# Variation in Orthologous Shell-Forming Proteins Contribute to Molluscan Shell Diversity

Daniel J. Jackson,*[,1,2] Laurin Reim,[3] Clemens Randow,[4] Nicolas Cerveau,[1] Bernard M. Degnan,[2] and Claudia Fleck[4]

[1]Department of Geobiology, Georg-August University of Göttingen, Göttingen, Germany

[2]School of Biological Sciences, University of Queensland, Brisbane, Australia

[3]Department of Earth- and Environmental Sciences, Ludwig-Maximilian University of Munich, München, Germany

[4]Department of Materials Engineering, Institute of Technology Berlin, Berlin, Germany

***Corresponding author:** E-mail: djackso@uni-goettingen.de.

**Associate editor:** Patricia Wittkopp

## Abstract

**Despite the evolutionary success and ancient heritage of the molluscan shell, little is known about the molecular details of its formation, evolutionary origins, or the interactions between the material properties of the shell and its organic constituents. In contrast to this dearth of information, a growing collection of molluscan shell-forming proteomes and transcriptomes suggest they are comprised of both deeply conserved, and lineage specific elements. Analyses of these sequence data sets have suggested that mechanisms such as exon shuffling, gene co-option, and gene family expansion facilitated the rapid evolution of shell-forming proteomes and supported the diversification of this phylum specific structure. In order to further investigate and test these ideas we have examined the molecular features and spatial expression patterns of two shell-forming genes (*Lustrin* and *ML1A2*) and coupled these observations with materials properties measurements of shells from a group of closely related gastropods (abalone). We find that the prominent "GS" domain of Lustrin, a domain believed to confer elastomeric properties to the shell, varies significantly in length between the species we investigated. Furthermore, the spatial expression patterns of *Lustrin* and *ML1A2* also vary significantly between species, suggesting that both protein architecture, and the regulation of spatial gene expression patterns, are important drivers of molluscan shell evolution. Variation in these molecular features might relate to certain materials properties of the shells of these species. These insights reveal an important and underappreciated source of variation within shell-forming proteomes that must contribute to the diversity of molluscan shell phenotypes.**

***Key words:*** **biomineralization, gene expression, evolution, mollusc, Lustrin, nacre, materials properties.**

## Introduction

The calcified molluscan shell serves a wide variety of functions and has done so for the majority of the evolutionary history of the Mollusca. The functional diversity of this evolutionarily successful structure is determined not only by its overall morphology, but also by the materials properties of the bio-ceramic from which it is constructed. These properties are conferred to the shell by an interaction between the mineral phase and a suite of biomolecules that are well known to differ significantly between species (Bédouet et al. 2001; Cartwright and Checa 2007; Farre and Dauphin 2009; Clark et al. 2010; Jackson et al. 2010a; Marin et al. 2012; Mann and Jackson 2014). Careful study of the proteins incorporated into the shell during its fabrication can provide deep insight into how these molecules bestow materials properties to the final structure that far exceed that of pure $CaCO_3$ (for example Chang and Evans 2015). It is also feasible to study these shell-forming proteins to infer both the evolutionary origins of the biological processes that build these structures, and the forces that drove their evolution (Jackson et al. 2010b, 2011; Zheng et al. 2015; Jackson and Degnan 2016; Arivalagan et al. 2017).

A recent survey of conciferan molluscs concluded that extensive gene co-option, lineage-specific gene family expansion and domain shuffling generate much of the diversity that can be observed in secreted, protein-coding mantle mRNAs (Aguilera et al. 2017). Using a similar approach, a sequence based similarity comparison of molluscan shell-forming proteins identified several highly conserved protein domains (for example, carbonic anhydrase, chitin binding, Ig-like, von Willenbrand) within the shells of a phylogenetically diverse collection of conchiferan molluscs (Feng et al. 2017). These kinds of broad in silico sequence-based comparisons have now been performed for many shell-forming data sets (Jackson et al. 2006, 2015; Mann and Jackson 2014; Arivalagan et al. 2017), and while informative, they make several fundamental assumptions that should be tested. For example, are shell-forming proteins that share sequence similarity deployed from their respective mantle tissues similarly, and are they likely to be employed to the same effect? Here,

Article

we begin to address these assumptions by investigating the primary sequence architectures and spatial expression patterns of two protein coding genes associated with nacre formation from several abalone species. In addition, we characterize several materials properties of the nacreous layer from a selection of these shells and interpret these differences within the context of our molecular investigations.

Nacre is a biomineral microstructure that has long received attention from scientists interested in a variety of its properties and mode of formation (Nakahara 1991; Lin and Meyers 2005; Bezares et al. 2008; Gilbert et al. 2008), however, much remains unknown. For example, Wang and Gupta (2011) highlight that little is understood of the interactions between nacre tablets and the immediate protein layers that affect its fracture deflection and deformation properties. A significant challenge for the field of biomineralogy is to integrate insights gained from the growing number of molecular surveys focused on the biomolecules that generate a biomineral, with an understanding of the materials properties of that biomineral. For molluscan shells this is challenging because the tools required to fully and accurately functionally characterize a specific shell-forming protein in vivo do not yet exist. One of the very first gene products to be fully isolated and characterized from a molluscan shell on a molecular level was Lustrin-A (Shen et al. 1997). This protein, originally isolated from the nacreous shell layer of the red abalone *Haliotis rufescens*, has several distinguishing features. The *H. rufescens* Lustrin protein is highly modular with ten cysteine-rich domains interspersed by eight proline-rich domains followed by a glycine/serine-rich domain. Using atomic force microscopy, Smith et al. (1999) demonstrated that when individual tablets of *H. rufescens* nacre are pulled apart the resulting force-extension curves display a characteristic saw tooth appearance with hysteretic recovery. This material property has been associated with the GS-rich domain of Lustrin where looped stacks of glycine and serine interspersed with aromatic residues may confer extensor-like properties to the mature molecule (Shen et al. 1997; Smith et al. 1999). In addition, Zhang et al. (2002) proposed that the N-terminal region of the ten C-rich domains may also confer this elastomeric-like property to the shell, while Wustman et al. (2002) suggest that the "RSKY" and "D4" regions may be responsible for enabling Lustrin-mineral or Lustrin-macromolecule interactions. However, these authors highlight the lack of evidence that can link Lustrin directly with these materials properties observations. In our previous efforts to characterize the shell-forming proteome of the tropical abalone (*Haliotis asinina*), we isolated a fragment of *Has-Lustrin* and another nacre-associated gene product, *Has-ML1A2* (Jackson et al. 2006). The derived *Has*-ML1A2 protein is spatially co-expressed with *Lustrin* in the proximal region of the outer mantle fold where nacre formation is thought to occur (Jackson et al. 2006, 2007).

Our desire to functionally characterize the major shell-forming components of molluscan shells led us to compare the primary structure and spatial expression of Lustrin and ML1A2 and the materials properties of the nacreous layer of three abalone species: the tropical abalone *H. asinina* (Linnaeus 1758); and temperate abalone species *Haliotis rubra* (Leach 1814) and *Haliotis laevigata* (Donovan 1808). These comparisons allow us to begin to test if orthologous shell-forming proteins confer similar materials properties to the shells of closely related species. We selected Lustrin because it has been relatively well studied, and is associated with a distinct region of the shell, the nacreous layer (Shen et al. 1997; Wustman et al. 2002, 2003a, 2003b; Zhang et al. 2002; Gaume et al. 2014). We also theorized that an advantage of using Lustrin in a multispecies comparison would be its modular nature as it should more readily reveal important molecular differences that could potentially be linked to differences in materials properties. While we cannot causally link interspecific differences in particular protein sequences or spatial gene expression patterns to differences in the materials properties of their respective shells, these differences in materials properties must ultimately be explained by interactions between biomolecules such as Lustrin and ML1A2 and the mineral phase.

