communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-07785-7

Proteomic analyses reveal the key role of gene co-option in the evolution of the scaly-foot snail scleritome

Wai Chuen Wong[1,2], Yick Hang Kwan[1,2,3], Xing He[4], Chong Chen[5], Shengling Xiang[6], Yao Xiao[1,2], Lexin Long[1,2], Kexin Gao[4], Ning Wang[6], Longjun Wu[2], Pei-Yuan Qian[1,2] ✉ & Jin Sun[4] ✉

Biomineralization, a key driving force underlying dramatic morphological diversity, is widely adopted by metazoans to incorporate inorganic minerals into their organic matrices. The scaly-foot snail *Chrysomallon squamiferum* from deep-sea hot vents uniquely possesses hundreds of sclerites on its foot in addition to a coiled shell, providing an exclusive case to study the formation of evolutionarily novel hard parts. Here, we identified the matrix proteins present in the exoskeletons of *C. squamiferum* and *Gigantopelta aegis*, a confamilial species from the same vent habitat but lacking sclerites, to uncover the genes and proteins presumably involved in the sclerite formation processes. Comparative multi-omics analyses suggest that *C. squamiferum* co-opted a diverse range of metazoan biocalcifying proteins through sclerite formation in a possibly deep homology scenario, and the up-regulated biomineralization-related genes in the foot imply alternative sources of sclerite proteins. The sclerite-secreting epithelium employs and utilizes genes considerably older than those in the mantle, which supports the predominant contribution of co-option in *C. squamiferum* sclerite formation. Our results highlight the importance of gene co-option in shaping novel hard parts in *C. squamiferum* and indicate that lineage-specific gene incorporation is a possible key factor leading to the rapid evolution of a novel hard structure in this vent-endemic species.

Organisms produce biominerals that serve multiple purposes, such as protection and structural support, usually through the deposition and solubilization of inorganic minerals on organic matrix frameworks[1,2]. The organic matrices assembled as mineral crystallization sites are the key to the overall "matrix-mediated" mineral crystallization[2–6]. Distinct biomineralization tissues simultaneously control the formation of diverse minerals[7], in which sophisticated molecular machineries guide the construction of organized crystalline microstructures. Biomineralization is an evolutionary novelty that originated from the emergence of major phyla during the Ediacaran–Cambrian transitions and have since been an integrative part of their diversification process[8,9]. Fossil records of the Cambrian explosion reveal the extensive acquisition of mineralized armature across animal groups[10], in which molluscs have evolved arguably the highest diversity of hard structures. Since the Early Cambrian, molluscs have developed a plethora of exoskeletal configurations[11], with the earliest fossils implying the subsequent derivation of calcareous contents from aragonite to calcite and calcite–aragonite composites[12–14].

The origin of calcified shells is linked to the predator–prey arms race and the need to detoxify excessive intracellular calcium[15,16]. Meanwhile, further calcified morphological novelties with various functions and distinctive morphological characteristics, such as statocysts, egg capsules, love darts, opercula, gizzard plates, and spicules, also emerged in various molluscan lineages, lending weight to the idea that independent armature development occurred through their evolution. Heightened concentrations of additional ions in the seawater[15] further facilitated the evolution of non-calcium biominerals[17]. The incorporation of inorganic materials, such as magnetite, ferrihydrite[11,12], iron phosphate[11], apatite[14], and lepidocrocite[13], contributes to the formation of the radula in some molluscs which exhibits notable mechanical strength. The establishment of a "biomineralization toolkit" pre-adapted from the common ancestor of the Lophotrochozoa

[1]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China. [2]Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong, China. [3]Department of Biology, HADAL & Nordcee, University of Southern Denmark, Odense, Denmark. [4]Key Laboratory of Evolution & Marine Biodiversity (Ministry of Education) and Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao, China. [5]X-STAR, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa, Japan. [6]Department of Physics, The Hong Kong University of Science and Technology, Hong Kong, China. ✉e-mail: boqianpy@ust.hk; jin_sun@ouc.edu.cn

gave rise to several universal skeletal features in the constituent phyla during evolution[9,18]. However, proteins participating in various molluscan biomineralizations processes exhibit distinct profiles that support their specific armature development[19].

Currently examined biominerals have shown a collaborative contribution composed of lineage- and species-specific proteins and proteins encoded by ancient genes preserved and co-opted among eukaryotes[20,21]. The deployment of distinct *hox* genes in morphologically similar molluscan shell fields suggests independent co-option at the shell-forming epithelium[22]. The extensive conservation of cellular arrangements in the embryonic stage further denotes a common and inherited molluscan biomineralization program from a distant period[23,24]. Conversely, studies have demonstrated the highly plastic nature of mantle secretomes and shell matrix proteomes in certain molluscs[25]; in addition, weak evolutionary constraints on mineral-interacting domains of acidic biomineralizing proteins further add to this plasticity[26]. Although novel and co-opted proteins account for varying proportional contributions across different molluscan groups, protein co-option has been suggested to play a dominant role in gastropods and bivalves[27].

The scaly-foot snail *Chrysomallon squamiferum* is a peltospirid gastropod endemic to deep-sea hydrothermal vents in Indian Ocean. Vent-endemic species have developed unique adaptations throughout evolution to help them physically and chemically thrive in their challenging deep-sea chemosynthetic habitats[28,29]. *Chrysomallon squamiferum* is the only living gastropod armored with dermal sclerites, which form on a β-chitin and protein matrix, while also retaining a coiled shell[30–32]. The *C. squamiferum* dermal scleritomes, although superficially similar to those of aculiferan molluscs such as chitons[31,33], are anatomically distinct and represent an evolutionary novelty unique to this species[33]; it likely evolved as a deposition site for sulfide waste from chemosymbiosis[34]. On the face of it, the scales of *C. squamiferum* closely resemble those of the Cambrian *Wiwaxia*[33] morphologically; transcriptomic analysis of genes possibly involved in its biomineralization indicated the co-option of an ancient biomineralization tool kit broadly utilized by lophotrochozoans indicative of deep homology[35]. The complete genome assembly of *C. squamiferum* also revealed an unexpectedly low gene family uniqueness (11%) compared to other lophotrochozoan genomes[35]. Thus, the formation of *C. squamiferum* sclerites relies on alterations in the expression of genes in a conserved biomineralization tool kit—a mechanism that also underlies its coiled shell and many other metazoan hard parts[35,36]. Iron sulfide is an unusual constituent infused in the *C. squamiferum* armature[34], resulting from the reaction of biologically secreted sulfur with iron ions from the surrounding environment[30,35,37]. The high expression of metal tolerance protein in individuals with iron sulfide-rich hard parts suggests that this adaptation is likely influenced by the specific environment with high iron concentration at certain hydrothermal vent sites[35]. Nevertheless, the actual components of protein matrices functioning at the sites of sclerite biomineralization and how *C. squamiferum* controls the formation of its hard parts remain obscure. Meanwhile, the confamilial snail *Gigantopelta aegis*, which is also endemic to hydrothermal vents on the mid-ocean ridges of the Indian Ocean, lacks dermal sclerites[38] and represents an ideal target for comparative studies aiming to uncover the critical factors contributing to sclerite formation in *C. squamiferum*.

