Parallel Evolution of Nacre Building Gene Sets in Molluscs

Daniel J. Jackson,[1,2] Carmel McDougall,[1] Ben Woodcroft,[1] Patrick Moase,[3] Robert A. Rose,[4] Michael Kube,[5] Richard Reinhardt,[5] Daniel S. Rokhsar,[6,7] Caroline Montagnani,[8] Caroline Joubert,[8] David Piquemal,[9] and Bernard M. Degnan*[,1]

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

[2]Courant Research Center Geobiology, Georg-August-Universität Göttingen, Göttingen, Germany

[3]Clipper Pearls and Autore Pearling, Broome, WA, Australia

[4]Pearl Oyster Propagators, Darwin, NT, Australia

[5]Max Planck Institute for Molecular Genetics, Berlin, Germany

[6]Center for Integrative Genomics and Department of Molecular and Cell Biology, University of California, Berkeley

[7]Department of Energy Joint Genome Institute, Walnut Creek, CA

[8]IFREMER, Biotechnology and Pearl Quality Laboratory, Centre Océanologique du pacifique, Taravao, Tahiti, French Polynesia

[9]Skuldtech, Amarante, Montpellier, France

***Corresponding author:** E-mail: b.degnan@uq.edu.au.

Associate editor: Claudia Kappen

Abstract

The capacity to biomineralize is closely linked to the rapid expansion of animal life during the early Cambrian, with many skeletonized phyla first appearing in the fossil record at this time. The appearance of disparate molluscan forms during this period leaves open the possibility that shells evolved independently and in parallel in at least some groups. To test this proposition and gain insight into the evolution of structural genes that contribute to shell fabrication, we compared genes expressed in nacre (mother-of-pearl) forming cells in the mantle of the bivalve *Pinctada maxima* and the gastropod *Haliotis asinina*. Despite both species having highly lustrous nacre, we find extensive differences in these expressed gene sets. Following the removal of housekeeping genes, less than 10% of all gene clusters are shared between these molluscs, with some being conserved biomineralization genes that are also found in deuterostomes. These differences extend to secreted proteins that may localize to the organic shell matrix, with less than 15% of this secretome being shared. Despite these differences, *H. asinina* and *P. maxima* both secrete proteins with repetitive low-complexity domains (RLCDs). *Pinctada maxima* RLCD proteins—for example, the shematrins—are predominated by silk/fibroin-like domains, which are absent from the *H. asinina* data set. Comparisons of *shematrin* genes across three species of *Pinctada* indicate that this gene family has undergone extensive divergent evolution within pearl oysters. We also detect fundamental bivalve–gastropod differences in extracellular matrix proteins involved in mollusc-shell formation. *Pinctada maxima* expresses a chitin synthase at high levels and several chitin deacetylation genes, whereas only one protein involved in chitin interactions is present in the *H. asinina* data set, suggesting that the organic matrix on which calcification proceeds differs fundamentally between these species. Large-scale differences in genes expressed in nacre-forming cells of *Pinctada* and *Haliotis* are compatible with the hypothesis that gastropod and bivalve nacre is the result of convergent evolution. The expression of novel biomineralizing RLCD proteins in each of these two molluscs and, interestingly, sea urchins suggests that the evolution of such structural proteins has occurred independently multiple times in the Metazoa.

Key words: biomineralization, nacre, EST, evolution, mollusc, pearl.

Introduction

The Mollusca is a highly successful group of metazoans in terms of species abundance, range of morphotypes, and diversity of habitats occupied. This success can in part be attributed to a mechanism of shell construction adopted early in the history of the clade during the late pre-Cambrian (545+ Ma). The diversity of gastropod and bivalve shell types observed among extinct and modern species is evidence of the evolutionary success of this structure. Formed by secretions of the underlying tissue (the mantle), molluscan shells are a biocomposite material consisting primarily of $CaCO_3$, polysaccharides, and proteins. The organically derived component of the shell controls the initiation, expansion, and termination of crystal growth in a cell autonomous fashion and can even control the polymorph of calcium carbonate that is precipitated, most often aragonite or calcite (Belcher et al. 1996). Analysis of the molecular (Bedouet et al. 2006; Jackson et al. 2006; Jackson, Wörheide, et al. 2007; Ma et al. 2007; Gong et al. 2008) and mineralogical (Metzler et al. 2007; Gilbert et al. 2008) processes that generate the shell is providing some insight into how this is achieved at biological pH, temperature, and pressure, a feat of great interest to materials scientists.

Although deep-level relationships between major molluscan clades remain the topic of long-standing debates and current research, monophyly for the major extant groups (Monoplacophora, Polyplacophora, Gastropoda, Bivalvia, Cephalopoda, and Scaphopoda) is generally accepted and supported by morphological and molecular data (Sigwart and Sutton 2007; Wilson et al. 2009). The Gastropoda (60,000+ species) and Bivalvia (30,000+ species) constitute the most diverse and abundant molluscan assemblages and are both members of the Conchifera (shelled molluscs). Divergence estimates for these two lineages range between 570 and 536 Ma (Colgan et al. 2000). Members of both clades are known to produce a "nacreous" (mother-of-pearl) material that is comprised of tablets of aragonite embedded in and perfused by an organic matrix. However, there are significant differences in the arrangement of the $CaCO_3$ crystals, the axes along which these crystals are arranged, and the way in which the crystals are deposited between the two clades (Chateigner et al. 2000; Cusack and Freer 2008; Meldrum and Cölfen 2008). Although it is known that the shells of bivalves and gastropods possess highly acidic proteins in the ethylenediaminetetraacetic acid (EDTA)–insoluble organic matrix (EIOM) (Gotliv et al. 2003; Fu et al. 2005) and that chitin is a component of the EIOM in both classes of molluscs (Weiner and Traub 1980; Keith et al. 1993; Weiss et al. 2006), deeper level homologies in the molecular processes that guide nacre formation are unknown.

A previous genomewide investigation of sea urchin (*Strongylocentrotus purpuratus*) biomineralization genes perhaps represents the most comprehensive molecular study of biomineralization to date (Livingston et al. 2006). That study found many of the proteins involved late in the biomineralization process differ between vertebrates and echinoderms, although similarities were detected in some of the proteins involved in the specification of skeleton-producing cells, including collagen-based proteins and transcription factors belonging to the *ETS* and *ALX* families that control the differentiation of matrix-secreting cells. Although those observations suggest that homologies exist in the skeletogenic program within deuterostomes, limited gene-expression data from other metazoan clades impede broader comparisons of the process of biomineralization.

To understand the origin and evolution of bilaterian and molluscan biomineralization, we have sequenced >12,000 expressed sequence tags (ESTs) from the biomineralizing tissues of the tropical abalone *Haliotis asinina* and the silver lipped pearl oyster *Pinctada maxima*. More specifically, these ESTs are derived from cells responsible for secreting the aragonitic polymorph of $CaCO_3$ that yields the mother-of-pearl nacreous layer. We expanded this comparison with ESTs derived from mantle-specific cDNA libraries of the limpet *Lottia gigantea* (Eogastropoda) and a subset of sequences derived from the mantle tissue of another oyster, *Pinctada margaritifera*. In the search for novel biomineralization proteins, an EST approach is unbiased by the difficulties of extracting proteins from calcified materials (Gotliv et al. 2003) and uncovers gene products that are secreted into the pallial space but not incorporated into the shell, proteins responsible for ion trafficking, gene transcription, and other upstream regulatory steps.

By directly comparing the transcriptomes of nacre-forming cells in a bivalve and a gastropod, we show that there are dramatic differences in the gene sets used to biofabricate the nacreous layer of the shell. This is particularly pronounced in the genes encoding secreted proteins, many of which are likely to contribute directly to shell formation (Jackson et al. 2006). These differences extend to comparisons within the Gastropoda (*Haliotis* vs. *Lottia*) and to comparisons within a single biomineralizing gene family (*shematrin*) across three species of *Pinctada*, lending support to our previous supposition that the molluscan shell–forming secretome is rapidly evolving (Jackson et al. 2006), and additionally that the parallel evolution of secreted proteins with repetitive low-complexity domains (RLCDs) is an important feature of conchiferan evolution.

Materials and Methods

RNA Isolation and Library Construction

Pinctada maxima mantle tissue was collected from a single healthy individual maintained at the Clipper Pearls/Autore Pearling farm lease, near Quondong Point, Broome, Western Australia, Australia. The anterior edge of the mantle (the mantle skirt) was removed, and the zone used as grafting tissue (donor tissue for pearling operations and responsible for nacre secretion) was dissected by an experienced pearl seeding technician, preserved in RNA later and frozen for transport to the University of Queensland. RNA was extracted with TriReagent. Tissue from *H. asinina* was derived from a single reproductively mature individual maintained at the Bribie Island Aquaculture Centre, Queensland, Australia. Based on previous studies (Jolly et al. 2004; Jackson et al. 2006; Jackson, Wörheide, et al. 2007) the region responsible for nacre deposition as recognized by a population of cells with a morphology distinct from cells that express a gene responsible for pigmenting the outer periostracum (Jackson et al. 2006), was dissected, and RNA extracted using TriReagent.

PolyA mRNA was isolated from both *P. maxima* and *H. asinina* samples using Dynabeads (according to the manufacturer's instructions) and shipped to the Max Planck Institute for Molecular Genetics, Berlin, Germany for EST sequencing. cDNA libraries were generated using the Cloneminer kit (Invitrogen) and were cloned into the pDONR222 vector.

ESTs from *P. margaritifera* were acquired from a cDNA library constructed from the mantle tissues of several *P. margaritifera* oysters. EST sequences were obtained according to the Genome Sequencer FLX's protocol (454/Roche manufacturer). A de novo assembly using the CAP3 assembly program from TGI Clustering tools was used to assemble 454 reads with overlapping identity percentage and minimum overlapping length parameters set to >90% and 60 bp, respectively. The *shematrin*-related EST sequences were obtained using a local BlastN and BlastP algorithm against the *P. margaritifera* EST data set.

Sequence Analysis and Bioinformatics

A total of 6,122 *H. asinina* and 6,737 *P. maxima* reads were acquired using standard Sanger sequencing chemistry (v3.1). Reads were vector and quality trimmed and clustered using TGCIL with cap3 (Huang and Madan 1999). EST data sets were annotated using annot8r (Schmid and Blaxter 2008). This pipeline takes a query collection of nucleotide sequences and uses the BlastX algorithm to search a UniProt database with gene ontology (GO) Enzyme Commission, and Kyoto Encyclopedia of Genes and Genomes annotations. We used an expect value of 10E−06 for these BlastX searches, and excluded inferred electronic annotations (IEA) in order to maintain confidence in these results. Where multiple annotations were returned the annotation with the best score and the highest proportion of hits was taken. GO-Diff (Chen et al. 2006) was then used to search for relative differences in the representation of GO-Slim terms as measured by EST abundance. The gdf. conf file was modified to run in execution mode 3, with a ratio cutoff of 2 and a false discovery rate of 0.1. Excluded evidence codes were IEA, ISS, NAS, ND, and NR.

In order to remove clusters derived from gene products not specifically involved in biocalcification we constructed a database of ribosomal RNAs and proteins from a range of metazoans. The database was initially established using ribosomal protein sequences as described for the poriferan *Suberites domuncula* (Perina et al. 2006). This set of reference sequences was then used to extract sequences from GenBank and the *L. gigantea* (Eogastropoda) genome in order to broaden taxonomic representation. rRNA sequences from a variety of metazoan taxa were collected from GenBank. The complete mitochondrial genomes of several lophotrochozoan, ecdysozoan, and vertebrate taxa were similarly used to identify mitochondrial encoded gene products in our data sets. These searches were conducted using cutoff values of 10E−12.