## Results and Discussion

### Lustrin Sequences

Cloning full length *Lustrin* transcripts from all haliotid species proved to be challenging. This was most likely due to the sequence features of *Lustrin* that make it so interesting; repetitive domains, low sequence complexity (in particular the GS domain), and its overall size (schematically represented in fig. 1). Indeed, Shen et al. (1997) in reporting the original *Lustrin* sequence for *H. rufescens* detected two *Lustrin* bands by northern blot (with sizes of 5.5 and 4.7 kb), and could only isolate a single sequence of 4.4 kb. We therefore wished to understand whether we were likely to have isolated full length *Lustrin* transcripts using our RACE methodology from all of the species we investigated. To this end we also performed next generation sequencing (NGS) using the Illumina HiSeq 2500 platform on adult mantle tissue RNA for *H. asinina* and *H. rubra* (data not shown). Despite collecting 137,903,588 and 140,416,586 100 bp paired end reads for *H. asinina* and *H. rubra* respectively, and employing a range of bioinformatic assembly parameters and algorithms (Cerveau and Jackson 2016), we were unable to assemble any *Lustrin* contig longer (and in most cases significantly shorter) than the "manually" RACE-isolated molecules reported here. We attribute this NGS assembly difficulty to the repetitive and low-complexity nature of *Lustrin*. In contrast, our NGS assemblies yielded contigs coding for full length ML1A2 protein sequences with 100% sequence similarity to those we isolated using the RACE methodology. We also performed northern blots against *Lustrin* for four species of abalone in order to determine whether we had isolated close to full length *Lustrin* gene products. Due to limitations in the amount of available sample material we could only perform Northern blots against *Lustrin* for the three commercially harvested species: *H. asinina*, *H. laevigata* and *H. rubra*. *Haliotis asinina* and *H. laevigata* shared the most similar Northern blot patterns with bands migrating at ~4,690 bp and 4,297 bp for *H. asinina*, and 4,600 bp and 4,290 for *H. laevigata* (fig. 2). *Haliotis rubra*

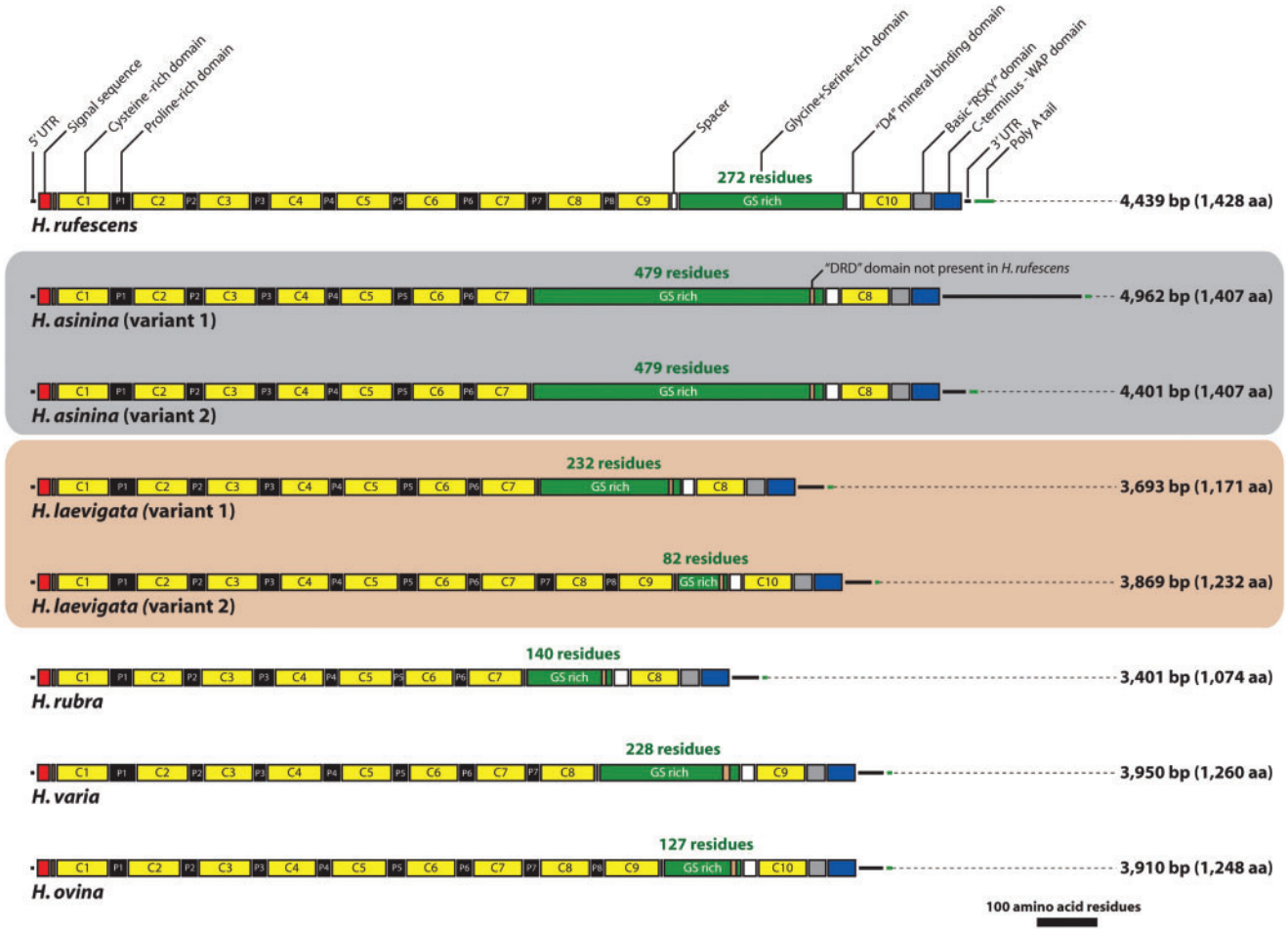


FIG. 1. A schematic overview of the primary architecture of Lustrin isoforms from six species of abalone. The first reported Lustrin sequence (Lustrin A from *Haliotis rufescens*) serves as a reference with the major domains and features labeled. Note the significantly longer "GS" domain in the two *Haliotis asinina* proteins compared with all other isoforms, and the significantly longer 3′ UTR in *H. asinina* transcript variant 1. All sequence lengths are to scale.

displayed the smallest *Lustrin* transcript at ~3,453 bp (fig. 2). Interestingly *H. rubra* appears to express only one *Lustrin* transcript. These patterns were consistent across multiple northern blot experiments with RNA isolated from multiple individuals. Northern blot results for *H. asinina* and *H. rubra* correspond well with the cloned sequence lengths (fig. 1 confer fig. 2), suggesting that these sequences are full length, however, for *H. laevigata* we have apparently not cloned the full length *Lustrin* transcripts even though we obtained molecules that encode complete open reading frames (ORFs).

Differences in the derived Lustrin protein sequences between species are striking. While the overall architecture of Lustrin in all cases was conserved, there are significant differences in the numbers of cysteine-rich and proline rich domains, and the lengths of the 3′ UTRs and the "GS" domains (fig. 1 and table 1). *Haliotis asinina* has a GS domain almost 3.5 times longer than that in *H. rubra*, and 1.8 times longer than that in the single *H. rufescens* sequence.

## ML1A2 Sequences

BLAST searches with ML1A2 protein sequences against nr, Swissprot and refseq_rna do not reveal significant similarity

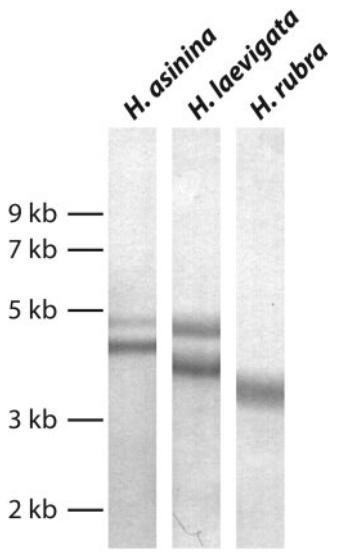


FIG. 2. Northern blot against *Lustrin* transcripts from three abalone species. Using a single probe with >85% homology to all *Lustrin* homologs, differences in both the number of *Lustrin* splice variants and/or paralogs and the sizes of these transcripts were revealed.

