).

Molecular phylogenetic analyses indicated that the hypothetical larval SMP, CGI_10007856 indeed clustered within the group comprised of the other three shell-chitobiases, suggesting that it is likely an SMP of the oyster larva (Figs. 3.6 A and B; Figs. 3.14 A and B; Figs. 3.15 A, B, C and D). The group comprised of larval chitobiases of *C. gigas* and *P. fucata* as well as the group comprised of adult chitobiases of *C. gigas* and *P. fucata*, agreed by the topologies of all four trees, explicitly suggests a duplication (Dp) event occured to the chitobiase in the larval or in the adult shell of the last common ancestor of two species before the speciation (Figs. 3.6 A and B; Figs. 3.14 A and B; Figs. 3.15 A, B, C and D). Thus, recruitments of chitobiases to the larval and adult shells of the last common ancestor of *C. gigas* and *P. fucata* were suggested.

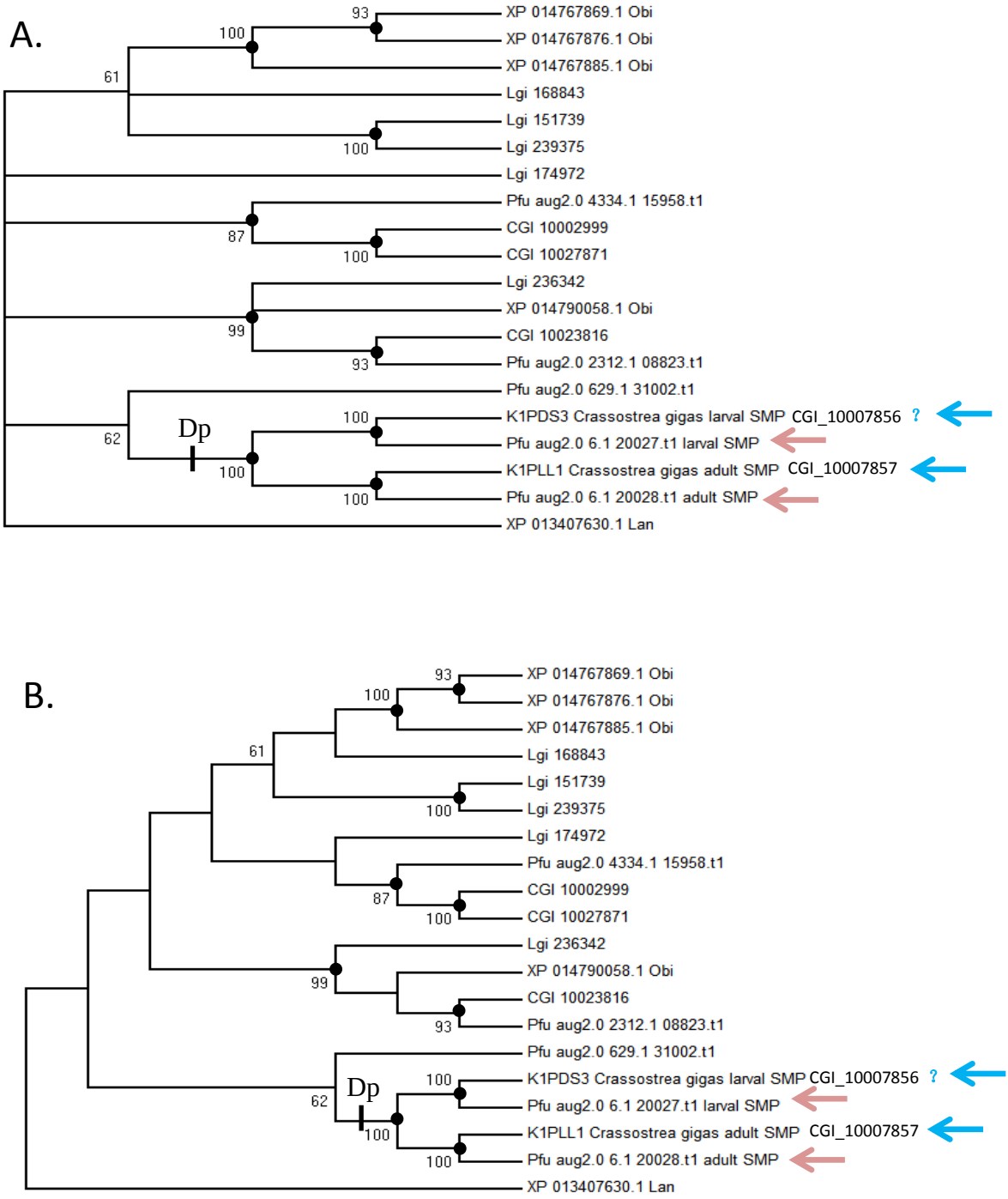


Fig. 3.6. Phylogenetic trees of chitobiases in molluscs via LG model on concatenated sequence of CHB_HEX domain (IPR004866), Glyco_hydro_20b domain (IPR015882), Glyco_hydro_20 domain (IPR015883) and CHB_HEX_C domain (IPR004867). A, Polychotomes are generated if the bootstrap value of a node is lower than 50. B, All dichotomes are preserved. Trees were generated on 523 amino acid residues. SMPs were indicated by blue (*C. gigas*) and Pink (*P. fucata*) arrowheads. Question mark indicates if the gene is a SMP is questionable. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides.* Lan, *Lingular anatina*.

3.4 Conclusions

Despite the complexity of the domain structures of VWA-CB dcps in molluscan shells exhibited by different species or even within single species, phylogenetic analyses performed separately on VWA, CB and Laminin G domains revealed that, regardless of the diversity in the number of VWA or CB domains, all VWA-CB dcps, including the multiple VWA domain-containing BMSPs, are inferred to have originated from a protein containing one VWA and two CB domains in the common ancestor of bivalves and gastropods. Possession of both the VWA-CB dcp with one VWA domain and the BMSP with multiple VWA domains by the common ancestor of bivalves and gastropods has been suggested. Phylogenetic analyses in this study also suggested that those VWA-CB dcps and BMSP as well as chitobiase were already contained in the larval and the adult shells of the common ancestor of bivalves before the speciation between *C. gigas* and *P. fucata*. On the other hand, in one of the most common SMPs which expanded widely among molluscs, the carbonic anhydrase SMPs, the gene duplications which gave rise to a separate deployment of larval and adult SMPs are inferred to have taken place after the divergence of those two bivalve species, which is a rather recent event than has hitherto been expected.

We discovered the dual proteomes of larval and the adult shells in *C. gigas* and in *P. fucata*, both exhibiting rather limited components common to both larval and adult SMPs (Zhao, et al. 2018). The fast evolution of the larval shell proteins makes the larval shell more than just a primitive state of molluscan common ancestor, yet it is evenhanded if we consider the larval stages also face the rigors of life (Garstang 1922) and new traits are allowed to appear at any developmental stage. Here we deeply scrutinized the evolutionary histories of several SMPs which might be important for the formation of both larval and adult shells, and inferred the time point when they were recruited by the molluscan shells using the proteomic and genomic data. However, origins and evolutionary scenarios might be more complicated than have been shown in this study. The potential “transitional phase” of heterochronic evolution manifested by the gene expression patterns of the transcript for the larval

SMP Pfu_aug2.0_421.1_04155.t1 points to a possibility that characterizations of the developmental gene networks controlling the formation of larval and adult shells may help to solve this question of antiquity of the larval shells in future. Systematic sampling of both larval and adult SMPs from more molluscan species and even across taxa should be considered in the future.

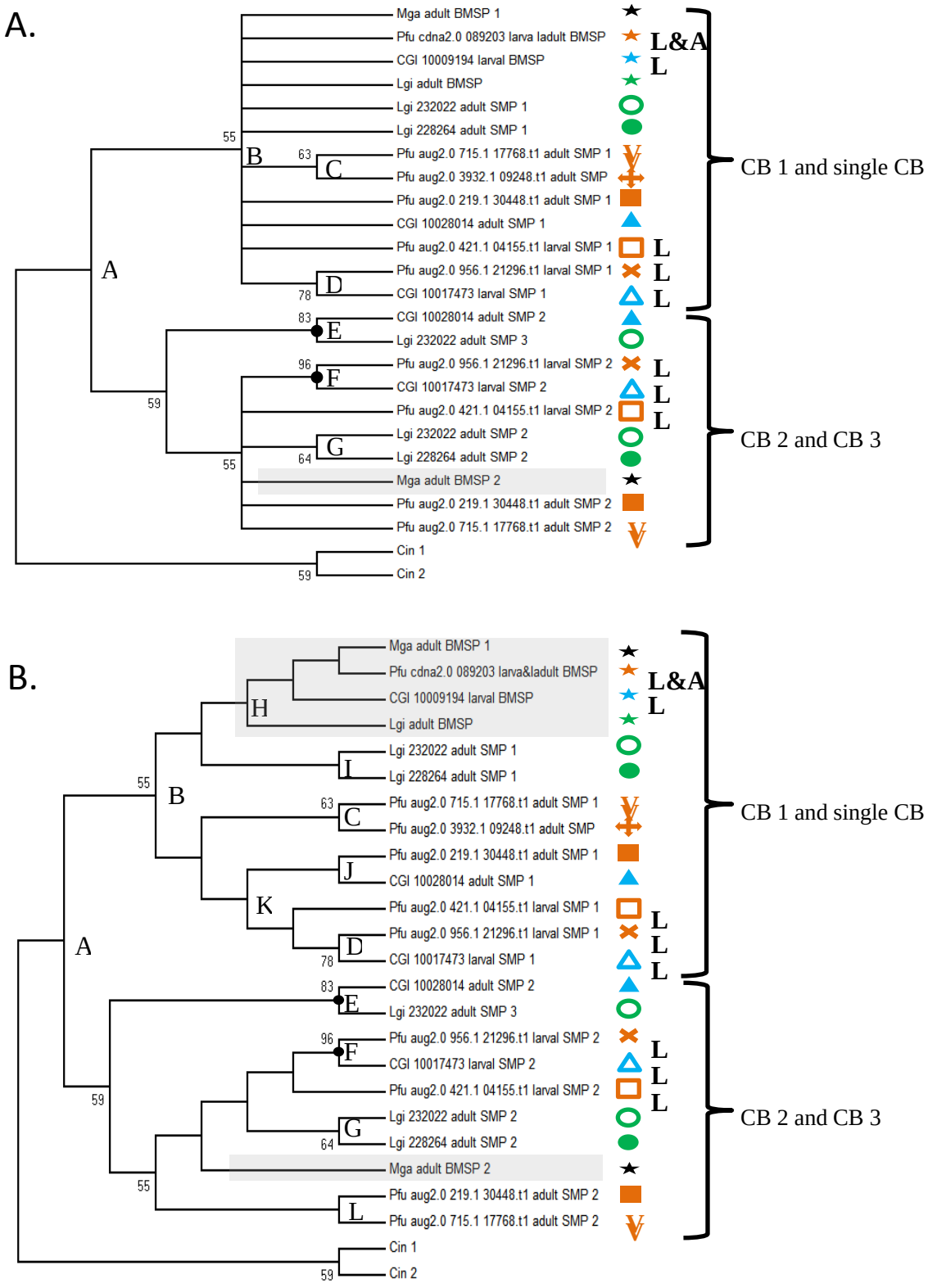
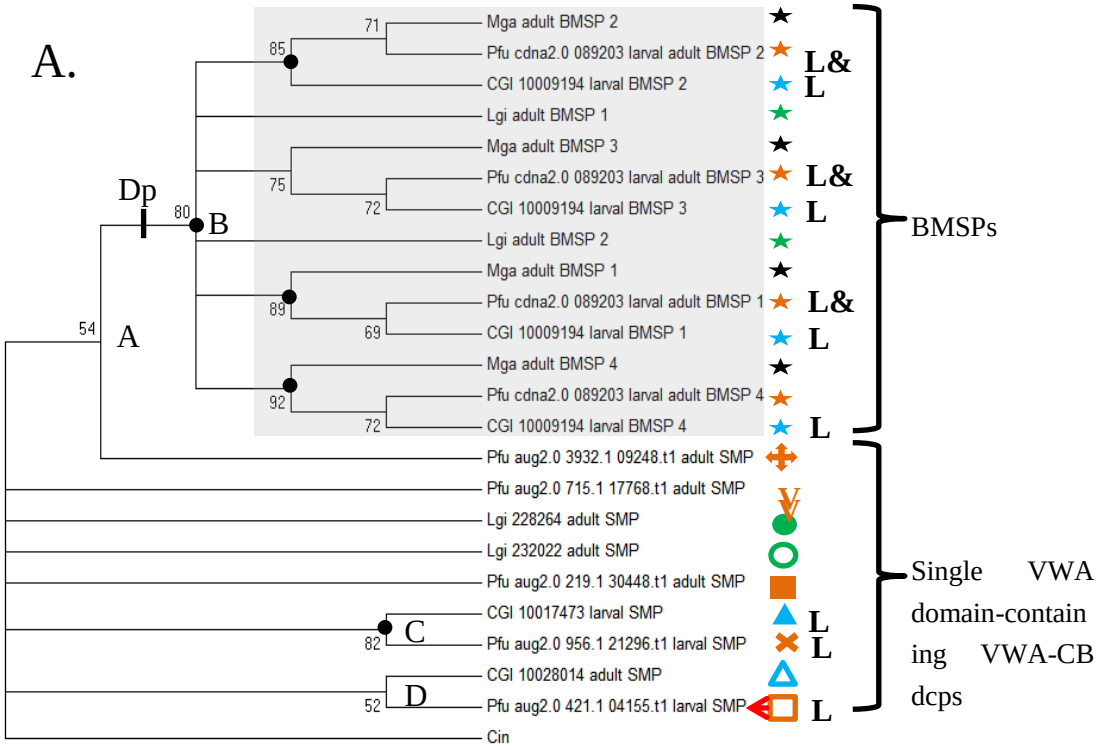
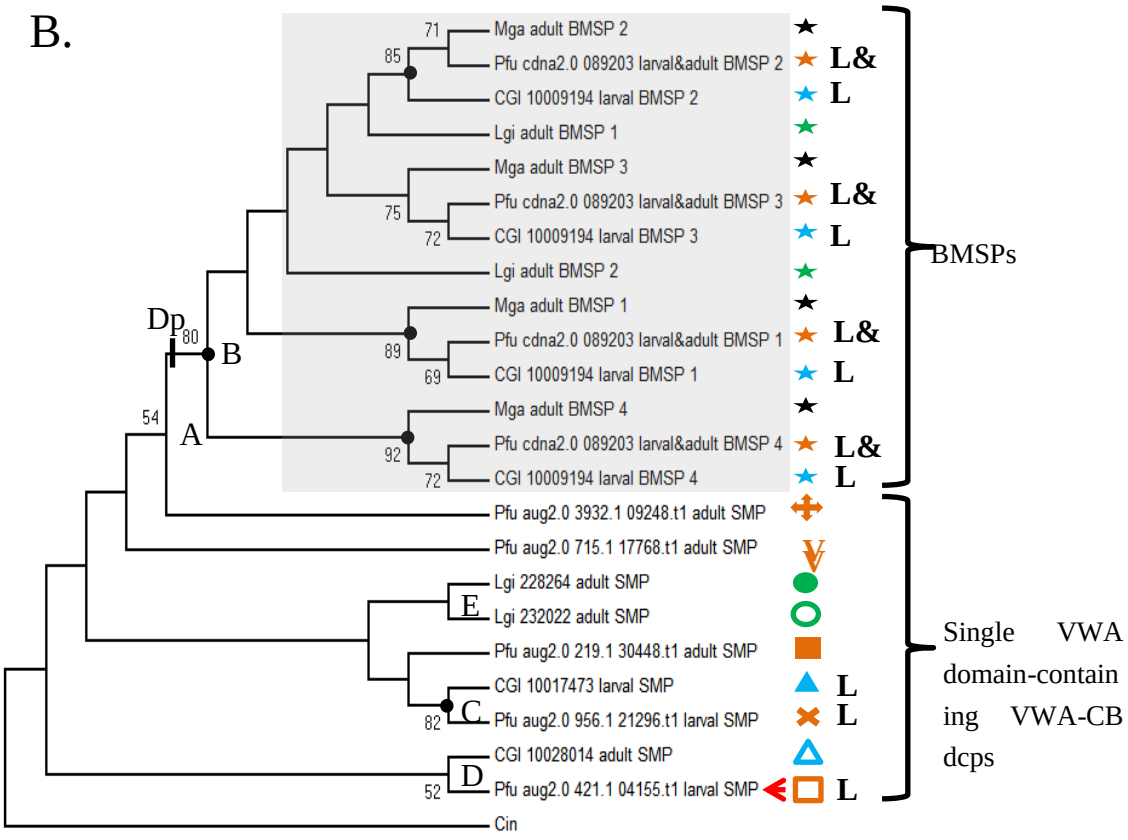


Fig. 3.7. Phylogenetic analyses of CB domains of VWA-CB dcps of molluscan shells through Poisson model on 37 amino acid residues. A, Polychotomes are generated if the bootstrap value of a node is lower than 50. Nodes A-G were indicated. B, All dichotomes are preserved. Bootstrap values are shown if over 50, and marked with black dots if over 80. Nodes A-L were indicated. In the tree region, same proteins were marked by same marks. BMSPs were marked by stars and grep squares. A SMP was marked by 'L' if it is a larval SMP. Pfu_cdna2.0_089203 was identified from both larval and adult shells of *P. fucata*. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Cin, *Ciona intestinalis*.


A.
Mga adult BMSP 2
Pfu cdna2.0 089203 larval adult BMSP 2
CGI 10009194 larval BMSP 2
Lgi adult BMSP 1
Mga adult BMSP 3
Pfu cdna2.0 089203 larval adult BMSP 3
CGI 10009194 larval BMSP 3
Lgi adult BMSP 2
Mga adult BMSP 1
Pfu cdna2.0 089203 larval adult BMSP 1
CGI 10009194 larval BMSP 1
Mga adult BMSP 4
Pfu cdna2.0 089203 larval adult BMSP 4
CGI 10009194 larval BMSP 4
Pfu aug2.0 3932.1 09248.t1 adult SMP
Pfu aug2.0 715.1 17768.t1 adult SMP
Lgi 228264 adult SMP
Lgi 232022 adult SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
Cin
Dp
A
B
C
D
L&
L
BMSPs
Single VWA domain-containing VWA-CB dcps


B.
Mga adult BMSP 2
Pfu cdna2.0 089203 larval&adult BMSP 2
CGI 10009194 larval BMSP 2
Lgi adult BMSP 1
Mga adult BMSP 3
Pfu cdna2.0 089203 larval&adult BMSP 3
CGI 10009194 larval BMSP 3
Lgi adult BMSP 2
Mga adult BMSP 1
Pfu cdna2.0 089203 larval&adult BMSP 1
CGI 10009194 larval BMSP 1
Mga adult BMSP 4
Pfu cdna2.0 089203 larval&adult BMSP 4
CGI 10009194 larval BMSP 4
Pfu aug2.0 3932.1 09248.t1 adult SMP
Pfu aug2.0 715.1 17768.t1 adult SMP
Lgi 228264 adult SMP
Lgi 232022 adult SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
Cin
Dp
A
B
C
D
E
L&
L
BMSPs
Single VWA domain-containing VWA-CB dcps

C.

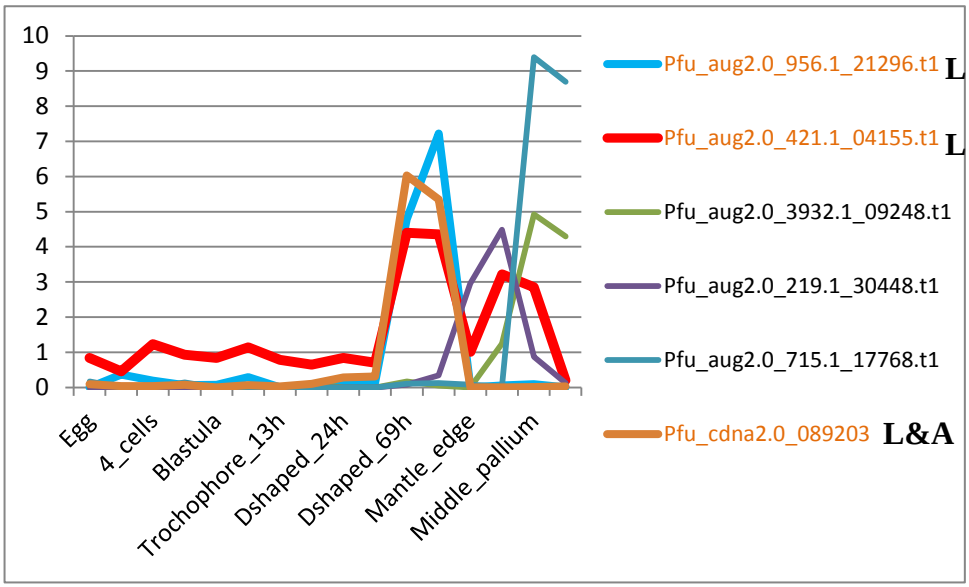
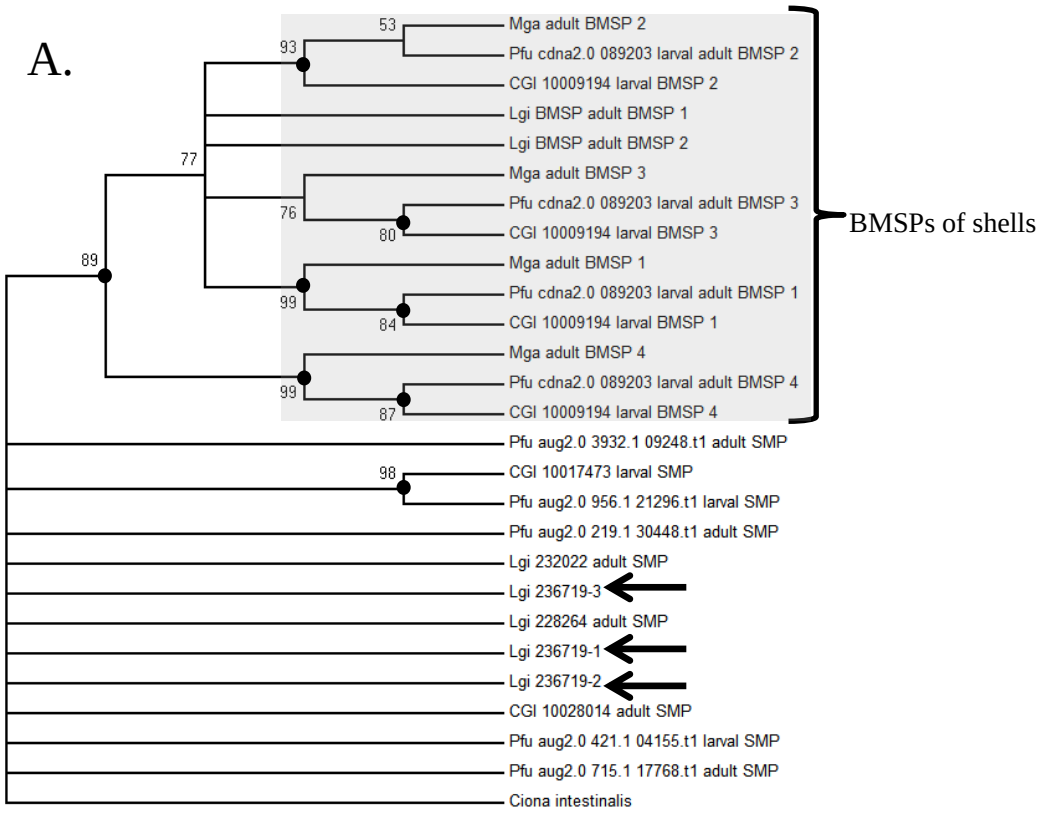
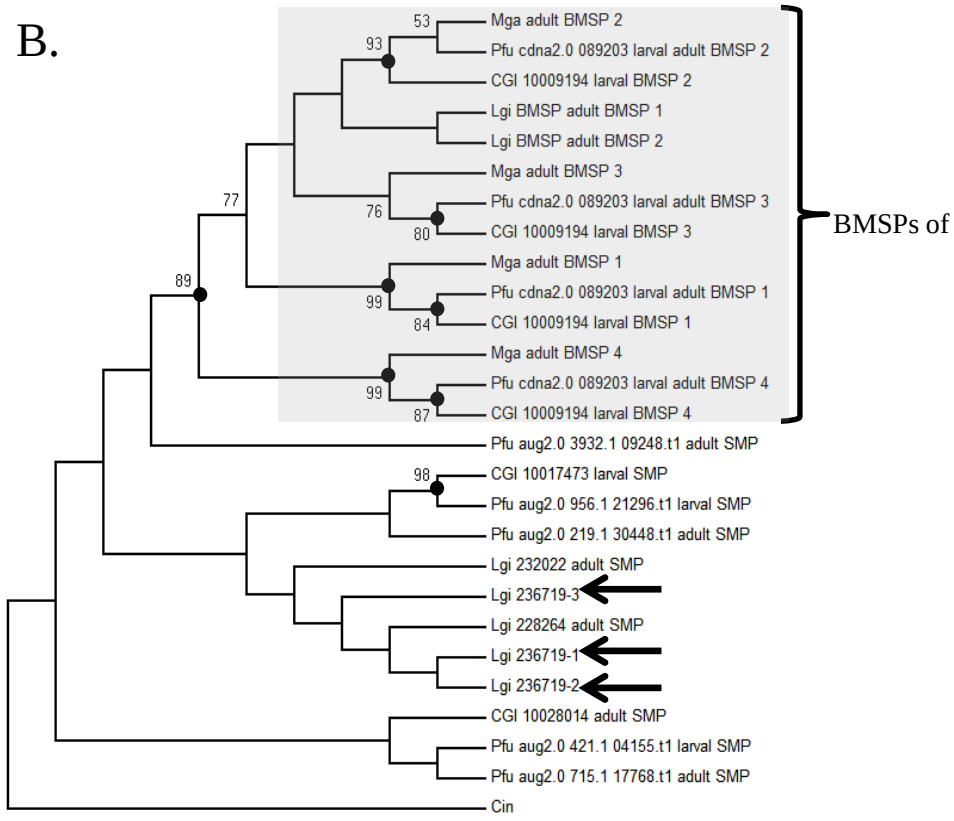


Fig. 3.8. Phylogenetic analyses of VWA domains of VWA-CB dcps of molluscan shells through Poisson model on 127 amino acid residues. A, Polychotomes are generated if the bootstrap value of a node is lower than 50. Nodes A-D were indicated. B, All dichotomes are preserved. Bootstrap values are shown if over 50, and marked with black dots if over 80. Nodes A-E were indicated. B, expression patterns of transcripts of VWA-CB dcps of *P. fucata*. Larval SMPs were indicated by bold line and orange characters. In the tree region, same proteins were marked by same marks. BMSPs were marked by stars and grep squares. A SMP was marked by 'L' if it is a larval SMP. Pfu_cdna2.0_089203 was identified from both larval and adult shells of *P. fucata*. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Cin, *Ciona intestinalis*.