In order to identify commonalties in the molecular processes occurring in the mantle tissues of both animals, an all-against-all search of *H. asinina* EST query sequences versus a *P. maxima* EST database and vice versa was conducted using the TBlastN algorithm under a local installation of Blast (v2.2.18). Low complexity filtering was turned on, and the expect value was set at 10E−06.

We next focused our search for gene products involved in calcification in both animals by searching for secreted gene products. To do this, we conceptually translated all contigs in all six possible reading frames and selected the longest reading frame beginning with a methionine residue. Open reading frames (ORFs) shorter than 50 amino acid residues and that did not possess a stop codon (in order to avoid missannotation of transmembrane proteins) were discarded. These potential full-length ORFs were searched for signal peptides using a local installation of SignalP v3 (Emanuelsson et al. 2007). Signal peptide–positive contigs were filtered for transmembrane domains using the TMHMM Server (v. 2.0). Glycosyl phosphatidyl inositol (GPI)-anchored proteins were detected using the GPI modification predictor (Eisenhaber et al. 1999). Proteins targeted for organelles such as mitochondria, the Golgi, or lysosomes were identified and removed by submitting the data set to TargetP (Emanuelsson et al. 2007). Abalone and oyster full-length secreted sequences were then searched against each other (TBlastX) and the following databases: *L. gigantea* JGI filtered protein models (BlastX); *L. gigantea* assembled whole-genome scaffolds (TBlastX); GenBank's nonredundant protein-sequence database (BlastX); GenBank's EST database (TBlastX).

In order to detect gene products with repetitive low-complexity features in our data sets, we employed XSTREAM (Newman and Cooper 2007). For the set of ESTs examined for tandem repeats (TRs), we included sequences coding for ORFs larger than 50 amino acids, predicted to be secreted by TargetP, but not necessarily full length. C-terminal lipid anchors and transmembrane regions in these non–full-length cases will not be detected. Although such membrane-associated gene products will not be located within the final calcified structure, we reason that they may still play a role in shell formation as they would be available to interact with and influence the biochemistry of the pallial fluid and the shell. We only investigated novel, or apparently mollusc-specific ESTs (i.e., removed sequences with broad taxonomic distributions unlikely to be involved in shell-forming processes). These criteria provided 124 *H. asinina* XSTREAM queries and 144 *P. maxima* queries. We also included in this analysis all available *L. gigantea* mantle ESTs (2,107 unigenes), which filtered down to 132 XSTREAM queries.

We also used XSTREAM to search for TRs in a pooled collection of shematrins from the *P. maxima* and *P. margaritifera* data sets and from previously published *P. fucata* shematrin sequences (Yano et al. 2006). The settings used in this analysis were degeneracy = 0, TR significance = high, and min consensus match = 0.8.

Phylogenetic Analyses

Alpha Carbonic Anhydrase (α-CA). We collected representatives of the major α-CA classes and compared these with the molluscan nacrein proteins and the two *P. maxima* sequences detected in the current data set. Because the nacreins contain a low-complexity Gly-Aln insertion within the α-CA domain that is not present in any other metazoan α-CA, we removed this domain from the nacreins and selected highly conserved regions of the alignment using Gblocks (Castresana 2000). α-CAs from a green alga and proteobacteria were set as outgroups. A Bayesian analysis (MrBayes v 3.1.2) of the resulting alignment was run for 2,500,000 generations with four chains each with eight runs heated to 0.2, by which time stationarity was achieved. The sampling frequency was 1,000, a proportion of sites were invariable with the remaining sampled from a gamma distribution, and the amino acid model was set to Whelan and Goldman (WAG) according to Protest results and our previous analyses (Jackson, Macis, et al. 2007).

Inference of Pinctada Relationships Using Cytochrome Oxidase I (COI) and 16S rRNA Genes. We selected several *Pinctada* species for which the commonly used phylogenetic markers *COI* and *16S rRNA* genes were available, combined these with available *P. maxima*, *P. fucata*, and *P. margaritifera* sequences and also included two outgroup bivalve species, *Crassostrea gigas* and *Pteria hirundo*. A protein-based alignment of the COI-coding sequence was performed and the resulting nucleotide alignment retrieved for a codon based Bayesian analysis (MrBayes v 3.1.2) using the metazoan mitochondrial genetic code. *16S rRNA* gene was included as a separate partition with a proportion of sites from both partitions set as invariable with the remaining sampled from a gamma distribution. The analysis was run for 1,000,000 generations with four chains each with two runs heated to 0.1 with a sampling frequency of 1,000, by which time stationarity was achieved.

Cyclophillin. A range of metazoan Cyclophilin protein sequences were retrieved from GenBank using the standard BlastP algorithm. Active sites as indicated by the conserved domain database (Marchler-Bauer et al. 2009) were mapped to an alignment against a human-reference sequence (Zhao and Ke 1996).

Secreted Protein Acidic Cysteine-Rich (SPARC). A broad representation of SPARC gene products based initially on the collection described by Livingston et al. (2006) were aligned using MUSCLE (Edgar 2004). Additional sequences retrieved from GenBank dbEST were included to increase the representation of lophotrochozoan taxa. A Bayesian analysis (MrBayes v 3.1.2) was run for 4,000,000 generations with eight chains each with four runs heated to 0.2 with a sampling frequency of 1,000, by which time stationarity was achieved. The resulting consensus tree was midpoint rooted due to the lack of an appropriate outgroup.

Results

Library Features

Employing an EST-sequencing approach, we have acquired 6,122 reads from the nacre-secreting mantle tissue of the tropical abalone *H. asinina* and 6,737 reads from the pearl oyster *P. maxima* (GenBank accessions: GT271630–GT284488, supplementary table 1, Supplementary Material online). Following quality trimming and clustering, these single pass 5′ reads produced 3,043 *H. asinina* contigs, which assembled into 3,005 unique clusters and 2,919 *P. maxima* contigs, which assembled into 2,850 unique clusters. Contigs with >1 EST member have been submitted to GenBank as a Transcriptome Shotgun Assembly (abalone GenBank accessions EZ420605–EZ421271, supplementary table 2, Supplementary Material online; pearl oyster GenBank accessions EZ420070–EZ420604, supplementary table 3, Supplementary Material online). Over 38% of the *H. asinina* reads were singletons with an overall library redundancy of 50%, and over 35% of the *P. maxima* reads were singletons with an overall library redundancy of 43% (fig. 1). The average contig length for *H. asinina* was 536 bp (min. 101 bp, max. 1,790 bp) and for *P. maxima* was 659 bp (min. 158 bp, max. 3,805 bp). We have also taken advantage of the information in the number of individual reads that constitute a contig, with the assumption that this is a reflection of the level of gene expression, as has been used elsewhere (Ewing et al. 1999; Nishikata et al. 2001; Satoh et al. 2003; Chao et al. 2006). The distribution of contig abundances between the two libraries indicates that *P. maxima* expresses a family of biomineralization genes (the *shematrins*) at levels about twice that of the most abundantly expressed *H. asinina* gene product, which is a novel, putative secreted gene product (fig. 1*B*).

Functional Categorization of Gene Products

An initial round of EST annotation using annot8r (Schmid and Blaxter 2008) indicated that a minority of clusters from both libraries returned hits with GO annotations from the UniProt database; 446 (14.8%) *H. asinina* annotations and 805 (28.3%) *P. maxima* annotations across the three GOA namespaces (biological process, molecular function, and cellular component) were returned (fig. 2).

We next removed all ribosomal RNAs, ribosomal proteins, and mitochondrial-encoded gene products from each library (104 contigs from *H. asinina* and 98 contigs from *P. maxima*), leaving a total of 2,939 unique *H. asinina* contigs and 2,821 unique *P. maxima* contigs. These data sets were searched reciprocally against each other (the *H. asinina* data set acted as a query set against the *P. maxima* database and vice versa) using TBlastX. Two hundred and five unique *H. asinina* clusters and 253 unique *P. maxima* clusters shared similarity with each other. These figures indicate that following removal of highly conserved ribosomal rRNAs, proteins and mitochondrial gene products, less than 10% of each library shared similarity with the other. The resulting sets of shared sequences were processed by annot8r in order to obtain GO annotations for shared biological processes, functions, and cellular localizations. These annotations revealed that the majority of shared genes with GO biological process annotations were involved in cellular processes, metabolic processes, and regulation of biological processes (fig. 3). Most annotations for molecular function were related to binding and catalysis, whereas the cellular localization of the majority of these annotated shared genes was intracellular.

These GO annotations were then processed by GO-Diff (Chen et al. 2006) to search for overall inferred expression-level differences in genes with characterized molecular functions (fig. 4). This revealed that *H. asinina* expressed more genes within the "enzyme regulator activity" category (predominantly protease inhibitors) and genes involved in the regulation and transport of ions, electrons, and proteins. In contrast, *P. maxima* had a higher representation of genes with "motor activity" and markedly higher levels of "lyase activity" primarily due to the abundant expression of two *α-CA* genes. Because α-CAs have previously been associated with biocalcification processes in various metazoan phyla (Miyamoto et al. 1996; Horne et al. 2002; Jackson, Macis, et al. 2007; Moya et al. 2008) we attempted

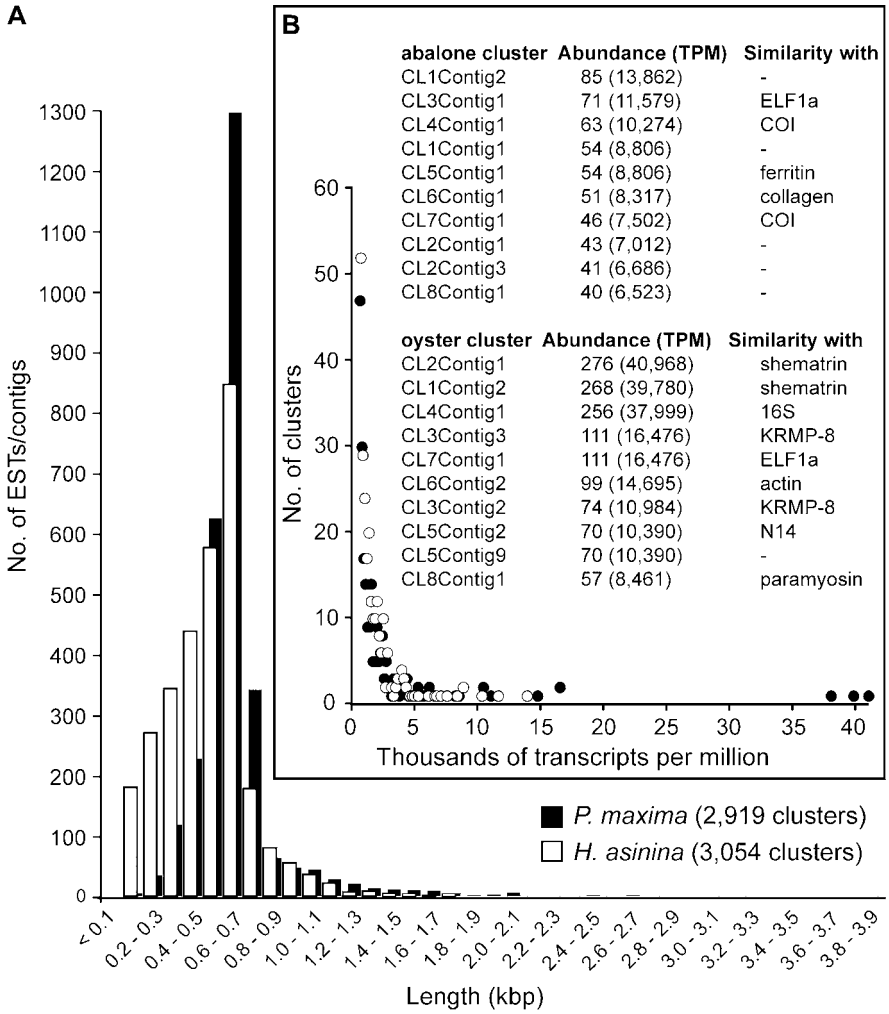

FIG. 1. Overview of abalone and pearl oyster nacre EST libraries. (*A*) Cluster-length frequencies for each library plotted in 100-bp bins. (*B*) The distribution of cluster abundances differed markedly with *Pinctada maxima* expressing known biomineralizing genes (*shematrin-* and *lysine-rich matrix protein—KRMP*) at levels more than twice that of the most abundantly expressed *Haliotis asinina* gene. The top 10 most abundant ESTs for each data set are listed with their corresponding annotations. Expression levels are presented as the raw number of reads and as normalized values for comparison (transcripts per million, TPM).

to assign these α-CAs to their respective classes. This analysis lacked resolution and returned low support values at several critical nodes (fig. 5) and is likely to be the result of insufficient taxon sampling within key clades. Although neither of the two α-CA *P. maxima* clusters could be assigned to any particular α-CA family, they were also excluded from a well-supported clade that included all molluscan nacrein proteins (fig. 5). The nacrein proteins are mollusc-specific enzymes that possess two α-CAs domains separated by a low-complexity domain rich in glycine and aspargine (Miyamoto et al. 2003) and are known to be involved in molluscan biocalcification (Miyamoto et al. 2005). This result suggests that bivalves express a diversity of α-CA enzymes within biomineralizing tissues.