**Table 1.** Lustrin Domain Lengths (in Amino Acid Residues).

| Domain | *H. rufescens* | *H. asinina variant 1* | *H. asinina variant 2* | *H. laevigata variant 1* | *H. laevigata variant 2* | *H. rubra* | *H. varia* | *H. ovina* |
|---|---|---|---|---|---|---|---|---|
| Signal peptide | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 |
| N-terminus | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| C1 | 84 | 84 | 84 | 84 | 84 | 84 | 84 | 84 |
| P1 | 30 | 34 | 34 | 39 | 39 | 33 | 40 | 26 |
| C2 | 85 | 84 | 84 | 84 | 84 | 81 | 84 | 84 |
| P2 | 17 | 25 | 25 | 24 | 24 | 24 | 20 | 24 |
| C3 | 84 | 84 | 84 | 84 | 84 | 84 | 79 | 84 |
| P3 | 26 | 28 | 28 | 30 | 30 | 30 | 18 | 21 |
| C4 | 84 | 79 | 79 | 79 | 79 | 79 | 88 | 78 |
| P4 | 18 | 18 | 18 | 18 | 18 | 18 | 26 | 19 |
| C5 | 88 | 85 | 85 | 88 | 88 | 88 | 81 | 88 |
| P5 | 18 | 26 | 26 | 26 | 26 | 26 | 23 | 26 |
| C6 | 84 | 79 | 79 | 79 | 79 | 79 | 79 | 79 |
| P6 | 30 | 18 | 18 | 18 | 18 | 18 | 23 | 23 |
| C7 | 76 | 85 | 85 | 88 | 88 | 88 | 81 | 79 |
| P7 | 29 | N/A | N/A | N/A | 26 | N/A | 18 | 23 |
| C8 | 88 | N/A | N/A | N/A | 79 | N/A | 88 | 79 |
| P8 | 19 | N/A | N/A | N/A | 18 | N/A | N/A | 18 |
| C9 | 85 | N/A | N/A | N/A | 88 | N/A | N/A | 88 |
| Spacer 1 | 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| GS domain | 272 | 478 | 478 | 232 | 82 | 144 | 228 | 127 |
| Spacer 2 "D4" | 24 | 21 | 21 | 19 | 19 | 19 | 21 | 21 |
| C10 | 79 | 79 | 79 | 79 | 79 | 79 | 79 | 79 |
| Basic domain | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 28 |
| C-terminus | 45 | 45 | 45 | 45 | 45 | 45 | 45 | 45 |
| Total | 1428 | 1407 | 1407 | 1171 | 1232 | 1074 | 1260 | 1248 |

Domain lengths that conform to the consensus are shaded grey.

to any other sequences. A BLAST search against dbEST returned significant hits against two other abalone species (*Hediste diversicolor*, *Haliotis discus hannai*) and an EST reported from the peanut plant *Arachis hypogaea* (GenBank accession number JK197538). There are no recognizable domains in ML1A2 other than a signal sequence, leaving us with little information regarding the function of this protein. We previously reported the spatial expression of ML1A2 in *H. asinina* (Jackson et al. 2006) and found that it has the same spatial expression profile as *Lustrin*, suggesting that it is also involved in nacre formation. We were able to isolate putative full length *ML1A2* sequences from all five abalone species and found that they all shared high levels of similarity on the amino acid level (fig. 3).

## Spatial Expression Patterns of *Lustrin* and *ML1A2*

*Lustrin* transcripts are clearly present in the mantle tissue of juvenile abalone from all three species investigated (fig. 4). As expected, the dominant signal for all species is found in the proximal region of the outer mantle fold where nacre formation is thought to take place (fig. 4*B*, *G*, and *L* boxed regions and arrows). However, *Lustrin* expression was also present in the epithelia covering the digestive gland of *H. laevigata*. Interestingly this signal was absent in *H. asinina* and *H. rubra* (dashed ovals in fig. 4*B*, *G*, and *L*). We can only speculate that the function of *Lustrin* expression in this region of *H. laevigata* is to further thicken the nacreous layer in the posterior region of the shell, however, it is also possible that Lustrin plays an as yet unknown pleiotropic role in this location. Closer inspection of the mantle tissue also revealed further differences in the spatial expression of *Lustrin* between species. In *H. asinina* and *H. laevigata* a population of *Lustrin*$^{+}$ cells at the anterior-most edge of the outer mantle fold could be seen in whole mount preparations and in thick sections of this material (red arrows in fig. 4*C*, *D*, *E*, *M*, *N*, and *O*). In contrast these cells appear to be completely absent *H. rubra* (fig. 4*H*, *I*, and *J*).

*ML1A2* transcripts are also clearly present in the mantle tissues of all three species (fig. 5), but in contrast to *Lustrin*, there was no apparent expression of *ML1A2* in the epithelia covering the digestive gland of *H. laevigata* (fig. 5*B*, *G*, and *L*). As with *Lustrin*, all three species express *ML1A2* in the proximal region of the outer mantle fold where nacre deposition is thought to take place, however, there are subtle differences between species. *H. asinina* displays no *ML1A2*$^{+}$ cells anterior to the nacre forming region (fig. 5*D* and *E*), while in *H. rubra* some *ML1A2*$^{+}$ cells can be detected in this region (fig. 5*I*) and in *H. laevigata* there are relatively many cells in this region (green arrows in fig. 5*M* and *N*), which are clearly visible in sections of this material (green arrow in fig. 5*O*).

## Materials Properties

The overall length and thickness of *H. laevigata* and *H. rubra* shells are remarkably different to those of *H. asinina*. Notably, the shell of *H. asinina* is very thin in comparison to the shells of the temperate species (table 2). These shells were all derived from fully mature individuals. The differences in the thickness of these shells was in part what drove us to investigate the features of Lustrin between these species; could

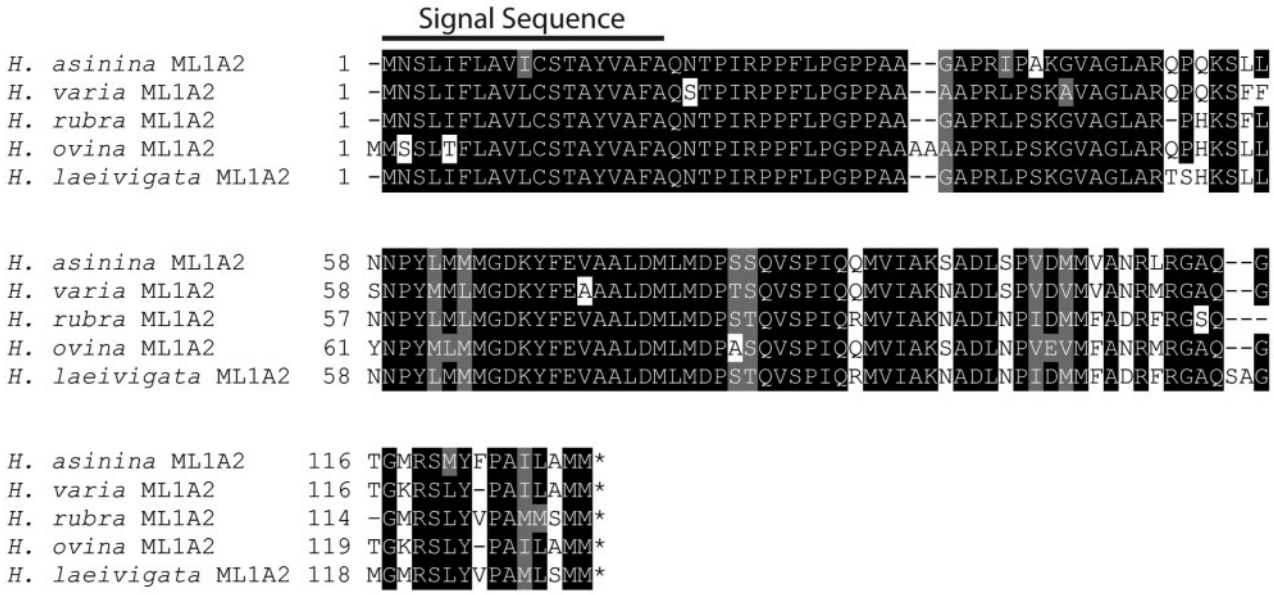


**Fig. 3.** Alignment of ML1A2 sequences derived from five species of abalone. The N-terminal signal sequence is indicated by a horizontal line. Residues with 100% identity are shaded black, residues with biochemical similarity are shaded grey, positions without conservation are unshaded, and gaps are represented by a dash.

