EVOLUTION & DEVELOPMENT 16:3, 139–148 (2014)
DOI: 10.1111/ede.12074

Horizontal transfer of the *msp130* gene supported the evolution of metazoan biomineralization

Charles A. Ettensohn*

Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA

*Author for correspondence (e-mail: ettensohn@andrew.cmu.edu)

SUMMARY It is widely accepted that biomineralized structures appeared independently in many metazoan clades during the Cambrian. How this occurred, and whether it involved the parallel co-option of a common set of biochemical and developmental pathways (i.e., a shared biomineralization "toolkit"), are questions that remain unanswered. Here, I provide evidence that horizontal gene transfer supported the evolution of biomineralization in some metazoans. I show that Msp130 proteins, first described as proteins expressed selectively by the biomineral-forming primary mesenchyme cells of the sea urchin embryo, have a much wider taxonomic distribution than was previously appreciated. Msp130 proteins are present in several invertebrate deuterostomes and in one protostome clade (molluscs). Surprisingly, closely related proteins are also present in many bacteria and several algae, and I propose that *msp130* genes were introduced into metazoan lineages via multiple, independent horizontal gene transfer events. Phylogenetic analysis shows that the introduction of an ancestral *msp130* gene occurred in the sea urchin lineage more than 250 million years ago and that *msp130* genes underwent independent, parallel duplications in each of the metazoan phyla in which these genes are found.

INTRODUCTION

The Cambrian explosion was characterized by the widespread and relatively synchronous emergence of biomineralization in many metazoan lineages (Knoll 2003; Murdock and Donohue 2011). These early evolutionary events, and the modification of biomineralization programs during the more than 500 million years that followed, have led to the remarkable diversity of biomineralized structures produced by modern metazoans. Reconstructing the evolution of biomineralization at the genetic, molecular, and developmental levels is a major challenge. Although it is widely accepted that biomineralized structures, in the strictest sense, appeared independently in many major metazoan groups, an important unanswered question concerns the extent to which this occurred by exploiting a common biomineralization "toolkit"—that is, a set of ancestral biochemical and developmental pathways that was independently co-opted for biomineral formation in diverse animal taxa (Westbroek and Marin 1998; Jackson et al. 2007; Murdock and Donohue 2011; Wörheide and Jackson 2011).

Recently, Jackson and co-workers reported that Spherulin, an abundant protein expressed selectively by the calcifying cells of the demosponge *Astrosclera willeyana*, is encoded by a gene that was introduced into the sponge genome from bacteria via horizontal gene transfer (HGT) (Jackson et al. 2011). Their findings suggest that HGT was an important mechanism that supported the evolution of biomineralization within this early-branching metazoan clade. The presence of Spherulin orthologs in two sponge lineages that diverged ~250 million years ago (MYA), but not in other metazoan species, suggests that this HGT event, though ancient, occurred after the diversification of animal phyla. HGT was probably facilitated by the close symbiotic relationships between sponges and diverse microbial communities (Hentschel et al. 2012). These relationships may have played a unique role in the evolution of biomineralization in sponges, as *A. willeyana* uses the degraded remains of bacteria to seed calcification (Jackson et al. 2010).

The formation of the $CaCO_3$-based endoskeleton of sea urchins (Phylum Echinodermata) has been studied intensively (Wilt and Ettensohn 2007; Killian and Wilt 2008; Ettensohn 2013). The lineage, differentiation, and morphogenesis of the cells that produce the embryonic skeleton (primary mesenchyme cells, or PMCs) are well understood. In particular, the program of gene expression in this embryonic lineage has been analyzed in considerable detail. A complex transcriptional network that controls the embryonic specification of PMCs has been elucidated (Oliveri et al. 2008; Rafiq et al. 2012, 2014), and many downstream effector genes that regulate the formation and patterning of the sea urchin skeleton have been identified, including many genes that are directly involved in biomineralization (Zhu et al. 2001; Illies et al. 2002; Livingston et al. 2006; Mann et al. 2010; Rafiq et al. 2012, 2014).

Msp130 (Mesenchyme-Specific-Protein, 130 KD) was the first PMC-specific gene product identified (Anstrom et al. 1987; Leaf et al. 1987). Subsequent studies revealed that Msp130 is a member of a small protein family that also includes Msp130-

related1, 2, 3, 4, 5, and 6 (Msp130rel1-6) (Illies et al. 2002; Livingston et al. 2006). Proteins of the Msp130 family contain an N-terminal signal peptide and are localized on the PMC surface via GPI linkages (Parr et al. 1990). Msp130, Msp130rel1, Msp130rel2, and Msp130rel3 are expressed zygotically during embryogenesis and are entirely restricted to PMCs (Leaf et al. 1987; Illies et al. 2002; Livingston et al. 2006). Transcript levels peak during early gastrulation, when *msp130* is one of the most abundant mRNAs in PMCs (~250 transcripts/cell). At the same stage, *msp130rel2* is expressed at an intermediate level (~150 transcripts/cell, and *msp130rel1* and *msp130rel3* at lower levels (~40 and ~10 transcripts/cell, respectively) (Rafiq et al., 2014). Early in gastrulation, transcripts of these genes are expressed relatively uniformly by all PMCs, but later in embryogenesis, mRNA expression is differentially regulated within the PMC syncytial network, with the highest levels of expression at sites of active skeletal rod growth (Harkey et al. 1992; Illies et al. 2002). Msp130 and Msp130rel1-3 are also the predominant Msp130 family proteins in biomineralized tissues of the adult (test, spines, and teeth) (Mann et al. 2008a, 2008b).

This study describes the phylogenetic distribution of Msp130 proteins. Multiple Msp130 family members are found in three invertebrate deuterostome clades (echinoderms, hemichordates, and cephalochordates) and in a single protostome clade (molluscs), where at least one Msp130 family member is expressed selectively by the shell-forming mantle. Remarkably, a single member of the Msp130 family is present in many bacterial species and in some algae. The gene is usually annotated as an atypical alkaline phosphatase, although its biochemical function is unknown. The phylogenetic analysis presented here suggests that independent HGT events transferred an ancestral *msp130* gene into multiple metazoan genomes, where the gene expanded by duplication and became functionally co-opted to support biomineralization. These findings, coupled with those of Jackson et al. (2011), show that multiple HGT events supported the evolution of biomineralization in metazoans.

