




# Evolution of Epidermal Growth Factor (EGF)-like and Zona Pellucida Domains Containing Shell Matrix Proteins in Mollusks

Keisuke Shimizu <sup>1</sup>, Takeshi Takeuchi <sup>2</sup>, Lumi Negishi,<sup>3</sup> Hitoshi Kurumizaka <sup>3</sup>, Isao Kuriyama,<sup>4</sup> Kazuyoshi Endo,<sup>5</sup> and Michio Suzuki<sup>\*1</sup>

<sup>1</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>2</sup>Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan

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

<sup>4</sup>Mie Prefecture Fisheries Research Institute, 3564-3 Hamajima, Hamajima-cho, Shima City, Mie 517-0404, Japan

<sup>5</sup>Department of Earth and Planetary Science, The University of Tokyo, 7-3-1 Hongo, Tokyo 113-0033, Japan

\*Corresponding author: E-mail: amichio@g.ecc.u-tokyo.ac.jp.

Associate editor: Yoko Satta

## Abstract

Several types of shell matrix proteins (SMPs) have been identified in molluskan shells. Their diversity is the consequence of various molecular processes, including domain shuffling and gene duplication. However, the evolutionary origin of most SMPs remains unclear. In this study, we investigated the evolutionary process EGF-like and zona pellucida (ZP) domains containing SMPs. Two types of the proteins (EGF-like protein (EGFL) and EGF-like and ZP domains containing protein (EGFZP)) were found in the pearl oyster, *Pinctada fucata*. In contrast, only EGFZP was identified in the gastropods. Phylogenetic analysis and genomic arrangement studies showed that EGFL and EGFZP formed a clade in bivalves, and their encoding genes were localized in tandem repeats on the same scaffold. In *P. fucata*, EGFL genes were expressed in the outer part of mantle epithelial cells are related to the calcitic shell formation. However, in both *P. fucata* and the limpet *Nipponacmea fuscoviridis*, EGFZP genes were expressed in the inner part of the mantle epithelial cells are related to aragonitic shell formation. Furthermore, our analysis showed that in *P. fucata*, the ZP domain interacts with eight SMPs that have various functions in the nacreous shell mineralization. The data suggest that the ZP domain can interact with other SMPs, and EGFL evolution in pterimorph bivalves represents an example of neo-functionalization that involves the acquisition of a novel protein through gene duplication.

**Key words:** Biomineralization, Neo-functionalization, Organic complex, Tandem duplication, Zona pellucida.

## Introduction

Understanding the origin of a novel gene is crucial for understanding the molecular basis of biodiversity in nature. Gene duplication is well known as an important avenue for providing new genetic material to species (Ohno 1970). A new paralogous gene resulting from gene duplication can acquire a novel function (neo-functionalization) or share the ancestral function with the original gene (sub-functionalization) (Lynch and Force 2000; Conant and Wolfe 2008; Kaessmann 2010). These paralogs can change their expression levels or spatial expression patterns when their regulatory regions, such as cis-regulatory modules, are changed by mutation (Necsulea and Kaessmann 2014). These mutations are also important in driving phenotypic evolution.

Some of the biomineralized structures in various metazoan lineages have been independently acquired in the Early Cambrian period (Murdock 2020). These mineralized structures are composed of inorganic crystals (e.g., calcium

carbonates and calcium phosphates) and numerous organic matrix (OM) components (Lowenstam and Weiner 1989). Molluskan shells are a good model for understanding biomineralization processes in invertebrates and have been well studied. They are composed of polymorphs of calcium carbonate crystals (aragonite and/or calcite) and show various microstructures with distinct features (e.g., prismatic, nacreous, and crossed-lamellar) (Carter 1990; Marin et al. 2012). Proteinaceous OM of shells called shell matrix proteins (SMPs) are secreted from the mantle tissue, which is covered by epithelial cells. Some SMPs form scaffold structures for mineralization (Suzuki and Nagasawa 2013), while others are involved in crystal growth, selection of crystal polymorphs, or organic-inorganic interactions by surrounding or being incorporated into the calcium carbonate crystals (Song et al. 2019; Cölfen and Mann 2003; Cölfen and Antonietti 2005; Marin et al. 2008). The SMP components are different between the microstructures (e.g., prismatic layer vs.

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

nacreous layer), and the genes encoding those SMPs are expressed in different zones of the mantle epithelium (Takeuchi and Endo 2006; Gardner et al. 2011; Marie et al. 2012). For instance, the nacreous layer-related SMP genes are expressed in the inner zone of the mantle (inner pallium), whereas the prismatic layer-related ones are expressed in the outer zone of the mantle (outer pallium and mantle edge) (Marie et al. 2012).

Numerous SMPs have been identified in the shells of mollusks. Interestingly, it was difficult to find homologues for most SMPs from other metazoans genomes due to rapid evolution, even in closely related species. For instance, the acidic SMP Aspein consists of many aspartic acids (60.41% of a full-length Aspein in *P. fucata*) (Tsukamoto et al. 2004; Takeuchi et al. 2008) and is diverse in other Pteriidae species (Isowa et al. 2012). A similar pattern was observed for other SMPs containing repetitive, low-complexity domains such as lysine [K]-rich mantle proteins, and shematrixins (McDougall et al. 2013). On the other hand, some SMPs contain specific domains that appear to be highly conserved among conchiferan mollusks. For instance, Pif was identified as a nacre protein in *P. fucata* containing one von Willebrand factor type A (VWA), one chitin-binding (ChtB), one chitin-binding-like (ChtBL), and one laminin G (LG) domain (Suzuki et al. 2009); these domains are conserved in other SMPs (Marie et al. 2017). BMSP (blue mussel shell protein) also contains four domains (VWA, ChtB, ChtBL, and LG) like Pif, but BMSP has four tandem VWA domains and evolved through the duplication of the VWA domain from the Pif (Suzuki Iwashima et al. 2011; Suzuki et al. 2013, 2017). The tyrosinase domain has been found in SMPs of various molluscan lineages, suggesting a common, ancient origin. However, the tyrosinase gene family is known to have expanded independently in the pearl oyster *Pinctada fucata* and the Pacific oyster *Crassostrea gigas* (Aguilera et al. 2014), and SMPs with a tyrosinase domain have also evolved in several gastropod species independently (Aguilera et al. 2014; Shimizu et al. 2019). A similar evolutionary history has been reported for carbonic anhydrase domain-containing SMPs. Gene duplication occurred in the carbonic anhydrase gene family, and the paralogs have been co-opted independently for skeletogenesis in multiple metazoan lineages (Le Roy et al. 2015). Likewise, the dermatopontin genes duplicated at least twice independently in several lineages of pulmonate snails (basommatophorans and stylommatophorans) and were independently co-opted as an SMP for their shell calcification (Sarashina et al. 2006). In the comparative genomic analyses of the bivalves *P. fucata* and *C. gigas*, lineage-specific gene family expansion was observed in the gene families involved in biotic and abiotic stress responses as well as in shell mineralization (Zhang et al. 2012; Takeuchi, Koyanagi, et al. 2016). The dynamic genome evolution such as expansion of gene families by gene duplications can help to accelerate the divergence of molluscan shell mineralization and to adapt to the highly stressful environments (Zhang et al. 2012; Takeuchi, Koyanagi, et al. 2016; Takeuchi, Yamada, et al. 2016; Kocot et al. 2016).

The epidermal growth factor (EGF)-like domain is a common domain in molluscan SMPs (Aguilera et al. 2017). In the Pacific oyster *C. gigas*, Gigasin-2, which contains a signal peptide and two EGF-like domains in tandem, was first identified by LC-MS/MS analysis (Marie, Zanella-Cléon, Guichard, et al. 2011). Later, similar EGF-like proteins (EGFL proteins) were found as SMPs from a number of bivalves (Marie et al. 2012; Zhang et al. 2012; Gao et al. 2015; Liao et al. 2015; Iwamoto et al. 2020). Although many proteome studies have been performed on molluscan SMPs identified EGFL proteins, the evolutionary origin of the EGFL proteins remains unclear.

Other type of EGF-like domain containing SMP was identified from the limpet *Lottia gigantea* (Mann et al. 2012; Marie et al. 2013; Mann and Edsinger 2014). This SMP called LUSP-17 is slightly different from the EGFL proteins in bivalves: LUSP-17 contains not only a signal peptide and two EGF-like domains in tandem in the N-terminus but also one zona pellucida (ZP) domain in the C-terminus (Marie et al. 2013). We named the proteins containing both EGF-like and ZP domains as EGF-like and ZP domains containing protein (EGFZP).

The ZP proteins containing the ZP domain are found in many eukaryotic EMCs and are often glycosylated (Bork and Sander 1992). ZP2 is expressed in mammalian oocytes and is involved in oogenesis and fertilization (Wassarman and Mortillo 1991). These ZP proteins play important roles in protein polymerization and extracellular protein-protein interactions (Jovine et al. 2002, 2005, 2006). In the abalone *Haliotis corrugata*, ZP proteins were reported as component of the vitelline envelope (egg coating) and, therefore, labeled VEZP (Aagaard et al. 2006). ZP proteins are also a part of the skeletal matrix proteins in corals (Ramos-Silva et al. 2013; Takeuchi, Yamada, et al. 2016) and gastropods (Mann et al. 2012; Marie et al. 2013; Mann and Edsinger 2014). However, the evolutionary relationships among those ZP proteins remain completely unclear.

A considerable number of SMPs have been reported from molluscan shells, showing a high degree of diversity. Although the roles of gene duplication for neo- or sub-functionalization in general are already known in various organisms, there has been no previous study that clearly demonstrated evolution of SMPs through neo-functionalization. Recent investigations on the transcriptomes and genomes of various mollusks provide an opportunity to understand the evolutionary origin(s) of SMPs. We thus focused on EGFL and EGFZP proteins and revealed that the EGFL gene was derived from the EGFZP gene through neo-functionalization. We searched EGFL and EGFZP proteins over the genome and transcriptome databases of 20 lophotrochozoans (see Materials and Methods for details). We analyzed the molecular phylogeny, genomic arrangements, and spatial gene expression patterns of the EGFL and EGFZP genes. Furthermore, we conducted the pull-down assay to find the SMPs interact with ZP domain in the pearl oyster *P. fucata*. This study provides new insights into understanding the evolution of key gene families involved in the divergence of molluscan shell mineralization.

