



# Diversification of von Willebrand Factor A and Chitin-Binding Domains in Pif/BMSPs Among Mollusks

Keisuke Shimizu<sup>1,2</sup> · Lumi Negishi<sup>3</sup> · Hitoshi Kurumizaka<sup>3</sup> · Michio Suzuki<sup>2</sup>

Received: 27 October 2023 / Accepted: 23 May 2024 / Published online: 12 June 2024  
© The Author(s) 2024

## Abstract

Pif is a shell matrix protein (SMP) identified in the nacreous layer of *Pinctada fucata* (Pfu) comprised two proteins, Pif97 and Pif 80. Pif97 contains a von Willebrand factor A (VWA) and chitin-binding domains, whereas Pif80 can bind calcium carbonate crystals. The VWA domain is conserved in the SMPs of various mollusk species; however, their phylogenetic relationship remains obscure. Furthermore, although the VWA domain participates in protein–protein interactions, its role in shell formation has not been established. Accordingly, in the current study, we investigate the phylogenetic relationship between PfuPif and other VWA domain-containing proteins in major mollusk species. The shell-related proteins containing VWA domains formed a large clade (the Pif/BMSP family) and were classified into eight subfamilies with unique sequential features, expression patterns, and taxa diversity. Furthermore, a pull-down assay using recombinant proteins containing the VWA domain of PfuPif 97 revealed that the VWA domain interacts with five nacreous layer-related SMPs of *P. fucata*, including Pif 80 and nacrein. Collectively, these results suggest that the VWA domain is important in the formation of organic complexes and participates in shell mineralisation.

**Keywords** Evolution · Mollusca · Mineralization · von Willebrand factor A

## Introduction

The biominerals produced by different organisms (e.g., sponges, corals, mollusks, brachiopods, and sea urchin) are diverse in their associated minerals, microstructures, and external morphology. They comprise inorganic minerals and specific organic molecules (reviewed in Liu and Zhang 2021). The diversification and functions of these organic molecules have been investigated in myriad applied research

fields, including medicine, dentistry, environmental science, and materials science.

Mollusk shells comprise calcium carbonate crystals (calcite and aragonite) and organic molecules, including polysaccharides, proteins, and lipids, that are secreted from the epithelial cells of the mantle and are incorporated into, or form scaffolds around, crystals (Belcher et al. 1996). Certain organic compounds also contribute to the formation of various shell microstructures, crystal nucleation, and crystal growth orientation (Belcher et al. 1996; Levi-Kalishman et al. 2001). The nacreous layer is lustrous and exists within the inner portion of mollusk shells, such as pearl oysters (*Bivalvia*), abalones (*Gastropoda*), and nautilus (*Cephalopoda*). This characteristic laminated composite structure is formed by aragonite crystals filling compartments separated by a chitin and insoluble matrix protein framework (Gregoire 1957; Wada 1961). The organic framework may serve as a scaffold for aragonite crystal growth, with various organic molecules partially guiding the crystal growth pattern, forming a nacreous layer (Weiner and Hood 1975; Falini et al. 1996; Belcher et al. 1996).

Numerous types of matrix proteins have been identified within the nacreous layer (e.g., nacrein [Miyamoto

---

Handling editor: **John Bracht**.

✉ Michio Suzuki  
amichiwo@mail.ecc.u-tokyo.ac.jp

<sup>1</sup> Research Institute for Global Change, Japan Agency for Marine–Earth Science and Technology, 2-15 Natsushima-Cho, Yokosuka, Kanagawa 237-0061, Japan

<sup>2</sup> Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

<sup>3</sup> Institute for Quantitative Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

et al. 1996], N16 [Samata et al. 1999], MSI60 [Sudo et al. 1997], and perlucin [Weiss et al. 2000]) of *Pinctada fucata*. In particular, Pif is a key protein in nacreous layer formation in the pearl oyster *P. fucata* (Suzuki et al. 2009). Pif in *P. fucata* (PfuPif) is cleaved into PfuPif 97 and PfuPif 80 at the dibasic cleavage site (RMKR); each protein with unique sequential features likely cooperates in nacreous layer formation (Suzuki et al. 2009). The cleaved proteins exhibit different features. PfuPif 97 is located within the N-terminal region of PfuPif and contains a von Willebrand factor A (VWA) domain, a chitin-binding (CB) domain, and a CB-like domain (Suzuki et al. 2009, 2013). The VWA domain is known to participate in protein–protein interactions (Tuckwell 1999; Whittaker and Hynes 2002). VWA domain-containing proteins (VWA dcps) have been reported in various metazoan skeletons, including corals (Ramos-Silva et al. 2013; Takeuchi et al. 2016b), mollusks (Zhang et al. 2012; Marie et al. 2012, 2013, 2017; Liu et al. 2015; Liao et al. 2015; Gao et al. 2015; Zhao et al. 2018; Mann et al. 2018; Shimizu et al. 2019, 2022b; Oudot et al. 2020; Takeuchi et al. 2021; Setiamarga et al. 2021; Zhang et al. 2021), and brachiopods (Luo et al. 2015). CB domain-containing proteins (CB dcps) participate in protein–chitin or other polysaccharide interactions and have been identified in shell matrix proteins (SMPs) in mollusks (Zhang et al. 2012; Marie et al. 2012, 2013, 2017; Liu et al. 2015; Liao et al. 2015; Gao et al. 2015; Zhao et al. 2018; Mann et al. 2018; Shimizu et al. 2019, 2022b; Oudot et al. 2020; Takeuchi et al. 2021; Setiamarga et al. 2021; Zhang et al. 2021). Thus, the VWA and CB domains of PfuPif 97 are thought to participate in organic framework formation in the nacreous layer (Suzuki et al. 2009).

PfuPif is cleaved into two proteins, PfuPif 97 and PfuPif 80, at the dibasic cleavage site (RMKR), and each protein with different sequential features is thought to cooperate in nacreous layer formation (Suzuki et al. 2009). The cleaved proteins exhibit different features. PfuPif 97 is the N-terminal part of PfuPif and has a von Willebrand factor A (VWA) domain, a CB domain, and a CB-like domain (Suzuki et al. 2009, 2013). The VWA domain is involved in protein–protein interactions and is composed of multiprotein complexes (Tuckwell 1999; Whittaker and Hynes 2002). VWA domain-containing proteins (VWA dcps) have been reported in various metazoan skeletons (corals: Ramos-Silva et al. 2013; Takeuchi et al. 2016b, mollusks: Zhang et al. 2012; Marie et al. 2012, 2013, 2017; Liu et al. 2015; Liao et al. 2015; Gao et al. 2015; Zhao et al. 2018; Mann et al. 2018; Shimizu et al. 2019, 2022b; Oudot et al. 2020; Takeuchi et al. 2021; Setiamarga et al. 2021; Zhang et al. 2021, brachiopod: Luo et al. 2015). CB domain-containing proteins (CB dcps) are involved in protein–chitin or other polysaccharide interactions and have also been reported as shell matrix proteins in mollusks (Zhang et al. 2012; Marie et al. 2012, 2013, 2017;

Liu et al. 2015; Liao et al. 2015; Gao et al. 2015; Zhao et al. 2018; Mann et al. 2018; Shimizu et al. 2019, 2022b; Oudot et al. 2020; Takeuchi et al. 2021; Setiamarga et al. 2021; Zhang et al. 2021). Thus, PfuPif 97 has both VWA and CB domains, which are thought to participate in organic framework formation in the nacreous layer (Suzuki et al. 2009).

PfuPif 80 is the C-terminal region of the PfuPif protein, containing partial laminin G (LG) domains and low-complexity regions with numerous acidic and basic amino acid residues (Asp, Lys, and Arg) (Suzuki et al. 2009, 2013). PfuPif 80 has aragonite crystal-binding ability and contributes to aragonite formation (Suzuki et al. 2009). Similarly, BMSP (Blue Mussel Shell Protein) has common domain components (VWA, CB, and LG domains); Pif has been identified as a calcium carbonate-binding protein in the nacreous layer of the blue mussel *Mytilus galloprovincialis* (Suzuki et al. 2011). MgaBMSP is also cleaved into two proteins: MgaBMSP 120 and MgaBMSP 100. Similar to Pif 80, MgaBMSP 100 can bind to calcium carbonate crystals (Suzuki et al. 2009, 2011). However, MgaBMSP differs from PfuPif in that it contains four VWA domains in tandem (Suzuki et al. 2011). Although PfuPif and MgaBMSP were initially identified as nacreous layer-specific matrix proteins, recent omics (genomics, transcriptomics, and proteomics) studies have revealed numerous types of Pif- and BMSP-like genes or proteins in various mollusks with and without nacreous layers (Marie et al. 2012, 2017; Feng et al. 2017; Gao et al. 2015; Liao et al. 2015; Zhang et al. 2018; Takeuchi et al. 2021; Sun et al. 2020; Varney et al. 2021; Shimizu et al. 2022b; Setiamarga et al. 2021; Zhang et al. 2021). For instance, BMSP-like proteins have four VWA domains, while CB domains have been identified in the shells of bivalves *Crassostrea gigas* (Zhao et al. 2018), *Atrina pectinata* (Shimizu et al. 2022b), and *Tridacna crocea* (Takeuchi et al. 2021). Similar to MgaBMSP 100, the BMSP-like protein in *A. pectinata* exhibits a calcium carbonate-binding ability (Shimizu et al. 2022b). Furthermore, the sequence of a BMSP-like protein, containing three VWA domains, a CB domain, and a CB-like domain, was detected in the genome of the gastropod *Lotia gigantea* (Suzuki et al. 2013) and later identified as an SMP (Marie et al. 2013).

VWA and CB domain-containing proteins (VWA–CB dcps) are diverse in lophotrochozoans, including mollusks, brachiopods, nemertean, and phoronids (Luo et al. 2015, 2017; Zhao et al. 2018). Proteomic analyses have identified six types of VWA–CB dcps in *P. fucata*, including PfuPif, within the adult shell layers (nacreous or prismatic layers) or the D-shaped larval shell (Zhao et al. 2018). One of these in *M. galloprovincialis* (MgaBMSP) contains four VWA domains (Suzuki et al. 2011). Previous studies have suggested that various VWA–CB dcps participate in shell formation; however, each protein has a unique role in different microstructures and developmental stages. Furthermore,

while many types of VWA or CB dcps have been defined as SMPs in various mollusks and designated ‘Pif-like,’ their evolutionary relationships remain unclear.

To understand the evolution of the Pif/BMSP family, in the current study, we conducted phylogenetic and functional analyses of the VWA domain. Phylogenetic analysis of VWA dcps identified a large clade designated Pif/BMSP family (Pif/BMSPf) comprising many shell-related VWA–CB dcps, including PfuPif and MgaBMSP. Pif/BMSPf was classified into eight subfamilies. We determined the spatial expression patterns of four Pif/BMSPf protein-encoding genes in the mantle tissue of *P. fucata*. Although the specific matrix proteins that interact with the VWA domains in Pif/BMSPf proteins have not been reported, we postulate that the VWA domains are key to the protein–protein interactions required for organic complex formation. Thus, we conducted pull-down assays using recombinant PfuPif proteins with altered VWA domains and SMPs extracted from the nectareous layer of *P. fucata*. Various SMPs that interact with the VWA domain of PfuPif were identified. Cumulatively, this study provides new insights into the function and evolution of Pif/BMSPf proteins in mollusks.

