



**HAL**  
open science

## **Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas.**

Ines Herlitze, Benjamin Marie, Frédéric Marin, Daniel J. Jackson

### ► **To cite this version:**

Ines Herlitze, Benjamin Marie, Frédéric Marin, Daniel J. Jackson. Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas.. *GigaScience*, 2018, 7 (6), pp.giy056. <10.1093/giga-science/giy056>. <hal-01821730>

**HAL Id: hal-01821730**

**<https://hal.science/hal-01821730v1>**

Submitted on 12 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## RESEARCH

# Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas

Ines Herlitze<sup>1</sup>, Benjamin Marie<sup>2</sup>, Frédéric Marin<sup>3</sup> and Daniel J. Jackson <sup>1,\*</sup>

<sup>1</sup>Department of Geobiology, Georg-August University of Göttingen, Goldschmidtstrasse 3, 37077 Göttingen, Germany, <sup>2</sup>UMR 7245 MNHN/CNRS Molécules de Communication et Adaptation des Micro-organismes, Département Aviv, Sorbonne Universités, Muséum National d'Histoire Naturelle, CP 39, 12 Rue Buffon, 75005 Paris, France and <sup>3</sup>UMR CNRS 6282 Biogéosciences, Université de Bourgogne - Franche-Comté, 6 Boulevard Gabriel, 21000 Dijon, France

\*Correspondence address. Daniel J. Jackson, Department of Geobiology, Georg-August University of Göttingen, Goldschmidtstrasse 3, 37077, Göttingen, Germany. Tel: +49 (0) 55139 14177; E-mail: [djackso@uni-goettingen.de](mailto:djackso@uni-goettingen.de)  <http://orcid.org/0000-0001-9045-381X>

## Abstract

**Background:** Conchiferan molluscs construct a biocalcified shell that likely supported much of their evolutionary success. However, beyond broad proteomic and transcriptomic surveys of molluscan shells and the shell-forming mantle tissue, little is known of the spatial and ontogenetic regulation of shell fabrication. In addition, most efforts have been focused on species that deposit nacre, which is at odds with the majority of conchiferan species that fabricate shells using a crossed-lamellar microstructure, *sensu lato*. **Results:** By combining proteomic and transcriptomic sequencing with *in situ* hybridization we have identified a suite of gene products associated with the production of the crossed-lamellar shell in *Lymnaea stagnalis*. With this spatial expression data we are able to generate novel hypotheses of how the adult mantle tissue coordinates the deposition of the calcified shell. These hypotheses include functional roles for unusual and otherwise difficult-to-study proteins such as those containing repetitive low-complexity domains. The spatial expression readouts of shell-forming genes also reveal cryptic patterns of asymmetry and modularity in the shell-forming cells of larvae and adult mantle tissue. **Conclusions:** This molecular modularity of the shell-forming mantle tissue hints at intimate associations between structure, function, and evolvability and may provide an elegant explanation for the evolutionary success of the second largest phylum among the Metazoa.

**Keywords:** mollusc; biomineralization; gene expression; asymmetry; modularity; evolution; shell; matrix protein; transcriptome; alternative splicing

## Introduction

Due to its evolutionary significance, impressive materials properties, and aesthetic beauty, the molluscan shell has long received attention from a wide variety of scientific disciplines [1–6]. Although molluscan shells are constructed from a complex mixture of calcium carbonate, carbohydrates [7, 8], and lipids [9], proteins have received the most attention arguably for two main reasons: they can provide deep insight into the evolutionary history of this composite structure and the techniques for the high-

throughput study of these molecules are well established and are technically straightforward. Much progress has been made in identifying the components of the shell-forming proteome from a variety of gastropod and (primarily) bivalve species (e.g., [10–17]) largely due to advances in nucleic acid sequencing technologies that, when coupled with high-throughput proteomic surveys of the biomineralized proteome, allow for the rapid generation of extensive lists of shell-associated proteins. However, without further validation, genes identified in this way should only be considered as candidate biomineralizing molecules. This

Received: 27 November 2017; Revised: 20 March 2018; Accepted: 9 May 2018

© The Author(s) 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

problem is often compounded by the fact that these proteins often share little to no sequence similarity with proteins from conventional model organisms, making any inference about their function very difficult. This bottleneck represents one of the current major challenges for scientists interested in understanding the mechanisms and evolution of molluscan biomineral formation. While knock-down of individual shell-forming genes via RNAi has been reported in some species of bivalves [10, 18, 19], these assays are rarely validated by protein immuno-detection, and levels of penetrance or statistical quantitation of knock-down phenotypes are rarely reported.

Another approach to gain insight into the function of shell-forming genes is to characterize their spatial expression patterns *in situ*. We previously adopted this approach in the tropical abalone *Haliotis asinina* with a Sanger Expressed Sequence Tag (EST) dataset and characterized the spatial expression patterns of more than 20 putative shell-forming genes in juvenile snails [16]. While a spatial expression pattern in the mantle is not direct evidence of a functional role in calcification, we were able to assign putative functions to genes involved in shell pigmentation [16] and ecological and mineralogical transitions [20]. Here, we have combined a next-generation sequencing (NGS) transcriptome analysis of adult mantle tissue with a proteomic survey of the adult shell of the freshwater pulmonate gastropod *Lymnaea stagnalis* in order to both compare the resulting data with other similar datasets and to generate the first *in situ*-validated ontogenetic transcriptome-scale dataset for a species that forms the most common molluscan shell microstructure, crossed lamellar [21–23]. The high-order structure of crossed-lamella, which allows it to efficiently deflect and arrest cracks [24–27], coupled with its extremely low organic content (typically <0.5%) has been suggested to be one reason it has enjoyed so much evolutionary success (reviewed in [28]). Recent proteomic studies have been reported for molluscs that build crossed-lamellae shells (*Helix aspersa maxima* [29] and *Cepaea nemoralis* [14]), however, those studies did not conduct any spatial expression analyses for the shell-forming proteins they identified. In addition to characterizing the spatial expression patterns of more than 30 shell-forming candidates in the adult mantle tissue of *L. stagnalis*, we have also investigated their spatial expression patterns during development.

Our analyses hint at the potential pleiotropic nature of some of these shell-forming genes and highlight the dynamic and asymmetric natures of their spatial regulation. A striking result of our analyses in the adult mantle tissue is the degree of spatial modularity displayed by distinct sets of genes. This general observation may contribute to an explanation of why the molluscan shell is apparently so evolvable. With the availability of a draft *L. stagnalis* genome and transcriptome data from a variety of adult tissues, we have also investigated the genetic architectures of our biomineralization candidates and explored to what extent alternative splicing plays a role in shell formation in *L. stagnalis*. These genes can also be compared with similar datasets from distantly related molluscs that build shells with alternative polymorphs of calcium carbonate (calcite vs. aragonite) and textures (prismatic vs. nacreous vs. crossed lamellae). Such comparisons can generate testable hypotheses regarding which components of the shell-forming toolkit contribute to these differences and which components are required for more fundamental aspects of shell formation.

## Methods

### Cultivation of adult *L. stagnalis*

*Lymnaea stagnalis* (Mollusca; Gastropoda; Heterobranchia; Euthyneura; Panpulmonata; Hygrophila; Lymnaeidae; Lymnaeidae; Lymnaea) does not fall under the German Animal Protection Act §8 and is listed as “least concern” under the International Union for Conservation of Nature (IUCN’s) list of threatened species. This work was therefore exempt from the University of Göttingen Ethics Committee. Adult specimens of *L. stagnalis* derived from animals originally collected from the Northeimer Seenplatte, Germany (51° 43′ 26.5368′, 9° 57′ 24.75′), and from a pond on the north campus of the University of Göttingen, Germany (51° 33′ 23.727′, 9° 57′ 25.617′), were kept in a stand-alone V30 unit (Aqua Schwarz) in demineralized water supplemented with ReMineral+ (Dennerle, 7036) to a conductivity of 200–220  $\mu$ S, 23°C, a pH of 7.5 to 7.9 and a 16:8 light regimen. Five to 10 individuals were kept in 3- or 5-liter boxes under a constant and low-flow rate. Snails were fed *ad libitum* with lettuce and a variety of other vegetables. Under this regime, adult snails lay egg masses year round.

### Organic matrix extraction from calcified shells

Twelve shells of adult *L. stagnalis* (larger than 3–4 cm in length) were selected for extraction. Prior to further treatment, the columella was delicately cut and removed from each shell. Superficial organic contaminants were removed by incubating pooled shell fragments in 10% v/v sodium hypochlorite (NaOCl) for 24 hours. Fragments were then thoroughly rinsed with water and subsequently ground into a fine powder that was sieved (>200  $\mu$ M). This biomineral powder was incubated in 5% v/v NaOCl for 5 hours and rinsed twice with MilliQ water. Powdered samples were decalcified overnight at 4°C in cold 5% v/v acetic acid that was slowly added by an automated burette (Titronic Universal, Mainz, Germany) at a flow rate of 100  $\mu$ L every 5 seconds. The solution (final pH ~4.2) was centrifuged at 3,900 g for 30 minutes. The resulting acid-insoluble matrix (AIM) pellet was rinsed six times with MilliQ water, freeze-dried, and weighed. The supernatant containing acetic acid-soluble matrix (ASM) was filtered (Millipore, 5  $\mu$ M) and concentrated in an Amicon ultrafiltration stirred cell (model 8400, 400 mL) on a Millipore membrane (10 kDa cutoff). The final solution (>5 mL) was extensively dialyzed against 1 L of MilliQ water (six water changes) before being freeze-dried and weighed.

### Sample preparation for proteomic analysis

In-solution digestion of unfractionated ASM (0.1 mg) and AIM (1 mg) material was performed as follows. Samples were reduced with 50  $\mu$ L of 10 mM dithiothreitol in 50 mM ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ) for 30 minutes at 50°C. Alkylation was performed with 50  $\mu$ L of 100 mM iodoacetamide in 50 mM  $\text{NH}_4\text{HCO}_3$  for 30 minutes at room temperature in the dark. The solution was then treated with 1  $\mu$ g of trypsin (proteomic grade; Promega) in 10  $\mu$ L of 50 mM  $\text{NH}_4\text{HCO}_3$  overnight at 37°C. Samples were then dried in a vacuum concentrator and re-suspended in 30  $\mu$ L of 0.1% trifluoroacetic acid and 2% acetonitrile ( $\text{CH}_3\text{CN}$ ).