MATERIALS AND METHODS

To identify homologs of Msp130, the Msp130 protein sequence from the purple sea urchin, *Strongylocentrotus purpuratus*, was used to query the current NCBI non-redundant protein database by BLAST-P, using a large number of taxon-specific searches (Table S1). The Ensemble (www.ensembl.org), JGI (genome.jgi.doe.gov), and SpBase (www.spbase.org) genome browsers were also used for BLAST-P searches of the most recent assemblies of many metazoan genomes.

A list of all protein sequences used for phylogenetic analysis is shown in Table S2. Phylogenetic trees were constructed using Guidance (http://guidance.tau.ac.il) (Penn et al. 2010) and MEGA5 (v. 5.2.1) (Tamura et al. 2011; Hall 2013). Briefly, guidance was used to generate MAFFT-based multiple sequence alignments (MSAs) and amino acids that could not be aligned with confidence (columns with Guidance scores < 0.25) were removed. MEGA5 was used to determine the optimal substitution model and to construct maximum likelihood (ML) trees without any further deletion of gaps and with a bootstrap value of 500.

RESULTS

The *msp130* gene family in sea urchins

To identify the complete set of Msp130-related proteins in the assembled genome of the purple sea urchin, *S. purpuratus*, the amino acid sequence of SpMsp130 was used to query the current collection of protein models in the Sea Urchin Genome Database (SpBase). This collection is based on the most recent assembly of the *S. purpuratus* genome (v3.1). It is supported by several EST collections and by a recent genome-wide transcriptome analysis that included 10 embryonic stages, six feeding larval and metamorphosed juvenile stages, and six adult tissues (Tu et al. 2012). The BLAST-P analysis identified all seven members of the Msp130 family that had been characterized previously (SpMsp130 and SpMsp130rel1-6) and one additional member, which I designated SpMsp130rel7 (SPU_021242, annotated as "Sp-Ap" in SpBase). The new member of the Msp130 protein family is predicted to contain 586 amino acids, including an N-terminal signal peptide (SignalP 4.1). The gene model for SpMsp130rel6 (SPU_014492) was previously suggested to be incomplete because the predicted protein lacked an N-terminal signal peptide (Livingston et al. 2006). The transcriptome analysis of Tu et al. (2012) identified several transcripts that overlap with the 5′ end of this gene model, including one (WHL22.405717.0) that encodes a putative signal peptide. The organization of this gene model, however, remains unclear. There appears to be a duplicate of the *msp130rel6* gene on a separate scaffold in the v3.1 assembly (SPU_015326, annotated as "Sp-Hypp_781"). Because the predicted amino acid sequence of Sp-Hypp_781 contained a full-length ORF, including a predicted signal peptide, the amino acid sequence of this protein was used as SpMsp130rel6 in our molecular phylogenetic analysis.

Each of the *msp130* family genes contains 12–13 exons, more than the genome-wide average (~8 exons/gene) in *S. purpuratus* (Tu et al. 2012). Five of the genes are found in two clusters located on separate genomic scaffolds, an organization which suggests that the family expanded relatively recently via gene duplication (Fig. 1). Each of these two scaffolds is ~700 kb in size and each tandem cluster of *msp130* genes is flanked by several unrelated genes. The remaining *msp130* genes, *msp130rel2*, *msp130rel5*, and *msp130rel7*, are located on scaffolds that are 330, 628, and 59 kb in size, respectively, and are flanked on each side by at least one unrelated gene. Any

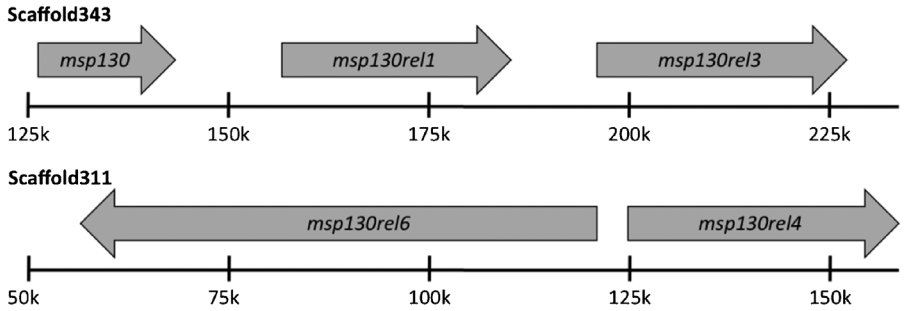

Fig. 1. Tandem clustering of Msp130 family genes in the sea urchin, *Strongylocentrotus purpuratus*. Scaffolds and coordinates are based on v. 3.1 of the *S. purpuratus* genome assembly (SpBase). For simplicity, the intron/exon organization of the genes is not shown.

higher order clustering of the *msp130* genes is currently obscured by the incomplete nature of the *S. purpuratus* assembly. Codon usage in the combined set of eight *msp130* family mRNAs closely mirrors that of other mRNAs in *S. purpuratus* (Kober and Pogson 2013) (Table S3).

Orthologs of Msp130 were previously identified in two other species of sea urchins, *Heliocidaris erythrogramma* and *H. tuberculata* (Klueg et al. 1997). These species and *S. purpuratus* are camarodont urchins, a clade that expanded within the past 150 Ma (Smith et al. 2006). A deeper phylogenetic signal was sought by analyzing a partial transcriptome assembly that has recently been generated from gastrula stage embryos of a cidaroid sea urchin, *Eucidaris tribuloides* (SpBase). This analysis identified transcripts that encoded clear orthologs of Msp130 and Msp130rel1-5 (Fig. 2). All modern sea urchins are thought to have descended from an ancestral stock of cidaroid-like sea urchins that survived the Permian extinction, and the evolutionary separation between *E. tribuloides* and the camardont species is estimated to be ~250 Ma (Smith et al. 2006). Therefore, a suite of at least six *msp130* family genes was present in the last common ancestor (LCA) of modern sea urchins.

Msp130 proteins in other metazoans

S. purpuratus Msp130 protein sequences were used to query the current collection of metazoan proteins, using the NCBI, Emsemble, and JGI browsers to search the complete nr protein databases and many taxon-specific protein sets (see Methods Section). In all cases, proteins that were identified as Msp130 family members yielded highly significant BLAST-P scores (*E*-values $< e^{-30}$) and back-BLASTs of these proteins against *S. purpuratus* proteins yielded only members of the Msp130 family as close matches (i.e., these matches yielded *E*-values $< e^{-30}$, while no other proteins yielded *E*-values $< e^{-5}$).