## Materials and Methods

### Identification of Pif/BMSP Family Proteins

VWA domain was identified from four molluscan genomes including two bivalves (*P. fucata* [Takeuchi et al. 2016a] and *C. gigas* [Zhang et al. 2012]), a gastropod *L. gigantea* [Simakov et al. 2013], a cephalopod *O. bimaculoides* [Albertin et al. 2015], and the transcriptome data of a schaphopod (*A. entalis*) (Bioproject Id: PRJNA506080) using HMMER v3.4 ( $e$ -values  $< 1.0e-5$ , <http://hmmmer.org>; last accessed August 30, 2023) with protein domain database Pfam (<https://pfam.xfam.org>; last accessed August 30, 2023). We then selected the VWA sequences with high homology to those of PfuPif using BLSTP ( $e$ -value  $< 1.0e-10$ ). We then conducted a molecular phylogenetic analysis using selected VWA sequences and VWA sequences from previously known SMPs or mantle-specific genes in molluscs (Suzuki et al. 2013; Liao et al. 2015; Marie et al. 2017; Zhang et al. 2018; Sun et al. 2020; Varney et al. 2021; Takeuchi et al. 2021; Shimizu et al. 2022b).

### Molecular Phylogeny

The sequences of the VWA domain regions of VWA dcps were aligned using the online version of MAFFT v7.511 (<https://mafft.cbrc.jp/alignment/server/>; last accessed August 30, 2023) (Kuraku et al. 2013; Katoh et al. 2019). We trimmed the alignments using TrimAl (1.2rev59) (gap

threshold set at 0.9) (Capella-Gutiérrez et al. 2009), and the remaining 161 residues were used for molecular phylogenetic analysis. The maximum likelihood tree was constructed with IQ-TREE v2.2.2.7 (<http://www.iqtree.org/>; last accessed August 30, 2023) (Minh et al. 2020) using the best-fit model (LG+G4) selected by ModelFinder (Kalyaana-moorthy et al. 2017) with 1,000 ultrafast bootstrap (UFboot) replicates and 1000 SH-aLRT branch tests.

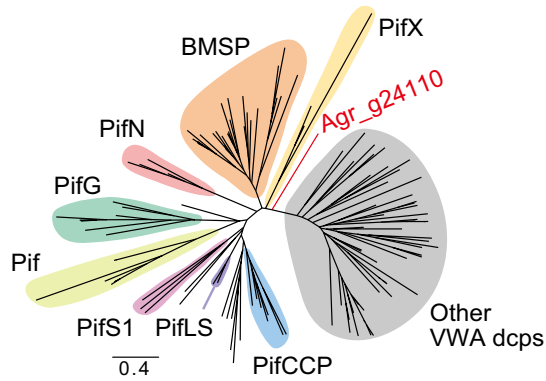
### Identification of Specific Domains and the Partial LG Domain

We identified specific domains in Pif/BMSP family proteins using HMMER v3.4 ( $e$ -values  $< 1.0e-5$ , <http://hmmmer.org>; last accessed August 30, 2023) with protein domain database Pfam (<https://pfam.xfam.org>; last accessed August 30, 2023). Signal peptide prediction and low complexity regions (LCR) (compositionally biased region) were conducted using SignalP (Petersen et al. 2011) and SEG (Wootton and Federhen 1996) ( $e$ -value  $< 1.0e-5$ ) included in the online version of Simple Modular Architecture Research Tool (SMART, <http://smart.embl-heidelberg.de>; last accessed October 2, 2023) (Letunic et al. 2015; Letunic and Bork 2018). Some of the Pif/BMSP family proteins were known to have one or more CB domain-like sequence (Suzuki et al. 2013). Thus, we also identified the CB Peritrophin-A (CBM\_14) like domain using HMMER v3.4 (<http://hmmmer.org>; last accessed August 30, 2023) with default setting and defined two types of CBM\_14-like domains (CBM\_14L1 and CBM\_14L2) according to their independent  $e$ -value ( $1e-5 \sim 1.0$  and  $1.0 \sim 10,000$ , respectively). To find the partial sequence of the LG domain in downstream of the CB domain of Pif/BMSP family proteins, we used conserved sequence (N-terminal of LG domain: AYFNGRAGLKIP-RFSGVPYGKSVFIKMKYKED, C-terminal of LG domain: WKTVSLKISNGHIRGRRDDREDKDVLDGDLKTTFS-GFQIQGASNKNFKGYMDEVYIYF) that have already reported in the previous study (Suzuki et al. 2013) and searched by BLASTN and the online version of MAFFT v7.511 (<https://mafft.cbrc.jp/alignment/server/>; last accessed August 30, 2023) (Kuraku et al. 2013; Katoh et al. 2019). Pif/BMSP family proteins were classified based on the similarity in amino acid composition between the N-terminal and C-terminal sides of the LG domain using Ward’s hierarchical clustering method in R v4.3.0. Silhouette analysis was conducted using R v4.3.0, and the optimal number of clusters was determined.

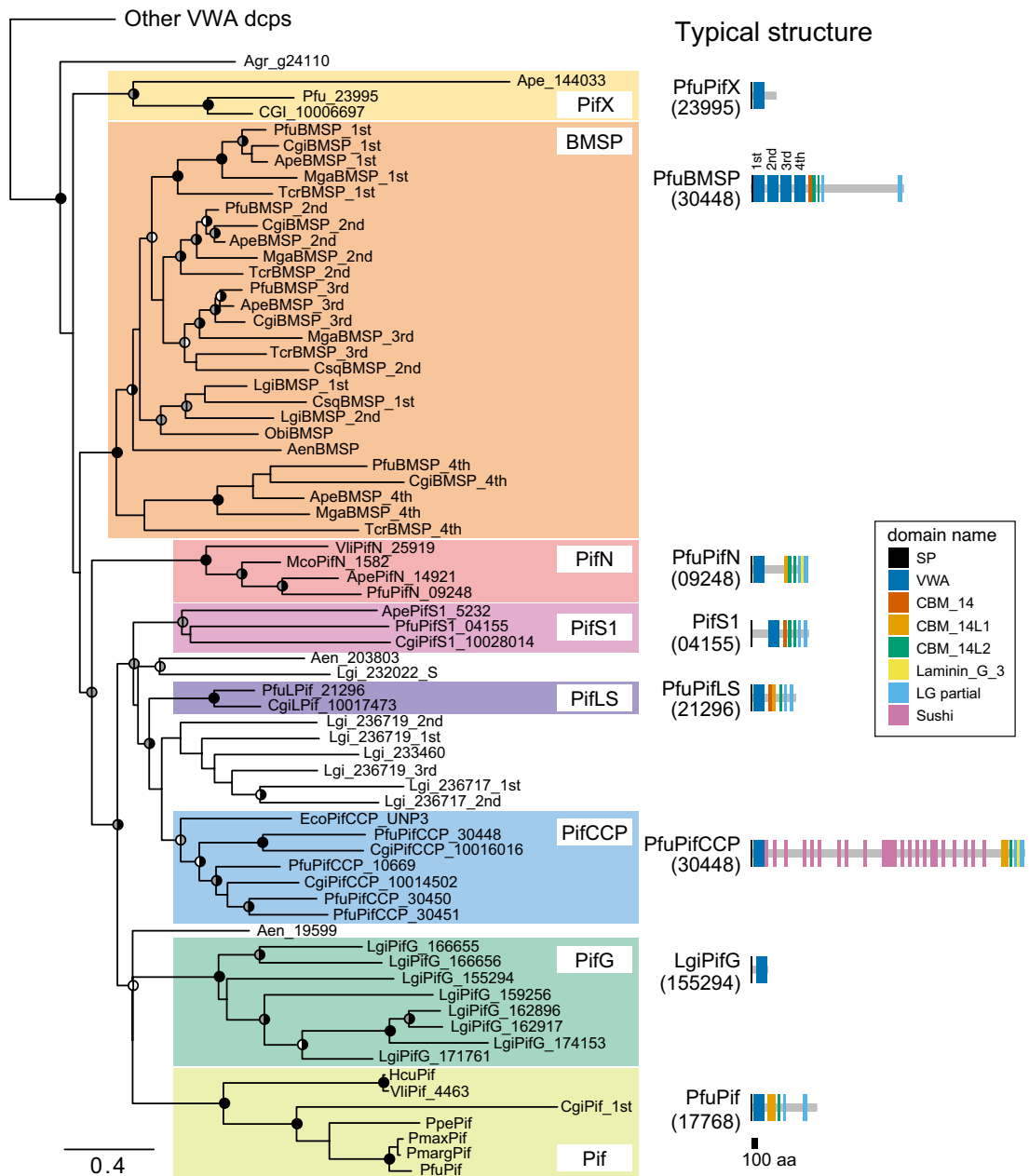
### In Situ Hybridization

The mantle tissues of *P. fucata* were dissected from fresh individuals and were gifted by the Mie Prefecture Fisheries Research Institute (Mie, Japan). RNA extraction and

A



B



**Fig. 1** Molecular phylogeny of molluscan VWA domain-containing proteins. **A, B** The maximum likelihood tree was inferred from 125 VWA domain sequences using the LG+G4 model (161 positions, 1,000 ultrafast bootstrap replicates, and SH-aLRT test). The branch lengths are proportional to the expected number of substitutions per site, as indicated by the scale bar. **B** Details of the Pif/BMSP family clade (except for the VWA dcps clade) in (A). Schematic representation of typical domain structure of several subfamilies are shown in right side of phylogenetic tree (Schematic representation of other Pif/BMSPs are shown in Fig. 2). Black, gray, and white left half circles on nodes indicate high ( $\geq 95\%$ ), medium ( $\geq 85\%$ ), and low ( $\geq 75\%$ ) ultrafast bootstrap values, respectively. Black, gray, and white right half circles on nodes also indicate high ( $\geq 85\%$ ), medium ( $\geq 65\%$ ), and low ( $\geq 45\%$ ) SH-aLRT values, respectively. See Supplementary Table S5 for species name abbreviations

cDNA synthesis were performed using Sepasol RNA I Super G (#09379-84, Nacalai Tesque Inc., Kyoto, Japan) and Prime Script RT reagent kit (#RR037A, Takara, Tokyo, Japan), respectively, according to the manufacturer's protocol. Partial Pif, BMSP, Pif in nacreous layer (PifN), and complement control protein (CCP) domain-containing Pif (PifCCP) sequences from *P. fucata* were amplified via PCR using specific primers (Supplementary Table S1). 2nd PCR was conducted using T7 or SP6 tailed primers (Supplementary Table S1) and the PCR products were used for probe synthesis. Probe synthesis was conducted using DIG RNA labeling mix (#11277073910, Roche), 10 mM dithiothreitol (DTT), RNase ribonuclease inhibitor (#SIN201, Toyobo), T7 or SP6 RNA polymerase in 1X transcription buffer (#10881767001 or #10810274001, Roche), and purified PCR products (500 ng per reaction) according to the manufacturer's protocol. RNA probes were purified using NucAway spin columns (#AM10070; Thermo Fisher Scientific). In situ hybridization was performed as previously described (Shimizu et al. 2020).