Here, we characterized the proteomes of the *C. squamiferum* shell and scale as well as the *G. aegis* shell, using high-sensitivity liquid chromatography with tandem mass spectrometry (LC-MS/MS). We profiled and compared proteins discovered in the biomineralized hard parts against the published whole genomes of the two species[35,39] and identified their conserved functional domains. Furthermore, we conducted interfamilial and interspecific comparative analysis on the biomineral proteome to construct a preliminary sclerite- and shell-formation toolkit in order to elucidate the differential selection of proteins involved in the biomineralization of the shell versus sclerites in *C. squamiferum*. Aside from the proteomics study, we also incorporated tissue-specific transcriptomic data and overall transcriptome expression patterns in a phylostratigraphic analysis[35,39] to provide a more inclusive dataset to explore the biological origins of sclerites compared to the shells of *C. squamiferum* and *G. aegis*. We reveal a proteome unique to the sclerites of *C. squamiferum*, providing insights into the deep homology of ancient biomineralization mechanisms and their contribution towards the novel-yet-rapid armature evolution in *C. squamiferum*.

Results

Genome cross-referencing and comparative transcriptomics

The published genomes of *Chrysomallon squamiferum* and *Gigantopelta aegis*[35,39] were re-analyzed to focus on the updated genomic synteny and gene family analyses. The conservative chromosomal synteny between the two species was demonstrated using chromosome-scale genome data of *C. squamiferum* and *G. aegis* (Supplementary Fig. 1; Supplementary Fig. 2), and the findings confirmed their overall highly conserved genomic synteny. Furthermore, 541 gene families were considerably expanded in *C. squamiferum* from protein-coding sequences of 14 molluscs (Supplementary Data 1), of which nine were *G. aegis* orthologues. Among these nine orthologues, hemicentin-1 isoform, cartilage matrix protein, troponin, chorion peroxidase, tubulin, and immunoglobulin domain-containing protein gene families have been discovered in previously reported shell-matrix proteomes[40–44], and are thus expected to be contributors to sclerite biomineralization processes. Gene Ontology (GO) enrichment analysis of the expanded gene families also suggested the expansion of the scavenger receptor (SR) activity category (Supplementary Fig. 3 and 4) in the transcriptome of *C. squamiferum* sclerites.

An expression heatmap of the corresponding matrix protein-encoding transcripts in various tissues revealed the possible sources of the resultant shell/sclerite matrix proteins extracted from shell/sclerite specimens and identified through LC-MS/MS (Supplementary Fig. 4). The components of the two investigated shells revealed dominating gene expressions in the shell-secreting mantle (Supplementary Fig. 4a and 4b; Supplementary Data 2). However, the expressions of sclerite matrix proteins indicated their scattered contributions from multiple soft tissues (Supplementary Fig. 4c; Supplementary Data 3). The expanded number of shell proteins emerging from non-epithelium sites implies that most proteins used in the sclerite formation perform alternative functions outside of this tissue.

Matrix protein profiling

Proteomics data acquired from LC-MS/MS sequencing of proteins extracted from the hard parts of *C. squamiferum* (shell and sclerite) and *G. aegis* (shell) were employed in matrix protein profiling and the subsequent proteomics analysis. Scanning electron microscopy (SEM) demonstrated the resemblance between the layered calcareous shells of *C. squamiferum* and *G. aegis* with minor discrepancies (Supplementary Fig. 5). The *G. aegis* shell features an outer granulose layer and a thicker, inner crossed-lamellar layer[45]; that of *C. squamiferum* is similar but has an additional thin prismatic layer on the innermost surface. In both species, numerous shell pores open from the inner surface, and they penetrate the crossed-lamellar layer. Cross-sections of *C. squamiferum* sclerites are mostly homogenous and without discernible layering (Supplementary Fig. 6). Previous studies on *C. squamiferum* sclerites from the iron-rich Kairei vent field revealed a layering corresponding to various levels of iron sulfide crystals[34,46]; however, given that our specimens from the Longqi vent field lacked crystallized iron sulfide, such a distinct layering was not observed. A total of 168 (*C. squamiferum* shell), 178 (*C. squamiferum* sclerite), and 156 (*G. aegis* shell) proteins containing multiple unique peptides with a false discovery rate (FDR) < 0.05 were identified. Supplementary Data 4–6 lists their resultant protein accessions and their detailed descriptions and corresponding SEQUEST. Acid-insoluble fractions of the matrices were extracted, but no protein was identified from the search.

On the basis of sclerite protein profiles, we predicted the signal peptide of a reduced fraction of proteins specific to sclerites (29 out of 178). A total of 86 *C. squamiferum* shell proteins carried secretion signals, whereas 49 shell proteins predicted in *G. aegis* had signal peptides. Transmembrane helices were predicted in *C. squamiferum* shell proteomes, but the proteins in *G.*

aegis acquired the transmembrane feature and signal peptide. More matrix proteins with tandem repeats were observed in both shell proteomes compared with that in the sclerite proteome (Supplementary Figs. 7–9). A high percentage of shell proteins with internal disorder (ID) regions remained uncharacterized, which implies an elevated modification rate. Notably, 62 out of 63 sclerite proteins with ID regions detected were annotated upon intrinsic disorder search.