Similarities and Differences between Gastropod and Bivalve Nacre-Forming Gene Products

In order to focus our comparisons on genes likely to be directly involved in nacre formation, we retrieved conceptually derived full-length gene products that possessed a signal sequence from each gene set. One hundred and twenty-nine *H. asinina* and 125 *P. maxima* sequences were identified and searched against each other and a variety of databases. The majority of these products were unique; 95 (74%) of the *H. asinina*–secreted products and 71 (57%) of the *P. maxima* products shared no similarity with sequences in GenBank nr and EST databases, nor the *L. gigantea* genome. Of the 54 *P. maxima*–secreted products that shared similarity with a previously described sequence, 12 of these were against bivalve-specific biomineralization proteins (fig. 6) (Yano et al. 2006; Zhang et al. 2006). In contrast, none of the 34 secreted abalone proteins that shared similarity with other sequences were mollusc specific (fig. 6). Secreted gene products of known function were mostly protease inhibitors and lectins. Within this secretome comparison, a notable difference was the range of inferred expression levels (fig. 6). *Pinctada maxima* expressed two families of previously characterized biocalcification

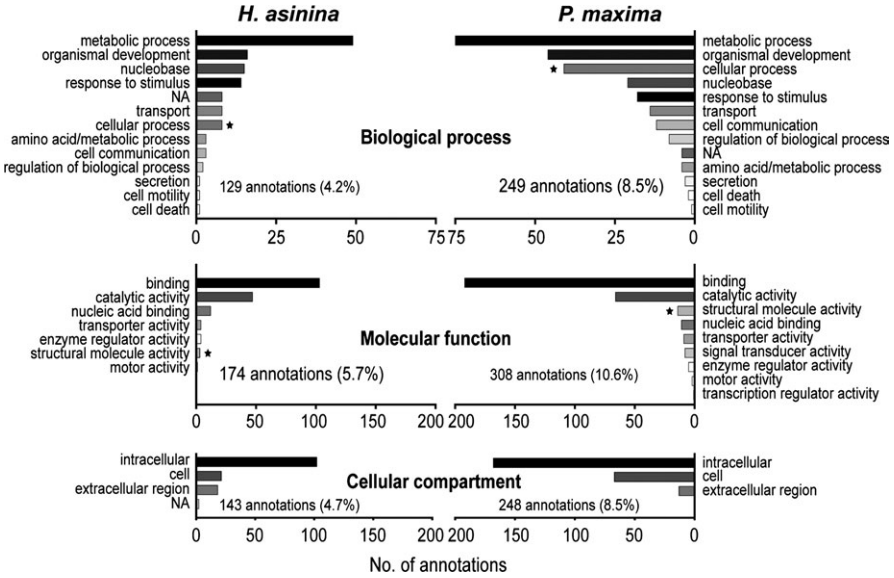

Fig. 2. GO annotations for all abalone and pearl oyster EST clusters using annot8r (Hotz-Wagenblatt et al. 2003). GO annotations for biological process, molecular function and cellular compartment were paired with their parent GO-SLIM identifiers. Three thousand and forty-three *Haliotis asinina* queries and 2,919 *Pinctada maxima* queries were processed. The total number of annotations for each category is indicated as an absolute value and as a percentage of the total number of queries. Notable differences between the two data sets are indicated with a star.

genes, the *shematrins* and *lysine-rich matrix proteins* (*KRMPs*) (Zhang et al. 2006) at levels in excess of 10,000 transcripts per million (TPM), that is, 4.1% and 1.6% of the sampled transcriptome. In comparison, the most abundantly expressed transcript in the *H. asinina* data set was a unique putative secreted gene product expressed at 1.4% of the sampled transcriptome. Only 6 novel abalone and 1 novel oyster secreted proteins shared similarity with protein models in the *Lottia* genome (supplementary fig. 1, Supplementary Material online).

RLCD-containing proteins

Several biomineral-influencing proteins are known to possess either repetitive motifs, or domains of low complexity (Marin and Luquet 2004). For example, Lustrin-A, one of the first molluscan shell proteins characterized at the molecular level, has 10 cysteine-rich domains, 8 proline-rich domains, and a Gly-Ser–rich domain of 250 residues (Shen et al. 1997). Similarly, nacrein is an α-CA present in both bivalves and gastropods that contains a low-complexity Gly-Asn domain (Miyamoto et al. 1996). We therefore searched our data sets for gene products containing such domains using XSTREAM (Newman and Cooper 2007). We isolated conceptually derived proteins from the abalone, limpet, and oyster data sets that possessed an ORF larger than 50 amino acid residues, a signal sequence, no detectable transmembrane domains, or GPI anchors and were predicted by TargetP to be secreted (Emanuelsson et al. 2007). From these collections we detected 14 sequences from the *P. maxima* data set and 4 sequences from both the *H. asinina* and *L. gigantea* data sets (fig. 7). A feature of many of these proteins was a high glycine content. *Pinctada maxima* expressed these genes at higher levels than *H. asinina*.

Given the abundance of RLCD gene products in the *P. maxima* data set and the apparent importance of shematrin gene products to bivalve nacre formation (Yano et al. 2006), we isolated *shematrin*-related genes from a 454/Roche GS Flex data set derived from the mantle of *P. margaritifera* and compared these with the *P. maxima* shematrin sequences reported here and previously reported sequences from *P. fucata* (Yano et al. 2006). Based on the presence of characteristic glycine- and tyrosine-rich domains, and a signature RKKKY, RRKKY, RRRKK, IRRKK, or PRKKY C-terminal motif, we were able to identify 24 shematrin-like gene products from the *P. margaritifera* data set. These 24 sequences were filtered from a collection of 92 contigs that returned positive shematrin Blast results.

Because alignment-based sequence comparisons between proteins that contain extensive low-complexity regions can violate assumptions of character homology (and are often phylogenetically uninformative), we adopted a comparative approach for the shematrin proteins independent of shematrin sequence alignment. We pooled all shematrin-like sequences and searched for TRs (Newman and Cooper 2007) and mapped these results onto a phylogram of *Pinctada* species constructed using *COI* and *16S rRNA* genes. Although this approach does not allow formal statements regarding rates or direction of shematrin evolution, we can observe broad trends of shematrin conservation and innovation. Across the three *Pinctada* species for which *shematrin* genes have been isolated, several similar motifs appear to be present in all three species (fig. 8), whereas an abundantly expressed *P. maxima shematrin* EST (CL1Contig1) contains a novel duplication of a large single repetitive domain (fig. 8). *Pinctada margaritifera*

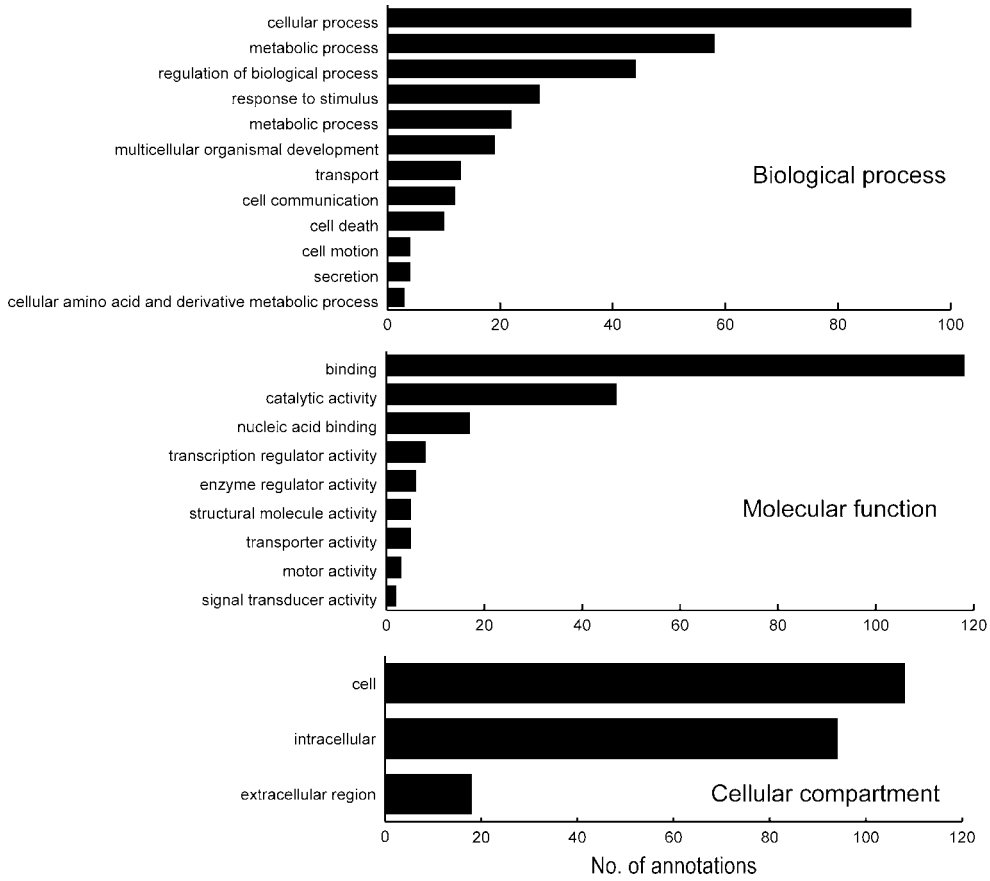

Fig. 3. GO annotations of ESTs shared between pearl oyster and abalone. Following the removal of genes encoding riboproteins, rRNAs, and mitochondrial proteins, each data set was searched against the other using TBlastX. The resulting list of shared ESTs was pooled and searched for GO annotations using annot8r (Schmid and Blaxter 2008). The resulting GO IDs were matched to their parent slim IDs.

and *P. fucata* also displayed a significant amount of lineage-specific diversity within these RLCDs. We acknowledge that these EST data sets are not exhaustive, and future efforts are likely to reveal further *shematrin* members. However, the finding that in terms of shematrin-sequence representation the two more extensively sampled species (*P. margaritifera* and *P. fucata*) share little shematrin sequence similarity supports the view of active *shematrin* expansion and diversification within the pearl oysters.