### Pull-Down Assay

The pull-down assay was performed as described previously (Shimizu et al. 2022a). The simple method is as follows: the recombinant VWA protein was prepared using the recombinant vector pET44a with the VWA sequence of PfuPif using the InFusion HD Cloning kit (#Z9648N, Takara) (Supplementary Table S1). The rVWA protein expressed in *Escherichia coli* (BL21) was purified using a Ni-column (#17531801, Ni Sepharose 6 Fast Flow, GE Healthcare, CHI, USA), and the purified proteins were confirmed by SDS-PAGE and CBB staining. Five micrograms of rVWA protein and tag-only protein (negative control) were bound to Ni-columns (#17531801, GE Healthcare), and each column was washed three times with 20 mM imidazole in wash buffer (0.5 M NaCl, 20 mM phosphate buffer pH 7.5). These columns were incubated with 50  $\mu$ g of proteins were extracted from nacreous layer of *P. fucata* using acetic acid

at 4°C for 18 h. After washing five times with 20 and 50 mM imidazole in wash buffer, the bound proteins were collected using wash buffer containing 500 mM imidazole. The binding protein solutions were concentrated in a Vivaspin 500–10 K (#VS0101, Sartorius, Göttingen, Germany) and used for peptide analysis.

### LC-MS/MS Analysis

Protein alkylation was conducted with alkylation buffer (7 M Guanidine Hydrochloride, 0.5 M Tris-HCl pH 8, 10 mM EDTA pH 8) containing 5 mM DTT for 30 min at 60 °C. After adding 2  $\mu$ L of 0.5 M Iodoacetamide and incubation for 1 h at 25 °C in dark, protein purification was performed using methanol-chloroform. The precipitation of protein was dissolved in 45  $\mu$ L of trypsin solution (11 ng/ $\mu$ L) (Trypsin Gold [V528A, Promega, WI, USA] in 50 mM  $\text{NH}_4\text{HCO}_3$  [pH 8.0]) and incubated at 37 °C for 18 h. After adding 5  $\mu$ L of 1% TFA (final concentration is 0.1%), samples were used for peptide analysis (LC-MS/MS) (Thermo Fisher, Orbitrap Fusion Tribrid Mass Spectrometer). The data from LC-MS/MS was analyzed using the soft of Proteome Discover 2.4 and the protein database from the predicted transcripts for genome assembly ver 2.0 of *P. fucata* (pfu\_aug2.0. AA.fasta) (Takeuchi et al. 2016a).

## Results

### Classification of Pif/BMSP Family Proteins in Mollusks

We identified 165, 124, 74, 111, and 32 VWA dcps in the genome and transcriptome databases of four mollusks: *P. fucata* (Bivalvia), *L. gigantea* (Gastropoda), *O. bimaculoides* (Cephalopoda), and *A. entails* (Schaphopoda), respectively (Supplementary Table S2). Within these proteins, a total of 255, 206, 94, 150, and 55 VWA domains were detected in *P. fucata*, *C. gigas*, *L. gigantea*, *O. bimaculoides*, and *A. entails*, respectively (Supplementary Table S3). To select VWA sequences for molecular phylogenetic analysis, we conducted a BLASTP homology search using the VWA domain sequence of PfuPif as a query sequence ( $e$ -value  $< 1.0e-10$ ). We obtained 20, 31, 25, 11, and 9 (total 96) VWA sequences with high homology to the PfuPif VWA domain among the 255 (*P. fucata*), 206 (*C. gigas*), 94 (*L. gigantea*), 150 (*O. bimaculoides*), and 55 (*A. entails*) VWA sequences, respectively (Supplementary Table S4). We then conducted a molecular phylogenetic analysis using the selected 96 VWA sequences from four mollusks combined with 29 VWA sequences from previously reported SMPs or mantle-specific genes in other mollusks. Hence, a total of

125 VWA sequences were included in our molecular phylogenetic analyses.

The result showed that 70 VWA sequences in 50 VWA dcps, including PfuPif, MgaBMSP, and other Pif-like proteins, formed a clade (i.e., the Pif/BMSP family) with strong nodal support (UFBoot  $\geq$  95 and SH-aLRT  $\geq$  85) (Fig. 1, electronic supplementary material, Figure S1, Table S5). The remaining 55 VWA sequences did not contain shell-related VWA dcps. Additionally, 50 Pif/BMSP family proteins were further classified into eight subfamilies with strong (UFBoot  $\geq$  95 and SH-aLRT  $\geq$  85) or medium (UFBoot  $\geq$  75) nodal support (Fig. 1). Subsequently, specific domain structures were identified for these 50 Pif/BMSP members using HMMER v3.3.2 (<http://hmmer.org/>; last accessed August 1, 2023) (Fig. 2). Most Pif/BMSPs (38/50, 76%) contained one or more CBM\_14 domains (independent  $e$ -value  $< 1e-5$ ) and CBM\_14-like domains (independent  $e$ -value  $< 1.0$  [CBM\_14L1] or 10,000 [CBM\_14L2]) in the downstream region of the VWA domain (Fig. 2, Supplementary Table S6).

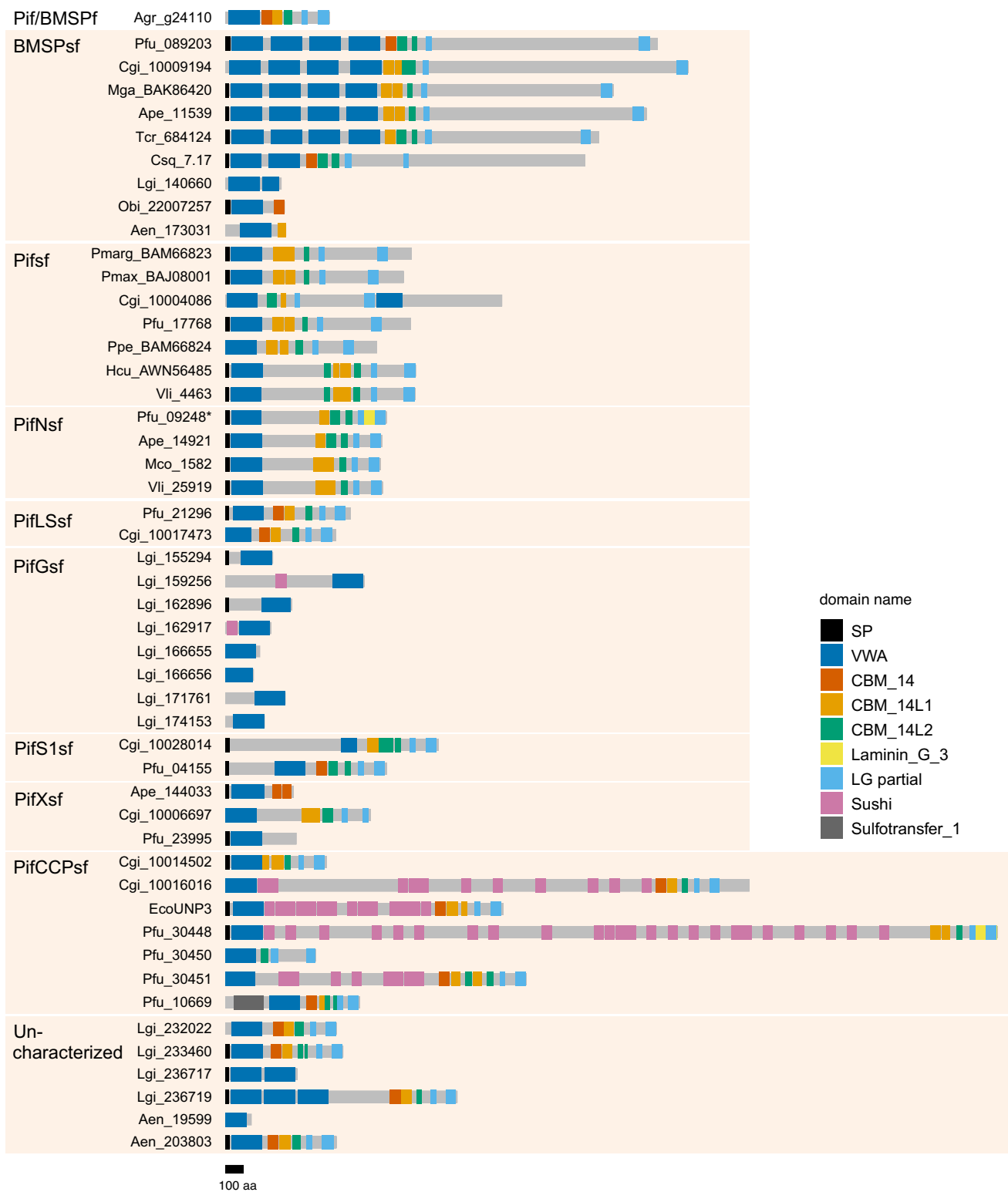
The BMSP subfamily (BMSPsf) comprises MgaBMSP and other proteins in bivalves, gastropods (*L. gigantea* and *Chrysomallon squamiferum*), cephalopods (*O. bimaculoides*), and schaphopods (*A. entalis*). All five BMSPs in bivalves (*P. fucata*, *A. pectinata*, *C. gigas*, *M. galloprovincialis*, and *Tridacna crocea*) contained four VWA domains tandemly arranged at the N-termini (Fig. 2), and these VWA domains formed a single clade (Fig. 1). The Pif subfamily (Pifsf) comprised PfuPif and other proteins identified as SMPs in the nacreous layer, except for the Pacific oyster *C. gigas* (CGI\_10004086) (Fig. 1B). Four of the bivalve VWA dcps formed the subfamily PifN (Pif-like in the nacreous layer) as they were also identified as SMPs from the nacreous layer of *Villosa lienosa* (Marie et al. 2017), *Mytilus coruscus* (Liao et al. 2015), *A. pectinata* (Shimizu et al. 2022b), and *P. fucata* (Zhao et al. 2018) like PfuPif (Suzuki et al. 2009) (Fig. 1B). Unlike other Pif/BMSPs (except for HcuPif and VliPif), PifNsf proteins have relatively long insert sequences between the VWA and CB domains (Fig. 2). Additionally, the PifLS (Pif-like in larval shell) subfamily comprises two proteins identified as larval SMPs in *P. fucata* and *C. gigas* (Zhao et al. 2018). Meanwhile, subfamily PifG (Pif-like in gastropoda), which are found only in the gastropod *L. gigantea* contained eight VWA dcps but lacked CBM\_14 and LG domains. Four of the eight PifGsf proteins had a VWA domain and CCP domains, while none were identified as SMPs of *L. gigantea* (Mann et al. 2012; Mann and Edsinger 2014). The remaining three subfamilies were only found in bivalves and had medium nodal support (UFBoot  $\geq$  75 and SH-aLRT  $\geq$  65); they were designated PifCCPsf (Pif-like with CCP domains), PifS1sf (Pif-like in shells 1), and PifXsf (Fig. 1). Members of PifCCPsf have VWA, CBM\_14, and LG domains similar to other BMSP/

Pif family proteins while also containing numerous Sushi/CCP domain repeats between the VWA and CB domains (Fig. 2). Two PifCCPsf proteins (Pfu\_30448 and EcoUNP3) were identified as SMPs from the prismatic layer of *P. fucata* (Zhao et al. 2018) and the nacreous layer of *Elliptio complanata* (Marie et al. 2017). PifS1sf comprises three proteins identified as SMPs in *P. fucata* (larval shell), *C. gigas* (adult shell), and *A. pectinata* (prismatic layer of adult shell) (Zhang et al. 2012; Zhao et al. 2018; Shimizu et al. 2022b). They had a relatively long insertion between the signal peptide and VWA domain, which differed from other Pif/BMSP family proteins (Fig. 2). Finally, PifXsf proteins did not have common features within this subfamily. Although the Ape\_144033 protein was identified as an SMP in the nacreous layer of *A. pectinata* (Shimizu et al. 2022b), other PifXsf proteins have not been previously reported as SMPs.