### Peptide fractionation and data acquisition

Mass spectrometry (MS) was performed using a Q-Star XL nanospray quadrupole/time-of-flight tandem mass spectrometer, nanospray-Qq-TOF-MS/MS (Applied Biosystems, Villebon-

sur-Yvette, France) coupled to an online nanoLC system (Ultimate Famos Switchos from Dionex, Amsterdam, The Netherlands). One microliter of each sample was loaded onto a trap column (PepMap100 C18; 5  $\mu\text{m}$ ; 100  $\text{\AA}$ ; 300  $\mu\text{M}$  x 5 mm; Dionex), washed for 3 minutes at 25  $\mu\text{L}/\text{min}$  with 0.05% trifluoroacetic acid/2% acetonitrile, then eluted onto a C18 reverse phase column (PepMap100 C18; 3  $\mu\text{m}$ ; 100  $\text{\AA}$ ; 75  $\mu\text{M}$  x 150 mm; Dionex). Peptides were separated at a flow rate of 0.300  $\mu\text{L}/\text{min}$  with a linear gradient of 5–80% acetonitrile in 0.1% formic acid over 120 minutes. MS data were acquired automatically using ANALYST QS 1.1 software (Applied Biosystems). Following an MS survey scan over  $m/z$  400–1600 range, MS/MS spectra were sequentially and dynamically acquired for the three most intense ions over  $m/z$  65–2000 range. The collision energy was set by the software according to the charge and mass of the precursor ion. MS and MS/MS data were recalibrated using internal reference ions from a trypsin autolysis peptide at  $m/z$  842.51  $[\text{M} + \text{H}]^+$  and  $m/z$  421.76  $[\text{M} + 2\text{H}]^{2+}$ .

### Mass spectrometry data analysis

Protein identification was performed using the MASCOT search engine (version 2.1; Matrix Science, London, UK) against translations in all six frames of our mantle transcriptomes, which possessed Benchmarking Universal Single-Copy Orthologs completeness scores of >98% (see below). Liquid chromatography (LC)-MS/MS data were searched using carbamido-methylation as a fixed modification and methionine oxidation as a variable modification. The peptide mass and fragment ion tolerances were set to 0.5 Da. The peptide hits (protein score >50; false discovery rate <0.05; 1 missed cleavage allowed) were manually confirmed by the observation of the raw LC-MS/MS spectra with ANALYST QS software (version 1.1). Quality criteria were the peptide MS value, the assignment of major peaks to uninterrupted y- and b-ion series of at least three to four consecutive amino acids, and the match with the *de novo* interpretations proposed by the software. All MS data has been deposited with the ProteomeXchange Consortium via PRIDE [30] with the dataset identifiers PXD008547 and 10.6019/PXD008547. Shell-forming candidates *Lstag-sfc-7*, *Lstag-sfc-8*, and *Lstag-sfc-9* were bioinformatically selected for analysis based on the presence of a signal peptide and their glycine-rich sequences (i.e., they were not detected using the proteomic methods described above).

### Bioinformatic analysis of protein sequences

Using the peptides identified from the proteomic survey described above, partial or, in most cases, full length coding sequences were isolated by standard or RACE polymerase chain reaction (PCR) as described in [31]. In some cases, Illumina transcriptome data (see below) were used to clarify the putative complete mRNA. Open reading frames were translated with the ExPASy translate tool [32]. Protein sequences were searched for signal sequences with SignalP 4.1 [33]. The theoretical pI, amino acid composition, and number of amino acids were determined using the ExPASy ProtParam tool [34]. Tandem repeats were identified with the T-REKS tool [35]. Sequence similarities searches were performed with the Basic Local Alignment Search Tool (BLAST) algorithm [36] with tBLASTx against nr and dbEST and BLASTx against SwissProt. Domain searches were performed with CD search [37]. Molecular function was predicted with InterProScan [38]. GalNAc O-glycosylation sites were predicted using the NetOGlyc 4.0 Server [39]. Scaled schematics of protein se-

quences were generated using Gene Structure Draw [40]. Intron-exon boundaries were mapped to a draft genome of *L. stagnalis* originally reported in [41] using Splign [42]. Similar transcripts were retrieved from the assembled transcriptomes of mantle zones 1–5 combined, mantle zone 5 alone, cephalic tentacle, cephalic lobe, central nervous system (CNS), foot, buccal mass, and larval stages 42 hours post first cleavage (hpfc), 52 hpfc, and 67 hpfc using BLASTn searches (see below for NGS details). All transcripts with complete open reading frames were considered. Only candidates yielding an mRNA coverage of >98% and an overall identity of >98% are documented. Scaled schematics of the gene architecture were generated using Gene Structure Draw [40]. Protein patterns were searched for using a modified local installation of PatMatch [43].

### NGS sequencing

Total RNA was extracted from the mantle edge and the proximal mantle tissue of a single adult *L. stagnalis* using TriReagent following the manufacturer's instructions. The resulting RNA was processed by the sequencing center at the IKMB at the University of Kiel (Germany). Paired end, stranded TrueSeq RNA libraries were constructed and sequenced for 101 bases from both ends using the Illumina HiSeq2000 platform (Illumina, CA, USA). More than 99 million and 100 million reads were generated from each of these libraries, respectively. These Illumina reads were processed using our pipeline as previously described [44]. Briefly, raw reads were adapter trimmed and quality filtered using Trimmomatic V0.32. Filtered reads were then assembled with Trinity V2.0.3, CLC Genomics Workbench *de novo* assembler (V8.5), and IDBA-tran. The resulting assemblies were then merged and filtered for redundancy using our pipeline [44]. Mantle transcriptome assemblies and cDNA and protein translations of the 34 shell-forming genes are available in the GigaScience Database, GigaDB [45]. In addition, transcriptomes from five adult tissues (cephalic tentacle, cephalic lobe, CNS, foot, and buccal mass) and three larval stages (42 hpfc, 52 hpfc, and 67 hpfc) were sequenced and assembled as described above. These transcriptomes were used to assess the tissue-specific alternative splicing characteristics of all shell-forming genes. All raw NGS data has been deposited in the Sequence Read Archive (SRA) with BioSample accession numbers SAMN08117214, SAMN08117215, SAMN08709370, SAMN08709371, SAMN08709372, SAMN08709373, SAMN08709374, SAMN08709375, SAMN08709376, and SAMN08709377.

### In situ hybridization on whole mounts and sections

Larvae were prepared for whole mount *in situ* hybridization as described in [46]. Sections (10  $\mu\text{M}$ ) were taken from *L. stagnalis* (shell length 10–50 mm) that had been fixed in formaldehyde for 1 hour and embedded in paraffin. Riboprobes were prepared as described in [16] and were used at concentrations of 100–500 ng/mL. Whole mounts and tissue sections were processed for hybridization, the color reaction developed, and photo-documented as described in [46].

### Comparisons of molluscan shell-forming proteomes

BLASTp-based comparisons of the *L. stagnalis* shell proteome were performed against a variety of calcifying proteomes reported in a wide phylogenetic range of metazoans as described in [14]. These included 42 proteins from the oyster *Pinctada max-*

ima reported in [47]; 78 proteins from the oyster *Pinctada margaritifera* reported in [47]; 94 proteins from the abalone *Haliotis asinina* reported in [17] and [16]; 80 proteins from the abalone *H. laevigata* reported in [48]; 63 proteins from the limpet *Lottia gigantea* reported in [49]; 53 proteins from the oyster *Crassostrea gigas* reported in [50]; 71 proteins from the mussel *Mya truncata* reported in [51]; 59 proteins from the grove snail *Cepaea nemoralis* reported in [14]; 44 proteins from the oyster *Pinctada fucata* reported in [52]; 53 proteins from the mussel *Mytilus coruscus* reported in [53]; 66 proteins from the brachiopod *Magellania venosa* reported in [54]; 139 proteins from the sea urchin *Strongylocentrotus purpuratus* reported in [55]; and 37 proteins from the coral *Acropora millepora* reported in [56].

### Analysis of the saccharide moieties of the shell matrix

The monosaccharide content of AIM and ASM was obtained by suspension and homogenization (vortex and ultrasound) of lyophilates in 2 M trifluoroacetic acid and subsequent hydrolysis at 105°C for 4 hours under a nitrogen atmosphere. This treatment allows for the release of most monosaccharides from complex mixtures, except sialic acids, which are destroyed, and the acetylated forms of glucosamine and galactosamine, which are converted to their respective non-acetylated forms. Samples were then centrifuged for 5 minutes at 15,000g and evaporated to dryness (using a SpeedVac) before being dissolved in 100 µL of 20 mM sodium hydroxide and homogenized. After a short centrifugation (2 minutes), 80 µL of supernatant was injected into the chromatograph system. The neutral, amino, and acidic sugar contents of hydrolysates were determined using high-pressure anion exchange-pulsed amperometric detection on a CarboPac PA 100 column (Dionex Corp., Sunnyvale, CA, USA). As blank controls, non-hydrolyzed AIMS were analyzed in order to detect potential free monosaccharides that may lead to an over-representation of some sugar residues.

## Results

### Proteomic analysis of the biomineralized matrix of *L. stagnalis* shells

More than 1,230 peptides were analyzed by High-performance Liquid Chromatography (HPLC)-MS and subsequently used for protein identification using Mascot against our translated mantle transcriptomes. Of these 1,230 peptides, 329 returned significant matches. From these 329 matches, a total of 40 shell-forming candidate transcripts were identified (see Additional file 1). Of these 40 gene products, 31 (78%) could be cloned and exhibit *in situ* hybridization signals compatible with a role in shell formation (either in larval stages and/or in the adult mantle tissue). Seven of these 40 candidates (18%) could be cloned from *L. stagnalis* cDNA but did not produce a positive or consistent *in situ* signal in any tissue. Three of the 40 candidate genes (8%) could not be amplified by gene-specific PCR or RACE PCR. In addition to the 31 proteomically identified candidates that generated positive *in situ* signals, three candidates that were identified via *in silico* methods (based purely on the presence of a signal sequence and their glycine-rich protein sequences) also generated *in situ* signals compatible with a role in shell formation and are reported here.