A.
Mga adult BMSP 2
Pfu cdna2.0 089203 larval adult BMSP 2
CGI 10009194 larval BMSP 2
Lgi BMSP adult BMSP 1
Lgi BMSP adult BMSP 2
Mga adult BMSP 3
Pfu cdna2.0 089203 larval adult BMSP 3
CGI 10009194 larval BMSP 3
Mga adult BMSP 1
Pfu cdna2.0 089203 larval adult BMSP 1
CGI 10009194 larval BMSP 1
Mga adult BMSP 4
Pfu cdna2.0 089203 larval adult BMSP 4
CGI 10009194 larval BMSP 4
BMSPs of shells
Pfu aug2.0 3932.1 09248.t1 adult SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
Lgi 232022 adult SMP
Lgi 236719-3
Lgi 228264 adult SMP
Lgi 236719-1
Lgi 236719-2
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
Pfu aug2.0 715.1 17768.t1 adult SMP
Ciona intestinalis


B.
Mga adult BMSP 2
Pfu cdna2.0 089203 larval adult BMSP 2
CGI 10009194 larval BMSP 2
Lgi BMSP adult BMSP 1
Lgi BMSP adult BMSP 2
Mga adult BMSP 3
Pfu cdna2.0 089203 larval adult BMSP 3
CGI 10009194 larval BMSP 3
Mga adult BMSP 1
Pfu cdna2.0 089203 larval adult BMSP 1
CGI 10009194 larval BMSP 1
Mga adult BMSP 4
Pfu cdna2.0 089203 larval adult BMSP 4
CGI 10009194 larval BMSP 4
BMSPs of
Pfu aug2.0 3932.1 09248.t1 adult SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
Lgi 232022 adult SMP
Lgi 236719-3
Lgi 228264 adult SMP
Lgi 236719-1
Lgi 236719-2
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
Pfu aug2.0 715.1 17768.t1 adult SMP
Cin

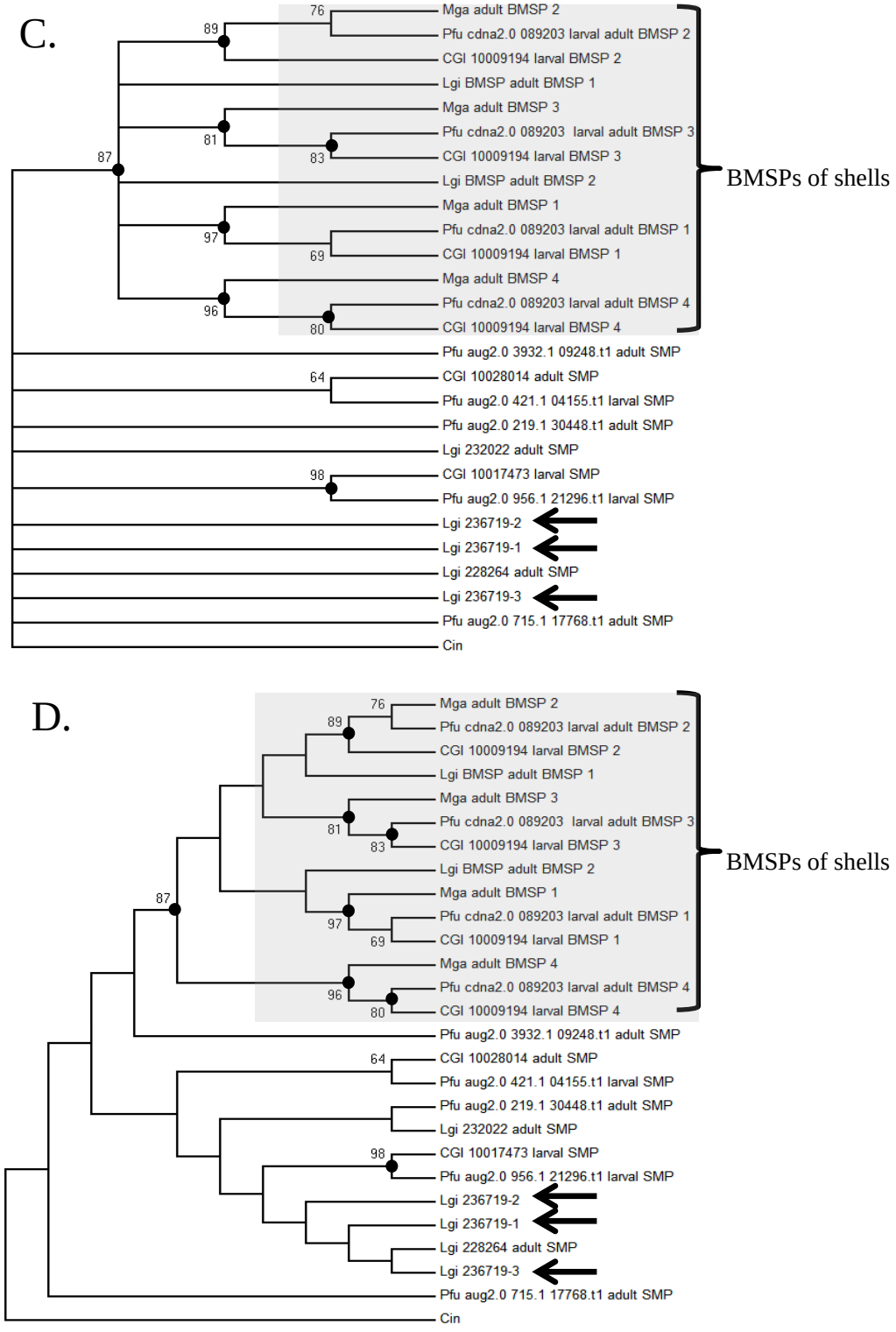
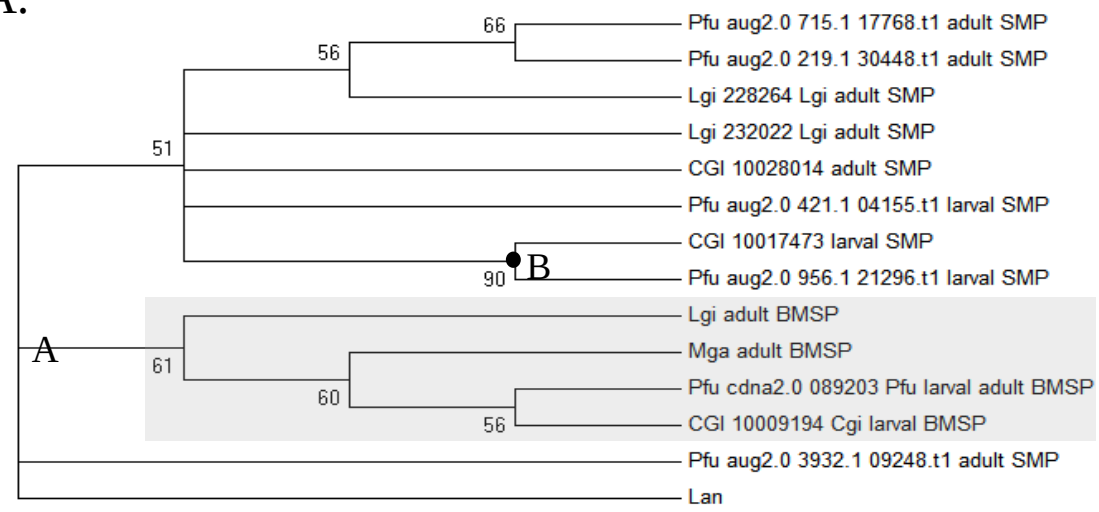
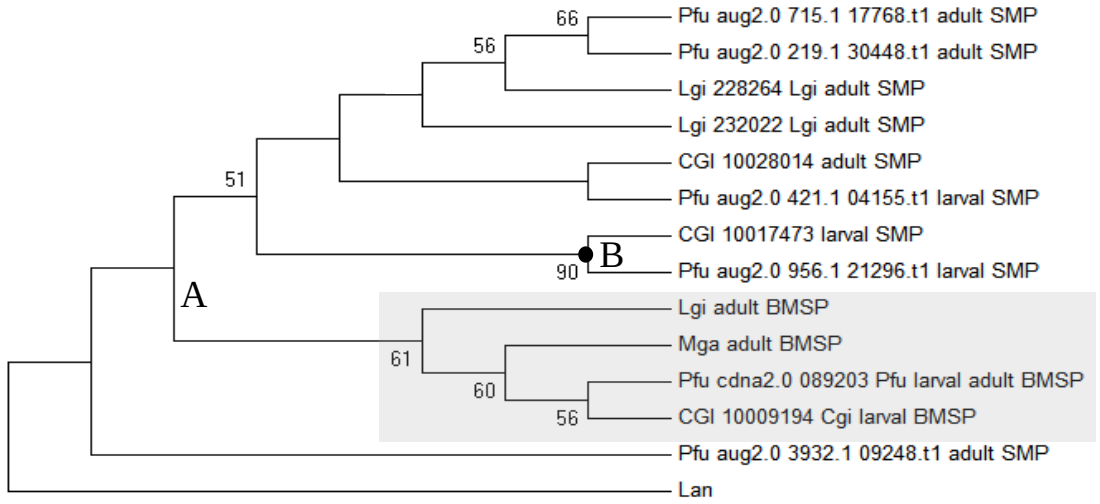


Fig. 3.9. Phylogenetic analyses of VWA domains of VWA-CB dcps of molluscan shells and LG236719 on 138 amino acid residues via LG (A, B) and Poisson (C, D). A, Polychotomes are generated if the bootstrap value of a node is lower than 50. All dichotomes are preserved. C, Polychotomes are generated if the bootstrap value of a node is lower than 50. D, All dichotomes are preserved. Bootstrap values are shown if over 50, and marked with black dots if over 80. VWA domains of LG236719 were indicated by arrowheads. The group formed by BMSPs are indicated by grey squares. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Cin, *Ciona intestinalis*.

A.


66
56
51
90
B
A
61
60
56
Pfu aug2.0 715.1 17768.t1 adult SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
Lgi 228264 Lgi adult SMP
Lgi 232022 Lgi adult SMP
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
Lgi adult BMSP
Mga adult BMSP
Pfu cdna2.0 089203 Pfu larval adult BMSP
CGI 10009194 Cgi larval BMSP
Pfu aug2.0 3932.1 09248.t1 adult SMP
Lan

B.


66
56
51
90
B
A
61
60
56
Pfu aug2.0 715.1 17768.t1 adult SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
Lgi 228264 Lgi adult SMP
Lgi 232022 Lgi adult SMP
CGI 10028014 adult SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
Lgi adult BMSP
Mga adult BMSP
Pfu cdna2.0 089203 Pfu larval adult BMSP
CGI 10009194 Cgi larval BMSP
Pfu aug2.0 3932.1 09248.t1 adult SMP
Lan

C.

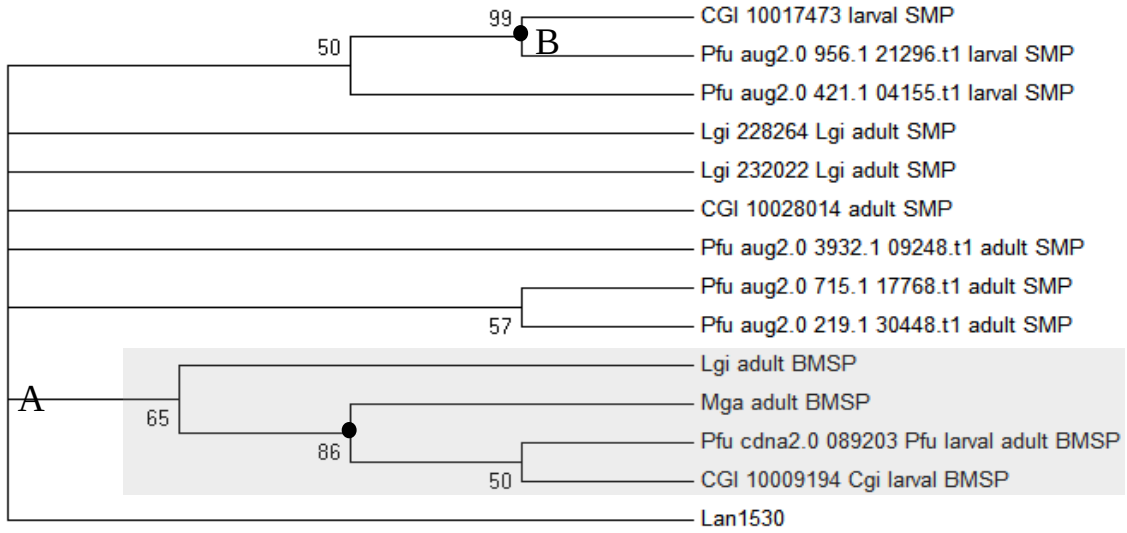
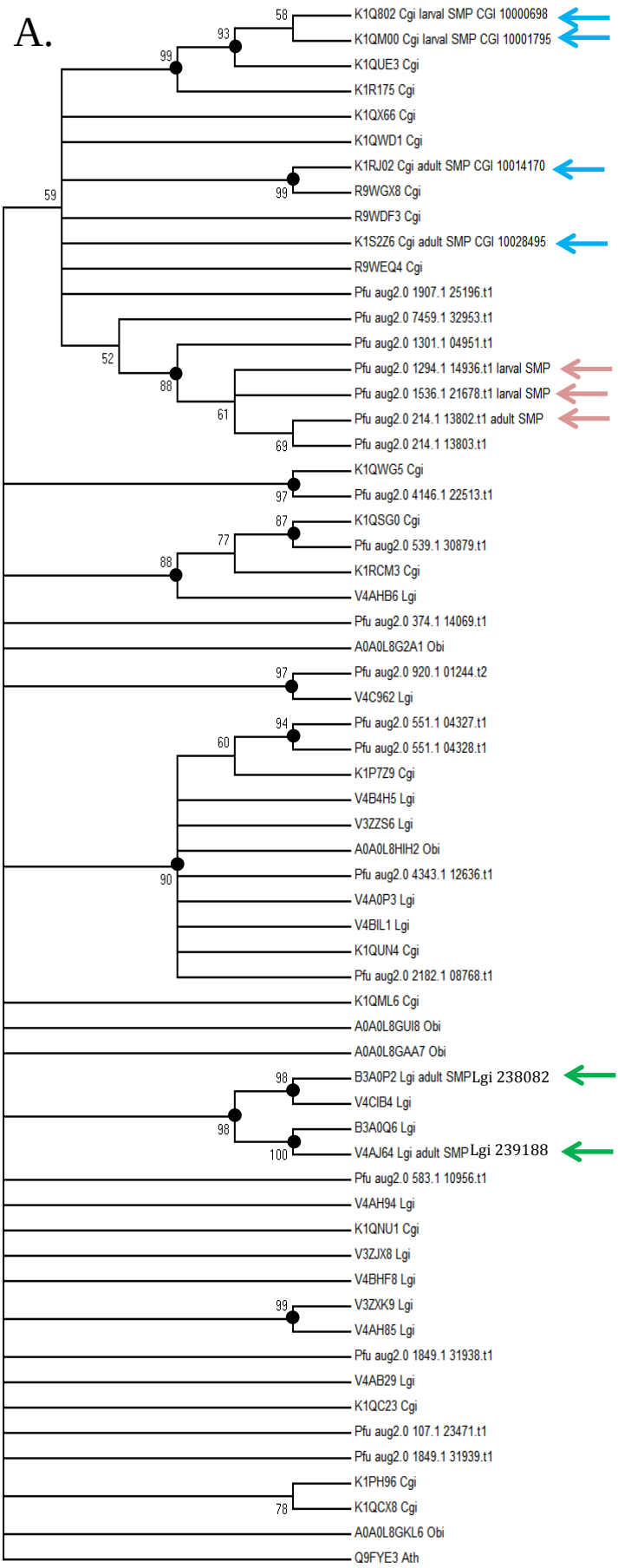

99
50
B
57
A
65
86
50
CGI 10017473 larval SMP
Pfu aug2.0 956.1 21296.t1 larval SMP
Pfu aug2.0 421.1 04155.t1 larval SMP
Lgi 228264 Lgi adult SMP
Lgi 232022 Lgi adult SMP
CGI 10028014 adult SMP
Pfu aug2.0 3932.1 09248.t1 adult SMP
Pfu aug2.0 715.1 17768.t1 adult SMP
Pfu aug2.0 219.1 30448.t1 adult SMP
Lgi adult BMSP
Mga adult BMSP
Pfu cdna2.0 089203 Pfu larval adult BMSP
CGI 10009194 Cgi larval BMSP
Lan1530

Fig. 3.10. Phylogenetic analyses of Laminin G domains on 57 aa and the concatenated sequence of a CB domain and the Laminin G domain on 86 aa of VWA-CB dcps of molluscan shells through Poisson model, respectively. A, Phlylogenetic tree performed on Laminin G domains. Polychotomes are generated if the bootstrap value of a node is lower than 50. B, Phlylogenetic tree performed on Laminin G domains. All dichotomes are preserved. C, Phlylogenetic tree performed on the concatenated sequence of a CB domain and the Laminin G domain. Polychotomes are generated if the bootstrap value of a node is lower than 50. D, Phlylogenetic tree performed on the concatenated sequence of a CB domain and the Laminin G domain. E Phylogram of Laminin G domain based on 57 amino acid residues via LG model. F Phylogram of Laminin G domain based on 57 amino acid residues via Poisson model. All dichotomes are preserved. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*. The group formed by BMSPs are indicated by grey squares. Nodes A-B were indicated.


A.
K1Q802 Cgi larval SMP CGI 10000698
K1QM00 Cgi larval SMP CGI 10001795
K1QUE3 Cgi
K1R175 Cgi
K1QX66 Cgi
K1QWD1 Cgi
K1RJ02 Cgi adult SMP CGI 10014170
R9WGX8 Cgi
R9WDF3 Cgi
K1S2Z6 Cgi adult SMP CGI 10028495
R9WEQ4 Cgi
Pfu aug2.0 1907.1 25196.t1
Pfu aug2.0 7459.1 32953.t1
Pfu aug2.0 1301.1 04951.t1
Pfu aug2.0 1294.1 14936.t1 larval SMP
Pfu aug2.0 1536.1 21678.t1 larval SMP
Pfu aug2.0 214.1 13802.t1 adult SMP
Pfu aug2.0 214.1 13803.t1
K1QWG5 Cgi
Pfu aug2.0 4146.1 22513.t1
K1QSG0 Cgi
Pfu aug2.0 539.1 30879.t1
K1RCM3 Cgi
V4AHB6 Lgi
Pfu aug2.0 374.1 14069.t1
A0A0L8G2A1 Obi
Pfu aug2.0 920.1 01244.t2
V4C962 Lgi
Pfu aug2.0 551.1 04327.t1
Pfu aug2.0 551.1 04328.t1
K1P7Z9 Cgi
V4B4H5 Lgi
V3ZZS6 Lgi
A0A0L8HIH2 Obi
Pfu aug2.0 4343.1 12636.t1
V4A0P3 Lgi
V4BIL1 Lgi
K1QUN4 Cgi
Pfu aug2.0 2182.1 08768.t1
K1QML6 Cgi
A0A0L8GUI8 Obi
A0A0L8GAA7 Obi
B3A0P2 Lgi adult SMP Lgi 238082
V4CIB4 Lgi
B3A0Q6 Lgi
V4AJ64 Lgi adult SMP Lgi 239188
Pfu aug2.0 583.1 10956.t1
V4AH94 Lgi
K1QNU1 Cgi
V3ZJX8 Lgi
V4BHF8 Lgi
V3ZXK9 Lgi
V4AH85 Lgi
Pfu aug2.0 1849.1 31938.t1
V4AB29 Lgi
K1QC23 Cgi
Pfu aug2.0 107.1 23471.t1
Pfu aug2.0 1849.1 31939.t1
K1PH96 Cgi
K1QCX8 Cgi
A0A0L8GKL6 Obi
Q9FYE3 Ath

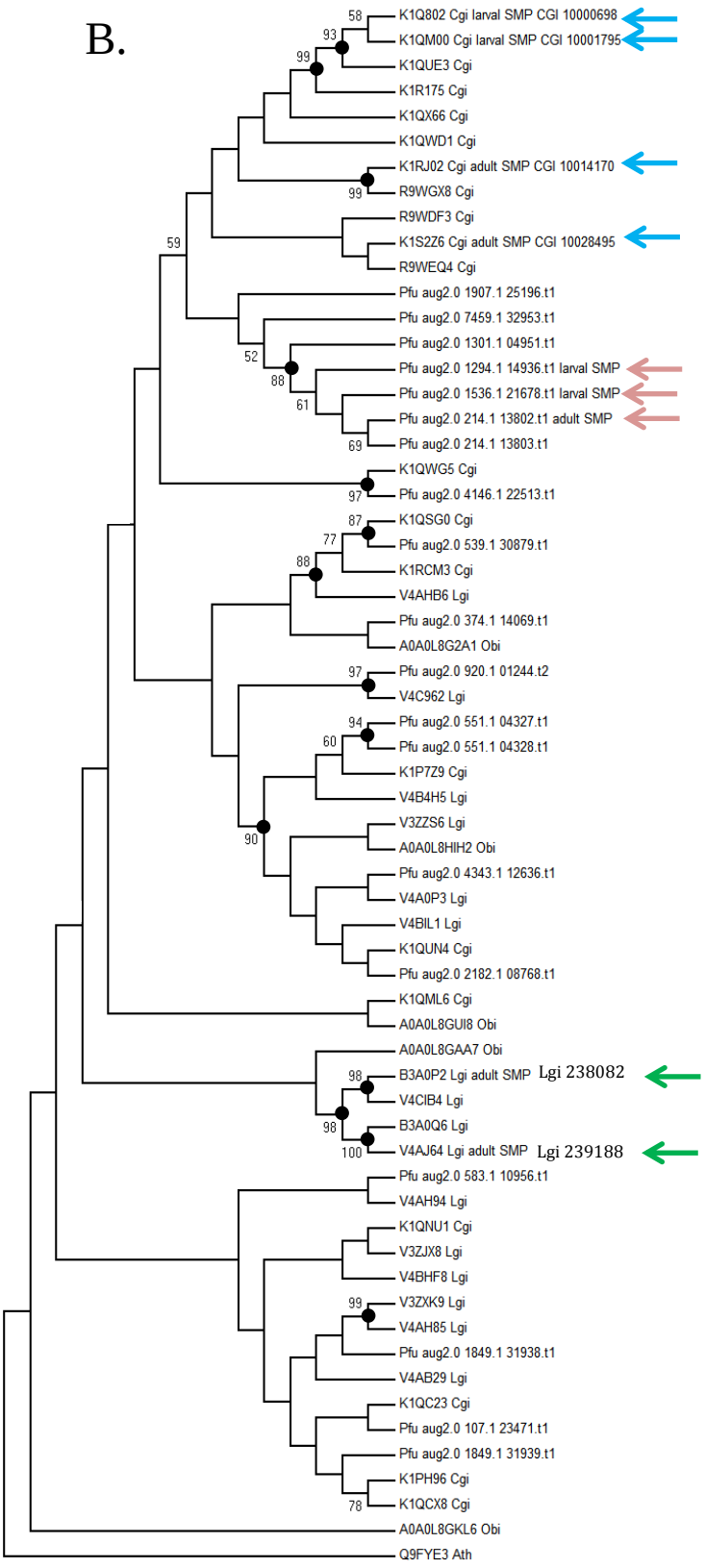
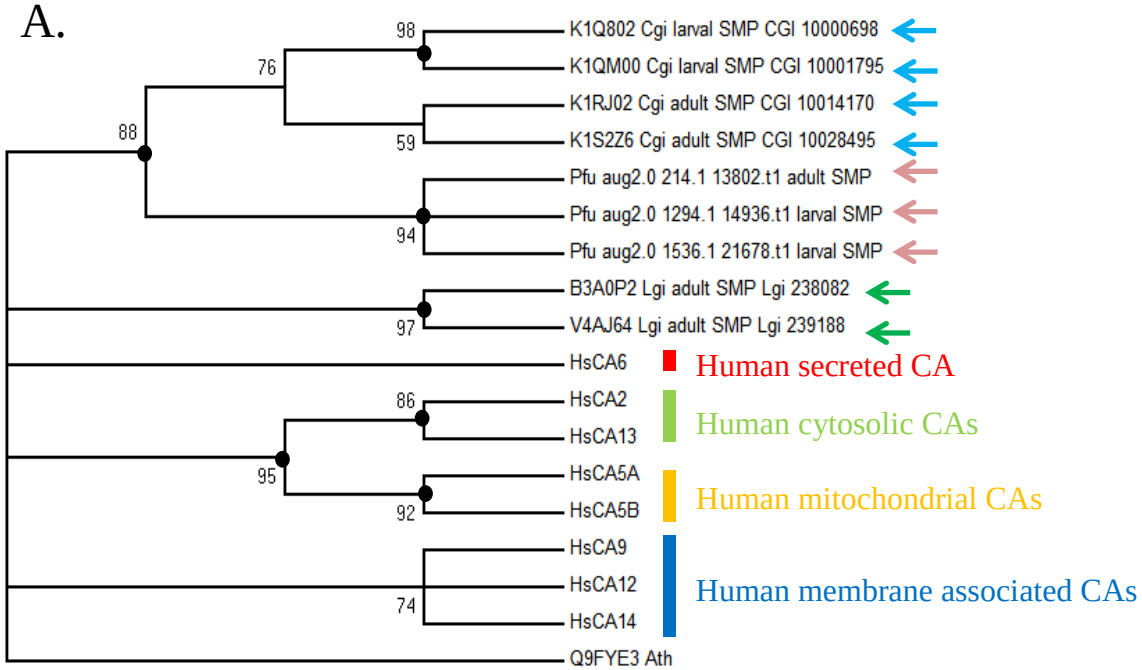
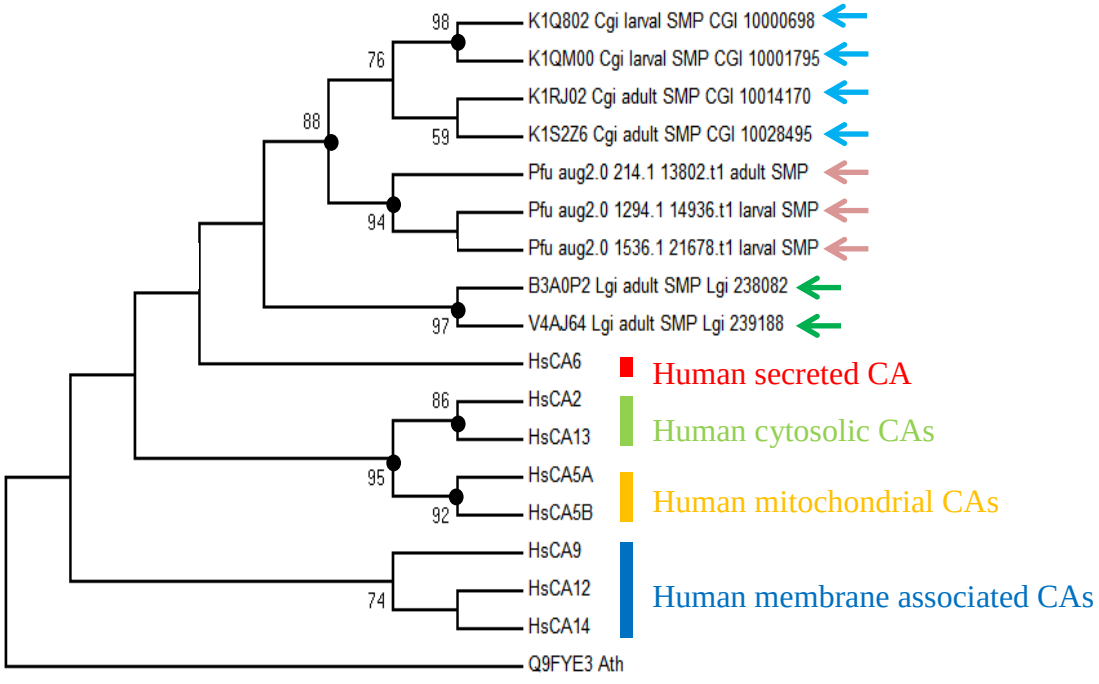


Fig. 3.11. Phylogenetic trees of CA domains in molluscs via Poisson model. A, Polychotomes are generated if the bootstrap value of a node is lower than 50. B, All dichotomes are preserved. Trees were generated on 82 amino acid residues. SMPs were indicated by blue (*C. gigas*), Pink (*P. fucata*) and green (*L. gigantea*) arrowheads. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides*. Ath, *Arabidopsis thaliana*.

A.


98
76
88
59
94
97
86
95
92
74
K1Q802 Cgi larval SMP CGI 10000698
K1QM00 Cgi larval SMP CGI 10001795
K1RJ02 Cgi adult SMP CGI 10014170
K1S2Z6 Cgi adult SMP CGI 10028495
Pfu aug2.0 214.1 13802.t1 adult SMP
Pfu aug2.0 1294.1 14936.t1 larval SMP
Pfu aug2.0 1536.1 21678.t1 larval SMP
B3A0P2 Lgi adult SMP Lgi 238082
V4AJ64 Lgi adult SMP Lgi 239188
HsCA6
HsCA2
HsCA13
HsCA5A
HsCA5B
HsCA9
HsCA12
HsCA14
Q9FYE3 Ath
Human secreted CA
Human cytosolic CAs
Human mitochondrial CAs
Human membrane associated CAs

B


98
76
88
59
94
97
86
95
92
74
K1Q802 Cgi larval SMP CGI 10000698
K1QM00 Cgi larval SMP CGI 10001795
K1RJ02 Cgi adult SMP CGI 10014170
K1S2Z6 Cgi adult SMP CGI 10028495
Pfu aug2.0 214.1 13802.t1 adult SMP
Pfu aug2.0 1294.1 14936.t1 larval SMP
Pfu aug2.0 1536.1 21678.t1 larval SMP
B3A0P2 Lgi adult SMP Lgi 238082
V4AJ64 Lgi adult SMP Lgi 239188
HsCA6
HsCA2
HsCA13
HsCA5A
HsCA5B
HsCA9
HsCA12
HsCA14
Q9FYE3 Ath
Human secreted CA
Human cytosolic CAs
Human mitochondrial CAs
Human membrane associated CAs

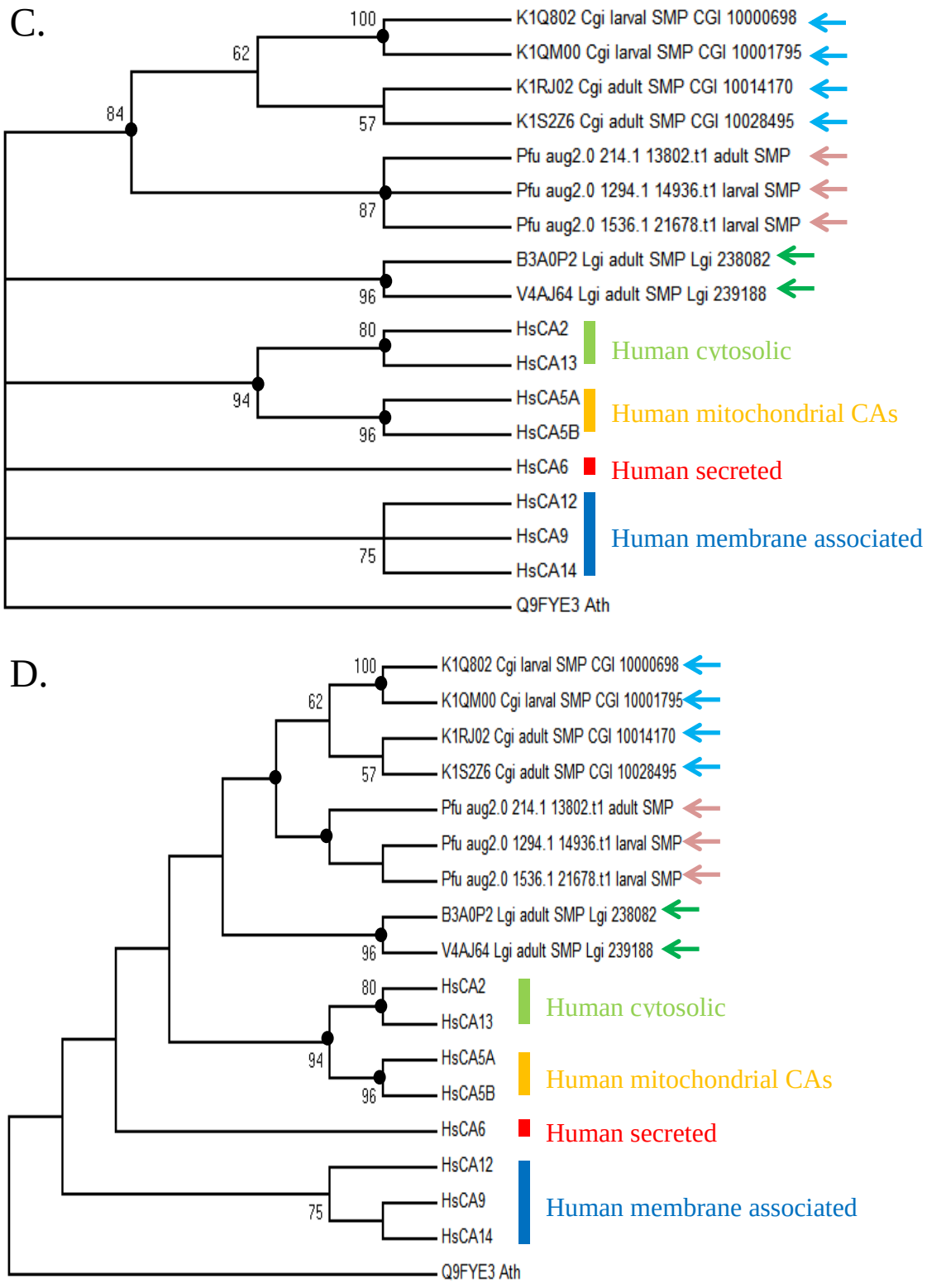


Fig. 3.12. Phylogenetic analyses of CAs in molluscan shells with CAs of human. Trees were generated on 55 amino acid residues based on LG (A, B) and Poisson (C, D) models, respectively. A, Polychotomes are generated if the bootstrap value of a node is lower than 50. B, All dichotomes are preserved. C, Polychotomes are generated if the bootstrap value of a node is lower than 50. D, All dichotomes are preserved. SMPs were indicated by blue (*C. gigas*), Pink (*P. fucata*) and green (*L. gigantea*) arrowheads. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Ath, *Arabidopsis thaliana*.