Proteins with identified conserved domains were classified under six categories possessing domains with putative functions in biomineralization (Supplementary Data 7–9). In addition to 80 *C. squamiferum* shell proteins, 20 sclerite proteins and 60 *G. aegis* shell proteins had insufficiently characterized functional domains (Supplementary Data 10–12). Foot protein 14, along with one neurogenic locus notch homolog protein 3 isoform X1, one predicted protein, and one uncharacterized protein were shell proteins found across both *C. squamiferum* biomineralized parts lacking an identified domain. The domain variety in the *C. squamiferum* sclerite proteome surpassed that of the *C. squamiferum* shell proteome in all categories and was more similar to that of the *G. aegis* shell proteome (Supplementary Data 13). Notably, immunity-related sclerite matrix proteins acquired a domain composition similar to that identified in both shells. In addition, proteins with immunoglobulin and fibronectin type III domains were specifically detected in *C. squamiferum* sclerites. The *C. squamiferum* sclerite proteome exclusively contained proteins carrying cadherin, collagen triple helix repeats, and laminin family domains, whereas both shell proteomes incorporated a different set of extracellular matrix proteins. Likewise, shell proteins categorized under cation interactions and enzymes exhibited a fairly poor similarity in domain composition among the three hard parts studied. Meanwhile, the shell and sclerite of *C. squamiferum* obtained IQ calmodulin-binding motifs, and the shells of *C. squamiferum* and *G. aegis* acquired C-type lectin domains. As EF-hand and epidermal growth factor (EGF) domains were the prevailing domains in all target biominerals, cation-interacting conserved domain spectrin/alpha-actinin and filamin/ABP280 were detected in the *C. squamiferum* sclerite matrix. Enzymatic proteins, despite the lack of a significant difference in the quantity between the *C. squamiferum* sclerite and *G. aegis* shell, exhibited a highly diversified composition of conserved domains across all three proteomes. Aside from the shared proteins, tyrosinase copper-binding domain, chitin-binding domain, and EF-hand domain were universal to these three hard-part proteomes.

Comparative proteomics

The matrix proteins of each hard part identified from LS-MS/MS served as independent inputs representing *C. squamiferum* shell, *C. squamiferum* sclerite, and *G. aegis* shell corresponding throughout interspecific comparisons. Figure 1 illustrates the results of interspecific comparisons of the shell proteins of the obtained matrix protein sequences and biomineral proteins collected from nine lophotrochozoan species (Supplementary Data 14) in the shell protein database. The results were categorized into four quartiles based on the percentage identity between matching sequences, and several examined shell proteins exhibited an efficient homology to the shell protein database (Supplementary Data 15–17). In *C. squamiferum* sclerites, 74 of 178 (41.6%) shell proteins achieved significant resemblance hits in the shell protein database. The sclerite matrix proteins had higher overall percentage similarity and sequence homology than the shell and biomineralization-related proteomes of other molluscan species. The *C. squamiferum* sclerite proteins with a sequence homology exceeding 75% revealed close affinities solely to the pulmonated land snail *Euhadra quaesita*, whereas the scleritome attained a fair homology with the mussel *Mytilus coruscus* (65.0%) and the brachiopod *Lingula anatina* (51.4%). The studied shell proteomes exhibited a lower percentage of similarity and degree of matching against other biomineralization-related proteomes of other molluscan species. The *G. aegis* shell had 53/156 (34.0%), and *C. squamiferum* shell had 46/168 (27.4%) shell proteins mapped to the database. The distantly related *Euhadra quaesita* had an actin sequence mapped to the target peltospirid biomineral proteomes with an unexpectedly high similarity above 90%. This finding indicates a high conservation of cytoplasmic proteins in the evolution of gastropods.

Given the possibility of cytoplasmic proteins inadvertently incorporating into biominerals, a BLAST search was further conducted to exclude proteins without signal peptides or previously documented biomineralization functions from the target shellomes/scleritome. The results reveal that 7.30% (13/178), 11.9% (20/168), and 9.62% (15/156) of the proteins from the scleritome of *C. squamiferum*, shellome of *C. squamiferum*, and shellome of *G. aegis*, respectively, exhibited sequence similarities with known shell matrix proteins (Supplementary Fig. 10; Supplementary Data 18–20). The observed reduction in sequence similarity compared with those of a previous search underscores the potential role of cellular proteins in the biomineralization of *C. squamiferum* and *G. aegis*. Notably, the increased matching rates of their shellomes compared with those of other shellomes suggest a preferential conservation of calcium carbonate-mineralizing proteins in gastropods. Furthermore, the predominance of protein similarity scores within the third quartile indicates a significant degree of plasticity among biomineralization-related proteins.

The lineage and species specificity of the unannotated shell/sclerite proteins obtained from LS-MS/MS were examined via interfamily analysis (Fig. 2a; Supplementary Data 21). A total of 12 out of 49 uncharacterized shell/sclerite proteins were specifically detected in *C. squamiferum*, whereas 19 were family-specific given that they were also present in *G. aegis*. Meanwhile, 7 out of the 178 profiled sclerite proteins were uncharacterized, and only one was confined to a lineage-specific proteome. Two of the seven uncharacterized sclerite proteins were shared within the examined gastropod family Peltospiridae, and the remaining ones had homologs across Mollusca. Orthologue analyses involved the genomes of *C. squamiferum* [35] and *G. aegis*[39]. More orthologous genes directly related to the exoskeletal proteomes were observed between *C. squamiferum* and *G. aegis* shells than between *C. squamiferum* sclerites and *G. aegis* shells (Fig. 2b; Supplementary Data 22 and 23). This finding suggests a high similarity of biomineralization genes between the two studied shells. The high number of ortholog groups found between the sclerite proteome and those of other lineages indicate co-option, and the originally nonbiomineralizing *G. aegis* proteome further revealed that *C. squamiferum* primarily utilizes co-opted genes which dominate the sclerite-formation process.

Evolutionary analysis of sclerite-secreting genes

A protein best-hit search between *C. squamiferum* and *G. aegis* resulted in 116 sclerite-secreting orthologues in *G. aegis*. These orthologues showed discrete expression patterns across *C. squamiferum* tissues and were classified under four orthologue clusters in relation to transcript per million (TPM) (Fig. 3) (Supplementary Data 24) using transcriptomic data. The expression patterns of these four clusters in *G. aegis* tissues were also examined, and the results revealed the increased abundance of orthologs in clusters I, II, and III in the *G. aegis* mantle. Clusters I and IV showed downregulated expressions in the *C. squamiferum* sclerite-secreting epithelium, whereas clusters II and III exhibited upregulation. More than half of the genes composing clusters I (27 out of 32) and 2 out of 5 composing cluster IV, which encode sclerite matrix proteins, were detected through LC-MS/MS. A number of orthologs in cluster I comprised genes encoding framework proteins, such as collagen matrix proteins, and cytoskeletal proteins, including calponin, filamin, and myosin. Cluster II exhibited elevated expressions of sclerite-related genes and their corresponding orthologs in all three armature-producing tissues. Although the *C. squamiferum* sclerite-secreting epithelium and mantle displayed comparable expression patterns in cluster II, mostly cytoskeletal/cytoplasmic *C. squamiferum* shell matrix protein-encoding genes were recovered. Orthologues in cluster III, which included chitin-related genes substantial to sclerite formation, showed exceptionally upregulated expressions in the sclerite-secreting epithelium. This linked pattern in gene upregulation in *G. aegis* shell-forming mantle and *C. squamiferum* sclerite-secreting epithelium suggests a co-option where genes responsible for the development of one type of hard part can be rapidly shifted for use in another type. Although present in all