We also searched abalone and pearl oyster data sets for the presence of known deuterostome biomineralizing genes. Based on a survey of the sea urchin genome by Livingston et al. (2006), we were able to detect several similarities. We retrieved seven cyclophilin gene products from the *P. maxima* data set and one from the *H. asinina* data set. Of the seven pearl oyster cyclophilins, two were expressed at moderately high levels (pmax_Cluster89Contig1 at 1,484 TPM and pmax_Cluster203Contig1 at 445 TPM). Interestingly three of the seven pearl oyster cyclophilins and the two sea urchin cyclophilins known to be expressed in biomineralizing cells (Amore and Davidson 2006; Livingston et al. 2006) have mutations at known active site positions (fig. 9). We also found a representative of the SPARC family in the pearl oyster data set. A Bayesian phylogenetic analysis of a broad sample of SPARC proteins generated two major clades composed of deuterostome and protostome members (fig. 10). Our *P. maxima* SPARC EST fell within a clade of bivalve EST sequences, whereas sea urchin SPARC and Osteonectin proteins, which are expressed in spicule-forming cells (Livingston et al. 2006), formed a sister group to chordate sequences. We also detected seven independent clusters associated with the production, deacetylation, and binding of chitin in the *P. maxima* data set (table 1). *Pinctada maxima* gene products associated with chitin biogenesis or modification were relatively highly expressed. Twenty-nine chitin synthase transcripts (4,305 TPM), and two independent clusters of chitin deacetylase (1,484 and 1,039 TPM, respectively) were detected. In contrast, we could only detect a single abalone EST with the potential to bind chitin; *H. asinina* EST P0009M03 returned a significant PFAM result for one peritrophin-A domain and a less significant match for a second domain (two peritrophin-A domains are necessary to bind chitin; Shen and Jacobs-Lorena 1998). We were also able to detect three *P. maxima* and three *H. asinina* gene products encoding collagen (table 1). Importantly, two of the abalone collagen gene clusters (hasinina_Cluster6Contig1 and hasinina_Cluster21Contig1) were expressed at extremely high levels (8,317 TPM and 4,240 TPM, respectively). In contrast, the oyster collagen clusters were expressed at much lower levels (1,484 and 1,187 TPM, respectively, table 1).

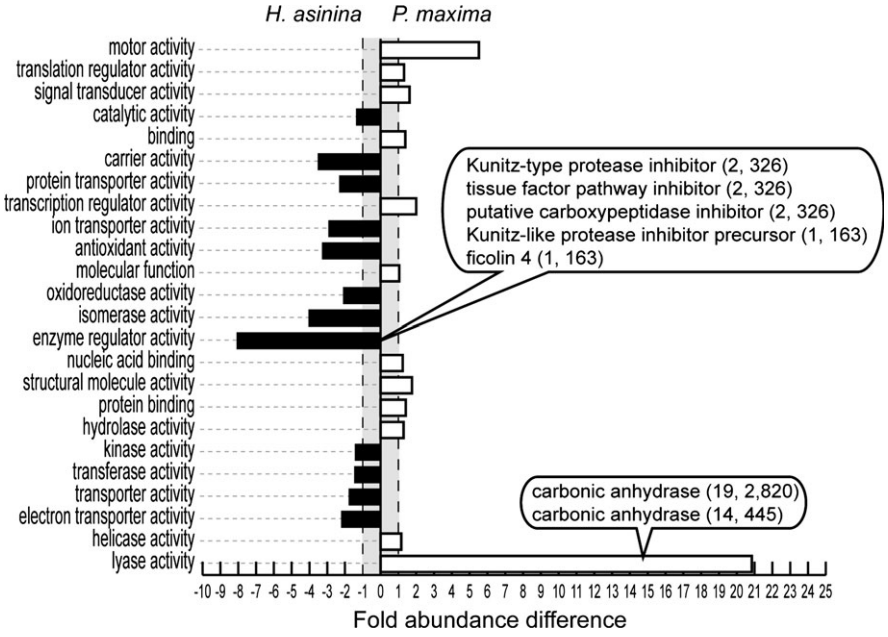

Fig. 4. Differences in the expression of GO "molecular function" annotations between pearl oyster and abalone EST data sets. Relative EST coverage level ratios for GO molecular function annotations were detected by GO-Diff (Chen et al. 2006). Black bars less than −1 indicate a more abundantly represented GO-Slim term in the *Haliotis asinina* data set, and white bars exceeding +1 a more abundantly represented term in the *Pinctada maxima* data set. Numbers in brackets indicate the raw number of ESTs representing that cluster and the normalized inferred expression level in TPM, respectively.

Discussion

The bivalve *P. maxima* and the gastropod *H. asinina* have been evolving independently since the Cambrian yet share features common and unique to molluscs. Both animals possess a shell with an inner nacreous layer, the fabrication of which is directed by a specific region of the mantle largely through secretion of compounds such as polysaccharides, peptides, and proteins. Many of these become integral structural components of the shell and confer species-specific architectures, patterns, and colors. Here, we have shown that the transcriptomes of nacre-producing mantle cells from *P. maxima* and *H. asinina* are markedly different, consistent with bivalve and gastropod nacre being a convergent trait.

High Levels of Gene Novelty in *H. asinina* and *P. maxima* Nacre-Forming Gene Sets

Both *P. maxima* and *H. asinina* EST data sets are comprised largely of sequences with no associated GO terms, similar to other nonmodel invertebrate EST data sets (Leu et al. 2007; Vera et al. 2008; Takahashi et al. 2009). It should be noted that the Blast based annotation pipeline we employed used a cutoff *e* value of 10E−06 and will have missed genes with low sequence similarity to known sequences. Of the <10% of clusters shared between the abalone and oyster (not including riboproteins, rRNAs, and mitochondrial genes), most correspond to conserved genes that are involved in processes beyond biocalcification. The number of GO annotations differs noticeably between these molluscs with more annotations in the *P. maxima* data set than *H. asinina* (fig. 2; cf. Jackson et al. 2006).

The marked differences in gene content and expression levels between *P. maxima* and *H. asinina* suggest that the molecular machinery used in shell construction by these animals is significantly different. Although our comparative EST analysis does not provide functional information for the novel sequences it identified, it does provide insight into the degree of transcriptome evolution between molluscan biomineralizing tissues. Gross differences in gene expression, both in composition and quantity, are consistent with independent evolutionary histories for nacre formation in these classes of molluscs.

Secreted RLCD Proteins

Previously, we characterized a selection of novel ESTs from the mantle tissue of *H. asinina* by in situ hybridization (Jackson et al. 2006). One of our EST selection criteria for in situ screening was the presence of both a signal sequence and a domain of low complexity. In all cases, such genes were spatially restricted to a discrete zone of the mantle tissue, strongly suggesting that these gene products are critical to the process of shell formation. In our current comparisons, we have found differences between *P. maxima* and *H. asinina* in both the abundance and diversity of gene products containing such RLCDs, suggesting that fundamental aspects of nacre deposition employed by the oyster and the abalone have evolved independently.

Fourteen gene products containing RLCDs as detected by XSTREAM were retrieved from the *P. maxima* data set (fig. 7), six of which shared no similarity with previously described sequences (novel), the remaining eight possessed similarity to two previously identified oyster

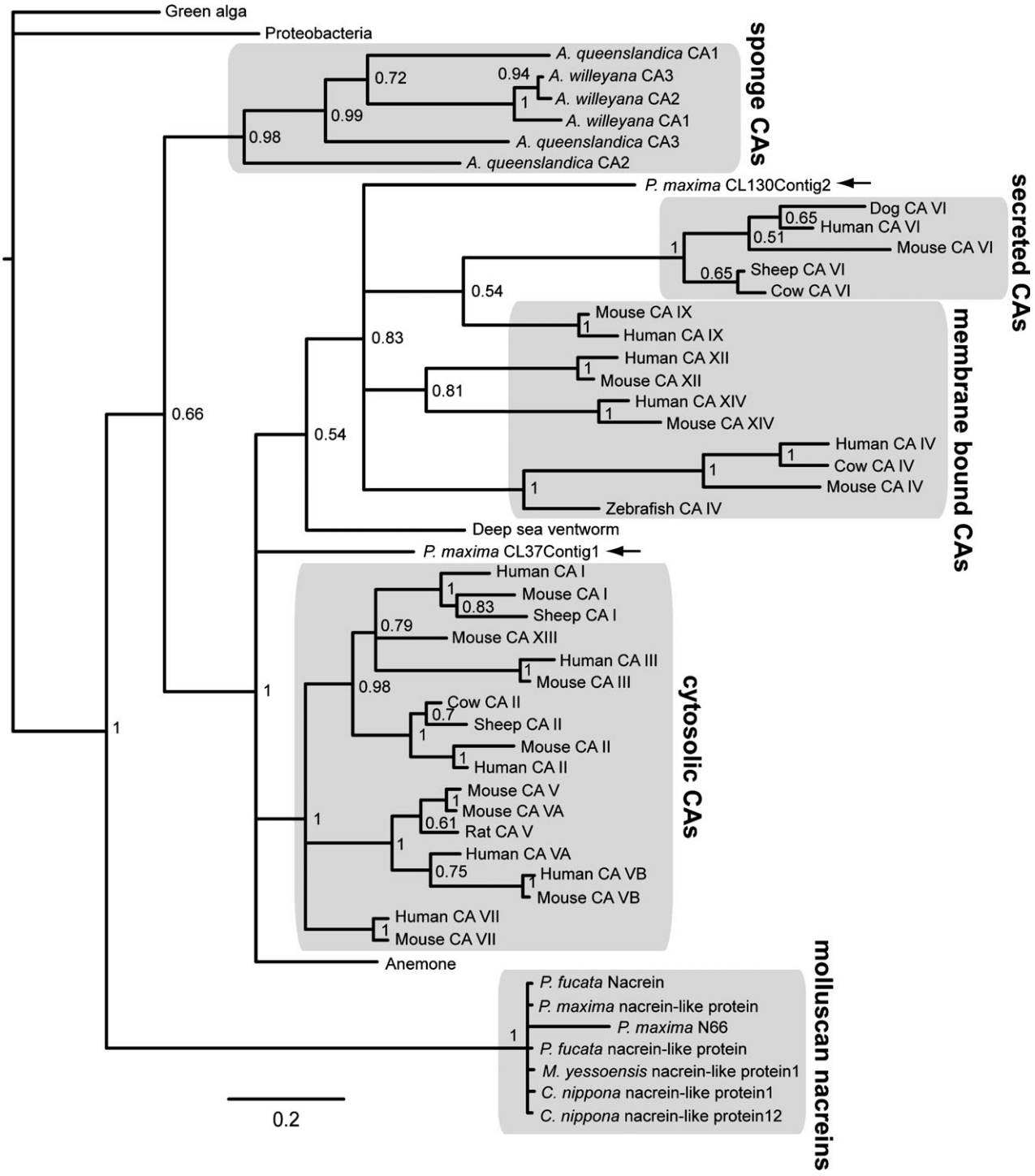

Fig. 5. Phylogram of α-CAs from various metazoan taxa. Included in the Bayesian analysis are the molluscan nacrein proteins known to be involved in biocalcification and two outgroup sequences (a bacterial and a green alga αCA). The topology shown is a 50% majority rule consensus tree. The two α-CAs retrieved from the oyster EST data set are indicated by black arrows.