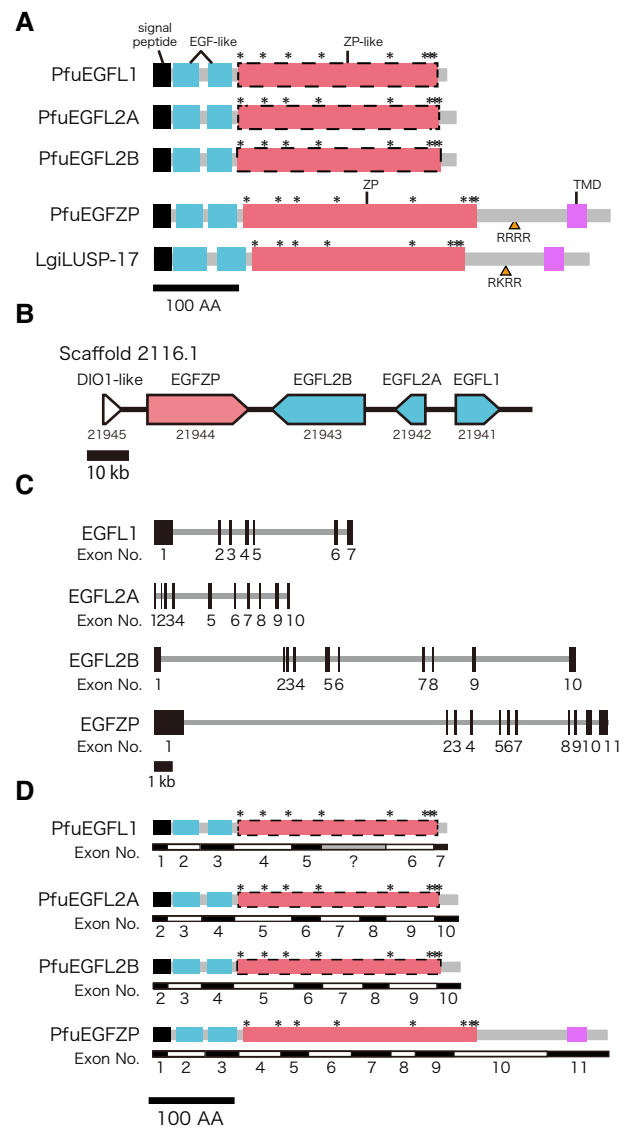
## Results

### Characterization of EGFL and EGFZP Proteins in *P. fucata*

To find homologs of EGFZP in the pearl oyster *P. fucata*, we searched for EGFZP proteins in the genome database, as well as in the mantle transcriptome data, of *P. fucata* (Takeuchi, Koyanagi, et al. 2016) by BLASTP (query protein was LgiLUSP-17) and found one copy of EGFZP-like protein (pfu\_cdna2.0\_064724 [gene model id: pfu\_aug2.0\_2116.1\_21944]). This protein (PfuEGFZP) contains a signal peptide, two EGF-like domains, one ZP domain, and one transmembrane domain (TMD) (fig. 1A). Although we could not identify the ZP domain in the three EGFL proteins of *P. fucata* (PfuEGFL1: pfu\_cdna2.0\_057745 [gene model id: pfu\_aug2.0\_2116.1\_21941], PfuEGFL2A: pfu\_cdna2.0\_061795 [gene model id: pfu\_aug2.0\_2116.1\_21942], and PfuEGFL2B: pfu\_cdna2.0\_061794 [gene model id: pfu\_aug2.0\_2116.1\_21943]) using the HMMER v3.3.2 (<http://hmmer.org/>; last accessed February 10, 2022) and the InterProScan (supplementary table S1, Supplementary Material online), we detected ZP domain-like sequences in their C-termini by the online version of Simple Modular Architecture Research Tool (SMART) (version 9.0) (Letunic et al. 2015; Letunic and Bork 2018; <http://smart.embl-heidelberg.de>; last accessed October 2, 2020) with non-significant e-value scores (fig. 1A, supplementary table S1, Supplementary Material online). We also found eight conserved cysteine (Cys) residues that are conserved in other ZP domain-containing proteins and are likely to participate in intramolecular disulfide bonding (Wassarman et al. 2001) from both the ZP domain and the ZP domain-like region in EGFZP proteins (PfuEGFZP, LgiEGFZP1, and LgiEGFZP2) and three PfuEGFL proteins, respectively (fig. 1; supplementary fig. S1, Supplementary Material online). We also identified one potential furin cleavage site (RRRR or RKRR) that is known in other ZP domain-containing proteins (Jovine et al. 2002) between the ZP domain and TMD in PfuEGFZP and LgiLUSP-17 (fig. 1A). The genes encoding the three EGFL and one EGFZP protein were arranged in tandem in the same scaffold (2216.1) (fig. 1B), although the gene structures (exon and intron distribution patterns) were different among them (fig. 1C). However, the boundaries of exons in EGFL2A and EGFL2B show similar patterns (fig. 1D). In addition, the three exons of these genes that are corresponding to the signal peptide and two EGF-like domains are well conserved (fig. 1D).

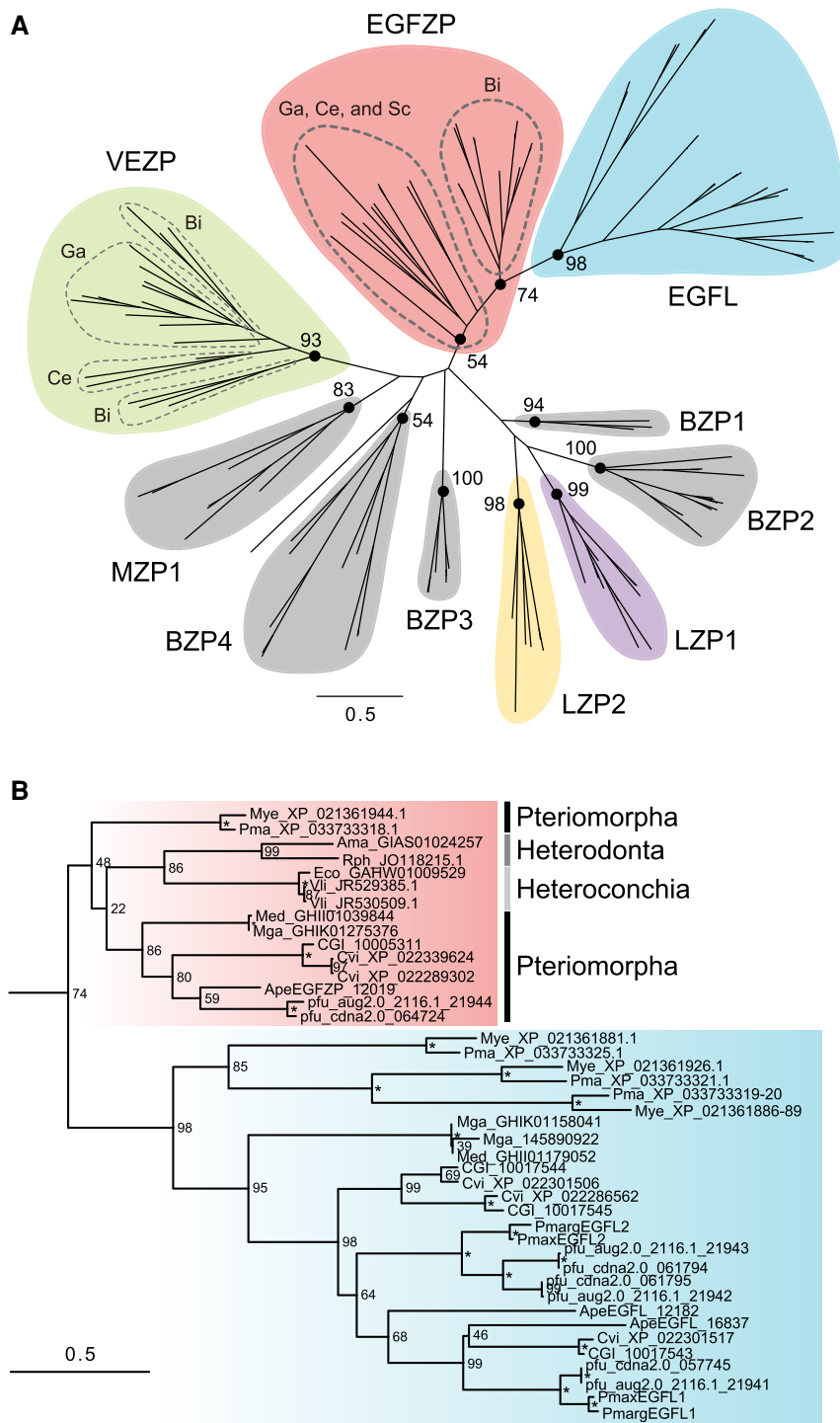
### Phylogenetic Analysis of ZP Proteins

By using BLASTP, the sequences of EGFL or EGFZP proteins were retrieved from a total 16 genomes and transcriptome databases. The ZP domain-containing proteins were also identified from the genome and transcriptome databases using InterProScan (supplementary table S2, Supplementary Material online), and they were used for molecular phylogenetic analysis. Then, we performed molecular



**Fig. 1.** Schematic representation of EGFL and EGFZP proteins. (A) Three EGFL proteins of *Pinctada fucata* (PfuEGFL1, 2A, and 2B) consist of a signal peptide (black box), two EGF-like domains, and one ZP-like domain. EGFZP proteins of *P. fucata* (PfuEGFZP) and *Lottia gigantea* (LUSP-17) are similar to EGFL proteins of *P. fucata*. However, they undoubtedly have one ZP domain and an additional TMD in their C-termini. Yellow arrowheads indicate the potential cleavage sites by furin enzyme (amino acid sequences are "RRRR" and "RKRR"). The Asterisks indicate the conserved Cys residues. (B, C) Genomic organization of EGFL and EGFZP in *P. fucata*. Arrows indicate the direction of the transcript (B), and black boxes represent exons (C). (D) The boundary of exons of EGFL and EGFZP. Black, blue, red, and pink boxes indicate signal peptide, EGF-like domain, ZP domain, and transmembrane, respectively. Asterisks indicate conserved cysteine residues. Gray box in the exon structure of PfuEGFL1 indicate the missing coding region in genome data (poly-N part). DIO1-like, type I iodothyronine deiodinase-like.

phylogenetic analysis to reveal the relationships among ZP proteins, including EGFL and EGFZP. The results showed that molluscan ZP proteins can be classified into at least nine families with high bootstrap accuracy (fig. 2A; supplementary figs. S2 and S3, Supplementary Material



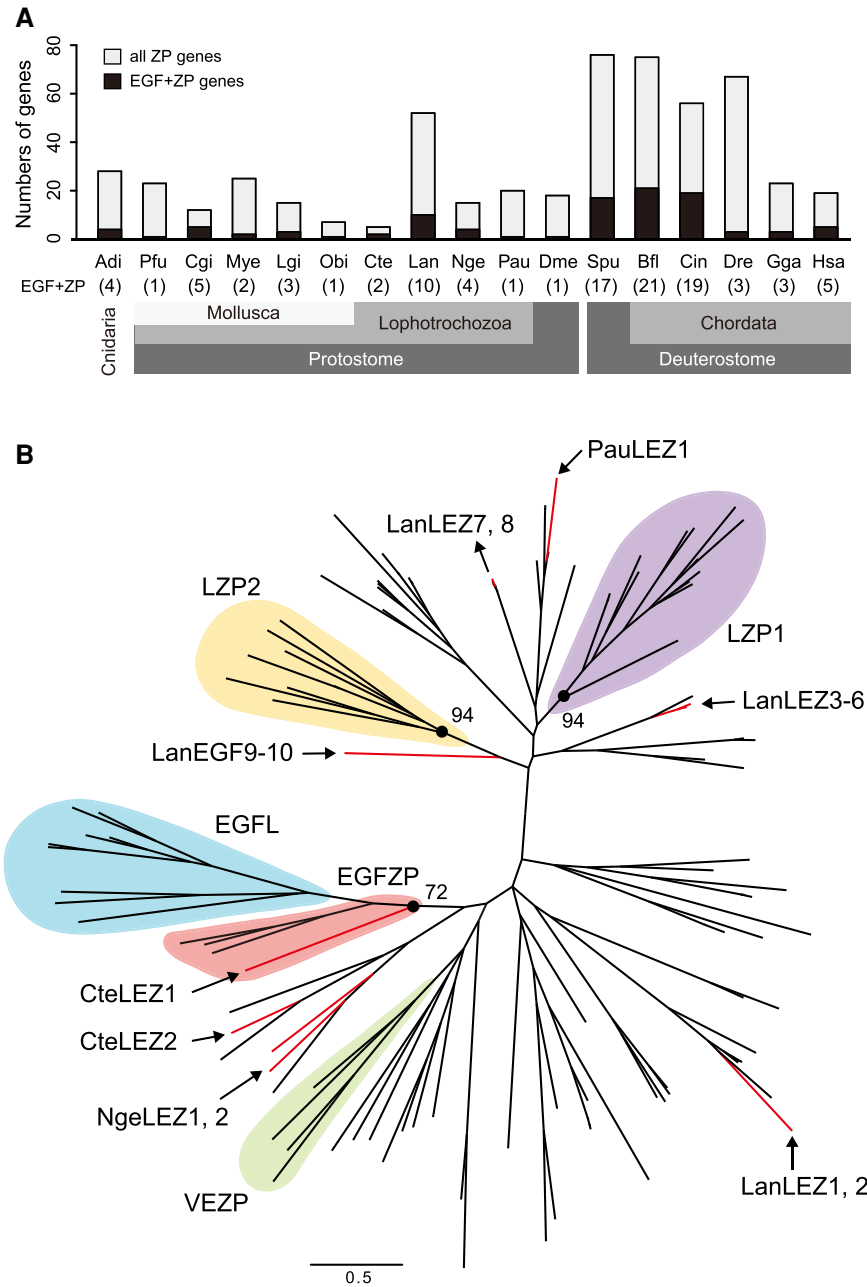
**FIG. 2.** Molecular phylogeny of molluscan ZP proteins. (A) The maximum likelihood tree was inferred from 137 ZP protein sequences under the WAG +  $\Gamma$  model (196 positions, 1000 bootstrap replicates). Branch lengths are proportional to the expected number of substitutions per site, as indicated by the scale bar. Numbers on nodes indicate the bootstrap values. (B) Details of the clade contain EGFZP and EGFL proteins in bivalves in (A). Asterisks indicate 100% bootstrap support. Ama, *Archivesica marissinica*; Ape, *Atrina pectinata*; Bi, Bivalvia; BZP, Bivalve ZP protein; Ce, Cephalopoda; CGI, *Crassostrea gigas*; Cvi, *Crassostrea virginica*; Eco, *Elliptio complanata*; Ga, Gastropoda; LZP, lophotrochozoan ZP protein; Med, *Mytilus edulis*; Mga, *Mytilus galloprovincialis*; Mye, *Mizuhopecten yessoensis*; MZP, molluscan ZP protein; Obi, *Octopus bimaculoides*; Ovu, *Octopus vulgaris*; Pfu, *Pinctada fucata*. Pmarg, *Pinctada margaritifera*; Pmax, *Pinctada maxima*; Pma, *Pecten maximus*; Rph, *Ruditapes philippinarum*; Sc, Scaphopoda; Vli, *Villosa lianosa*; VEZP, Vitelline envelope ZP protein.