Msp130 proteins were identified in two species of invertebrate deuterostomes other than echinoderms: *Saccoglossus kowalevskii* (a hemichordate) and *Branchiostomia floridae* (a cephalochordate). Orthologs of these proteins were not identified, however, in vertebrates or urochordates. Cameron and Bishop (2012) recently described calcium carbonate (aragonite) biominerals in hemichordates and provided evidence that Msp130 family members are expressed during embryogenesis in *S. kowalevskii*. Phylogenetic analysis of Msp130 family proteins in the three invertebrate deuterostomes (*S. purpuratus*, *S. kowalevskii*, and *B. floridae*) showed that these proteins were more similar to other members of the Msp130 family within the same species than to Msp130-related proteins in the other two taxa (Fig. 3). This suggests that the *msp130* gene family expanded independently in the sea urchin, hemichordate, and cephalochordate lineages sometime after the divergence of these groups, which occurred 650–570 Ma (Swalla and Smith 2008; Erwin et al. 2011).

Among protostomes, Msp130 proteins were identified in two molluscs, the Pacific oyster (*Crassostrea gigas*) and the owl limpet (*Lottia gigantea*), both of which have well-assembled genomes (Zhang et al. 2012; Simakov et al. 2013) (Fig. 3). Msp130 proteins were not found, however, in several other protostomes with high-quality genome assemblies, including nematodes, Drosophila, honeybee (*Apis mellifera*) and three other recently analyzed protostomes: *Daphnia pulex* (a crustacean) and the annelids *Helobdella robusta* and *Capitella teleta* (Simakov et al. 2013). The identification of Msp130 proteins in molluscs, which produce a $CaCO_3$-based shell (Weiss et al. 2002), hints at a conserved function of these proteins in mineralization. Support for this hypothesis comes from a recent analysis of mRNAs that are enriched in the shell-forming mantle tissue of *C. gigas*, many of which likely encode proteins that mediate biomineralization (Table S2 in Zhang et al. 2012). This collection of mantle-enriched mRNAs includes a member of the Msp130 family (NCBI Accession No. EKC42376.1).

An *msp130* gene was introduced into the metazoan genome via HGT

Unexpectedly, Msp130 family proteins were also identified in representatives of many major bacterial clades, including α, β, δ, and γ-proteobacteria, acidobacteria, cyanobacteria, planctomycetes, actinobacteria, and archaebacteria (Fig. 3,

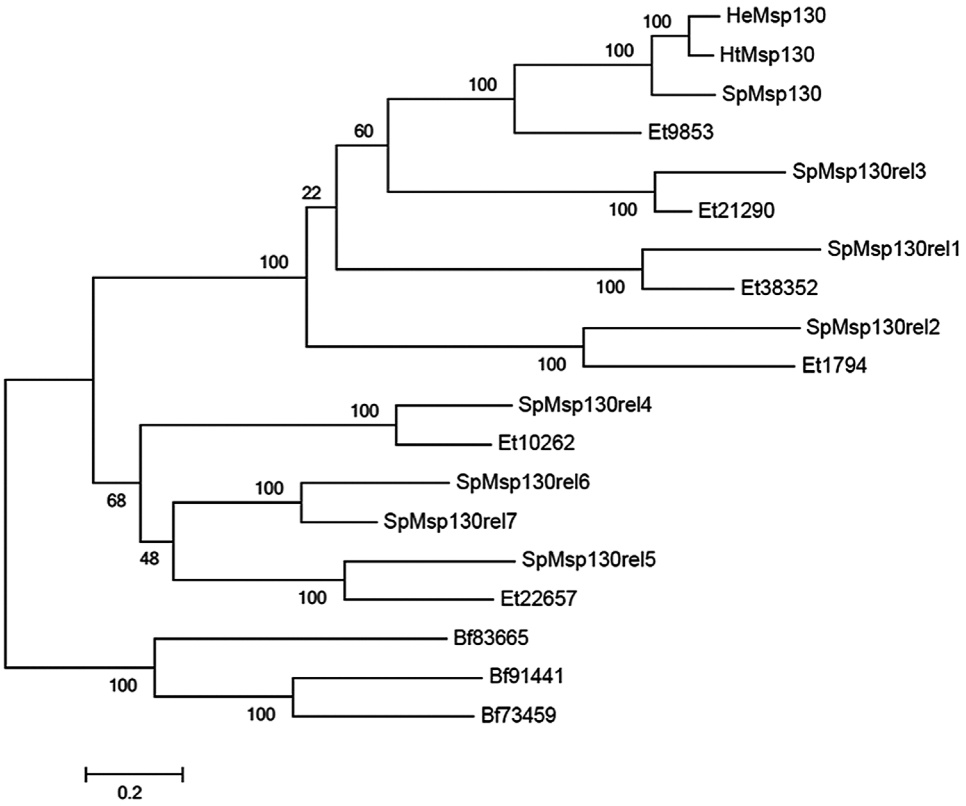

Fig. 2. Maximum likelihood tree of MSP130 family proteins from sea urchins (Echinodermata: Echinoidea). Guidance (http://guidance.tau.ac.il) (Penn et al. 2010) was used to generate a MAFFT-based multiple sequence alignment and amino acids that could not be aligned with confidence (columns with Guidance scores < 0.25) were removed. MEGA5 (v. 5.2.1) (Tamura et al. 2011; Hall 2013). was used to determine the optimal substitution model and to construct the tree, without any further deletion of gaps and with a bootstrap value of 500. The tree was rooted on Msp130 protein sequences from *Branchiostomia floridae*, a cephalochordate. Msp130 proteins from a cidaroid sea urchin, *Eucidaris tribuloides* can be unambiguously identified as orthologs of specific MSP130 family members from *Strongylocentrotus purpuratus*, a euechinoid, indicating that the expansion of Msp130 family genes in sea urchins predated the cidaroid-euechinoid divergence, dated at ~250 Ma (Smith et al. 2006). Bf, *Branchiostomia floridae*; Et, *Eucidaris tribuloides*; He, *Heliocidaris erythrogramma*; Ht, *Heliocidaris tuberculata*; Sp, *Strongylocentrotus purpuratus*.