biomineralization gene families, the *shematrins*, and the *KRMPs* (Yano et al. 2006; Zhang et al. 2006); that is, none of the RLCDs matched with genes outside the genus *Pinctada*. Pearl oyster ESTs representing these eight proteins were highly abundant in our data set (497 ESTs in total). The *shematrin* gene family was first identified in *P. fucata*, where it was proposed that these proteins form part of the insoluble organic framework within the prismatic layer, which overlies the nacreous layer (Yano et al. 2006). Our data suggest that the shematrin and KRMP proteins are also an important component of the nacreous layer of *P maxima*. Although we have not specifically investigated the expression levels of these genes within the mantle edge (thought to be the region responsible for prismatic layer deposition), or for the presence of their protein products within the shell itself, these differences suggest that

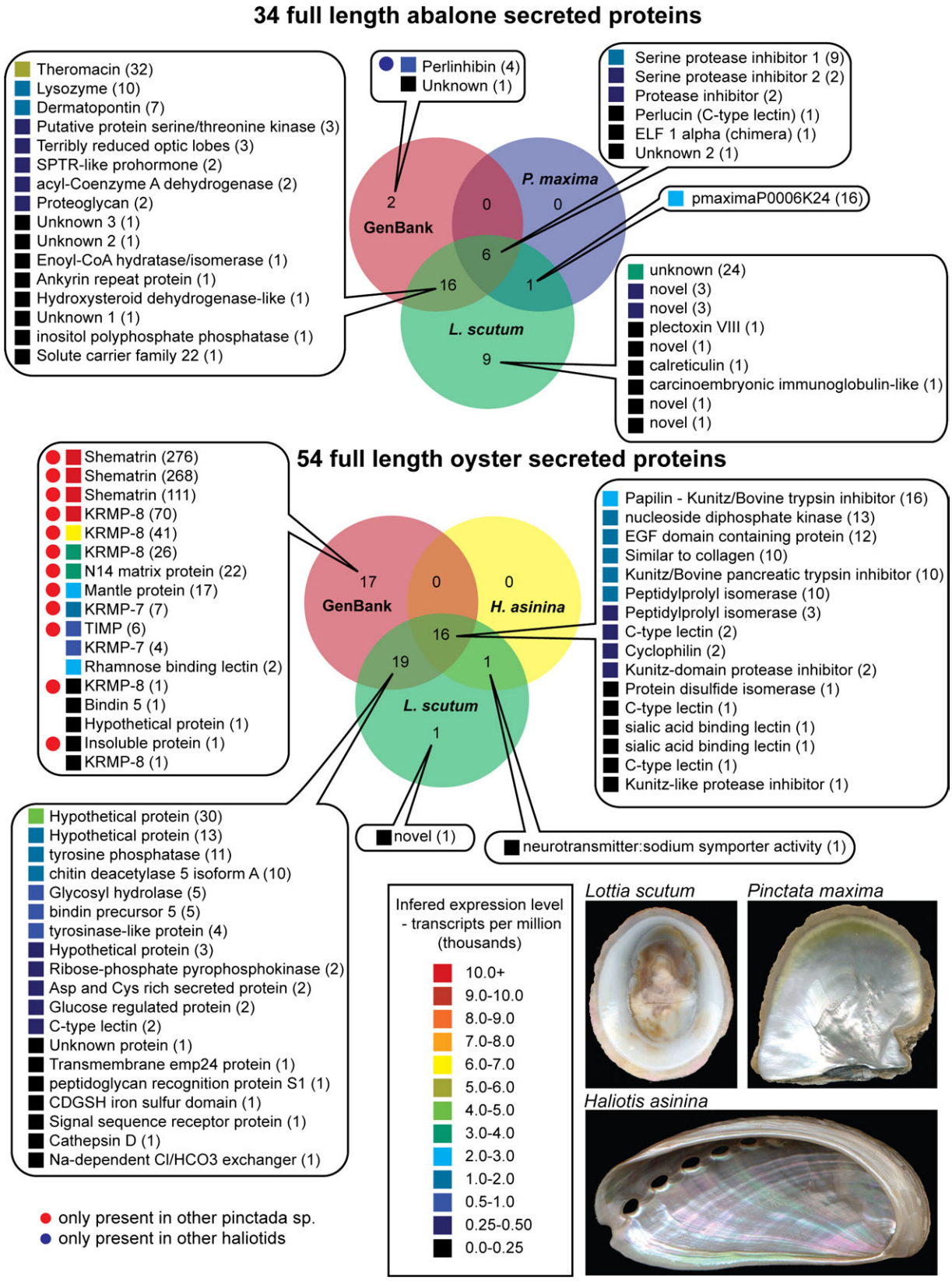

Fig. 6. Comparison of abalone and pearl oyster conceptually derived secreted transcriptomes. From each data set, putative full-length ORFs (ORFs >50 amino acids with a methionine and stop codon) possessing a signal sequence, lacking a trans-membrane domain and GPI anchor and predicted to be secreted by TargetP (Emanuelsson et al. 2007) were searched against the following databases: GenBank nr and dbEST; *Lottia* filtered model transcripts draft assembly v1.0; and *Lottia* mantle ESTs. The secreted abalone set was also searched against the secreted oyster set and vice versa using TBlastX. Normalized expression levels (transcripts per million, TPM) based on transcript abundance are indicated with a colorimetric scale. Numbers in brackets are the raw EST abundance counts. Views of the inner (ventral) surface of representative shells of each species are also shown.

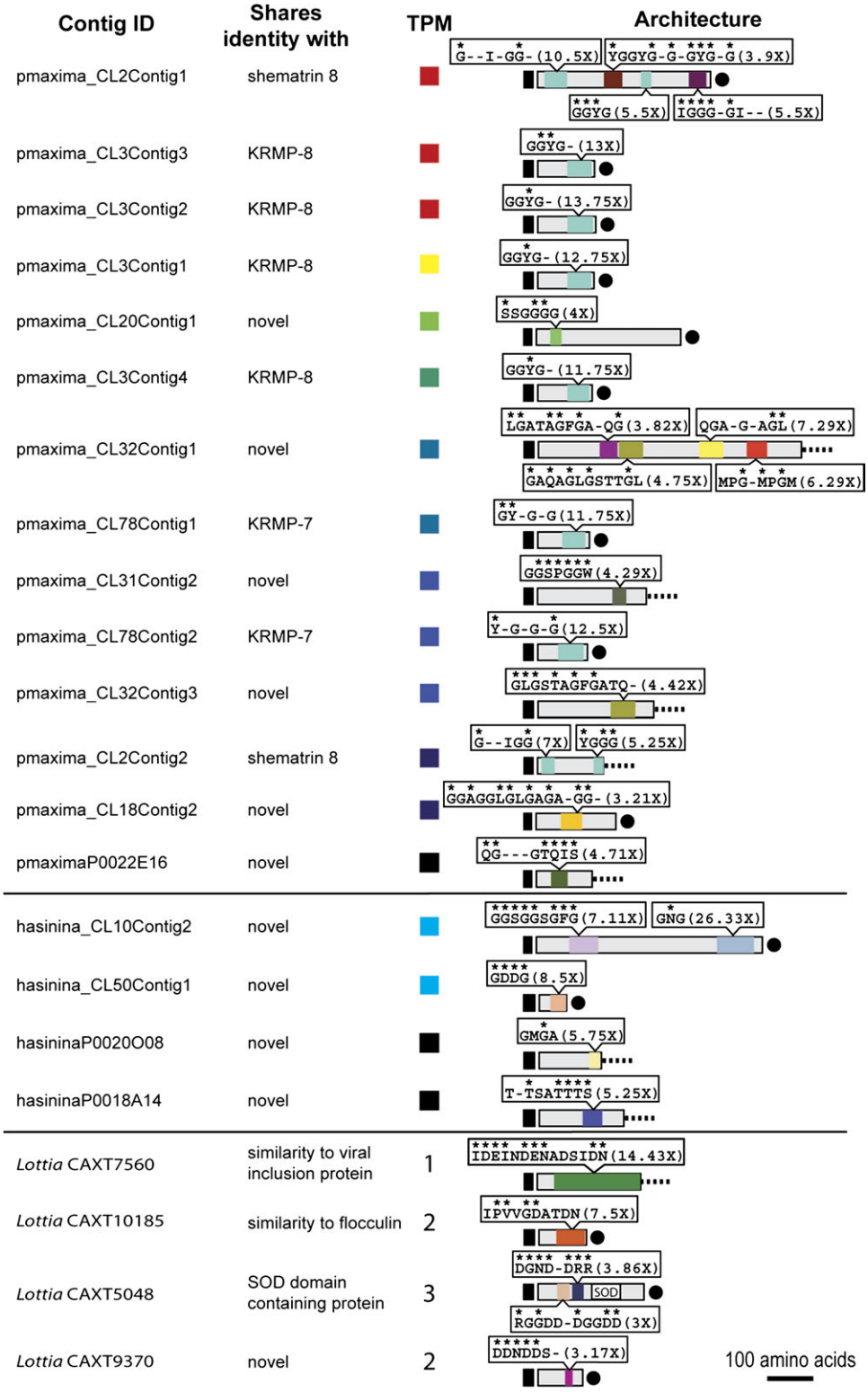

FIG. 7. Schematic representation of secreted abalone and pearl oyster proteins possessing TRs. Inferred expression levels are normalized to thousands of TPM (see fig. 5 for the corresponding colormetric legend), and the molecular architecture of the derived protein is represented to scale (scale bar lower right). Each derived protein sequence has a signal sequence (indicated as an amino-terminal black box), followed by the conceptually mature protein. Putative full-length ORFs are followed by a black circle at the carboxyl terminus and incomplete ORFs by a dashed line. TRs of similar architecture and sequence composition as determined by XSTREAM (Newman and Cooper 2007) share the same color. For each case, the consensus TR sequence is shown followed by the copy number (in brackets). Residues within each TR that are conserved 100% are indicated by an asterisk. Transcript abundances were not available for *Lottia gigantea*. Instead the presence/absence of each of the four limpet-derived proteins in various libraries is indicated. 1 = present in mantle, head, foot, visceral mass (radula), and heart cDNA libraries. 2 = present in mantle and female gonad cDNA libraries. 3 = present only in mantle cDNA libraries.