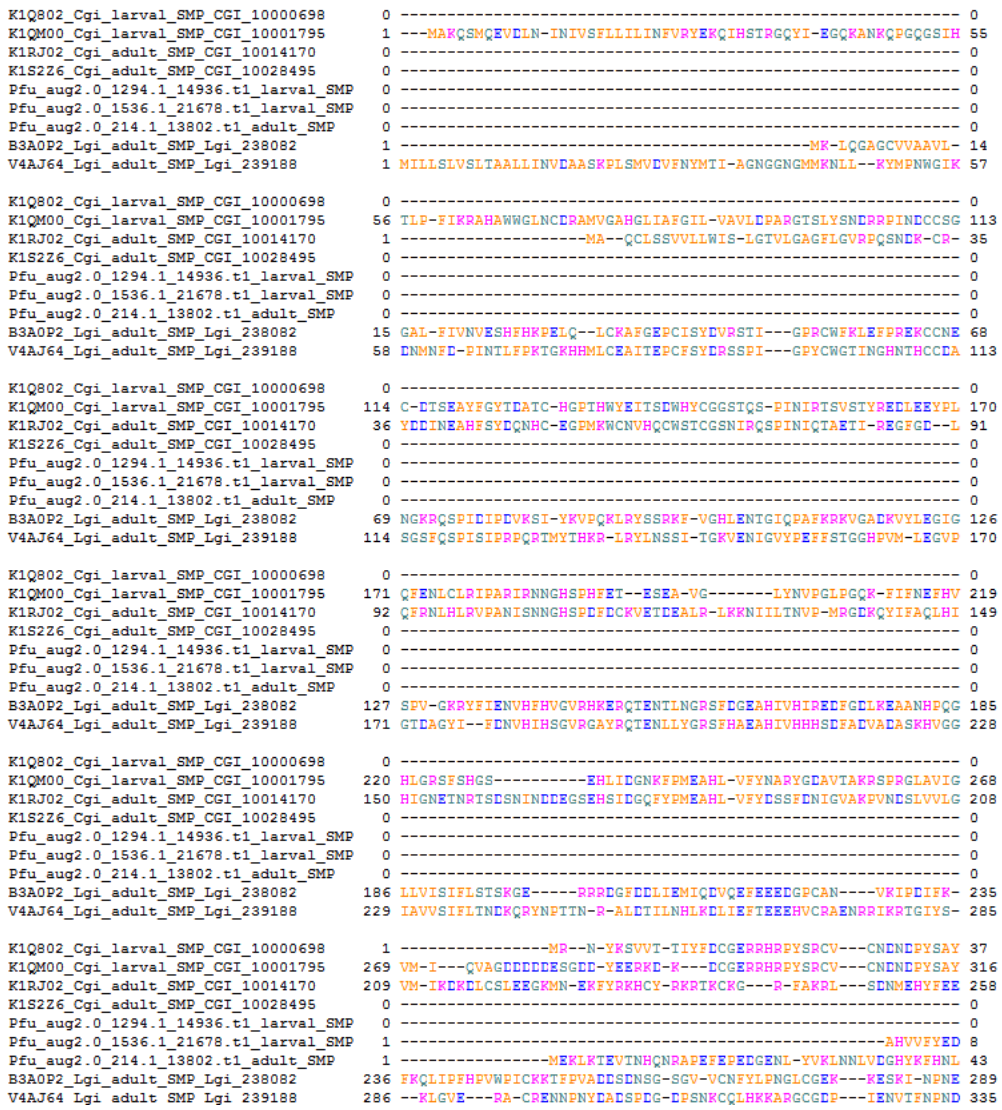

K1Q802_Cgi_larval_SMP_CGI_10000698 0 ---------- 0
K1QM00_Cgi_larval_SMP_CGI_10001795 1 ---MAKQSMQEVDLN-INIVSFLLILINFVRYEKQIHSTRGQYI-EGQKANKQPGQGSIH 55
K1RJ02_Cgi_adult_SMP_CGI_10014170 0 ---------- 0
K1S2Z6_Cgi_adult_SMP_CGI_10028495 0 ---------- 0
Pfu_aug2.0_1294.1_14936.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_1536.1_21678.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_214.1_13802.t1_adult_SMP 0 ---------- 0
B3A0P2_Lgi_adult_SMP_Lgi_238082 1 ----------MK-LQGAGCVVAAVL- 14
V4AJ64_Lgi_adult_SMP_Lgi_239188 1 MILLSLVSLTAALLINVDAASKPLSMVDVFNYMTI-AGNGGNGMMKNLL--KYMPNWGIK 57
K1Q802_Cgi_larval_SMP_CGI_10000698 0 ---------- 0
K1QM00_Cgi_larval_SMP_CGI_10001795 56 TLP-FIKRAHAWWGLNCDRAMVGAHGLIAFGIL-VAVLDPARGTSLYSNDRRPINDCCSG 113
K1RJ02_Cgi_adult_SMP_CGI_10014170 1 ----------MA--QCLSSVVLLWIS-LGTVLGAGFLGVRPQSNDK-CR- 35
K1S2Z6_Cgi_adult_SMP_CGI_10028495 0 ---------- 0
Pfu_aug2.0_1294.1_14936.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_1536.1_21678.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_214.1_13802.t1_adult_SMP 0 ---------- 0
B3A0P2_Lgi_adult_SMP_Lgi_238082 15 GAL-FIVNVESHFHKPELQ--LCKAFGEPCISYDVRSTI---GPRCWFKLEFPREKCCNE 68
V4AJ64_Lgi_adult_SMP_Lgi_239188 58 DNMNFD-PINILFPKTGKHHMLCEAITEPCFSYDRSSPI---GPYCWGTINGHNTHCCDA 113
K1Q802_Cgi_larval_SMP_CGI_10000698 0 ---------- 0
K1QM00_Cgi_larval_SMP_CGI_10001795 114 C-DTSEAYFGYTDATC-HGPTHWYEITSDWHYCGGSTQS-PINIRTSVSTYREDLEEYPL 170
K1RJ02_Cgi_adult_SMP_CGI_10014170 36 YDDINEAHFSYDQNHC-EGPMKWCNVHQCWSTCGSNIRQSPINIQTAETI-REGFGD--L 91
K1S2Z6_Cgi_adult_SMP_CGI_10028495 0 ---------- 0
Pfu_aug2.0_1294.1_14936.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_1536.1_21678.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_214.1_13802.t1_adult_SMP 0 ---------- 0
B3A0P2_Lgi_adult_SMP_Lgi_238082 69 NGKRQSPIDIPDVKSI-YKVPQKLRYSSRKF-VGHLENTGIQPAFKRKVGADKVYLEGIG 126
V4AJ64_Lgi_adult_SMP_Lgi_239188 114 SGSFQSPISIPRPQRTMYTHKR-LRYLNSSI-TGKVENIGVYPEFFSTGGHPVM-LEGVP 170
K1Q802_Cgi_larval_SMP_CGI_10000698 0 ---------- 0
K1QM00_Cgi_larval_SMP_CGI_10001795 171 QFENLCLRIPARIRNNGHSPHFET--ESEA-VG-------LYNVPGLPGQK-FIFNEFHV 219
K1RJ02_Cgi_adult_SMP_CGI_10014170 92 QFRNLHLRVPANISNNGHSPDFDCKVETDEALR-LKKNIILTNVP-MRGDKQYIFAQLHI 149
K1S2Z6_Cgi_adult_SMP_CGI_10028495 0 ---------- 0
Pfu_aug2.0_1294.1_14936.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_1536.1_21678.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_214.1_13802.t1_adult_SMP 0 ---------- 0
B3A0P2_Lgi_adult_SMP_Lgi_238082 127 SPV-GKRYFIENVHFHVGVRHKERQTENTLNGRSFDGEAHIVHIRDDFGDLKEAANHPQG 185
V4AJ64_Lgi_adult_SMP_Lgi_239188 171 GTDAGYI--FDNVHIHSGVRGAYPQTENLLYGRSFHAEAHIVHHHSDFADVADASKHVGG 228
K1Q802_Cgi_larval_SMP_CGI_10000698 0 ---------- 0
K1QM00_Cgi_larval_SMP_CGI_10001795 220 HLGRSFSHGS----------EHLIDGNKFPMEAHL-VFYNARYGDAVIAKRSPRGLAVIG 268
K1RJ02_Cgi_adult_SMP_CGI_10014170 150 HIGNETNRTSDSNINDDEGSEHSIDGQFYPMEAHL-VFYDSSFDNIGVAKPVNDSLVVLG 208
K1S2Z6_Cgi_adult_SMP_CGI_10028495 0 ---------- 0
Pfu_aug2.0_1294.1_14936.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_1536.1_21678.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_214.1_13802.t1_adult_SMP 0 ---------- 0
B3A0P2_Lgi_adult_SMP_Lgi_238082 186 LLVISIFLSTSKGE-----RRRDGFDDLIEMIQDVQEFEEEDGPCAN----VKIPDIFK- 235
V4AJ64_Lgi_adult_SMP_Lgi_239188 229 IAVVSIFLTNDKQRYNPTTN-R-ALDTILNHLKDLIEFTEEEHVCRAENRRIKRTGIYS- 285
K1Q802_Cgi_larval_SMP_CGI_10000698 1 ----------MR--N-YKSVVT-TIYFDCGERRHRPYSRCV---CNDNDPYSAY 37
K1QM00_Cgi_larval_SMP_CGI_10001795 269 VM-I---QVAGDDDDESGDD-YEERKD-K---DCGERRHRPYSRCV---CNDNDPYSAY 316
K1RJ02_Cgi_adult_SMP_CGI_10014170 209 VM-IKDKDLCSLEEGKMN-EKFYRKHCY-RKRTKCKG---R-FAKRL---SDNMEHYFEE 258
K1S2Z6_Cgi_adult_SMP_CGI_10028495 0 ---------- 0
Pfu_aug2.0_1294.1_14936.t1_larval_SMP 0 ---------- 0
Pfu_aug2.0_1536.1_21678.t1_larval_SMP 1 ----------AHVVFYED 8
Pfu_aug2.0_214.1_13802.t1_adult_SMP 1 ----------MEKLKTEVTNHQNRAPEFEPEDGENL-YVKLNNLVDGHYKFHNL 43
B3A0P2_Lgi_adult_SMP_Lgi_238082 236 FKQLIPFHPVWPICKKTFPVADDSDNSG-SGV-VCNFYLPNGLCGEK---KESKI-NPNE 289
V4AJ64_Lgi_adult_SMP_Lgi_239188 286 --KLGVE---RA-CPENNPNYDADSPDG-DPSNKCQLHKKARGCGDP---IENVTFNPND 335

Fig. 3.13. Alignment of molluscan CAs identified in the shells. The NG-repeat domains are indicated by the box. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*.

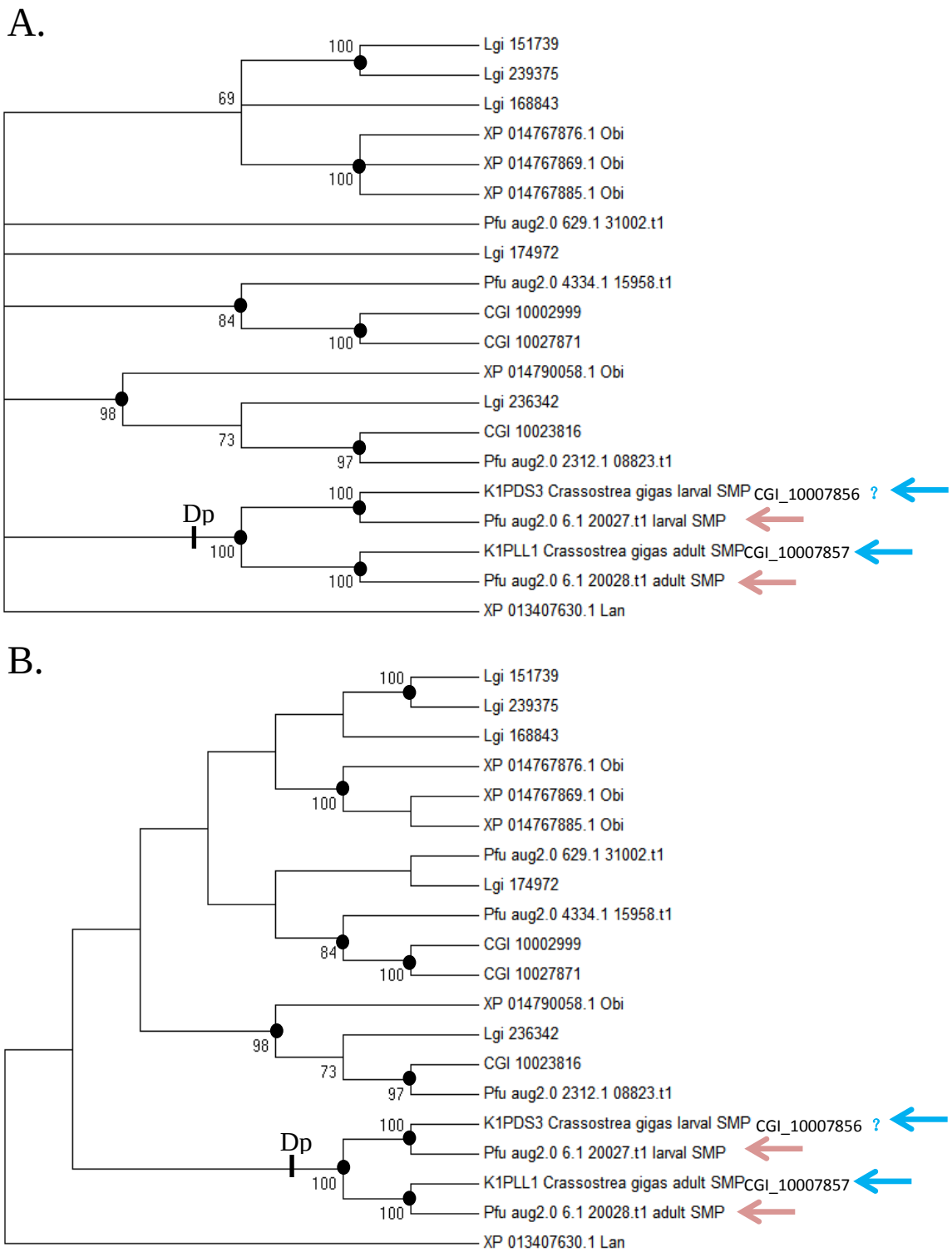
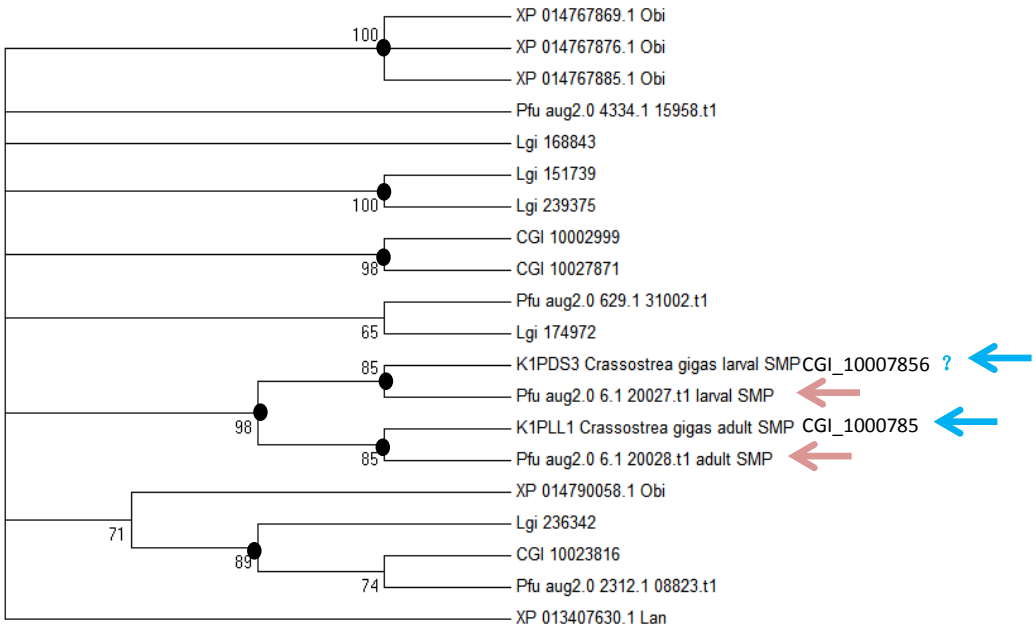
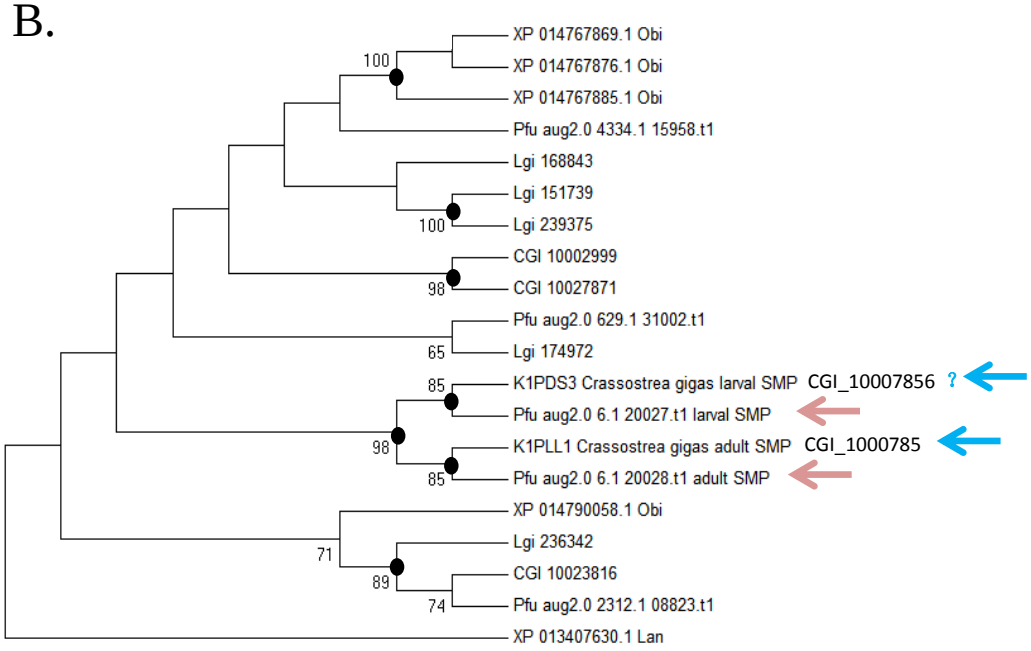


Fig. 3.14. Phylogenetic trees of chitobiases in molluscs via Poisson model on concatenated sequence of CHB_HEX domain (IPR004866), Glyco_hydro_20b domain (IPR015882), Glyco_hydro_20 domain (IPR015883) and CHB_HEX_C domain (IPR004867). A, Polychotomes are generated if the bootstrap value of a node is lower than 50. B, All dichotomes are preserved. Trees were generated on 523 amino acid residues. SMPs were indicated by blue (*C. gigas*) and Pink (*P. fucata*) arrowheads. Question mark indicates if the gene is a SMP is questionable. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides.* Lan, *Lingular anatina*.

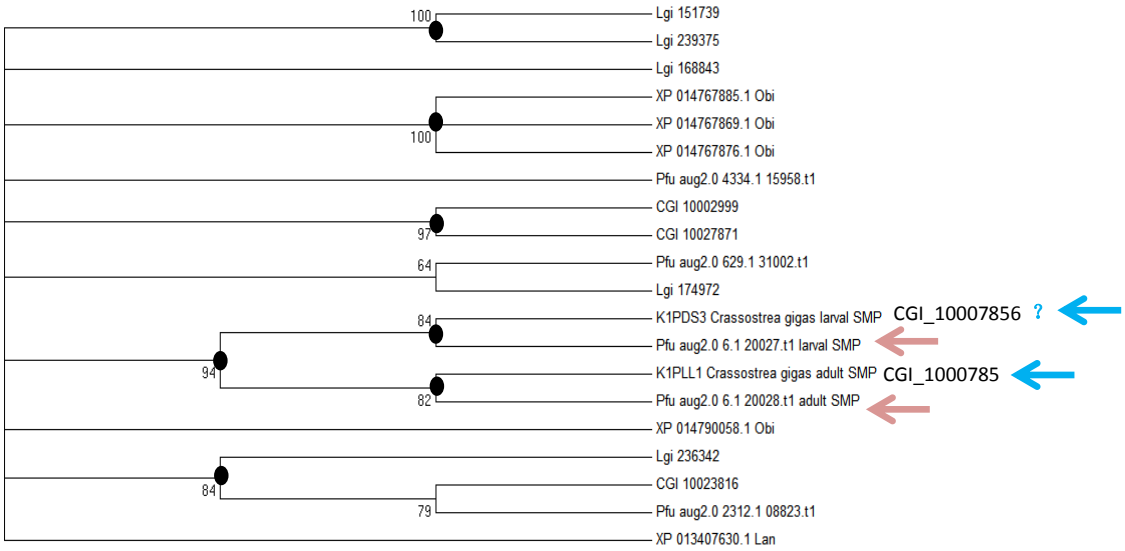
A.


XP 014767869.1 Obi
XP 014767876.1 Obi
XP 014767885.1 Obi
Pfu aug2.0 4334.1 15958.t1
Lgi 168843
Lgi 151739
Lgi 239375
CGI 10002999
CGI 10027871
Pfu aug2.0 629.1 31002.t1
Lgi 174972
K1PDS3 Crassostrea gigas larval SMP
CGI_10007856 ?
Pfu aug2.0 6.1 20027.t1 larval SMP
K1PLL1 Crassostrea gigas adult SMP
CGI_1000785
Pfu aug2.0 6.1 20028.t1 adult SMP
XP 014790058.1 Obi
Lgi 236342
CGI 10023816
Pfu aug2.0 2312.1 08823.t1
XP 013407630.1 Lan
100
100
98
65
85
98
85
71
89
74

B.


XP 014767869.1 Obi
XP 014767876.1 Obi
XP 014767885.1 Obi
Pfu aug2.0 4334.1 15958.t1
Lgi 168843
Lgi 151739
Lgi 239375
CGI 10002999
CGI 10027871
Pfu aug2.0 629.1 31002.t1
Lgi 174972
K1PDS3 Crassostrea gigas larval SMP
CGI_10007856 ?
Pfu aug2.0 6.1 20027.t1 larval SMP
K1PLL1 Crassostrea gigas adult SMP
CGI_1000785
Pfu aug2.0 6.1 20028.t1 adult SMP
XP 014790058.1 Obi
Lgi 236342
CGI 10023816
Pfu aug2.0 2312.1 08823.t1
XP 013407630.1 Lan
100
100
98
65
85
98
85
71
89
74

C.



D.

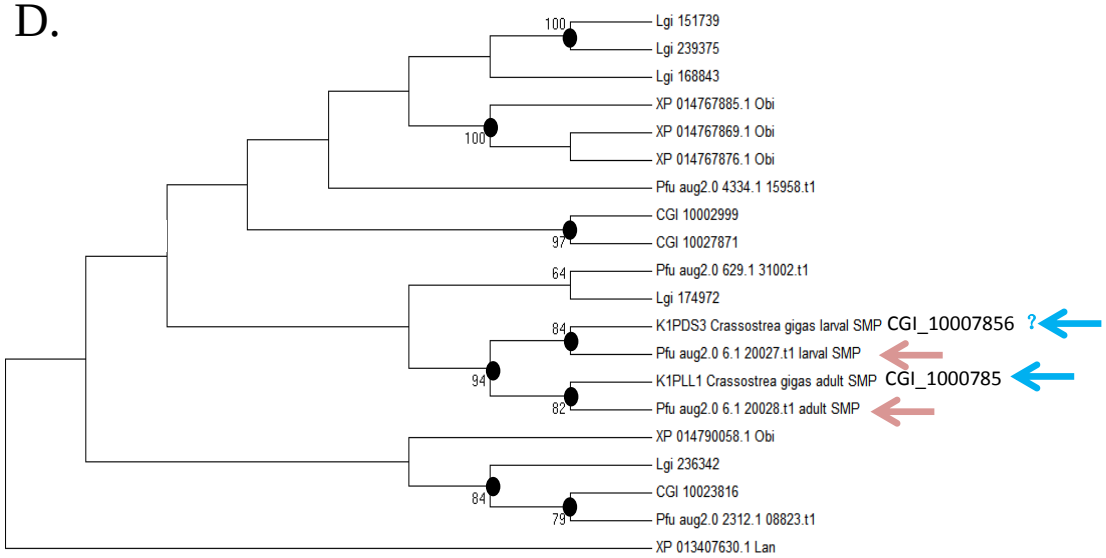


Fig. 3.15. Phylogenetic trees of chitobiases in molluscs via LG (A, B) and Poisson (C, D) models on concatenated sequence of CHB_HEX domain (IPR004866), Glyco_hydro_20b domain (IPR015882), Glyco_hydro_20 domain (IPR015883) and CHB_HEX_C domain (IPR004867). A, Polychotomes are generated if the bootstrap value of a node is lower than 50. B, All dichotomes are preserved. C, Polychotomes are generated if the bootstrap value of a node is lower than 50. D, All dichotomes are preserved. Trees were generated on 106 amino acid residues. SMPs were indicated by blue (*C. gigas*) and Pink (*P. fucata*) arrowheads. Question mark indicates if the gene is a SMP is questionable. Bootstrap values are shown if over 50, and marked with black dots if over 80. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides.* Lan, *Lingular anatina*.

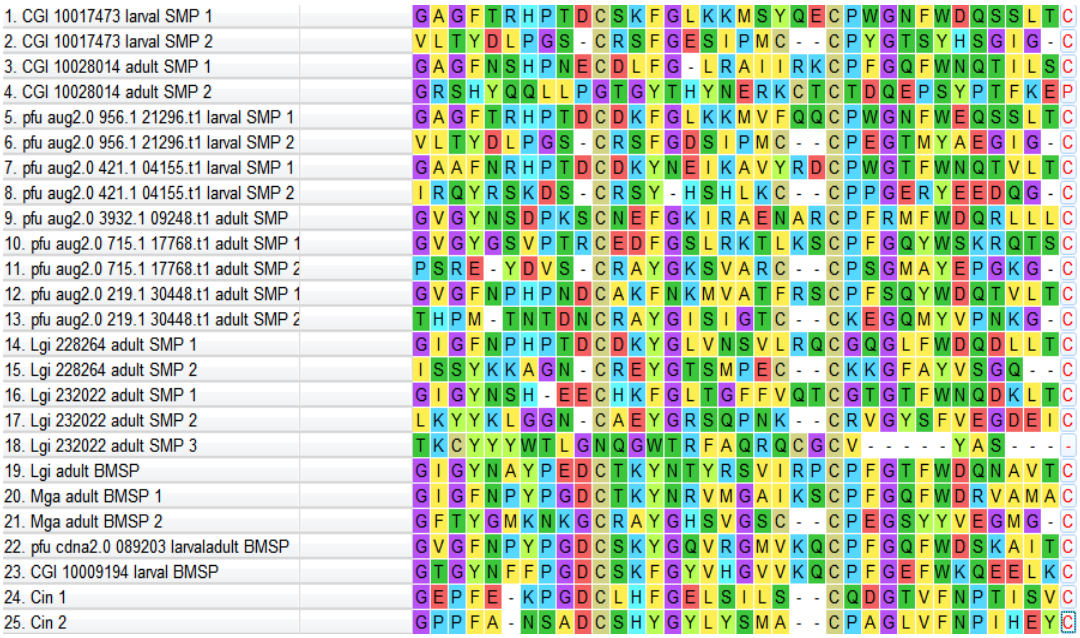


Fig. 3.16. Alignment of molluscan CB domains with that of *C. intestinalis* on 37 amino acid residues. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Cin,

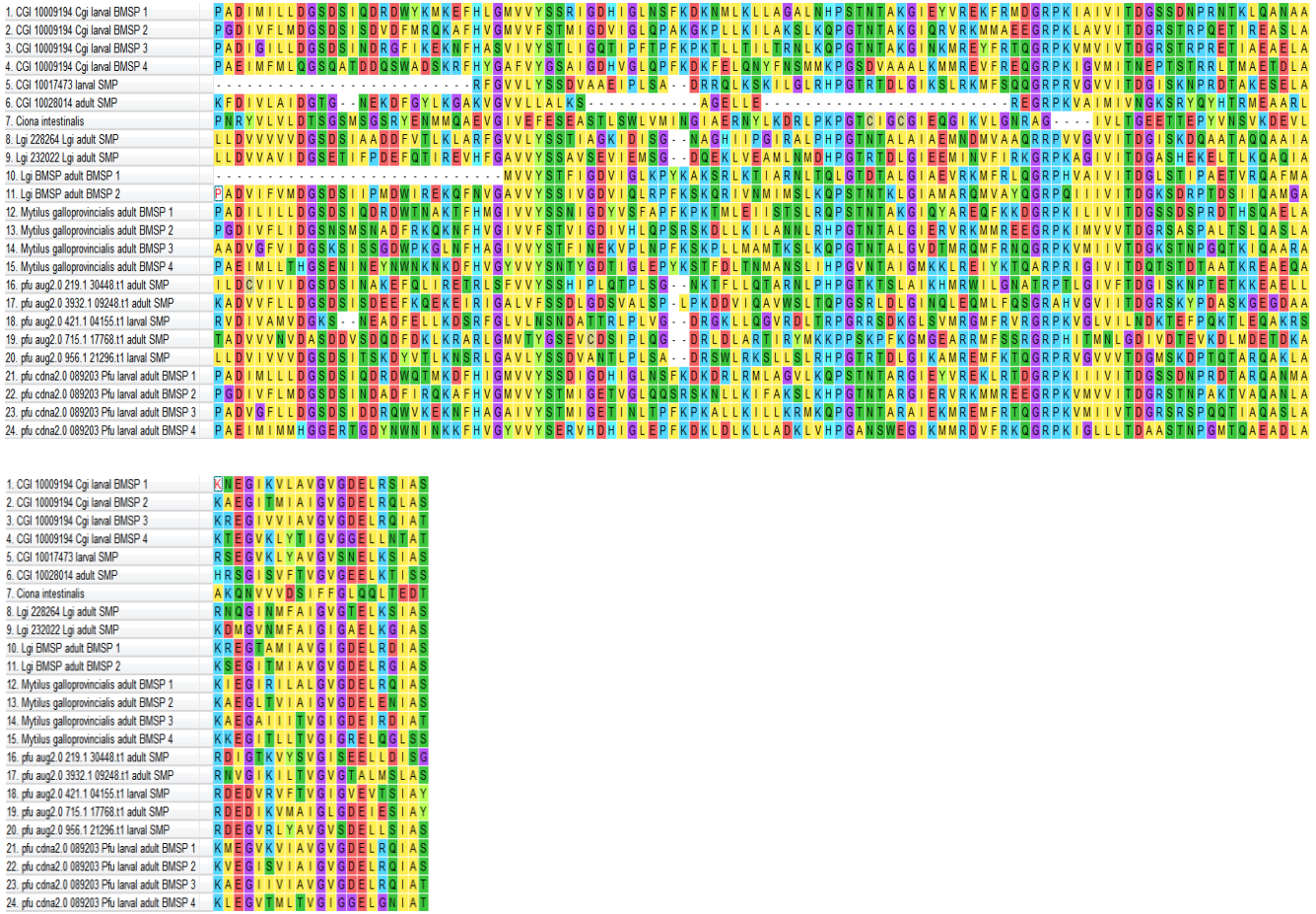


Fig. 3.17. Alignment of molluscan VWA domains with that of *C. intestinalis* on 127 amino acid residues. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*.