Tables S1 and S2). For each of these bacterial species, a query with the SpMsp130 protein sequence identified a single, unambiguous ortholog with a BLAST-P score of e^{-25}–e^{-38}, while no other bacterial proteins yielded an *E*-value < 1.0. The bacterial Msp130 proteins were usually annotated as "alkaline phosphatase," but this appears problematical, as the bacterial protein sequences (like the Msp130 orthologs in eukaryotes) contained no alkaline phosphatase domains or any other identifiable protein domains in the Conserved Domains (NCBI), Pfam (Sanger Institute), or Smart databases (EMBL), nor did they show significant similarity to *bona fide* alkaline phosphatases from bacteria or eukaryotes. Single members of the Msp130 family were also identified in two species of green algae, *Chlamydomonas reinhardtii* and *Volvox carteri*, and in a brown alga, *Ectocarpus siliculosis*, but not in higher plants (Fig. 3, Table S1). Clustal alignment of the bacterial proteins with their metazoan homologs revealed clusters of conserved residues throughout the protein sequences, but with distinctly greater conservation in the central region, interrupted in the case of the *S. purpuratus* proteins by highly repetitive, glycine/proline/glutamine-rich sequences that are found in some, but not all, members of the Msp130 protein family (Fig. 4).

DISCUSSION

Functions of Msp130 proteins

The biochemical functions of the Msp130 proteins are unknown. These proteins lack any recognizable domains in the Conserved Domains (NCBI), Pfam (Sanger Institute), or Smart databases (EMBL). The proteins have an N-terminal signal peptide and are found on the cell surface, probably via GPI-linkages (Parr et al. 1990). In sea urchins, early work showed that a monoclonal antibody (mAb 1223) directed against Msp130 blocked calcium uptake and skeletogenesis by cultured PMCs (Carson et al. 1985; Kabakoff et al. 1992). This antibody was shown to recognize an

N-linked oligosaccharide chain on Msp130 that binds divalent cations, including calcium (Farach-Carson et al. 1989). Later work, however, showed that the oligosaccharide moiety recognized by mAb 1223 was present not only on Msp130 (130 kD), but also on two larger molecules of 205 and 250 kD, the identity of remains unknown. The developmental functions of Msp130 proteins have not yet been explored by means of morpholino-based knockdowns in sea urchins, a strategy which is complicated by the fact that

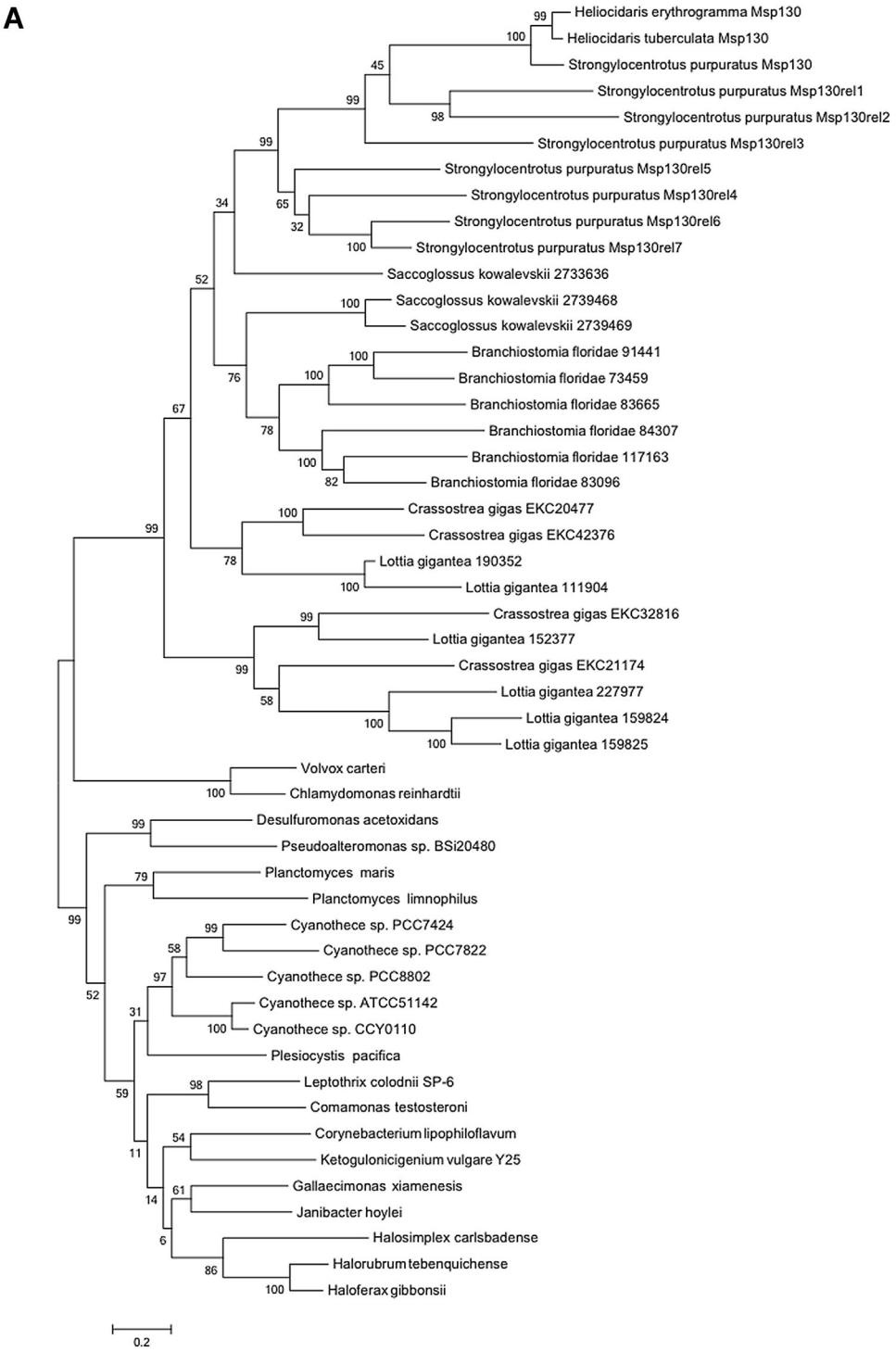

Fig. 3. Continued.