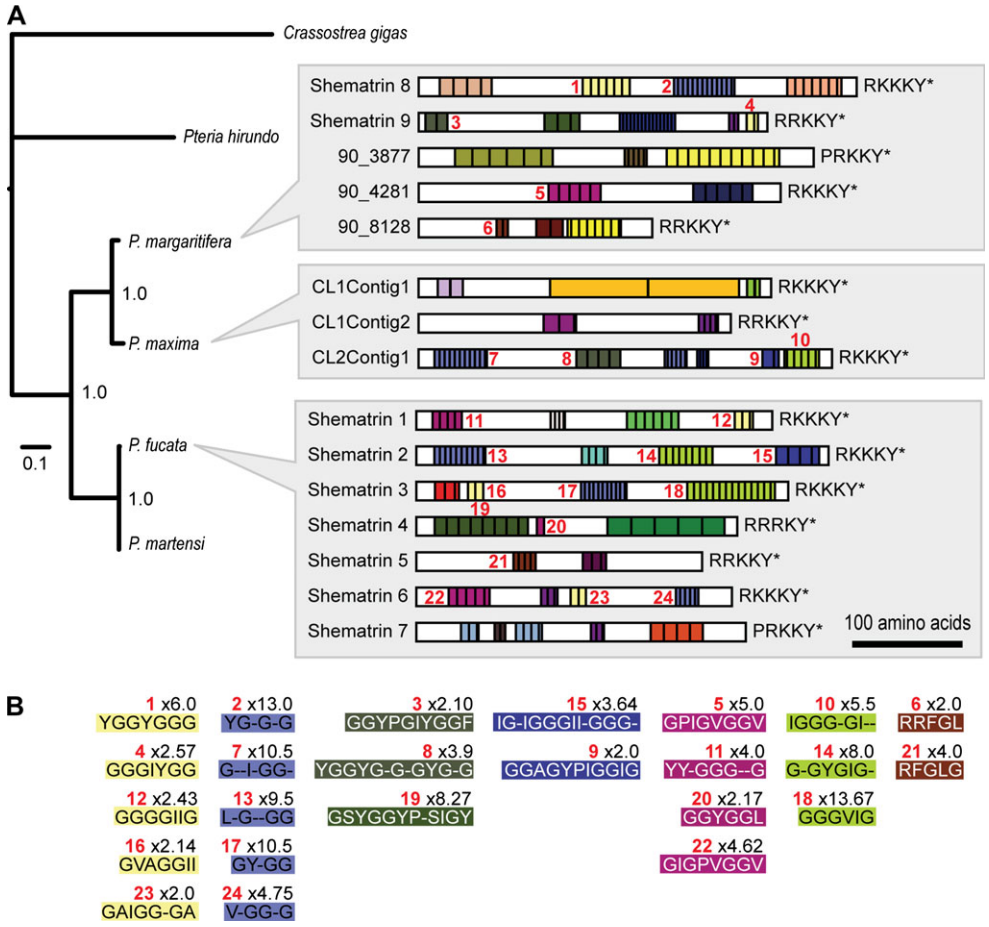

Fig. 8. Shematrin evolution across three pearl oyster lineages. (*A*) The relationships of four Pinctada species and two bivalve outgroups illustrated by a partitioned Bayesian analysis of COI and *16S rRNA*. Shematrin protein sequences from this study (*Pinctada maxima* and *Pinctada margaritifera*) and previously reported sequences (*Pinctada fucata*) are represented to scale. TR domains as detected by XSTREAM (Newman and Cooper 2007) within shematrin-like sequences are highlighted according to similarity. The novel *P. maxima* and *P. margaritifera* shematrins reported here are putatively full-length with a stop codon following the carboxy terminal tyrosine residue. Two *P. margaritifera* clusters shared similarity with shematrin 8 and 9. (*B*) Individual TR-sequence motifs detected by XSTREAM. The copy number of each repeat is given to the top right of each motif (black font). The identity of each TR is given to the top left of each motif (red font) and can be mapped to the appropriate schematic in (*A*).

shematrin and KRMP proteins are not restricted to the insoluble matrix of one particular shell layer.

Families of fast-evolving RLCD proteins in secreted structures appear common throughout the Metazoa. Arachnid silk fibroins and lepidopteran fibroins (which appear to have evolved independently) are comprised of highly repetitive regions (Mita et al. 1994; Gatesy et al. 2001). Similar RLCDs are found in byssal proteins in mussels (Qin et al. 1997) and insect chorion proteins (Burke and Eickbush 1986; Waring et al. 1990). The presence of these proteins in disparate taxa, the lack of sequence similarity in flanking nonrepetitive regions between closely related taxa, and the function of these proteins in taxon-specific, extracellular structural novelties suggests that selective pressures exist for the generation of these RCLD proteins. Once present, RLCD proteins appear to evolve rapidly, which may confer new mechanical properties to the structure they comprise (Gosline et al. 1999; Hayashi and Lewis 2000). This observation extends to the *shematrin* gene family present within *Pinctada* species. There appears to be conservation of at least two core motifs across the three oyster species surveyed, with lineage specific RLCD novelty within shematrin members reflecting phylogenetic distance between species (fig. 8). If we assume that all shematrin gene products fulfill the same function, it appears that relaxed purifying selection has permitted extensive shematrin sequence diversification, further supporting the notion that molluscan-specific shell-building genes are rapidly evolving.

Differences in Molluscan Nacre Ultrastructure and Its Organic Matrix

On a gross scale, the aragonitic innermost layer of an oyster shell and an abalone shell appear similar and are both described as being nacreous. However, there are significant differences in the arrangement of the $CaCO_3$ crystals, the axes along which these crystals are arranged (the crystal textures), and the way in which they are deposited. Studies by Hedegaard and colleagues (Hedegaard 1997; Hedegaard and Wenk 1998; Chateigner et al. 2000) have described gastropod nacre as "columnar," with stacks of tablets forming

FIG. 9. Alignment of cyclophilin genes from representative metazoan taxa. (*A*) Functionally characterized cyclophilin (derived from human) with active site residues (indicated by #) acts as the reference sequence. Sequences with active sites that differ from the reference sequence are in bold. Two sea urchin sequences known to be involved in biomineralization are indicated with asterisks.

along common axes, and bivalve nacre as "sheet nacre," with each crystal tablet out of register with its neighbors in a brick wall-like manner. Furthermore, the arrangement of the three axes that describe a crystal's orientation (*a*, *b*, and *c*) differ between abalone and oyster; the *a* and *b* faces of each $CaCO_3$ crystal of gastropod aragonite freely rotate about the *c* axis, whereas in bivalve sheet nacre, all crystal axes are aligned with the *b* axis oriented in the direction of shell growth (Hedegaard and Wenk 1998; Chateigner et al. 2000). Both gastropod and bivalve nacre orient the *c* axis perpendicular to the shell surface. These crystallographic observations suggest that the molecular mechanisms of $CaCO_3$ deposition differ between these two molluscan clades. The large-scale qualitative and quantitative differences in EST sets between *P. maxima* and *H. asinina* are compatible with these observed crystallographic differences.

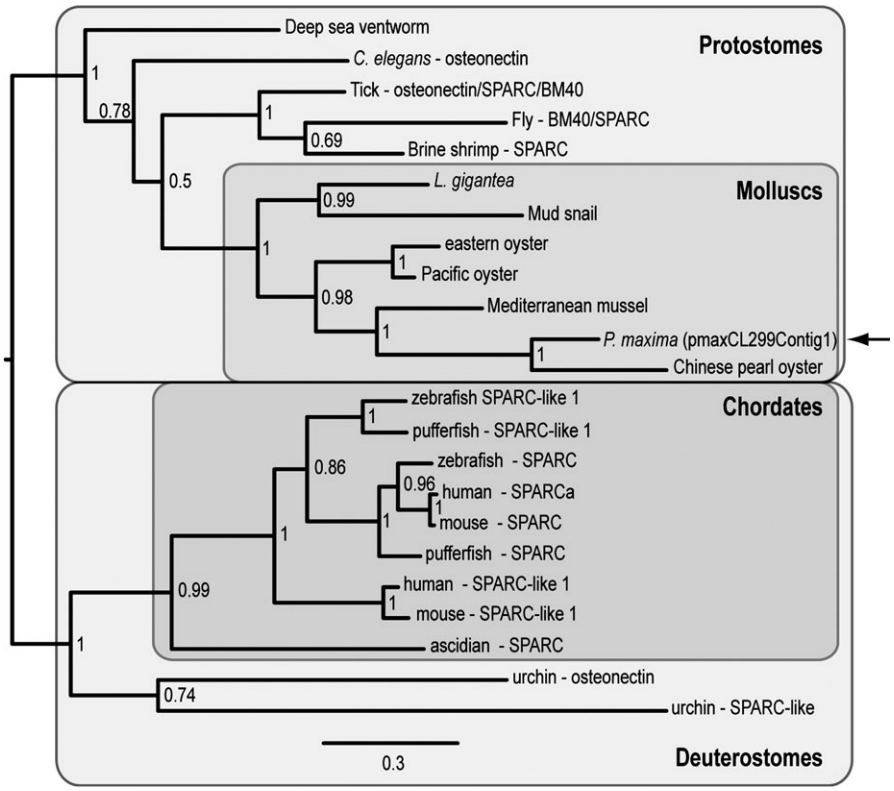

FIG. 10. Phylogram of SPARC/osteonectin proteins. Posterior probability values retrieved from eight runs with four chains each heated to 0.2 after 8 million generations are indicated. Trees were sampled every 1,000 generations. The topology shown is a 50% majority rule consensus tree and is midpoint rooted. The pearl oyster SPARC-like protein retrieved from the current data set is indicated by a black arrow.

Table 1. *Pinctada maxima* and *Haliotis asinina* Gene Products Associated with the Production, Modification, or Interaction with Chitin and Collagen, Their Associated Expression Levels, and Annotation Results.

Contig ID	Expression Level (TPM)	Similarity to (Best Hit GenBank Accession Number)	PFAM Domains
pmaxima_CL7Contig1	111 (16,476)	ELF-1α (BAD35019)	Elongation factor Tu GTP binding domain PF00009
pmaxima_CL22Contig1[a]	27 (4,008)	Chitin synthase (BAF73720)	Chitin synthase PF03142
pmaxima_CL90Contig1	10 (1,484)	Chitin deacetylase (NP_001103739)	Polysaccharide deacetylase PF01522
pmaxima_CL112Contig1	7 (1,039)	Chitin deacetylase (NP_001103739)	Polysaccharide deacetylase PF01522
pmaxima_CL147Contig1	5 (742)	Chitinase-like protein (CAI96024)	Glycosyl hydrolases family 18 PF00704
pmaxima_CL210Contig1	3 (445)	Septin (XP_001650957)	Septin PF00735
pmaxima_CL355Contig1	2 (297)	Chitinase (NP_777124)	Glycosyl hydrolases family 18 PF00704
pmaxima_CL475Contig1[a]	2 (297)	Chitin synthase (BAF73720)	—
pmaxima_CL74Contig1	10 (1,484)	Collagen type 1 α1 (NP_001116390)	Collagen PF01391
pmaxima_CL101Contig1	8 (1,187)	Collagen pro α-chain (BAA75669)	Collagen PF01391; COLF1 PF01410
pmaximaP0011J08	1 (148)	Polymorphic collagen-like (ABY19404)	Collagen PF01391
hasinina_CL3Contig1	71 (11,579)	ELF-1α (AAZ30688)	Elongation factor Tu GTP binding domain PF00009
hasininaP0009M03	1 (163)	Peritrophin-A (AAR02437)	Chitin-binding Peritrophin-A domain PF01607
hasinina_CL6Contig1	51 (8,317)	Collagen pro α-chain (BAA75668)	Collagen PF01391; fibrillar collagen C-terminal domain PF01410
hasinina_CL21Contig1	26 (4,240)	Collagen pro α-chain (BAA75669)	Fibrillar collagen C-terminal domain PF01410
hasininaP0002E06	1 (163)	Collagen pro α-chain (BAA75669)	Collagen PF01391

As an expression level reference, the housekeeping gene ELF-1α from the abalone and oyster data sets are included. Expression levels are presented as both the raw number of ESTs and as a normalized value (TPM = transcripts per million).

[a] According to alignments with chitin synthases, these two contigs are likely to represent the same gene product but have not assembled into a single contig due to insufficient overlap.