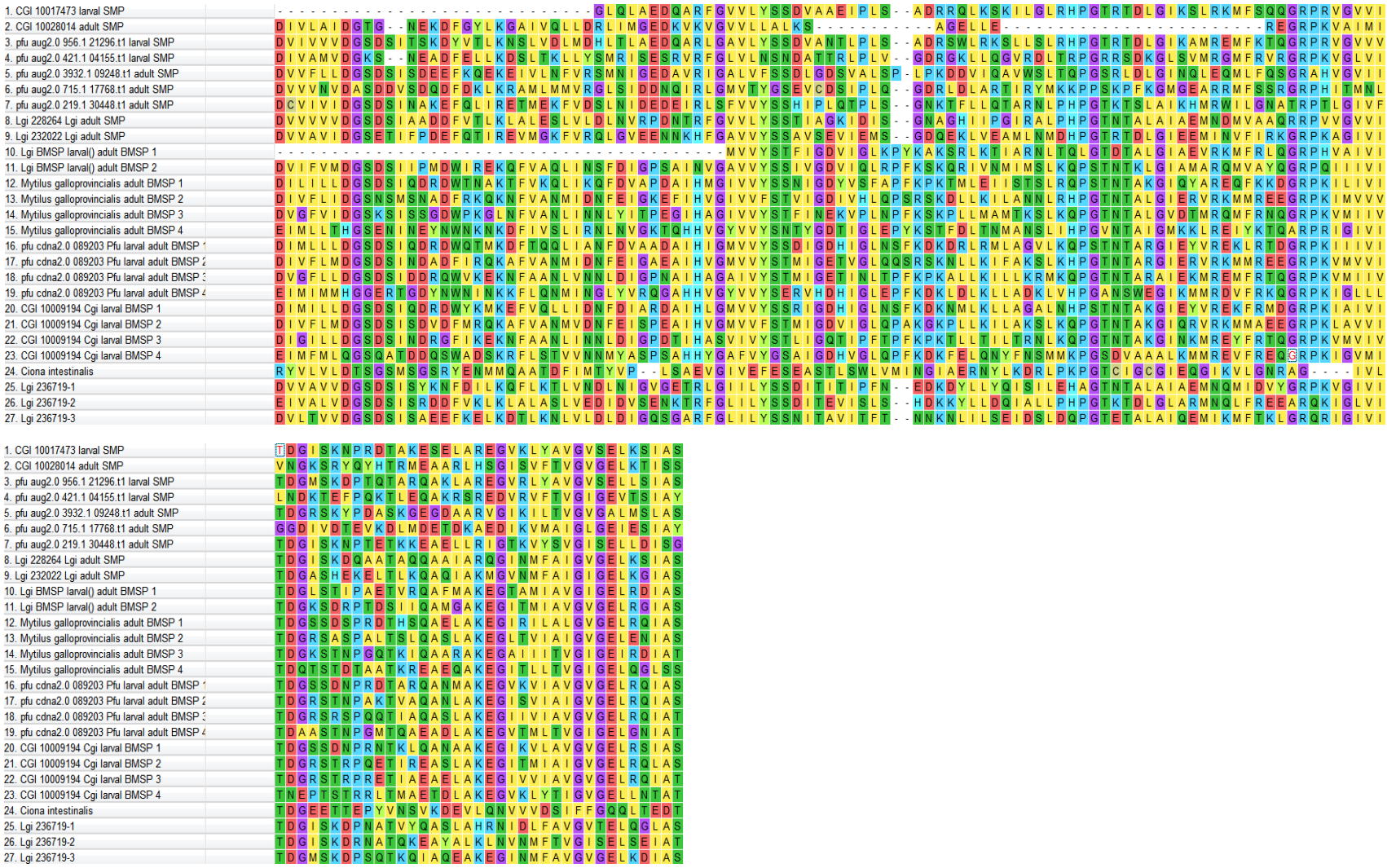


Fig. 3.18. Alignment of VWA domains of VWA-CB dcps of molluscan shells and LG236719 on 138 amino acid residues. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*.

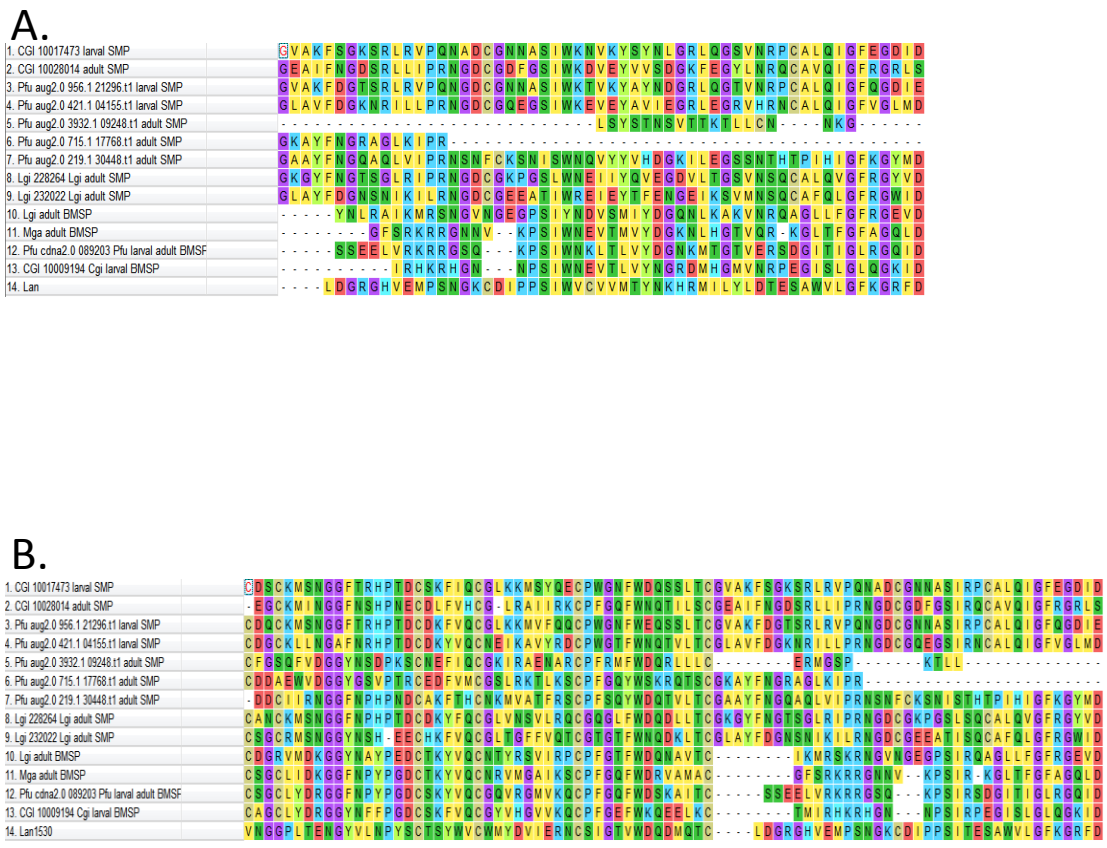


Fig. 3.19. Alignments of Laminin G domains on 57 amino acid residues (A) and the concatenated sequence of CB domain and Laminin G domain on 86 amino acid residues (B). CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*.

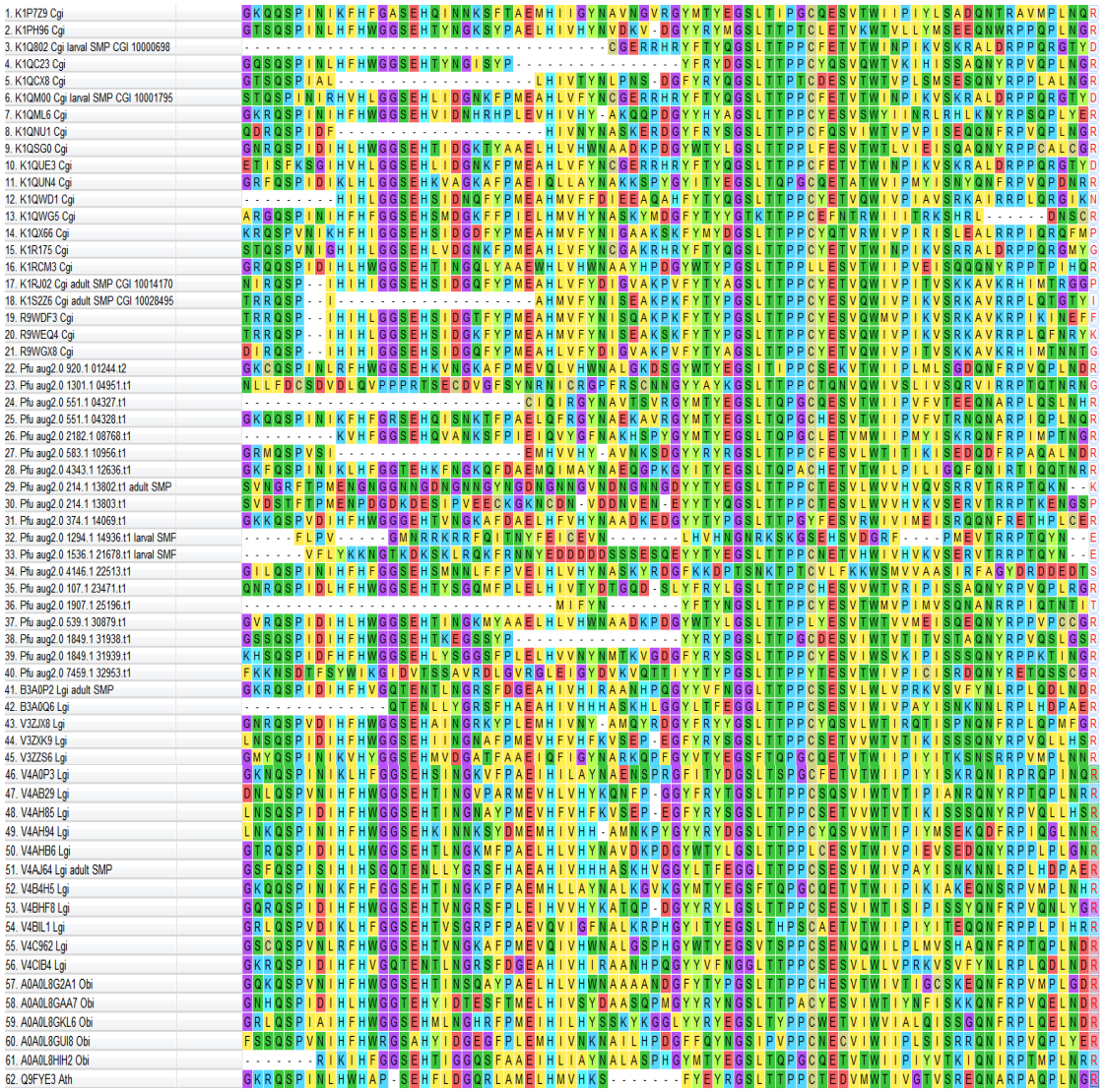


Fig. 3.20. Alignment of molluscan CA domains with that of *Arabidopsis thaliana* on 82 amino acid residues. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides.* Ath, *Arabidopsis thaliana*.

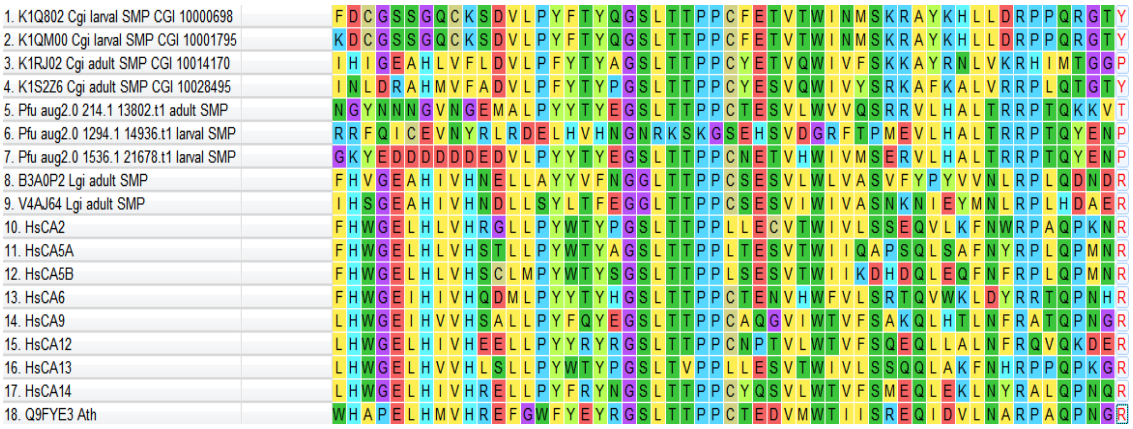
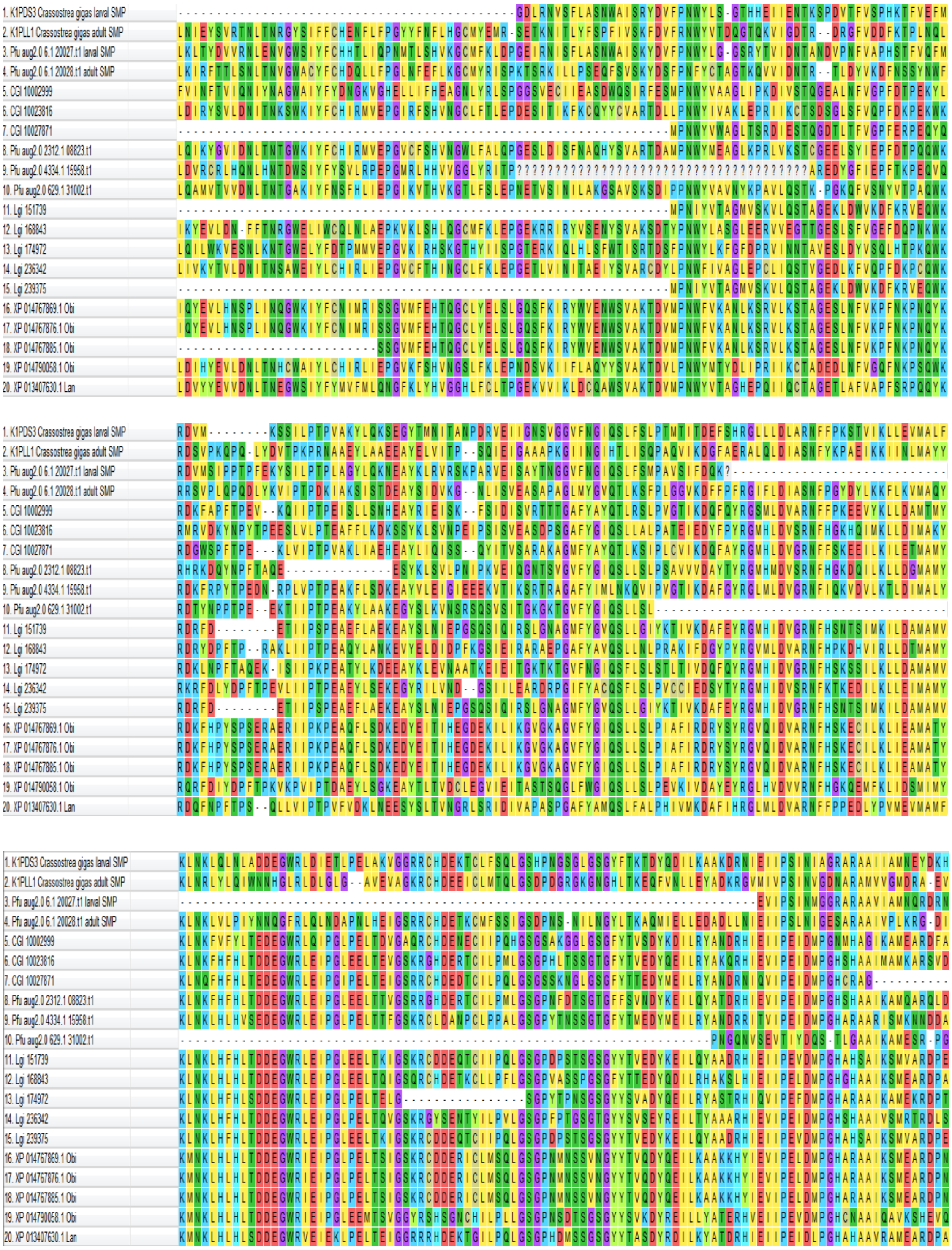


Fig. 3.21. Alignment of CA domains of molluscs and human with that of *Arabidopsis thaliana* on 55 amino acid residues. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides.* Ath, *Arabidopsis thaliana*.


1. K1PDS3 Crassostrea gigas larval SMP
2. K1PLL1 Crassostrea gigas adult SMP
3. Pfu aug2.0 6.1 20027.t1 larval SMP
4. Pfu aug2.0 6.1 20028.t1 adult SMP
5. CGI 10002999
6. CGI 10023816
7. CGI 10027871
8. Pfu aug2.0 2312.1 08823.t1
9. Pfu aug2.0 4334.1 15958.t1
10. Pfu aug2.0 629.1 31002.t1
11. Lgi 151739
12. Lgi 168843
13. Lgi 174972
14. Lgi 236342
15. Lgi 239375
16. XP 014767869.1 Obi
17. XP 014767876.1 Obi
18. XP 014767885.1 Obi
19. XP 014790058.1 Obi
20. XP 013407630.1 Lan

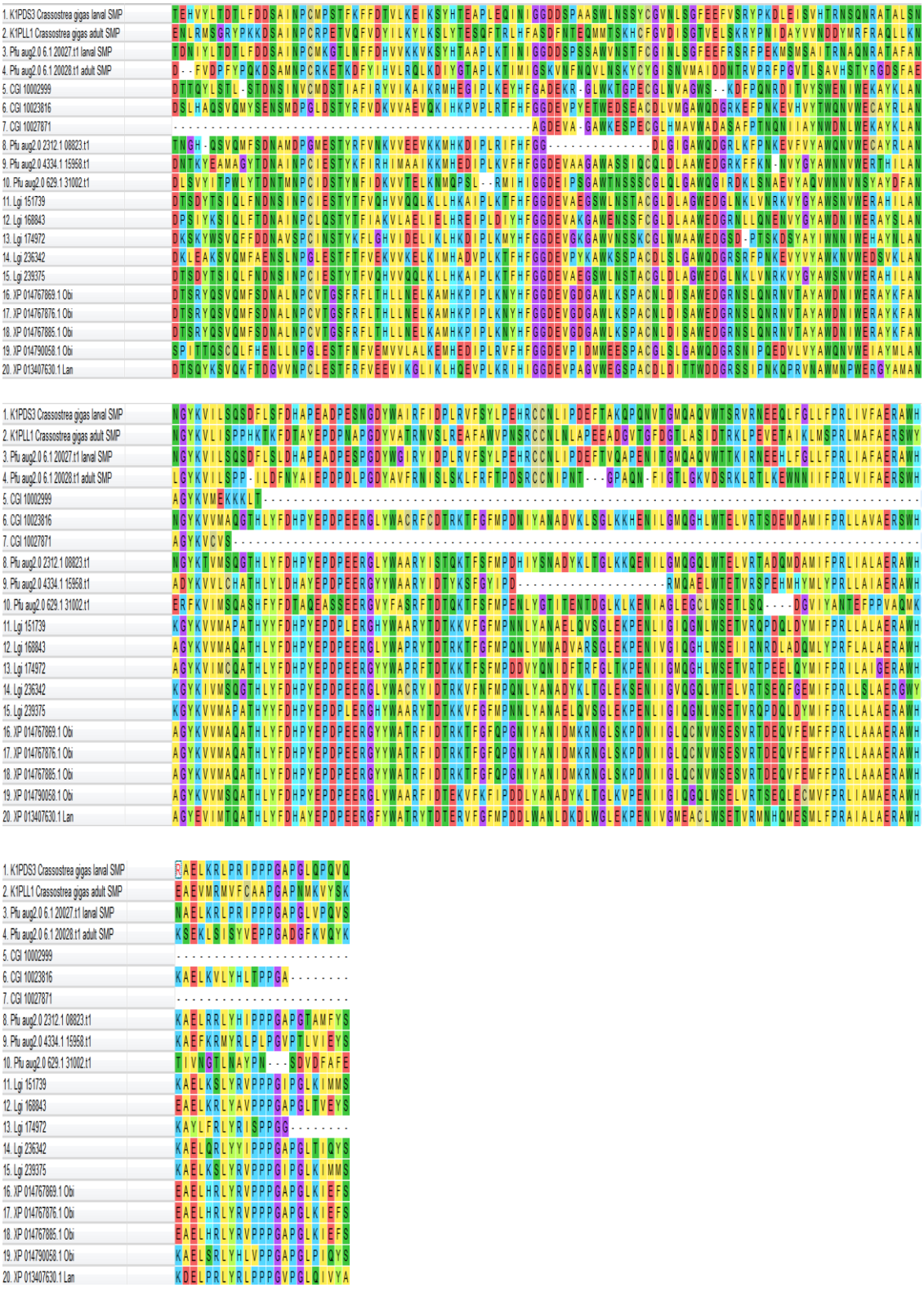


Fig. 3.22. Alignment of concatenated sequence of CHB_HEX domain (IPR004866), Glyco_hydro_20b domain (IPR015882), Glyco_hydro_20 domain (IPR015883) and CHB_HEX_C domain (IPR004867) on 523 amino acid residues. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides*. Lan, *Lingula anatina*.

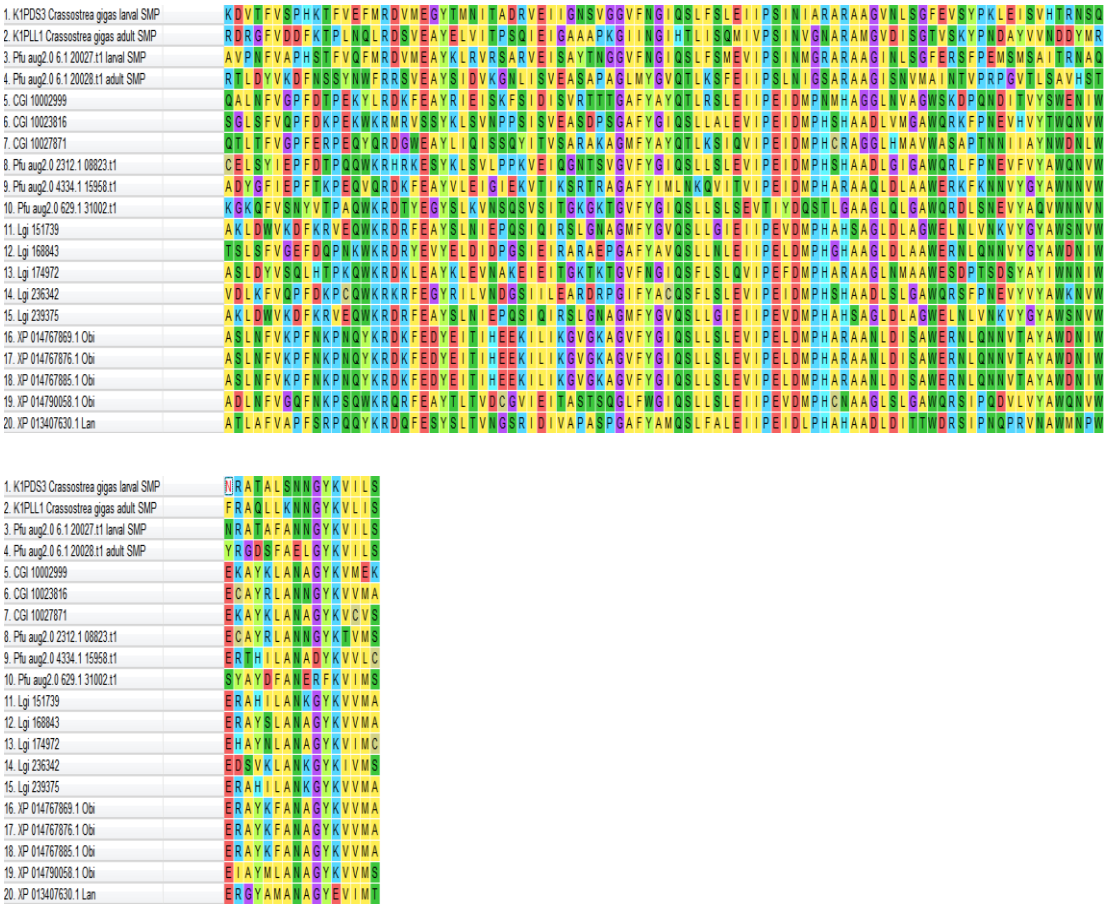


Fig. 3.23. Alignment of concatenated sequence of CHB_HEX domain (IPR004866), Glyco_hydro_20b domain (IPR015882), Glyco_hydro_20 domain (IPR015883) and CHB_HEX_C domain (IPR004867) on 106 amino acid residues by removing all gaps. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides*. Lan, *Lingula anatina.*

Table 3.1 List of VWA-CB dcps in this study.

Protein	Larval SMP	Adult SMP	Species	Accession code	Reference/Resource
PIF	Yes		*Crassostrea gigas*	CGI_10017473	Zhao, et al. 2018
PIF		Yes	*Crassostrea gigas*	CGI_10028014	Zhang, et al. 2012
BMSP	Yes		*Crassostrea gigas*	CGI_10009194	Zhao, et al. 2018
PIF	Yes		*Pinctada fucata*	Pfu_aug2.0_956.1_21296.t1	Zhao, et al. 2018
PIF	Yes		*Pinctada fucata*	Pfu_aug2.0_421.1_04155.t1	Zhao, et al. 2018
PIF		Yes	*Pinctada fucata*	Pfu_aug2.0_3932.1_09248.t1	Zhao, et al. 2018
PIF		Yes	*Pinctada fucata*	Pfu_aug2.0_715.1_17768.t1	Zhao, et al. 2018
Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1		Yes	*Pinctada fucata*	Pfu_aug2.0_219.1_30448.t1	Zhao, et al. 2018
Collagen alpha-5(VI) chain	Yes	Yes	*Pinctada fucata*	Pfu_cdna2.0_089203	Zhao, et al. 2018
BMSP		Yes	*Mitulus galloprovincialis*	BAK86420.1	Suzuki, et al. 2011
Pif/BMSP-like protein		Yes	*Lottia gigantea*	Lgi_228264	Mann, et al. 2012
Pif/BMSP-like protein		Yes	*Lottia gigantea*	Lgi_232022	Mann, et al. 2012
BMSP		Yes	*Lottia gigantea*	Lgi_173137, 173138, 173139	Marie, et al. 2013
Calcium-activated chloride channel regulator 4-like			*Ciona intestinalis*	A0A1W3JQS8	InterProScan database

Table 3.2 List of proteins containing carbonic anhydrase domain in this study.

Protein	Larval SMP	Adult SMP	Species	Accession code	Reference/Resource
Carbonic anhydrase	Yes		*Crassostrea gigas*	K1Q802/CGI_10000698	Zhang, et al. 2012
Carbonic anhydrase	Yes		*Crassostrea gigas*	K1QM00/CGI_10001795	Zhang, et al. 2012
Carbonic anhydrase		Yes	*Crassostrea gigas*	K1RJ02/CGI_10014170	Zhang, et al. 2012
Carbonic anhydrase		Yes	*Crassostrea gigas*	K1S2Z6/CGI_10028495	Zhang, et al. 2012
Uncharacterized protein			*Crassostrea gigas*	K1P7Z9	Zhang, et al. 2012
Carbonic anhydrase			*Crassostrea gigas*	K1PH96	Zhang, et al. 2012
Carbonic anhydrase 1			*Crassostrea gigas*	K1QC23	Zhang, et al. 2012
Carbonic anhydrase 12			*Crassostrea gigas*	K1QCX8	Zhang, et al. 2012
Carbonic anhydrase 2			*Crassostrea gigas*	K1QNU1	Zhang, et al. 2012
Carbonic anhydrase 2			*Crassostrea gigas*	K1QSG0	Zhang, et al. 2012
Nacrein-like protein P2			*Crassostrea gigas*	K1QUE3	Zhang, et al. 2012
DnaJ-like protein subfamily C member 7			*Crassostrea gigas*	K1QUN4	Zhang, et al. 2012
Carbonic anhydrase 2			*Crassostrea gigas*	K1QWD1	Zhang, et al. 2012
Carbonic anhydrase 2			*Crassostrea gigas*	K1QX66	Zhang, et al. 2012
Carbonic anhydrase 7			*Crassostrea gigas*	K1R175	Zhang, et al. 2012
Carbonic anhydrase 2			*Crassostrea gigas*	K1RCM3	Zhang, et al. 2012
Carbonic anhydrase 13			*Crassostrea gigas*	K1RKE6	Zhang, et al. 2012
Nacrein-like protein F2			*Crassostrea gigas*	R9WDF3	Song, et al. 2014
Nacrein-like protein			*Crassostrea gigas*	R9WEQ4	Song and Wang, 2013
Nacrein-like protein F1			*Crassostrea gigas*	R9WGX8	Song, et al. 2014
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_920.1_01244.t2	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_1301.1_04951.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_551.1_04327.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_551.1_04328.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_2182.1_08768.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_583.1_10956.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_4343.1_12636.t1	Takeuchi, et al. 2016
Uncharacterized protein		Yes	*Pinctada fucata*	Pfu_aug2.0_214.1_13802.t1	Zhao, et al. 2018
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_214.1_13803.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_374.1_14069.t1	Takeuchi, et al. 2016
Carbonic anhydrase	YES		*Pinctada fucata*	Pfu_aug2.0_1294.1_14936.t1	Zhao, et al. 2018
Carbonic anhydrase	YES		*Pinctada fucata*	Pfu_aug2.0_1536.1_21678.t1	Zhao, et al. 2018
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_4146.1_22513.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_107.1_23471.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_1907.1_25196.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_539.1_30879.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_1849.1_31938.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_1849.1_31939.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	Pfu_aug2.0_7459.1_32953.t1	Takeuchi, et al. 2016
Putative carbonic		Yes	*Lottia gigantea*	B3A0P2	Marie, et al. 2013

anhydrase 1				
Putative carbonic anhydrase 2		*Lottia gigantea*	B3A0Q6	Marie, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V3ZJX8	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V3ZXK9	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V3ZZS6	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4A0P3	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4AB29	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4AH85	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4AH94	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4AHB6	Simakov, et al. 2013
Uncharacterized protein	Yes	*Lottia gigantea*	V4AJ64	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4B4H5	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4BHF8	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4BIL1	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4C962	Simakov, et al. 2013
Uncharacterized protein		*Lottia gigantea*	V4CIB4	Simakov, et al. 2013
Uncharacterized protein		*Octopus bimaculoides*	A0A0L8G2A1	Albertin, et al. 2015
Uncharacterized protein		*Octopus bimaculoides*	A0A0L8GAA7	Albertin, et al. 2015
Uncharacterized protein		*Octopus bimaculoides*	A0A0L8GKL6	Albertin, et al. 2015
Uncharacterized protein		*Octopus bimaculoides*	A0A0L8GUI8	Albertin, et al. 2015
Uncharacterized protein		*Octopus bimaculoides*	A0A0L8HIH2	Albertin, et al. 2015
Alpha carbonate dehydratase 3		*Arabidopsis thaliana*	Q9FYE3	Institute, Kazusa DNA Research, et al. 2000
CA2		*Homo sapiens*	P00918	Murakami, et al. 1987
CA5A		*Homo sapiens*	P35218	Nagao, et al. 1993
CA5B		*Homo sapiens*	Q9Y2D0	Fujikawa-Adachi, et al. 1999
CA6		*Homo sapiens*	P23280	Aldred, et al. 1991
CA9		*Homo sapiens*	Q16790	Pastorek, et al. 1994
CA12		*Homo sapiens*	O43570	Ota, et al. 2004
CA13		*Homo sapiens*	Q8N1Q1	Ota, et al. 2004
CA14		*Homo sapiens*	Q9ULX7	Ota, et al. 2004

Table 3.3 Chitobiase-like proteins applied to phylogenetic analyses in this study.