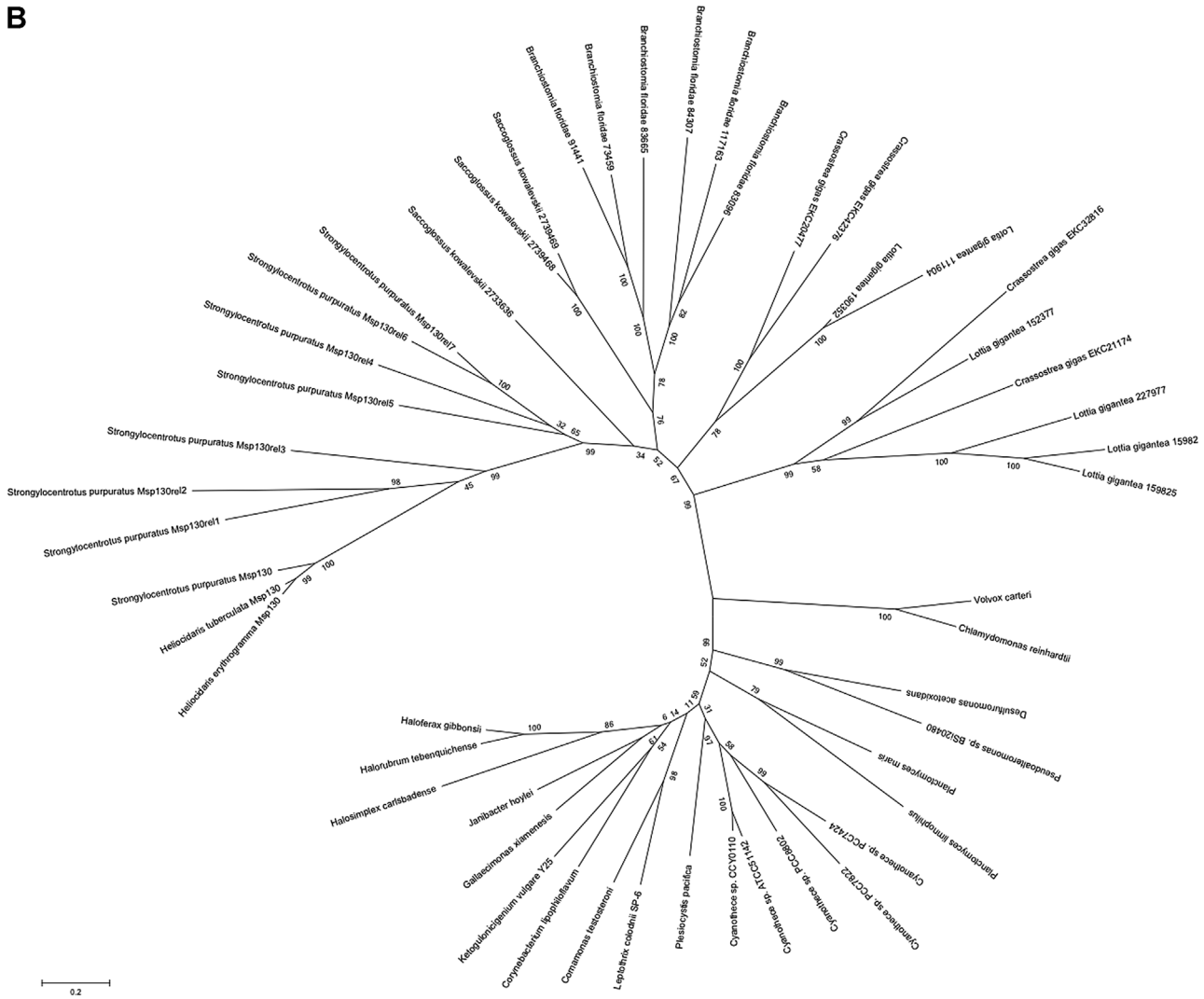

Fig. 3. Maximum likelihood trees of metazoan Msp130 family proteins and several representatives from bacteria. Tree construction methods are described in Materials and Methods and in the legend to Fig. 2. (A) Tree rooted on bacteria. (B) Unrooted tree. Although there are a number of weakly supported nodes, most of which lie within the partially reconstructed bacterial phylogeny shown here, this analysis strongly supports the view that *msp130* family genes expanded independently within the echinoderms (*Strongylocentrotus* + *Heliocidaris*), hemichordates (*Saccoglossus*), cephalochordates (*Branchiostomia*), and molluscs (*Crassotrea* + *Lottia*). *Volvox* and *Chlamydomonas* are algae; the remaining taxa are bacteria.

several Msp130 family members, possibly with redundant functions, are expressed during embryogenesis.

Although functional data are lacking, the expression of *msp130* genes in echinoderms (Anstrom et al. 1987; Leaf et al. 1987; Illies et al. 2002; Livingston et al. 2006) and molluscs (see Results Section) selectively in cells that deposit calcium carbonate-based biomineral points strongly to a role for these proteins in biomineralization. It would be of considerable interest to analyze the expression of *msp130* genes in amphioxus, which lacks a biomineralized skeleton but forms cartilage (Meulemans and Bronner-Fraser 2007). We also note that it has recently been shown that a basal clade of cyanobacteria produces intracellular inclusions of amorphous calcium carbonates, suggesting that intracellular biomineralization machinery may have been present in ancient cyanobacteria (Couradeau et al. 2012). We identified Msp family members in several species of cyanobacteria and it is possible that the biochemical function(s) of Msp130 proteins in bacteria are also related in some way to biomineralization.

The evolution of *msp130* genes

There are increasing numbers of examples of the introduction of genes into eukaryotic genomes by HGT, usually from bacterial symbionts (Keeling 2009; Dunning Hotopp 2011; Jackson et al. 2011; Azad and Lawrence 2012). The findings reported here strongly suggest that *msp130* genes were introduced into