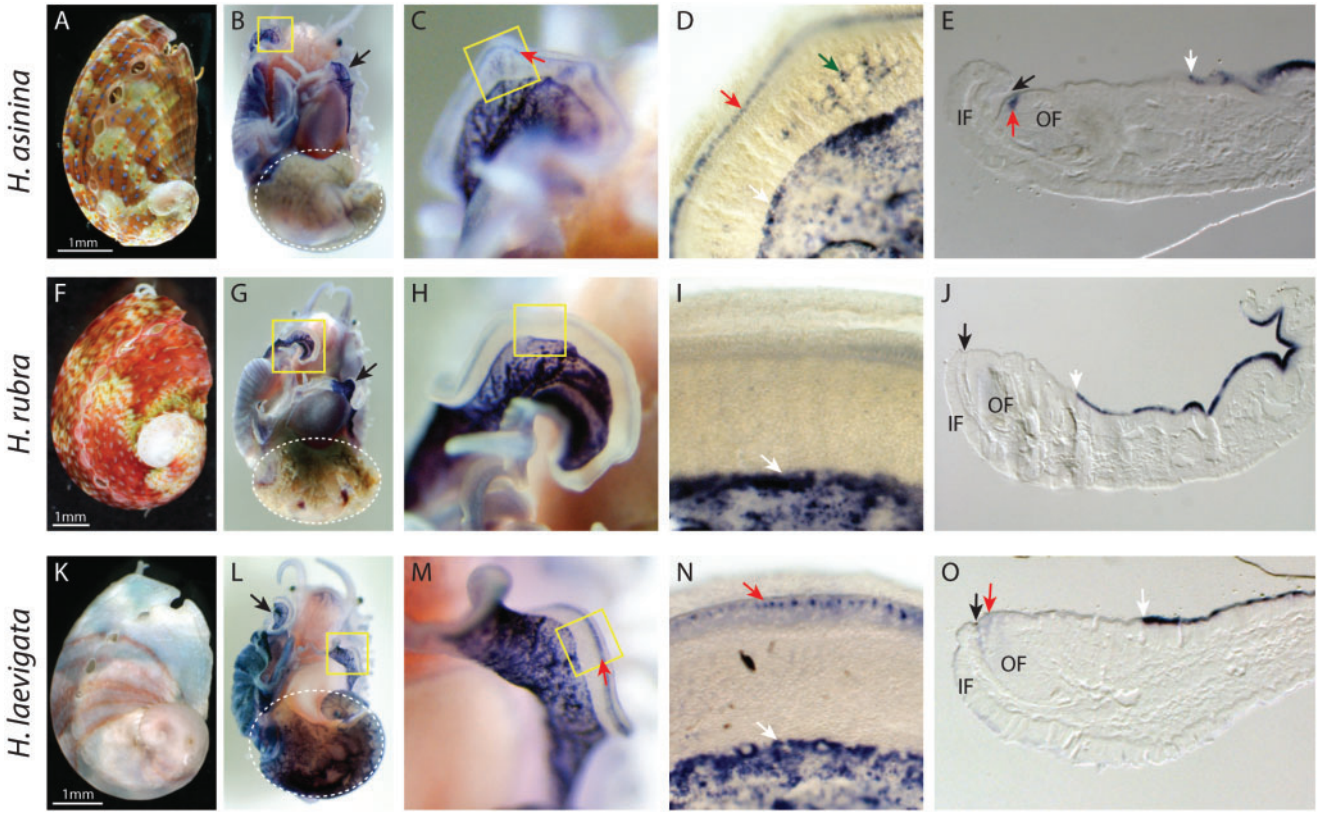


**Fig. 4.** WMISH in three abalone species reveals differences in the spatial expression of *Lustrin*. In all three species *Lustrin* is abundantly expressed in the proximal region of the outer mantle fold where nacre deposition is likely to take place. The boundary between these nacre-forming *Lustrin*$^{+}$ cells and *Lustrin*$^{-}$ cells is sharp (indicated by white arrows in *D–E*, *I–J* and *N–O*). In *Haliotis asinina* and *Haliotis rubra* there is a distinct lack of *Lustrin* signal in the epithelia overlying the digestive gland (dashed ovals in *B* and *G*), whereas in *Haliotis laevigata* this signal is strong (dashed oval in *L*). Additional differences across species exist in the mantle tissue where a line of *Lustrin*$^{+}$ cells can be seen in cells at the anterior edge of the outer mantle fold in *H. asinina* and *H. laevigata* (red arrows in *C–E*, *M–O*) but not in *H. rubra* (*H–J*). An additional population of *Lustrin*$^{+}$ cells in *H. asinina* can be found between the nacre forming region and the edge of the mantle (green arrow in *D*). The inner fold (IF) and outer fold (OF) of the mantle can be clearly seen in sections of the mantle tissue (*E*, *J*, and *O*), as well as the periostracal groove from which the periostracum is secreted (black arrows in *E*, *J*, and *O*). The anterior most limit of *Lustrin*$^{+}$ cells in the putative nacre forming region is indicated by white arrows (*D–E*, *I–J*, and *N–O*).