Primarily based on investigations of bivalve shells, a model of the organic matrix within which nacre deposition occurs has been developed (Levi-Kalisman et al. 2001; Falini et al. 2003; Addadi et al. 2006). This model describes interlamellar layers of the polysaccharide β-chitin between which a hydrogel of silk-like fibroin proteins maintains the required hydration and kinetic conditions required for aragonitic crystal growth to proceed (Addadi et al. 1987, 2006; Nudelman et al. 2006). Additionally, the aligned *a* and *b* axes of bivalve aragonite, as opposed to the randomly oriented axes of gastropod nacre, appear to result from the ordered arrangement of chitin fibrils and associated silk-fibroin-like proteins (Weiner and Traub 1984; Addadi and Weiner 1992). Interestingly, the presence of chitin in gastropod shells is a more contentious issue. Chitin has previously been detected in gastropod shells (Peters 1972; Poulicek et al. 1986) by histochemical and chromatographic techniques known to produce false positives (Keith et al. 1993). Furthermore, reports range from very low levels of chitin in gastropod shells (Marxen et al. 1998; Furuhashi et al. 2009) to no chitin in the shell of *H. rufescens* (Keith et al. 1993). In order to infer from a genetic perspective whether *H. asinina* and *P. maxima*, both employ chitin as a biocalcification scaffolding molecule we searched for genes involved in chitin production, modification, and interaction. Interestingly, we found nine contigs within the *P. maxima* data set related to chitin biosynthesis and/or modification (table 1). The two most abundantly expressed chitin-related gene products within the *P. maxima* data set were a chitin synthase and a chitin deacetylase. Transcripts of chitin synthase in *P. maxima* were detected at expression levels 24.3% that of elongation factor 1 alpha (ELF1α), a constitutively transcribed housekeeping gene required for protein synthesis. In contrast, we could identify only a single EST with a potential chitin binding capacity in the *H. asinina* data set (table 1). This supports previous observations that chitin is an integral component of the organic matrix of pearl oyster nacre (Weiss and Schönitzer 2006; Weiss et al. 2006; Suzuki et al. 2007), whereas for *H. asinina*, chitin is likely to play a less significant role in scaffolding the calcification process. This observation may explain the differences in the crystallographic arrangement of *a* and *b* axes between pearl oyster and abalone nacre; chitin and associated fibroin-silk proteins are thought to align the *a* and *b axes* of oyster nacre (Weiner and Traub 1980,1984), whereas crystals of abalone nacre that may lack the orientating influence of a chitin scaffold would be free to rotate about the *c* axis (Hedegaard 1997; Hedegaard and Wenk 1998; Chateigner et al. 2000).

We also detected a marked difference in lyase activity between the two data sets, primarily due to the relatively high expression of two α-CA isoforms in *P. maxima* (fig. 4). Despite the fact that we could not assign these *P. maxima* gene products to any particular α-CA class, they were robustly excluded from a clade composed of the molluscan nacrein proteins (fig. 5). Interestingly, no α-CA isoforms were detected in the *H. asinina* gene set.

A Conserved Bilaterian Biomineralization "Toolkit"

Using the recently sequenced sea urchin genome, Livingston et al. (2006) conducted targeted searches for genes implicated in biomineralization from both echinoderms and vertebrates. The identification of certain shared genes in the early biomineralization pathway of both of these groups led the authors to propose the presence of a common toolkit underlying biomineralization mechanisms in distantly related taxa. Building on this work, we searched our molluscan data sets for genes implicated in deuterostome biomineralization. Cyclophilins are

peptidyl-prolyl cis-trans isomerases that are found in both prokaryotes and eukaryotes and facilitate protein folding. Although their specific role in calcification is not known, Livingston et al. found eight cyclophilin genes expressed in sea urchin primary mesenchyme cells (PMCs), which are responsible for the initiation of spiculogenesis (Yamasu and Wilt 1999). One of these genes, *Spcyp-1*, has been shown to be a member of the skeletogenic gene battery (Amore and Davidson 2006). An additional cyclophilin domain–containing gene product was demonstrated by Livingston et al. (2006) to be highly expressed in PMC cells during the early stages of spicule formation. We found seven cyclophilin genes in *P. maxima* and one in *H. asinina* (fig. 9). Interestingly, when we aligned a representative set of cyclophilins to a reference sequence for which active site residues have been determined (Zhao and Ke 1996), we found that two of most highly expressed sea urchin cyclophilin-like proteins in PMC cells had mutations at active site positions (fig. 9). We also detected similar critical mutations in three of the seven *P. maxima* cyclophilin-like proteins. This suggests that an ancestral cyclophilin gene may have been exapted from a deuterostome–protostome ancestor into a biomineralization role and no longer possess peptidyl-prolyl cis-trans isomerase activity.

Secreted calcium-binding phosphoproteins are involved in a large number of biocalcification roles in vertebrates and are thought to be the result of vertebrate-specific duplications of an ancestral SPARC gene (Kawasaki et al. 2004; Kawasaki and Weiss 2006). We have detected a *P. maxima* gene with similarity to SPARC in our data set, and additional searches of dbEST revealed several other lophotrochozoan SPARC-like gene products. This is the first time a SPARC gene has been reported from a lophotrochozoan biomineralizing tissue. Due to this increased taxon sampling, our phylogenetic reconstruction of this family of proteins returns a slightly different result with both sea urchin sequences forming a sister group to the chordates. This result supports the suggestion that the sea urchin SPARC/osteonectin genes are paralogs and that multiple chordate SPARC genes have also arisen by lineage-specific duplications. As our tree is midpoint rooted (due to the lack of a suitable outgroup sequence), we are unable to infer the direction of early SPARC evolution. Searches of the *Amphimedon queenslandica* (demosponge) genome returned a moderately significant hit (1E−04); however, including this protein model in SPARC alignments greatly reduced alignment quality.

Type I collagen is the major matrix protein of vertebrate dentin and bone (Butler and Ritchie 1995) and is known to be involved in sea urchin spiculogenesis (Benson et al. 1990; Wessel et al. 1991). Collagen is present in the extracellular matrix of many tissues, and although we cannot specifically assign a biomineralization role to the collagens identified here, the fact that one of these gene products is the fifth most abundant transcript in the abalone data set (another is also highly expressed—15th most abundant) suggests that collagen may play a role in nacre deposition in the abalone shell. A collagen-like domain in the haliotid biocalcification gene Lustrin-A was observed by Shen et al. (1997); however, this domain lacks key regular third glycine residues, and no hydroxyproline was detected by amino acid analysis, both of which are two key features of collagen. Blank et al. (2003) suggested on the basis of morphology that organic fibrils observed using atomic force microscopy of abalone nacre were collagen; however, a specific biochemical test for collagen was not conducted. Our EST data support their model of collagen fibrils surrounding *a* and *b* faces of each nacre tablet. The relative paucity of collagen transcripts in the *P. maxima* data set further suggests that the organic framework of these two molluscan shells is fundamentally different.

Implications for Nacre Evolution

Despite the long fossil record of shelled molluscs (Colgan et al. 2000), a clear understanding of conchiferan evolution is complicated by the fact that shell characters alone usually do not provide adequate phylogenetic signal (Ponder and Lindberg 2008). Furthermore, Cambrian molluscs are known to be morphologically and phylogenetically distinct from contemporary forms (Parkhaev 2008). Molecular and morphological analyses of extant groups often yield conflicting phylogenetic topologies generating much debate concerning sister-group relationships within the Mollusca (Ponder and Lindberg 2008); however, larger molecular data sets that include difficult to sample early branching taxa are yielding increasingly resolved topologies (Wilson et al. 2009). Until these phylogenetic issues are resolved, a complete understanding of the evolutionary origins of nacre will remain obscure. Nonetheless, imprints of nacreous microstructures can be found in small shelly fossils common to Cambrian strata, indicating that this material has long been used by molluscs (Parkhaev 2008). Among contemporary molluscan taxa, nacre is restricted to gastropods (columnar), bivalves (sheet), cephalopods (columnar and sheet) and recently has been shown to be present in possibly the earliest branching conchiferans, the monoplacophorans (Checa et al. 2009). Interestingly, foliated aragonite, which Checa et al. (2009) propose has replaced true nacre in the Neopilinidae, displays significant crystallographic differences to both bivalve and gastropod nacre; in monoplacophorans, the *a* axis is oriented perpendicular to the direction of growth, in bivalves, it is the *b* axis, and in gastropods, the *a* and *b* axes rotate randomly about the *c* axis. These observations suggest the molecular mechanisms that guide the deposition of the variants of nacre and its derivatives across the Mollusca are fundamentally different.

Conclusion

The results of the general and targeted comparisons described here suggest that there may be significant differences in the molecular mechanisms used by *P. maxima* and *H. asinina* for nacre deposition. This indicates that the Bivalvia and Gastropoda have either independently evolved the ability to deposit nacre or that subsequent to the genesis of this ability in a common ancestor, bivalves and gastropods have significantly modified the molecular mechanisms that guide this process. We suggest that these

molecular differences are likely to relate to the crystallographic differences observed between the two classes (Hedegaard and Wenk 1998; Chateigner et al. 2000), further highlighting the importance of organic molecules in the process of biocalcification. The degree of gene novelty and differences between the molluscs analyzed here also highlights the importance of the evolution of coding sequences to the generation of metazoan morphological novelty. In particular, the evolution and diversification of novel RLCD proteins is apparently a key feature of molluscan shell evolution. These features provide a genome-level explanation for the diversity of shell structures and patterns. We anticipate that the comparative approach used here will assist in the identification of genes specifically necessary for the deposition of nacre, a substance of significant interest to materials scientists.

Supplementary Material

Supplementary figure 1 and supplementary tables 1–3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org).

Acknowledgments

We acknowledge the JGI Community Sequencing Program for providing public access to a draft assembly of the *L. gigantea* genome and Eric Edsinger Gonzales for discussions regarding *Lottia* sequence data. This work was supported by funding from the Australian Research Council to B.M.D. and the Deutsche Forschungsgemeinschaft to D.J.J. through the Courant Research Centre for Geobiology (Göttingen) and the German Initiative of Excellence.

References

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

Addadi L, Moradian J, Shay E, Maroudas NG, Weiner S. 1987. A chemical-model for the cooperation of sulfates and carboxylates in calcite crystal nucleation— relevance to biomineralization. *Proc Natl Acad Sci USA*. 84:2732–2736.

Addadi L, Weiner S. 1992. Control and design principles in biological mineralization. *Angew Chem*. 31:153–169.

Amore G, Davidson EH. 2006. *cis*-Regulatory control of cyclophilin, a member of the ETS-DRI skeletogenic gene battery in the sea urchin embryo. *Dev Biol*. 293:555–564.

Bedouet LB, Rusconi F, Rousseau M, Duplat D, Marie A, Dubost L, Le Ny K, Berland S, Péduzzi J, Lopez E. 2006. Identification of low molecular weight molecules as new components of the nacre organic matrix. *Comp Biochem Physiol B*. 144:532–543.

Belcher AMB, Wu XH, Christensen RJ, Hansma PK, Stucky GD, Morse DE. 1996. Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature* 381:56–58.

Benson S, Smith L, Wilt F, Shaw R. 1990. The synthesis and secretion of collagen by cultured sea urchin micromeres. *Exp Cell Res*. 188:141–146.

Blank S, Arnoldi M, Khoshnavaz S, Treccani L, Kuntz M, Mann K, Grathwohl G, Fritz M. 2003. The nacre protein perlucin nucleates growth of calcium carbonate crystals. *J Microsc*. 212:280–291.

Burke WD, Eickbush TH. 1986. The silkmoth late chorion locus. I. Variation within two paired multigene families. *J Mol Biol*. 190:343–356.

Butler WT, Ritchie H. 1995. The nature and functional significance of dentin extracellular matrix proteins. *Int J Dev Biol*. 39:169–179.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552.

Chao S, Lazo GR, You F, et al. (20 co-authors). 2006. Use of a large-scale Triticeae expressed sequence tag resource to reveal gene expression profiles in hexaploid wheat (*Triticum aestivum* L.). *Genome* 49:531–544.

Chateigner D, Hedegaard C, Wenk H. 2000. Mollusc shell microstructures and crystallographic textures. *J Struct Geol*. 22:1723–1735.

Checa A, Ramírez-Rico J, González-Segura A, Sánchez-Navas A. 2009. Nacre and false nacre (foliated aragonite) in extant monoplacophorans (= Tryblidiida: Mollusca). *Naturwissenschaften*. 96:111–122.

Chen Z, Wang W, Ling XB, Liu JJ, Chen L. 2006. GO-Diff: mining functional differentiation between EST-based transcriptomes. *BMC Bioinform*. 7:72.

Colgan DJ, Ponder WF, Eggler PE. 2000. Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28S rDNA and histone H3 sequences. *Zool Scripta*. 29:29–63.