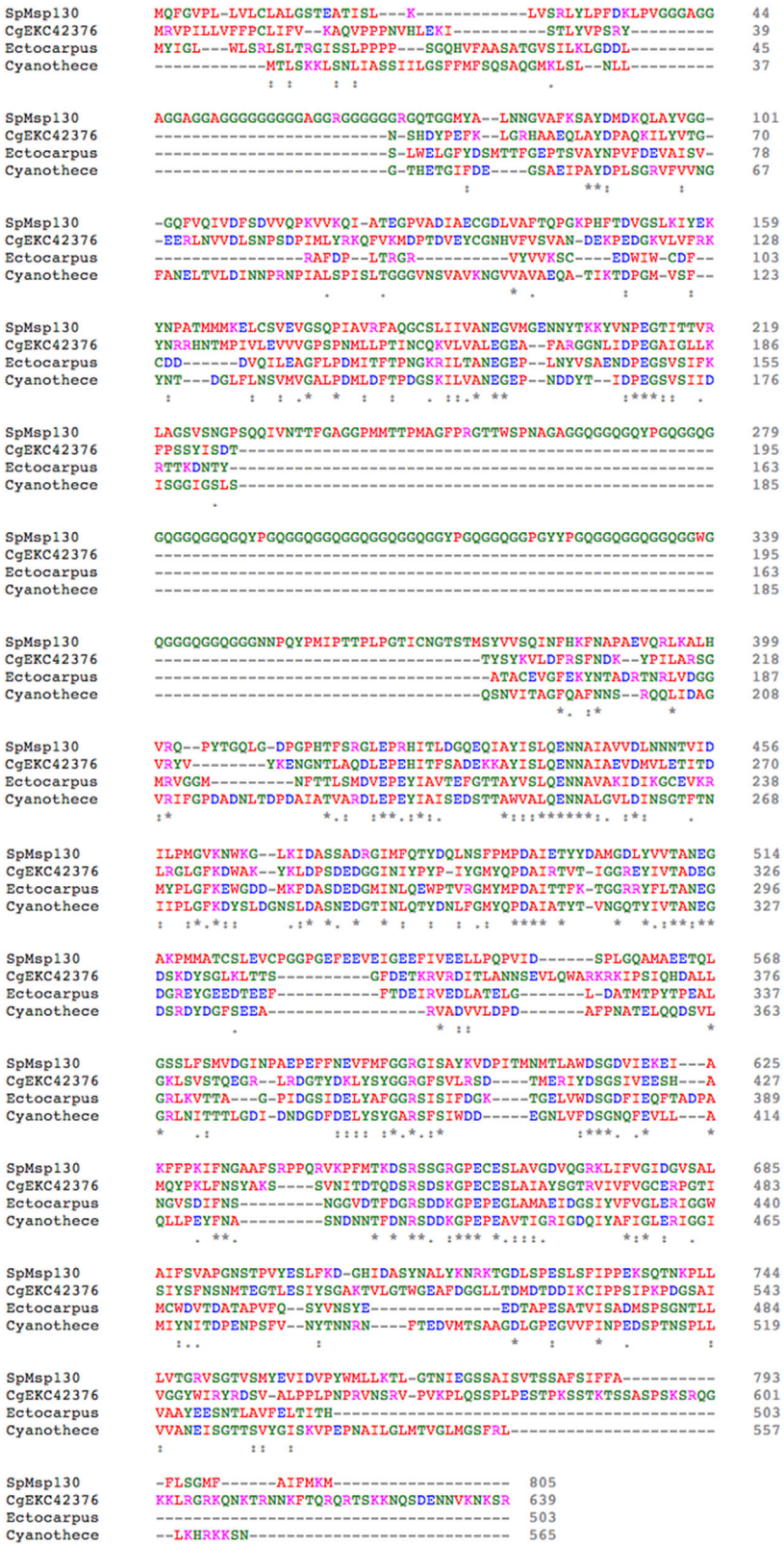

Fig. 4. Clustal Omega (Sievers et al. 2011) alignment of Msp130 proteins from the sea urchin *S. purpuratus* (SpMsp130, NCBI Acc. No. NP_001116986.1), the mollusc *Crassostrea gigas* (CgEKC42376, NCBI Acc. No. EKC42376.1), the brown alga *Ectocarpus siliculosis* (NCBI Acc. No. CBJ25800.1) and the cyanobacterium *Cyanothece sp. ATCC 51142* (YP_001804430.1). The alignment was carried out using the EMBL-EBI webserver (http://www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters. Asterisks indicate identical amino acids, two dots indicate highly similar amino acids, and single dots represent moderately similar amino acids. There are clusters of conserved residues throughout the protein sequences, but distinctly greater conservation is apparent in the central region of the proteins, with the exception of highly repetitive, glycine/proline/glutamine-rich sequences that are found in some, but not all, members of the Msp130 protein family in sea urchins. Pairwise BLAST-P scores support the striking similarity between metazoan and bacterial msp130 family proteins; for example, BLAST-P alignment of the SpMsp130 and the Cyanothece proteins yields a total score of 199 and an *E* value of $5e^{-42}$.

metazoans via HGT, either directly from bacteria or indirectly via an algal intermediary. Algae are a principal food source for larval and adult sea urchins and the intestinal tracts of adult sea urchins are populated by a diverse collection of bacterial species, many of which aid in digestion (Lawrence et al. 2013). Perhaps this close relationship facilitated gene transfer in some way, as is postulated in the case of sponges and their bacterial symbionts (Jackson et al. 2011). The possibility that gene transfer occurred in the opposite direction (i.e., from metazoans to bacteria) seems implausible given the extremely wide phylogenetic distribution of *msp130* genes among different bacterial clades, which diverged prior to the appearance of metazoans. Such a distribution would only be possible if the gene had been transferred horizontally into each of the bacterial lineages independently, which is highly improbable.

The principal evidence in support of HGT is an incongruence of the taxonomic distribution of homologous genes with the expected pattern of vertical inheritance based on accepted phylogenetic relationships (Azad and Lawrence 2012). The presence of unambiguous *msp130* genes in several invertebrate deuterostomes, but only a single protostome phylum (molluscs), suggests that *msp130* was introduced into the genomes of molluscs and deuterostomes separately, via independent HGT events. It is noteworthy in this regard that *msp130* genes are absent from several non-bilaterian metazoans, including two cnidarian species (*Hydra magnipapillata* and *Nematostella vectensis*), a sponge (*Amphimedon queenslandica*), a placozoan (*Tricoplax adhaerens*), and a ctenophore (*Mnemiopsis leidyi*), an observation which argues against the possibility that an *msp130* gene was present in the LCA of all metazoans. Despite these various considerations, it remains possible that a single HGT event occurred early in metazoan evolution (e.g., in the LCA of bilaterians) and was followed by gene losses in many lineages. Such scenarios are more probable if gain of *msp130* by HGT was a very unlikely event, while losses occurred readily. At present, an important limitation in distinguishing among these evolutionary scenarios is that the number of fully sequenced metazoan genomes outside the deuterostomes is quite small, and our picture of the distribution of *msp130* genes will likely change as more genomes become available. In particular, it will be important to determine whether *msp130* genes are present in those protostomes and non-bilaterian metazoans that form calcium carbonate-based biominerals; groups that include the calcifying sponges, scleractinian corals, and various protostome clades, including calcifying annelids, arthropods, and bryozoans (Knoll 2003).