Protein	Larval SMP	Adult SMP	Species	Accession code	Reference
N,*N*'-diacetylchitobiase	Yes		*Crassostrea gigas*	CGI_10007856	Zhao, et al. 2018
Chitobiase		Yes	*Crassostrea gigas*	CGI_10007857	Zhang, et al. 2012
N,*N*'-diacetylchitobiase			*Crassostrea gigas*	CGI_10002999	Zhang, et al. 2012
Chitobiase			*Crassostrea gigas*	CGI_10023816	Zhang, et al. 2012
N,*N*'-diacetylchitobiase			*Crassostrea gigas*	CGI_10027871	Zhang, et al. 2012
N,*N*'-diacetylchitobiase	Yes		*Pinctada fucata*	pfu_aug2.0_6.1_20027.t1	Zhao, et al. 2018
Beta-hexosaminidase		Yes	*Pinctada fucata*	pfu_aug2.0_6.1_20028.t1	Zhao, et al. 2018
Uncharacterized protein			*Pinctada fucata*	pfu_aug2.0_2312.1_08823.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	pfu_aug2.0_4334.1_15958.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Pinctada fucata*	pfu_aug2.0_629.1_31002.t1	Takeuchi, et al. 2016
Uncharacterized protein			*Lottia gigantea*	Lgi_151739	Simakov, et al. 2013
Uncharacterized protein			*Lottia gigantea*	Lgi_168843	Simakov, et al. 2013
Uncharacterized protein			*Lottia gigantea*	Lgi_174972	Simakov, et al. 2013
Uncharacterized protein			*Lottia gigantea*	Lgi_236342	Simakov, et al. 2013
Uncharacterized protein			*Lottia gigantea*	Lgi_239375	Simakov, et al. 2013
Beta-hexosaminidase-like isoform X1			*Octopus bimaculoides*	XP_014767869.1	Albertin, et al. 2015
Beta-hexosaminidase-like isoform X2			*Octopus bimaculoides*	XP_014767876.1	Albertin, et al. 2015
Beta-hexosaminidase-like isoform X3			*Octopus bimaculoides*	XP_014767885.1	Albertin, et al. 2015
N,*N*'-diacetylchitobiase-like			*Octopus bimaculoides*	XP_014790058.1	Albertin, et al. 2015
Uncharacterized protein			*Lingula anatina*	XP_013407630.1	Luo, et al. 2015

Table 3.4 Proteins containing Glyco_18 domain (IPR011583) or Glyco_hydro_20 domain (IPR015883) identified in the shell proteome data and genome data of *C. gigas* and *P. fucata*.

	Glyco_18 domain (IPR011583)	Glyco_hydro_20 domain (IPR015883)	Reference
Larval SMP of *C. gigas*		CGI_10007856 ?	Zhao, et al. 2018
Adult SMP of *C. gigas*	CGI_10026605	CGI_10007857	Zhang, et al. 2012
Larval SMP of *P. fucata*		pfu_aug2.0_6.1_20027.t1	Zhao, et al. 2018
Adult SMP of *P. fucata*	pfu_aug2.0_194.1_13762.t1	pfu_aug2.0_6.1_20028.t1	Zhao, et al. 2018
	pfu_aug2.0_194.1_13763.t1		Zhao, et al. 2018
Genonome of *C. gigas*	CGI_10002421	CGI_10001718	Zhang, et al. 2012
	CGI_10006211	CGI_10002246	Zhang, et al. 2012
	CGI_10022102	CGI_10002999	Zhang, et al. 2012
	CGI_10022487	CGI_10004764	Zhang, et al. 2012
	CGI_10024867	CGI_10007196	Zhang, et al. 2012
	CGI_10024868	CGI_10007856	Zhang, et al. 2012
	CGI_10024869	CGI_10007857	Zhang, et al. 2012
	CGI_10024870	CGI_10008618	Zhang, et al. 2012
	CGI_10026598	CGI_10017246	Zhang, et al. 2012
	CGI_10026599	CGI_10022781	Zhang, et al. 2012
	CGI_10026600	CGI_10023604	Zhang, et al. 2012
	CGI_10026601	CGI_10023605	Zhang, et al. 2012
	CGI_10026602	CGI_10023816	Zhang, et al. 2012
	CGI_10026603	CGI_10024502	Zhang, et al. 2012
	CGI_10026604	CGI_10027871	Zhang, et al. 2012
	CGI_10026605	CGI_10028163	Zhang, et al. 2012
	CGI_10026762		Zhang, et al. 2012
Genonome of *P. fucata*	pfu_aug2.0_11054.1_16443.t1	pfu_aug2.0_2312.1_08823.t1	Takeuchi, et al. 2016
	pfu_aug2.0_1382.1_08349.t1	pfu_aug2.0_25.1_16729.t1	Takeuchi, et al. 2016
	pfu_aug2.0_14217.1_26475.t1	pfu_aug2.0_2583.1_12215.t1	Takeuchi, et al. 2016
	pfu_aug2.0_1660.1_01750.t1	pfu_aug2.0_32.1_06783.t1	Takeuchi, et al. 2016
	pfu_aug2.0_1664.1_15180.t1	pfu_aug2.0_4334.1_15958.t1	Takeuchi, et al. 2016
	pfu_aug2.0_1921.1_05318.t1	pfu_aug2.0_6.1_20027.t1	Takeuchi, et al. 2016
	pfu_aug2.0_194.1_13761.t1	pfu_aug2.0_6.1_20028.t1	Takeuchi, et al. 2016
	pfu_aug2.0_194.1_13762.t1	pfu_aug2.0_629.1_31002.t1	Takeuchi, et al. 2016
	pfu_aug2.0_194.1_13763.t1	pfu_aug2.0_63.1_10205.t2	Takeuchi, et al. 2016
	pfu_aug2.0_194.1_13766.t1		Takeuchi, et al. 2016
	pfu_aug2.0_21.1_03434.t1		Takeuchi, et al. 2016
	pfu_aug2.0_21.1_03435.t1		Takeuchi, et al. 2016
	pfu_aug2.0_2404.1_15511.t1		Takeuchi, et al. 2016
	pfu_aug2.0_2934.1_15667.t1		Takeuchi, et al. 2016
	pfu_aug2.0_3570.1_02442.t1		Takeuchi, et al. 2016
	pfu_aug2.0_620.1_00959.t1		Takeuchi, et al. 2016
	pfu_aug2.0_70.1_00174.t1		Takeuchi, et al. 2016
	pfu_aug2.0_84.1_13566.t1		Takeuchi, et al. 2016
	pfu_aug2.0_84.1_13567.t1		Takeuchi, et al. 2016

Question mark indicates if the gene is a SMP is questionable.

Chapter 4 Exploration of performing functional analysis of shell matrix proteins via transgenic molluscs

Key words: transgenic, shell matrix protein (SMP), electroporation, CRISPR/Cas9, *SPARC*.

4.1 Introduction

With the benefits of new techniques such as high-resolution transmission electron microscopy and next-generation DNA sequencing, current researchers on molluscan biomineralization appear to be enthusiastic with the characterization of the nanostructures of the shells and identification of an increasing number of shell matrix proteins. It is generally conceded an opinion that the shell matrix is where bionimeralization occurs, and the main components of which, proteins, glycoproteins, acidic polysaccharides control this dynamic process. However, questions such as exactly which gene product or gene products control the transforming of the calcium carbonate polymorphs between calcite and aragonite (Suzuki, et al. 2009; Takeuchi, et al. 2008), or one of the most intriguing phenomenon in nature, and how the fine aragonitic microstructure of prodissoconch change abruptly to relatively coarse prismatic and foliated calcitic microstructures in the dissoconch (juvenile-adult shell) (Carriker 1979; Waller 1981) are still not explicitly answered.

Genetic transformations aiming at knock-out and knock-in of certain shell protein genes could be a powerful strategy to address these problems. Nevertheless, although integration of new traits has been widely investigated in mammals and fish, only a few similar studies have been undertaken on invertebrates, and even less on molluscs. Gene transformation has been reported in abalones via electroporation into the embryos of the red abalone *Haliotis rufescens* (Powers, et al. 1995) and using sperm as carrier via electroporation in the Japanese abalone *Haliotis diversicolor* (Tsai, et al. 1997). Gene transfer has been reported in the Pacific oyster *Crassostrea gigas*, by

particle bombardment into the embryos (Cadoret, et al. 1997a) and microinjection into fertilized eggs (Cadoret, et al. 1997b). Transgenic antibiotic resistant embryos of the Eastern oyster *Crassostrea virginica* were produced by electroporation and chemical transfection (Buchanan, et al. 2001).

Considering the small sizes of the eggs of the pearl oysters, typical nacro-prismatic mollusc, used in this study, two foreign DNA delivery methods, i.e., electroporation and chemical transfection were tested on the fertilized eggs of the pearl oyster *Pinctada fucata*. As foreign DNAs, *green fluorescent protein* (*GFP*) expression plasmids integrated with an endogenous *beta-tubulin* promoter or an exogenous *heat-shock* protein promoter of the fruit fly *Drosophila melanogaster* were tested. Foreign DNAs were detected within the total DNA extraction of embryos after transfection with either method, which indicated the effectiveness of the delivery of DNAs. However, fluorescent individuals could not be observed under microscope, which suggested some failures in the promoter sequences or the incompatibility of the codon usage of *GFP* that is not optimized for molluscs.

Instead of spending time on the hard work of traditional gene-targeting method, designer nucleases (ZFN and TALEN) have been applied to various organisms to direct site-specific DNA double-strand breaks (DSBs) in the coding region of a target gene (Beumer, et al. 2013; Sakuma, et al. 2013). Nonhomologous end joining (NHEJ) repair after the DSBs at the cut site lead to shift-frame mutations of the target gene. For both ZFN and TALEN, generation of the modules that recognize the target chromosomal site can be a significant investment. Recently, clustered, regularly interspaced short, palindromic repeats (CRISPR) technology combined with RNA-guided Cas9 nuclease changed this situation. This tool originated from bacterial defense mechanisms has been successfully introduced into various model and non-model systems, including mouse (Wang, et al. 2013a), human (Cong, et al. 2013), zebrafish (Hwang, et al. 2013), *Xenopus* (Blitz, et al. 2013), nematode (Friedland, et al. 2013), *Drosophila* (Gratz, et al. 2013) and the mollusc *Crepidula fornicata* (Perry

and Henry 2015). ZEN and TALEN utilize a nuclease designed for each target site, while in CRISPR/Cas9, the nuclease domain of Cas9 is guided to the target site by a small RNA construct (gRNA). Therefore, different sites can be targeted by simply modifying the gRNA without modifying the nuclease.

SPARC (Osteonectin), was identified to be a major protein in the acid-soluble shell matrix proteome and a minor component in the acid-insoluble fraction in the limpet *Lottia gigantea* (Mann, et al. 2012). Homologous proteins were identified in *Haliotis discus* and *Pinctada fucata* (H. Miyamoto and F. Asada, unpublished, UniprotKB/TrEMBL accessions F2Z9K1_PINFU and F2Z9K2_HALDI). Regulatory roles of SPARC in some biomineralization processes of mammals have been reported (Wallin, et al. 2001), while whether or not it possesses similar functions in molluscs is unknown.

As a hopeful way for generating transgenic molluscs, and also as a trial for a direct in vivo functional analysis of shell proteins, CRISPR/Cas9 gene editing technique was applied to an SMP gene of limpets, for which delivery of foreign DNA to embryos is known to be possible through microinjections. In this study, first, the SPARC gene of the limpet *Nipponacmea fuscoviridus* was characterized, and its expression in the larval stage was confirmed. Next, Cas9-coding RNA and gRNA targeted on the SPARC (Osteonectin) gene of *N. fuscoviridus* produced by *in vitro* transcription were microinjected into the fertilized eggs of *N. fuscoviridus.* Several merits of this study can be expected compared with the former way of generating transgenic lineage of *P. fucata*: 1, the bigger size of the eggs of *N. fuscoviridus* make the microinjection feasible, thus the delivery of DNA can be visually confirmed. 2, directly applying the Cas9-coding RNA and gRNA avoids the problems caused by the promoters. 3, fertilized eggs are more accessible compared with *P. fucata*, whose mating season is restricted to April-July. *N. fuscoviridus* has a longer breeding season (April-June and September to November). Also, the adult animals are easier to be reared in the laboratory.

4.2 Materials and methods

4.2.1 Artificial fertilization

The pearl oyster, *Pinctada fucata*, was a gift from Mikimoto Pearl Research Institute. The oyster, *Crassostrea gigas*, was purchased at local fish market, Yoshiike. The artificial fertilization of both species followed the same protocol. The female and the male were separately cultured in aquarium with artificial sea water at 22 ℃ and 18 ℃, respectively. On the day of artificial fertilization, the gonad of the female is firstly removed to a piece of clean gauze (15 cm * 15 cm). After being cut into several pieces, the gonad wrapped in gauze was immerced into 500 mL ammonia filtered artificial sea water (FASW; 25 ℃) containing in a ratio of 0.75 ml 1M ammonia in 1 L FASW, with shaking to release eggs, then, left for 40 min till the germinal vesicles were destroyed. The sperm was obtained from the male gonads, and then was added into ammonia FASW as above for 5 min to be activated. Next, 500 mL FASW containing the egg without germinal vesicle was mixed with 100 mL FASW containing the active sperm for 5 min. Being collected in 20 N nylon net, the fertilized egg were washed to remove the sperm and were incubated in 20 L FASW at 24 ℃ for 7 h. Trochophore stage larvae that were floating on the surface of the water were collected and were incubated in 20 L FASW at 24~26 ℃ overnight. About 24 h after fertilization, the successfully fertilized eggs that turned into D-shape larva were collected and incubated in FASW at a density of 10 individual/mL at 24~26 ℃ for further study.

The marine gastropod *Nipponacmea fuscoviridis* was captured from a rocky shore intertidal locality in Hiraiso kaigan, Oarai, Ibaraki Prefecture, Japan, during mating periods from April to July and from September to November. Artificial fertilization was performed following the previous method (Deguchi 2007). Embryos were cultured in FASW at 25 ℃.

4.2.2 Cloning of β-tubulin and heat-shock protein promoters into the GFP

expression plasmid

A 3.2 kb nucleotide sequence before the start codon of beta-tubulin gene located on the scaffold7719.1|size90101 (Pearl oyster genome, v 1.0) was cloned into the plasmid pSP-nEGFP (a gift from Prof. Sasakura, Shimoda Marine Research Center, University of Tsukuba) to generate the GFP expression plasmid containing the pearl oyster promoter, pBTGFP. The promoter sequence was tailed with NotI restriction enzyme site at either side by PCR amplification. The genomic DNA extracted by CTAB method (Fisher and Skibinski 1990) from the muscle of *P. fucata* was used as the template. PCR amplification catalyzed by PrimeSTAR® GXL DNA Polymerase (Takara Inc.), consisted of 40 cycles of 98 ℃ for 10 s, 63 ℃ for 15 s and 68 ℃ for 4.5 min. Primers NotIBTNotI S (ATTTGCGGCCGCCCTTGTCAATCAATAGGGTTCATCC) and NotIBTNotI A (ATTTGCGGCCGCTTGTGTCGGCGTTTCTCTTTGCTGT) were designed based on the complementary sequences franking the promoter region in the scaffold7719.1|size90101. Both the PCR product and the plasmid pSP-nEGFP were digested by restriction enzyme NotI. The linearized plasmid was treated with phosphatase in order to prevent self-ligation. Next, the promoter and the plasmid were ligated in Ligation Mix (Takara, Inc.) at 4 ℃ overnight, and then were transfected into the *E. coli* DH5 competent cell. Colonies were picked as the template for PCR amplification to confirm the insertion with the primer BT to psp-eGFP S (TCGAGCAGCTGAAGCTTGCATGCCT) and BT to psp-eGFP A (GTCATTTTTTCTGAGCGCCGTACCC), which were designed based on the sequence of the plasmid psp-nEGFP. PCR catalyzed by PrimeSTAR® GXL DNA Polymerase (Takara Inc.), consisted of 40 cycles of 98 ℃ for 10 s, 62 ℃ for 15 s and 68 ℃ for 5 min. The PCR product was sequenced to confirm successful insertion of the promoter sequence.

A heat-shock protein promoter was also cloned into the GFP expression plasmid, which generated the plasmid, pHsp70eGFP. BamHI restriction enzyme sites tailed specific PCR primers, BamHIhsp70pro S (1-21,

GCAATCGGATCCATCCCCCTAGAATCCCAAAAC) and BamHIhsp70pro A (572-592, GCAATCGGATCCCTGCTGGGACTCCGTGGATAC). The template was the plasmid pBS-Hsp70-Cas9. PCR was catalyzed by Extaq DNA polymerase (Takara, Inc.), consisted of 25 cycles of 94 ℃ for 40 s, 65 ℃ for 45 s and 72 ℃ for 1 min. The PCR product, or the heat-shock protein promoter sequence tailed by BamHI sites, was purified by ethanol precipitation. Both this promoter sequence and the plasmid pSP-nEGFP were digested by the restriction enzyme BamHI and ligated in Ligation Mix (Takara, Inc.) overnight, and then transfected into *E. coli*. To identify the colony carrying the plasmid pHsp70eGFP, colonies were picked as the template for PCR amplification. PCR performed with Extaq DNA polymerase (Takara, Inc.) consisted of 25 cycles of 94 ℃ for 40 s, 60 ℃ for 45 s and 72 ℃ for 2 min. Primers, Insert-to-eGFP S (3397-3416, AAATAGGCGTATCACGAGGC) and Insert-to-eGFP A (550-569, CGTTGTGGCTGTTGTAGTTG), were designed based on the sequence of the plasmid pSP-nEGFP. The PCR product was purified and sequenced to verify successful ligation of the insert. The *E. coli* carrying the pHsp70eGFP was cultured before the plasmid was purified by GenElute^{TM} Plasmid Miniprep Kit (SIGMA-ALDRICH, Inc.).

4.2.3 Introduction of foreign DNA into the embryo

Electroporation was performed following the previous method (Buchanan, et al. 2001). Three hours after fertilization, about 10,000 embryos were collected in 0.8 mL FASW with 60 μg DNA in 0.4-cm electroporation cuvettes (Bio-Rad Laboratories, Inc.). The cuvette containing embryos was set in a Gene Pulser Xcell^{TM} Electroporation System (Bio-Rad Laboratories, Inc.), which generates an exponential decay-type electrical field. In accordance with the instruction manual, a time constant protocol was applied, in which the electrical field strength was reported as a time constant, τ, which means the length of time for the electrical field to decay about 37% (1/e). Since τ equals resistance multiplied by capacitance, for a sample whose resistance is settled, raising the capacitance raises the electrical field duration. To evaluate the breakage ratios of embryos under the electrical field duration, different

capacitance values of 10, 25, 50, 75 and 100 μF, each at 200, 260, 300 V, were tested. Embryos mixed with DNA without pulse were taken as control. After the pulse, embryos were cultured with DNA in 24-well plate for 1 h, and then were transferred into 10 cm shell plates for incubation at room temperature. All treatments were performed in triplicate.

To measure the embryo breakage ratios after electroporation, 0.4 mL embryos after pulse were pipetted into 12-well plate with 1 mL FASW. Three figures of random areas of each sample were taken by VHX-900 digital microscope (Keyence, Inc.). Numbers of the broken and the normal embryo were counted. Two hours later, embryos were transferred to 40 mL FASW in 50 mL tube at 25 ℃ for further incubation.

4.2.4 Chemically mediated transfection on embryos

Chemical transfection was performed following the previous method (Buchanan, et al. 2001). At thirty minutes and at three hours after fertilization, about 2,500 embryos of *P. fucata* were collected in 24-well tissue culture plate, respectively. Plasmid pHsp70eGFP was diluted in oyster saline solution (0.48 g/L CaCl2, 1.45 g/L MgSO4, 2.18 g/L MgCl2 • 6H2O, 0.31 KCl g/L, 11.61 g/L NaCl, 0.35 g/L NaHCO3). To test the transfection effect, 100μl of DNA and SuperFect (Qiagen, Inc.) with DNA-to-SuperFect ratios of 1:3, 1:6, 1:9 (μg/μl) with 2.5 or 5μg DNA were prepared and incubated for 10 min at 25 ℃ according to the manual. Then, embryos were transferred to a well with fresh FASW for further incubation. Control groups are embryos mixed with 2.5μg and 5μg DNA alone per well.

4.2.5 Examination of transfection effect

To extract the total DNA of the embryos after transfection treatments, 20 embryos of each group were collected in 20 N nylon net and washed with fresh FASW for 10 times to remove the free DNA in the environment. Then, total DNA was extracted following the forensic DNA extraction method (Edwards and HoY 1993). The

extracted DNA was applied as the template for PCR catalyzed by Extaq DNA polymerase (Takara, Inc.), consisting of 30 cycles of 94 ℃ for 40 s, 60 ℃ for 1 min and 72 ℃ for 2 min. Specific primers GFP confirm S: 187-206, AAACGGCCACAAGTTCAGCG and GFP confirm A: 756-775, CATGTGATCGCGCTTCTCGT are designed based on the sequence of psp-nEGFP. For the positive control group of PCR, plasmid pHsp70eGFP construct was used as the template. For the negative control, eggs accepting no pulse were cultured with the DNA for 1h, being washed to remove the free DNA. Then the total DNA was extracted to be used as the template for PCR.

4.2.6 Characterization of the SPARC gene of *Nipponacmea fuscoviridus*

To identify potential target sequences within the SPARC gene for CRISPR/Cas9, touchdown PCR amplification was performed. cDNAs synthesized from the mRNA extracted from the veliger stage embryos and the mantle edge of the adult *N. fuscoviridus* were applied as the template for PCR. Mix primers, Sparc F1 (ATGMGNAARTGGATHGTNGC), Sparc F2 (ATGMGNAARACNATHGTNGC), Sparc R1 (ARRCARTGYTCCMWNGCYTT), and Sparc R2 (AGGAAGGGMACRAGRCAGTG) were designed based on the alignment of the amino acid sequences of SPARC gene among the three marine gastropod species, *Haliotis discus*, *Patella vulgate and Lottia gigantea* (Fig. 4.1). Touchdown PCR, catalyzed by Extaq DNA polymerase (Takara Inc.), consisted of 2 cycles of 94 ℃ for 1 min, 60 ℃ (ramped to 46 ℃) for 1 min and 72 ℃ for 1 min and 28 cycles of 94 ℃ for 1 min, 45 ℃ for 1 min and 72 ℃ for 2 min. PCR products were sequenced (Fig. 4.2A).

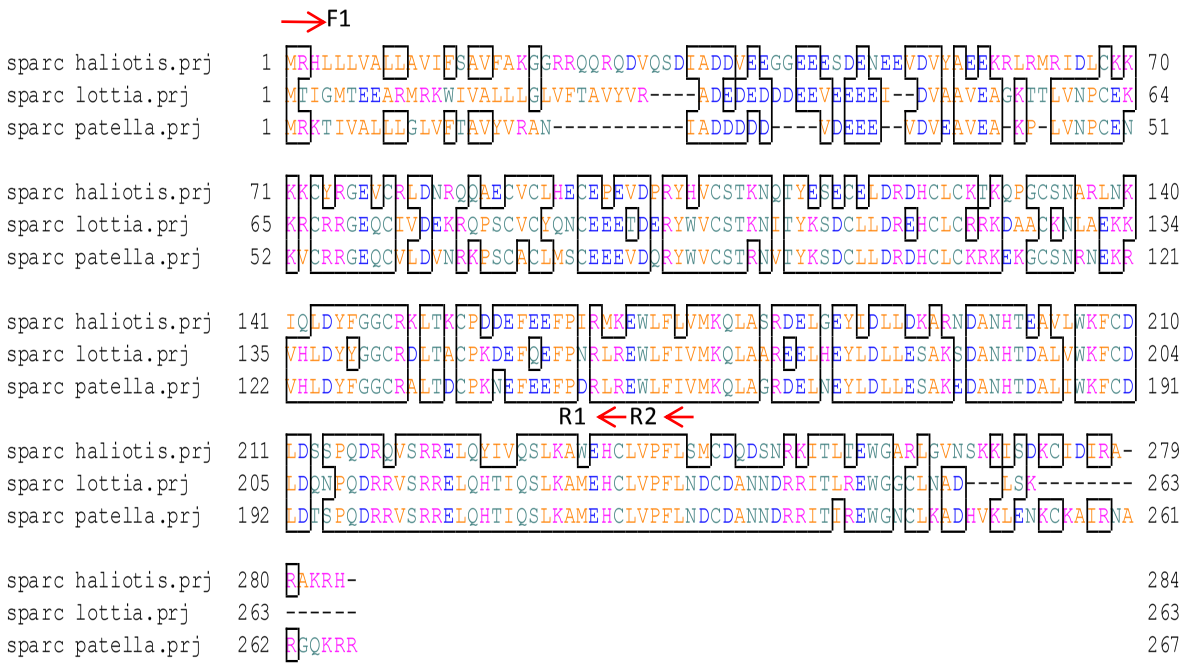
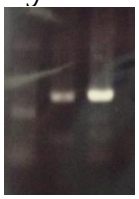


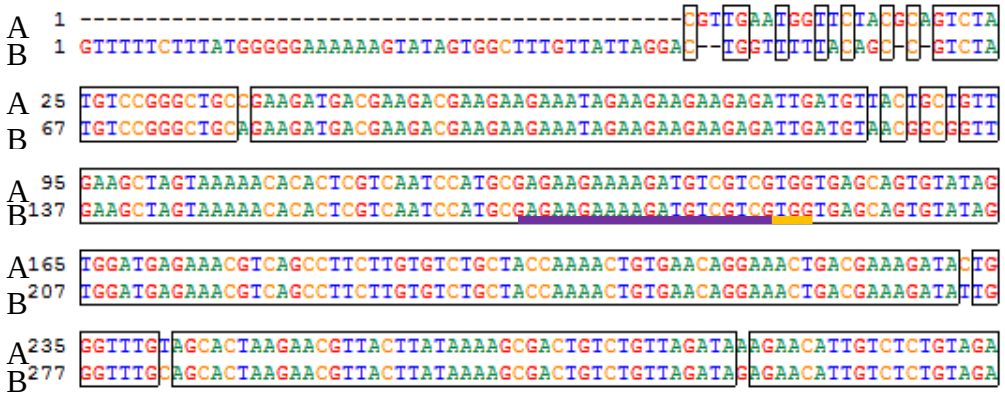
Fig. 4.1 Alignment of the amino acid sequence of SPARC gene of marine gastropod. Arches indicate the position for mix primers.

A

PCR by mix primers



A B



B

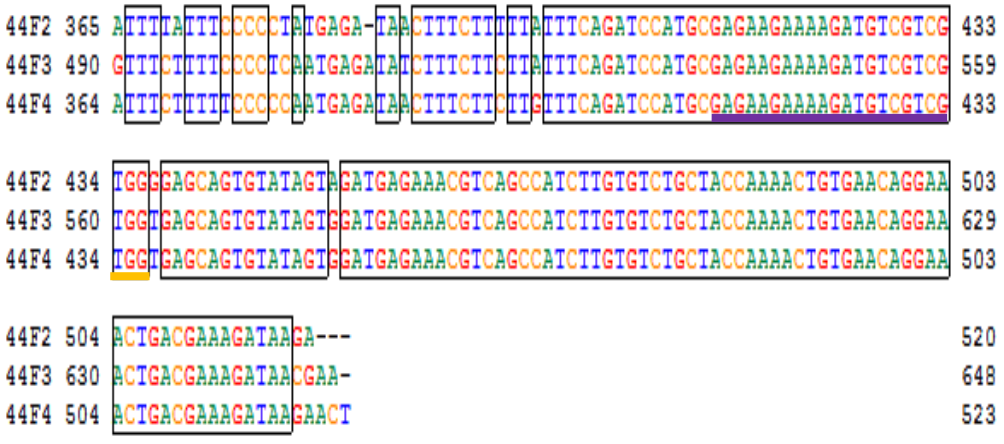


Fig. 4.2 Characterization of SPARC gene of *N. fuscoviridus*. A. Coding sequence of SPARC gene. B. Genomic DNA sequence of N-terminal region of SPARC gene. Purple and yellow line indicate the target sequence and 'PAM' sequence, respectively.

PCR amplification was performed to characterize the sequence of SPARC gene of the genomic DNA. The template was the genomic DNA extracted via CTAB method (ref.) from the muscle of three adult individuals of *N. fuscoviridus*. Primers Fdc 4 (CTAGTAAAAACACACTCGTC) and Rdc 4 (ATCTTTCGTCAGTTTCCTGT) were designed based on the coding sequence of SPARC gene of *N. fuscoviridus*. PCR catalyzed with PrimeSTAR® GXL DNA Polymerase (Takara Inc.) consisted of 28 cycles of 98 ℃ for 10 s, 48 ℃ for 15 s and 68 ℃ for 5 min. The PCR products were sequenced (Fig. 4.2B).

4.2.7 Generation of gRNA containing plasmid pU6-BbsI-chiRNA-SPARC

Plasmids pBS-Hsp70-Cas9, Peft-3::cas9-SV40_NLS::tbb-2 3'UTR and pU6-BbsI-chiRNA were purchased from Addgene (https://www.addgene.org/). gRNA coding plasmid targets on the SPARC gene was generated according to the protocol for generating gRNA in the Addgene site (https://www.addgene.org/) (Fig. 4.3). The insertion of the targeted gene into the pU6-BbsI-chiRNA was confirmed by PCR amplification using primers M13/pUC-forward (588-610, CCCAGTCACGACGTTGTAAAACG) and M13/pUC-reverse (1325-1347, AGCGGATAACAATTTCACACAGG). PCR catalyzed by Extaq DNA polymerase (Takara Inc.) consisted of 28 cycles of 94 ℃ for 40 s, 60 ℃ for 1 min and 72 ℃ for 2 min. The PCR product was purified by ethanol precipitation and sequenced (Fig. 4.3).

Protocol for generating plasmid pU6-BbsI-chiRNA
targeted on SPARC gene

1. Oligos were designed as follows:
   sense oligo: 5' – **CTTCG** (19 nt) – 3'
   antisense oligo: 3' – C (19 nt) **CAAA** – 5'

2. Anneal oligos:
   Oligo were diluted to 100 μM in TE buffer. 5 μL of both oligos were mixed. Then, the following thermocycler was performed:

   95 ℃ for 5 min, then ramp to 25 ℃ at a rate of -5 ℃ /min.