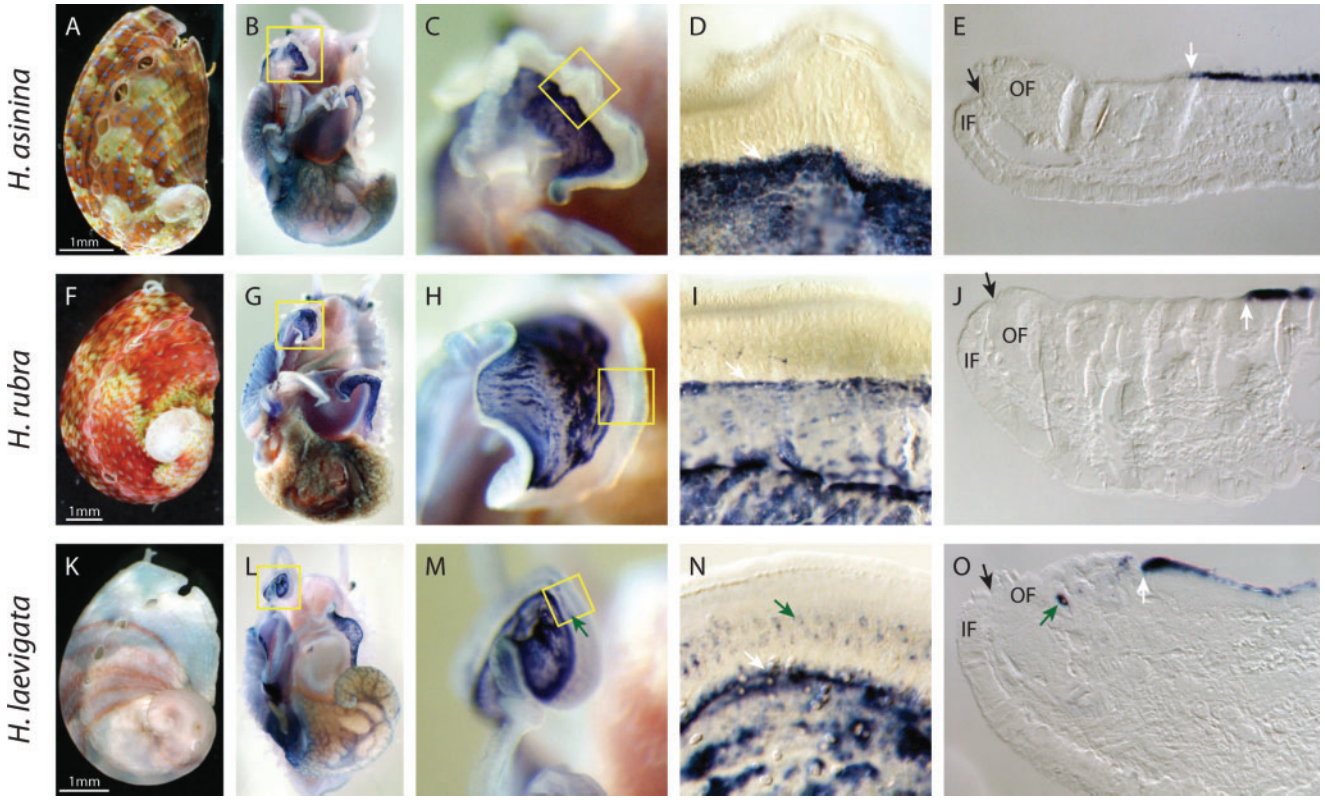


**Fig. 5.** WMISH in three abalone species reveals differences in the spatial expression of *ML1A2*. In all three species *ML1A2* is abundantly expressed in the proximal region of the outer mantle fold where nacre deposition is likely to take place. The boundary between these nacre-forming *ML1A2*$^{+}$ cells and *ML1A2*$^{-}$ cells is sharp (indicated by white arrows in *D–E*, *I–J*, and *N–O*). In *Haliotis asinina* and *Haliotis rubra* there is a lack of *ML1A2* signal in cells lying between the proximal nacre-forming region and the distal edge of the outer mantle fold, whereas in *Haliotis laevigata* a population of cells in this region are *ML1A2*$^{+}$ (green arrows in *4N–O*). The inner fold (IF) and outer fold (OF) of the mantle can be clearly seen in sections of the mantle tissue (*E*, *J*, and *O*), as well as the periostracal groove from which the periostracum is secreted (black arrows in *E*, *J*, and *O*).

**Table 2.** Shell Length and Thickness for the Samples Used in the Materials Properties Testing.

| | *H. rubra* | *H. laevigata* | *H. asinina* |
|---|---|---|---|
| **Shell length (mm)** | **90–100** | **90–110** | **60–80** |
| **Shell thickness (μm)** | **250–300** | **200–250** | **50–80** |

interspecific differences in the features of shell forming proteins potentially compensate for the significant differences in the thickness of these shells, and are there any differences in the materials properties of these shells? This line of molecular investigation is qualitative, as we have not quantitated the total organic content of these shells, nor the amounts of Lustrin or ML1A2 protein in the mantle or shell.

The nacreous microstructure was similar for all three abalone species, with a typical arrangement of nacre tablets stacked in columns (Cartwright and Checa 2007; Metzler et al. 2007; Gilbert et al. 2008; Checa et al. 2009). While we collected nano-indentation data for both the outer prismatic and the inner nacreous layers (supplementary figs. S1 and S3, Supplementary Material online) we focus here on the data obtained from the nacreous region (given our interest in Lustrin and ML1A2). Using both nanoindentation and three point bending experiments, we observed differences in the materials properties of nacre derived from the shells of the three abalone species we investigated. Because nacre is a biogenic material (and therefore contains a relatively high degree of heterogeneity) we were aware that it would be necessary to take multiple measurements of each property in order to achieve an impression of the overall materials property behavior for each shell. As expected, there is indeed considerable variation in all of the measurements we made, however, some interesting patterns are apparent.

The bending strength, fracture strain and Young's modulus as determined by three-point bending tests and the hardness properties from nanoindentation tests are presented in figure 6*A–D* respectively. Nacre derived from *H. asinina* appears to be able to withstand greater stresses than nacre derived from *H. rubra* or *H. laevigata*. Interestingly we could not detect any statistically significant differences in the fracture strain properties of any of the nacres we investigated (fig. 6*B*). Nacre from *H. asinina* was significantly stiffer than one of the *H. rubra* samples, however, this *H. rubra* sample also differed significantly from one of the other *H. rubra* samples indicating significant intraspecies variation in this material property (fig. 6*C*). Nanoindentation measurements revealed subtle but significant differences across all three species in the hardness of their nacres, with *H. asinina* displaying the hardest nacre and *H. laevigata* the softest (fig. 6*D*). We also found little difference in these values when surfaces were prepared from