online), and one of them can be classified into two subfamilies EGFZP and EGFL (fig. 2A; supplementary fig. S2, Supplementary Material online). We named these families as follows: EGFZP, which include EGFZP and EGFL proteins, VEZP, one molluscan ZP proteins (MZP1), four bivalve ZP proteins (BZP1-4), and two lophotrochozoan ZP proteins (LZP1 and 2) (fig. 2A; supplementary figs. S2 and S3, Supplementary Material online). EGFZP and EGFL proteins of bivalves formed a clade (fig. 2B). The clade generated by combination of the EGFZP- and EGFL-related clade was a

sister to the clade of EGFZP proteins in gastropods, cephalopods, and scaphopods (fig. 2A; supplementary fig. S2, Supplementary Material online). The VEZP family contained the vitelline envelope (egg coating) proteins of the abalone *H. corrugata* (Aagaard et al. 2006). The MZP1 proteins were found in the genome databases of three mollusks *P. fucata* (Bivalvia), *L. gigantea* (Gastropoda), and *O. bimaculoides* (Cephalopoda) (supplementary fig. S2, Supplementary Material online). The four BZP families (BZP1-4) were found only in bivalves (*P. fucata*, *C. gigas*, *Myzohpecten yessoensis*,



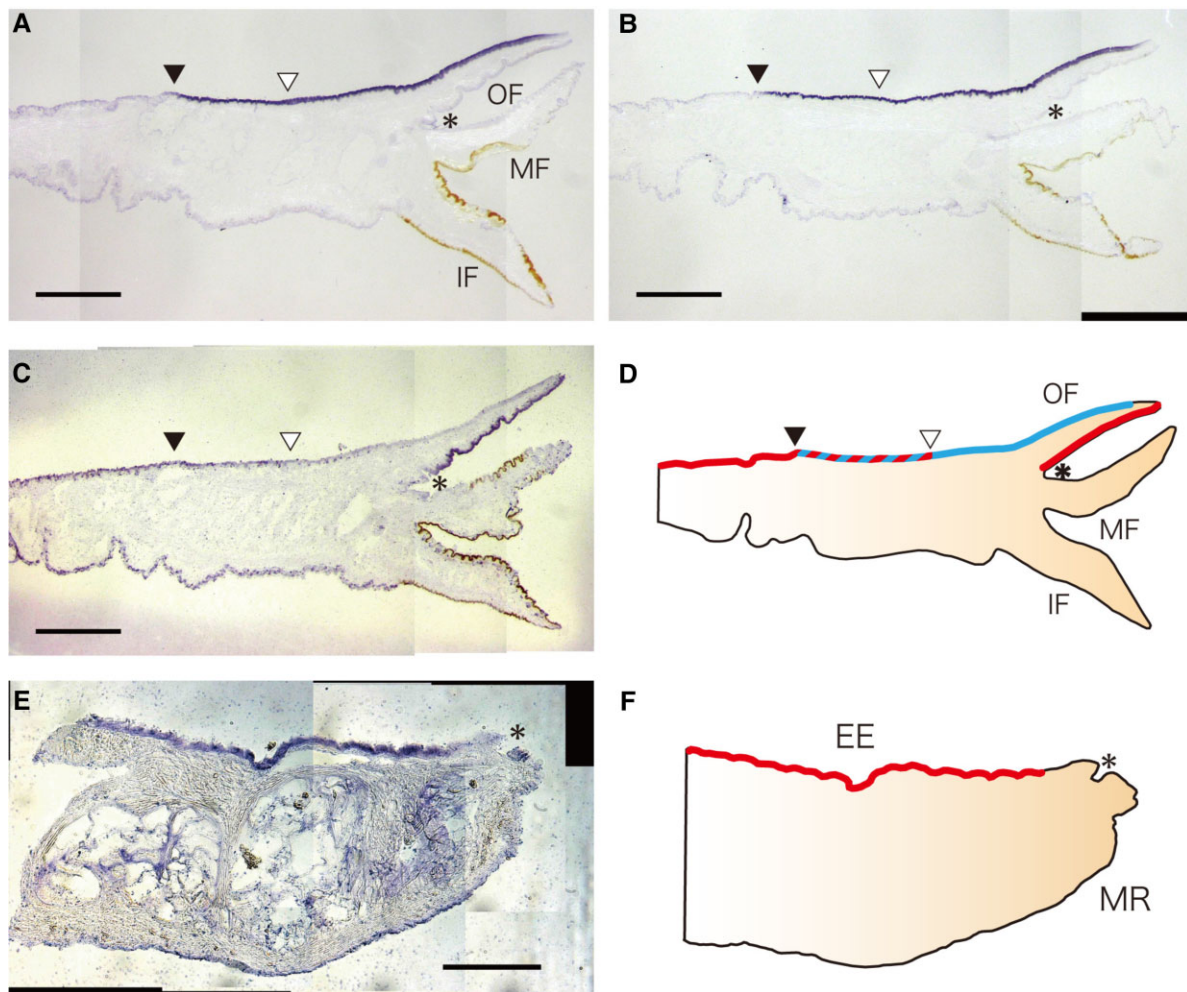
**Fig. 3.** Evolution of EGF-like and ZP domains-containing proteins in animals. (A) Numbers of EGF-like and/or ZP domains-containing genes in 17 animal genomes. Numbers indicate the number of both EGF-like and ZP domains-containing genes (EGF + ZP). Adi, *Acropora digitifera*; Bfl, *Branchiostoma floridae*; Cgi, *Crassostrea gigas*; Cin, *Ciona intestinalis*; Cte, *Capitella teleta*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Gga, *Gallus gallus*; Hsa, *Homo sapiens*; Lan, *Lingula anatina*; Lgi, *Lottia gigantea*; Mye, *Mizuhopecten yessoensis*; Nge, *Notospermus geniculatus*; Obi, *Octopus bimaculoides*; Pau, *Phoronis australis*; Pfu, *Pinctada fucata*; Spu, *Strongylocentrotus purpuratus*. (B) Molecular phylogeny of lophotrochozoan ZP proteins. The maximum likelihood tree was inferred from 128 ZP proteins sequences under the WAG + Z model (221 positions, 1000 bootstrap replicates). Branch lengths are proportional to the expected number of substitutions per site, as indicated by the scale bar. Red lines indicate ZP proteins with EGF-like domains in four lophotrochozoans (*C. teleta*, *L. anatina*, *N. geniculatus*, and *P. australis*). Domain structures of these proteins are shown in [supplementary fig. S3, Supplementary Material](#) online. LEZ, lophotrochozoan EGF-like and ZP domains-containing protein; LZP, lophotrochozoan ZP protein; VEZP, Viteline envelope ZP protein.



or *Pecten maximus*) ([supplementary fig. S2, Supplementary Material](#) online). The two LZP families (LZP1 and 2) were found in other lophotrochozoans, as well as mollusks (details are shown in the next section).

We then extended the investigation of ZP proteins to the genome databases of 17 animals using InterProScan. We found many ZP proteins with EGF-like domains ([fig. 3A; supplementary table S3, Supplementary Material](#) online). In four lophotrochozoans, the annelid *Capitella teleta*, brachiopod *Lingula anatina*, nemertea *Notospermus geniculatus*, and phoronid *Phoronis australis*, both EGF-like and ZP domains were present in 2, 10, 4, and 1 protein/s, respectively ([fig. 3A](#)). These ZP proteins were called lophotrochozoan EGF-like and ZP domain-containing proteins (LEZ proteins). Some of LEZ proteins contained other

domains, such as scavenger receptor cysteine-rich (SRCR), low-density lipoprotein receptor domain class A, complement protein C1r/C1s, Uegf, and Bmp1, and zinc-dependent metalloprotease domains, in addition to the EGF-like and ZP domains ([supplementary fig. S4, Supplementary Material](#) online). To understand the evolution of ZP proteins in lophotrochozoa, we analyzed ZP proteins from four non-molluscan lophotrochozoans genomes (the annelid *C. teleta*, brachiopod *L. anatina*, nemertea *N. geniculatus*, and phoronid *P. australis*) and three molluscan genomes (*P. fucata*, *C. gigas*, and *M. yessoensis*). Molecular phylogenetic analysis indicated a diverse origin of ZP proteins in lophotrochozoans. While molluscan EGFZP and EGFL proteins formed a clade, any LEZ proteins of other lophotrochozoans except for Cte\_210134 (*C. teleta*) do not



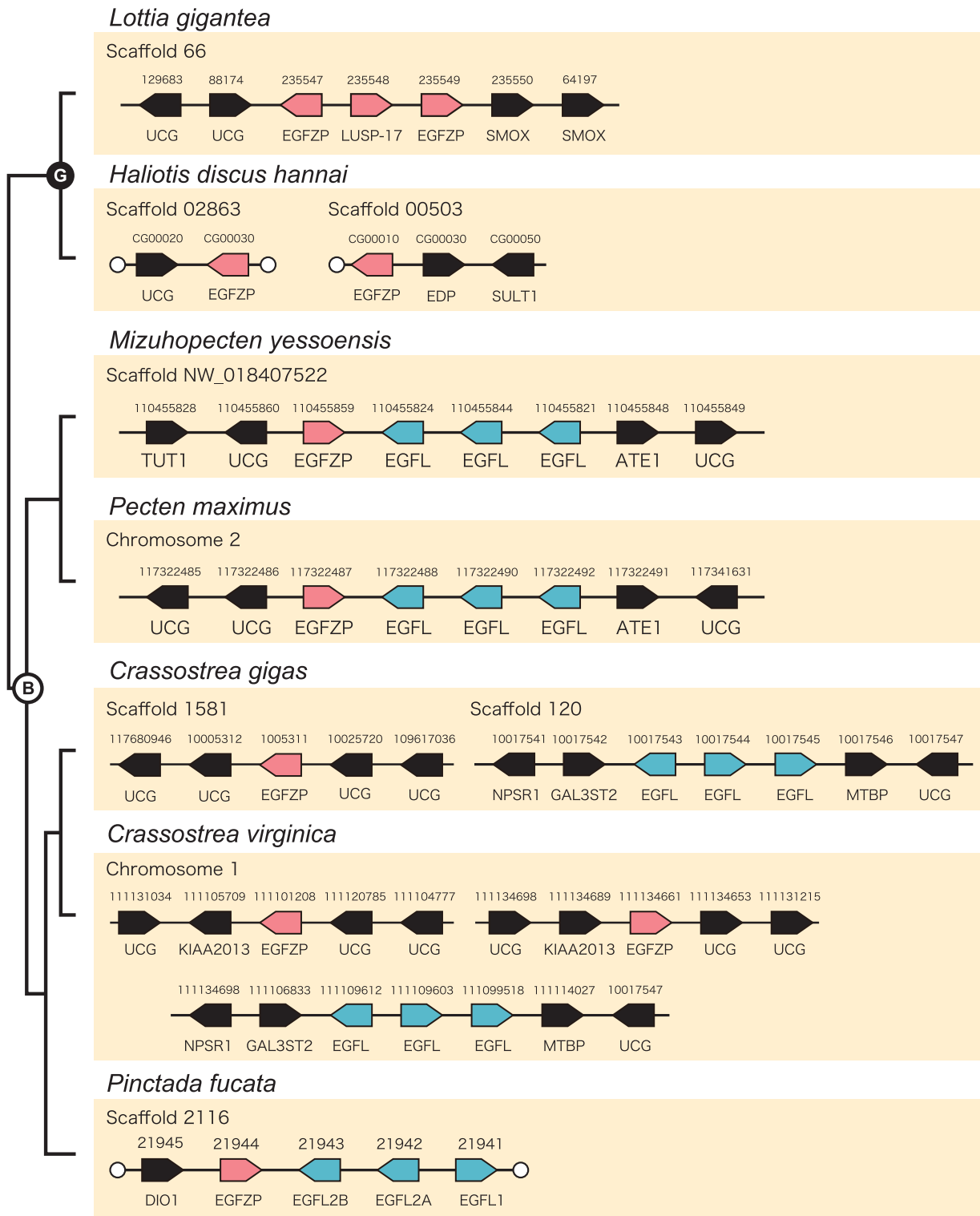
**Fig. 4.** Spatial expression patterns of EGFL and EGFZP genes in the mantle tissue. (A–C) Expression of PfuEGFL1 (A), PfuEGFL2A (B), and PfuEGFZP (C) in the mantle epithelium of *P. fucata*. Scale bar, 500  $\mu$ m. (D) Schematic representation of these expression patterns in the mantle epithelium. Red and blue colors indicate the specific expression regions of PfuEGFZP and PfuEGFLs, respectively. Black and white arrowheads indicate the end of the expression region of PfuEGFLs and PfuEGFZP, respectively. The asterisk indicates the periostracal groove. (E) Expression of NfuEGFZP in the mantle epithelium of *N. fuscoviridis*. Scale bar, 200  $\mu$ m. (F) Schematic representation of NfuEGFZP in the mantle epithelium. Red color indicates the expression regions of NfuEGFZP. The asterisk indicates the periostracal groove. EE, external epithelium; IF, inner fold; MF, middle fold; MR, mantle rim; OF, outer fold.