3. Plasmid pU6-BbsI-chiRNA was cut with restriction enzyme BbsI (Biolabs Inc.) and de-phospharylated.

4. Annealed oligos and plasmid were mixed at ratios of 1 : 1 and 2 : 1 and ligated in Ligation Mix (Takara Inc.) overnight, and then transfected into *E. coli.*

5. Confirmation of successful generation of the plasmid containing targeted gene by PCR amplification and sequencing.

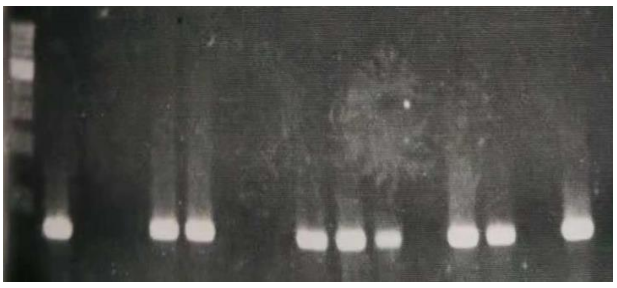


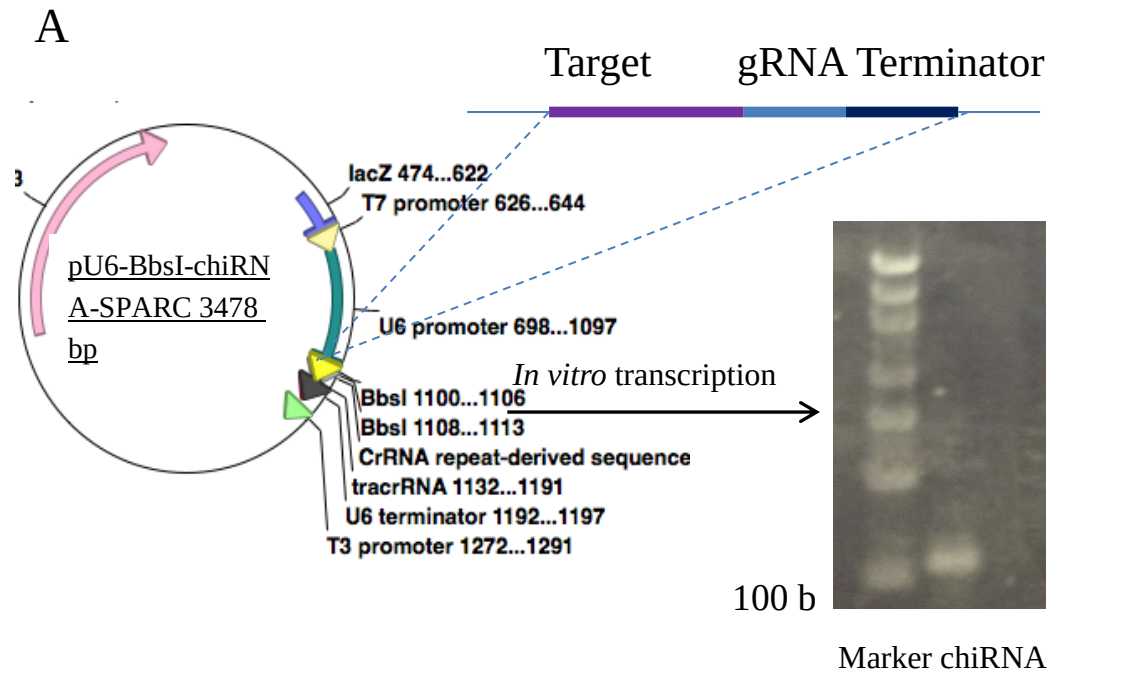
Fig. 4.3 Protocol for cloning the target sequence into the gRNA containing plasmid and verification by PCR amplification. The result of PCR products indicated that all the band were correctly ligated sequence.

4.2.8 Generation of plasmid Peft-3::cas9 SP6

In order to obtain *Cas9* RNA via *in vitro* transcription, a SP6 promoter was integrated into the *Cas9* coding plasmid, Peft-3::cas9-SV40_NLS::tbb-2 3'UTR. First, BsrGI restriction enzyme site and SP6 promoter containing oligos, BsrGIsp6 F (GTACAATTTAGGTGACACTATAGAAT) and BsrGIsp6 R (GTACATTCTATAGTGTCACCTAAATT), were annealed through the thermocycler that was applied for generating chiRNA. Next, the annealed oligos were ligated with the plasmid Peft-3::cas9-SV40_NLS::tbb-2 3'UTR that was cut by the restriction enzyme BsrGI to form plasmid Peft-3::cas9 SP6 (Fig. 4.4B).

4.2.9 *In vitro* transcription

After being digested at the unique restriction enzyme HindIII site, the linearized Peft-3::cas9-SV40_NLS::tbb-2 3'UTR SP6 was used as the template to synthesize *Cas9* RNA using mMessage mMachine SP6 RNA transcription kit (Life Technologies Inc.). In order to obtain a linearized template for synthesize chiRNA, the target sequence (19 nt) containing pU6-BbsI-chiRNA was applied as the template for PCR amplification. SP6 promoter containing specialized primers, Sp6sparcchiRNA F (GCAATCATTTAGGTGACACTATAGAAGCGGAGAAGAAAAGATGTCGT) and Sp6sparcchiRNA R (TCGATAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACG) were applied to the PCR amplification, which consisted of 27 cycles of 94 ℃ for 40 s, 63 ℃ for 45 s and 72 ℃ for 1 min. The PCR product contains a SP6 promoter, the target sequence and a U6 terminator, which was purified and used as the template for *in vitro* transcription via the method mentioned above. Yielded *Cas9* RNA and chiRNA were treated with Turbo DNaseI at 37 ℃ for 15 min, purified by RNeasy MinElute cleanup kit (Qiagen Inc.), eluted in nuclease-free water. The sequence length of synthesized *Cas9* RNA and chiRNA were confirmed through gel electrophoresis (Figs. 4.4 A and B).



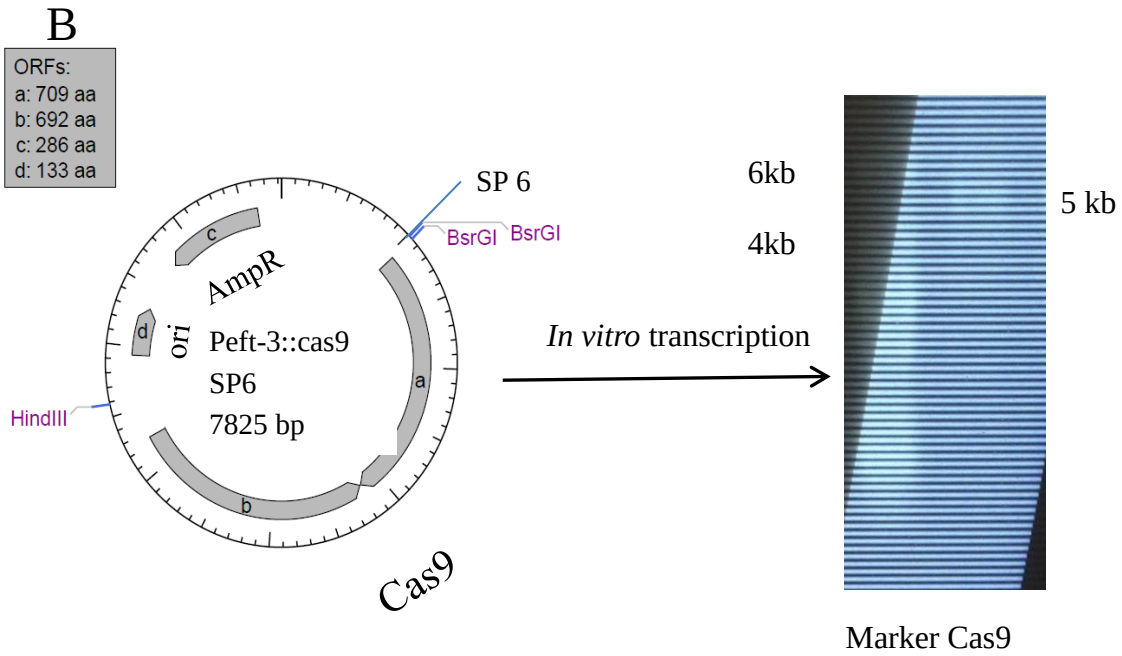


Fig. 4.4 Schematics of CRISPR/Cas9 constructs and electrophoresis images of mRNA of gRNA and *Cas9*. A. The 20 nt targeted sequence was cloned into the gRNA containing plasmid via BbsI restriction enzyme sites. The messenger RNA containing the sequence of the gRNA, the target sequence, and a terminator was obtained. B. A SP6 promoter was inserted before the *Cas9* coding sequence. The messenger RNA of *Cas9* of about 5 k b was

4.2.10 Microinjection

Before the first cleavage, fertilized eggs of *N. fuscoviridis* were microinjected with micromanipulators (Narishige, Inc.) and FemtoJet microinjector (Eppendorf, Inc.). To visualize the introduction of the solution into the cell, sterilized filtered 0.5% phenol red solution (Sigma, Life Science) was mixed with RNA solution (1 part phenol red solution to 3 part RNA solution). Final concentrations of *Cas9* RNA and chiRNA in the mixture were 400 ng/μl and 200 ng/μl, respectively. Phenol red solution with either *Cas9* or chiRNA alone was used as control mixture. After microinjection, cells were incubated in FASW at room temperature. Twenty hours after fertilization, larvae were observed under BX51 upright microscope (Olympus).

4.2.11 Examination of the expression of foreign DNA in molluscan larvae

24 and 48 hour after fertilization, larvae introduced with pBTGFP were examined by Leica MZ10F fluorescent microscope (Leica, microsystems) to check the expression of green fluorescent protein. 24 h after fertilization, larvae introduced with pHsp70eGFP were given with a heat shock at 35-37 ℃ for 1 h, and then were incubated at room temperature. 48 h and 72 h after fertilization, these larvae were observed under the fluorescent microscope.

4.3 Results

4.3.1 Artificial fertilization and larva sustaining in *Pinctada fucata*

Artificial fertilization of *P. fucata* was performed with the previous method (Fujimura, et al. 1995). Early development of fertilization was observed under microscope. Treated by ammonia filtered artificial sea water (FASW), the germinal vesicle of the egg of irregular shape was removed, which accelerated the maturation of eggs for fertilization (Figs. 4.5 A and B). Ammonia FASW also stimulated the activity of the sperm, which was confirmed by the acute motions of the sperms. The first and the second cleavage occurred at 20-30 min and about 1 h after fertilization, respectively (Figs. 4.5 C and D). Six hours after fertilization, the cavity of blastocoel could be observed (Fig. 4.5 E); and the blastula stage embryos began to swim into the surface of the water. Swimming-ciliated trochophore stage larvae were formed 12 h after fertilization (Fig. 4.5 F). Twenty hours after fertilization, early D-shape larva whose soft tissues were covered by the larval shells could be newly observed (Fig. 4.5 G). Forty-eight hours later after fertilization, the cilia of D-shape larva could be observed and the shape of inside organs became clear (Fig. 4.5 H). From 24 h after fertilization, the planktonic diatome *Pavlova lutheri* (Pearl Research Institute, Mikimoto, Japan), was given to the larva at a dose of 500 cell/individual /day as food. The water was changed at the day 6, day 9, day 12, day 15, and day 20 after fertilization. Artificial fertilization and larva sustaining thereafter were repeated more than 3 times. Larvae survived more than three weeks after fertilization without metamorphosis.

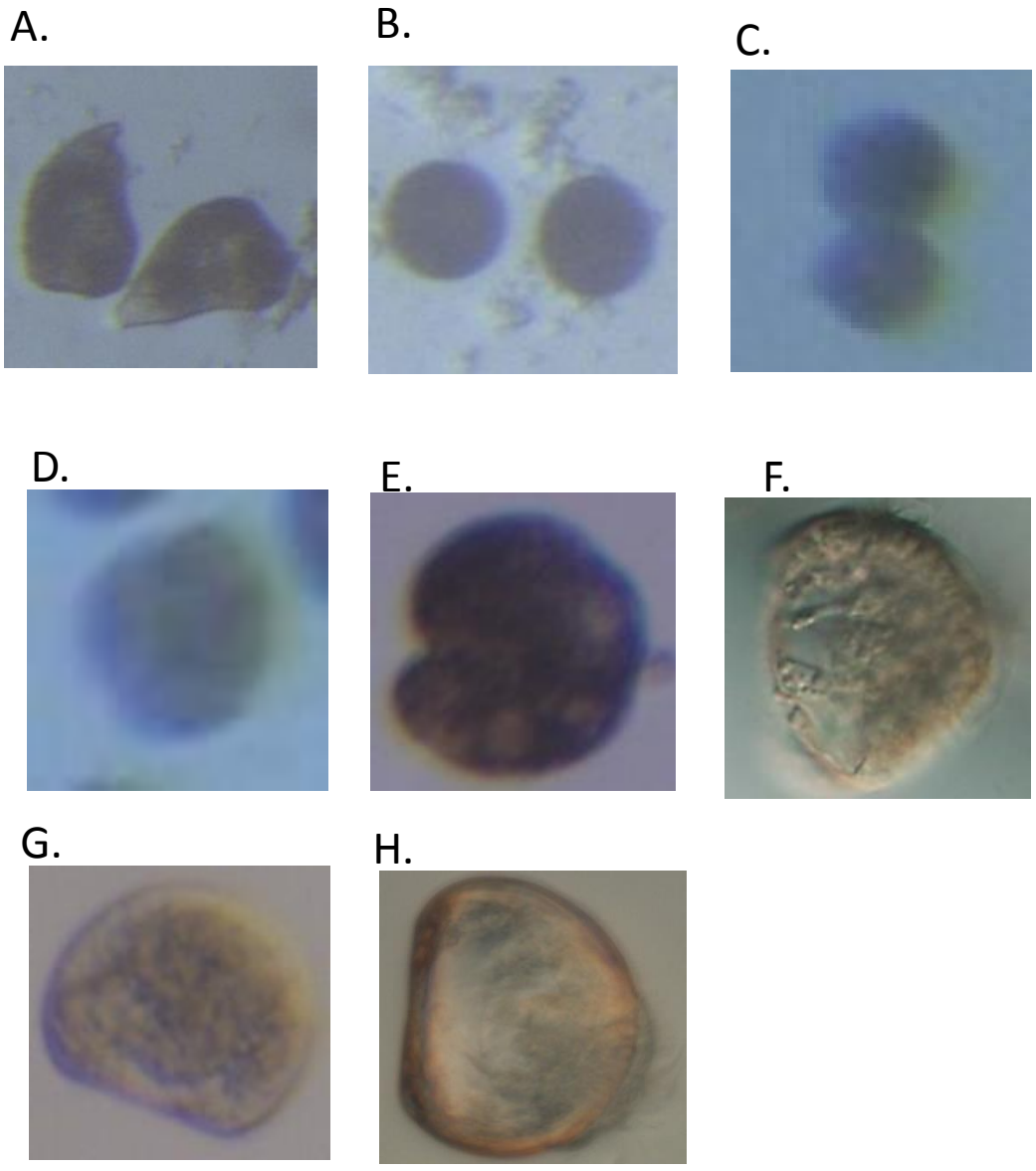
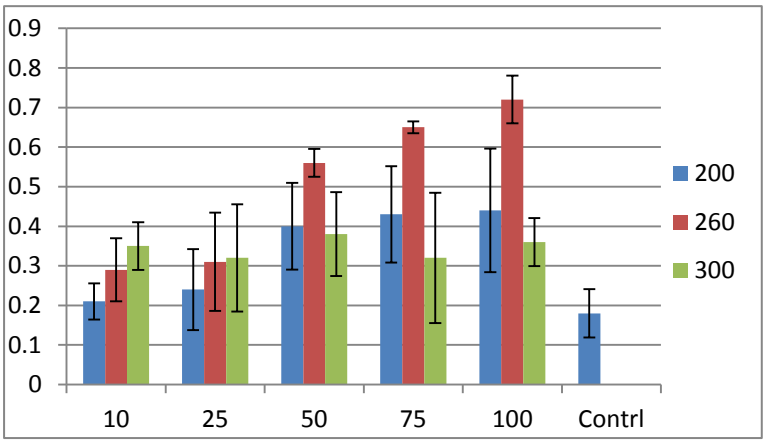


Fig. 4.5 Early embryo development of *P. fucata*. A. Eggs separated from the gonad. B. Germinal vesical of the egg was removed by incubating with ammonia FASW for 40 min. C. The first cleavage of the egg. D. The second cleavage of the egg. E. Blastula stage of the embryo. F. Trochophore stage of the embryo. G. Early D-shape larva at 20 h after fertilization. H. D-shape larva at 48 h after fertilization.

4.3.2 Introduction and detection of the foreign DNA in the fertilized eggs of *Pinctada fucata*

Because of the small diameter (about 50 μm) of the eggs of *P. fucata*, electroporation, rather than microinjection, was performed to introduce foreign DNA. Plasmids pHsp70eGFP and pBTGFP were tested as foreign DNA. The conditions of electroporation followed the previously described method (Buchanan, et al. 2001). The breakage ratios of the eggs were measured to evaluate the effect of electric field imposed to the eggs (Fig. 4.6 A). We noticed that when the breakage ratio exceeded 40%, a large amount of embryos abnormally developed in the following incubation. Meanwhile, in the control group without pulse, about 20% breakage ratio was observed, which possibly resulted from the operation after removing the germinal vesicle. Therefore, we focused on those combinations of capacitance and voltage which generated about 30-40% breakage ratio in the following tests. The foreign DNA was clearly detected via PCR amplification in the D-shape larva 24 h after giving the pulse at 100 μF and 100-200 V, which was confirmed by repeated experiments (Figs. 4.6 B and C). However, fluorescent individuals could not be detected under microscope.

A.



B.

C.

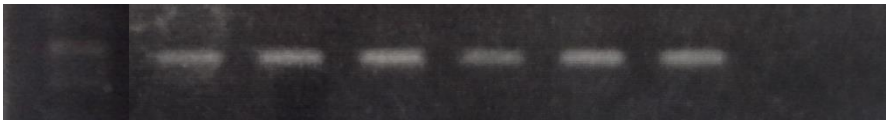


Fig. 4.6 Introduction of the foreign DNA into the egg of *P. fucata* through electroporation. A. The effect of electroporation to the breakage ratio of the fertilized egg. B. The foreign DNA was detected after giving the electric pulse at 75 μF and 100 μF with the voltage of 100-200 v. PC, positive control. NC, negative control. C. The introduction of foreign DNA was confirmed after the pulse at 100 μF and 100-200 v.

4.3.3 Chemical mediated transfection

Embryos of *P. fucata* thirty minutes and three hours after fertilization were transfected with SuperFect. Our results indicated that regardless of the ratios between DNA and SuperFect and the stage of embryos, higher levels of transfection effect were detected in the 5μg DNA group than that in the 2.5μg DNA group (Fig. 4.4). Meanwhile, the DNA cannot be detected in the control group in which embryos were cultured with DNA alone. Nevertheless, fluorescent light of the larva could not be detected under microscope.

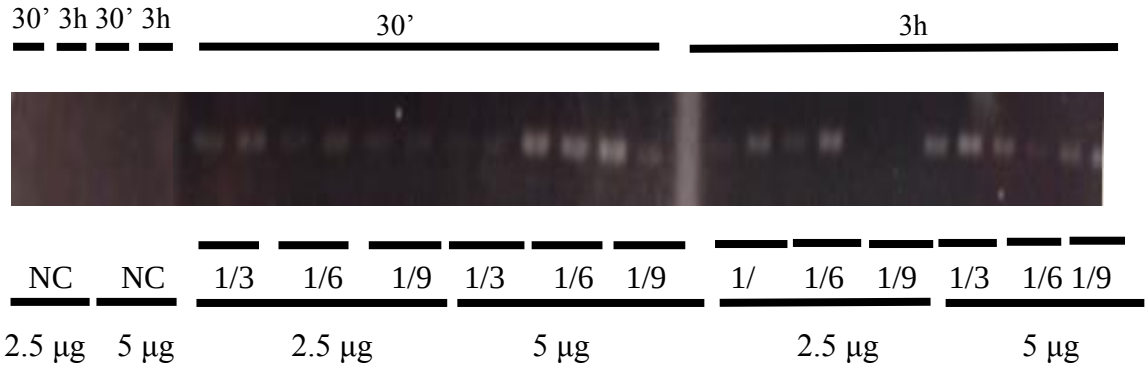


Fig. 4.7 Detection of the foreign DNA after transfection via SuperFect. Thirty minutes and three hours after fertilization, embryos were incubated with different ratios of DNA-SuperFect mixture. 2.5 and 5 μg of DNA were tested. NC, negtive control groups in which embryos were incubated with DNA alone.

4.3.4 Identification of the targeted nucleotide sequence for gRNA

One of the shell matrix proteins, SPARC, was selected as the target gene to test the CRISPR/Cas9 effect on the marine gastropod *Nipponacmea fuscoviridus*. The SPARC gene of *N. fuscoviridus* was characterized for the first time in order to search the targeted sequence sites. By PCR amplification with mixed primers designed based on the alignment of the amino acid sequences of the same gene from three marine gastropods *Haliotis discus discus*, *Lottia gigantea*, and *Patella vulgata* (Fig. 4.1), the N-terminal coding sequence of the SPARC gene in *N. fuscoviridus* was identified (Fig. 4.1A). Also, sequence determinations of these PCR products confirmed the coherency of this region being not disturbed by any intron (Fig. 4.1B). Next, a 20 nt sequence followed by a 'NGG' was chosen as the target sequence for gRNA. In the end, the messenger RNA of the endonuclease *Cas9* and the gRNA targeting on the SPARC gene were obtained via *in vitro* transcription method (Krieg and Melton 1984) (Fig. 4.4).

4.3.5 Introduction of the CRISRP/Cas9 into the fertilized egg of *N. fuscoviridis*

A mixture of messenger RNA constructs of the endonuclease *Cas9*, the gRNA targeting on the SPARC gene and the phenol red solution was microinjected into the fertilized eggs of *N. fuscoviridis* before the first cell cleavage. As the first trials of microinjection, the introduction of the RNA mixture was confirmed by the color change and the volume inflation of the egg (Fig. 4.8). However, pulling the needle out of the egg was accompanied with an ejection of cytoplasm. Also, a low survival rate after microinjection was observed.

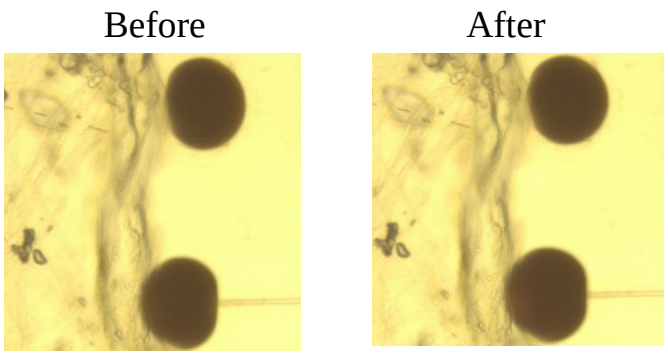


Fig. 4.8 Introduction of CRISPR/Cas9 constructs into the fertilized egg of *N. fuscoviridus*. The delivery of DNA was confirmed by the color and volume change of the cell.

4.4 Discussion

In order to generate a transgenic molluscan platform to perform the functional analysis of SMPs, artificial fertilization and larva sustaining of *P. fucata* systems were firstly established in the laboratory. The D-shape larva was sustained more than three weeks in our experiment, although the umbonal stage larva observed in previous reports did not form, and infection by microbes occurred to the larva. Thus, the larva generated in these systems can be applied at least to the studies of larval shell formation. Considering the differences between the small scale culturing conditions and the large scale ones applied by previous researches, the frequency of changing water and the quality as well as the dose of the plankton food given to the larva should be further modified to improve the sustaining condition.

Although electroporation has been widely utilized in mammalian cells as well as in bacteria, its effects on the eggs of marine molluscs are still poorly understood. Moreover, the delivery of foreign DNA through electroporation cannot be visually confirmed and the transfection efficiency cannot be determined unlike in the microinjection method. Based on a time constant protocol, combinations of a variety of capacitances and voltages were tested, and relatively effective conditions for introducing the foreign DNA into the egg were identified, which was confirmed by repeated experiments and the following PCR amplifications. Meanwhile, chemical mediated transfection towards the embryos of *P. fucata* seems also effective. However, hitherto the expression of the foreign DNA could not be verified via fluorescent microscope, which might be attributed to several reasons: 1, promoters. There is no previous report of any confirmed effective promoter for expressing foreign DNA in *P. fucata*. Also, because the intrinsic deficiency of the genomic data of *P. fucata*, it is impossible to identify endogenous promoters region from the most commonly used house-keeping genes, such as Elongation factor 1 alpha (EF-1α) and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Meanwhile, transgenic studies on non-model organisms, such as molluscs, are limited. 2, *GFP*-coding sequence. I applied a GFP gene that was proved to be effective in *Ciona intestinalis*. However, it is not

guaranteed to be expressed correctly in molluscan species. 3, DNA delivery efficiency. Although the delivery of DNA into the embryos by electroporation was confirmed by PCR amplification, it is not certain if the introduced DNA reached the enough amounts to be effective. Therefore, in the future, the experiment can be improved mainly in three ways: 1, other promoters of other house-keeping genes or different regions of the beta-tubulin gene of *P. fucata* should be tested; 2, different nucleotide sequences encoding the GFP gene should be tested; 3, other electroporation conditions should be tested for higher efficiency of DNA delivery.

We characterized the coding sequence of SPARC gene of *N. fuscoviridus* for the first time and successfully generated relevant plasmids and RNA constructs of CRISRP/Cas9. However, before further discussion about the effect of this novel gene knock-out technique on marine molluscs, there are still some works needed to be done concerning microinjections to guarantee that those constructs are delivered to the embryos frequently without causing any side-effects.

Chapter 5 General discussion and future perspectives

5.1 The first larval shell proteomes

In a first attempt to extract shell matrices from molluscan larval shells, in chapter 2, a total of 111 and 71 SMPs were identified from the larval shells of *C. gigas* and *P. fucata*, respectively. Surprisingly, only four SMPs were shared by the larval and adult shells in each species. Though the BLAST results indicated that Nacrein-like and Pif/BMSP-like proteins are contained in both the larval and adult SMP repertoires in the two species, gene IDs of these protein genes are different between the larval and the adult SMPs, in other words, different homologs of these gene families are used by the larval and the adult shells in each species. Carbohydrate-binding domain 14 (CBM_14), von Willebrand factor type A (VWA), carbonic anhydrase (CA) and EF-hand domains were identified in both the larval and adult shells of the two species. These observations indicated their probable important roles in the generation of both larval and adult shells in those molluscs. With only a few exceptions, the expression patterns of the transcripts for most of the larval and adult SMPs, including the ones containing those common domains for larval and adult shells, exhibited a larval or adult stage-specific feature. Taken together, these intrinsic differences between the repertories of SMPs of larval and adult molluscs suggest that the larval SMPs and adult SMPs have evolved independently with each other, and that the larval and adult shells are subjected to different milieux of adaptation.

Despite some domains are shared by larval shells of two species, the larval SMP of the two species are as different as those of their adult counterparts, thus, it proved difficult to test the hypothesis given that the larva may preserve more ancestral features than the adult (Davidson 1990; Davidson, et al. 1995; Peterson, et al. 1997) by comparing the components of the SMP repertoires of larval shells. However, the potential “transitional phase” of heterochronic evolution manifested by the gene

expression patterns of the transcript for the larval SMP Pfu_aug2.0_421.1_04155.t1 (Chapter 3) points to a possibility that characterizations of the developmental gene networks controlling the formation of larval and adult shells may help to solve this question of antiquity of the larval shells in future.

Compared with those of *P. fucata* (Zhao, et al. 2018), more house-keeping proteins were found in the adult shell of *C. gigas* (Zhang, et al. 2012), a similar condition was observed in the larval shell proteomes, which suggested a higher extent of involvement of cells for the larval shell formation in *C. gigas* than in *P. fucata*, and therefore the cell-mediated shell formation model (Mount, et al. 2004; Wang, et al. 2013b) could be more adequate to explain the shell formation mechanisms than the shell matrix mediated model for both larval and adult shells of *C. gigas*. However, as it was mentioned in Chapter 2, the possibility of contamination of the sample by unremoved soft tissues cannot be totally excluded, despite the fact that soft tissues could not be observed after the NaOH treatment under the microscope. Also, considering the mode of development of the larval shells involving shell gland, it appears difficult to apply the cell-mediated model to the processes of larval shell formation.

5.2 Phylogenetic analyses performed on SMPs containing common domains shared by larval and adult shells of two pteriomorph bivalves, *C. gigas* and *P. fucata*

In order to test the previous hypotheses of whether the calcification in molluscs was originated from the ancestral functions or was independently obtained by different metazoan lineages (Marin, et al. 2007), in Chapter 3, phylogenetic analyses were performed to the common proteins shared by both the larval and adult shells in the two pteriomorph species, *C. gigas* ad *P. fucata* (Zhao, et al. 2018). The results indicated that at least the carbonic anhydrase (CA) proteins were recruited to the shell much more recently than expected. As the hydrolytic enzyme of carbon dioxide in the equation, $CO_2 + H_2O \rightarrow HCO_3^- + H^+$, CA is important for providing HCO_3^- that reacts

with Ca^{2+} to form $CaCO_3$. It is well expanded in molluscan species as well as in other metazoan taxa, which bear calcium carbonate skeletons (Zhao, et al. 2018). Therefore, it is surprising to see that after the speciation of the ancestors of *C. gigas* and *P. fucata*, a single CA gene, which may have or may not have encoded an SMP, gave rise to multiple copies of CA genes in each lineage, and some of them were deployed as adult SMPs, while others were deployed as larval SMPs in each lineage independently. Therefore, although CA was once taken as an evidence to support the "ancient heritage" scenario of the origin of calcification of molluscs (Marin, et al. 2007), it turned out to support the "recent heritage and fast evolution" scenario (Marin, et al. 2007) based on the results obtained in this study.

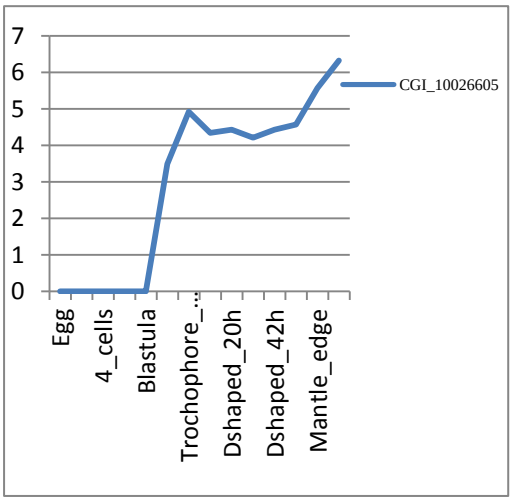
In Chapter 3, phylogenetic analyses performed on the VWA, CB and Laminin G domains of VWA-CB dcps indicated that BMSPs and other VWA-CB dcps were recruited to the shell of the common ancestor of bivalves and gastropods before the divergence, which is a different conclusion to the history of domain duplications in BMSPs previously published by Suzuki, et al. 2013.

EF-hand domain-containing proteins, CGI_10012241 and pfu_aug2.0_3796.1_22455.t1, are also shared by the larval shells of the two bivalve species, both of which are not identical to any previously reported protein (Zhao, et al. 2018) and both exhibited explicitly larval stage-specific expression patterns (Fig. 2.6). Although data are not shown, a signal peptide was found in CGI_10012241, but not in pfu_aug2.0_3796.1_22455.t1, a fact which indicates that at least the former is a secreted protein. EF-hand proteins were reported to be expressed specially by the cells of mineralization-related tissues and to concentrate calcium ions, therefore are thought to play important roles in the regulation of shell formation (Huang, et al. 2007; Liu, et al. 2007). Similar functions of these proteins in the larval shell can be expected.