### Amino Acids Compositions of the LG Domain Insert Sequences and Low-Complexity Regions

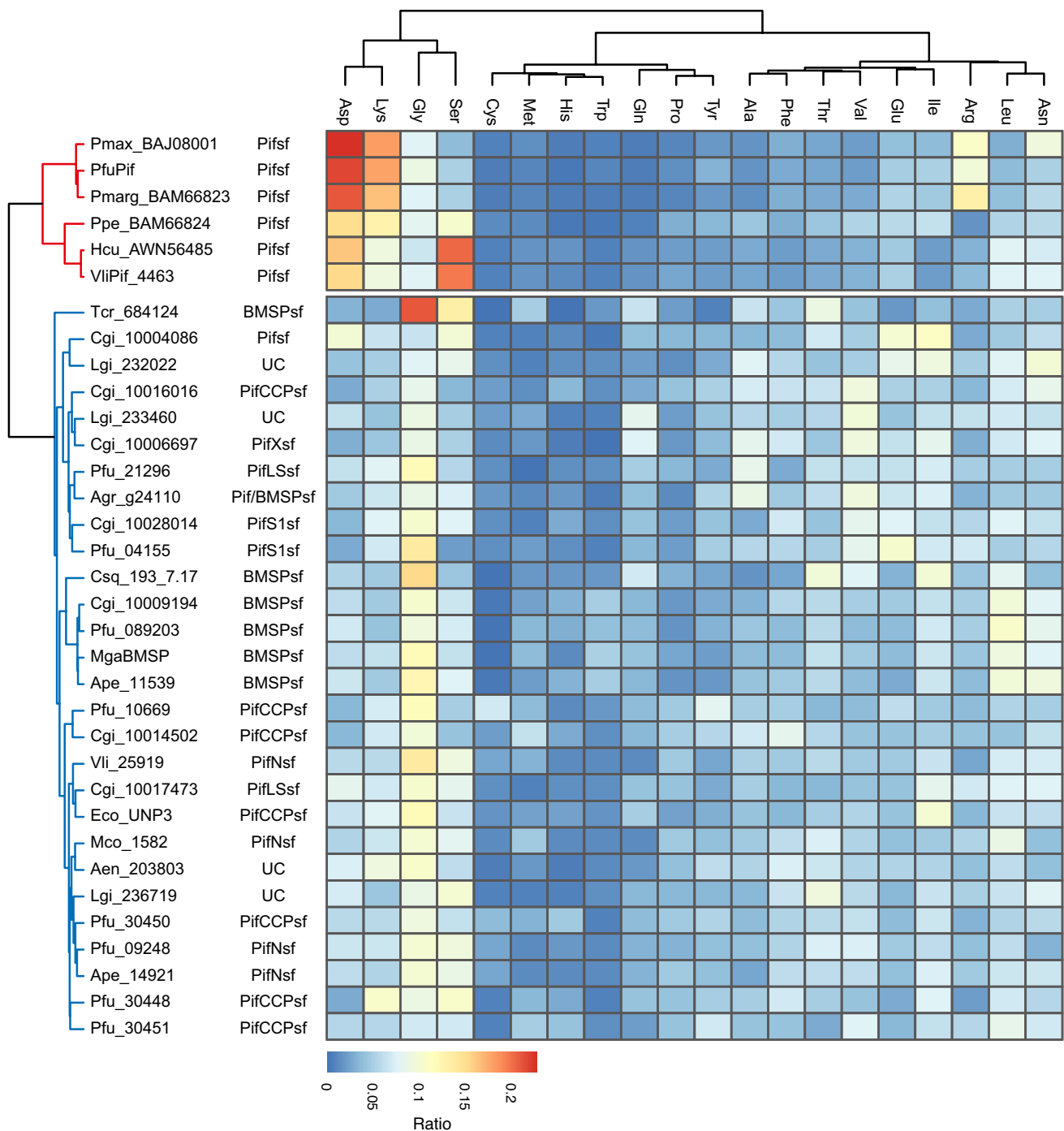
Some Pif and BMSP proteins that contain partial sequences of the N-terminal and C-terminal regions of the LG domain in the downstream region of the CB domain are well-conserved (Suzuki et al. 2013). We searched for partial LG domain sequences from the 50 Pif/BMSPs identified in this study. Partial sequences of the N- and C-terminal regions of the LG domain were identified in the downstream region of the CB domain for 34 Pif/BMSPs (Fig. 2); complete sequences of the LG domain were obtained for 2/34 Pif/BMSPs (PfuPifCCP and PfuPifN) (Fig. 2). Considering that sequences of various lengths were inserted (amino acids 138–1,438), we compared the amino acid compositions of the insert sequences for 34 Pif/BMSPs. The cluster analysis results identified two clusters: six Pifsf proteins (PfuPif, PmargPif, PmaxPif, PpePif, HcuPif, and VliPif) and another 28 Pif/BMSPs (Fig. 2, Supplementary Figure S2). Furthermore, the insert regions of the Pifsf proteins were classified into two subclusters (*Pinctada* spp. and others). The insert sequence of the LG domain Pif of *Pinctada* spp. (*P. fucata*, *P. maxima*, and *P. margaritifera*) exhibited unique features with many polar amino acids (Asp, Arg, and Lys) (Fig. 3). In contrast, the Pif of freshwater pearl oysters (*H. cumingii* and *V. lienosa*) contained Ser and Asp residues in the insert region (Fig. 3). The Pif of *P. penguin* had intermediate features (Asp-, Lys-, and Ser-rich) (Fig. 3). Only CgiPif (Cgi\_10004088) was separated from the Pif family proteins and classified with other Pif/BMSPs (Fig. 3).

Certain Pif/BMSPs contained one or more low-complexity regions (LCRs); 119 LCRs were detected in the 30 Pif/BMSPs in four distinct regions: zone 1, upstream of the VWA domain; zone 2, between tandem VWA domains; zone 3, VWA and CB domains; zone 4, N- and C-terminal regions of the LG domain (Fig. 4, Supplementary Table S7). The



**Fig. 2** Schematic representation of VWA dcps in Pif/BMSP family. The three types of CBM\_14 (CBM\_14, CBM\_14L1, and CBM\_14L2) show differences in independent  $e$ -value ( $< 1e-5$ , 1.0, and 10,000, respectively) (Supplementary Table S6). *CBM\_14* chitin-

binding peritrophin A domain, Laminin\_G\_3, *LG* laminin G, *Sushi*, sushi repeat domain, *SP* signal peptide, *VWA* von Willebrand factor type A domain, *Sulfotransfer\_1* sulfotransferase domain; see Supplementary Table S5 for species name abbreviations



**Fig. 3** Amino acid component of insert region between partial LG domains. Heatmap shows amino acid ratio in the insert region. The result of cluster analysis shows on the left. Red and blue colored lines indicate cluster 1 (Pif subfamily proteins except for CgiPif

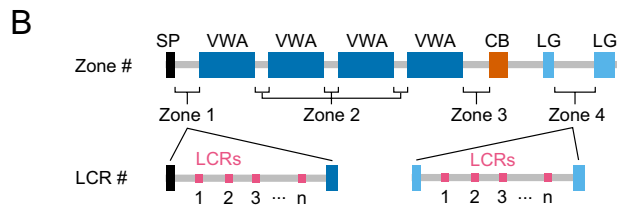
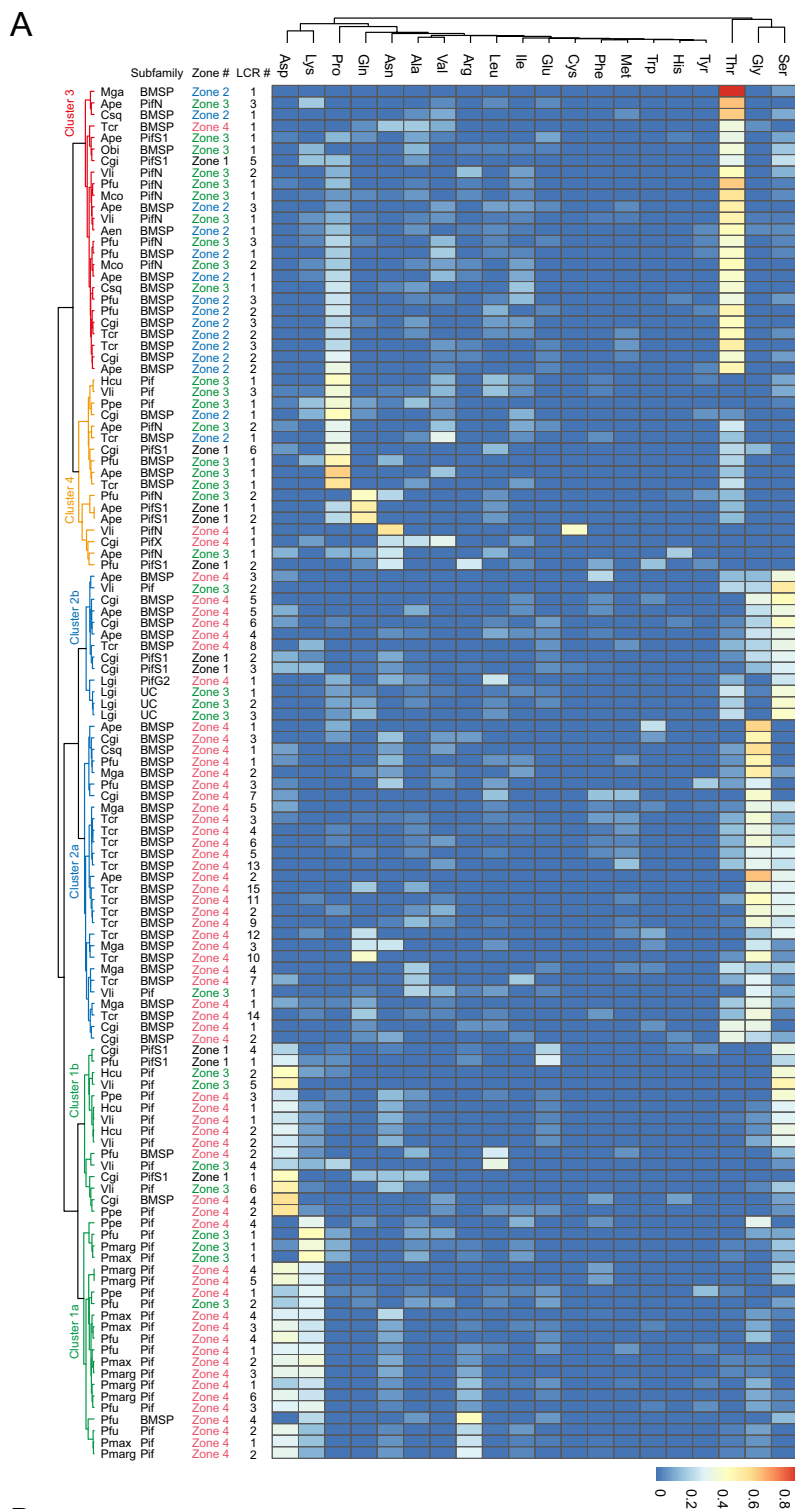
[Cgi\_10004086]) and 2 (other proteins), respectively. See Supplementary Table S5 for species name abbreviations. *UC* uncharacterized Pif/BMSP protein