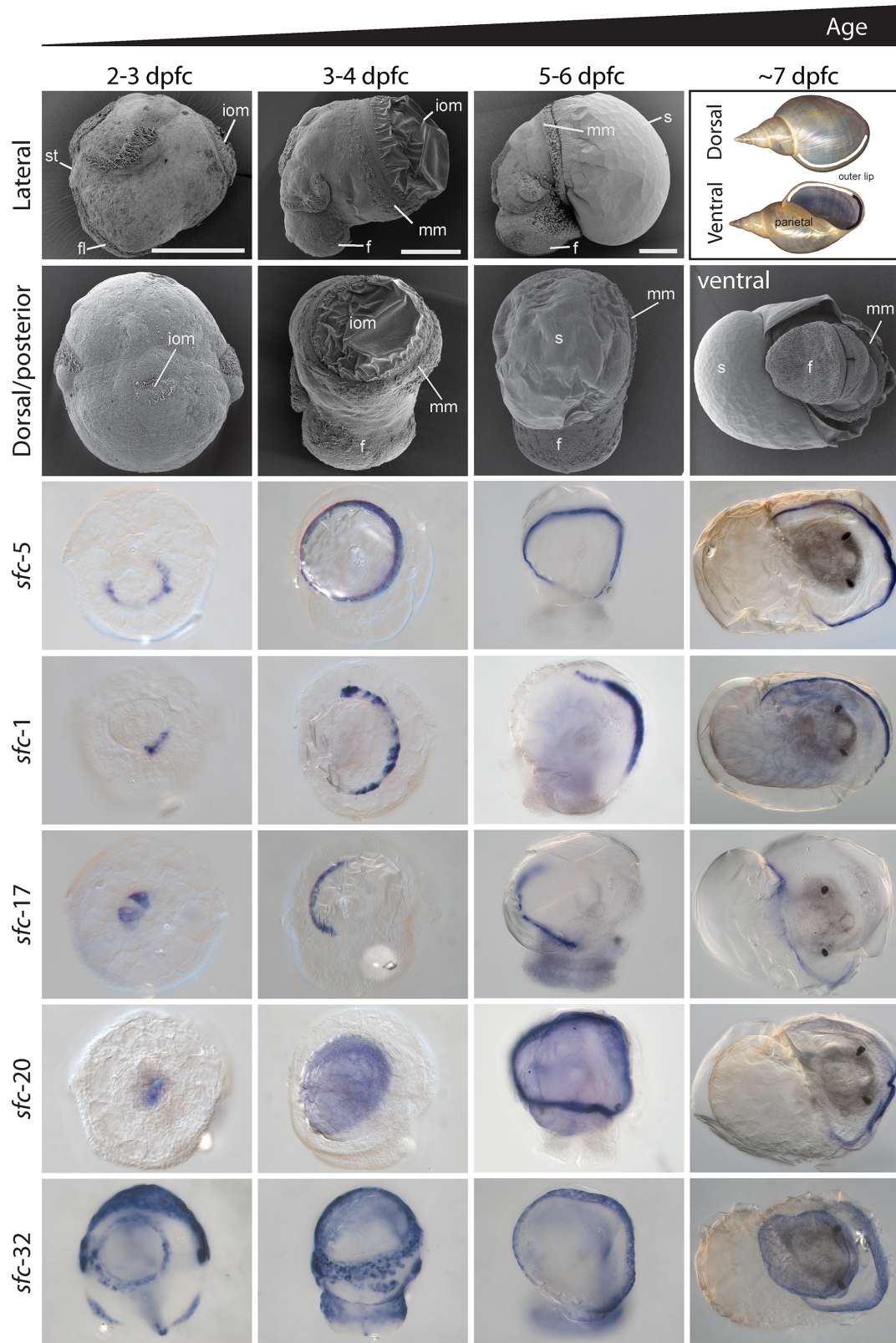
### A brief morphological description of *L. stagnalis* shell ontogeny and the adult mantle

We previously described the ontogeny of the shell gland and shell field in *L. stagnalis* [57]. In order to aid the interpretation of our *in situ* patterns, the following is a summary of the main developmental stages that we focused on. The first visible sign of differentiation of the shell-forming tissue in *L. stagnalis* is a thickening of the dorsal ectoderm that begins at approximately 29 hpf [57, 58]. These cells subsequently invaginate and by 2 days post first cleavage (dpfc) a clearly visible “shell gland” is present [57, 58]. By 3 dpfc, the shell gland has formed a sealed lumen and displays the first signs of outward signs of asymmetry [57]. The marginal cells that border the shell gland remain uninvaginated and form a ring-like structure, the rosette [2]. During this time, the first extracellular organic material is secreted and is clearly visible by Scanning Electron Microscopy (SEM) (Fig. 1; [57]). By 3 dpfc, the shell gland has evaginated to form the shell field. The former rosette cells remain highly elongated while the central cells take on a low columnar appearance. Over the next several days, the shell field continues to expand until it has overgrown the visceral mass and will eventually become the adult mantle tissue [2, 57, 58].

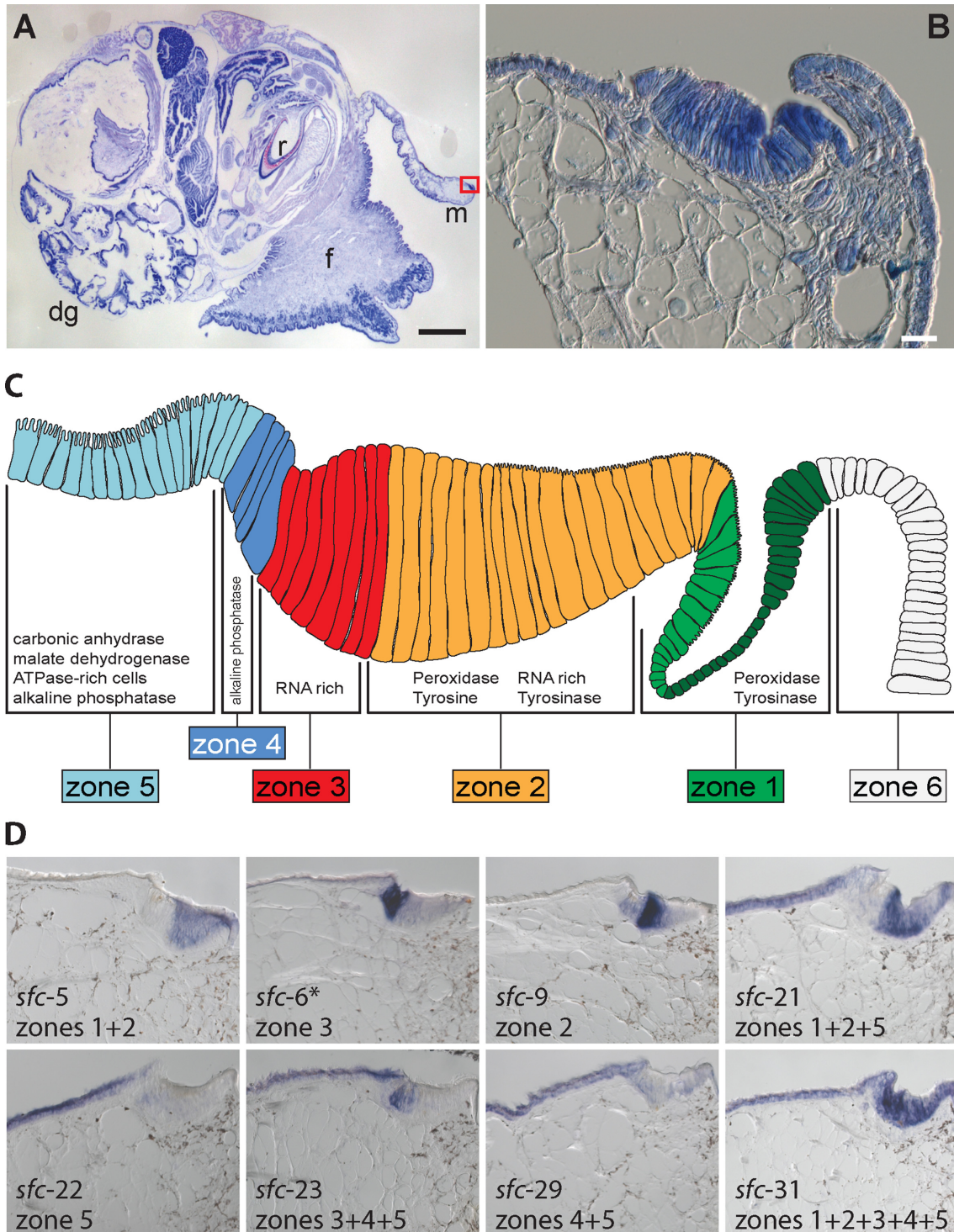
The adult mantle covers the inner surface of the shell and is responsible for shell growth and repair. The free edge of the mantle is responsible for the growth of the outer lip of the shell. Timmermans conducted an extensive histochemical characterization of the mantle tissue of *L. stagnalis* and was able to categorize the free edge of the adult mantle into six distinct zones based on their morphology, enzymatic activities, and biochemical signatures [59]. We largely follow this categorization of the adult mantle tissue. Parallel to the mantle edge runs the mantle groove (also known as the pallial groove) defined as zone 1 (Fig. 2). Several high-resolution microscopy and histological studies on a variety of molluscs have demonstrated that it is from within the pallial groove that the periostracum is formed and secreted [59–64]. We detected a sub-regionalization of the pallial groove (zone 1) into proximal and distal zones. Immediately adjacent to the pallial groove is a broad region of high columnar cells referred to by Timmermans [59] as the “belt” that can be subdivided into three distinct zones (zones 2–4). Zone 2 is immediately adjacent to the posterior wall of the pallial groove and comprises the anterior (or distal) portion of the belt (Fig. 2). Zone 3 consists of the posterior portion of the belt, while zone 4 represents the transitional zone between the high columnar cells of the belt proper and the more posterior low columnar cells of the outer epithelium, which comprise zone 5 (Fig. 2) [59].

### Spatial expression patterns and molecular features of shell-forming candidate genes

We performed *in situ* hybridization for 34 distinct shell-forming genes on four distinct developmental stages and on adult mantle tissue. The detailed results of these analyses are presented in Additional files 2–35, with an extensive summary presented in Additional file 36 (the raw image files that constitute these figures are available in the associated GigaDB repository [45]). In Fig. 1 (for larvae) and Fig. 2 (for adult mantle tissue) we present a selection of these results that highlight some prominent features of these expression patterns. In trochophore and veliger larval stages (2–6 dpfc), all genes could be categorized either as being: expressed in cells that symmetrically or asymmetrically border the shell gland or shell field (15/34); in cells that lay within the shell gland or shell field (9/34); a pattern that did not fit into