The identification of unambiguous, orthologous pairs of Msp130 family members in cidaroid and euechinoid sea urchins indicates that an ancestral *msp130* gene must have been present (and had already undergone multiple duplications) in the LCA ancestor of sea urchins. This establishes a minimum date of the HGT event in this lineage at 250 Ma. An ancient HGT event in the sea urchin lineage in consistent with the observation that the pattern of synonymous codon usage in *S. purpuratus msp130* mRNAs closely resembles that of other *S. purpuratus* mRNAs (Table S3).

With these various considerations in mind, one plausible and parsimonious model is illustrated in Fig. 5. This model postulates two independent HGT events, one that introduced an ancestral *msp130*-like gene into molluscs and another that introduced the gene into an early deuterostome, followed by a loss in the (vertebrate + urochordate) lineage. A variation of this scenario is that two independent HGT events occurred in deuterostomes, one that introduced the *msp130* gene into the LCA of (echinoderms + hemichordates) and another that introduced the gene into the cephalochordate lineage. An even simpler model, which requires only a single introduction of the gene in deuterostomes and no losses, arises if cephalochordates are grouped with echinoderms rather than chordates. This grouping has been suggested by some (Delsuc et al. 2006) but most recent studies have supported the vertebrate–urochordate–cephalochordate grouping, with the cephalochordates a sister group to the chordates (Bourlat et al. 2006; Delsuc et al.

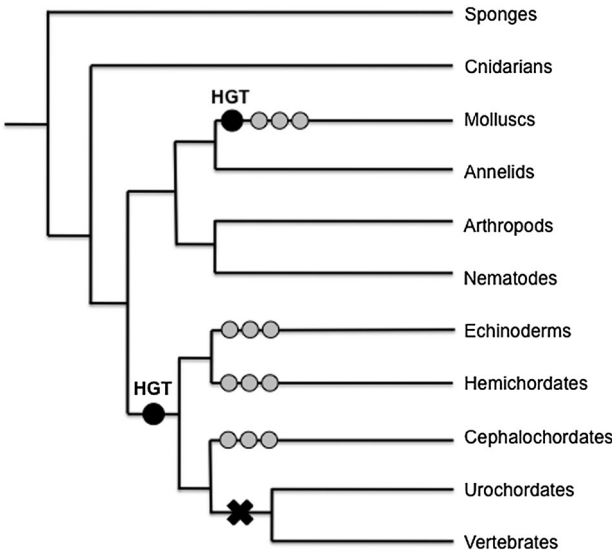

Fig. 5. A provisional model of the evolution of *Msp130* genes in metazoans. A partial metazoan phylogeny is shown, after Erwin et al. (2011). Black circles represent introduction of an ancestral bacterial *Msp130* gene by HGT. This might have occurred directly from bacteria, or indirectly via an algal intermediary. Small gray circles represent an indeterminate number of gene duplication events. The cross represents gene loss. Although the model shown here requires relatively few gene transfers + losses, other scenarios are possible. For example, there may have been a single HGT event in an early bilaterian ancestor, followed by gene loss in many protostome clades and in the urochordate/vertebrate lineage. Note that the number of fully sequenced protostome genomes is small, which limits the resolution of the analysis in that group. Expansion of the *Msp130* gene family occurred independently in the four animal lineages. In echinoderms, this occurred before the cidaroid-euechinoid split, ~250 Ma (Smith et al. 2006).

2006, 2008; Swalla and Smith 2008). Although the precise number and timing of HGT events cannot be determined with certainty, it seems likely that Msp130 proteins proved to be useful in mediating biomineralization in several metazoan lineages and that multiple, independent duplications of these genes occurred.

Acknowledgments

I thank Dr. Dannie Durand for many valuable discussions concerning tree construction and HGT. This work was supported by NSF Grant IOS-1021805.

REFERENCES

Anstrom, J. A., Chin, J. E., Leaf, D. S., Parks, A. L., and Raff, R. A. 1987. Localization and expression of Msp130, a primary mesenchyme lineage-specific cell surface protein in the sea urchin embryo. *Development* 101: 255–265.

Azad, R. K., and Lawrence, J. G. 2012. Detecting laterally transferred genes. *Methods Mol. Biol.* 855: 281–308.

Bourlat, S. J., et al. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444: 85–88.

Cameron, C. B., and Bishop, C. D. 2012. Biomineral ultrastructure, elemental constitution and genomic analysis of biomineralization-related proteins in hemichordates. *Proc. Biol. Sci.* 279: 3041–3048.

Carson, D. D., Farach, M. C., Earles, D. S., Decker, G. L., and Lennarz, W. J. 1985. A monoclonal antibody inhibits calcium accumulation and skeleton formation in cultured embryonic cells of the sea urchin. *Cell* 41: 639–648.

Couradeau, E., et al. 2012. An early-branching microbialite cyanobacterium forms intracellular carbonates. *Science* 336: 459–462.

Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439: 965–968.

Delsuc, F., Tsagkogeorga, G., Lartillot, N., and Philippe, H. 2008. Additional molecular support for the new chordate phylogeny. *Genesis* 46: 592–604.

Dunning Hotopp, J. C. 2011. Horizontal gene transfer between bacteria and animals. *Trends. Genet.* 27: 157–163.

Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D., and Peterson, K. J. 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334: 1091–1097.

Ettensohn, C. A. 2013. Encoding anatomy: developmental gene regulatory networks and morphogenesis. *Genesis* 51: 383–409.

Farach-Carson, M. C., Carson, D. D., Collier, J. L., Lennarz, W. J., Park, H. R., and Wright, G. C. 1989. A calcium-binding, asparagine-linked oligosaccharide is involved in skeleton formation in the sea urchin embryo. *J. Cell Biol.* 109: 1289–1299.

Hall, B. G. 2013. Building phylogenetic trees from molecular data with MEGA. *Mol. Biol Evol.* 30: 1229–1235.