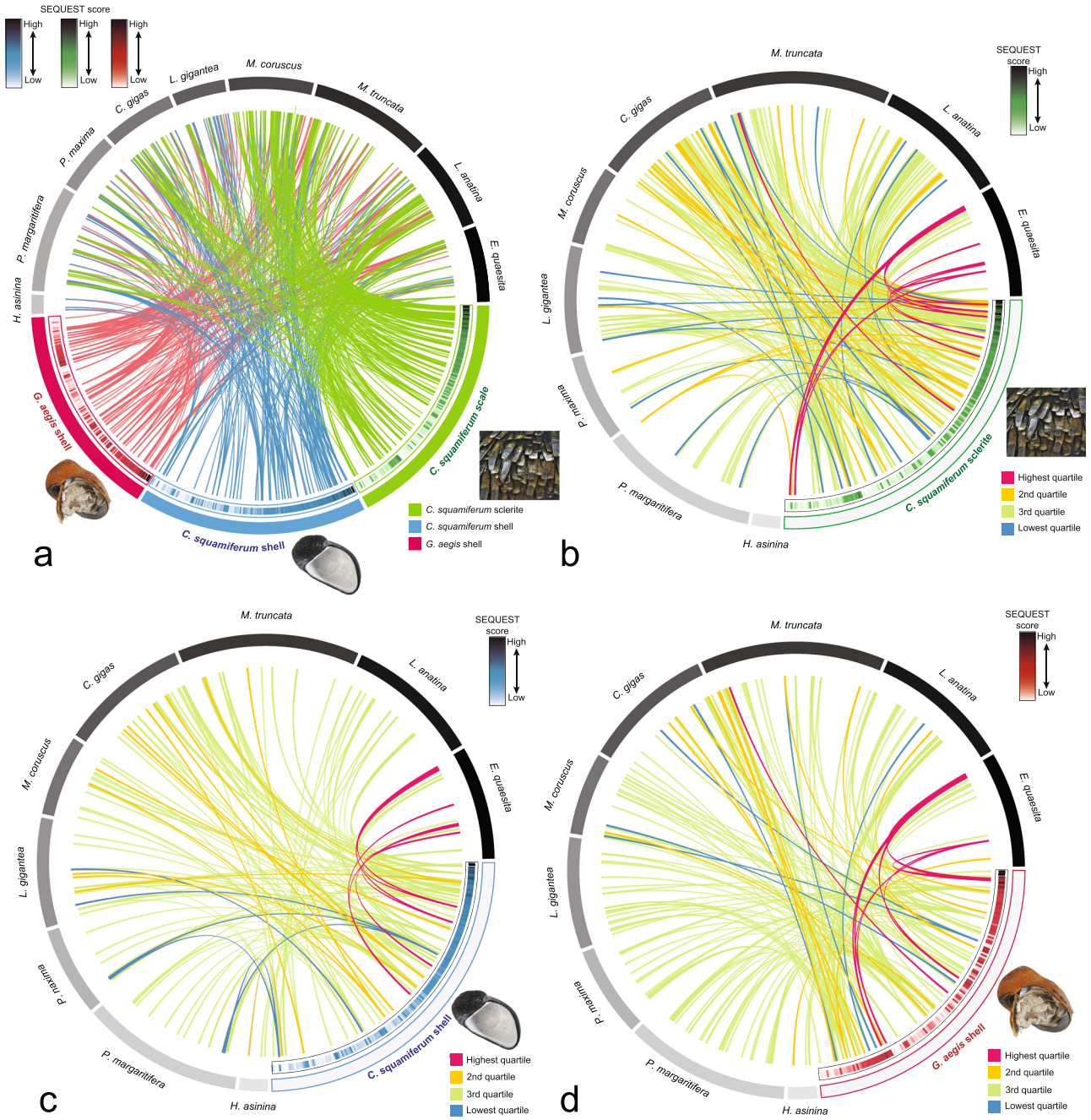

Fig. 1 | Interspecific comparison of the *Chrysomallon squamiferum* shell and sclerite and the *Gigantopelta aegis* shell proteome against 489 SMPs adopted from three gastropods (*Haliotis asinina, Lottia gigantea*, and *Euhadra quaesita*), five bivalves (*Mytilus unguiculatus* (=*M. coruscus*), *Mya truncata, Pinctada margaritifera*, *Pinctada maxima, Crassostrea gigas*), and one brachiopod (*Lingula anatina*). a Overall results of interspecific comparison of the target specimen against the SMP database. Green lines connect proteins of significant homologs between the *C. squamiferum* sclerite and the database, blue lines connect the proteins of significant homolog between the *C. squamiferum* shell and the database, and red lines connect the proteins of significant homologs between the *G. aegis* shell and the database. Colored bars in the inner circle above the target specimen indicate SEQUEST score of the proteins obtained, **b** interspecific comparison of the *C. squamiferum* sclerite proteome and SMP database, **c** interspecific comparison of the *C. squamiferum* shell proteome and SMP database, and **d** interspecific comparison of the *G. aegis* shell proteome and SMP database. Proteins joined by spanning lines across the ideogram indicate their significant similarity of e-value < $10e^{-6}$. Blue lines connect proteins with similarity falling into the lowest quartile, green lines with next highest quartile of similarity, yellow lines with that of the next highest, and red lines with the highest quartile of similarity. Colored bars in the inner circle above the target specimen indicate the SEQUEST score of proteins obtained, and the gradient color indicator displayed at the top right corner of each figure indicates the expression level.

biomineralized parts, foot protein 14 in cluster III also notably exhibited this pattern. Genes in cluster IV were persistently abundant in non-biomineralizing tissues, although the shift in their expressions from foot to esophageal gland not only suggests co-option but also the possible link between symbiosis sclerite biomineralization. Notably, in contrast to the prevailing presence of sclerite-secreting genes, the overall deprivation of shell matrix proteins in these clusters implies the co-option of biomineralization genes.