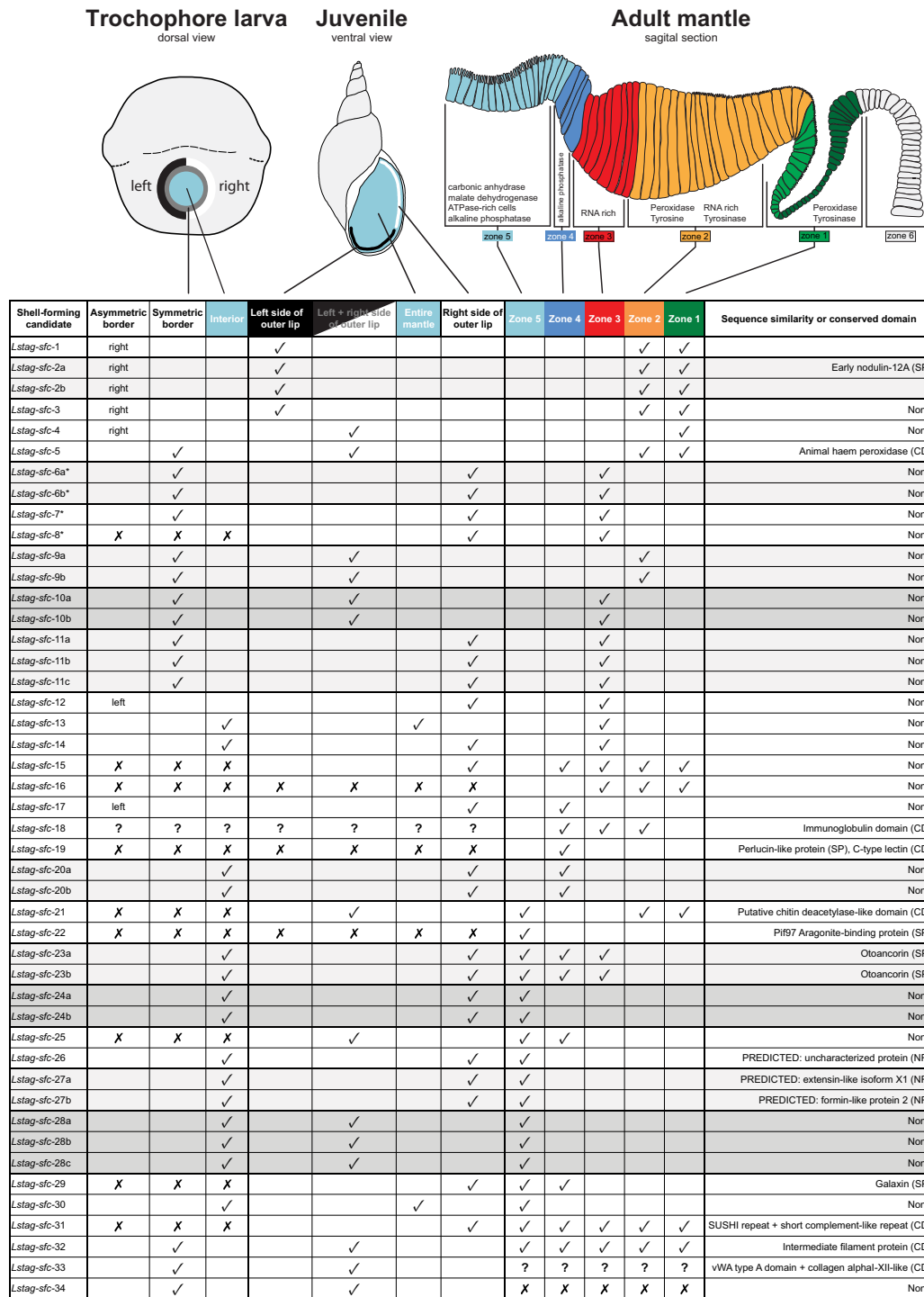
**Figure 1:** Overview of four developmental stages and representative shell-forming gene expression patterns in *L. stagnalis*. The first two rows provide a set of reference SEM images and adult shells (top-right-most panel) against which the images of the *in situ* results can be oriented. All *in situ* panels are from a dorsal view except the right-most column, which is from a ventral view. Here, we present the expression patterns of a selection of five shell-forming genes. These include genes with expression patterns in shell-forming cells that display evidence of symmetry (*sfc-5*), right asymmetry (*sfc-1*), left-asymmetry (*sfc-17*), expression entirely throughout the shell field and dorsal mantle epithelium (*sfc-20*), and expression in additional non-shell-forming cells. This last expression pattern provides evidence of genes involved in shell formation that have pleiotropic functions. The scale bars in the first row are 100  $\mu$ m. Indicated in the SEM images are the positions of the foot lobe (fl), foot (f), mantle margin (mm), calcified shell (s), stomodeum (st), and insoluble organic material (iom) of the shell.



**Figure 2:** Overview of the adult *L. stagnalis* shell-forming mantle tissue and representative shell-forming gene expression patterns that reveal its molecular modularity. (A) A semi-thin sagittal section of an adult *L. stagnalis* stained with Giemsa. The foot (f), mantle (m), digestive gland (dg), and radula (r) are indicated. The mantle tissue is a thin sheet of epithelium that covers the dorsal surface of the adult animal and is responsible for fabricating the shell. (B) A magnified view of the red-boxed region in part A reveals some of the cellular morphology of the adult mantle tissue. (C) A schematic representation of the mantle tissue divided into six zones as described by Timmermans [59]. The spatial distribution of enzymatic activities and biochemicals indicated in this schematic are adapted from [59]. We detect a sub-regionalization of the pallial groove (zone 1) into proximal (light green) and distal (dark green) zones. (D) The spatial expression patterns of eight representative shell-forming genes in the adult mantle tissue. The asterisk indicates that *sfc-6* was identified using *in silico* methods rather than proteomic methods.

our classification scheme (1/34); or were not expressed in any detectable way (9/34). In later stages (~7 dpfc), all genes were either: expressed uniformly along the outer edge of the mantle

(10/34); asymmetrically in the outer edge of the mantle (18/34); throughout the entire mantle tissue (2/34); a pattern that did not fit into our classification scheme (1/34); or were not expressed in



**Figure 3:** Summary of the spatial gene expression profiles and conserved features of 34 *L. stagnalis* shell-forming candidates. Schematically represented in a trochophore larva are genes with an asymmetric expression profile (dark gray), as well as genes expressed broadly across the shell field (light blue). Cells in this region of the trochophore are likely to give rise to cells in zone 5 of the adult mantle, and we have maintained that color scheme to suggest this. Although we schematically present a trochophore larva here (2–3 dpfc), the summarized expression patterns also include veliger stages (3–6 dpfc). Cells bordering the larval shell gland and shell field (black ring in the trochophore) are likely to give rise to one or more zones 1–4 in the adult mantle. In juveniles (~7 dpfc) many genes were expressed in the left, right, or continuously throughout the free edge of the mantle that produces the outer lip of the shell. Question marks indicate expression patterns that could not be categorized according to our scheme. An “x” indicates no expression was detected. A “?” indicates that the expression pattern could not be categorized according to our scheme. The names of enzymes and other molecular features indicated in zones 1–5 on the schematic of the adult mantle are summarized from [59] and [3]. Sequence similarity and conserved domains in the final column of the table are summarized from a number of BLAST searches against SwissProt, the non-redundant National Center for Biotechnology Information database, and the Conserved Domain database. See Additional files 40–42 for the results of all BLAST and domain searches. The lineages of the top BLAST hits are listed in Additional file 43. A version of this figure that includes a more complete summary of the molecular features of each gene is provided in Additional file 36. The asterisks indicate that sfc-6, -7, and -8 were identified using *in silico* methods rather than proteomic methods (as was the case for all other gene products presented here).

any detectable way (3/34). Finally, in adult mantle tissue, genes were either: expressed in one or more of the five zones described by Timmermans [59] (32/34); a pattern that did not fit into our classification scheme (1/34); or were not expressed in any detectable way (1/34). We have schematically summarized all of these results in Fig. 3.

### Alternative splicing increases the diversity of shell-forming proteins

Via alternative splicing of mRNAs, transcripts with a variety of functions can be generated from a single genomic locus [65]. With a draft genome for *L. stagnalis* available, we were able to perform some preliminary investigations into alternative splicing of our shell-forming candidates. While some candidate genes displayed the same or very similar exon-splicing patterns in all surveyed tissues (e.g., Additional files 5, 9, 14, 15, and 16), most candidates are apparently alternatively spliced depending on the tissue they are expressed in (Additional files 4, 11, 12, 18, 19, 22, 23, 25, and 33). Striking examples include *Lsttag-sfc-21* and *Lsttag-sfc-24*, which are expressed in many tissues but display significant alternative splicing in the adult mantle (Additional files 22 and 25). All splice variants of candidate *Lsttag-sfc-24* encode proteins with the same aspartic acid-rich motif (Additional files 25, 37, and 38). Aspartic acid-rich proteins have been suggested to act as an organic template for epitaxial crystal growth [66, 67]. It is tempting to speculate that the three additional domains only present in adult mantle *Lsttag-sfc-24* contigs confer a specific shell-forming function to this protein. The putative chitin-interacting candidate *Lsttag-sfc-21* presented in Additional file 22 carries a signal sequence and is predicted to possess a catalytic activity. Intriguingly, a number of splice variants of this gene within the adult mantle are predicted to lack a signal sequence, the chitin-binding or catalytic ability (Additional file 37).

A number of shell-forming gene candidates produce alternatively spliced transcripts that encode proteins with differences regarding the presence/absence of a signal sequence (Additional files 11, 12, 22, 25, and 37). Some shell-forming genes also produce alternatively spliced transcripts that encode proteins with similar coding features but radically different 5' or 3' untranslated regions (UTRs) (Additional files 12, 19, 33, and 37). While UTRs do not contain protein-coding information, they can be critical for localization of the mRNA [68, 69] and post-transcriptional gene regulation by molecules such as miRNAs [70]. Indeed, several miRNAs have now been associated with the targeting and regulation of biomineralizing proteins [71, 72].

### Comparisons of molluscan shell-forming proteomes

We conducted a broad comparison of our *L. stagnalis* shell-forming genes against a wide phylogenetic range of 12 other biomineralizing proteomes comprising 879 proteins (sequences used in this analysis are provided in Additional file 39). Of all *L. stagnalis* shell proteins, 27 shared significant sequence similarity with 1 or more of these 12 proteomes (Fig. 4). The highest degree of overall similarity was found with the shell-forming proteome of the common groove snail *C. nemoralis* (Fig. 4), the closest phylogenetic relative to *L. stagnalis* of all species in this comparison. The next highest level of similarity shared with the *L. stagnalis* shell-forming proteome (15.9% of the *L. gigantea* shell-forming proteome) was markedly lower. *Lymnaea stagnalis* shell-forming proteins that shared significant similarity with biomineralizing proteins from other species and that also returned a significant match against a SwissProt entry included *Lsttag-sfc-*

32 (with similarity to *C. nemoralis* contig 572), which appears to be an intermediate filament protein. *Lsttag-sfc-22* (a gene expressed exclusively in zone 5, Additional file 23) shared relatively weak similarity with *C. nemoralis* contig 821 and shares significant sequence similarity with PIF, an aragonite-binding protein reported to be involved in nacre formation in the oyster *P. fucata* [19]. Interestingly, of the 12 candidates expressed in the matrix-secreting zone 5, 9 showed similarity with other shell proteins (Fig. 4 and Additional files 23–33). Eight of these were shared with *C. nemoralis* (Fig. 4). In contrast, none of the asymmetrically expressed or glycine-rich candidates were found in any of the other biomineralizing proteomes (low-complexity filtering was inactivated in these comparisons; Fig. 4).