LCRs in Pifsf and BMSPsf existed almost exclusively in zone 4 (Fig. 4). In contrast, those in PifS1sf and PifNsf were primarily located in zones 1 and 3, respectively (Fig. 4). Most PifCCPsf proteins were inserted by CCP domains rather than LCRs in zone 3.

We then conducted a clustering analysis based on the amino acid composition of the 119 LCRs, generating four clusters (Fig. 4 and Supplementary Figure S3). Cluster 1 comprised LCRs in zones 3 and 4 of the Pifsf proteins and was further classified into two subclusters (Fig. 4A). Cluster



**Fig. 4** Amino acid component of low complexity regions in Pif/BMSP family proteins. **A** Heat map shows amino acid ratio in the low complexity regions. The result of cluster analysis shows on the left. See Supplementary Figure S3 for the detail of cluster analysis. Zone 1, 2, 3, and 4 indicate the upstream region of VWA domain, between tandem VWA domains, between VWA and CB domains, and between conserved LG domains, respectively (see Fig. 4B). LCR # indicates the order of LCRs in each zone (see Fig. 4B). See Supplementary Tables S5 and S7 for species name abbreviations. **B** Schematic representation of zones 1–4 in Pif/BMSP protein (upper) and the LCR # in each zone (lower). UC uncharacterized Pif/BMSP protein



1a consisted of LCRs in zone 3 of the *Pinctada* spp. Pifsf proteins; these regions contained Asp-, Lys-, or Arg-rich sequences (Fig. 4A). In contrast, cluster 1b comprised zones 3 and 4 of Pifsf proteins in freshwater pearl oysters (*V. lienosa* and *H. cumingii*) (Fig. 4A). They also contain Asp residues. Furthermore, Ser residues were present in zone 4 of Pifsf proteins of freshwater pearl oysters (Fig. 4A). Cluster 2 was classified into two subclusters (2a and 2b), comprising LCRs in zone 4 of BMSPsf and zone 1 of PifS1sf proteins, respectively, and containing Thr residues (Fig. 4A). The former contained Gly-rich LCRs, whereas the latter contained Ser-rich LCRs (Fig. 4A). Cluster 3 primarily included LCRs in zone 2 of the BMSPsf and zone 3 of the PifNsf and contained Thr residues (Fig. 4A). Cluster 4 comprised LCRs in zone 3 of Pifsf and BMSPsf and zone 1 of PifS1. The former was characterised by Pro-rich LCRs, whereas the latter contained Gln- or Asn-rich sequences (Fig. 4A).

### Spatial Expression of Pif/BMSP Genes in the Mantle

We investigated the expression regions of four adult shell-related Pif/BMSP genes (*PfuPif*, *PfuBMSP*, *PfuPifN*, and *PfuPifCCP*) in *P. fucata* using in situ hybridization. They were expressed in two distinct areas of the dorsal outer epithelium of the mantle (Fig. 5). Three genes, *PfuPif*, *PfuBMSP*, and *PfuPifN*, were expressed in the mantle pallium corresponding to the nacreous layer-forming region (Fig. 5A–C). The *PfuPif* signal showed the highest intensity and was clearly expressed only in the mantle pallium (Fig. 5A). In contrast, signals of *PfuBMSP* and *PfuPifN* were detected in the mantle pallium and the inner surface of the outer fold, located near the periostracal groove, and on the surface epithelium of the middle and outer folds. The expression of *PfuPifCCP* was relatively weak and ambiguous compared with *PfuPif*, *PfuBMSP*, and *PfuPifN* (Fig. 5D). *PfuPifCCP* was expressed at the mantle edge corresponding to the prismatic layer-forming region (Fig. 5D).

### Protein Interaction Between the VWA Domain and Shell Matrix Proteins

To investigate the function of the Pif VWA domain, we prepared a recombinant PfuPifVWA protein and conducted a pull-down assay. We inserted the VWA domain sequence of PfuPif into the pET-44(+) vector and purified the recombinant protein (r-PifVWA) using a Ni column (Supplementary Figure S4). Pull-down assays were conducted using r-PifVWA; proteins were extracted from the nacreous layer of *P. fucata* using acetic acid. Four SMPs, namely, PfuPif 80 (C-terminal region of Pif), nacrein, and serine proteinase inhibitors (SPIs) (pfu\_aug2.0\_1101.1\_04821.t1 and pfu\_aug2.0\_283.1\_10559.t1) were identified as the major members of the rPifVWA-binding SMPs (more than two

unique peptides) in the nacreous layer (Table 1, Supplementary Table S8).

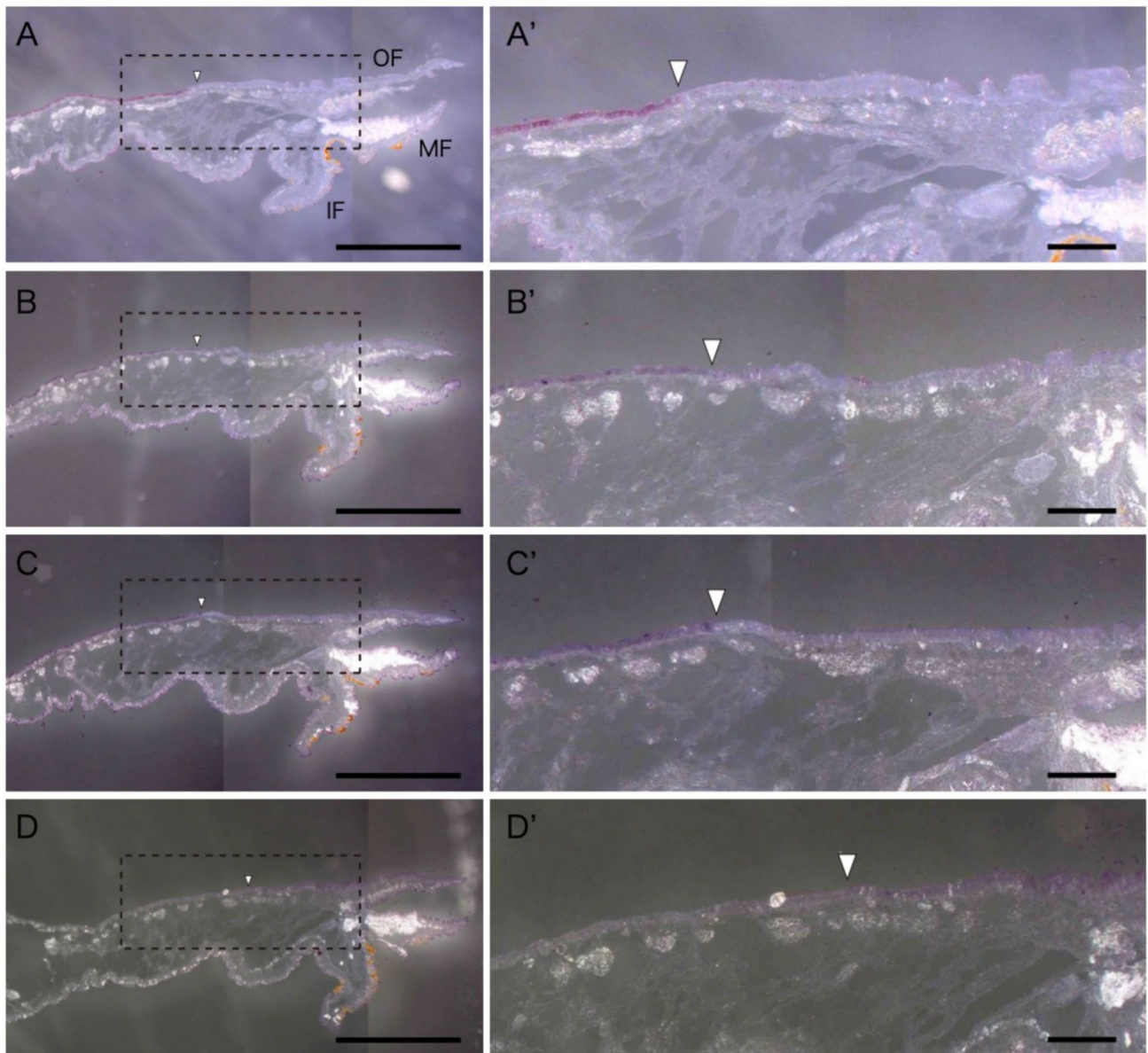
## Discussion

### Classification of Pif/BMSPs

Both VWA and CB dcps are unique proteins in mollusks and various lophotrochozoans (Nemertea, Phoronida, and Brachiopoda) (Zhao et al. 2018; Luo et al. 2015, 2017). These proteins may expand within each phylum, resulting in domain and motif shuffling (Suzuki et al. 2013; Kocot et al. 2016). Although many Pif/BMSP-like genes or proteins containing VWA and CB domains have been reported in mollusks, their evolutionary relationships remain unclear. Our molecular phylogenetic analysis of VWA dcps in molluscs showed that shell-related Pif/BMSP families, including PfuPif and MgaBMSP, formed a clade with strong nodal support in the conchifera (Bivalvia, Gastropoda, Cephalopoda, and Scaphopoda; Fig. 1). This suggests that domain shuffling occurred in the last common ancestor of mollusks and that Pif/BMSP family members are diverse among several taxa.