Cusack M, Freer A. 2008. Biomineralization: elemental and organic influence in carbonate systems. *Chem Rev*. 108:4433–4454.

Edgar R. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform*. 5:113.

Eisenhaber B, Bork P, Eisenhaber F. 1999. Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol*. 292:741–758.

Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Prot*. 2:953–971.

Ewing R, Kahla A, Poirot O, Lopez F, Audic S, Claverie J. 1999. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res*. 9:950–959.

Falini G, Weiner S, Addadi L. 2003. Chitin-silk fibroin interactions: relevance to calcium carbonate formation in invertebrates. *Cal Tiss Int*. 72:548–554.

Fu G, Valiyaveettil S, Wopenka B, Morse DE. 2005. $CaCO_3$ biomineralization: acidic 8-kDa proteins isolated from aragonitic abalone shell nacre can specifically modify calcite crystal morphology. *Biomacromolecules* 6:1289–1298.

Furuhashi T, Beran A, Blazso M, Czegeny Z, Schwarzinger C, Steiner G. 2009. Pyrolysis GC/MS and IR spectroscopy in chitin analysis of molluscan shells. *Biosci Biotechnol Biochem*. 73:93–103.

Gatesy J, Hayashi C, Motriuk D, Woods J, Lewis R. 2001. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* 291:2603–2605.

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*. 130:17519–17527.

Gong N, Shangguan J, Liu X, Yan Z, Ma Z, Xie L, Zhang R. 2008. Immunolocalization of matrix proteins in nacre lamellae and their in vivo effects on aragonitic tablet growth. *J Struct Biol*. 164:33–40.

Gosline JM, Guerette PA, Ortlepp CS, Savage KN. 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J Exp Biol*. 202:3295–3303.

Gotliv BA, Addadi L, Weiner S. 2003. Mollusk shell acidic proteins: in search of individual functions. *Chem Biochem*. 4:522–529.

Hayashi CY, Lewis RV. 2000. Molecular architecture and evolution of a modular spider silk protein gene. *Science* 287:1477–1479.

Hedegaard C. 1997. Shell structures of the recent Vetigastropoda. *J Moll Stud*. 63:369–377.

Hedegaard C, Wenk H. 1998. Microstructure and texture patterns of mollusc shells. *J Moll Stud*. 64:133–136.

Horne F, Tarsitano S, Lavalli KL. 2002. Carbonic anhydrase in mineralization of the crayfish cuticle. *Crustaceana* 75:1067–1081.

Hotz-Wagenblatt A, Hankeln T, Ernst P, Glatting KH, Schmidt ER, Suhai S. 2003. ESTAnnotator: a tool for high throughput EST annotation. *Nucleic Acid Res*. 31:3716–3719.

Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res*. 9:868–877.

Jackson DJ, Macis L, Reitner J, Degnan BM, Wörheide G. 2007. Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. *Science* 316:1893–1895.

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

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

Jolly C, Berland S, Milet C, Borzeix S, Lopez E, Doumenc D. 2004. Zonal localization of shell matrix proteins in mantle of *Haliotis tuberculata* (Mollusca, Gastropoda). *Mar Biotech*. 6:541–551.

Kawasaki K, Suzuki T, Weiss KM. 2004. Genetic basis for the evolution of vertebrate mineralized tissue. *Proc Natl Acad Sci USA*. 101:11356–11361.

Kawasaki K, Weiss KM. 2006. Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *J Exp Biol*. 306:295–316.

Keith J, Stockwell S, Ball D, Remillard K, Kaplan D, Thannhauser T, Sherwood R. 1993. Comparative analysis of macromolecules in mollusc shells. *Comp Biochem Physiol B*. 105:487–496.

Leu J, Chang C, Wu J, Hsu C, Hirono I, Aoki T, Juan H, Lo C, Kou G, Huang H. 2007. Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected *Penaeus monodon*. *BMC Gen*. 8:120.

Levi-Kalisman Y, Falini G, Addadi L, Weiner S. 2001. Structure of the nacreous organic matrix of a bivalve mollusk shell examined in the hydrated state using cryo-TEM. *J Struct Biol*. 135:8–17.

Livingston BT, Killian CE, Wilt H, Cameron A, Landrum MJ, Ermolaeva O, Sapojnikov V, Maglott DR, Buchanan AM, Ettensohn CA. 2006. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. *Dev Biol*. 300:335–348.

Ma Z, Huang J, Sun J, Wang G, Li C, Xie L, Zhang R. 2007. A novel extrapallial fluid protein controls the morphology of nacre lamellae in the pearl oyster, *Pinctada fucata*. *J Biol Chem*. 282:23253–23263.

Marchler-Bauer A, Anderson JB, Chitsaz F, et al. (28 co-authors) 2009. CDD: specific functional annotation with the Conserved Domain Database. *Nuc Acid Res*. 37:D205–D10.

Marin F, Luquet G. 2004. Molluscan shell proteins. *Comptes Rendus Palevol*. 3:469–492.

Marxen J, Hammer M, Gehrke T, Becker W. 1998. Carbohydrates of the organic shell matrix and the shell-forming tissue of the snail *Biomphalaria glabrata* (Say). *Biol Bull*. 194:231–240.

Meldrum FC, Coölfen H. 2008. Controlling mineral morphologies and structures in biological and synthetic systems. *Chem Rev*. 108:4332–4432.

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*. 98:268102.

Mita K, Ichimura S, James TC. 1994. Highly repetitive structure and its organization of the silk fibroin gene. *J Mol Evol*. 38:583–592.

Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A. 1996. A carbonic anhydrase from the nacreous layer in oyster pearls. *Proc Natl Acad Sci USA*. 93:9657–9660.

Miyamoto H, Miyoshi F, Kohno J. 2005. The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*. *Zool Sci*. 22:311–315.

Miyamoto H, Yano M, Miyashita T. 2003. Similarities in the structure of nacrein, the shell-matrix protein, in a bivalve and a gastropod. *J Moll Stud*. 69:87–89.

Moya A, Tambutté S, Bertucci A, Tambutté E, Lotto S, Vullo D, Supuran CT, Allemand D, Zoccola D. 2008. Carbonic anhydrase in the scleractinian coral *Stylophora pistillata*: characterization, localization, and role in biomineralization. *J Biol Chem*. 283: 25475–25484.

Newman AM, Cooper JB. 2007. XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinform*. 8:382.

Nishikata T, Yamada L, Mochizuki Y, Satou Y, Shin IT, Kohara Y, Satoh N. 2001. Profiles of maternally expressed genes in fertilized eggs of *Ciona intestinalis*. *Dev Biol*. 238:315–331.

Nudelman F, Gotliv BA, Addadi L, Weiner S. 2006. Mollusk shell formation: mapping the distribution of organic matrix components underlying a single aragonitic tablet in nacre. *J Struct Biol*. 153:176–187.

Parkhaev PY. 2008. The early Cambrian radiation of Mollusca. In: Ponder WF, Lindberg DR, editors. Phylogeny and evolution of the Mollusca. Berkeley (CA): University of California Press. p. 33–69.

Perina D, Cetkovic H, Harcet M, Premzl M, Lukic-Bilela L, Müller WEGM, Gamulin V. 2006. The complete set of ribosomal proteins from the marine sponge *Suberites domuncula*. *Gene* 366:275–284.

Peters W. 1972. Occurrence of chitin in Mollusca. *Comp Biochem Physiol B*. 41:541–550.

Ponder WF, Lindberg DR. 2008. Phylogeny and evolution of the Mollusca. Berkeley (CA): University of California Press.

Poulicek M, Voss-Foucart MF, Jeniaux C. 1986. Chitinoproteic complexes and mineralization in mollusk skeletal structures. In: Muzzarelli R, Jeuniaux C, Gooday GW, editors. Chitin in nature and technology. New York: Plenum Press. p. 7–12.

Qin XX, Coyne KJ, Waite JH. 1997. Tough tendons. Mussel byssus has collagen with silk-like domains. *J Biol Chem*. 272:32623–32627.

Satoh N, Satou Y, Davidson B, Levine M. 2003. *Ciona intestinalis*: an emerging model for whole-genome analyses. *Trends Genet*. 19:376–381.

Schmid R, Blaxter ML. 2008. annot8r: gO, EC and KEGG annotation of EST datasets. *BMC Bioinform*. 9:180.

Shen X, Belcher AM, Hansma PK, Stucky GD, Morse DE. 1997. Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *J Biol Chem*. 272:32472–32481.

Shen Z, Jacobs-Lorena M. 1998. A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin. *J Biol Chem*. 273:17665–17670.

Sigwart J, Sutton M. 2007. Deep molluscan phylogeny: synthesis of palaeontological and neontological data. *Proc Biol Sci*. 274: 2413–2419.

Suzuki M, Sakuda S, Nagasawa H. 2007. Identification of chitin in the prismatic layer of the shell and a chitin synthase gene from the Japanese pearl oyster, *Pinctada fucata*. *Biosci Biotechnol Biochem*. 71:1735–1744.

Takahashi T, McDougall C, Troscianko J, Chen WC, Jayaraman-Nagarajan A, Shimeld SM, Ferrier DE. 2009. An EST screen from the annelid *Pomatoceros lamarckii* reveals patterns of gene loss and gain in animals. *BMC Evol Biol*. 9:240.

Vera JC, Wheat C, Fescemyer H, Frilander M, Crawford D, Hanski I, Marden J. 2008. Rapid transcriptome characterization for a non-model organism using 454 pyrosequencing. *Mol Ecol*. 17:1636–1647.

Waring GL, Hawley RJ, Schoenfeld T. 1990. Multiple proteins are produced from the dec-1 eggshell gene in *Drosophila* by alternative RNA splicing and proteolytic cleavage events. *Dev Biol*. 142:1–12.

Weiner S, Traub W. 1980. X-ray diffraction study of the insoluble organic matrix of mollusk shells. *FEBS Lett*. 111:311–316.

Weiner S, Traub W. 1984. Macromolecules in mollusk shells and their functions in biomineralization. *Philos Trans R Soc Lond B*. 304:425–434.

Weiss I, Schönitzer V. 2006. The distribution of chitin in larval shells of the bivalve mollusk *Mytilus galloprovincialis*. *J Struct Biol*. 153:264–277.

Weiss I, Schönitzer V, Eichner N, Sumper M. 2006. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. *FEBS Lett*. 580:1846–1852.

Wessel GM, Etkin M, Benson S. 1991. Primary mesenchyme cells of the sea urchin embryo require an autonomously produced, nonfibrillar collagen for spiculogenesis. *Dev Biol*. 148:261–272.

Wilson NG, Rouse GW, Giribet G. 2009. Assessing the molluscan hypothesis Serialia (Monoplacophora + Polyplacophora) using novel molecular data. *Mol Phylo Evol*. doi:10.1016/j.ympev.2009.07.028.

Yamasu K, Wilt FH. 1999. Functional organization of DNA elements regulating SM30alpha, a spicule matrix gene of sea urchin embryos. *Dev Growth Differ*. 41:81–91.

Yano M, Nagai K, Morimoto K, Miyamoto H. 2006. Shematrin: a family of glycine-rich structural proteins in the shell of the pearl oyster *Pinctada fucata*. *Comp Biochem Physiol B*. 144: 254–262.

Zhang C, Xie L, Huang J, Liu X, Zhang R. 2006. A novel matrix protein family participating in the prismatic layer framework formation of pearl oyster, *Pinctada fucata*. *Biochem Biophys Res Commun*. 344:735–740.

Zhao Y, Ke H. 1996. Crystal structure implies that cyclophilin predominantly catalyzes the *trans* to *cis* isomerization. *Biochemistry* 35:7356–7361.