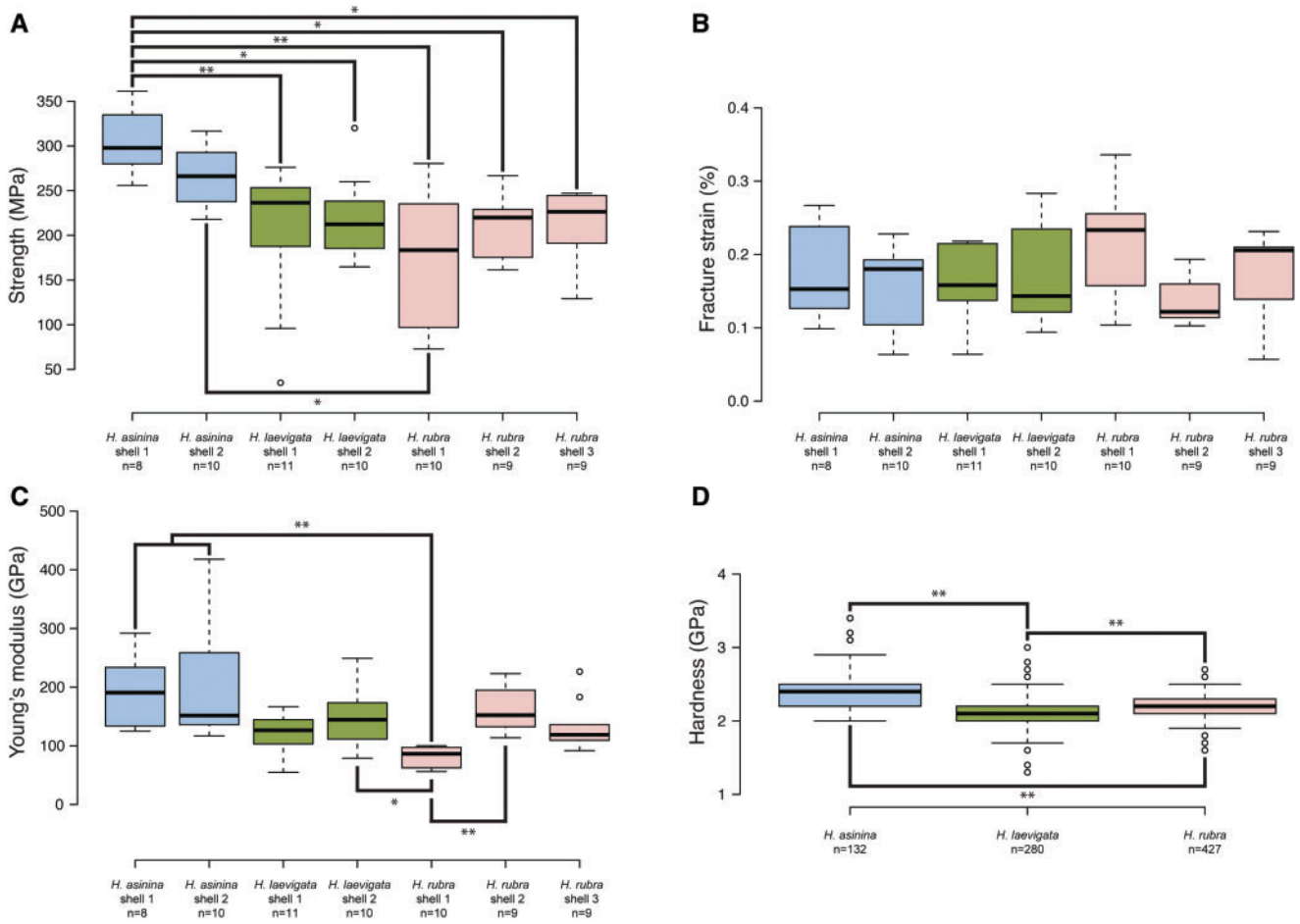


FIG. 6. Three point bending stress–strain responses and nanoindentation measures of hardness of nacre derived from *Haliotis laevigata*, *Haliotis rubra*, and *Haliotis asinina* shells. Centre lines of box plots show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots. Significant differences (as determined by Tukey post hoc tests) are indicated by bold black lines with differences at the $P = 0.05$ level represented by one asterisk, and $P = 0.01$ represented by two asterisks. The data for Young's modulus was log transformed to satisfy assumptions of normal data distribution and homogeneity of variance prior to statistical analysis, but is presented here untransformed. (*A*) *H. asinina* shell 1 displayed significantly higher bending strength values than all *H. rubra* and *H. laevigata* shells, with no differences detected between *H. rubra* and *H. laevigata* shells. (*B*) No significant differences were detected in the fracture strain values of any of the shells investigated. (*C*) Both *H. asinina* shells displayed significantly higher Young's moduli, as measured by three-point bending, than shell 1 from *H. rubra*, however, *H. rubra* shell 2 also exhibited a significantly higher Young's modulus than *H. rubra* shell 1 (as did *H. laevigata* shell 2). (*D*) The hardness values of all shells, derived from nanoindentation measurements, differed significantly from each other at $P = 0.01$ level.

either longitudinal or transverse faces of prepared nacre from *H. asinina*, suggesting that anisotropic effects contribute negligibly to these results.

It would be satisfying to correlate the patterns we observe in our data with the evolutionary relationships of the species we investigated. Unfortunately, the phylogenetic relationships of these abalone species are not fully or robustly resolved. For example, a study by Degnan et al. (2006) using the COII gene could not resolve the relationship between *H. asinina* and *H. laevigata* + *H. rubra*. Similarly, de Merwe (2012) combined mitochondrial NADH-dehydrogenase subunit 1 with hemocyanin to study the relationships of 18 abalone species and could also not resolve the relationship between *H. asinina* and *H. laevigata* + *H. rubra*. Despite this lack of phylogenetic resolution, it is clear *H. laevigata* and *H. rubra* share a more recent common ancestor than either does with *H. asinina*. This phylogenetic pattern is reflected in our measurements of shell strength with *H. rubra* and *H. laevigata* displaying equally strong shells, and one of the two *H. asinina* shells displaying a significantly stronger shell than the temperate species (fig. 6*A*). The ecological interactions and environments occupied by these three species have undoubtedly influenced the evolutionary histories of the shells they fabricate today. A challenge for the future is to identify the connections between these factors and specific materials properties of the shell and ultimately, with the genes that influence those properties.

It should also be noted that certain factors could not be controlled for in our experimental design. For example, none of our molecular analyses were quantitative. One factor that could therefore confound our materials properties comparisons would be differences in the amounts of Lustrin, ML1A2,

or other shell-forming proteins deposited in each shell. We are also assuming that proteins such as Lustrin and ML1A2 are uniformly distributed throughout the nacreous region of the shell, and that therefore our materials properties measurements are representative of the characteristics that these proteins confer to the shell. These assumptions could be addressed by raising antibodies against these proteins and then localizing and quantifying their expression profiles.

## Conclusions

We have identified striking interspecies differences in the primary sequence architecture of what is likely to be a critical shell-forming protein, Lustrin. Furthermore, the spatial expression patterns of *Lustrin* and *ML1A2* display differences that are likely to affect the physical properties of the shells they are associated with. While we cannot directly link these molecular differences with disparities in the materials properties of the shells that they fabricate, it is clear that the materials properties of these (and all) conchiferan shells are affected by differences in the organic matrices that interact with, and guide the deposition of $CaCO_3$. A pressing challenge for the field of conchiferan biomineralogy is to develop accurate gene knock-out assays for the in vivo functional characterization of these kinds of proteins. Such assays will support a stronger integration of molecular biology, crystallography and materials properties—an integration that we believe is essential for a holistic understanding of any biomineralization process. However, it is now clear that even orthologous shell-forming proteins can harbor significant variation at the genus level, and that variation in these proteins and their expression patterns must also contribute to the diversity of molluscan shell morphologies observable amongst extant and extinct taxa.

## Materials and Methods

### Sample Collection, RNA Extraction, cDNA Synthesis and RACE Library Construction

Live animals were collected from the following locations: *H. asinina*, *H. ovina*, and *H. varia* from Heron Island, Great Barrier Reef, Australia; *H. laevigata* and *H. rubra* from Ocean Wave Seafoods, Lara, Victoria, Australia. Total RNA was extracted from the mantle tissue of all species using Tri Reagent (MRC #118) according to the manufacturer's instructions. The optional high salt precipitation step described by the manufacturer was included to minimize the coprecipitation of proteoglycans and polysaccharides. All RNA extractions were quantified using a Nanodrop 1000 spectrophotometer (running version 3.8.1 software) and qualified by running 500 ng on a nondenaturing agarose gel. cDNA synthesis for standard PCRs was performed by combining 1 μg of total RNA, 5 μl of a 10 μM solution of oligo dT primer, 5 μL of 5X Promega MMLV-RT buffer, 1 μL of 10 mM dNTPs, 8 μL of nuclease free water, and 1 μL of Promega MMLV-RT H⁻. The RNA, primer and water were heated to 70 °C for 10 min, and then the buffer, dNTPs and enzyme added and the reaction incubated at 42 °C for 1 h. The reverse transcriptase was then inactivated for 15 min at 70 °C. 5′ and 3′ RACE libraries were constructed using a protocol based on that described by Zhu et al. (2001).

### RACE PCR

A fragment of *Lustrin* derived from *H. asinina* (*Has-Lustrin*) was initially isolated using primers designed to the 5′ end of the gene with standard cDNA as the template (Jackson et al. 2006). This fragment was then used to design primers to perform 5′ and 3′ RACE PCR in order to isolated the full length *Has-Lustrin* sequence. Full length *Lustrin* sequences were subsequently isolated in a similar manner from *H. ovina*, *H. varia*, *H. laevigata* and *H. rubra*. We previously identified *ML1A2* in a screen for shell-forming gene products in *H. asinina* (Jackson et al. 2006). This sequence was used to design primers for RACE PCR on *H. asinina*, *H. rubra* and *H. laevigata*. All fragments were cloned into pGEM-T (Promega #A1360) and sequenced using standard Sanger chemistry. All sequences reported here have been deposited in GenBank under accession numbers KX687863–KX687874.

### Northern Blots

Five microgram of total mantle RNA isolated from *H. asinina*, *H. rubra*, and *H. laevigata* was denatured in 10 μL of loading buffer (0.8 mM EDTA pH 8.0, 0.5% formaldehyde, 4% glycerol, 5.6% formamide, 0.8× MOPS buffer, 0.003% Bromophenol blue) and 2 μg of ethidium bromide at 75 °C for 10 min. This was then loaded onto a MOPS/formaldehyde 1.2% agarose gel and electrophoresed at 50 V for ∼1 h. RNA was visualized in the gel against a fluorescent ruler, and the gel then soaked in sterile RO $H_2O$ 3× for 15 min each. The RNA was then transferred to a charged nylon membrane by downwards capillary transfer overnight in 20× SSC. RNA was cross-linked to the membrane by exposure to UV light for 2 min, and the membrane then briefly rinsed in RO $H_2O$. The membrane was prehybridized in 2.5 ml of hybridization buffer (5× SSC; 5 mM EDTA; 50% formamide; 100 μg/ml heparin; 0.1% Tween-20; 100 μg/ml sonicated salmon sperm) at 60 °C for 3 h. After this time 1 μl of an 844 bp probe (420 ng) corresponding to the N terminus and the first three conserved Cys-rich domains (see fig. 1) was added to the hybridization buffer and allowed to hybridize at 60 °C overnight. This probe has on average 89% identity with all haliotid *Lustrin* sequences reported here. Following hybridization the membrane was washed three times for 15 min each with a solution of 0.1× SSC and 1% SDS at 60 °C. The membrane was then rinsed in water and blocked with 5 ml of a 2% block solution (Roche #11096176001) in maleic acid buffer at room temperature for 30 min. To this blocking solution 1 μl of antiDIG Fab antibody fragments (Roche #11093274910) was added and incubated for a further 30 min at room temperature. Unbound antibody was then removed with three washes in a solution of maleic acid buffer + 0.3% Tween-20 for 15 min each. The membrane was then equilibrated in color development buffer (0.1 M Tris pH 9.5; 0.1 M NaCl), and then a solution of color development buffer with 225 μg/ml NBT and 175 μg/ml BCIP was applied to the membrane. The color reaction was allowed to proceed until discrete bands were clearly visible. The membrane was then washed in water, air dried and scanned.