BMSP is a subclade of the Pif/BMSP family and is well-conserved in the major conchifera (Bivalvia, Gastropoda, Cephalopoda, and Scaphopoda). BMSPs were first identified in the nacreous layer of the blue mussel *M. galloprovincialis* (Suzuki et al. 2011) and are primarily characterised by four tandem VWA domains. In the gastropod *L. gigantea*, the BMSP-like (Lgi236719) has two tandem VWA domains and has been detected in the shell (Marie et al. 2013; Mann and Edsinger 2014). Meanwhile, our results show that Lgi236719 did not form a clade with other BMSP subfamily members (Fig. 1). Rather, Lgi140660 was identified within the BMSP clade and also contained two tandem VWA domains (Figs. 1 and 2). In addition, the four VWA domains of bivalve BMSPs formed each clade while differing from the two gastropod VWA domains (Fig. 1). In cephalopods (*O. bimaculoides*) and scaphopods (*A. entalis*), the BMSPs contained a single VWA domain (Figs. 1 and 2). These results suggest that the ancestral BMSP contained one VWA domain, the duplication of which occurred independently in the leading bivalves and gastropods.

BMSP expression in *P. fucata* was observed in the dorsal region of the outer epithelium of the mantle pallium, corresponding to the nacreous layer-forming region (Fig. 5). The expression of BMSPs in the gastropod *C. squamiferum* (Csq193\_7.17) has also been reported specifically in mantle epithelial cells (Sun et al. 2020), similar to bivalve BMSPs (Fig. 5). Thus, the ancestral BMSP in conchifera might have been expressed in mantle epithelial cells and participated in



**Fig. 5** Spatial expression of shell-related Pif/BMSP genes in the mantle tissue. **A–D** Expression of PfuPif (**A**), PfuBMSP (**B**), PfuPifN (**C**), and PfuPifCCP (**D**) in the mantle epithelium of *P. fucata*. Positive cells are stained in purple. Upper and lower sides of pictures

show dorsal and ventral sides, respectively. Scale bar, 1 mm. **A'–D'** Enlargement of the broken lines square in (**A–D**), respectively. Scale bar, 200  $\mu\text{m}$ . White arrowheads indicate the boundary of gene expression. *IF* inner fold, *MF* middle fold, *MR* mantle rim, *OF* outer fold

**Table 1** The lists of the SMPs that interacted with rPfuPifVWA

Gene_id	Protein name	Layers	Domains	Function	Reference
pfu_aug2.0_715.1_17768	Pif 80	N	LCR	Calcium binding	Suzuki et al. (2009)
pfu_aug2.0_214.1_13802	naerein	N, S	CA	$\text{HCO}_3^-$ production, Calcium binding	Miyamoto et al. (1996), Liu et al. (2015), Zhao et al. (2018)
pfu_aug2.0_1101.1_04821	SPI	N, S	Kunitz_BPTI	Proteinase inhibition	Zhao et al. (2018)
pfu_aug2.0_283.1_10559	Kazal type SPI	N, S	Kazal	Proteinase inhibition	Zhao et al. (2018)
pfu_aug2.0_2218.1_28718	Collagen-like	N	Collagen	Framework interaction	Zhao et al. (2018)

the nacreous layer as well as in the formation of various shell microstructures, such as crossed lamellar.

Pif is a well-known SMP found in the nacreous layers of *P. fucata* (Suzuki et al. 2009). Many Pif-like proteins containing VWA and CB domains have been identified as SMPs (or mantle-specific genes) in bivalves and other mollusks (gastropods, cephalopods, and polyplacophora) (Sun et al. 2020). However, molecular phylogenetic analysis has shown that Pifsf is a unique Pteriomorphia (Suzuki et al. 2009, 2013; Marie et al. 2012) and Unionida subfamily, including *V. lienosa* (Marie et al. 2017) and *H. cumingii* (Zhang et al. 2018). Other Pif-like proteins belonging to the Pif/BMSP family that are not homologues of PfuPif have been previously annotated (Fig. 1). All Pif homologues have been identified in the nacreous layer (Suzuki et al. 2009; Marie et al. 2012, 2017; Zhang et al. 2018), with the PfuPif-encoding gene highly expressed in the dorsal region of the outer epithelium of the mantle pallium, corresponding to the nacreous layer-forming region (Fig. 5).

We annotated a new nacreous layer-related Pif protein (i.e., PifN) that is an SMP found in the nacreous layer of four bivalves: *P. fucata* (Zhao et al. 2018), *A. pectinate* (Shimizu et al. 2022b), *M. coruscus* (Liao et al. 2015), and *V. lienosa* (Marie et al. 2017). They have relatively long insert sequences, including the poly-threonine sequence between the VWA and CB domains, compared to other Pif/BMSP family proteins, excluding PifCCP (Figs. 2, 4). Most Pif/BMSP family members have partial LG domain sequences in the C-terminus, whereas PfuPifN has a complete sequence (Fig. 2). Hence, the ancestral features of Pif/BMSP family proteins may remain in PfuPifN, with a unique sequence, including low-complexity regions, inserted.

Most Pif/BMSP family members comprise only VWA, CB, and LG domains. However, PifCCP contains many repeating shushi/CCP domains between the VWA and CB domains (Fig. 2) (Marie et al. 2017; Zhao et al. 2018). Unlike the PfuPif, PfuPifN, and PfuBMSP proteins that have been identified in the nacreous layer, PfuPifCCP is an SMP that exists in the prismatic layer of *P. fucata* (Zhao et al. 2018). Its expression pattern also differs from that of PfuPif, PfuPifN, and PfuBMSP; that is, PfuPifCCP is expressed in the mantle edge corresponding to the prismatic layer-forming region rather than the mantle pallium corresponding to the nacreous layer-forming region (Fig. 5). This spatial expression pattern is consistent with previous transcriptome analysis results (Zhao et al. 2018). In other pearl oysters, *P. margaritifera*, PUSP11 and PUSP12 comprise two and four Sushi/CCP repeats without VWA, CB, or LG domains, respectively, and are located within the prismatic layer (Marie et al. 2012). Given the incomplete assembly of the gene database, these proteins may represent partial sequences of PifCCP proteins in *P. margaritifera*. Other Sushi/CCP domain-containing proteins have also

been reported as SMPs in mollusks (Marie et al. 2012, 2013, 2017; Mann and Edsinger 2014; Mann et al. 2018; Shimizu et al. 2019) and brachiopods (Jackson et al. 2015; Luo et al. 2015). Although Sushi/CCP repeats may be related to shell formation, their function in biomineralization remains unknown. Nevertheless, with the exception of PifCCP, the combination of the Sushi/CCP domain with the VWA, CB, and LG domains has not been previously reported as an SMP in mollusks or brachiopods.

Considering that PifLS1 proteins have been identified in the larval shells of *P. fucata* and *C. gigas*, they may be involved in larval shell formation in bivalves (Zhao et al. 2018). However, it remains unclear how these proteins are employed during larval shell formation. In addition, homologs of these larval Pif proteins were not identified in a gastropod genome database for *L. gigantea*. Although proteomic analysis of the gastropod larval shells has not yet been reported, this result suggests that PifLS1sf evolved in bivalves (or pteriomorph bivalves). Indeed, the results of the molecular phylogenetic analysis showed that only BMSPsf were present in the major conchifera (Bivalvia, Gastropoda, Cephalopoda, and Scaphopoda), while members of other subfamilies were likely diverse across each class (Bivalvia and Gastropoda; Fig. 1). One Pif/BMSP subfamily, PifG1, and one unknown clade were detected in the gastropod *L. gigantea*, one of which contained SMPs (Lgi\_236719 and Lgi\_232022) identified in the adult shells of this species (Mann et al. 2012; Mann and Edsinger 2014).

### Possible Function of Pif/BMSP Family Proteins in Shell Formation

Various extracellular matrix proteins have been identified in biominerals, including bones and teeth in vertebrates and external skeletons in invertebrates. Most VWA dcps are components of the extracellular matrix (Tuckwell 1999). However, although they are commonly found in biominerals as skeleton matrix proteins (Liu and Zhang 2021), their functions in biomineralization remain unclear. The Von Willebrand factor (VWF) mediates platelet adhesion to collagen (Tuckwell 1999). In general, VWA DCs, such as VWF, are involved in cell adhesion and multiprotein complex formation (Tuckwell 1999; Whittaker and Hynes 2002). Thus, most previous studies on biomineralization have hypothesised that VWA dcps participate in the protein–protein interactions among other matrix proteins. However, matrix proteins that interact with the VWA domain of biomineral proteins to form organic complexes have never been reported.

Using a pull-down assay of rPfuPifVWA, we revealed that the VWA domain region of PfuPif specifically interacts with four previously known SMPs (Pif 80, nacrein, and two serine proteinase inhibitors) (Supplementary Table S8). PfuPif is

cleaved into PfuPif 97 and PfuPif 80 (Suzuki et al. 2009). PfuPif 97 consists of VWA and CB domains that can interact with chitin, whereas PfuPif 80 has an aragonite-binding ability (Suzuki et al. 2009). Furthermore, the results of gel filtration high-performance liquid chromatography (HPLC) and SDS–PAGE analyses showed that PfuPif 97, PfuPif 80, and PfuN16—the major proteins in the nacreous layer (Samata et al. 1999)—may form a high-molecular-weight complex (Suzuki et al. 2009). In addition, Suzuki et al. (2009) hypothesised that PfuPif 97 and PfuPif 80 form a complex, resulting in interactions with chitin (an organic sheet) and the nucleation of aragonite crystals. Our pull-down assay results support these findings; PfuPif 97 can bind to PfuPif 80 via the VWA domain region and likely forms a complex. In contrast, the interaction between PfuN16 and PfuPif 97 or PfuPif 80 was likely not mediated through the VWA domain.

We also identified Pfunacrein as a protein that complexes with PfuPif 97, in addition to PfuPif 80. Nacrein is a shell-related carbonic anhydrase (CA) in the nacreous and prismatic layers of *P. fucata* (Miyamoto et al. 1996; Miyashita et al. 2002). This enzyme catalyses the hydration of carbon dioxide and provides bicarbonate ions. Nacrein contains a specific Gly–Xaa–Asn repeat (Xaa = Asp, Asn, or Glu) between the CA subdomains, which interacts with calcium ions and participates in calcium carbonate crystal formation (Miyamoto et al. 1996). However, nacrein is an acid- or EDTA-soluble protein that may not readily interact directly with the insoluble organic matrix sheets composed of chitin fibrils, which form the basis of aragonite crystal growth in the nacreous layer. Therefore, PfuPif 97 with VWA and CB domains can bridge the nacrein and the organic matrix sheets to efficiently promote the nucleation and growth of calcium carbonate crystals.