Gene age investigation with the transcriptome age index (TAI)

Gene age investigation involved the TPM of *C. squamiferum* and *G. aegis* transcriptomes. Genes of *C. squamiferum* and *G. aegis* were assigned to 12

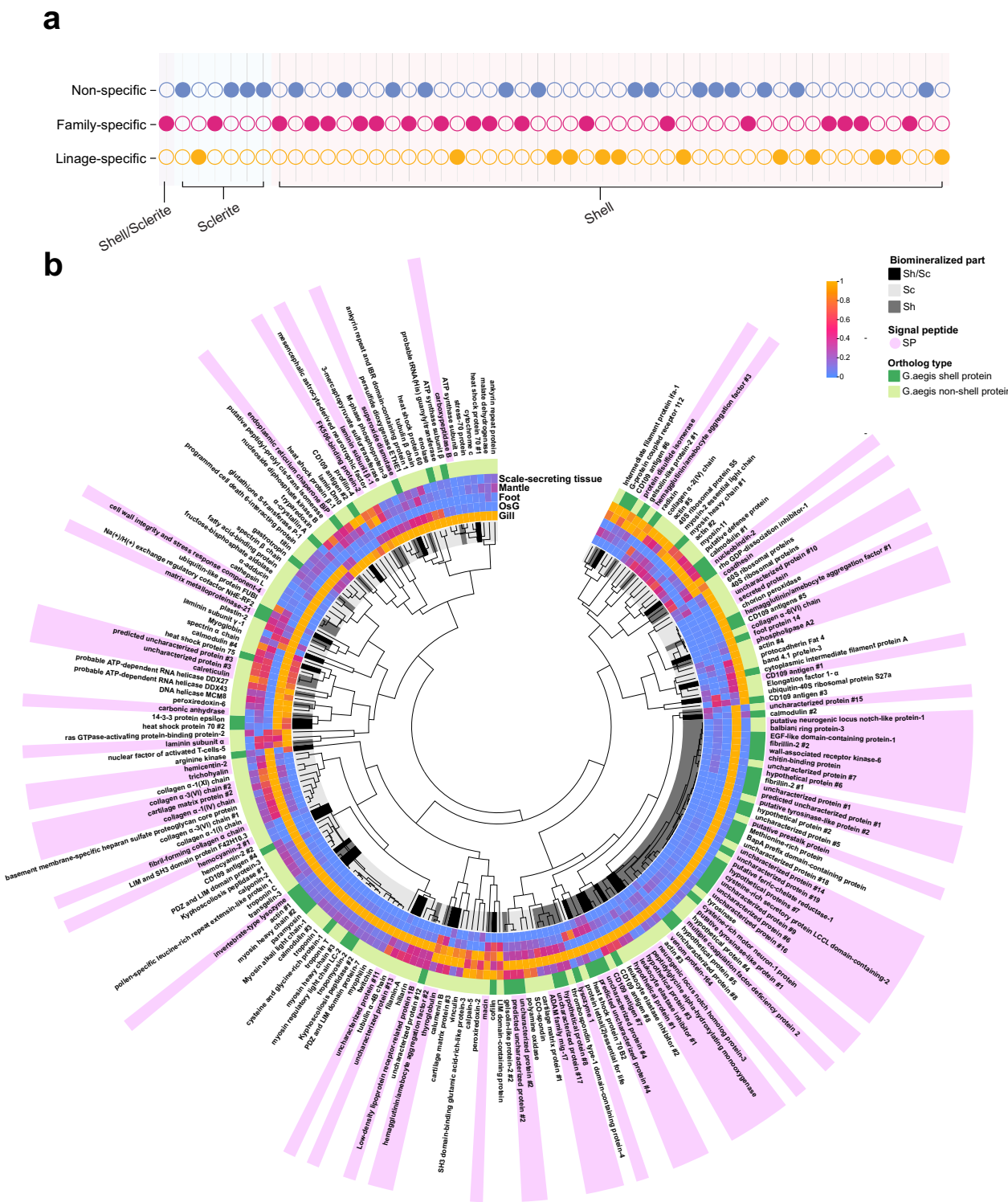

Fig. 2 | Interfamilial comparison of the *C. squamiferum* shell and sclerite and the *G. aegis* shell proteome against their complete proteomes. a Linage specificity of uncharacterized proteins discovered in the *C. squamiferum* sclerite and shell and the *G. aegis* shell. Solid dots indicate the lineage-specific to the uncharacterized proteins. **b** Abundance heatmap of the *C. squamiferum* shell/sclerite proteins that acquired *G. aegis* orthologs in various tissues. Those whose names are highlighted with pink indicate that the corresponding shell/sclerite protein carries signal peptides. The outer ring implies that the ortholog is either *G. aegis* shell or nonshell protein. The inner rings marked with tissue names refer to the tissue proteome of *C. squamiferum* the protein belongs to. The innermost tree was constructed based on the similarity of expression patterns of the corresponding genes in tissues (Ma: mantle; F: foot; OsG: esophageal gland; Gi: gill; Sc: sclerite-secreting epithelium).

phylostrata (cellular organisms–Eukaryota–Opisthokonta–Metazoa–Eumetazoa–Bilateria–Protostomia–Lophotrochozoa–Mollusca–Gastropoda–Peltospiridae–*C. squamiferum*/*G. aegis*) according to the latest common ancestor of homologous genes (Supplementary Fig. 11). Genes nonhomologous to a known sequence were considered lineage-specific (Supplementary Data 25 and 26). In both species, the mantle served as the tissue hosting comparatively younger transcriptome sets, and the gill had the oldest transcriptome sets (Fig. 4). The relative expression levels of each tissue