Family 20 chitobiase, Pfu_aug2.0_6.1_20027.t1 was found in the larval shell of *P.*

fucata (Zhao, et al. 2018), and later, CGI_10007856 was suggested to be a chitobiase candidate in the larval shell of *C. gigas* (Chapter 3). It is curious that no Family 18 chitinase was identified from the larval shells of either species, because chitobiases cannot work without chitinases (Suzuki, et al. 2007). The expression patterns of transcripts show a strong expression at the stages later than the blastula stage of Family 18 chitinase, or CGI_10026605, which is the adult SMP with Glyco_18 domain (IPR011583) in *C. gigas*, and a detectable expression of pfu_aug2.0_194.1_13763.t1, one of the adult SMPs with a Glyco_18 domain (IPR011583) in *P. fucata* at the trochophore stage (as well as at the stage between the 4 cells and blastula stages and at the adult stage), respectively (Figs. 5.1 A and B). Those observations suggest that SMPs of Family 18 chitinases possibly exist in larval shells, but have not been detected.

A.



B.

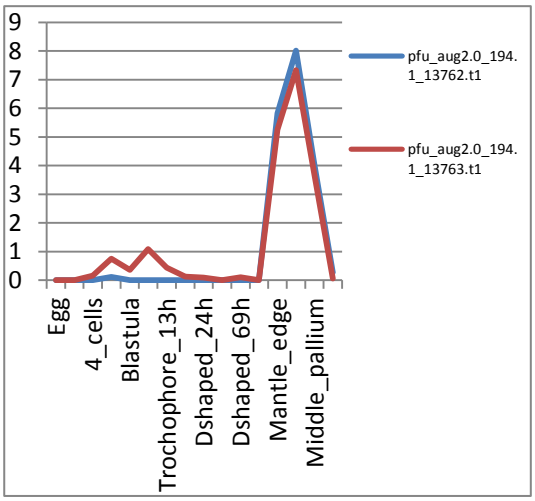


Fig. 5.1. Expression pattern of adult SMPs containing Glyco_18 domain (IPR011583) in *C. gigas* and *P. fucata*, respectively. CGI_10026605 (A). pfu_aug2.0_194.1_13762.t1 and pfu_aug2.0_194.1_13763.t1 (B) based on transcriptomic and proteomic data (Takeuchi, et al. 2012; Takeuchi, et al. 2016; Zhang, et al. 2012; Zhao, et al. 2018).

5.3 Trials on generating transgenic molluscan lineages

In chapter 4, electroporation and chemical mediated transfection were tested on the fertilized egg of *P. fucata* in an attempt to introduce *GFP*-coding plasmids, and the successful introduction was demonstrated by PCR amplification. However, fluorescent embryos were not detected after the transfection treatments. On the other hand, constructs of CRISPR/Cas9 that target on the SPARC gene of *Nipponacmea fuscoviridus* were generated and introduced into the fertilized eggs. But until now, no positive results of successful introduction of the constructs have been obtained. Next, in order to obtain the expression of *GFP*, different plasmids containing other promoters and different *GPF*-coding sequences based on the codon usage of *P. fucata* will be tested. Also, the delivery of CRISPR/Cas9 constructs into fertilized eggs of *N. fuscoviridus* will be continued.

5. 4 Future perspectives

With the help of the combination of proteomic, genomic, and transcriptomic techniques, my study revealed that the larval and adult molluscan shells are controlled by distinct gene repertories and have different recruitment histories of SMPs. Some parts of these conclusions are far beyond the general expectations. However, just like many other researches of molluscan SMPs, this study is restricted to the commercial species of only two pteriomorph bivalves, the oyster *C. gigas* and the pearl oyster *P. fucata.* Our knowledge about the larval SMPs of other molluscan species is still poor. In the future, different groups of molluscs, such as the gastropods, should be included in the larval shell proteome project, though the study might be impeded by the size of molluscan larval shells. More comprehensive information on the larval SMPs from the systematic sampling will definitely add to our understanding of the whole processes of shell formation and help us date back to the origin of this physiologically controlled mineralization with fine and diverse structures.

In the project of functional analyses of SMPs via transgenic molluscs, no conclusive results have been gained at this moment. However, I still have obtained some promising data. In the next step, different *GFP*-coding sequences and promoters need to be tested. Also, the distinct gene repertoires of the larval SMPs, one of the most important knowledge obtained from chapter 2 of this study, should be taken into consideration, for example, as a reference for selecting the target to perform CRISPR/Cas9 gene knockout.

References

Addadi L, Joester D, Nudelman F, Weiner S 2006. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. Chemistry–A European Journal 12: 980-987.

Beumer KJ, Trautman JK, Christian M, Dahlem TJ, Lake CM, Hawley RS, Grunwald DJ, Voytas DF, Carroll D 2013. Comparing ZFNs and TALENs for gene targeting in Drosophila. G3: Genes, Genomes, Genetics: g3. 113.007260.

Blitz IL, Biesinger J, Xie X, Cho KW 2013. Biallelic genome modification in F0 Xenopus tropicalis embryos using the CRISPR/Cas system. Genesis 51: 827-834.

Boggild OB 1930. The shell structure of the mollusks. Det Kongelige Danske Videnskabernes Selskabs Skrifter. Naturvidenskabelig og Mathematisk Afdeling, Raekke 9 2: 231-326.

Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Bonar DB 1976. Molluscan metamorphosis: a study in tissue transformation. American Zoologist 16: 573-591.

Buchanan JT, Nickens AD, Cooper RK, Tiersch TR 2001. Transfection of eastern oyster (Crassotrea virginica) embryos. Marine Biotechnology 3: 322-335.

Cadoret J-P, Boulo V, Gendreau S, Mialhe E 1997a. Promoters from Drosophila heat shock protein and cytomegalovirus drive transient expression of luciferase introduced by particle bombardment into embryos of the oyster Crassostrea gigas. Journal of Biotechnology 56: 183-189.

Cadoret J, Gendreau S, Delecheneau J, Rousseau C, Mialhe E 1997b. Microinjection of bivalve eggs: application in genetics. Molecular marine biology and biotechnology 6: 72-77.

Carriker MR editor. Proc Natl Shellfish Assoc. 1979.

Carter JG 1990. Evolutionary significance of shell microstructure in the Palaeotaxodonta, Pteriomorphia and Isofilibranchia (Bivalvia: Mollusca). Skeletal biomineralization: patterns, processes and evolutionary trends 1: 135-296.

Carter JG, Clark GR 1985. Classification and phylogenetic significance of molluscan shell microstructure. Studies in Geology, Notes for a Short Course 13: 50-71.

Cather JN 1967. Cellular interactions in the development of the shell gland of the gastropod, Ilyanassa. Journal of Experimental Zoology 166: 205-223.

Cather JN, Verdonk NH, RENÉ DOHMEN M 1976. Role of the vegetal body in the regulation of development in Bithynia tentaculata (Prosobranchia, Gastropoda). American Zoologist 16: 455-468.

Chateigner D, Hedegaard C, Wenk H-R 2000. Mollusc shell microstructures and crystallographic textures. Journal of Structural Geology 22: 1723-1735.

Choi JW, Sutor SL, Lindquist L, Evans GL, Madden BJ, Bergen III HR, Hefferan TE, Yaszemski MJ, Bram RJ 2009. Severe osteogenesis imperfecta in cyclophilin B–deficient mice. PLoS genetics 5: e1000750.

Chomczynski P, Sacchi N 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical biochemistry 162: 156-159.

Chomez P, De Backer O, Bertrand M, De Plaen E, Boon T, Lucas S 2001. An overview of the MAGE gene family with the identification of all human members of the family. Cancer research 61: 5544-5551.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini L 2013. Multiplex genome engineering using CRISPR/Cas systems. Science: 1231143.

Cunniffe GM, O'Brien FJ 2011. Collagen scaffolds for orthopedic regenerative medicine. Jom 63: 66.

Davidson EH 1990. How embryos work: a comparative view of diverse modes of cell fate specification.

Development 108: 365-389.
Davidson EH, Peterson KJ, Cameron RA 1995. Origin of bilaterian body plans: evolution of developmental regulatory mechanisms. Science 270: 1319-1325.
Deguchi R 2007. Fertilization causes a single Ca2+ increase that fully depends on Ca2+ influx in oocytes of limpets (Phylum Mollusca, Class Gastropoda). Developmental biology 304: 652-663.
Drake JL, Mass T, Haramaty L, Zelzion E, Bhattacharya D, Falkowski PG 2013. Proteomic analysis of skeletal organic matrix from the stony coral Stylophora pistillata. Proceedings of the National Academy of Sciences: 201301419.
Edwards OR, HoY MA 1993. Polymorphism in two parasitoids detected using random amplified polymorphic DNA polymerase chain reaction. Biological Control 3: 243-257.
Eyster L editor. American Zoologist. 1982.
Eyster LS 1986. Shell inorganic composition and onset of shell mineralization during bivalve and gastropod embryogenesis. The Biological Bulletin 170: 211-231.
Eyster LS 1983. Ultrastructure of early embryonic shell formation in the opisthobranch gastropod *Aeolidia papillosa*. The Biological Bulletin 165: 394-408. doi: 10.2307/1541204
Eyster LS, Morse MP 1984. Early shell formation during molluscan embryogenesis, with new studies on the surf clam, Spisula solidissima. American Zoologist 24: 871-882.
Feng W, Sun W 2003. Phosphate replicated and replaced microstructure of molluscan shells from the earliest Cambrian of China. Acta Palaeontologica Polonica 48.
Fisher C, Skibinski DO 1990. Sex-biased mitochondrial DNA heteroplasmy in the marine mussel Mytilus. Proc. R. Soc. Lond. B 242: 149-156.
Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA 2013. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nature methods 10: 741.
Fujii N, Minetti C, Nakhasi H, Chen S, Barbehenn E, Nunes P, Nguyen N 1992. Isolation, cDNA cloning, and characterization of an 18-kDa hemagglutinin and amebocyte aggregation factor from Limulus polyphemus. Journal of Biological Chemistry 267: 22452-22459.
Fujimura T, Wada K, Iwaki T 1995. Development and morphology of the pearl oyster larvae, Pinctada fucata. Japanese Journal of Malacology (Japan).
Fukazawa H, Hirai H, Hori H, Roberts RD, Nukaya H, Ishida H, Tsuji K 2001. Induction of abalone larval metamorphosis by thyroid hormones. Fisheries Science 67: 985-988.
Furuhashi T, Schwarzinger C, Miksik I, Smrz M, Beran A 2009. Molluscan shell evolution with review of shell calcification hypothesis. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 154: 351-371.
Gardin A, White J 2011. The Sanger Mouse Genetics Programme: high throughput characterisation of knockout mice. Acta Ophthalmologica 89.
Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468: 67.
Garstang W 1922. The Theory of Recapitulation: A Critical Re‐statement of the Biogenetic Law. Journal of the Linnean Society of London, Zoology 35: 81-101.
Gasiunas G, Barrangou R, Horvath P, Siksnys V 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences 109: E2579-E2586.
Gotliv BA, Kessler N, Sumerel JL, Morse DE, Tuross N, Addadi L, Weiner S 2005. Asprich: A novel

aspartic acid‐rich protein family from the prismatic shell matrix of the bivalve Atrina rigida. ChemBioChem 6: 304-314.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29: 644.
Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics: genetics. 113.152710.
Gropper SS, Smith JL. 2012. Advanced nutrition and human metabolism: Cengage Learning.
Hasegawa Y, Uchiyama K 2005. cDNA clonings of shell matrix proteins from scallop shell. Fisheries Science 71: 1174-1178.
Hashimoto N, Kurita Y, Wada H 2012. Developmental role of dpp in the gastropod shell plate and co-option of the dpp signaling pathway in the evolution of the operculum. Developmental biology 366: 367-373.
Haviv F, Bradley MF, Kalvin DM, Schneider AJ, Davidson DJ, Majest SM, McKay LM, Haskell CJ, Bell RL, Nguyen B 2005. Thrombospondin-1 mimetic peptide inhibitors of angiogenesis and tumor growth: design, synthesis, and optimization of pharmacokinetics and biological activities. Journal of medicinal chemistry 48: 2838-2846.
Hinman VF, O'Brien EK, Richards GS, Degnan BM 2003. Expression of anterior Hox genes during larval development of the gastropod Haliotis asinina. Evolution & development 5: 508-521.
Huang J, Zhang C, Ma Z, Xie L, Zhang R 2007. A novel extracellular EF-hand protein involved in the shell formation of pearl oyster. Biochimica et Biophysica Acta (BBA)-General Subjects 1770: 1037-1044.
Huber O, Sumper M 1994. Algal‐CAMs: isoforms of a cell adhesion molecule in embryos of the alga Volvox with homology to Drosophila fasciclin I. The EMBO Journal 13: 4212-4222.
Hwang WY, Fu Y, Reyon D, Maeder ML, Kaini P, Sander JD, Joung JK, Peterson RT, Yeh J-RJ 2013. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PloS one 8: e68708.
Isowa Y, Sarashina I, Oshima K, Kito K, Hattori M, Endo K 2015. Proteome analysis of shell matrix proteins in the brachiopod Laqueus rubellus. Proteome science 13: 21.
Ivester M editor. American Zoologist. 1972.
Iwata K 1980. Mineralization and architecture of the larval shell of *Haliotis discus hannai* Ino,(Archaeogastropoda). Journal of the Faculty of Science, Hokkaido University. Series 4, Geology and mineralogy 19: 305-320.
Iwata K, Akamatsu M 1975. A study on the prodissoconch of Patinopecten yessoensis. Ann. Rept. His. Mus. Hokkaido 10: 11-18.
Jablonski D 1980. Molluscan larval shell morphology, ecological and paleontological applications. Skeletal growth of aquatic organisms: 323-377.
Jackson DJ, Mann K, Häussermann V, Schilhabel MB, Lüter C, Griesshaber E, Schmahl W, Wörheide G 2015. The Magellania venosa biomineralizing proteome: a window into brachiopod shell evolution. Genome biology and evolution 7: 1349-1362.
Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, Degnan BM 2006. A rapidly evolving secretome builds and patterns a sea shell. BMC biology 4: 40.
Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C 2009. Parallel evolution of nacre building gene sets in molluscs. Molecular biology and evolution 27: 591-608.

Jacobs DK, Wray CG, Wedeen CJ, Kostriken R, DeSalle R, Staton JL, Gates RD, Lindberg DR 2000. Molluscan engrailed expression, serial organization, and shell evolution. Evolution & development 2: 340-347.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science: 1225829.

Jolly C, Berland S, Milet C, Borzeix S, Lopez E, Doumenc D 2004. Zona localization of shell matrix proteins in mantle of Haliotis tuberculata (Mollusca, Gastropoda). Marine Biotechnology 6: 541-551.

Kawasaki K, Suzuki T, Weiss KM 2004. Genetic basis for the evolution of vertebrate mineralized tissue. Proceedings of the National Academy of Sciences of the United States of America 101: 11356-11361. doi: 10.1073/pnas.0404279101

Killian CE, Wilt FH 2008. Molecular aspects of biomineralization of the echinoderm endoskeleton. Chemical reviews 108: 4463-4474.

Kim J-E, Kim S-J, Lee B-H, Park R-W, Kim K-S, Kim I-S 2000. Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-β-induced gene, βig-h3. Journal of Biological Chemistry 275: 30907-30915.

Kintsu H, Okumura T, Negishi L, Ifuku S, Kogure T, Sakuda S, Suzuki M 2017. Crystal defects induced by chitin and chitinolytic enzymes in the prismatic layer of Pinctada fucata. Biochemical and biophysical research communications 489: 89-95.

Kniprath E 1980. Larval development of the shell and the shell gland inMytilus (Bivalvia). Wilhelm Roux's archives of developmental biology 188: 201-204.

KNIPRATH E 1981. Ontogeny of the molluscan shell field: a review. Zoologica Scripta 10: 61-79.

Kniprath E 1977. Zur Ontogenese des Schalenfeldes vonLymnaea stagnalis. Roux's archives of developmental biology 181: 11-30.

Knoll AH, Carroll SB 1999. Early animal evolution: emerging views from comparative biology and geology. Science 284: 2129-2137.

Kobayashi I 1969. Internal microstructure of the shell of bivalve molluscs. American Zoologist 9: 663-672.

Kono M, Hayashi N, Samata T 2000. Molecular mechanism of the nacreous layer formation in Pinctada maxima. Biochemical and biophysical research communications 269: 213-218.

Kouchinsky A 2000. Shell microstructures in Early Cambrian molluscs. Acta Palaeontologica Polonica 45.

Kretsinger RH 1976. Calcium-binding proteins. Annual review of biochemistry 45: 239-266.

Krieg PA, Melton D 1984. Functional messenger RNAs are produced by SP6 in vitro transcription of cloned cDNAs. Nucleic acids research 12: 7057-7070.

Kumar S, Stecher G, Tamura K 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution 33: 1870-1874.

LaBarbera M 1974. Calcification of the first larval shell of Tridacna squamosa (Tridacnidae: Bivalvia). Marine Biology 25: 233-238.

Langmead B, Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357.

Lecointre G, Le Guyader H 2001. Classification phylogenetique du vivant.— Paris. France: Belin.

Liu C, Li S, Kong J, Liu Y, Wang T, Xie L, Zhang R 2015. In-depth proteomic analysis of shell matrix proteins of Pinctada fucata. Scientific Reports 5: 17269. doi: 10.1038/srep17269

Liu H-L, Liu S-F, Ge Y-J, Liu J, Wang X-Y, Xie L-P, Zhang R-Q, Wang Z 2007. Identification and characterization of a biomineralization related gene PFMG1 highly expressed in the mantle of Pinctada

fucata. Biochemistry 46: 844-851.
Lowenstam HA, Weiner S. 1989. On biomineralization: Oxford University Press on Demand.
Luo Y-J, Takeuchi T, Koyanagi R, Yamada L, Kanda M, Khalturina M, Fujie M, Yamasaki S-i, Endo K, Satoh N 2015. The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. Nature communications 6: 8301.
Lydie M, GOLUBIC S, LE CAMPION-ALSUMARD T, PAYRI C 2001. Developmental aspects of biomineralisation in the Polynesian pearl oyster Pinctada margaritifera var. cumingii. Oceanologica acta 24: 37-49.
Malbon CC 2004. Frizzleds: new members of the superfamily of G-protein-coupled receptors. Front Biosci 9: 1048-1058.
Mann K, Edsinger-Gonzales E, Mann M 2012. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet Lottia gigantea. Proteome science 10: 28.
Mann S 1988. Molecular recognition in biomineralization. Nature 332: 119.
Marie B, Jackson DJ, Ramos‐Silva P, Zanella‐Cléon I, Guichard N, Marin F 2013. The shell‐forming proteome of Lottia gigantea reveals both deep conservations and lineage‐specific novelties. The FEBS journal 280: 214-232.
Marie B, Joubert C, Tayalé A, Zanella-Cléon I, Belliard C, Piquemal D, Cochennec-Laureau N, Marin F, Gueguen Y, Montagnani C 2012. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of Sciences 109: 20986-20991.
Marie B, Le Roy N, Zanella-Cléon I, Becchi M, Marin F 2011a. Molecular evolution of mollusc shell proteins: insights from proteomic analysis of the edible mussel Mytilus. Journal of molecular evolution 72: 531-546.
Marie B, Trinkler N, Zanella-Cleon I, Guichard N, Becchi M, Paillard C, Marin F 2011b. Proteomic identification of novel proteins from the calcifying shell matrix of the Manila clam Venerupis philippinarum. Marine Biotechnology 13: 955-962.
Marin F, Layrolle P, De Groot K, Westbroek P 2003. The origin of metazoan skeleton. Biomineralization: Formation, Diversity, Evolution, and Application. Eds: I Kobayashi, H Ozawa, Tokai University Press, Kanagawa.
Marin F, Luquet G, Marie B, Medakovic D 2007. Molluscan shell proteins: primary structure, origin, and evolution. Current topics in developmental biology 80: 209-276.
Martoja M. 1995. Mollusques: Institut océanographique.
Marxen JC, Becker W 1997. The organic shell matrix of the freshwater snail Biomphalaria glabrata. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 118: 23-33.
Marxen JC, Nimtz M, Becker W, Mann K 2003. The major soluble 19.6 kDa protein of the organic shell matrix of the freshwater snail Biomphalaria glabrata is an N-glycosylated dermatopontin. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1650: 92-98.
Mass T, Drake JL, Peters EC, Jiang W, Falkowski PG 2014. Immunolocalization of skeletal matrix proteins in tissue and mineral of the coral Stylophora pistillata. Proceedings of the National Academy of Sciences 111: 12728-12733.
Maurer P, Hohenester E, Engel J 1996. Extracellular calcium-binding proteins. Current opinion in cell biology 8: 609-617.
Medaković D 2000. Carbonic anhydrase activity and biomineralization process in embryos, larvae and adult blue mussels Mytilus edulis L. Helgoland Marine Research 54: 1.

Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A 1996. A carbonic anhydrase from the nacreous layer in oyster pearls. Proceedings of the National Academy of Sciences 93: 9657-9660.
Miyamoto H, Miyoshi F, Kohno J 2005. The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc Pinctada fucata. Zoological science 22: 311-315.
Miyamoto H, Yano M, Miyashita T 2003. Similarities in the structure of nacrein, the shell-matrix protein, in a bivalve and a gastropod. Journal of molluscan studies 69: 87-89.
Miyashita T 2000. Identical carbonic anhydrase contributes to nacreous or prismatic layer formation in Pinctada fucata (Mollusca Bivalvia). Veliger 45: 250-255.
Morris AH, Kyriakides TR 2014. Matricellular proteins and biomaterials. Matrix Biology 37: 183-191.
Morris SC. 1998. The crucible of creation: the Burgess Shale and the rise of animals: Peterson's.
Moshel SM, Levine M, Collier J 1998. Shell differentiation and engrailed expression in the Ilyanassa embryo. Development genes and evolution 208: 135-141.
Mount AS, Wheeler A, Paradkar RP, Snider D 2004. Hemocyte-mediated shell mineralization in the eastern oyster. Science 304: 297-300.
Nagai K, Yano M, Morimoto K, Miyamoto H 2007. Tyrosinase localization in mollusc shells. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 146: 207-214.
Neame P, Choi H, Rosenberg L 1989. The isolation and primary structure of a 22-kDa extracellular matrix protein from bovine skin. Journal of Biological Chemistry 264: 5474-5479.
Nederbragt AJ, van Loon AE, Dictus WJ 2002. Expression of Patella vulgata orthologs of engrailed and dpp-BMP2/4 in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. Developmental biology 246: 341-355.
Norizuki M, Samata T 2008. Distribution and function of the nacrein-related proteins inferred from structural analysis. Marine Biotechnology 10: 234-241.
Peifer M, Berg S, Reynolds AB 1994. A repeating amino acid motif shared by proteins with diverse cellular roles. cell 76: 789-791.
Peifer M, Rauskolb C, Williams M, Riggleman B, Wieschaus E 1991. The segment polarity gene armadillo interacts with the wingless signaling pathway in both embryonic and adult pattern formation. Development 111: 1029-1043.
Perry KJ, Henry JQ 2015. CRISPR/C as9‐mediated genome modification in the mollusc, Crepidula fornicata. Genesis 53: 237-244.
Peterson KJ, Cameron RA, Davidson EH 1997. Set‐aside cells in maximal indirect development: evolutionary and developmental significance. BioEssays 19: 623-631.
Powers D, Kirby V, Cole T, Hereford L 1995. Electroporation as an effective means of introducing DNA into abalone (Haliotis rufescens) embryos. Molecular marine biology and biotechnology 4: 369-375.
Rahman MA, Shinjo R, Oomori T, Wörheide G 2013. Analysis of the proteinaceous components of the organic matrix of calcitic sclerites from the soft coral Sinularia sp. PloS one 8: e58781.
Roberts A, Pachter L 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. Nature methods 10: 71.
Rosado CJ, Kondos S, Bull TE, Kuiper MJ, Law RH, Buckle AM, Voskoboinik I, Bird PI, Trapani JA, Whisstock JC 2008. The MACPF/CDC family of pore‐forming toxins. Cellular microbiology 10: 1765-1774.
Runnegar B 1996. Early evolution of the Mollusca: the fossil record. Origin and evolutionary radiation

of the Mollusca.
Sakuma T, Hosoi S, Woltjen K, Suzuki Ki, Kashiwagi K, Wada H, Ochiai H, Miyamoto T, Kawai N, Sasakura Y 2013. Efficient TALEN construction and evaluation methods for human cell and animal applications. Genes to Cells 18: 315-326.
Saleuddin A, PETIT HP. 1983. The mode of formation and the structure of the periostracum. In. The Mollusca, Volume 4: Elsevier. p. 199-234.
Sarashina I, Endo K 2001. The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop Patinopecten yessoensis. Marine Biotechnology 3: 362-369.
Sarashina I, Endo K 1998. Primary structure of a soluble matrix protein of scallop shell; implications for calcium carbonate biomineralization. American Mineralogist 83: 1510-1515.
Sarashina I, Yamaguchi H, Haga T, Iijima M, Chiba S, Endo K 2006. Molecular evolution and functionally important structures of molluscan dermatopontin: Implications for the origins of molluscan shell matrix proteins. Journal of molecular evolution 62: 307-318.
Schütze J, Skorokhod A, Müller IM, Müller WE 2001. Molecular evolution of the metazoan extracellular matrix: cloning and expression of structural proteins from the demosponges Suberites domuncula and Geodia cydonium. Journal of molecular evolution 53: 402-415.
Seed R 1980. Shell growth and form in the Bivalvia. " Skeletal growth of aquatic organisms": 23-61.
Shubin NH, Marshall CR 2000. Fossils, genes, and the origin of novelty. Paleobiology: 324-340.
Simkiss KW, Wilbur K 1989. K.. 1989. Biomineralization. Cell Biology and Mineral Deposition, Academic Press, San Diego.
Smith‐Keune C, Jerry DR 2009. High levels of intra‐specific variation in the NG repeat region of the Pinctada maxima N66 organic matrix protein. Aquaculture research 40: 1054-1063.
Stenzel HB 1964. Oysters: composition of the larval shell. Science 145: 155-156.
Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakashima K, Takahashi T 1997. Structures of mollusc shell framework proteins. Nature 387: 563.
Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, Lan Y, Fields CJ, Hui JHL, Zhang W 2017. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature ecology & evolution 1: 0121.
Suzuki M, Iwashima A, Kimura M, Kogure T, Nagasawa H 2013. The molecular evolution of the Pif family proteins in various species of mollusks. Marine Biotechnology 15: 145-158.
Suzuki M, Iwashima A, Tsutsui N, Ohira T, Kogure T, Nagasawa H 2011. Identification and Characterisation of a Calcium Carbonate‐Binding Protein, Blue Mussel Shell Protein (BMSP), from the Nacreous Layer. ChemBioChem 12: 2478-2487.
Suzuki M, Murayama E, Inoue H, Ozaki N, Tohse H, Kogure T, Nagasawa H 2004. Characterization of Prismalin-14, a novel matrix protein from the prismatic layer of the Japanese pearl oyster (Pinctada fucata). Biochemical Journal 382: 205-213.
Suzuki M, Sakuda S, Nagasawa H 2007. Identification of chitin in the prismatic layer of the shell and a chitin synthase gene from the Japanese pearl oyster, Pinctada fucata. Bioscience, biotechnology, and biochemistry 71: 1735-1744.
Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T, Nagasawa H 2009. An acidic matrix protein, Pif, is a key macromolecule for nacre formation. Science 325: 1388-1390.
Takeuchi T, Endo K 2006. Biphasic and dually coordinated expression of the genes encoding major shell matrix proteins in the pearl oyster Pinctada fucata. Marine Biotechnology 8: 52-61.

Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T, Shoguchi E, Fujiwara M, Shinzato C, Hisata K 2012. Draft genome of the pearl oyster Pinctada fucata: a platform for understanding bivalve biology. DNA research 19: 117-130.

Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M, Goto H, Yamasaki S, Nagai K, Morino Y 2016. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. Zoological letters 2: 3.

Takeuchi T, Sarashina I, Iijima M, Endo K 2008. In vitro regulation of CaCO3 crystal polymorphism by the highly acidic molluscan shell protein Aspein. FEBS letters 582: 591-596.

Taylor JD 1973. The structural evolution of the bivalve shell. Palaeontology 16: 519-534.

Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE 1996. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay–Sachs disease. Nature Structural and Molecular Biology 3: 638.

Thomas R, Shearman RM, Stewart GW 2000. Evolutionary exploitation of design options by the first animals with hard skeletons. Science 288: 1239-1242.

Tian J, Cai T, Yuan Z, Wang H, Liu L, Haas M, Maksimova E, Huang X-Y, Xie Z-J 2006. Binding of Src to Na+/K+-ATPase forms a functional signaling complex. Molecular biology of the cell 17: 317-326.

Timmermans LP 1968. Studies on shell formation in molluscs. Netherlands Journal of Zoology 19: 413-523.

Timpl R, Rohde H, Robey PG, Rennard SI, Foidart J-M, Martin GR 1979. Laminin--a glycoprotein from basement membranes. Journal of Biological Chemistry 254: 9933-9937.

Tracey S, Todd JA, Erwin DH, Benton M 1993. The fossil record 2.

Tsai H-J, Lai C-H, Yang H-S 1997. Sperm as a carrier to introduce an exogenous DNA fragment into the oocyte of Japanese abalone (Haliotis divorsicolor suportexta). Transgenic research 6: 85-95.

Tsukamoto D, Sarashina I, Endo K 2004. Structure and expression of an unusually acidic matrix protein of pearl oyster shells. Biochemical and biophysical research communications 320: 1175-1180.

Uozumi S 1981. The evolution of shell structures in the bivalvia. Study of Molluscan Paleobiology, Professor Masae Omori Memorial Volume: 63-77.

Voskoboinik I, Smyth MJ, Trapani JA 2006. Perforin-mediated target-cell death and immune homeostasis. Nature Reviews Immunology 6: 940.

Waller TR 1981. Functional morphology and development of veliger larvae of the European oyster, Ostrea edulis Linné.

Wallin R, Wajih N, Greenwood GT, Sane DC 2001. Arterial calcification: a review of mechanisms, animal models, and the prospects for therapy. Medicinal research reviews 21: 274-301.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R 2013a. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. cell 153: 910-918.

Wang X, Li L, Zhu Y, Du Y, Song X, Chen Y, Huang R, Que H, Fang X, Zhang G 2013b. Oyster shell proteins originate from multiple organs and their probable transport pathway to the shell formation front. PloS one 8: e66522.

Wanninger A, Haszprunar G 2001. The expression of an engrailed protein during embryonic shell formation of the tusk‐shell, Antalis entalis (Mollusca, Scaphopoda). Evolution & development 3: 312-321.

Watanabe T, Noji S, Mito T 2014. Gene knockout by targeted mutagenesis in a hemimetabolous insect, the two-spotted cricket Gryllus bimaculatus, using TALENs. Methods 69: 17-21.