form a clade with them (fig. 3B; supplementary fig. S5, Supplementary Material online). In contrast, at least two ZP families were conserved in the five lophotrochozoans (fig. 3B; supplementary fig. S5, Supplementary Material online). One ZP family named LZP1 consisted of a signal peptide, one or more SRCR domains (PF00530), ZP domain, and TMD; the other family, named LZP2, consisted of a signal peptide, a ZP domain, and TMD (supplementary fig. S6, Supplementary Material online).

#### Spatial Expression of EGFL and EGFZP Genes in the Mantle

We analyzed the expression regions of PfuEGFLs and PfuEGFZP in the mantle using *in situ* hybridization. The signals of PfuEGFL1 and PfuEGFL2A were observed in the dorsal region of the outer epithelium of the mantle corresponding to the mantle edge and the outer pallium

regions (fig. 4A and B). On the other hand, PfuEGFZP is expressed in two regions: the middle and outer pallial regions of the outer epithelium as well as in the inner surface of the outer fold and the inner surface of the outer fold, which is located near the periostracal groove, is involved in periostracum formation (fig. 4C). The signals of PfuEGFZP in the outer surface of the outer fold were weaker than in other regions (fig. 4C). No signal was detected in the ventral region of the mantle or in the middle and the inner folds for PfuEGFL1, EGFL2, or EGFZP (fig. 4A–D). To compare the expression pattern of EGFZP between the bivalve and the gastropod, we performed *in situ* hybridization using the mantle of the limpet *N. fuscoviridis*. NfuEGFZP signal was observed in the dorsal region of the outer epithelium, except in the outermost edge part (fig. 4E and F). No signal was detected in the ventral region of the mantle (fig. 4E and F).



**Fig. 5.** Genomic arrangement of EGFL and EGFZP genes in mollusks. Arrangement of EGFL and EGFZP genes among two gastropods (*Lottia gigantea* and *Haliotis discus hannai*) and five bivalves (*Mizuhopecten yessoensis*, *Pecten maximus*, *Crassostrea gigas*, *C. virginica*, and *Pinctada fucata*). Arrows indicate the direction of the transcripts. White circles indicate the end of the scaffolds. B, bivalves; DIO1-like, type I iodothyronine deiodinase-like; EDP, EGF-like domain containing protein; G, gastropods; Mal1: molybdenum cofactor sulfurase; NPSR1, neuropeptide S receptor; SMOX, Spermine oxidase; SULT1, sulfotransferase 1; TUT1, speckle targeted PIP5K1A-regulated poly(A) polymerase-like.

#### Genomic Arrangement of EGFL and EGFZP Genes

One EGFL1, two EGFL2, and one EGFZP encoding gene were found in the whole genome assembly of *P. fucata*. These genes were located in tandem in the same scaffold

(fig. 5). The direction of these genes was different, with PfuEGFL2A and PfuEGFL2B being forward, and PfuEGFL1 and PfuEGFZP being reversed. In the oyster *C. gigas*, EGFL and EGFZP genes exist in the same scaffold, and three

EGFL genes (one EGFL1 and two EGFL2) exist in tandem (fig. 5). However, EGFZP genes are not arranged in tandem with EGFL genes (fig. 5). This pattern is also observed in the closely related species *C. virginica* (fig. 5). The directions of EGFL1 and EGFL2 are different: EGFL2A and EGFL2B are forward, and EGFL1 is reverse both in *C. gigas* and *C. virginica*. The two scallops *M. yessoensis* and *P. maximus* have three EGFL1/2 and one EGFZP encoding gene, and all of them are located in tandem in the homologous scaffold or chromosome (fig. 5). The direction of the genes was different for EGFL1/2 and EGFZP in both *M. yessoensis* and *P. maximus* (fig. 5). In gastropods *L. gigantea* and *H. discus hannai*, we also found three and two copies of the EGFZP gene, respectively (fig. 2A; supplementary fig. S2, Supplementary Material online). The EGFZP genes in *L. gigantea* were located in tandem in the same scaffold (fig. 5), while those in *H. discus hannai* were found in different scaffolds (fig. 5). We also confirmed that two flanking genes of EGFL and EGFZP genes are conserved only between the two closely related species of Pacific oysters (*C. gigas* and *C. virginica*) and between the scallops (*Mizuhopecten yessoensis* and *Pecten maximus*) (fig. 5); in the Pacific oysters, the flanking genes of EGFL genes at 5' and 3' sides are galactose-3-O-sulfotransferase 2 (GAL3ST2) and mdm2-binding protein (MTBP), respectively, and in the scallops, the flanking genes of EGFZP and EGFL genes at 5' and 3' sides are uncharacterized gene (UCG), and arginyl-tRNA-protein transferase 1-like (ATE1), respectively (fig. 5).

#### Protein Interaction between ZP Domain and SMPs

The ZP domain is known to be involved in protein-protein interactions, especially between egg and sperm, and is probably able to interact with other extracellular proteins such as SMPs in mollusks. We inserted the ZP domain sequence of PfuEGFZP into the pET-44(+) vector and prepared recombinant protein (r-PfuZP) (supplementary fig. S7, Supplementary Material online). The Protein-binding assay was conducted using r-PfuZP and SMPs, which were extracted from nacreous layer of *P. fucata*. A total

of 20 proteins were identified as the r-PfuZP-binding proteins, and eight of them have already been reported as SMPs in the nacreous layer (Table 1; supplementary tables S4 and S5, Supplementary Material online). Four of those eight SMPs contained specific domains, including serine proteinase inhibitors (Kunitz\_BPTI, Antistasin, or Kazal), and BMSP100 (Lam\_G). The remaining four SMPs included Shematin (Yano et al., 2006), nacre uncharacterized shell protein 16-like (Marie et al. 2012), and two uncharacterized shell proteins (Zhao et al. 2018).

## Discussion

### EGF-like Domain-containing Proteins in EMCs

Proteins containing the EGF-like domain are known as extracellular matrix components (EMCs) in metazoans. Various functions of EGF-like domains have been reported, including EGF signaling, Ca<sup>2+</sup> binding ability, protein aggregation, and protein/protein recognition (Maurer and Hohebest 1997), and some EGF-like domain-containing EMCs are probably involved in metazoan biomineralization. For instance, EGF-like repeats and discoidin I-like domain-containing protein 3 (EDIL3) was identified in the chicken eggshell (Mann et al. 2006). EDIL3 contains three EGF-like domains in the N-terminus and two coagulation factor 5/8 domains and is thought to be involved in eggshell mineralization (Marie et al. 2015) and calcium transport during eggshell formation (Stapane et al. 2019). In the brachiopod *L. anatina*, fibrillar collagens that were identified from the shell contain EGF domains and two or more collagen domains (Luo et al. 2015). These collagens are different from those found in vertebrate bones in origin, and those EGF-domain containing collagens became diverse in the *Lingula* lineage by domain shuffling (Luo et al. 2015). In another brachiopod, *Magellania venosa*, an EGF-like domain-containing protein was identified from the shells (Jackson et al. 2015). However, it is different from the EGF-like domain-containing SMPs in the brachiopod *L. anatina* or in

**Table 1.** The Lists of the SMPs That Interacted with r-PfuZP.

Gene_id	Protein name	Structure	Domains	Function	Reference
pfu_aug2.0_1101.1_04821.t1	SPI	N, P	Kunitz_BPTI	Proteinase inhibition	Liu et al. 2015; Zhao et al. 2018
pfu_aug2.0_2907.1_25578.t1	NSPI3/4	N, P	Antistasin, Kunitz_BPTI	Proteinase inhibition	Marie et al. 2012; Zhao et al. 2018
pfu_aug2.0_283.1_10559.t1	Kazal type SPI	N, P	Kazal	Proteinase inhibition	Zhao et al. 2018
pfu_aug2.0_1726.1_21763.t1	BMSP100	N, P	LamG, LCR	Calcium binding	Suzuki et al. 2013; Liu et al. 2015; Zhao et al. 2018
pfu_aug2.0_663.1_11059.t1	Shematin8	N, P	Gly-rich domain, LCR	Framework interaction	Yano et al. 2006; Liu et al. 2015; Zhao et al. 2018
pfu_aug2.0_53.1_10183.t1	NUSP16	N	Gly (16%), Met (13%), Gln (13%), Asn (12%), and Pro (12%)	unknown	Marie et al. 2012; Zhao et al. 2018
pfu_aug2.0_899.1_31259.t1	USP	N, P	LCR, coiled coil	unknown	Zhao et al. 2018
pfu_aug2.0_583.1_10957.t1	USP	N	LCR	unknown	Zhao et al. 2018

N, nacreous layer; P, prismatic layer.



mollusks in evolutionary origins. Previous studies suggest that EGF-like domain-containing proteins are relatively common in the skeletal matrix proteins in metazoans but may have evolved independently through domain shuffling in several metazoan lineages.

### EGFL and EGFZP Proteins in Mollusks

Proteins containing two EGF-like domains in tandem have been reported as SMPs in many bivalves (Marie, Zanella-Cléon, Corneillat, et al. 2011; Marie et al. 2012; Zhang et al. 2012; Gao et al. 2015; Liao et al. 2015; Liu et al. 2015). This type of EGF-like protein was not found to be a skeletal matrix protein in other invertebrates such as sea urchins (Mann et al. 2010) and brachiopods (Luo et al. 2015; Jackson et al. 2015). Gigasin-2, annotated under the ID number CGI\_100543 in the *C. gigas* genome, was first identified from the shells of the Pacific oyster *C. gigas* (Marie, Zanella-Cléon, Corneillat, et al. 2011). Two kinds of EGF-like proteins (EGFL1 and EGFL 2) were identified from the prismatic layer in the two pearl oysters *P. maxima* and *P. margaritifera* (Marie et al. 2012). In *P. margaritifera*, EGFL1 and EGFL2 encoding genes are highly expressed in the mantle edge corresponding to the prismatic layer forming region than in the mantle pallium, corresponding to the nacreous layer forming region (Marie et al. 2012). Recently, CgELC (EGFL2), CGI\_10017544 in the *C. gigas* database, was found in the chalky layer of *C. gigas* (Iwamoto et al. 2020). The results of *in vitro* crystallization using CgELC suggest that this protein is incorporated into the calcite crystals and is involved in the aggregation of polycrystalline calcite (Iwamoto et al. 2020). LUSP-17 (EGFZP) was also found in the shells of *L. gigantea*, but this protein contains one ZP domain in their C-terminus (Marie et al. 2013). However, this type of EGF-like domain-containing protein has not been identified as an SMP in other gastropods (Mann and Jackson 2014; Shimizu et al. 2019; Ishikawa et al. 2020). In the pond snail, *Lymnaea stagnalis*, EGF-like domain-containing protein was identified as an SMP, but it contains only one EGF-like domain in the N-terminus and two whey acidic protein (WAP) domains in the C-terminus (Ishikawa et al. 2020). Thus, the EGFZP protein possibly has evolved as SMPs only in the Patellogastropod lineage.