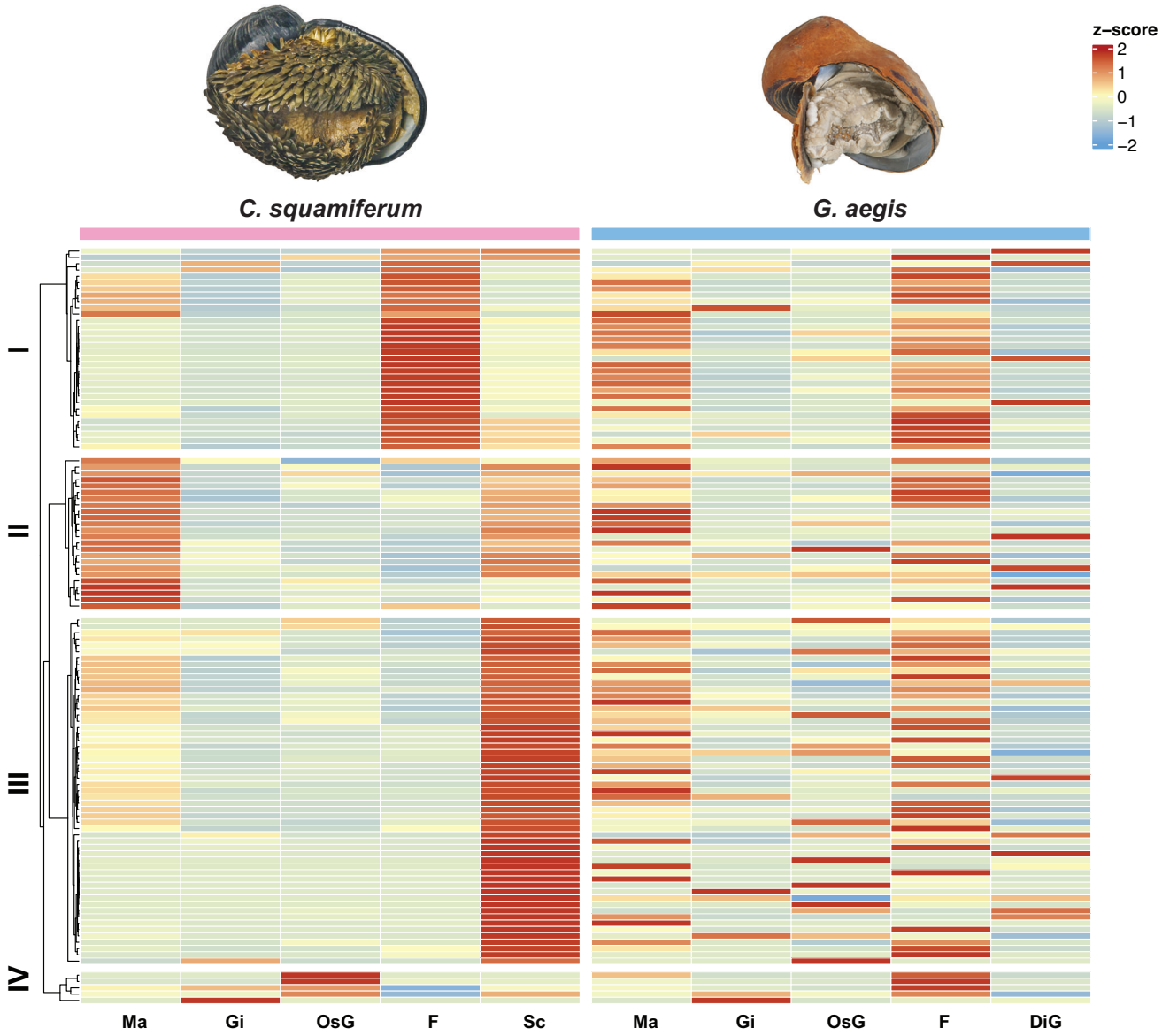

Fig. 3 | Abundance heatmap of sclerite-related matrix protein encoding genes in *C. squamiferum* tissues and the corresponding orthologous genes in *G. aegis* tissues. The orthologues[1] were clustered based on their TPM values in each tissue, and tissues (column) were grouped according to species, with *C. squamiferum* tissues on the left and *G. aegis* tissues on the right. Color gradient indicates the expression level of a particular gene, and the clusters were split into four groups based on the expression patterns of specific genes in *C. squamiferum* tissues with k-means = 4. Cluster I includes sclerite-related matrix protein encoding genes highly expressed in *C. squamiferum* foot tissue. Cluster II comprises sclerite-related genes highly expressed in the *C. squamiferum* mantle and sclerite-secreting epithelium. Cluster III consists of sclerite-related genes particularly expressed in sclerite-secreting epithelium, and Cluster IV includes sclerite-related genes highly expressed in non-biomineralizing tissues (Ma: mantle; F: foot; OsG: esophageal gland; Gi: gill; Sc: sclerite-secreting epithelium).

further indicate the contribution of genes from discrete phylostrata (Fig. 5a, c). In *C. squamiferum*, although most of the genes in most organs emerged during the evolution of Peltospiridae, the contribution of lineage-restricted genes dominated the mantle regulatory network. Similarly, in *G. aegis*, the weighted relative expression profiles of the latest three phylostrata, which are linked to the number of genes allocated, implied the possible active participation of these genes in basic mantle functions. The increased expressions of relatively young genes show a relation to the novelty of shell proteins given that the mantle is the preliminary site of shell secretion. Figure 5b and d show the partial contribution of each phylostratum to the overall mantle TAI. The figures indicate the high utilization of younger genes comprising the shell-biomineralizing proteomes of both species.

The sclerite-secreting epithelium of *C. squamiferum* obtained a discernibly low TAI value, which is consistent with our speculation based on the proteomic analysis. Based on the old mean transcriptomic age, the most highly expressed genes in the mantle emerged close to the phylostratigraphic levels of Mollusca to Gastropoda[47]. To further investigate whether sclerites were products of newly emerged genes when *C. squamiferum* evolved, we calculated the relative expression levels and partial contributions of the sclerite-secreting tissue transcriptome. The relative expression levels revealed an incompatible pattern between the two biomineralized tissues. Still, almost parallel expressions of gene sets were also observed in the mantle and foot tissues throughout the phylostrata (Fig. 5a). The relative expression level profiles of the sclerite-secreting tissue showed the increased participation of genes that emerged from Eukaryota, Opisthokonta, Eumetazoa,

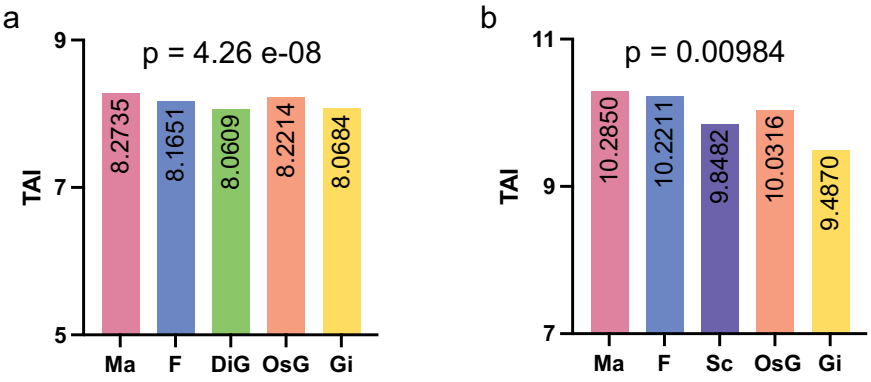

Fig. 4 | Transcriptome age index (TAI) values calculated for *C. squamiferum* and *G. aegis* soft tissues. (Ma: mantle; F: foot; DiG: digestive gland; OsG: esophageal gland; Gi: gill; Sc: sclerite-secreting epithelium). **a** TAI values calculated for *G. aegis* soft tissues. **b** TAI values calculated for *C. squamiferum* soft tissues. Column height and number displayed indicate the TAI value of each tissue. The *p*-value quantifies the significance of the phylostran-scriptomic patterns of employed tissues.