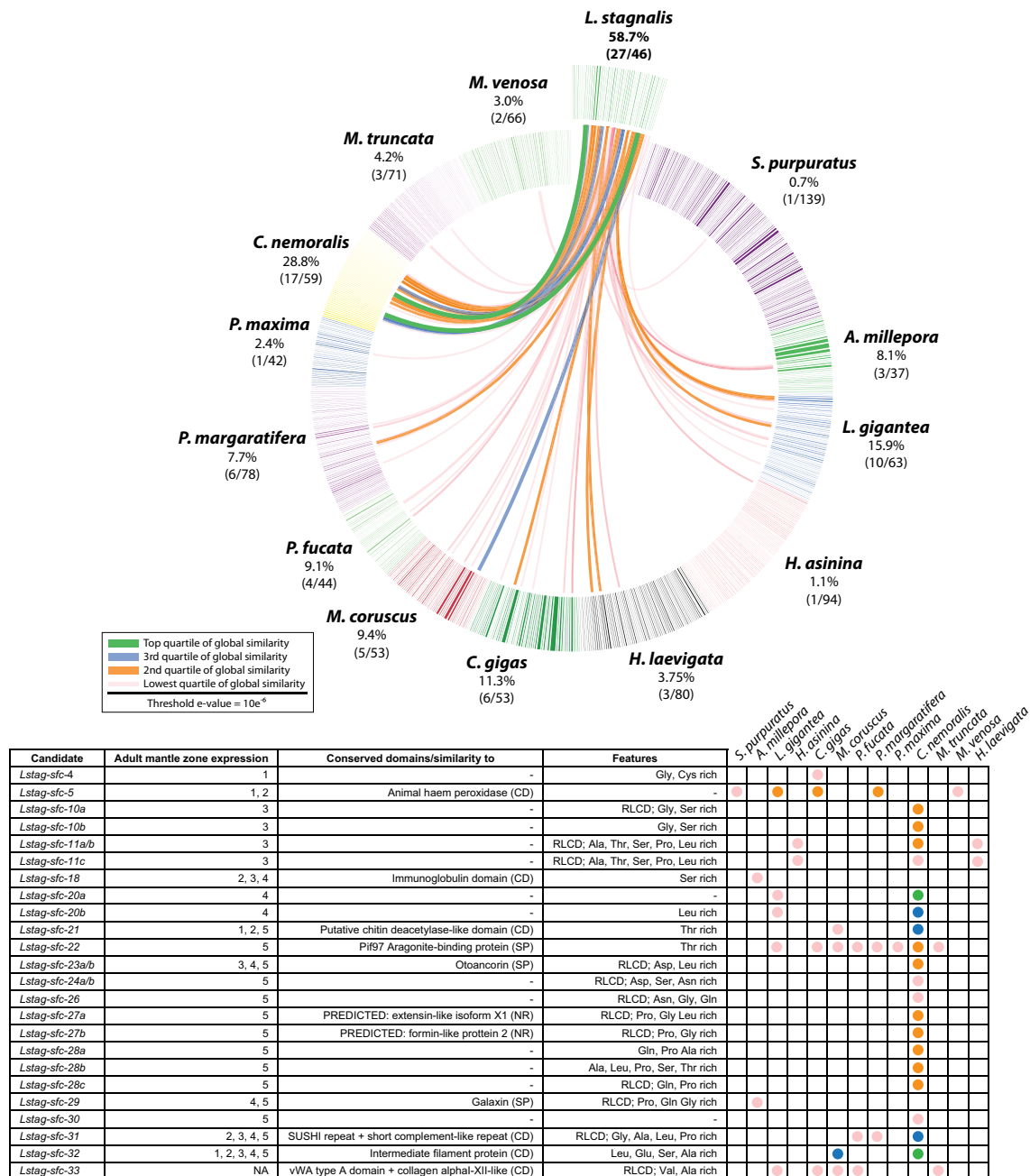
### Glycosylation patterns of the shell matrix

The monosaccharides profiles of ASMs and AIMs extracted from adult *L. stagnalis* shells were peculiar, with less than half of the dozen standard monosaccharides represented. For ASM, these include galactosamine, glucosamine, galactose, glucuronic acid, and glucose, while AIM lacked glucuronic acid (Table 1). We also found marked differences in the glycosylation rates of ASMs and AIMs. In general, the absolute degree of glycosylation of the ASM was 1–3 orders of magnitude greater than that of the AIM. The largest difference was in the amount of galactosamine (10.6 ng/ $\mu$ L vs. 0.03 ng/ $\mu$ L). While glucosamine was more abundant absolutely in the ASM (9.41 ng/ $\mu$ L vs. 0.13 ng/ $\mu$ L), the proportional difference was not so extreme (34.8% vs. 54.2%).

## Discussion

### Molecular modularity of the adult molluscan mantle

Two of the most striking features of the phylum Mollusca are its size (in terms of number of species) and its diversity. Widely accepted to be second only to the Arthropoda in terms of number of living species [73, 74], molluscs arguably display the greatest diversity of body forms of all metazoan phyla and have successfully colonized all kinds of environments. While there currently exists no consensus as to why molluscs have enjoyed such deep evolutionary success (one interesting suggestion includes a plastic nervous system [75]) we believe the mantle tissue (an apomorphy of the phylum) and its ability to prolifically evolve new shell phenotypes must contribute to an explanation of this success. A logical extension of this question would therefore be, “what is it about the molluscan mantle tissue that makes it so evolutionarily plastic?” For arthropods, segmentation and body plan modularization (and the underlying gene regulatory networks that control appendage identity within each segment) are widely thought to have played leading roles in supporting the diversification of insects, spiders, and crustaceans [76]. The importance of establishing segmentation at a very early developmental age in prominent phyla such as annelids, chordates, and arthropods has caused much effort to be spent on identifying the causal molecular mechanisms that may have common evolutionary histories [77–79]. As recently reviewed by Esteve-Altava [80], the presence of morphological modules can help us to understand the evolvability of body form, but the identification of such modules has so far been biased toward mammals, arthropods, and plants. Following Esteve-Altava’s and Eble’s [81] definition of a morphological module (a group of body parts that are more integrated among themselves than they are to other parts outside the group), we propose that the molluscan mantle is a prime example of such a morphological module. This modular



**Figure 4:** BLASTp comparisons of the *L. stagnalis* shell proteome against 879 biocalcifying proteins derived from six bivalves, four gastropods, one brachiopod, one sea urchin, and one coral. Individual lines spanning the ideogram connect proteins that share significant similarity ( $e$ -values  $< 10e^{-6}$ ). Transparent red lines connect proteins with the lowest quartile of similarity (with a threshold of  $10e^{-6}$ ) and green lines with the highest quartile of similarity. The percentage of each shell proteome that shared similarity with the *L. stagnalis* proteome is indicated. The table provides further information for those candidates that share sequence similarity with another species. Abbreviations: CD: conserved domain database; NR: GenBank non-redundant protein database; SP: SwissProt database.

nature of the molluscan mantle is not unique to *Lymnaea* [82–84]. Although the precise functions of these zones (and of the individual gene products that define them) await the development of targeted genome editing methods, it is clear that they must act in a coordinated way to deposit the shell. We predict that there are related modules of gene regulatory networks (GRNs) that act to specify each zone of the molluscan mantle and that it is the modular nature of these GRNs and the resulting morphological modularity of the mantle tissue that supported the diversification of the phylum Mollusca. Characterization of the

spatial expression patterns of shell-forming proteomes from a selection of other molluscan lineages will contribute to a more refined understanding of molluscan shell evolution.

### Ontogenetic expression of shell-forming candidates

A prominent outcome of our survey of the adult shell proteome is that many of the genes that encode these proteins are not only regulated spatially but also temporally. Many shell-forming candidates are expressed in the invaginated larval shell gland of the

**Table 1:** Glycosylation analysis of acid-soluble and acid-insoluble matrices extracted from adult *L. stagnalis* shells

Monosaccharide	ASM		AIM	
	ng/ $\mu$ g	%	ng/ $\mu$ g	%
Fucose	ND	-	TR	-
Rhamnose	ND	-	TR	-
Galactosamine	10.60	39.2	0.03	12.5
Arabinose	ND	-	ND	-
Glucosamine	9.41	34.8	0.13	54.2
Galactose	4.34	16.0	0.04	16.7
Glucose	0.32	1.2	0.04	16.7
Mannose	ND	-	TR	-
Xylose	ND	-	ND	-
Galacturonic acid	ND	-	ND	-
Glucuronic acid	2.39	8.8	ND	-
<b>Total</b>	<b>27.06</b>	<b>100.0</b>	<b>0.24</b>	<b>100.0</b>

ND = not detected, TR = trace.

trochophore (Additional files 14, 15, 21, 24, 25, 27, and 28) or in cells that border it (Additional files 2, 3, 6, 7, 8, 10, 11, and 18). Only two candidates were solely expressed in the adult mantle tissue and not in any larval stage (Additional files 17 and 23). Timmermans [59] concluded that the spatial patterning of larval shell-forming cells persists throughout development and foreshadows the zonation observable in the adult mantle. We also observed this phenomenon at the molecular level. All candidate genes that were expressed in the margin of the shell gland or the shell field were expressed in the belt (zones 2, 3, and 4) of the adult mantle (summarized in Fig. 3, Additional files 2–8, 10–12, and 18). Most candidates expressed in the invaginated cells of the shell gland or throughout the developing shell field were subsequently expressed in the low columnar outer epithelium of zone 5 in adult mantle tissue (summarized in Fig. 3, Additional files 24, 25, 27–29, and 31). However, three genes conspicuously deviate from this pattern. *Lsttag-sfc-13*, *-14*, and *-20* display a broad expression pattern in the invaginated cells of the shell gland and throughout the entire shell field in larvae but were not detected in the low columnar outer epithelium (zone 5) of the adult mantle tissue (Additional files 14, 15, and 21). However, we should point out that for all candidate shell-forming genes, we did not consider the potential effect of a diurnal rhythm on gene expression. All samples for *in situ* hybridization were taken during daylight hours, and so genes with activity during the night would be missed.

### Asymmetric expression of shell-forming genes

The expression of *Lsttag-sfc-1*, *Lsttag-sfc-2*, and *Lsttag-sfc-3* in zones 1 and 2 of the adult mantle suggests they may be involved in the formation of the periostracum (Fig. 1 and Additional files 2–4); however, it is their larval expression patterns that are more striking. *Lsttag-sfc-1*, *-2*, and *-3* display a right-sided asymmetric expression pattern in cells bordering the shell gland and shell field. In contrast, *Lsttag-sfc-17* is expressed on the left side (Fig. 1 and Additional file 18). Following the expression of these genes ontogenetically into older larvae that begin to display the coiled phenotype of the adult, it is apparent that right-sided cells in the trochophore are likely to be those that give rise to the right + anterior region of the adult mantle that will produce the outer lip of the shell, while left-sided cells will give rise to posterior mantle tissue responsible for forming the left + parietal region of the shell

(Fig. 1). We therefore suggest that *Lsttag-sfc-1*, *-2*, and *-3* are in some way associated with producing thinner, more rapidly produced shell at the outer shell lip than in the thicker parietal region, while *Lsttag-sfc-17* may inhibit the rapid deposition of shell. Exactly how this is achieved awaits the development of gene-specific function assays.

In addition to the trochophore left/right asymmetry corresponding to the left + parietal/right + outer lip regions of the shell, there is a second axis of symmetry that becomes apparent in 7-day-old juvenile snails. Many shell-forming candidates are initially symmetrically expressed in or surrounding the shell gland of 2- to 3-dpfc trochophores but then become asymmetrically expressed in the mantle of older animals. For example, *Lsttag-sfc-6*, *-7*, *-8*, *-12*, *-14*, *-15*, *-17*, *-18*, *-20*, *-23*, *-24*, *-26*, *-27*, *-29* and *-31* are expressed in the left side of the free mantle edge in 7-dpfc juveniles (summarized in Fig. 3). In contrast, relatively few shell-forming gene candidates (*Lsttag-sfc-5*, *-9*, *-10*, *-21*, and *-25*) are expressed evenly along the free edge of the mantle in 7-dpfc juveniles (summarized in Fig. 3).