### Evolution of ZP Proteins

ZP proteins were first identified as egg coat proteins in mammals (Wassarman and Mortillo 1991). Later, ZP proteins became to be commonly found throughout the animal phyla, and some of them were identified as egg coat proteins in invertebrates, namely: the urochordate *Ciona intestinalis* (Kürn et al. 2007) and the mollusk *H. corrugata* (Aagaard et al. 2006). ZP proteins have also been identified as other components of other structures in invertebrates (reviewed in Jovine et al. 2005), namely: the mucous house in urochordates (Thompson et al. 2001), cuticles in arthropods (DiBartolomeis et al. 2002; Roch et al. 2003) and

nematodes (Fujimoto and Kanaya 1973; Sebastiano et al. 1991), skeletons in cnidarians (Ramos-Silva et al. 2013; Takeuchi, Yamada, et al. 2016), and shells in mollusks (Marie et al., 2011, 2013). However, the function and evolutionary relationship of ZP proteins in invertebrates remain unclear.

The results of molecular phylogenetic analysis indicate that ZP proteins have diverse origins in lophotrochozoans. In bivalves, ZP proteins were classified into at least nine families (fig. 2A), and one of them was further divided into two subfamilies EGFZP and EGFL (fig. 2A). EGFZP proteins are conserved in other mollusks including gastropods, scaphopods, and cephalopods (fig. 2A). Although some ZP proteins with EGF-like domains exist in four non-molluscan lophotrochozoans (*C. teleta*, *L. anatina*, *N. geniculatus*, and *P. australis*), their origins greatly differ from the origin of molluscan EGFZP except for Cte\_210134 (*C. teleta*) (fig. 3B). Domain shuffling of EGF-like and ZP domains independently occurred many times in several taxa. Although it should be noted that our phylogenetic analysis is based on the genome or transcriptome data of a limited number of species, and the transcriptomes do not cover all mRNA encoded by a genome, our results suggested that EGFL proteins are found only in pteriomorph bivalves and that they are derived from EGFZP proteins. VEZP proteins are also conserved in mollusks but are absent in other lophotrochozoans. It is still unknown whether VE-related ZP proteins exist in other lophotrochozoans. The results of the phylogenetic analysis showed that VEZP evolved in the common ancestor of conchiferans or mollusks (fig. 2A; supplementary fig. S2, Supplementary Material online), and no homologous proteins were present in the other four non-molluscan lophotrochozoans (*C. teleta*, *L. anatina*, *N. geniculatus*, and *P. australis*) (fig. 3B; supplementary fig. S5, Supplementary Material online). In contrast, two ZP protein families (LZP1 and LZP2) were conserved in those lophotrochozoans (fig. 3B; supplementary fig. S4, Supplementary Material online). LZP1 proteins contain one or more SRCR domains, as well as the ZP domain. Many SRZP domain-containing proteins are membrane proteins that are known to be involved in the immune system (Bowdish and Gordon 2009). It can be said that the LZP1 family evolved only in lophotrochozoans among protostomes, because ZP proteins with SRCR domains were absent in ecdysozoans (the nematode *C. elegans* and arthropod *D. melanogaster*). The domain composition of LZP2 proteins is simple, being comprised of signal peptide, ZP domain, and TMD. This protein is probably distributed on the surface of cell membranes, but its function is still unknown.

### Possible Function of ZP Proteins in Biomineralization

Many types of matrix proteins have already been identified in the animal skeletons or shells. They likely play roles in different steps of the calcification processes, such as providing organic framework, crystal nucleation, crystal growth, and regulation of crystal polymorphs. In addition,

the mediation of interactions among those proteins would also be important to regulate shell mineralization, because the matrix proteins would often work as organic complexes that are formed by the framework-related proteins, crystal-related proteins, and others involved in accurate regulation of the formation of various biomineral structures. The ZP proteins are thought as being involved in extracellular protein–protein interactions and protein polymerization (Jovine et al. 2002, 2005, 2006). Most studies on ZP proteins focused on the sperm-egg interaction. However, the egg coat ZP proteins are only a part of the whole ZP proteins, and other ZP proteins play roles in other extracellular structures such as exoskeletons. In the corals *Acropora millepora* and *A. digitifera*, ZP proteins were identified as the skeletal OM proteins called SOMPs (Ramos-Silva et al. 2013; Takeuchi, Yamada, et al., 2016). Molluscan EGFZP proteins were also found in the shells of the limpet *L. gigantea* (Mann et al. 2012; Marie et al. 2013; Mann and Edsinger 2014) and the nautilus *Nautilus macromphalus* (Marie, Zanella-Cléon, Corneillat et al. 2011). These proteins possibly interact with other skeletons or SMPs. However, it remains unclear which SMPs actually interact with the ZP protein to form the organic complexes.

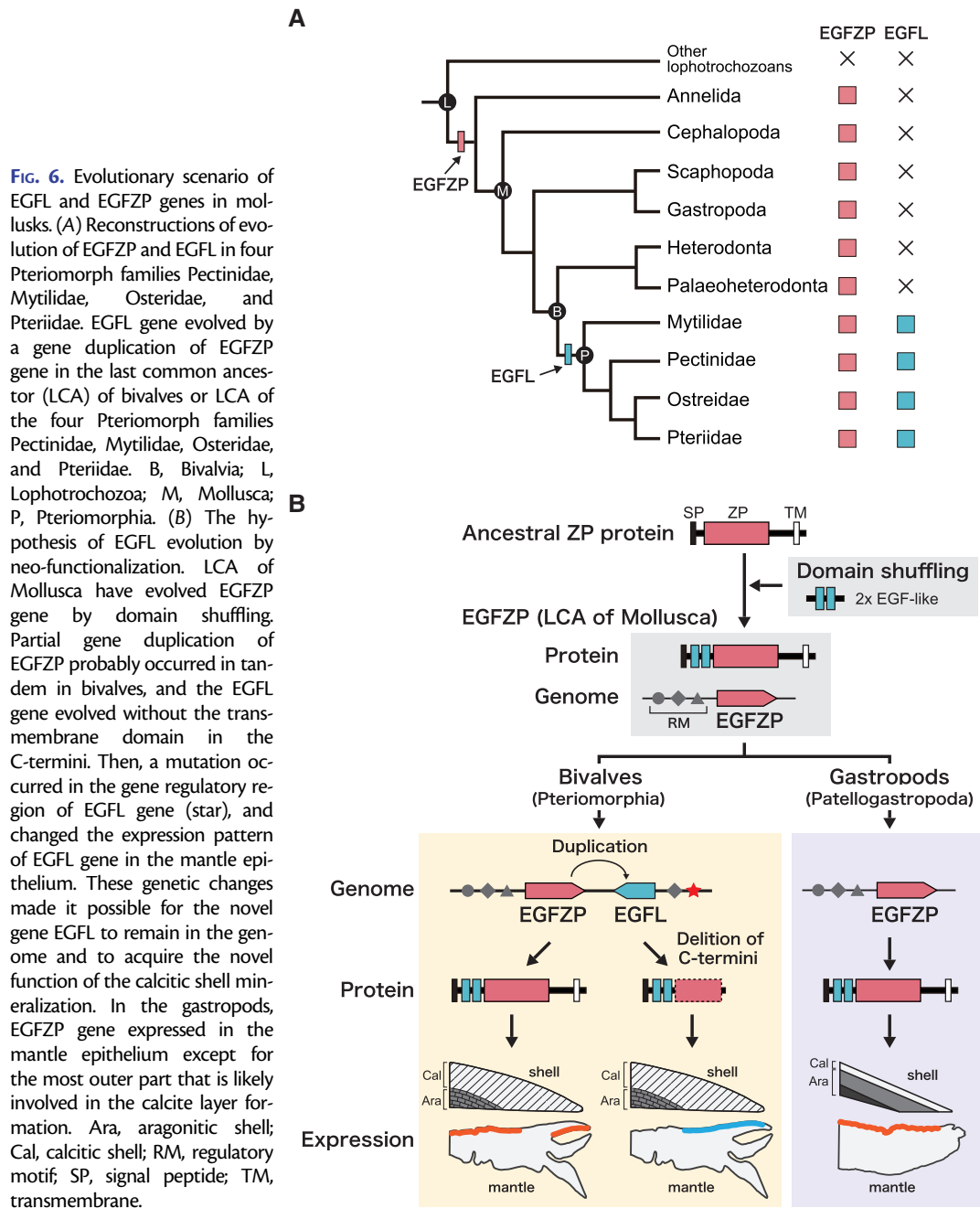
In our study, we isolated 20 proteins that specifically interact with the ZP domain in PfuEGFZP via a pull-down assay followed by proteomic identification. Eight of these proteins have already been reported as SMPs in *P. fucata* and are known to have various functions in shell formation; framework interaction (Shematin-8), calcium carbonate binding (BMSP100), and proteinase inhibition. These proteins are secreted to the extrapallial space and are interpreted to be bound by the EGFZP protein, which is anchored on the surface of the mantle epithelial cells by TMD. Those proteins around EGFZP probably form organic complexes, which are collectively involved in the shell formation. This function is possibly remained by EGFL proteins, which are inferred to have derived from EGFZP proteins because the former is nested within the cluster formed by the latter in the phylogenetic trees. Moreover, since EGFL proteins lose the anchor connecting to the cell membrane, they can be directly involved in shell mineralization. Furthermore, EGFZP proteins have a potential furin cleavage site (RRRR or RKRR) between the ZP domain and TMD (fig. 1), as previously identified in vertebrates ZP proteins (Jovine et al. 2002); however, most EGFZP proteins are likely to remain around the cell surfaces. Some EGFZP proteins have been identified as SMPs in the limpet *L. gigantea* and the nautilus *Nautilus pompilius* (Marie, Zanella-Cléon, Corneillat, et al. 2011, 2013), suggesting that they may be present in the extrapallial space due to furin digestion or other mechanisms. ZP proteins were also identified in the skeletons of corals (Ramos-Silva et al. 2013; Takeuchi, Yamada, et al., 2016). The ZP domain-containing SOMPs consist of a signal peptide, ZP domain, and TMD. The ZP domain-containing SOMPs in corals would also be able to interact with other coral SOMPs. It could be that the ancestral ZP proteins in the lineage

leading to bilaterians can mediate protein-protein interactions. However, the phylogenetic distributions clearly indicate that the mineralization-related ZP proteins evolved in corals, in mollusks, and in other phyla several times independently.

### Evolutionary Scenario of EGF-like Proteins in Mollusks

EGFL proteins have been identified as SMPs in bivalves (Marie, Zanella-Cléon, Corneillat, et al. 2011; Marie et al. 2012; Zhang et al. 2012; Liu et al. 2015; Gao et al. 2015; Liao et al. 2015), and all have been found in the calcitic shells of various shell microstructures (e.g. prismatic layer, fibrous prismatic layer, and chalky layer). In heterodonta and palaeoheterodonta bivalves, the homologues of EGFL were not found at least from the transcriptome databases (figs. 2 and 6). In gastropods, EGFL proteins have never been found in shells or genome databases (*L. gigantea* and *H. discus hannai*). Instead, EGFZP protein has been found in the shells of the limpet *L. gigantea* (LUSP-17) (Mann et al. 2012; Marie et al. 2013; Mann and Edsinger 2014). Homologs of EGFZP proteins were identified not only in gastropods but also in other mollusks, including bivalves, a scaphopod and cephalopods (figs. 2 and 6). The results of phylogenetic analysis suggested that the EGFL gene evolved from the EGFZP gene (fig. 2), and EGFZP gene evolved from ZP protein by domain shuffling (fig. 6). First, a reverse tandem duplication of the EGFZP gene occurred in the latest common ancestor of Pectinidae, Mytilidae, Osteridae, and Pteriidae, producing the EGFL1/2 gene (figs. 2, 5 and 6), and the region encoding the C-terminal transmembrane domain in the EGFL1/2 gene was deleted (figs. 1 and 6). Then, tandem duplication of the EGFL1/2 gene occurred in the latest common ancestor of Osteridae and Pteriidae, producing paralogs with a reversed gene order (figs. 2, 5, and 6). Finally, EGFL2 was independently duplicated in each of the families Pteriidae and Osteridae (figs. 2, 5, and 6).