Finally, rPfuPifVWA also interacted with two types of SPIs, Kunitz-type (pfu\_aug2.0\_1101.1\_04821) and Kazal-type (pfu\_aug2.0\_283.1\_10559) (Table 1). Many Kunitz-type SPIs have been identified within the nacreous or prismatic layers of pearl oysters, *P. fucata*, *P. margaritifera*, *P. maxima* (Marie et al. 2012; Liu et al. 2015; Zhao et al. 2018), and other bivalves (Gao et al. 2015; Liao et al. 2015, 2019; Marie et al. 2017; Shimizu et al. 2022b) and were clustered in the genome of *P. fucata* (Takeuchi et al. 2016a). Although the function of SPIs in shell formation is unclear, their diversification likely indicates their involvement in biomineralisation (e.g., regulation of matrix maturation) (Arivalagan et al. 2017; Dombre et al. 2016). The zona pellucida (ZP) domain is also involved in extracellular protein–protein interactions, such as the VWA domain (Jovine et al. 2002, 2005, 2006). Various ZP domain-containing proteins have been identified from skeletal matrix proteins in corals (Ramos-Silva et al. 2013; Takeuchi et al. 2016b) and gastropods (Marie et al. 2013; Mann and Edsinger 2014), while a portion of the ZP domain has also been identified as SMPs in bivalves

(Shimizu et al. 2022a). The results of the pull-down assay using recombinant ZP protein and SMPs extracted from the nacreous layer of *P. fucata* showed that the ZP protein interacts with other SMPs, such as proteinase inhibitors (SPIs) and BMSP 100 (Shimizu et al. 2022a). However, the SMPs interacting with the ZP domain (Shimizu et al. 2022a) and the VWA domain (this study) differ slightly. These results suggest that different types of organic complexes are formed in the shell-forming region by different protein mediators (e.g., VWA dcps and ZP dcp) with roles in the formation of various microstructures.

## Evolution of Pif/BMSP Family Members in Mollusks

In the pearl oyster, *P. fucata*, 17 types of VWA–CB dcps were found in the genome database, six of which were identified in adult and larval shells (Zhao et al. 2018). Moreover, a BLAST search revealed that the VWA domains of nine VWA–CB dcps exhibited similarity with the VWA of PfuPif (*e*-value < 1.0e–10). Thus, the remaining seven VWA–CB dcps likely differed from the ‘Pif/BMSP family’ (Supplementary Figure S5). Phylogenetic analysis of VWA domains further showed that six shell-related VWA–CB dcps and three other VWA–CB dcps (total nine) formed a ‘Pif/BMSP family’ clade. Thus, the origin of the VWA domain in the Pif/BMSP family differs from that in other VWA–CB dcps. Hence, VWA–CB dcps likely evolved at least twice independently by domain shuffling. Pif/BMSP family members have evolved in the common Mollusca ancestor and are diverse in several taxa. This novel protein likely participates in shell formation as most Pif/BMSP family genes are expressed in the mantle tissue, and some have been identified as SMPs (Suzuki et al. 2009, 2013; Zhang et al. 2012; Marie et al. 2013, 2017; Mann and Edsinger 2014; Liao et al. 2015; Gao et al. 2015; Takeuchi et al. 2021; Shimizu et al. 2022b). Ancestral molluscs are thought to have mineralised structures such as shell(s) and sclerites (Scherholz et al. 2013; Vinther et al. 2017; Giribet and Edgecombe 2020), which are composed of organic matrices, such as extracellular matrix proteins, and polysaccharides, such as chitin. A previous genome study of a chiton *A. granulata* (Varney et al. 2021) showed that a VWA–CB dcp (Agr\_g24110) called Pif-like (Varney et al. 2021) was highly expressed in the girdle, which involves sclerites formation compared with other tissues. The phylogenetic analysis further revealed that Agr\_g24110 is located in the base of Pif/BMSP family members in mollusks (Fig. 1). Thus, the common ancestor of mollusks contains Pif/BMSP, which may also be involved in the formation of mineralised structures (shells and sclerites).

Aside from bivalves, BMSPs have also been identified in other conchifera (Gastropoda, Cephalopoda, and Schaphopoda) (Fig. 1). This suggests that BMSPs are ancestral Pif/BMSPs in the common conchifera ancestor and are involved

in shell formation. Various microstructures (e.g., crossed lamellar, prismatic, and nacreous layers) have been observed in the shells of several conchifera (Marin et al. 2012), with diverse Pif/BMSPs in several groups (Fig. 1). In fact, six Pif/BMSP-encoding genes in *P. fucata* were expressed in different parts of the mantle tissue (mantle edge or mantle pallium; Fig. 5) at different developmental stages (D-shaped larva or adult) and within the shells of adults and larva (Zhao et al. 2018).

PfuPif was first identified as a nacreous layer-specific matrix protein in the pearl oyster *P. fucata* (Suzuki et al. 2009) and is specifically expressed in the epithelial cells of the mantle pallium (Fig. 5). Nacre is commonly observed in other bivalves (paleotaxodonts, pteriomorphids, paleoheterodonts, and anomalodesmata) and has evolved independently (Carter 1990). However, homologues of Pif proteins were identified as SMPs in the nacreous layers of the freshwater mussels *V. lienosa* and *H. cumingii* (Marie et al. 2017; Zhang et al. 2018) (Fig. 1). The C-terminal region of Pif/BMSPs contains two conserved regions of the LG domain (Suzuki et al. 2013). The insertion sequences in the Pif subfamily differ from those in other subfamilies (Fig. 3). Although the domain structures of Pif/BMSPs are similar (Fig. 2), the insertion sequences, including many low-complexity regions between the conserved partial sequences of the LG domain in the Pifsf, are quite different from those of other subfamilies (Figs. 3, 4). Furthermore, they differed slightly between the genus *Pinctada* and the order Unionida. Many basic residues, Lys (K) and Arg (R), exist in *Pinctada* spp. (17–19% and 10–15%, respectively) but are relatively minor in Unionida (around 8% and 2–4%, respectively). Instead of basic residues, many Ser (S) residues exist in Unionida (25–26%), not *Pinctada* spp. (3–4%; Fig. 3). In contrast, the acidic residue Asp (D) is the major residue in the *Pinctada* and Unionida (24–26% and 21–23%, respectively; Fig. 3); hence, the Asp-rich sequence is likely important for the interaction with calcium ions and shell mineralisation. Meanwhile, the insert sequence in *P. penguin* was intermediate between those of *Pinctada* and Unionida (Fig. 3). This suggests that Pif 80 and Pifs in other *Pinctada* species (*P. margaritifera* and *P. maxima*) evolved by accumulating mutations and through natural selection in each lineage. Moreover, these proteins were likely involved in independently evolved well-ordered nacreous layer formation.

## Conclusion

Various microstructures exist within the molluscan shells; their formation is precisely regulated by secretions of organic matrix components from the mantle epithelial cells. In the current study, we focused on PfuPif—a well-known SMP in the nacreous layer of *P. fucata*—revealing a portion of the function

and evolution of Pif/BMSP family members in mollusks. PfuPif is cleaved to form PfuPif 97 and PfuPif 80. The former comprises VWA and CB domains that are highly conserved in other mollusk SMPs. We identified SMPs that interact with the VWA domain of PfuPif 97, including PfuPif 80, which binds to calcium carbonate crystals. Our findings support the hypothesis that PfuPif 97 and PfuPif 80 form a complex, resulting in interactions with chitin (an organic sheet) and the nucleation of aragonite crystals (Suzuki et al. 2009).

We also conducted a molecular phylogenetic analysis using VWA domain sequences because Pif/BMSPf proteins that contain VWA and CB domains were detected within the nacreous layer and other microstructures; however, their molecular evolution is unclear. We found that Pif/BMSPf proteins were classified into eight subfamilies (BMSPsf, Pifsf, PifNsf, PifLSsf, PifGsf, PifS1sf, PifXsf, and PifCCPsf) that exhibited different sequential features, expression patterns, and diversity in different taxa. Furthermore, Pif/BMSPf proteins contained one or more LCRs with amino acid components and insert regions that differed among subfamilies. Our results showed that Pif proteins in pearl oysters (genus *Pinctada*) and freshwater pearl oysters (Unionida) have independently evolved more characteristic LCRs (D, S, K, or R-rich) than other Pif/BMSP family members; these features probably correlate with well-ordered nacreous layer formation. Collectively, the findings of this study provide useful insights regarding the molecular mechanisms of shell mineralisation as well as applied research using biomimetic mineralisation in medicine, dentistry, or materials science.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00239-024-10180-1>.

**Acknowledgements** We thank two professional editors who are native speakers of English for English editing of this manuscript.

**Author Contributions** KS and MS conceived and designed the study. KS, LN, HK, and MS conducted the experiments and analyzed the data. KS and MS prepared the manuscript. All the authors have read and approved the final manuscript.