Harkey, M. A., Whiteley, H. R., and Whiteley, A. H. 1992. Differential expression of the Msp130 gene among skeletal lineage cells in the sea urchin embryo: a three dimensional in situ hybridization analysis. *Mech. Dev.* 37: 173–184.

Hentschel, U., Piel, J., Degnan, S. M., and Taylor, M. W. 2012. Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* 10: 641–654.

Illies, M. R., Peeler, M. T., Dechtiaruk, A. M., and Ettensohn, C. A. 2002. Identification and developmental expression of new biomineralization proteins in the sea urchin *Strongylocentrotus purpuratus*. *Dev. Genes Evol.* 212: 419–431.

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

Jackson, D. J., Macis, L., Reitner, J., and Wörheide, G. 2011. A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. *BMC Evol. Biol.* 11: 238–243.

Jackson, D. J., Thiel, V., and Wörheide, G. 2010. An evolutionary fast-track to biocalcification. *Geobiology* 8: 191–196.

Kabakoff, B., Hwang, S. P., and Lennarz, W. J. 1992. Characterization of post-translational modifications common to three primary mesenchyme cell-specific glycoproteins involved in sea urchin embryonic skeleton formation. *Dev. Biol.* 150: 294–305.

Keeling, P. J. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr. Opin. Genet. Dev.* 19: 613–619.

Killian, C. E., and Wilt, F. H. 2008. Molecular aspects of biomineralization of the echinoderm endoskeleton. *Chem. Rev.* 108: 4463–4474.

Knoll, A. H. 2003. Biomineralization and evolutionary history. *Rev. Min. Geochem.* 54: 329–356.

Klueg, K. M., Harkey, M. A., and Raff, R. A. 1997. Mechanisms of evolutionary changes in timing, spatial expression, and mRNA processing in the Msp130 gene in a direct-developing sea urchin, Heliocidaris erythrogramma. *Dev. Biol.* 182: 121–133.

Kober, K. M., and Pogson, G. H. 2013. Genome-wide patterns of codon bias are shaped by natural selection in the purple sea urchin, *Strongylocentrotus purpuratus*. *G3 (Bethesda)* 3: 1069–1083.

Lawrence, J. M., Lawrence, A. L., and Watts, S. A. 2013. Feeding, digestion, and digestibility of sea urchins. In J. M. Lawrence (ed.). *Sea Urchins, Biology and Ecology.* Chapter 9. Elsevier BV, pp. 135–154.

Leaf, D. S., Anstrom, J. A., Chin, J. E., Harkey, M. A., Showman, R. M., and Raff, R. A. 1987. Antibodies to a fusion protein identify a cDNA clone encoding Msp130, a primary mesenchyme-specific cell surface protein of the sea urchin embryo. *Dev. Biol.* 121: 29–40.

Livingston, B. T., et al. 2006. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. *Dev. Biol.* 300: 335–348.

Mann, K., Poustka, A. J., and Mann, M. 2008. In-depth, high-accuracy proteomics of sea urchin tooth organic matrix. *Proteome Sci.* 6: 33.

Mann, K., Poustka, A. J., and Mann, M. 2008. The sea urchin (*Strongylocentrotus purpuratus*) test and spine proteomes. *Proteome Sci.* 6: 22.

Mann, K., Wilt, F. H., and Poustka, A. J. 2010. Proteomic analysis of sea urchin (*Strongylocentrotus purpuratus*) spicule matrix. *Proteome Sci.* 8: 33.

Meulemans, D., and Bronner-Fraser, M. 2007. Insights from amphioxus into the evolution of vertebrate cartilage. *PLoS ONE* 2: e787.

Murdock, D. J. E., and Donohue, P. 2011. Evolutionary origins of animal skeletal biomineralization. *Cell Tiss. Org.* 94: 98–102.

Oliveri, P., Tu, Q., and Davidson, E. H. 2008. Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl. Acad. Sci. U.S.A.* 105: 5955–5962.

Parr, B. A., Parks, A. L., and Raff, R. A. 1990. Promoter structure and protein sequence of Msp130, a lipid-anchored sea urchin glycoprotein. *J. Biol. Chem.* 265: 1408–1413.

Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., and Pupko, T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 38: W23–W28.

Rafiq, K., Cheers, M. S., and Ettensohn, C. A. 2012. The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development* 139: 579–590.

Rafiq, K., Shashikant, T., McManus, C. J., and Ettensohn, C. A. 2014. Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins. *Development* 141: 950–961.

Sievers, F., et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7: 539.

Simakov, O., et al. 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature* 493: 526–531.

Smith, A. B., Pisani, D., Mackenzie-Dodds, J. A., Stockley, B., Webster, B. L., and Littlewood, D. T. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol. Biol. Evol.* 23: 1832–1851.

Swalla, B. J., and Smith, A. B. 2008. Deciphering deuterostome phylogeny: molecular, morphological, and palaeontological perspectives. *Phil. Trans. R. Soc. B* 363: 1557–1568.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.

Tu, Q., Cameron, R. A., Worley, K. C., Gibbs, R. A., and Davidson, E. H. 2012. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res.* 22: 2079–2087.

Weiss, I. M., Tuross, N., Addadi, L., and Weiner, S. 2002. Mollusc larval shell formation: amorphous calcium carbonate is a precursor phase for aragonite. *J. Exp. Zool.* 293: 478–491.

Westbroek, P., and Marin, F. 1998. A marriage of bone and nacre. *Nature* 392: 861–862.

Wilt, F. H., and Ettensohn, C. A. 2007. The morphogenesis and biomineralization of the sea urchin larval skeleton. In E. Bauerlein (ed.). *Handbook of Biomineralization*. Wiley-VCH Press, pp. 183–210.

Wörheide, G., and Jackson, D. J. 2011. Animal biocalcification, evolution. In J. Reitner and and V. Thiel (eds.). *Encyclopedia of Geobiology.* Springer, Berlin, pp. 53–58.

Zhang, G., et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490: 49–54.

Zhu, X., Mahairas, G., Illies, M., Cameron, R. A., Davidson, E. H., and Ettensohn, C. A. 2001. A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. *Development* 128: 2615–2627.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Table S1. Taxonomic terms used in clade-specific BLAST-P searches of the NCBI non-redundant protein database

Table S2. Protein sequences used for the construction of phylogenetic trees

Table S3. Codon usage bias in *S. purpuratus msp130* mRNAs