Interestingly, EGFL proteins have been identified only in the calcitic parts of the bivalve shells, including the prismatic layer of *Pinctada* species (Marie et al. 2012; Liu et al. 2015; Zhao et al. 2018), the fibrous prismatic layer of *Mytilus* species (Gao et al. 2015; Liao et al. 2015), and the chalky layer of *C. gigas* (Iwamoto et al. 2020). The EGFL genes of *P. fucata* and the pen shell *Atrina pectinata* are expressed in epithelial cells of the outer pallium and the mantle edge, which are involved in prismatic layer formation (figs. 4 and 6; Shimizu et al. 2020). On the other hand, EGFZP protein has been identified from the aragonitic shells of the limpet *L. gigantea* (Mann et al. 2012; Marie et al. 2013; Mann and Edsinger 2014). The inner parts of the limpet shells consist of crossed lamellar layers made of aragonite crystals (Fuchigami and Sasaki 2005; Suzuki et al. 2010; Suzuki Kogure et al. 2011), and the outermost layer (M + 3) consists of a very thin mosaic microstructure made of calcite crystals (Suzuki et al. 2010). We found that the NfuEGFZP gene is expressed in the epithelial cells of



**FIG. 6.** Evolutionary scenario of EGFL and EGFZP genes in mollusks. (A) Reconstructions of evolution of EGFZP and EGFL in four Pteriormorph families Pectinidae, Mytilidae, Ostreidae, and Pteriidae. EGFL gene evolved by a gene duplication of EGFZP gene in the last common ancestor (LCA) of bivalves or LCA of the four Pteriormorph families Pectinidae, Mytilidae, Ostreidae, and Pteriidae. B, Bivalvia; L, Lophotrochozoa; M, Mollusca; P, Pteriormorphia. (B) The hypothesis of EGFL evolution by neo-functionalization. LCA of Mollusca have evolved EGFZP gene by domain shuffling. Partial gene duplication of EGFZP probably occurred in tandem in bivalves, and the EGFL gene evolved without the transmembrane domain in the C-termini. Then, a mutation occurred in the gene regulatory region of EGFL gene (star), and changed the expression pattern of EGFL gene in the mantle epithelium. These genetic changes made it possible for the novel gene EGFL to remain in the genome and to acquire the novel function of the calcitic shell mineralization. In the gastropods, EGFZP gene expressed in the mantle epithelium except for the most outer part that is likely involved in the calcite layer formation. Ara, aragonitic shell; Cal, calcitic shell; RM, regulatory motif; SP, signal peptide; TM, transmembrane.

the shell-facing side of the mantle tissue, except for the edge part, in another limpet *N. fuscoviridis* (figs. 4 and 6). The edge part of the mantle epithelial cells, where EGFZP is not expressed, is likely involved in the formation of the outermost calcite layer, while the remaining part of the mantle epithelial cells, where EGFZP is expressed, would be involved in the aragonitic shell (the crossed-lamellar layer) formation. In the pearl oyster *P. fucata*, EGFZP is expressed in the epithelial cells of the inner pallium, which is involved in the formation of the aragonite shell (the nacreous layer) (figs. 4 and 6) and in the inner surface of the outer fold, which is possibly involved in periostracum formation (figs. 4 and 6). Although further genome sequencing and gene expression analyses are needed, these observations

strongly suggest that a mutation occurred in the gene regulatory region of the nascent EGFL gene after gene duplication of the ancestral EGFZP gene, changing the spatial expression pattern of the EGFL gene to be expressed in the neighboring epithelial cells located in the mantle edge. These processes enabled a duplicated EGFL gene to acquire a novel function involved in the shell formation (neo-functionalization), allowing the EGFL gene to be maintained in the genome.

## Conclusion

In this study, we demonstrated the evolutionary relationship of the EGFL gene family and investigated its function in shell formation in mollusks. The EGFZP proteins,



commonly found in mollusks, seem to retain ancestral domain architecture and play a role in aragonitic crystallization. The EGFL gene family was expanded in the pteriomorph bivalve lineage by tandem gene duplication, and the genes acquired novel functions for calcitic shell formation, possibly due to changes in gene expression patterns in the mantle tissue and rapid molecular evolution in the ZP domain. The EGFL gene family represents a prominent example of neo-functionalization that is acquisition of a novel protein involves in the shell formation through gene duplication. This study represents the first study to show that certain SMPs evolved through neo-functionalization.

## Materials and Methods

### Identification of EGFL and EGFZP Proteins

We searched EGFL and EGFZP proteins from 12 lophotrochozoan genomes including five bivalves (*P. fucata*, *C. gigas*, *C. virginica*, *M. yessoensis*, and *P. maximus*) (Takeuchi, Koyanagi, et al. 2016; Zhang et al. 2012; Gómez-Chiarri et al. 2015; Wang et al. 2017; Kenny et al. 2020), two gastropods (*L. gigantea* and *H. discus hannai*) (Simakov et al. 2013; Nam et al. 2017), a cephalopod (*O. bimaculoides*), an annelid (*C. teleta*) (Simakov et al. 2013), a brachiopod (*L. anatina*) (Luo et al. 2015), a nemertean (*N. geniculatus*) (Luo et al. 2017), and a phoronid (*P. australis*) (Luo et al. 2017) and the transcriptome data of 7 bivalves (*A. pectinata*, *E. complanata*, *V. lienosa*, *Archivesica marissinica*, *Ruditapes philippinarum*, *M. edulis*, and *M. galloprovincialis*) (Bioproject Id: PRJDB9333, PRJNA194430, PRJNA75063, PRJNA471131, PRJNA664867, PRJNA525607, and PRJNA525609, respectively) (Wang et al. 2012; Ghiselli et al. 2012; Cornman et al. 2014; Björnmark et al. 2016; Knöbel et al. 2020; Shimizu et al. 2020; Ip et al. 2021) and a schaphopod (*Antalis entalis*) (Bioproject Id: PRJNA506080) using BLASTP program. These transcriptome data (assembled sequences) are available as Transcriptome Shotgun Assembly (TSA) in DDBJ/ENA/GenBank (ICPQ01000000, GAHW01000000, GAEH00000000, GIAS01000000, GHII01000000, and GHIK01000000). Query sequences were three EGFL proteins in the *P. fucata* transcriptome and genome (PfuEGFL1: pfu\_cdna2.0\_057745 [gene model id: pfu\_aug2.0\_2116.1\_21941], PfuEGFL2A: pfu\_cdna2.0\_061795 [gene model id: pfu\_aug2.0\_2116.1\_21942], and PfuEGFL2B: pfu\_cdna2.0\_061794 [gene model id: pfu\_aug2.0\_2116.1\_21943]) (Takeuchi, Koyanagi, et al. 2016) and two EGFZP proteins of the limpet *L. gigantea* (LgiLUSP-17 and LgiLUSP-24) (Marie et al. 2013) for EGFL and EGFZP searches, respectively (e-value < 1.0E-10). The domain organization of the protein sequences were identified using the online version of Simple Modular Architecture Research Tool (SMART; Letunic et al. 2015; Letunic and Bork 2018; <http://smart.embl-heidelberg.de>; last accessed October 2, 2020), including signal peptide prediction (SignalP; Petersen et al. 2011), Pfam domain search (Finn et al. 2016), transmembrane helices prediction (TMHMM; Krogh et al. 2001), and compositionally biased regions prediction (SEG; Wootton and

Federhen 1996) (e-value < 1.0e-5). We also conducted a comprehensive conserved domain search using InterProScan (ver.5.45-80.0) platform (Jones et al. 2014), including analyses with Pfam, database of protein domains, families and functional sites, and SMART against lophotrochozoan genome and transcriptome datasets, and identified the ZP domain-containing proteins. For molecular phylogenetic analysis, we used all the ZP domain-containing proteins that were identified from the lophotrochozoan genomes by InterProScan as well as the proteins similar to EGFZP and EGFL that were identified from the genome and transcriptome datasets by BLASTP. Sequence alignment of ZP domain regions of ZP proteins was conducted with the online version of MAFFT (v7.310; <http://mafft.cbrc.jp/alignment/server/index>, last accessed May 21, 2020; Katoh et al. 2002). The gap regions were trimmed by TrimAl (1.2rev59) (gap threshold set 0.7) (Capella-Gutiérrez et al. 2009), and remained 196 and 221 residues were used for molecular phylogenetic analysis of Molluskan ZP proteins (fig. 2) and Lophotrochozoan ZP proteins (fig. 3), respectively. The best-fit amino acid substitution model was searched using MEGA (v10.1.7) (Kumar et al. 2018). The maximum likelihood trees were constructed with the RAXML v8.2 (raxmlHPC-AVX-v8) (Kozlov et al. 2019) using WAG +  $\Gamma$  model with 1000 bootstrap replications.

### RNA Extraction, cDNA Synthesis, and Gene Cloning

Adult individuals of *P. fucata* were a gift from the Mie Prefecture Fisheries Research Institute (Mie, Japan), and those of *N. fuscoviridis* were collected in the rocky shore of Hiraiso (Hitachinaka, Ibaraki, Japan). Total RNA was extracted from the adult mantle tissues of *P. fucata* and *N. fuscoviridis* using Sepasol RNA I Super G (#09379-84, Nacalai Tesque Inc, Kyoto, Japan) according to the manufacturer's protocol. cDNA synthesis was conducted from 500 ng of total RNA using Prime Script RT reagent kit (#RR037A, Takara, Tokyo, Japan) according to the manufacturer's protocol. Partial sequences (around 700 bp) of the genes encoding EGF-like domain containing protein of *P. fucata* (Gene model IDs: pfu\_aug2.0\_2116.1\_21941, pfu\_aug2.0\_2116.1\_21942, and pfu\_aug2.0\_2116.1\_21944) were amplified with PCR using specific primers designed with reference to the genomic data of *P. fucata* (supplementary table S6, Supplementary Material online). In *N. fuscoviridis*, EGFZP gene was amplified with PCR using specific primers designed with reference to the genomic data of *L. gigantea* (Gene ID: 235548 [LgiLUSP-17]; supplementary table S6, Supplementary Material online). These PCR amplicons were purified by the QIAquick PCR Purification Kit (#28104, Qiagen, Hilden, Germany) and were ligated into the pGEM-T easy vectors using a DNA ligation kit (#A1360, Promega). The vectors were then transformed into the competent *Escherichia coli* BL21 cells. GenBank/EMBL/DDBJ accession numbers are as follows: LC582815 for PfuEGFL1 (pfu\_aug2.0\_2116.1\_21941), LC582816 for PfuEGFL2A (pfu\_aug2.0\_2116.1\_21942), LC582817 for PfuEGFZP (pfu\_aug2.0\_2116.1\_21944), and LC582814 for NfuEGFZP.



### Probe Synthesis and Section in Situ Hybridization

Antisense probes were synthesized using DIG RNA labeling Mix (#11277073910, Roche), 10 mM dithiothreitol (DTT), RNase ribonuclease inhibitor (#SIN201, Toyobo), T7 or SP6 RNA polymerase with 1X transcription buffer (#10881767001 or #10810274001, Roche), and purified PCR products (500 ng per reaction) according to manufacturer's protocol. Probe synthesis reactions were performed at 37°C for at least 3 hours and then were treated with DNase I (#M6101, Promega) at 37°C for 1 hour. Synthesized probes were purified using NucAway spin columns (#AM10070, Thermo Fisher Scientific) and stored at -20°C.

Adult mantle tissues of *P. fucata* and *N. fuscoviridis* were fixed with fixation buffer (4% paraformaldehyde, 0.5 M NaCl, 0.1 M MOPS, and 2 mM EGTA) overnight at 4°C. After washing with PBS, samples were dehydrated with 80% ethanol and stored in 80% ethanol at -20°C. In situ hybridization was performed as described previously (Shimizu et al. 2020).