**Funding** Open Access funding provided by The University of Tokyo. This research was partially funded by Grant-in-Aid for Scientific Research B (JP19H03045), Grant-in-Aid for Scientific Research A (23H00339), Grant-in-Aid for Scientific Research on Innovative Areas IBmS: JSPS KAKENHI Grant Number JP19H05771, Environmental Restoration and Conservation Agency (1CN-2201), New Energy and Industrial Technology Development Organization (NEDO). This research was also partially supported by Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP23ama121009.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR et al (2015) The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524(7564):220–224. <https://doi.org/10.1038/nature14668>
- Arivalagan J, Yarra T, Marie B et al (2017) Insights from the shell proteome: biomineralization to adaptation. *Mol Biol Evol* 34:66–77. <https://doi.org/10.1016/j.molgen.2016.03.005>
- Belcher A, Wu X, Christensen R, Hansma PK, Stucky GD, Morse DE (1996) Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature* 381:56–58. <https://doi.org/10.1038/381056a0>
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Carter JG (1990) Chapter 10. Evolutionary significance of shell microstructure in the Palaeotaxodonta, Pteriomorpha and Isofilibranchia (Bivalvia: Mollusca). In: Carter JG (ed) *Skeletal biomineralization: patterns, processes and evolutionary trends*. Van Nostrand Reinhold, New York, pp 135–296
- Dombre C, Guyot N, Moreau T, Monget P, Da Silva M, Gautron J, Réhault-Godbert S (2016) Eggsepins: the chicken and/or the egg dilemma. *Semin Cell Dev Biol*. <https://doi.org/10.1016/j.semcdb.2016.08.019>
- Falini G, Albeck S, Weiner S, Addadi L (1996) Control of aragonite or calcite polymorphism by mollusk shell macromolecules. *Science* 271:67–69. <https://doi.org/10.1126/science.271.5245.67>
- Feng D, Li Q, Yu H, Kong L, Du S (2017) Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. *Sci Rep* 7:45754. <https://doi.org/10.1038/srep45754>
- Gao P, Liao Z, Wang XX, Bao LF, Fan MH et al (2015) Layer-by-layer proteomic analysis of *Mytilus galloprovincialis* shell. *PLoS ONE* 10:e0133913. <https://doi.org/10.1371/journal.pone.0133913>
- Giribet G, Edgecombe GD (2020) Mollusca. In: Giribet G, Edgecombe GD (eds) *The invertebrate tree of life*. Princeton University Press, Oxford, pp 358–390
- Gregoire C (1957) Topography of the organic components in mother-of-pearl. *J Biophys Biochem Cytol* 3(5):797–808. <https://doi.org/10.1083/jcb.3.5.797>
- Jackson DJ, Mann K, Häussermann V, Schilhabel MB, Lüter C et al (2015) The *Magellania venosa* biomineralizing proteome: a window into brachiopod shell evolution. *Genome Biol Evol* 7:1349–1362. <https://doi.org/10.1093/gbe/evv074>
- Jovine L, Qi H, Williams Z, Litscher E, Wassarman PM (2002) The ZP domain is a conserved module for polymerization of extracellular proteins. *Nat Cell Biol* 4:457–461. <https://doi.org/10.1038/ncb802>
- Jovine L, Darie CC, Litscher ES, Wassarman PM (2005) Zona pellucida domain proteins. *Annu Rev Biochem* 74:83–114. <https://doi.org/10.1146/annurev.biochem.74.082803.133039>
- Jovine L, Janssen WG, Litscher ES, Wassarman PM (2006) The PLAC1-homology region of the ZP domain is sufficient for protein polymerization. *BMC Biochem* 7(1):11. <https://doi.org/10.1186/1471-2091-7-11>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160–1166. <https://doi.org/10.1093/bib/bbx108>
- Kocot KM, Aguilera F, McDougall C, Jackson DJ, Degnan BM (2016) Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. *Front Zool* 13:23. <https://doi.org/10.1186/s12983-016-0155-z>
- Kuraku S, Zmasek CM, Nishimura O, Katoh K (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res* 41:W22–W28. <https://doi.org/10.1093/nar/gkt389>
- Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* 46(D1):D493–D496
- Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* 43(Database issue):D257–D260
- Levi-Kalishman Y, Falini G, Addadi L, Weiner S (2001) Structure of the nacreous organic matrix of a bivalve mollusk shell examined in the hydrated state using cryo-TEM. *J Struct Biol* 135:8–17. <https://doi.org/10.1006/jsbi.2001.4372>
- Liao Z, Bao LF, Fan MH, Gao P, Wang XX et al (2015) In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *J Proteomics* 122:26–40. <https://doi.org/10.1016/j.jprot.2015.03.027>
- Liao Z, Jiang YT, Sun Q, Fan MH, Wang JX, Liang HY (2019) Microstructure and in-depth proteomic analysis of *Perna viridis* shell. *PLoS ONE* 14(7):e0219699. <https://doi.org/10.1371/journal.pone.0219699>
- Liu C, Zhang R (2021) Biomineral proteomics: a tool for multiple disciplinary studies. *J Proteome* 238:104171. <https://doi.org/10.1016/j.jprot.2021.104171>
- Liu C, Li S, Kong J, Liu Y, Wang T et al (2015) In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Sci Rep* 5:17269. <https://doi.org/10.1038/srep17269>
- Luo YJ, Takeuchi T, Koyanagi R, Yamada L, Kanda M et al (2015) The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat Commun* 6:8301. <https://doi.org/10.1038/ncomms9301>
- Luo YJ, Kanda M, Koyanagi R, Hisata K, Akiyama T, Sakamoto H, Sakamoto T, Satoh N (2017) Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nat Ecol Evol* 2:141–151. <https://doi.org/10.1038/s41559-017-0389-y>
- Mann K, Edsinger E (2014) The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. *Proteome Sci* 12(12):28. <https://doi.org/10.1186/1477-5956-12-28>
- 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 Sci* 10:28. <https://doi.org/10.1186/1477-5956-10-28>
- Mann K, Cerveau N, Gummich M, Fritz M, Mann M, Jackson DJ (2018) In-depth proteomic analyses of *Haliotis laevigata* (greenlip abalone) nacre and prismatic organic shell matrix. *Proteome Sci* 16:11. <https://doi.org/10.1186/s12953-018-0139-3>
- Marie B, Joubert C, Tayalé A, Zanella-Cléon I, Belliard C et al (2012) Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *Proc Natl Acad Sci USA* 109:20986–20991. <https://doi.org/10.1073/pnas.1210552109>

- Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cléon I, Guichard N et al (2013) The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. *FEBS J* 280:214–232. <https://doi.org/10.1111/febs.12062>
- Marie B, Arivalagan J, Mathéron L, Bolbach G, Berland S, Marie A et al (2017) Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the unionoida *Elliptio complanata* and *Villosa lianosa*. *J R Soc Interface* 14:20160846. <https://doi.org/10.1098/rsif.2016.0846>
- Marin F, Le Roy N, Marie B (2012) The formation and mineralization of mollusk shell. *Front Biosci* 1(4):1099–1125. <https://doi.org/10.2741/s321>
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A (1996) A carbonic anhydrase from the nacreous layer in oyster pearls. *Proc Natl Acad Sci USA* 93:9657–9660. <https://doi.org/10.1073/pnas.93.18.9657>
- Miyashita T, Takagi R, Miyamoto H, Matsushiro A (2002) Identical carbonic anhydrase contributes to nacreous or prismatic layer formation in *Pinctada fucata* (Mollusca: bivalvia). *Veliger* 45:250–255
- Oudot M, Neige P, Shir IB, Schmidt A, Strugnell JM et al (2020) The shell matrix and microstructure of the Ram's horn squid: molecular and structural characterization. *J Struct Biol* 211:107507. <https://doi.org/10.1016/j.jsb.2020.107507>
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>
- Ramos-Silva P, Kaandorp J, Huisman L et al (2013) The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. *Mol Biol Evol* 30(9):2099–2112
- Samata T, Hayashi N, Kono M, Hasegawa K, Horita C, Akera S (1999) A new matrix protein family related to the nacreous layer formation of *Pinctada fucata*. *FEBS Lett* 462:225–229. [https://doi.org/10.1016/s0014-5793\(99\)01387-3](https://doi.org/10.1016/s0014-5793(99)01387-3)
- Scherholz M, Redl E, Wollesen T, Todt C, Wanninger A (2013) Aplousophoran mollusks evolved from ancestors with polyplacophoran-like features. *Curr Biol* 23(21):2130–2134. <https://doi.org/10.1016/j.cub.2013.08.056>
- Setiamarga DHE, Hirota K, Yoshida M-a, Takeda Y, Kito K et al (2021) Hydrophilic shell matrix proteins of *Nautilus pompilius* and the identification of a core set of conchiferan domains. *Genes* 12(12):1925. <https://doi.org/10.3390/genes12121925>
- Shimizu K, Kimura K, Isowa Y, Oshima K, Ishikawa M et al (2019) Insights into the evolution of shell and love dart of land snail revealed from their matrix proteins. *Genome Biol Evol* 11:380–397. <https://doi.org/10.1093/gbe/evy242>
- Shimizu K, Kintsu H, Awaji M, Matumoto T, Suzuki M (2020) Evolution of biomineralization genes in the prismatic layer of the pen shell *Atrina pectinata*. *J Mol Evol* 88(10):742–758. <https://doi.org/10.1007/s00239-020-09977-7>
- Shimizu K, Negishi L, Ito T, Touma S, Matsumoto T et al (2022a) Evolution of nacre- and prisms-related shell matrix proteins in the pen shell, *Atrina pectinata*. *Comp Biochem Physiol Part D Genom Proteom* 44:101025. <https://doi.org/10.1016/j.cbd.2022.101025>
- Shimizu K, Takeuchi T, Negishi L, Kurumizaka H, Kuriyama I, Endo K, Suzuki M (2022b) Evolution of epidermal growth factor (EGF)-like and zona pellucida domains containing shell matrix proteins in mollusks. *Mol Biol Evol* 39:msac148. <https://doi.org/10.1093/molbev/msac148>
- Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P et al (2013) Insights into bilateral evolution from three spiralian genomes. *Nature* 493:526–531. <https://doi.org/10.1038/nature11696>
- Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M et al (1997) Structures of mollusc shell framework proteins. *Nature* 387:563–564. <https://doi.org/10.1038/42391>
- Sun Q, Jiang Y, Fan M, Zhang X, Xu H, Liao Z (2020) Characterization of a novel shell matrix protein with vWA domain from *Mytilus coruscus*. *Biosci Biotechnol Biochem* 84:1629–1644. <https://doi.org/10.1080/09168451.2020.1756735>
- Suzuki M, Saruwatari K, Kogure T et al (2009) An acidic matrix protein, Pif, is a key macromolecule for nacre formation. *Science* 11:1388–1390. <https://doi.org/10.1126/science.1173793>
- 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(16):2478–2487. <https://doi.org/10.1002/cbic.201100317>
- Suzuki M, Iwashima A, Kimura M, Kogure T, Nagasawa H (2013) The molecular evolution of the Pif family proteins in various species of mollusks. *Mar Biotechnol* 15:145–158. <https://doi.org/10.1007/s10126-012-9471-2>
- Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K et al (2016a) Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zool Lett* 2:3. <https://doi.org/10.1186/s40851-016-0039-2>
- Takeuchi T, Yamada L, Shinzato C, Sawada H, Satoh N (2016b) Stepwise evolution of coral biomineralization revealed with genome-wide proteomics and transcriptomics. *PLoS ONE* 11:e0156424. <https://doi.org/10.1371/journal.pone.0156424>
- Takeuchi T, Fujie M, Koyanagi R, Plasseraud L, Ziegler-Devlin I, Brosse N et al (2021) The ‘shellome’ of the crocus clam *Tridacna crocea* emphasizes essential components of mollusk shell biomineralization. *Front Genet* 12:674539. <https://doi.org/10.3389/fgene.2021.674539>
- Tuckwell D (1999) Evolution of von Willebrand factor A (VWA) domains. *Biochem Soc Trans* 27:835–840
- Varney RM, Speiser DI, McDougall C, Degnan BM, Kocot KM (2021) The iron-responsive genome of the chiton *Acanthopleura granulata*. *Genome Biol Evol* 13:evaa263. <https://doi.org/10.1093/gbe/evaa263>
- Vinther J, Parry L, Briggs DEG, Van Roy P (2017) Ancestral morphology of crown-group molluscs revealed by a new Ordovician stem aculiferan. *Nature* 542(7642):471–474
- Wada K (1961) Crystal growth of molluscan shells. *Bull Natl Pearl Res Lab Jpn* 7:703–828
- Weiner S, Hood L (1975) Soluble protein of the organic matrix of mollusk shells: a potential template for shell formation. *Science* 190(4218):987–989. <https://doi.org/10.1126/science.1188379>
- Weiss IM, Kaufmann S, Mann K, Fritz M (2000) Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliotis laevigata*. *Biochem Biophys Res Commun* 267:17–21. <https://doi.org/10.1006/bbrc.1999.1907>
- Whittaker CA, Hynes RO (2002) Distribution and evolution of von Willebrand/integrin a domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* 13:3369–3387
- Wootton JC, Federhen S (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* 266:554–571
- Zhang G, Fang X, Guo X, Li L, Luo R et al (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49–54. <https://doi.org/10.1038/nature11413>
- Zhang R, Qin M, Shi J, Tan L, Xu J et al (2018) Molecular cloning and characterization of Pif gene from pearl mussel, *Hyriopsis cumingii*, and the gene expression analysis during pearl formation. *3 Biotech* 8:214. <https://doi.org/10.1007/s13205-018-1233-z>



- Zhang Y, Mao F, Mu H, Huang M, Bao Y et al (2021) The genome of *Nautilus pompilius* illuminates eye evolution and biomineralization. *Nat Ecol Evol* 5(7):927–938. <https://doi.org/10.1038/s41559-021-01448-6>
- Zhao R, Takeuchi T, Luo YJ, Ishikawa A, Kobayashi T et al (2018) Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation. *Mol Biol Evol* 35:2751–2761. <https://doi.org/10.1093/molbev/msy172>