### The spatial expression of a peroxidase in the adult mantle allows a model of shell formation to be developed

In agreement with Timmermans histochemical study of peroxidase activity [59], the expression of *Lsttag-sfc-5*, a shell-forming candidate with an “animal heme-dependent peroxidase” domain (Pfam PF03098; Additional file 41B) is localized to zones 1 and 2. Peroxidases may be involved in periostracum formation by cross-linking fibrous proteins rich in reactive quinones to form water insoluble, protease-resistant polymers [85–87]. This process, also referred to as tanning or sclerotization, can also be catalyzed by tyrosinase (also known as catechol oxidase, catecholase, polyphenoloxidase, phenoloxidase, and phenolase). Within the molluscan biomineralization literature, sclerotization by tyrosinase appears to be the more commonly assumed mechanism, rather than by peroxidase. Nonetheless, Timmermans demonstrated that heat inactivation clears the periostracal groove and belt of both peroxidase activity and the ability to form melanin (a typical assay used to test for tyrosinase activity), while specific tyrosinase inhibitors sodium bisulphite and potassium cyanide (NaHSO<sub>3</sub> and KCN) did not affect its ability to produce melanin [59]. The spatial expression pattern of *Lsttag-sfc-5*, coupled with the observations that newly secreted perio-

tracum itself also displays peroxidase activity [57] and Timmermans experiments [59], strongly suggests that the peroxidase we have identified here plays a key role in cross-linking the periostracum in *L. stagnalis* rather than a tyrosinase, as also supposed for other gastropods such as *Lottia* [15].

### Glycine-rich shell-forming candidates are likely to be substrates for the peroxidase

An important aspect of scleroprotein formation is its spatial coordination. The cross-linking reaction often generates cytotoxic intermediates, and the end products cannot be easily degraded or resorbed [88]. Furthermore, the uncontrolled formation of extensive scleroprotein polymers prior to secretion would clearly be detrimental to the cell. One common strategy to avoid these events is to compartmentalize the scleroprotein precursor (that is unable to spontaneously polymerize) away from the cross-linking enzyme. Following secretion, the precursors are activated and enzymatically cross-linked [88]. Such a scenario would suggest that the substrate upon which the peroxidase acts is not located within the same cells.

Three candidates expressed in zone three (*Lstag-sfc-6*, *-7*, and *-8*) encode secreted, basic proteins that are dominated by repetitive low-complexity domains (RLCDs) and anomalous amino acid contents (high glycine, tyrosine, asparagine, and leucine contents; Additional files 37 and 38). All of these glycine-rich proteins carry tyrosine residues flanked by glycine. This arrangement has been shown to be favorable for the formation of cross-links between tyrosine residues by peroxidase [89]. Waite, in his review of natural quinone-tanned glues [85], highlighted the typical L-3,4-dihydroxyphenylalanine (DOPA)-containing consensus precursor peptide sequences from a number of marine invertebrates. Allowing for a single mismatch, these substrate peptides (VGGYGYGK, GGGFGGYGK, and GGGYGGYGK, cross-linking tyrosine residues in bold) can be found within *Lstag-sfc-6*, *Lstag-sfc-7*, and *Lstag-sfc-8*. Interestingly, these glycine-rich candidates are expressed exclusively in zone 3 (Additional files 7–9) immediately adjacent to zone 2, the region in which the peroxidase *Lstag-sfc-5* is expressed (Additional file 6). Theoretically, once these proteins are secreted, the secreted peroxidase would be in very close proximity to the glycine-rich proteins and could act on the favored tyrosine residues to form di-tyrosine cross-links extracellularly.

### A role for immunity and signaling in shell formation

*Lstag-sfc-18* contains two Ig superfamily domains (Additional file 19; [90]) and displays sequence similarity with the IMP-L2-like proteins (Additional files 40 and 41), an insulin-like growth factor-binding protein (IGF-BP) that carries two immunoglobulin-like domains and is able to bind IGF [91]. Several studies by Dogterom and colleagues demonstrated the influence of a growth hormone secreted by the cerebral ganglia specifically on shell formation in *L. stagnalis* [92–95]. The authors conclude that this growth hormone acts on cells in the belt region to control shell extension and periostracum formation but not on shell thickening. Interestingly, perlustrin, a protein associated with nacre in abalone shells, contains an IGF-BP domain and was also shown to bind IGFs and insulin [96]. An intriguing idea for the presence of IGF-BP in the abalone shell is that it would allow the shell to signal to the underlying mantle epithelium. According to this hypothesis, IGFs present in the extrapallial fluid are bound by IGF-BP during calcification and incorporated into the shell. Should the shell dissolve or be locally

damaged, these IGFs would be released and subsequently stimulate the underlying mantle epithelium to re-calcify. One line of evidence that strongly supports this hypothesis is provided by the osteogenic activity of mollusc shells [97]. This hypothesis implies that although the shell is acellular, it is able to actively communicate and provide real-time feedback to the mantle epithelium [98]. This interesting hypothesis awaits the development of gene-specific knock-down or knock-out assays.

### RLCDs are an abundant feature of *L. stagnalis* shell proteins

Proteins containing RLCDs are a prominent feature of molluscan shell-forming proteomes [15, 99, 100], and *L. stagnalis* is no exception. More than half of the *L. stagnalis* shell-forming candidates we identified possess RLCDs. Proteins containing these domains were present in the belt and the low columnar outer epithelium of the adult mantle and in a wide variety of patterns of the larval stages we investigated. The motif complexity, motif length, and number of motif repeats can vary greatly, from stretches consisting of a single amino acid (e.g., Additional file 24) to motifs that exceed 10 amino acids (e.g., Additional files 3, 4, 29 and 34). In some cases, almost the whole protein is composed of RLCDs (Additional files 7–10). Repeated motifs are a common feature of structural proteins such as collagens, keratins, silk, and cell wall proteins, as well as structural modules in functional proteins such as receptors, histones, ion channels, and transcription factors [101, 102]. RLCDs are often part of intrinsically unstructured regions that lack a fixed or ordered three-dimensional structure [101]. In some cases, these regions define the functionality of the protein. As a general rule, unstructured proteins interact readily with other proteins [103], and the highly repetitive, modular, and biased amino acid compositions can confer strength and elasticity [104]. It will be extremely informative to selectively remove RLCDs from shell-forming proteins and to study the resulting shell phenotypes once genome modification tools become broadly available to molluscs.

### Broad sequence similarity comparisons of metazoan biomineralizing proteomes

The crossed-lamellar microstructure is fabricated by phylogenetically diverse molluscan taxa and is by far the most commonly employed shell design of the Conchifera [21, 22, 28]. While much attention has been dedicated to the characterization of nacre-forming bivalve shell proteomes, technical advances in nucleic acid sequencing and proteome-scale surveys have seen a rapid growth in the number and diversity of molluscan shell-forming proteomes and allow broad comparisons of these datasets to be performed. These comparisons can provide insight into the degree of evolutionary conservation that exists across shell-forming proteomes [50]. In general, molluscan shell-forming proteomes are markedly different, with some deeply conserved elements such as alkaline phosphatases, peroxidases, and carbonic anhydrases [28, 57, 105, 106]. The significant diversity of molluscan shell ultrastructures, crystal textures, colors, and materials properties therefore cannot be explained by the use of the same genes in different ways. Rather, each lineage has uniquely evolved a large fraction of its shell-forming proteome [14–16, 100, 107]. To expand on this comparative theme, we collected 879 biomineralizing proteins validated by proteomics from 10 molluscs, 1 brachiopod, 1 sea urchin, and 1 coral and performed sequence similarity comparisons against

our *L. stagnalis* dataset. Two of the 10 molluscs, *Cepaea nemoralis* and *Mya truncata* [51, 108], construct shells that contain crossed-lamellar texture. Interestingly, our comparative analyses show that *L. stagnalis* and *M. truncata* have only three proteins that share relatively low degrees of sequence similarity, while *L. stagnalis* and *C. nemoralis* share 17 proteins (some of these with very high degrees of sequence similarity), the highest extent of similarity between all species surveyed (Fig. 4). Both *L. stagnalis* and *C. nemoralis* inhabit non-marine environments, and the similarities in their shell proteomes may either be a manifestation of this and/or a reflection of their crossed-lamellar shells. The similarity of their shell protein content may also reflect the relatively recent divergence time (Meso-Cenozoic) of these two clades (Stylommatophoran, i.e., *C. nemoralis* vs. Hygrophilid, i.e., *L. stagnalis*) within the monophyletic order of pulmonate gastropods, in comparison to the other species. One of the most striking observations we made in these comparisons was that almost all *L. stagnalis* shell-forming candidates expressed in zone 5 share sequence similarity with *C. nemoralis*. Conversely, *L. stagnalis* shell-forming candidates expressed asymmetrically on the right side in larvae were not present in any other biomineralizing proteomes.

Some *L. stagnalis* shell-forming proteins contain domains found in a number of the biomineralizing proteins present in the dataset we assembled or are known to play a role in processes other than biomineralization such as the Sushi domain, the von Willebrand factor A domain, the immunoglobulin domain, and the filament protein domain [14, 109]. The Pif-like protein is prevalent in both bivalve and gastropod nacreous shell proteomes and is known to bind aragonite crystals and to regulate nacre formation [110]. However, limpets, which construct crossed-lamellar structures, also contain Pif in their shells [15, 110, 111]. Our results further demonstrate that Pif is not limited to nacreous matrices and that it is likely to be a deeply conserved element of the molluscan biomineralizing proteome.

### Glycosylation patterns

Our preliminary analysis of the sugar moieties associated with shell-forming proteins revealed an interesting dichotomy between the ASM and AIM; the population of ASM proteins appear to be far more glycosylated than AIM proteins (Table 1). Whether this difference is generated by a heavily glycosylated subset of the ASM or if it reflects a general trend of most ASM proteins being glycosylated remains unknown. We also cannot determine whether there are any spatial biases within the adult mantle tissue with regards to the location of glycosylated proteins. The high percentage of glucosamine identified in AIM and ASM suggests that chitin, or its deacetylated derivative chitosan, is present in both extracts, but this hypothesis requires further testing. Despite their likely importance to the functional mechanisms of shell formation, post-translational modifications of molluscan shell-forming proteins remain relatively understudied, and we predict that research efforts in these directions would yield interesting functional insights into the mechanisms of shell fabrication.

### Conclusion

By characterizing the spatial expression patterns of 34 genes associated with shell formation, we have revealed patterns of asymmetry that presumably contribute to the coiled phenotype of *Lymnaea*'s shell. Our broad survey of these genes in the adult mantle tissue also highlight the morphological modularity of

this phylum-specific organ and allude to an explanation as to why the Mollusca have evolved so many successful shell morphologies. While gene co-option, domain shuffling, and gene family expansion are mechanisms that have clearly contributed to the great diversity of molluscan shell-forming proteins, our analyses also suggest that alternative splicing acts to significantly expand the shell-forming molecular repertoire. Comparing the results of spatial gene expression surveys focused on shell-formation from a broad range of molluscan taxa will shed further light on the evolutionary story of this fascinating structure.