### Extraction of SMPs from the Nacreous Layer

The prismatic layer that is an outer layer of the shell was completely removed by 30% sodium hypochlorite solution. The remained nacreous layer that is an inner layer of the shell was washed with distilled water and decalcified with 1 M of acetic acid. The acid soluble fraction was concentrated in the Amicon Ultra-15 centrifugation devices 10 kDa (UFC901024, Millipore) and desalted three times with distilled water, and the final product was used for pull-down assay as acid soluble matrices (ASM). The remained acid insoluble fraction was treated with extraction buffer (1% SDS, 10 mM of DTT, and 50 mM of Tris-HCl pH8.0) at 100°C for 10 min. After centrifugation, the aqueous part was concentrated in the Amicon Ultra-15 centrifugation devices 10 kDa (UFC901024, Millipore) and washed three times with distilled water, and the final product was used for pull-down assay as acid insoluble matrices (AIM).

### Preparation of Recombinant Proteins

The sequence of ZP domain of PfEGFZP (PfuZP) was amplified by PCR using specific primers (supplementary table S6, Supplementary Material online) and was inserted into pET-44a vector using InFusion HD Cloning kit (#Z9648N, Takara). The vectors, pET-44a with and without PfuZP, were transformed to BL21 (DE3) competent cells. The expressions of target proteins were induced by 1 mM of IPTG treatment and cultured on Luria-Bertani broth at 20°C for 24 h. They were centrifuged at 4,000 g at 4°C for 15 min and were removed their supernatants. The pellets were homogenized into 1X PBS on ice using ultrasonic homogenizer (IKA U200S control, IKA Japan, Osaka, Japan). After centrifuge at 4,000 g at 4°C for 15 min, we collected the soluble fraction, and they were applied SDS-PAGE to confirm these proteins presence by CBB staining. These proteins were then purified using Ni-column (#17531801, Ni Sepharose 6 Fast Flow, GE Healthcare, CHI, USA) according to the manufacture's

protocol and confirmed purified proteins using SDS-PAGE and CBB staining. These purified samples were concentrated in the Amicon Ultra-15 centrifugation devices 10 kDa (UFC901024, Millipore, Billerica, CA, USA) and removed imidazole using wash buffer (0.5 M NaCl, 20 mM of phosphate buffer pH 7.5).

### Pull-down Assay and Peptide Analysis

The r-PfuZP protein and tag-only protein (5 µg each) were bind to Ni-columns (#17531801, GE Healthcare), and each column was washed three times with 20 mM of imidazole in wash buffer (0.5 M NaCl, 20 mM of phosphate buffer pH 7.5). 50 µg of ASM or AIM that were extracted from the nacreous layer were incubated with these columns at 4°C for 18 h and were washed five times with 20 mM of imidazole in wash buffer. After washed with five times with 50 mM of imidazole in wash buffer, the binding proteins were eluted with 500 mM of imidazole in wash buffer. The binding protein solutions were concentrated in the Vivaspin 500-10K (#VS0101, Sartorius, Göttingen, German) and removed imidazole using wash buffer (0.5 M NaCl, 20 mM of phosphate buffer pH 7.5).

After freeze-drying, samples were dissolved with 100 µL of Alkylation buffer (7 M Guanidine Hydrochloride, 0.5 M Tris-HCl pH 8, 10 mM EDTA pH 8) and added 1 µL of 0.5 M DTT. After incubating for 30 min at 60°C, we added 2 µL of 0.5 M Iodoacetamide, then mixed by vortexing and incubated for 1 h at room temperature (RT) in dark. MeOH (400 µL), Chloroform (100 µL), and distilled water (300 µL) were added and mixed by vortexing one-by-one and samples were centrifuged for 5 min at 14,000 rpm at 4°C. After removing the supernatant, 300 µL of methanol were added on ice and mixed by inverting gently, and samples were centrifuged for 3 min at 14,000 rpm at 4°C. Supernatant were removed and added 200 µL of 70% EtOH. After centrifuging for 3 min at 14,000 rpm at 4°C, supernatant were removed and solved with 40 µL of 33 mM NH<sub>4</sub>HCO<sub>3</sub> (pH8.0). Samples were treated with 5 µL of trypsin solution (100 ng/µL of Trypsin Gold [V528A, Promega, WI, USA] in 50 mM NH<sub>4</sub>HCO<sub>3</sub> [pH8.0]) at 37°C for 18 h. After adding 5 µL 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 2.1 and the protein database from the predicted transcripts for genome assembly ver 2.0 of *P. fucata* (pfu\_aug2.0.AA.fasta, Takeuchi, Koyanagi, et al. 2016).

### Acknowledgments

We thank two professional editors, who are both English native speakers in Editage ([www.editage.com](http://www.editage.com)), for editing English in this manuscript. This research was partially funded by Grant-in-Aid for Scientific Research B (JP19H03045), Grant-in-Aid for Scientific Research on Innovative Areas IBmS: JSPS KAKENHI (JP19H05771), Sasakawa Scientific

Research Grant from the Japan Science Society (2020-4040), Platform Project for Supporting Drug Discovery and Life Science Research (BINDS) from AMED JP21am0101076 to H.K, Environmental Restoration and Conservation Agency (1CN-2201), and New Energy and Industrial Technology Development Organization (NEDO).

## Author Contributions

K.S., T.T., K.E., and M.S. conceived and designed the experiments. K.S., T.T., L.N., and H.K. performed experiments and analyzed data. I.K. and M.S. contributed reagents and materials. K.S., T.T., K.E., and M.S. wrote the manuscript.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Data Availability

The data underlying this article are available in the GenBank/EMBL/DDBJ database at <https://www.ddbj.nig.ac.jp/index-e.html>, and can be accessed with accession numbers LC582814-LC582817.

## References

- Aagaard JE, Yi X, MacCoss MJ, Swanson WJ. 2006. Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. *Proc Natl Acad Sci USA* **103**: 17302–17307. doi:10.1073/pnas.0603125103
- Aguilera F, McDougall C, Degnan BM. 2014. Evolution of the tyrosinase gene family in bivalve molluscs: independent expansion of the mantle gene repertoire. *Acta Biomater* **10**:3855–3865. doi: 10.1016/j.actbio.2014.03.031
- Aguilera F, McDougall C, Degnan BM. 2017. Co-option and de novo gene evolution underlie molluscan shell diversity. *Mol Biol Evol* **34**(4):779–792. doi:10.1093/molbev/msw294
- Björnmark NA, Yarra T, Churcher AM, Felix RC, Clark MS, Power DM. 2016. Transcriptomics provides insight into *Mytilus galloprovincialis* (Mollusca: Bivalvia) mantle function and its role in biomineralisation. *Mar Genomics* **27**:37–45. doi: 10.1016/j.margen.2016.03.004
- Bork P, Sander C. 1992. A large domain common to sperm receptors (Zp2 and Zp3) and TGF-beta type III receptor. *FEBS Lett* **300**: 237–240.
- Bowdish DM, Gordon S. 2009. Conserved domains of the class A scavenger receptors: evolution and function. *Immunol Rev* **227**:19–31. doi:10.1111/j.1600-065X.2008.00728.x
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973. doi:10.1093/bioinformatics/btp348
- Carter JG. 1990. Evolutionary significance of shell microstructure in the Palaeotaxodonta, Pteriomorphia and Isofilibranchia (Bivalvia: Mollusca). In: Carter JG, editors. *Skeletal biomineralization: patterns, processes and evolutionary trends*. New York: Van Nostrand Rein- hold, Chapter 10. p. 135–296.
- Cölfen H, Mann S. 2003. Higher-order organization by mesoscale self-assembly and transformation of hybrid nanostructures. *Angew Chem* **42**:2350–2365. doi:10.1002/anie.200200562
- Cölfen H, Antonietti M. 2005. Mesocrystals: inorganic superstructures made by highly parallel crystallization and controlled alignment. *Angew Chem* **44**:5576–5591. doi:10.1002/anie.200500496
- Combosch DJ, Collins TM, Glover EA, Graf DL, Harper EM, Healy JM, Kawauchi GY, Lemer S, McIntyre E, Strong EE, et al. 2017. A family-level tree of life for bivalves based on a Sanger-sequencing approach. *Mol Phylogenet Evol* **107**: 191–208. doi: 10.1016/j.ymp.2016.11.003
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* **9**:938–950. doi: 10.1038/nrg2482
- Cornman RS, Robertson LS, Galbraith H, Blakeslee C. 2014. Transcriptomic analysis of the mussel *Elliptio complanata* identifies candidate stress- response genes and an abundance of novel or noncoding transcripts. *PLoS ONE* **9**(11):e112420. doi:10.1371/journal.pone.0112420
- DiBartolomeis SM, Akten B, Genova G, Roberts MA, Jackson FR. 2002. Molecular analysis of the *Drosophila* miniature-dusky (m-dy) gene complex: m-dy mRNAs encode transmembrane proteins with similarity to *C. elegans* cuticulin. *Mol Genet Genom* **267**:564–576. doi:10.1007/s00438-002-0700-7
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**:D279–D285.
- Fuchigami T, Sasaki T. 2005. The shell structure of the recent Patellogastropoda (Mollusca: Gastropoda). *Paleontol Res* **9**:143–168.
- Fujimoto D, Kanaya S. 1973. Cuticlin: a noncollagen structural protein from *Ascaris* cuticle. *Arch Biochem Biophys* **157**:1–6. doi: 10.1016/0003-9861(73)90382-2
- Gao P, Liao Z, Wang XX, Bao LF, Fan MH, Li X-m, Wu C-w, Xia S-w, Lin S. 2015. Layer-by-layer proteomic analysis of *Mytilus galloprovincialis* shell. *PLoS ONE* **10**:e0133913. doi:10.1371/journal.pone.0133913.
- Gardner LD, Mills D, Wiegand A, Leavesley D, Elizur A. 2011. Spatial analysis of biomineralization associated gene expression from the mantle organ of the pearl oyster *Pinctada maxima*. *BMC Genomics* **12**:455. doi:10.1186/1471-2164-12-455
- Ghiselli F, Milani L, Chang PL, Hedgecock D, Davis JP, Nuzhdin SV, Passamonti M. 2012. De Novo assembly of the Manila clam *Ruditapes philippinarum* transcriptome provides new insights into expression bias, mitochondrial doubly uniparental inheritance and sex determination. *Mol Biol Evol* **29**(2):771–86. doi: 10.1093/molbev/msr248
- Gómez-Chiarri M, Warren WC, Guo X, Proestou D. 2015. Developing tools for the study of molluscan immunity: the sequencing of the genome of the eastern oyster, *Crassostrea virginica*. *Fish Shellfish Immunol* **46**:2–4. doi:10.1016/j.fsi.2015.05.004
- Ip JC, Xu T, Sun J, Li R, Chen C, Lan Y, Han Z, Zhang H, Wei J, Wang H, et al. 2021. Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. *Mol Biol Evol* **38**(2):502–518. doi: 10.1093/molbev/msaa241
- Ishikawa A, Shimizu K, Isowa Y, Takeuchi T, Zhao R, Kito K, Fujie M, Satoh N, Endo K. 2020. Functional shell matrix proteins tentatively identified by asymmetric snail shell morphology. *Sci Rep* **10**:9768. doi:10.1038/s41598-020-66021-w
- Isowa Y, Sarashina I, Setiamarga DHE, Endo K. 2012. A comparative study of the shell matrix protein aspein in pteroid bivalves. *J Mol Evol* **75**:11–18. doi:10.1007/s00239-012-9514-3
- Iwamoto S, Shimizu K, Negishi L, Suzuki N, Nagata K, Suzuki M. 2020. Characterization of the chalky layer-derived EGF-like domain-containing protein (CgELC) in the pacific oyster, *Crassostrea gigas*. *J Struct Biol* **212**:107594. doi:10.1016/j.jsb.2020.107594
- 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 Biol Evol* **7**:1349–1362. doi:10.1093/gbe/evv074
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5:

- genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240. doi: 10.1093/bioinformatics/btu031
- Jovine L, Darie CC, Litscher ES, Wassarman PM. 2005. Zona pellucida domain proteins. *Annu Rev Biochem*. **74**:83–114. doi: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 polymerisation. *BMC Biochem*. **7**(1):11. doi: 10.1186/1471-2091-7-11
- 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. doi:10.1038/ncb802
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. **20**:1313–1326. doi: 10.1101/gr.101386.109
- Katoh K, Misawa K, Ki K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. **30**:3059–3066. doi:10.1093/nar/gkf436
- Kenny NJ, McCarthy SA, Dudchenko O, James K, Betteridge E, Corton C, Dolucan J, Mead D, Oliver K, Omer AD, et al. 2020. The gene-rich genome of the scallop *Pecten maximus*. *Gigascience* **9**:giaa037. doi:10.1093/gigascience/giaa037
- Knöbel L, Breusing C, Bayer T, Sharma V, Hiller M, Melzner F, Stuckas H. 2020. Comparative de novo assembly and annotation of mantle tissue transcriptomes from the *Mytilus edulis* species complex (*M. edulis*, *M. galloprovincialis*, *M. trossulus*). *Marine Genomics* **51**: 100700. <https://doi.org/10.1016/j.margen.2019.100700>
- 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. doi: 10.1186/s12983-016-0155-z
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAXML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**:4453–4455. doi:10.1093/bioinformatics/btz305
- Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. **305**(3):567–580.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. **35**:1547–1549.
- Kürn U, Sommer F, Bosch TC, Khalturin K. 2007. In the urochordate *Ciona intestinalis* zona pellucida domain proteins vary among individuals. *Dev Comp Immunol*. **31**:1242–1254. doi:10.1016/j.dci.2007.03.011
- Lemer S, Bieler R, Giribet G. 2019. Resolving the relationships of clams and cockles: dense transcriptome sampling drastically improves the bivalve tree of life. *Proc Biol Sci*. **286**:20182684. doi: 10.1098/rspb.2018.2684
- Le Roy N, Jackson DJ, Marie B, Ramos-Silva P, Marin F. 2015. The evolution of metazoan  $\alpha$ -carbonic anhydrases and their roles in calcium carbonate biomineralization. *Front Zool*. **11**:75.
- 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**:D257–D260.
- Liao Z, Bao LF, Fan MH, Gao P, Wang XX, Qin C-l, Li X-m. 2015. In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *J. Proteomics*. **122**:26–40. doi:10.1016/j.jprot.2015.03.027
- Liu C, Li S, Kong J, Wang T, Liping XL, Zhang R. 2015. In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Sci Rep*. **5**:17269. doi:10.1038/srep17269
- Lowenstam HA, Weiner S. 1989. *On Biomineralization*. London: Oxford University Press. p. 7–23.
- 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. doi:10.1038/s41559-017-0389-y
- Luo YJ, Takeuchi T, Koyanagi R, Yamada L, Kanda M, Khalturina M, Fujie M, Yamasaki S, Endo K, Satoh N. 2015. The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat Commun*. **6**:8301. doi:10.1038/ncomms9301
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**:459–473.
- 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. doi:10.1186/1477-5956-10-28
- Mann K, Edsinger E. 2014. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. *Proteome Sci*. **2014**(12):28. doi:10.1186/1477-5956-12-28
- Mann K, Jackson DJ. 2014. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genom*. **15**:249. doi:10.1186/1471-2164-15-249
- Mann K, Macek B, Olsen JV. 2006. Proteomic analysis of the acid-soluble organic matrix of the chicken calcified eggshell layer. *Proteomics* **6**:3801–3810. doi:10.1002/pmic.200600120
- Mann K, Wilt FH, Poustka AJ. 2010. Proteomic analysis of sea urchin (*Strongylocentrotus purpuratus*) spicule matrix. *Proteome Sci*. **8**: 33. doi:10.1186/1477-5956-8-33
- Marie B, Arivalagan J, Mathéron L, Bolbach G, Berland S, Marie A, Marin F. 2017. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the Unionoida *Elliptio complanata* and *Villosa lienosa*. *J R Soc Interface*. **14**. doi:10.1098/rsif.2016.0846
- 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. *FEBS J*. **280**:214–232. doi:10.1111/febs.12062
- 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. *Proc Natl Acad Sci USA* **109**:20986–20991. doi:10.1073/pnas.1210552109
- Marie P, Labas V, Brionne A, Harichaux G, Hennequet-Antier C, Nys Y, Gautron J. 2015. Quantitative proteomics and bioinformatic analysis provide new insight into protein function during avian eggshell biomineralization. *J Proteomics* **113**:178–193. doi: 10.1016/j.jprot.2014.09.024
- Marie B, Zanella-Cléon I, Corneillat M, Becchi M, Alcaraz G, Plasseraud L, Luquet G, Marin F. 2011. Nautilin-63, a novel acidic glycoprotein from the shell nacre of *Nautilus macromphalus*. *FEBS J*. **278**(12):2117–30. doi: 10.1111/j.1742-4658.2011.08129.x
- Marie B, Zanella-Cléon I, Guichard N, Becchi M, Marin F. 2011. Novel proteins from the calcifying shell matrix of the Pacific oyster *Crassostrea gigas*. *Mar Biotechnol*. **13**:1159–1168. doi:10.1007/s10126-011-9379-2
- Marin F, Le Roy N, Marie B. 2012. The formation and mineralization of mollusk shell. *Front Biosci*. **1**(4):1099–125. doi: 10.2741/s321
- Marin F, Luquet G, Marie B, Medakovic D. 2008. Molluscan shell proteins: primary structure, origin, and evolution. *Curr Top Dev Biol*. **80**:209–276. doi:10.1016/S0070-2153(07)80006-8
- Maurer P, Hohebest E. 1997. Structural and functional aspects of calcium binding in extracellular matrix proteins. *Matrix Biol*. **15**:569–580. doi: 10.1016/s0945-053x(97)90033-0
- McDougall C, Aguilera F, Degnan BM. 2013. Rapid evolution of pearl oyster shell matrix proteins with repetitive, low-complexity domains. *J R Soc Interface* **10**(82):20130041. doi:10.1098/rsif.2013.0041
- Murdock DJE. 2020. The “biomineralization toolkit” and the origin of animal skeletons. *Biol Rev Camb Philos Soc*. **95**:1372–1392. doi: 10.1111/brv.12614
- Nam BH, Kwak W, Kim YO, Kim DG, Kong HJ, Kim W-J, Kang J-H, Park JY, An CM, Moon J-Y, et al. 2017. Genome sequence of pacific abalone (*Haliotis discus hannai*): the first draft genome in



- family Haliotidae. *Gigascience* **6**:1–8. doi:10.1093/gigascience/gix014
- Nesculea A, Kaessmann H. 2014. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet.* **15**:734–748. doi:10.1038/nrg3802
- Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**(10):785–786.
- Ramos-Silva P, Kaandorp J, Huisman L, Marie B, Zanella-Cléon I, Guichard N, Miller DJ, Marin F. 2013. The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuf-fling. *Mol Biol Evol.* **30**(9):2099–2112.
- Roch F, Alonso CR, Akam M. 2003. Drosophila miniature and dusky encode ZP proteins required for cytoskeletal reorganisation during wing morphogenesis. *J Cell Sci.* **116**:1199–1207. doi:10.1242/jcs.00298
- 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. *J Mol Evol.* **62**:307–18.
- Sebastiano M, Lassandro F, Bazzicalupo P. 1991. cut-1 a *Caenorhabditis elegans* gene coding for a dauer-specific noncollagenous component of the cuticle. *Dev Biol.* **146**:519–530. doi:10.1016/0012-1606(91)90253-y
- Shimizu K, Kimura K, Isowa Y, Oshima K, Ishikawa M, Kagi H, Kito K, Hattori M, Chiba S, Endo K. 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. doi:10.1007/s00239-020-09977-7
- Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T, Lv J, Arendt D, et al. 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**:526–531. doi:10.1038/nature11696
- Song RQ, Cölfen H. 2010. Mesocrystals-ordered nanoparticle superstructures. *Adv Mater.* **22**:1301–1330.
- Song X, Liu Z, Wang L, Song L. 2019. Recent advances of shell matrix proteins and cellular orchestration in marine molluscan shell biomineralization. *Front Mar Sci.* **6**:41. doi:10.3389/fmars.2019.00041
- Stapane L, Le Roy N, Hincke MT, Gautron J. 2019. The glycoproteins EDIL3 and MFG8 regulate vesicle-mediated eggshell calcification in a new model for avian biomineralization. *J Biol Chem.* **294**:14526–14545. doi:10.1074/jbc.RA119.009799
- 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. doi:10.1007/s10126-012-9471-2
- Suzuki M, Iwashima A, Tsutsui N, Ohira T, Kogure T, Nagasawa H. 2011a. Identification and characterisation of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from the nacreous layer. *Chembiochem* **12**(16):2478–2487. doi:10.1002/cbic.201100317
- Suzuki M, Kameda J, Sasaki T, Saruwatari K, Nagasawa H, Kogure T. 2010. Characterization of the multilayered shell of a limpet, *Lottia kogamogai* (Mollusca: Patellogastropoda), using SEM-EBSD and FIB-TEM techniques. *J Struct Biol.* **171**:223–230. doi:10.1016/j.jssb.2010.04.008
- Suzuki M, Kogure T, Nagasawa H. 2017. Studies on the Chemical Structures of Organic Matrices and Their Functions in the Biomineralization Processes of Molluscan Shells. *AGRI-Biosci Monogr.* **7**:25–39. doi:10.5047/agbm.2017.00702.0025
- Suzuki M, Kogure T, Weiner S, Addadi L. 2011b. Formation of Aragonite Crystals in the Crossed Lamellar Microstructure of Limpet Shells. *Cryst Growth Des.* **11**(11):4850–4859. doi:10.1021/cg2010997
- Suzuki M, Nagasawa H. 2013. Mollusk shell structures and their formation mechanism. *Can J Zool.* **91**:349–366. doi:10.1139/cjz-2012-0333
- 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* **11**:1388–1390. doi:10.1126/science.1173793
- Takeuchi T, Endo K. 2006. Biphasic and dually coordinated expression of the genes encoding major shell matrix proteins in the pearl oyster *Pinctada fucata*. *Mar Biotechnol (NY)* **8**:52–61. doi:10.1007/s10126-005-5037-x
- Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M, Goto H, Yamasaki S, Nagai K, Morino Y, et al. 2016a. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Lett.* **2**:3. doi:10.1186/s40851-016-0039-2
- Takeuchi T, Sarashina I, Iijima M, Endo K. 2008. In vitro regulation of CaCO<sub>3</sub> crystal polymorphism by the highly acidic molluscan shell protein Aspein. *FEBS Lett.* **582**:591–596. doi:10.1016/j.febslet.2008.01.026
- 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. doi:10.1371/journal.pone.0156424
- Thompson EM, Kallesøe T, Spada F. 2001. Diverse genes expressed in distinct regions of the trunk epithelium define a monolayer cellular template for construction of the oikopleurid house. *Dev Biol.* **238**:260–273. doi:10.1006/dbio.2001.0414
- Tsukamoto D, Sarashina I, Endo K. 2004. Structure and expression of an unusually acidic matrix protein of pearl oyster shells. *Biochem Biophys Res Commun.* **320**:1175–1180. doi:10.1016/j.bbrc.2004.06.072
- Wang R, Li C, Stoeckel J, Moyer G, Liu Z, Peatman E. 2012. Rapid development of molecular resources for a freshwater mussel, *Villosa liososa* (Bivalvia:Unionidae), using an RNA-seq-based approach. *Freshwater Sci.* **31**(3):695–708. doi:10.1899/11-149.1
- Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, et al. 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol.* **1**:120. doi:10.1038/s41559-017-0120
- Wassarman PM, Jovine L, Litscher ES. 2001. A profile of fertilization in mammals. *Nat Cell Biol.* **3**:E59–E64. doi:10.1038/35055178
- Wassarman PM, Mortillo S. 1991. Structure of the mouse egg extracellular coat, the zona pellucida. *Int Rev Cytol.* **130**:85–110. doi:10.1016/s0074-7696(08)61502-8
- Wootton JC, Federhen S. 1996. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **266**:554–571.
- 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*. *Comp Biochem Physiol B Biochem Mol Biol.* **144**:254–262. doi:10.1016/j.cbpb.2006.03.004
- Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**:49–54. doi:10.1038/nature11413
- Zhao R, Takeuchi T, Luo YJ, Ishikawa A, Kobayashi T, Koyanagi R, Villar-Briones A, Yamada L, Sawada H, Iwanaga S, et al. 2018. Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation. *Mol Biol Evol.* **35**:2751–2761. doi:10.1093/molbev/msy172