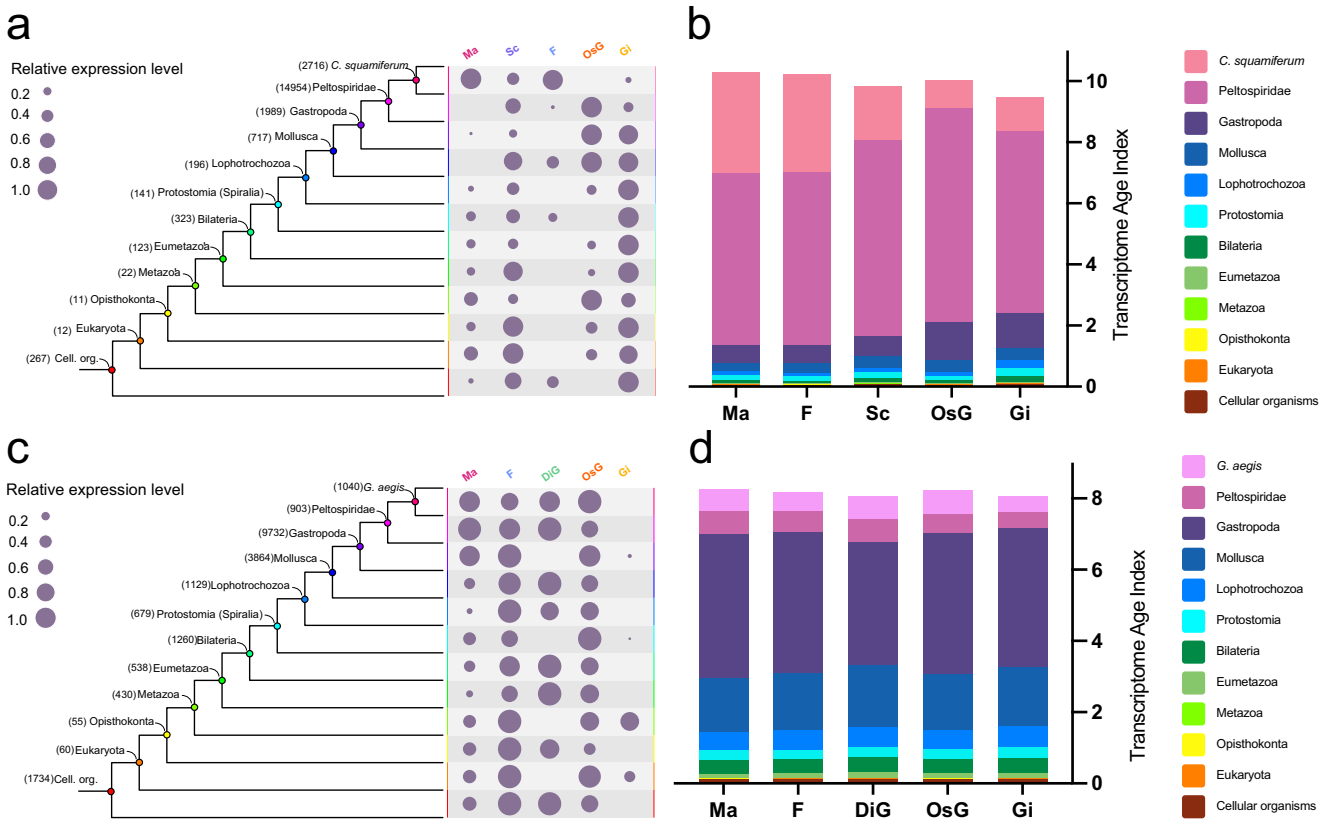

Fig. 5 | Gene number distribution, relative expression level, and partial TAI contribution of different tissues in *C. squamiferum* and *G. aegis*. **a** Phylostratigraphic levels of *C. squamiferum* and the number of genes assigned to the corresponding phylostratum; the dots on the right represent the relative expression level of genes belonging to that phylostratum within the particular tissue. Large dot denotes the high relative expression level of gene sets in that tissue compared with others. **b** Partial TAI of each phylostratum in each *C. squamiferum* tissue, which together make up the overall TAI of the tissue. Different colors refer to phylostrata, and the height of fractions indicates degree of contribution to the overall tissue TAI. **c** Phylostratigraphic levels of *G. aegis* and the number of genes assigned to the corresponding phylostratum; the dots on the right represent the relative expression level of genes belonging to that phylostratum within the particular tissue. Larger dot denotes higher relative expression level of gene sets in that tissue compared with others. **d** Partial TAI of each phylostratum in each *G. aegis* tissue, which together make up the overall TAI of the tissue. Different colors refer to phylostrata, and the height of fractions indicates degree of contribution to the overall tissue TAI. (Ma: mantle; F: foot; DiG: digestive gland; OsG: esophageal gland; Gi: gill; Sc: sclerite-secreting epithelium).

and Mollusca regardless of the gene number. Although the TAI fraction of *C. squamiferum* points to the introduction of novel genes at the origin of family Peltospiridae, the contribution of lineage-restricted genes in the sclerite-secreting tissue remained outnumbered by those in the mantle and foot tissues.

Discussion

In the present work, we provided important findings that emerged from the combined analysis of previously published genomic and transcriptomic data[37,39] and newly acquired downstream proteomic data. This integrated approach offers comprehensive insights into the biomineralization process of the *Chrysomallon squamiferum* sclerites. A former study[38] suggested that *C. squamiferum* utilizes a conserved "biomineralization toolkit" during exoskeleton formation based on genomic evidence, and our results further reveal that the morphologically conserved coiled shell has integrated more recently evolved biomineralization genes. Intraspecific and interspecific analyses of the scleritome and shellomes revealed the dominant contributions of co-opted genes in the rapidly evolving sclerite biomineralization

process of *C. squamiferum*. Furthermore, we utilized data from subsequently published studies on closely related species that lack sclerites to conduct TAI analysis, which further clarified the principal molecular distinctions influencing sclerite production.