### Availability of supporting data

All raw NGS data has been deposited with the SRA with BioSample accession numbers SAMN08117214, SAMN08117215, SAMN08709370, SAMN08709371, SAMN08709372, SAMN08709373, SAMN08709374, SAMN08709375, SAMN08709376, and SAMN08709377. Individual image files for the *in situ* hybridization gene expression patterns and the sense strand cDNA sequences used to generate the *in situ* hybridization riboprobes can be accessed from the associated GigaDB repository [45]. The mantle transcriptome assemblies are also available via GigaDB [45] (file names: C2844\_CLC.idba.Trinity\_for\_annotation.fasta and C2845\_CLC.idba.Trinity\_for\_annotation.fasta). All MS data have been deposited with the ProteomeXchange Consortium with the dataset identifiers PXD008547 and 10.6019/PXD008547. Other supporting data are available from additional files, also including an extended description of the *in situ* hybridization results (see additional file 47).

### Additional files

- Additional file 1.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_1.xlsx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_1.xlsx). Summarized results of MASCOT searches.
- Additional files 2–35.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files). Whole mount *in situ* hybridisation results and molecular features of 34 shell-forming gene candidates.
- Additional file 36.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_36.pdf](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_36.pdf). A more comprehensive summary of the results presented in Fig. 3.
- Additional file 37.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_37.xlsx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_37.xlsx). Detailed table of the molecular features of all shell-forming protein candidates.
- Additional file 38.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_38.xlsx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_38.xlsx). Results of repetitive motif searches using Xstream.
- Additional file 39.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_39.txt](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_39.txt). A FASTA formatted file of the 879 protein sequences used to construct Fig. 4.
- Additional file 40.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_40.xlsx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_40.xlsx). Detailed results of tBLASTx similarity searches for all shell-forming candidate genes against nr database.
- Additional file 41.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_41.xlsx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_41.xlsx). Detailed results of protein family and protein domain

similarity searches for all shell-forming candidate genes against CDD database.

**Additional file 42.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_42.xlsx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_42.xlsx). Detailed results of BLAST similarity searches for all shell-forming candidate genes against SwissProt database.

**Additional file 43.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_43.xlsx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_43.xlsx). Lineages for all shell-forming candidates that returned positive BLAST results.

**Additional file 44.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_44.txt](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_44.txt). Nucleotide sequences of 34 families of shell forming candidate genes.

**Additional file 45.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_45.txt](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_45.txt). Translated sequences of 34 families of shell forming candidate genes.

**Additional file 46.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_46.txt](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_46.txt). mRNA regions targeted by riboprobes.

**Additional file 47.** [ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS\\_supplemental\\_files/Additional\\_file\\_47.docx](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001.101000/100436/MS_supplemental_files/Additional_file_47.docx). Extended Results and Discussion.

## Abbreviation

AIM: acid-insoluble matrix; ASM: acid-soluble matrix; BLAST: Basic Local Alignment Search Tool; CNS: central nervous system; dpfc: days post first cleavage; GRN: gene regulatory network; hpfc: hours post first cleavage; IGF-BP: insulin-like growth factor-binding protein; LC: Liquid chromatography; MS: mass spectrometry; NGS: next-generation sequencing; PCR: polymerase chain reaction; RLCD: repetitive low-complexity domain; SEM: Scanning Electron Microscopy; SRA: Sequence Read Archive; UTR: untranslated region.

## Competing interests

The authors declare that they have no competing interests.

## Author contributions

I.H. carried out the molecular work, bioinformatic analyses, and co-wrote and drafted the manuscript. F.M. and B.M. performed the proteomic analyses and drafted the manuscript. D.J.J. conceived and supervised the study, contributed to the molecular work, contributed to the bioinformatic analyses, performed the histological sections, and co-wrote and drafted the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We are grateful to Wolfgang Dröse for assistance and advice with *in situ* sectioning, Isabelle Zanella-Cleon from IBCP (Lyon) for MS analysis, and Jennifer Hohagen and Dorothea Hause-Reitner who generated SEM images. Illumina sequencing was performed by Markus B. Schilhabel and his team at the Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel. This work was funded by DFG (JA 2108/2-1 and JA 2108/6-1) and VolkswagenStiftung (92075) grants to D.J.J.

## References

- Mao L-B, Gao H-L, Yao H-B, et al. Synthetic nacre by predesigned matrix-directed mineralization. *Science* 2016;**354**:107–10.
- Kniprath E. Ontogeny of the molluscan shell field: a review. *Zool Script* 1981;**10**:61–79.
- Jackson DJ, Degnan BM. The importance of Evo-Devo to an integrated understanding of molluscan biomineralisation. *J Struct Biol* 2016;**196**:67–74.
- Okabe T, Yoshimura J. Optimal designs of mollusk shells from bivalves to snails. *Sci Rep* 2017;**7**:42445.
- Aguilera F, McDougall C, Degnan BM. Co-option and de novo gene evolution underlie molluscan shell diversity. *Mol Biol Evol* 2017;**34**:779–92.
- Liang J, Xie J, Gao J et al. Identification and characterization of the lysine-rich matrix protein family in *Pinctada fucata*: indicative of roles in shell formation. *Mar Biotech* 2016;**18**:645–58.
- Marxen JC, Hammer M, Gehrke T, et al. Carbohydrates of the organic shell matrix and the shell-forming tissue of the snail *Biomphalaria glabrata* (Say). *Biol Bull* 1998;**194**:231–40.
- Arias JL, Fernández MS. Polysaccharides and proteoglycans in calcium carbonate-based biomineralization. *Chem Rev* 2008;**108**:4475–82.
- Farre B, Dauphin Y. Lipids from the nacreous and prismatic layers of two Pteriomorpha mollusc shells. *Comp Biochem Physiol B* 2009;**152**:103–9.
- Zhao M, He M, Huang X, et al. Functional characterization and molecular mechanism exploration of three granulin epithelin precursor splice variants in biomineralization of the pearl oyster *Pinctada fucata*. *Mol Genet Genom* 2016;**291**:399–409.
- Wang J, Gao J, Xie J, et al. Cloning and mineralization-related functions of the calponin gene in *Chlamys farreri*. *Comp Biochem Physiol B* 2016;**201**:53–58.
- Yarra T, Gharbi K, Blaxter M, et al. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. *Mar Geonomics* 2016;**27**:9–15.
- Gao P, Liao Z, Wang X-X, et al. Layer-by-layer proteomic analysis of *Mytilus galloprovincialis* shell. *PLoS One* 2015;**10**:e0133913.
- Mann K, Jackson DJ. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genom* 2014;**15**:249.
- Marie B, Jackson DJ, Ramos-Silva P, et al. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. *FEBS J* 2013;**280**:214–32.
- Jackson DJ, McDougall C, Green K, et al. A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol* 2006;**4**:40.
- Marie B, Marie A, Jackson DJ, et al. Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. *Prot Sci* 2010;**8**:54.
- Zhao M, He M, Huang X, et al. A homeodomain transcription factor gene, *PfMSX*, activates expression of *Pif* gene in the pearl oyster *Pinctada fucata*. *PLoS One* 2014;**9**:e103830.
- Suzuki M, Saruwatari K, Kogure T, et al. An acidic matrix protein, *Pif*, is a key macromolecule for nacre formation. *Science* 2009;**325**:1388–90.
- Jackson DJ, Wörheide G, Degnan BM. Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. *BMC Evol Biol* 2007;**7**:160.
- Dauphin Y, Denis A. Structure and composition of the aragonitic crossed lamellar layers in six species of Bivalvia and Gastropoda. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 2000;**126**:367–77.
- de Paula SM, Silveira M. Studies on molluscan shells: contri-

- butions from microscopic and analytical methods. *Micron* 2009;**40**:669–90.
23. Almagro I, Drzymala P, Berent K, et al. New crystallographic relationships in biogenic aragonite: the crossed-lamellar microstructures of mollusks. *Crystal Growth & Design* 2016.
  24. Kuhn-Spearing LT, Kessler H, Chateau E, et al. Fracture mechanisms of the *Strombus gigas* conch shell: implications for the design of brittle laminates. *J Mater Sci* 1996;**31**:6583–94.
  25. Kamat S, Su X, Ballarini R, et al. Structural basis for the fracture toughness of the shell of the conch *Strombus gigas*. *Nature* 2000;**405**:1036–40.
  26. Pokroy B, Zolotoyabko E. Microstructure of natural plywood-like ceramics: a study by high-resolution electron microscopy and energy-variable X-ray diffraction. *J Mater Chem* 2003;**13**:682–8.
  27. Rodriguez-Navarro AB, Checa A, Willinger M-G, et al. Crystallographic relationships in the crossed lamellar microstructure of the shell of the gastropod *Conus marmoreus*. *Acta Biomater* 2012;**8**:830–5.
  28. Marin F, Luquet G, Marie B, et al. Molluscan shell proteins: primary structure, origin, and evolution. *Curr Top Dev Biol* 2008;**80**:209–76.
  29. Pavat C, Zanella-Cléon I, Becchi M, et al. The shell matrix of the pulmonate land snail *Helix aspersa maxima*. *Comp Biochem Physiol B* 2012;**161**:303–14.
  30. Vizcaíno JA, Csordas A, Del-Toro N, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 2016;**44**:D447–56.
  31. Jackson DJ, Ellemor N, Degnan BM. Correlating gene expression with larval competence, and the effect of age and parentage on metamorphosis in the tropical abalone *Haliotis asinina*. *Mar Biol* 2005;**147**:681–97.
  32. Gasteiger E, Gattiker A, Hoogland C, et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nuc Acids Res* 2003;**31**:3784–8.
  33. Petersen TN, Brunak S, von Heijne G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth* 2011;**8**:785–6.
  34. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. Humana Press. Totowa, NJ; 2005.
  35. Jorda J, Kajava AV. T-REKS: identification of tandem REpeats in sequences with a K-meanS based algorithm. *Bioinformatics* 2009;**25**:2632–8.
  36. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.
  37. Marchler-Bauer A, Derbyshire MK, Gonzales NR, et al. CDD: NCBI's conserved domain database. *Nuc Acids Res* 2015;**43**:D222–6.
  38. Mitchell A, Chang H-Y, Daugherty L, et al. The InterPro protein families database: the classification resource after 15 years. *Nuc Acids Res* 2015;**43**:D213–21.
  39. Steentoft C, Vakhrushev SY, Joshi HJ, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* 2013;**32**:1478–88.
  40. Gene Structure Draw. [<http://www.compgen.uni-muenster.de/tools/strdraw/index.hbi?>]. Accessed 10th April 2018
  41. Davison A, McDowell GS, Holden JM, et al. Formin is associated with left-right asymmetry in the pond snail and the frog. *Curr Biol* 2016;**26**:654–60.
  42. Kapustin Y, Tózsér J, Souvorov A, et al. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biology Direct* 2008;**3**:20.
  43. Yan T, Yoo D, Berardini TZ, et al. PatMatch: a program for finding patterns in peptide and nucleotide sequences. *Nuc Acids Res* 2005;**33**:W262–6.
  44. Cerveau N, Jackson DJ. Combining independent de novo assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms. *BMC Bioinformatics* 2016;**17**:525.
  45. Herlitze I, Marie B, Marin F, et al. Supporting data for “Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas.” *GigaScience Database* 2018. <http://dx.doi.org/10.5524/100436>.
  46. Jackson DJ, Herlitze I, Hohagen J. A whole mount in situ hybridization method for the gastropod mollusc *Lymnaea stagnalis*. *J Vis Exp* 2016;(109). doi: 10.3791/53968.
  47. Marie B, Joubert C, Tayaléa A, et al. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *Proc Natl Acad Sci* 2012;**109**:20986–91.
  48. Mann K, Cerveau N, Gummich M, et al., In-depth proteomic analyses of *Haliotis laevis* (greenlip abalone) nacre and prismatic organic shell matrix, *Proteome Science*, 2018. Submitted.
  49. Mann K, Edsinger E. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. *Prot Sci* 2014;**12**:28.
  50. Feng D, Li Q, Yu H, et al. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. *Sci Rep* 2017;**7**:45754.
  51. Arivalagan J, Marie B, Sleight VA, et al. Shell matrix proteins of the clam, *Mya truncata*: roles beyond shell formation through proteomic study. *Mar Geomics* 2016;**27**:69–74.
  52. Liu C, Li S, Kong J, et al. In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Sci Rep* 2015;**5**:17269.
  53. Liao Z, Bao L-F, Fan M-H, et al. In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *J Proteomics* 2015;**122**:26–40.
  54. Jackson DJ, Mann K, Häussermann V, et al. The *Magellania venosa* biomineralizing proteome: a window into brachiopod shell evolution. *Gen Biol Evol* 2015;**7**:1349–62.
  55. Mann K, Poustka AJ, Mann M. In-depth, high-accuracy proteomics of sea urchin tooth organic matrix. *Prot Sci* 2008;**6**:33.
  56. Ramos-Silva P, Kaandorp J, Herbst F, et al. The skeleton of the staghorn coral *Acropora millepora*: molecular and structural characterization. *PLoS One* 2014;**9**:e97454.
  57. Hohagen J, Jackson DJ. An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. *BMC Dev Biol* 2013;**13**:27.
  58. Kniprath E. Zur Ontogenese des Schalenfeldes von *Lymnaea stagnalis*. *Wilhelm Roux's Archives of Developmental Biology* 1977;**181**:11–30.
  59. Timmermans LPM. Studies on shell formation in molluscs. *Neth J Zool* 1969;**19**:413–523.
  60. Saleuddin ASM. An electron microscopic study on the formation of the periostracum in *Helisoma* (Mollusca). *Calcif Tissue Int* 1975;**18**:297–310.
  61. Kniprath E. Formation and structure of the periostracum in *Lymnaea stagnalis*. *Cal Tiss Res* 1972;**9**:260–71.
  62. Bubel A. An electron-microscope study of periostracum formation in some marine bivalves. I. The origin of the periostracum. *Mar Biol* 1973;**20**:213–21.
  63. Bevelander G, Nakahara H. An electron microscope study of the formation of the periostracum of *Macrocallista maculata*.