Weiner S 1979. Aspartic acid-rich proteins: major components of the soluble organic matrix of mollusk shells. Calcified Tissue International 29: 163-167.
Weiner S 1983. Mollusk shell formation: isolation of two organic matrix proteins associated with calcite deposition in the bivalve Mytilus californianus. Biochemistry 22: 4139-4145.
Weiner S, Hood L 1975. Soluble protein of the organic matrix of mollusk shells: a potential template for shell formation. Science 190: 987-989.
Weiner S, Traub W 1984. Macromolecules in mollusc shells and their functions in biomineralization. Phil. Trans. R. Soc. Lond. B 304: 425-434.
Weiss IM, Schönitzer V, Eichner N, Sumper M 2006. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. FEBS letters 580: 1846-1852.
Weiss IM, Tuross N, Addadi L, Weiner S 2002. Mollusc larval shell formation: amorphous calcium carbonate is a precursor phase for aragonite. Journal of Experimental Zoology 293: 478-491.
WILBUR KM, Saleuddin A. 1983. Shell formation. In. The Mollusca, Volume 4: Elsevier. p. 235-287.
Wintergerst ES, Maggini S, Hornig DH 2006. Immune-enhancing role of vitamin C and zinc and effect on clinical conditions. Annals of Nutrition and Metabolism 50: 85-94.
Yano M, Nagai K, Morimoto K, Miyamoto H 2006. Shematrin: a family of glycine-rich structural proteins in the shell of the pearl oyster Pinctada fucata. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 144: 254-262.
Zhang C, Xie L, Huang J, Liu X, Zhang R 2006. A novel matrix protein family participating in the prismatic layer framework formation of pearl oyster, Pinctada fucata. Biochemical and biophysical research communications 344: 735-740.
Zhang C, Zhang R 2006. Matrix proteins in the outer shells of molluscs. Marine Biotechnology 8: 572-586.
Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, Xu Z, Liu Y, Domazet-Lošo T, Du Y, Sun X, Zhang S, Liu B, Cheng P, Jiang X, Li J, Fan D, Wang W, Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li N, Wang J, Chen M, He Y, Tan F, Song X, Zheng Q, Huang R, Yang H, Du X, Chen L, Yang M, Gaffney PM, Wang S, Luo L, She Z, Ming Y, Huang W, Zhang S, Huang B, Zhang Y, Qu T, Ni P, Miao G, Wang J, Wang Q, Steinberg CEW, Wang H, Li N, Qian L, Zhang G, Li Y, Yang H, Liu X, Wang J, Yin Y, Wang J 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490: 49. doi: 10.1038/nature11413
https://www.nature.com/articles/nature11413#supplementary-information
Zhao R, Takeuchi T, Luo Y-J, Ishikawa A, Kobayashi T, Koyanagi R, Villar-Briones A, Yamada L, Sawada H, Iwanaga S 2018. Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation. Molecular biology and evolution 35: 2751-2761.
Zylbertal A, Yarom Y, Wagner S 2017. The Slow Dynamics of Intracellular Sodium Concentration Increase the Time Window of Neuronal Integration: A Simulation Study. Frontiers in computational neuroscience 11: 85.

Appendix I. Detailed information of peptides used for the identification of larval shell matrix proteins in *C. gigas*.

SMPs of *C. gigas*

						Soluble			Insoluble					
Accession	ΣCoverage	Σ# Proteins	Σ# Unique Peptides	Σ# Peptides	Σ# PSMs	Coverage A2	# Peptides A2	# PSM A2	Coverage B2	# Peptides B2	# PSM B2	# AAs	MW [kDa]	calc. pI
TRINITY_DN11263_c0_g1_i1	69.16	1	23	23	408	30.84	4	5	69.16	23	403	107	11.9	11.34
TRINITY_DN27156_c0_g1_i1	41.61	3	13	16	101	8.03	3	4	41.61	16	97	137	15.5	11.37
TRINITY_DN9130_c0_g2_i1	41.67	1	12	16	89	5.45	3	3	41.67	16	86	312	36.0	8.27
TRINITY_DN10003_c0_g1_i1	35.29	1	13	14	81	29.41	12	39	35.29	14	42	136	14.8	9.25
TRINITY_DN7741_c0_g1_i1	17.11	4	7	9	73	2.19	1	1	17.11	9	72	456	51.2	4.93
TRINITY_DN30273_c0_g1_i1	63.37	1	7	10	66				63.37	10	66	101	12.5	9.98
TRINITY_DN25821_c0_g1_i1	59.56	1	8	14	61	6.62	1	1	59.56	14	60	136	15.0	10.49
TRINITY_DN6144_c0_g1_i1	49.62	1	17	22	58	6.44	2	3	49.62	22	55	264	30.3	10.39
TRINITY_DN7858_c0_g1_i1	28.61	1	9	11	56				28.61	11	56	416	48.0	4.94
TRINITY_DN5856_c0_g1_i1	47.01	4	6	8	53	6.72	1	1	47.01	8	52	134	14.3	10.73
TRINITY_DN9353_c0_g1_i1	30.99	2	18	25	48				30.99	25	48	797	89.8	7.91
TRINITY_DN20596_c0_g1_i1	38.62	1	11	15	42				38.62	15	42	290	31.9	10.32
TRINITY_DN8071_c0_g1_i1	15.13	1	11	11	40				15.13	11	40	522	59.5	9.70
TRINITY_DN33014_c0_g1_i1	63.03	1	3	6	39				63.03	6	39	119	12.9	9.31
TRINITY_DN36948_c0_g1_i1	16.89	1	9	9	37				16.89	9	37	533	57.5	5.43
TRINITY_DN9441_c0_g5_i1	50.00	1	8	9	36				50.00	9	36	186	20.9	5.19
TRINITY_DN9248_c0_g1_i1	17.89	1	7	7	31	2.45	1	1	17.89	7	30	408	44.0	4.81
TRINITY_DN7058_c0_g1_i1	36.71	1	5	7	30	7.59	1	1	36.71	7	29	158	16.7	12.34
TRINITY_DN3983_c0_g1_i1	37.68	1	3	4	30				37.68	4	30	138	16.0	5.95
TRINITY_DN11315_c0_g1_i1	33.98	1	11	12	28				33.98	12	28	259	27.8	10.18
TRINITY_DN9441_c0_g3_i1	12.01	1	5	6	25				12.01	6	25	283	32.0	8.09
TRINITY_DN32132_c0_g1_i1	24.37	1	1	3	24				24.37	3	24	119	13.5	9.51
TRINITY_DN17428_c0_g1_i1	37.09	1	5	6	21				37.09	6	21	151	16.5	5.25
TRINITY_DN4535_c0_g1_i1	33.50	1	5	6	21				33.50	6	21	200	20.2	9.31
TRINITY_DN5666_c0_g1_i1	24.75	1	8	8	20				24.75	8	20	198	23.2	10.78
TRINITY_DN9062_c0_g1_i1	29.43	1	5	9	19	7.36	2	2	29.43	9	17	299	33.2	9.07
TRINITY_DN6468_c0_g1_i1	25.33	1	8	11	19				25.33	11	19	458	50.9	5.07
TRINITY_DN7441_c0_g1_i1	18.18	1	4	4	18				18.18	4	18	154	16.9	9.55
TRINITY_DN7441_c1_g1_i1	64.44	1	6	7	18				64.44	7	18	90	9.8	5.11
TRINITY_DN4623_c0_g1_i1	25.98	1	5	5	16				25.98	5	16	127	13.5	7.06
TRINITY_DN31972_c0_g1_i1	25.00	1	3	7	15				25.00	7	15	248	29.4	10.51
TRINITY_DN25349_c0_g1_i1	20.13	1	4	6	14	2.60	1	1	20.13	6	13	308	33.6	9.52
TRINITY_DN14712_c0_g2_i1	23.70	1	3	5	14				23.70	5	14	211	22.8	7.88
TRINITY_DN9353_c0_g2_i1	49.06	1	6	7	14				49.06	7	14	106	11.7	9.09
TRINITY_DN10332_c0_g1_i1	16.46	1	5	5	13	3.70	1	1	16.46	5	12	243	26.2	8.16
TRINITY_DN25899_c0_g1_i1	23.24	1	7	8	13				23.24	8	13	241	26.5	9.58
TRINITY_DN8042_c0_g1_i1	16.24	1	6	7	13				16.24	7	13	271	30.8	10.35
TRINITY_DN7367_c0_g2_i1	8.27	2	2	2	12	8.27	2	3	8.27	2	9	254	26.6	10.67
TRINITY_DN1225_c0_g1_i1	34.96	1	5	6	12				34.96	6	12	226	24.5	9.42
TRINITY_DN14296_c0_g1_i1	26.62	1	4	6	11				26.62	6	11	154	17.7	10.51
TRINITY_DN37303_c0_g1_i1	35.81	1	3	7	11				35.81	7	11	148	16.8	10.67
TRINITY_DN4797_c0_g1_i1	29.67	1	4	6	9				29.67	6	9	209	23.8	9.96
TRINITY_DN7147_c0_g1_i1	68.87	2	5	5	9				68.87	5	9	151	17.0	6.58
TRINITY_DN9130_c0_g1_i1	88.57	1	1	4	9				88.57	4	9	35	4.2	8.22
TRINITY_DN15802_c0_g1_i1	14.78	1	2	2	9				14.78	2	9	115	12.7	12.73
TRINITY_DN317_c0_g2_i1	20.15	1	2	2	9				20.15	2	9	134	14.1	8.48
TRINITY_DN37813_c0_g1_i1	13.50	1	2	2	9				13.50	2	9	274	28.7	7.42
TRINITY_DN2055_c0_g1_i1	20.28	1	3	4	8				20.28	4	8	217	25.2	10.26
TRINITY_DN7940_c0_g1_i1	6.94	1	4	6	8				6.94	6	8	1052	116.4	5.50
TRINITY_DN8163_c0_g4_i1	32.88	1	4	5	8				32.88	5	8	146	17.8	8.28
TRINITY_DN4963_c0_g1_i1	29.70	1	3	3	8				29.70	3	8	101	10.7	11.21
TRINITY_DN8587_c0_g1_i1	14.91	1	4	4	8				14.91	4	8	389	45.2	9.54
TRINITY_DN38090_c0_g1_i1	20.87	1	3	3	7	6.96	1	1	20.87	3	6	115	12.5	12.16
TRINITY_DN3662_c0_g1_i1	12.26	1	3	3	7				12.26	3	7	155	18.1	10.65
TRINITY_DN8163_c0_g2_i1	25.00	1	3	3	7				25.00	3	7	168	19.9	5.73
TRINITY_DN8163_c0_g3_i1	13.26	1	4	4	7				13.26	4	7	347	38.2	5.74
TRINITY_DN9269_c0_g1_i1	8.29	1	2	4	7				8.29	4	7	543	59.7	6.80
TRINITY_DN4256_c0_g1_i1	16.11	1	3	4	7				16.11	4	7	329	36.0	5.12
TRINITY_DN1462_c0_g1_i1	12.36	1	1	4	7				12.36	4	7	259	28.4	10.83
TRINITY_DN5792_c0_g2_i1	8.42	1	4	5	7				8.42	5	7	404	45.9	4.83
TRINITY_DN35361_c0_g1_i1	13.98	1	3	4	7				13.98	4	7	322	35.0	8.94
TRINITY_DN7927_c0_g1_i1	12.39	1	2	4	6				12.39	4	6	355	39.0	9.07
TRINITY_DN8911_c0_g3_i1	18.53	1	1	5	6				18.53	5	6	259	28.4	8.98
TRINITY_DN2725_c0_g1_i1	3.01	1	1	1	6				3.01	1	6	365	39.0	8.28

TRINITY_DN5521_c0_g2_i1	35.51	1	4	4	6				35.51	4	6	107	12.6	9.41
TRINITY_DN8727_c0_g1_i1	25.77	2	2	3	6				25.77	3	6	97	10.6	8.76
TRINITY_DN20542_c0_g1_i1	24.49	1	3	5	6				24.49	5	6	196	22.4	11.80
TRINITY_DN416_c0_g1_i1	26.62	1	3	4	6				26.62	4	6	154	17.5	9.99
TRINITY_DN30798_c0_g1_i1	10.43	1	2	2	6				10.43	2	6	115	12.1	9.29
TRINITY_DN31349_c0_g1_i1	36.63	1	2	5	6				36.63	5	6	101	11.4	10.52
TRINITY_DN3632_c0_g1_i1	8.57	1	2	2	5				8.57	2	5	105	12.2	9.36
TRINITY_DN8853_c0_g1_i1	6.91	1	1	2	5				6.91	2	5	246	28.2	10.96
TRINITY_DN9854_c0_g1_i1	17.75	1	2	4	5				17.75	4	5	169	19.0	9.88
TRINITY_DN27158_c0_g1_i1	7.09	1	1	1	5				7.09	1	5	141	15.5	10.21
TRINITY_DN36507_c0_g1_i1	14.23	1	1	3	5				14.23	3	5	239	26.3	10.05
TRINITY_DN8776_c0_g1_i1	7.40	1	1	3	5				7.40	3	5	622	64.6	8.68
TRINITY_DN20458_c0_g1_i1	12.43	1	2	2	5				12.43	2	5	169	19.6	4.75
TRINITY_DN7464_c0_g1_i1	17.51	1	1	4	5				17.51	4	5	257	28.4	9.32
TRINITY_DN9366_c0_g1_i1	5.93	1	2	2	5				5.93	2	5	405	43.6	5.07
TRINITY_DN8315_c0_g1_i1	5.47	1	3	3	5				5.47	3	5	402	45.5	11.43
TRINITY_DN4557_c0_g1_i1	4.84	1	1	3	5				4.84	3	5	516	55.9	6.76
TRINITY_DN30198_c0_g1_i1	16.36	1	2	2	5				16.36	2	5	165	18.0	6.10
TRINITY_DN7785_c0_g1_i1	4.75	1	1	2	4				4.75	2	4	400	44.3	9.99
TRINITY_DN9190_c0_g1_i1	10.26	1	1	2	4				10.26	2	4	273	29.8	10.35
TRINITY_DN11013_c0_g1_i1	20.20	2	1	3	4				20.20	3	4	99	11.0	6.04
TRINITY_DN12181_c0_g1_i1	15.85	1	1	2	4				15.85	2	4	82	9.9	10.65
TRINITY_DN20809_c0_g1_i1	7.44	1	2	4	4				7.44	4	4	511	58.6	7.56
TRINITY_DN15724_c0_g1_i1	12.87	1	1	2	4				12.87	2	4	202	23.5	9.22
TRINITY_DN4224_c0_g1_i1	7.08	1	1	2	4				7.08	2	4	325	37.0	8.22
TRINITY_DN2302_c0_g1_i1	10.78	1	2	2	4				10.78	2	4	167	17.9	8.18
TRINITY_DN3283_c0_g1_i1	10.14	1	2	2	4				10.14	2	4	217	23.5	9.77
TRINITY_DN7433_c0_g2_i1	9.66	1	2	4	4				9.66	4	4	476	51.3	4.53
TRINITY_DN16131_c0_g1_i1	3.82	1	1	2	4				3.82	2	4	314	33.4	12.54
TRINITY_DN26411_c0_g1_i1	16.67	1	2	2	4				16.67	2	4	102	11.3	7.39
TRINITY_DN28596_c0_g1_i1	10.34	1	1	1	4				10.34	1	4	116	13.4	11.19
TRINITY_DN32170_c1_g1_i1	4.57	1	1	1	4				4.57	1	4	219	25.3	6.05
TRINITY_DN39563_c0_g1_i1	14.63	1	1	2	4				14.63	2	4	82	9.0	9.86
TRINITY_DN15129_c0_g1_i1	4.32	1	1	2	3	4.32	2	2	2.07	1	1	532	53.4	9.58
TRINITY_DN1319_c0_g1_i1	7.94	1	1	1	3				7.94	1	3	126	14.9	11.65
TRINITY_DN3026_c0_g1_i1	10.14	1	1	2	3				10.14	2	3	276	30.0	10.11
TRINITY_DN7946_c0_g1_i1	1.75	1	2	3	3				1.75	3	3	1944	223.2	5.68
TRINITY_DN8279_c0_g1_i1	5.66	1	1	2	3				5.66	2	3	371	39.7	5.36
TRINITY_DN21460_c0_g1_i1	7.17	1	2	3	3				7.17	3	3	446	47.8	9.14
TRINITY_DN23896_c0_g1_i1	14.52	1	1	3	3				14.52	3	3	186	21.1	6.58
TRINITY_DN25573_c0_g1_i1	4.07	1	2	2	3				4.07	2	3	467	50.9	9.04
TRINITY_DN20813_c0_g1_i1	4.29	1	2	2	3				4.29	2	3	303	32.0	7.90
TRINITY_DN11341_c0_g1_i1	9.68	1	1	3	3				9.68	3	3	217	25.1	11.00
TRINITY_DN209_c0_g2_i1	5.47	1	1	1	3				5.47	1	3	201	22.4	6.54
TRINITY_DN15328_c0_g1_i1	9.92	1	1	2	3				9.92	2	3	242	27.5	5.26
TRINITY_DN25335_c0_g1_i1	7.27	1	3	3	3				7.27	3	3	275	31.2	9.23
TRINITY_DN359_c0_g1_i1	12.33	1	2	3	3				12.33	3	3	219	24.9	9.25
TRINITY_DN1791_c0_g1_i1	4.90	1	1	1	3				4.90	1	3	408	46.2	4.72
TRINITY_DN16477_c0_g1_i1	5.25	1	1	2	3				5.25	2	3	400	45.4	5.41
TRINITY_DN3407_c0_g1_i1	5.99	1	1	1	3				5.99	1	3	167	19.4	10.62
TRINITY_DN35716_c0_g1_i1	10.69	1	1	2	3				10.69	2	3	131	14.6	7.27
TRINITY_DN595_c0_g1_i1	11.48	1	1	1	2				11.48	1	2	183	21.0	5.39
TRINITY_DN1116_c0_g1_i1	25.66	1	1	1	2				25.66	1	2	113	12.2	9.45
TRINITY_DN2223_c0_g2_i1	3.16	1	1	1	2				3.16	1	2	253	27.3	7.47
TRINITY_DN3351_c0_g1_i1	4.76	1	1	2	2				4.76	2	2	210	23.6	9.66
TRINITY_DN3875_c0_g2_i1	1.90	1	1	1	2				1.90	1	2	633	69.9	6.57
TRINITY_DN8949_c0_g1_i1	12.68	1	2	2	2				12.68	2	2	276	30.7	6.96
TRINITY_DN8901_c0_g1_i1	7.78	1	1	1	2				7.78	1	2	167	18.4	8.56
TRINITY_DN9423_c0_g3_i1	10.14	1	1	1	2				10.14	1	2	69	7.9	4.72
TRINITY_DN9885_c0_g1_i1	67.86	1	1	1	2				67.86	1	2	28	3.1	10.74
TRINITY_DN27372_c0_g1_i1	42.86	1	1	2	2				42.86	2	2	28	2.9	11.84
TRINITY_DN29591_c0_g1_i1	16.00	1	1	1	2				16.00	1	2	50	5.6	9.29
TRINITY_DN33637_c0_g1_i1	13.13	1	2	2	2				13.13	2	2	99	10.2	10.15
TRINITY_DN35640_c0_g1_i1	7.09	1	2	2	2				7.09	2	2	127	13.8	4.41
TRINITY_DN2791_c0_g1_i1	7.69	1	1	1	2				7.69	1	2	208	23.1	8.73
TRINITY_DN4187_c0_g1_i1	6.90	1	1	2	2				6.90	2	2	319	36.0	5.87
TRINITY_DN8766_c0_g1_i1	4.18	2	1	2	2				4.18	2	2	670	74.1	5.14
TRINITY_DN7863_c0_g1_i1	6.11	1	2	2	2				6.11	2	2	311	34.0	6.80

TRINITY_DN2935_c0_g1_i1	16.83	1	1	2	2				16.83	2	2	101	11.8	5.48
TRINITY_DN1085_c0_g1_i1	9.71	1	1	1	2				9.71	1	2	103	9.4	6.34
TRINITY_DN6669_c0_g1_i1	2.58	1	1	1	2				2.58	1	2	387	43.8	8.21
TRINITY_DN15380_c0_g1_i1	18.52	1	1	1	2				18.52	1	2	108	11.8	11.91
TRINITY_DN384_c0_g1_i1	9.70	1	1	2	2				9.70	2	2	165	17.0	10.10
TRINITY_DN6441_c0_g2_i1	7.43	1	2	2	2				7.43	2	2	296	31.6	9.94
TRINITY_DN9324_c0_g2_i1	1.74	1	1	2	2				1.74	2	2	518	56.9	6.06
TRINITY_DN8350_c0_g1_i2	3.16	2	1	1	2				3.16	1	2	507	57.3	8.57
TRINITY_DN8351_c1_g1_i1	11.57	1	2	2	2				11.57	2	2	216	23.9	5.41
TRINITY_DN8396_c1_g1_i1	5.83	1	1	2	2				5.83	2	2	309	35.2	6.54
TRINITY_DN8311_c0_g1_i1	5.10	1	1	2	2				5.10	2	2	510	57.2	4.68
TRINITY_DN6545_c0_g2_i1	3.75	1	1	1	2				3.75	1	2	267	31.2	8.48
TRINITY_DN6590_c0_g1_i1	1.55	1	1	1	2				1.55	1	2	710	79.7	7.85
TRINITY_DN21232_c0_g1_i1	4.49	1	1	2	2				4.49	2	2	423	48.3	8.06
TRINITY_DN8029_c0_g1_i1	3.30	1	1	1	2				3.30	1	2	212	23.3	9.67
TRINITY_DN25901_c0_g1_i1	2.24	1	1	1	2				2.24	1	2	490	53.9	8.85
TRINITY_DN30398_c0_g1_i1	22.62	1	1	2	2				22.62	2	2	84	9.5	8.50
TRINITY_DN30510_c0_g1_i1	14.59	1	2	2	2				14.59	2	2	185	21.9	10.81
TRINITY_DN31110_c0_g1_i1	6.40	1	1	1	2				6.40	1	2	250	28.0	5.53
TRINITY_DN27445_c0_g1_i1	52.38	1	1	1	1	52.38	1	1				21	2.3	12.00
TRINITY_DN7801_c0_g1_i1	2.77	1	1	1	1	2.77	1	1				289	30.1	4.39
TRINITY_DN15309_c0_g1_i1	43.33	1	1	1	1	43.33	1	1				30	3.6	10.29
TRINITY_DN7506_c0_g1_i1	4.85	1	1	1	1	4.85	1	1				227	24.2	5.31
TRINITY_DN14035_c0_g1_i1	25.00	1	1	1	1	25.00	1	1				36	3.8	9.14
TRINITY_DN2605_c0_g2_i1	13.64	1	1	1	1	13.64	1	1				88	10.3	11.72
TRINITY_DN1589_c0_g1_i1	5.59	1	1	1	1				5.59	1	1	143	16.7	10.81
TRINITY_DN3305_c0_g1_i1	7.43	1	1	1	1				7.43	1	1	148	16.5	10.32
TRINITY_DN3807_c1_g1_i1	6.54	2	1	1	1				6.54	1	1	214	24.7	7.90
TRINITY_DN4754_c0_g1_i1	9.85	1	1	1	1				9.85	1	1	132	14.9	10.96
TRINITY_DN6950_c0_g4_i1	14.06	1	1	1	1				14.06	1	1	64	6.8	10.21
TRINITY_DN6973_c0_g2_i1	11.49	1	1	1	1				11.49	1	1	87	10.0	9.31
TRINITY_DN6990_c0_g1_i1	66.67	1	1	1	1				66.67	1	1	36	3.8	8.73
TRINITY_DN7925_c0_g1_i1	8.40	1	1	1	1				8.40	1	1	131	15.0	4.93
TRINITY_DN8690_c0_g1_i1	6.37	1	1	1	1				6.37	1	1	157	18.0	10.78
TRINITY_DN8808_c0_g1_i1	1.95	1	1	1	1				1.95	1	1	512	60.7	5.44
TRINITY_DN8840_c0_g1_i1	4.95	1	1	1	1				4.95	1	1	222	25.2	9.16
TRINITY_DN8812_c0_g1_i1	3.63	1	1	1	1				3.63	1	1	413	49.0	8.79
TRINITY_DN8950_c0_g1_i1	1.11	1	1	1	1				1.11	1	1	719	82.6	5.16
TRINITY_DN9524_c1_g1_i1	8.61	1	1	1	1				8.61	1	1	151	17.1	11.68
TRINITY_DN9518_c0_g1_i1	1.68	1	1	1	1				1.68	1	1	536	62.0	9.14
TRINITY_DN9562_c0_g1_i1	1.95	1	1	1	1				1.95	1	1	564	61.9	6.32
TRINITY_DN11530_c0_g1_i1	6.71	1	1	1	1				6.71	1	1	164	19.2	5.19
TRINITY_DN11605_c0_g1_i1	16.39	1	1	1	1				16.39	1	1	122	13.9	5.22
TRINITY_DN11641_c0_g1_i1	34.38	1	1	1	1				34.38	1	1	32	3.6	9.58
TRINITY_DN14718_c0_g1_i1	12.22	1	1	1	1				12.22	1	1	90	10.1	6.60
TRINITY_DN17104_c0_g1_i1	7.38	1	1	1	1				7.38	1	1	271	29.2	9.96
TRINITY_DN17296_c0_g1_i1	17.36	1	1	1	1				17.36	1	1	121	14.4	5.48
TRINITY_DN21129_c0_g1_i1	24.19	1	1	1	1				24.19	1	1	62	6.9	11.08
TRINITY_DN22716_c0_g1_i1	28.21	1	1	1	1				28.21	1	1	39	3.9	11.46
TRINITY_DN22884_c0_g2_i1	7.75	1	1	1	1				7.75	1	1	142	16.4	10.78
TRINITY_DN25478_c0_g1_i1	8.39	1	1	1	1				8.39	1	1	155	17.8	10.07
TRINITY_DN25411_c0_g1_i1	6.03	1	1	1	1				6.03	1	1	232	26.1	7.14
TRINITY_DN31993_c0_g1_i1	3.38	1	1	1	1				3.38	1	1	296	33.0	7.65
TRINITY_DN32421_c0_g1_i1	18.09	1	1	1	1				18.09	1	1	94	9.5	7.34
TRINITY_DN35561_c0_g2_i1	59.38	1	1	1	1				59.38	1	1	32	3.7	7.17
TRINITY_DN8118_c0_g1_i1	1.81	1	1	1	1				1.81	1	1	719	76.1	6.58
TRINITY_DN8131_c0_g2_i1	4.29	2	1	1	1				4.29	1	1	280	31.4	7.12
TRINITY_DN8151_c0_g1_i1	4.40	1	1	1	1				4.40	1	1	250	28.6	6.05
TRINITY_DN11457_c0_g1_i1	35.56	1	1	1	1				35.56	1	1	45	5.0	6.51
TRINITY_DN9201_c0_g1_i1	4.37	1	1	1	1				4.37	1	1	229	25.1	8.25
TRINITY_DN3163_c0_g1_i1	48.39	1	1	1	1				48.39	1	1	31	3.6	10.56
TRINITY_DN8727_c0_g2_i1	3.37	1	1	1	1				3.37	1	1	504	55.8	5.12
TRINITY_DN20090_c0_g1_i1	31.58	1	1	1	1				31.58	1	1	38	4.5	12.31
TRINITY_DN20404_c0_g1_i1	6.57	1	1	1	1				6.57	1	1	137	16.0	9.86
TRINITY_DN6101_c0_g1_i1	3.01	1	1	1	1				3.01	1	1	399	44.5	5.35
TRINITY_DN6104_c0_g1_i1	5.92	1	1	1	1				5.92	1	1	169	18.8	10.20
TRINITY_DN19931_c0_g2_i1	4.61	1	1	1	1				4.61	1	1	282	32.1	6.14
TRINITY_DN7505_c0_g2_i1	6.39	1	1	1	1				6.39	1	1	219	24.0	7.46

TRINITY_DN25363_c0_g1_i1	7.33	1	1	1	1				7.33	1	1	191	21.5	5.01
TRINITY_DN6452_c0_g3_i1	2.54	1	1	1	1				2.54	1	1	433	49.3	6.11
TRINITY_DN15629_c0_g1_i1	4.73	1	1	1	1				4.73	1	1	169	19.7	10.87
TRINITY_DN7606_c0_g1_i1	3.72	1	1	1	1				3.72	1	1	215	24.4	5.33
TRINITY_DN7494_c0_g3_i1	11.22	1	1	1	1				11.22	1	1	98	10.9	6.06
TRINITY_DN9348_c0_g4_i1	2.22	1	1	1	1				2.22	1	1	495	57.0	6.35
TRINITY_DN9356_c0_g1_i1	5.42	1	1	1	1				5.42	1	1	166	18.7	7.81
TRINITY_DN6289_c0_g1_i1	9.62	1	1	1	1				9.62	1	1	104	11.6	9.89
TRINITY_DN475_c0_g1_i1	7.25	1	1	1	1				7.25	1	1	138	15.6	10.08
TRINITY_DN21887_c0_g1_i1	6.76	1	1	1	1				6.76	1	1	148	17.0	9.99
TRINITY_DN6731_c0_g1_i1	2.29	1	1	1	1				2.29	1	1	480	54.0	8.00
TRINITY_DN6738_c0_g1_i1	1.42	1	1	1	1				1.42	1	1	563	61.8	5.27
TRINITY_DN8558_c0_g1_i1	18.18	1	1	1	1				18.18	1	1	55	5.9	9.99
TRINITY_DN22005_c0_g1_i1	4.02	1	1	1	1				4.02	1	1	199	22.3	6.81
TRINITY_DN5838_c0_g1_i1	1.98	1	1	1	1				1.98	1	1	405	42.9	7.87
TRINITY_DN23380_c0_g1_i1	3.52	1	1	1	1				3.52	1	1	256	28.3	8.15
TRINITY_DN26305_c0_g1_i1	6.19	1	1	1	1				6.19	1	1	226	26.0	8.13
TRINITY_DN26554_c0_g1_i1	7.04	1	1	1	1				7.04	1	1	142	15.8	7.81
TRINITY_DN26903_c0_g1_i1	3.98	1	1	1	1				3.98	1	1	251	28.5	5.34
TRINITY_DN27081_c0_g1_i1	14.12	1	1	1	1				14.12	1	1	85	9.2	9.35
TRINITY_DN27093_c0_g1_i1	19.15	1	1	1	1				19.15	1	1	94	10.1	11.46
TRINITY_DN27610_c0_g1_i1	10.84	1	1	1	1				10.84	1	1	83	9.3	10.51
TRINITY_DN28195_c0_g1_i1	12.34	1	1	1	1				12.34	1	1	154	17.3	9.83
TRINITY_DN30509_c0_g1_i1	4.76	1	1	1	1				4.76	1	1	189	22.4	11.50
TRINITY_DN31727_c0_g1_i1	47.22	1	1	1	1				47.22	1	1	36	3.6	11.09
TRINITY_DN33526_c0_g1_i1	9.47	1	1	1	1				9.47	1	1	190	21.6	6.38
TRINITY_DN35087_c0_g1_i1	7.31	1	1	1	1				7.31	1	1	219	24.7	5.06
TRINITY_DN35187_c0_g1_i1	13.68	1	1	1	1				13.68	1	1	117	12.5	5.39
TRINITY_DN36079_c0_g1_i1	4.02	1	1	1	1				4.02	1	1	199	22.3	4.75