### Whole Mount In Situ Hybridization (WMISH)

Juvenile *H. asinina*, *H. rubra*, and *H. laevigata* were first relaxed by gently adding 1 M $MgCl_2$. Once fully anaesthetized, all animals were fixed for 30 min in a 4% paraformaldehyde (PFA) buffered solution (0.5 M MOPS; 10 mM $MgSO_4$; 5 mM EDTA; 2.5 M NaCl). Samples were then dehydrated through an ethanol series and stored at −20 °C. When ready for processing, samples were equilibrated to room temperature and rehydrated into 1× phosphate buffered saline (PBS) and then decalcified in a solution of 350 mM EDTA, 4% PFA, and 1× PBS. Once thoroughly decalcified, the insoluble periostracum and proteinaceous shell material was manually removed with fine forceps, and the animals washed in 1× PBS-Tween and stepped into hybridization buffer (0.75 M sodium chloride, 75 mM sodium citrate, 5 mM EDTA, 50% formamide, 50 μg/ml heparin, 0.1% Tween-20, 0.1 mg/ml yeast total RNA). Samples were slowly brought to the hybridization temperature (50 °C) and prehybridized for 3–4 h. Riboprobes were denatured at 75 °C for 10 min and added to the prehybridized tissue. Hybridization was carried out at 55 °C for 12–16 h. Samples were washed at 55 °C with an increasingly stringent series of washes consisting of three washes with 4× wash (50% formamide, 4× SSC, 0.1% Tween-20), three washes with 2× wash (50% formamide, 2× SSC, 0.1% Tween-20) and three washes with 1× wash (50% formamide, 1× SSC, 0.1% Tween-20). All wash solutions were brought to 60 °C prior to use. Samples were then brought to room temperature and washed several times with 1× SSC, 0.1% Tween-20. To prevent nonspecific binding of the antiDIG antibody, tissue was blocked for 3–4 h at room temperature with a solution of 1× Roche blocking reagent. The DIG hapten was detected by incubation in a solution of antiDIG fab fragments (Roche) diluted 1: 5,000 in 1× Roche blocking reagent for 12–16 h. Unbound antibody was removed by extensive washing in PBS-Tween. Colorimetric detection of DIG bound alkaline phosphatase conjugated fragments was carried out in a solution of alkaline phosphatase detection buffer and NBT/BCIP (Roche) as per the manufactures instructions. The alkaline phosphatase reaction was allowed to proceed until signal intensity reached an appropriate strength, and was then stopped with several washes in stop buffer (0.1 M Glycine pH 2). The tissue was then washed in PBTw, dehydrated through an ethanol series, and photographed whole mount. Some whole mounts were then embedded in paraffin, sectioned at 5 μm, de-parafinized and mounted for photography.

### Nanoindentation and Three Point Bending

Because of the size of the shells and the thickness of the nacreous layers, only relatively thin and small samples were available for mechanical characterization. Nanoindentation is a method specifically suitable to probe the hardness and stiffness on a very small length-scale, and, therefore, to identify property gradients in small specimens. The hardness is the maximum load related to the surface area of the indent. The reduced Young's modulus is calculated from the slope of the force/displacement curve upon unloading, which is related to the stiffness of the specimen under consideration of the stiffness of the setup. While nanoindentation gives information on the elastic properties and hardness depending on the local microstructure in the nm- to μm-range, it does not give information on the macroscopic elastic and fracture properties resulting from the interplay between different regions. The nanoindentation tests were therefore complemented by three point bending tests. Bending loading was chosen because accurate and reproducible machining of small bending bars from a biological material is less challenging than preparing samples for classical tensile measurement. Furthermore, clamping these small bending bars in place for measurement is much easier than samples of comparable size for tensile measurement. In order to minimize variation in the degree of hydration between samples which can influence the materials properties of nacre (Barthelat et al. 2007), nanoindentation and three point bending tests were performed on all samples in the dry state. For nanoindentation, shells of *H. asinina*, *H. laevigata*, and *H. rubra* were cut longitudinally from their anterior edge (i.e., the youngest part of the shell) along the median plane. In addition, for *H. asinina*, samples were taken in a transverse plane to investigate any possible effects of structural anisotropy. Approximately 1 cm thick shell wafers were obtained, embedded into an epoxy resin and mechanically polished to a grain size of 1 μm to achieve a planar surface. Nanoindentation tests were conducted with a "UNAT" indentation device from Asamec (University of Augsburg), equipped with a standard Berkovich tip. Hardness and the reduced Young's modulus were determined with a large number (>100) of indentations attained in five adjacent rows across the entire cross-section of the shell on three different shell portions (supplementary fig. S1, Supplementary Material online). This degree of within-sample replication was intended to control for issues of compositional heterogeneity, and variation in the crystallographic orientation of the nacre tablets.

Three point bending experiments were performed on a total of 67 specimens from seven different shells: Two shells each from *H. asinina*; two from *H. laevigata*; and three shells from *H. rubra*. Parallelepipeds, 2–4 mm wide and 18–22 mm long were ground to ∼0.2–1.2 mm thick to control for interspecies differences in the thickness of the nacreous layers (see supplementary fig. S1, Supplementary Material online). These parallelepipeds were obtained from comparable regions of these shells (supplementary fig. S2, Supplementary Material online). The inner-most, nacreous part was extracted by grinding, and all surfaces were polished to a high finish. Care was taken to ensure that the outer calcitic and prismatic layers were completely removed. For this purpose, cross sections of each specimen were imaged with a light microscope prior and subsequent to the flexural stress–strain measurements. The bending tests were performed on a custom made three point bending device equipped with a fixed inner bearing and adjustable outer bearings. The outer bearings were connected to each other, and each of these bearings was connected to one of two identical load cells (tension–compression type Althen ALF 250 with a nominal force range of ± 100 N, resolution 0.1 N) for the measurement of normal forces. The load signals were read and added by an

analogue–digital converter to obtain the resulting force value. The vertical bearing position and thus the deflection of the specimens was controlled and read by a stepping motor (PD-013/TMCM-013, Trinamic Motion Control, Hamburg, Germany; resolution below 1 nm) via custom-made Labview software. Surface stress and strain were calculated from the force and deflection readings by applying beam theory. From the stress/strain curves, the Young's modulus, the bending strength and the fracture strain were determined. The Young's modulus is a measure of the material stiffness. It is calculated from the ratio of stress and strain in the elastic, that is the fully reversible, loading range. Strictly speaking, beam theory only applies to homogeneous and isotropic materials; therefore, and due to the different stress state, the elastic modulus determined by bending usually differs somewhat from the elastic (Young's) modulus determined in tension. Flexural modulus would therefore be the correct term; for simplicity, we stick to the term "Young's modulus". The bending strength is the highest surface stress reached before fracture of the specimen, and the fracture strain is the permanent strain after fracture of the specimen.

### Statistical Analyses

Box plots were constructed using BoxPlotR (Krzywinski and Altman 2014). Statistical analyses were performed using StatPlus mac (version 6.0.3). All data sets were tested for normality and homogeneity of variance and where necessary, were log transformed prior to analysis by ANOVA. The nanoindentation data could not be made to satisfy these assumptions and so a nonparametric Mann–Whitney test was used to compare means.