- Cal Tiss Res 1967;1:55–67.
64. Bubel A. An electron-microscope study of periostracum formation in some marine bivalves. II. The cells lining the periostracal groove. *Mar Biol* 1973;20:222–34.
  65. Ast G. How did alternative splicing evolve. *Nat Rev Gen* 2004;5:773–82.
  66. Weiner S, Traub W, Parker SB. Macromolecules in mollusc shells and their functions in biomineralization [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences* 1984;304:425–34.
  67. Addadi L, Weiner S. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. *Proc Natl Acad Sci* 1985;82:4110–4.
  68. Kingsley EP, Chan XY, Duan Y, et al. Widespread RNA segregation in a spiralian embryo. *Evol Dev* 2007;9:527–39.
  69. Lambert JD, Nagy LM. Asymmetric inheritance of centrosomally localized mRNAs during embryonic cleavages. *Nature* 2002;420:682–6.
  70. Bartel DP, Chen C-Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Gen* 2004;5:396–400.
  71. Zheng Z, Jiao Y, Du X, et al. Computational prediction of candidate miRNAs and their potential functions in biomineralization in pearl oyster *Pinctada martensii*. *Saudi Journal of Biological Sciences* 2016;23:372–8.
  72. Jiao Y, Zheng Z, Tian R, et al. MicroRNA, pm-miR-2305, participates in nacre formation by targeting pearl in pearl oyster *Pinctada martensii*. *Int J Molec Sci* 2015;16:21442–53.
  73. Brusca RC, Brusca GJ. *Invertebrates*. 2nd edition. Sinauer Associates: Sunderland, Massachusetts; 2002.
  74. Rosenberg G. A new critical estimate of named species-level diversity of the recent Mollusca. *Am Malacol Bull* 2014;32:308–22.
  75. Hochner B, Glanzman DL. Evolution of highly diverse forms of behavior in molluscs. *Curr Biol* 2016;26:R965–71.
  76. Williams TA, Nagy LM. Developmental modularity and the evolutionary diversification of arthropod limbs. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* 2001;291:241–57.
  77. Peel AD, Chipman AD, Akam M. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat Rev Gen* 2005;6:905–16.
  78. Tautz D. Segmentation. *Dev Cell* 2004;7:301–12.
  79. Raff RA. *The shape of life: genes, development, and the evolution of animal form*. University of Chicago Press: Chicago, Illinois; 1996.
  80. Esteve-Altava B. In search of morphological modules: a systematic review. *Biol Rev* 2017;92:1332–47.
  81. Eble GJ. Morphological modularity and macroevolution. In: Callebaut W, Rasskin-Gutman D, eds. *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. MIT Press, Cambridge, Massachusetts; 2005. pp. 221–38.
  82. McDougall C, Green K, Jackson DJ, et al. Ultrastructure of the mantle of the gastropod *Haliotis asinina* and mechanisms of shell regionalization. *Cells Tissues Organs* 2011;194:103–7.
  83. Jolly C, Berland S, Milet C, et al. Zonal localization of shell matrix proteins in mantle of *Haliotis tuberculata* (Mollusca, Gastropoda). *Mar Biotech* 2004;6:541–51.
  84. Sleight VA, Marie B, Jackson DJ, et al. An Antarctic molluscan biomineralisation tool-kit. *Sci Rep* 2016;6:36978.
  85. Waite JH. The phylogeny and chemical diversity of quinone-tanned glues and varnishes. *Comp Biochem Physiol B* 1990;97:19–29.
  86. Waite JH. Evidence for the mode of sclerotization in a molluscan periostracum. *Comp Biochem Physiol B* 1977;58:157–62.
  87. Waite JH, Wilbur KM. Phenoloxidase in the periostracum of the marine bivalve *Modiolus demissus* Dillwyn. *J Exp Zool* 1976;195:359–67.
  88. Waite JH. Quinone-tanned scleroproteins. In: Saleuddin ASM, Wilbur KM, eds. *Physiology*. Academic Press: New York; 1983. pp. 467–504.
  89. Michon T, Chenu M, Kellershon N, et al. Horseradish peroxidase oxidation of tyrosine-containing peptides and their subsequent polymerization: a kinetic study. *Biochem* 1997;36:8504–13.
  90. Halaby DM, Mornon JPE. The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. *J Mol Evol* 1998;46:389–400.
  91. Honegger B, Galic M, Köhler K, et al. Imp-L2, a putative homolog of vertebrate IGF-binding protein 7, counteracts insulin signaling in *Drosophila* and is essential for starvation resistance. *Journal of Biology* 2008;7:1.
  92. Dogterom AA, Doderer A. A hormone dependent calcium-binding protein in the mantle edge of the freshwater snail *Lymnaea stagnalis*. *Calcif Tissue Int* 1981;33:505–8.
  93. Dogterom AA, Jentjens T. The effect of the growth hormone of the pond snail *Lymnaea stagnalis* on periostracum formation. *Comparative Biochemistry and Physiology Part A: Physiology* 1980;66:687–90.
  94. Dogterom AA, van Loenhout H, van der Schors RC. The effect of the growth hormone of *Lymnaea stagnalis* on shell calcification. *Gen Comp Endocrinol* 1979;39:63–68.
  95. Dogterom AA, van der Schors RC. The effect of the growth hormone of *Lymnaea stagnalis* on (bi) carbonate movements, especially with regard to shell formation. *Gen Comp Endocrinol* 1980;41:334–9.
  96. Weiss IM, Göhring W, Fritz M, et al. Perlustrin, a *Haliotis laevigata* (abalone) nacre protein, is homologous to the insulin-like growth factor binding protein N-terminal module of vertebrates. *Biochem Biophys Res Commun* 2001;285:244–9.
  97. Zhang G, Willemin AS, Brion A, et al. A new method for the separation and purification of the osteogenic compounds of nacre ethanol soluble matrix. *J Struct Biol* 2016;196:127–37.
  98. Marin F, Luquet G. Molluscan shell proteins. *CR Palevol* 2004;3:469–92.
  99. Shen X, Belcher AM, Hansma PK, et al. Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *J Biol Chem* 1997;272:32472–81.
  100. Kocot KM, Aguilera F, McDougall C, et al. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. *Front Zool* 2016;13:23.
  101. Luo H, Nijveen H. Understanding and identifying amino acid repeats. *Brief Bioinform* 2014;15:582–91.
  102. Albà M, Tompa P, Veitia R. Amino acid repeats and the structure and evolution of proteins. In: Volff J-N (ed) *Basel, Karger. Gene and Protein Evolution*. Karger Publishers; 2007. pp. 119–30.
  103. Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 5th revised edition. Garland Science: New York; 2007.
  104. McDougall C, Woodcroft BJ, Degnan BM. The widespread prevalence and functional significance of silk-like structural proteins in metazoan biological materials. *PLoS One* 2016;11:e0159128.
  105. Le Roy N, Jackson D, Marie B, et al. Carbonic anhydrase and

- metazoan biocalcification: a focus on molluscs. In: Marin F, Brümmer F, Checa A, Furtos G, Lesci IG, Šiller L, eds. *Biomaterialization: From Fundamentals to Biomaterials & Environmental Issues*. Trans Tech Publications Ltd; 2015. pp. 151–7.
106. Sun X, Yang A, Wu B, et al. Characterization of the mantle transcriptome of Yesso scallop (*Patinopecten yessoensis*): identification of genes potentially involved in biomineralization and pigmentation. *PLoS One* 2015;**10**:e0122967.
  107. Jackson DJ, McDougall C, Woodcroft B, et al. Parallel evolution of nacre building gene sets in molluscs. *Mol Biol Evol* 2010;**27**:591–608.
  108. Sleight VA, Thorne MAS, Peck LS, et al. Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam, *Mya truncata*. *Mar Geonomics* 2016;**27**:47–55.
  109. Arivalagan J, Yarra T, Marie B, et al. Insights from the shell proteome: biomineralization to adaptation. *Mol Biol Evol* 2017;**34**(1):66–77.
  110. Suzuki M, Kogure T, Weiner S, et al. Formation of aragonite crystals in the crossed lamellar microstructure of limpet shells. *Crystal Growth & Design* 2011;**11**:4850–9.
  111. Mann K, Edsinger-Gonzales E, Mann M. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. *Prot Sci* 2012;**10**:28.