## Supplementary Material

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

## Acknowledgments

We gratefully acknowledge Dr Erika Griesshaber and Professor Wolfgang Schmahl for assistance with the materials properties analyses. This work was supported by Deutsche Forschungsgemeinschaft (DFG) funding to D.J.J. through the CRC Geobiology and the German Excellence Initiative and DFG project JA2108/2-1 and Australian Research Council funding to B.M.D.

## Authors' Contributions

D.J.J. conceived, designed, and coordinated the project, collected shells for materials properties analyses, carried out all of the molecular work, sequence analyses, performed the statistical analysis of the materials properties data, wrote and drafted the manuscript. L.R. and C.R. prepared shell samples and generated the materials properties data. N.C. generated transcriptome assemblies and drafted the manuscript. B.M.D. conceived the project, interpreted the results and drafted the manuscript. C.F. coordinated and interpreted the results of the materials properties analyses and drafted the manuscript. All authors gave final approval for publication.

## References

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.

Arivalagan J, Yarra T, Marie B, Sleight VA, Duvernois-Berthet E, Clark MS, Marie A, Berland S. 2017. Insights from the shell proteome: biomineralization to adaptation. *Mol Biol Evol*. 34(1):66–77.

Barthelat F, Tang H, Zavattieri PD, Li CM, Espinosa HD. 2007. On the mechanics of mother-of-pearl: a key feature in the material hierarchical structure. *J Mech Phys Solids*. 55(2):306–337.

Bédouet L, Schuller MJ, Marin F, Milet C, Lopez E, Giraud M. 2001. Soluble proteins of the nacre of the giant oyster *Pinctada maxima* and of the abalone *Haliotis tuberculata*: extraction and partial analysis of nacre proteins. *Comp Biochem Physiol B*. 128(3):389–400.

Bezares J, Asaro R, Hawley M. 2008. Macromolecular structure of the organic framework of nacre in *Haliotis rufescens:* implications for growth and mechanical behavior. *J Struct Biol*. 163(1):61–75.

Cartwright JHE, Checa AG. 2007. The dynamics of nacre self-assembly. *J R Soc Interface*. 4(14):491–504.

Cerveau N, Jackson DJ. 2016. Combining independent *de novo* assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms. *BMC Bioinformatics*. 1–13.

Chang EP, Evans JS. 2015. Pif97, a von Willebrand and Peritrophin biomineralization protein, organizes mineral nanoparticles and creates intracrystalline nanochambers. *Biochemistry* 54(34):5348–5355.

Checa AG, Cartwright JHE, Willinger M-G. 2009. The key role of the surface membrane in why gastropod nacre grows in towers. *Proc Natl Acad Sci U S A*. 106(1):38–43.

Clark MS, Thorne MA, Vieira FA, Cardoso JC, Power DM, Peck LS. 2010. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. *BMC Genomics*. 11(1):362.

Degnan SM, Imron, Geiger DLG, Degnan BMD. 2006. Evolution in temperate and tropical seas: disparate patterns in southern hemisphere abalone (Mollusca: Vetigastropoda: Haliotidae). *Mol Phyl Evol*. 41:249–256.

de Merwe AEB-V, D'Amato ME, Swart BL, Rouvay R-W. 2012. Molecular phylogeny of South African abalone, its origin and evolution as revealed by two genes. *Mar Biol Res*. 8(8):727–736.

Farre B, Dauphin Y. 2009. Lipids from the nacreous and prismatic layers of two Pteriomorpha Mollusc shells. *Comp Biochem Physiol B*. 152(2):103–109.

Feng D, Li Q, Yu H, Kong L, Du S. 2017. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. *Sci Rep*. 7:45754.

Gaume B, Denis F, Van Wormhoudt A, Huchette S, Jackson DJ, Avignon S, Auzoux-Bordenave S. 2014. Characterisation and expression of the biomineralising gene Lustrin A during shell formation of the European abalone *Haliotis tuberculata*. *Comp Biochem Physiol B*. 169:1–8.

Gilbert PU, Metzler RA, Zhou D, Scholl A, Doran A, Young A, Kunz M, Tamura N, Coppersmith SN. 2008. Gradual ordering in red abalone nacre. *J Am Chem Soc*. 130(51):17519–17527.

Jackson DJ, McDougall C, Green KM, Simpson F, Wörheide G, Degnan BM. 2006. A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol*. 4:40.

Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C, et al. 2010. Parallel evolution of nacre building gene sets in molluscs. *Mol Biol Evol*. 27(3):591–608.,

Jackson DJ, Thiel V, Wörheide G. 2010. An evolutionary fast-track to biocalcification. *Geobiology* 8(3):191–196.

Jackson DJ, Wörheide G, Degnan BM. 2007. Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. *BMC Evol Biol*. 7:160.

Jackson DJ, Macis L, Reitner J, Worheide G. 2011. A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. *BMC Evol Biol*. 11:238.

Jackson DJ, Degnan BM. 2016. The importance of Evo-Devo to an integrated understanding of molluscan biomineralisation. *J Struct Biol*. 196(2):67–74.

Jackson DJ, Mann K, Häussermann V, Schilhabel MB, Luter C, Griesshaber E, Schmahl W, Worheide G. 2015. The *Magellania venosa* biomineralizing proteome: a window into brachiopod shell evolution. *Gen Biol Evol*. 7(5):1349–1362.

Krzywinski M, Altman N. 2014. Visualizing samples with box plots. *Nat Meth*. 11(2):119–120.

Lin A, Meyers MA. 2005. Growth and structure in abalone shell. *Mater Sci Eng A: Struct Mater Prop Microstruct Process*. 390(1–2):27–41.

Mann K, Jackson DJ. 2014. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genom*. 15:249.

Marin F, Le Roy N, Marie B. 2012. The formation and mineralization of mollusk shell. *Front Biosci*. 4:1099–1125.

Metzler RA, Abrecht M, Olabisi RM, Ariosa D, Johnson CJ, Frazer BH, Coppersmith SN, Gilbert PU. 2007. Architecture of columnar nacre, and implications for its formation mechanism. *Phys Rev Lett*. 98(26):268102.

Nakahara H. 1991. Nacre formation in bivalve and gastropod molluscs. In: Suga S, Nakahara H, editors. Mechanisms and phylogeny of mineralization in biological systems. Tokyo: Springer. p. 343–350.

Shen X, Belcher AMB, Hansma PK, Stucky GD, Morse DEM. 1997. Molecular cloning and characterization of Lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *J Biol Chem*. 272(51):32472–32481.

Smith BL, Schaffer TE, Viani M, Thompson JB, Frederick NA, Kindt J, Belcher A, Stucky GD, Morse DE, Hansma PK. 1999. Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature* 399(6738):761–763.

Wang R, Gupta HS. 2011. Deformation and fracture mechanisms of bone and nacre. *Ann Rev Mater Res*. 41(1):41–73.

Wustman BA, Morse DE, Evans JS. 2002. Structural analyses of polyelectrolyte sequence domains within the adhesive elastomeric biomineralization protein Lustrin A. *Langmuir*. 18(25):9901–9906.

Wustman BA, Weaver JC, Morse DE, Evans JS. 2003a. Structure-function studies of the Lustrin A polyelectrolyte domains, RKSY and D4. *Conn Tiss Res*. 44:10–15.

Wustman BA, Weaver JC, Morse DE, Evans JS. 2003b. Characterization of a Ca(II)-, mineral-interactive polyelectrolyte sequence from the adhesive elastomeric biomineralization protein lustrin A. *Langmuir* 19:9373–9381.

Zhang B, Wustman BA, Morse D, Evans JS. 2002. Model peptide studies of sequence regions in the elastomeric biomineralization protein, lustrin A. I. The C-domain consensus-PG-, -NVNCT-motif. *Biopolymers* 63(6):358–369.

Zheng X, Cheng M, Xiang L, Liang J, Xie L, Zhang R. 2015. The AP-1 transcription factor homolog Pf-AP-1 activates transcription of multiple biomineral proteins and potentially participates in *Pinctada fucata* biomineralization. *Sci Rep*. 5:14408.

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* 30:892–897.