The alternative upstream activation of biomineralization toolkit genes in *C. squamiferum* has been suggested to be the key factor leading to the development of the two distinct hard parts[35]. Sclerite-secreting epithelial cells are considered the sole regulation and construction sites of sulfur-rich columns along the longitudinal growth direction on a protein–chitin matrix[34]; however, the expression profile of these cells remains unclear for sclerite formation. The hierarchically clustered genes displayed scattered expression patterns along with a limited signal peptide predicted in the sclerite proteome (Fig. 2), and as a result, the sclerite organic framework obtains disparate biomolecule sources from multiple tissues during formation. Although the presence of corresponding orthologues in the scaleless *Gigantopelta aegis* was enriched in the mantle tissue, *C. squamiferum* further rapidly co-opted shell-building genes into its sclerite formation process.

The morphologically unique *C. squamiferum* sclerites employ proteins that emerged after the divergence of this organism from other scale-less peltospirid snails. Whilst the co-option of ancient genes into the biomineralization-related secretome has a high occurrence throughout molluscan evolution[27], our proteomic study combined with transcriptomic validation indicated that the scaly-foot snail utilizes broadly conserved biomineralization genes in sclerite formation, in line with previous genomic evidence[35]. The synteny between *C. squamiferum* and *G. aegis* implies the lack of chromosomal rearrangement, consistent with the subsequent sclerite gene expression on *G. aegis* mantle and foot, and supports the swift evolution of scales within the family through co-option.

The final sclerite matrix proteome largely reflects the co-option of shell-forming proteins into the sclerite biomineralization process. Despite the lack of calcareous content of sclerites, they possess more proteins with calcium-interactive domains apart from those incorporating the EF-hand 49; in addition, most of these proteins had been recorded in previously studied molluscan shellomes. Calmodulin contributes to the regulation of oyster calcium metabolism, and our finding on calmodulin with multiple entries in the *C. squamiferum* sclerites implies a possible alternative role of this calcium-binding protein[48]. Tyrosinase, on the other hand, is thought to be highly conserved among molluscan shellomes[49,50]; it functions diversely in the construction of periostracum or shell deposition in molluscs[5,51] and is also present in the *C. squamiferum* scleritome. The discovery of multiple hemocyanins in *C. squamiferum* exoskeletons concurs with past knowledge that blood participates in mollusc biomineralization[52]. This scenario may be linked to the elaborate cardiovascular system of *C. squamiferum*[53], where the primary role of hemocyanins in oxygen transportation has been considerably preserved[53]. Furthermore, despite the functional obscurity of foot protein 14, it is likely another case of co-option based on its gene abundancy pattern in the hard-part secreting tissues of *C. squamiferum* and *G. aegis* (Fig. 3). More interestingly, carbonic anhydrase 7 exhibited a substantially a reduced expression in both *C. squamiferum* biomineralizing tissues but showed an increased expression in symbiont-hosting esophageal gland. This finding may indicate either the participation of symbionts in *C. squamiferum* biomineralization processes or conversely the unknown functions of carbonic anhydrase 7 in symbiosis (Supplementary Fig. 12). Given that carbonic anhydrase is essential for the biomineralization process in calcareous sponges[54], other SMPs may compensate for carbonic anhydrase 7 at the shell-forming site, which suggests the functional redundancy of this deeply conserved enzyme. The emergence of functionally redundant regulatory proteins during biomineralization can increase the flexibility and specialization of genes in cases of co-option[55].

Meanwhile, in *C. squamiferum* sclerites, the Pif-like protein likely gave rise to the uniaxially oriented β-chitin columns for sulfur-nanoparticle transportation and detoxication of sulfur-enriched domains[30,34]; chitin-binding domains have been markedly conserved in the molluscan biomineralization toolkit throughout evolution. Pif and Pif-like proteins obtaining various copies and distinct arrangements of vWA and chitin-binding domains have been observed in various molluscan species (Supplementary Fig. 13), such as the pearl oysters *Pinctada margaritifera*, *P. maxima*, and *P. penguin*, where they contribute to nacre formation[56]. The wide distribution of Pif and Pif-like proteins across Mollusca indicates their importance in shell biomineralization and their constructional and functional novelty in sclerite construction, which is a hard structure distinct from the shell, and therefore corroborates the key role of co-option. Furthermore, the absence of signal peptides and the ubiquitous expression of these SMPs in multiple tissues implies their other functions in addition to biomineralization. These functions are likely retained in protein co-option events or employ mechanisms that are excluded in classic extracellular secretion pathways due to the lack of secretion signals. Although the involvement of cellular proteins in molluscan biomineralization has been identified primarily through indicative omics evidence and their critical roles in hard part formation remain uncertain, their repeated emergence in molluscan shellomes of various lineages suggests their likely participation in the biomineralization process[40,57–60].

The resemblance of proteomes with relatively phylogenetically distant animals implies the asymmetrical gain and loss of biomineralization-related proteins in evolution[61]. However, the conservation rate of sclerite proteome surpasses that of another morphologically conserved exoskeleton of the same species, which indicates the incorporation of ancient gene sets in the formation of a completely novel structure. Although the scales were initially assumed to be protective, especially with their iron infusion[32,62], its true function appears to be sulfur detoxification[30,34]. Thus, in *C. squamiferum*, which has a novel scleritome and a conserved shell, the co-option of ancient genes in biomineralization probably led to the evolution of sclerites, which perform a specific function in its adaptation to the hot-vent environment. This controlled acquisition of a novel skeletal tissue by *C. squamiferum* further improves our comprehension of the coordinated adaptation of conserved, ancestrally inherited regulatory elements into distinct biomineralized tissues across distant lineages beyond a single taxonomic class[63].

The mantle secretomes across Mollusca have long been defined as "rapidly evolving"[24] and in combination with organismal or environmental factors, they directly result in the diversity of molluscan shells[24,61]. The shells of the two peltospirid species are not unusual in appearance for gastropods, with the presence of conserved biomineralization-related proteins, such as EGF-like and zona pellucida-containing, and whey acidic protein four-disulfide core proteins that regulate matrix assembly and control crystal nucleation[49,64], collagen isoforms that construct matrix frameworks[65], lipocalins, SR, and alpha-macroglobulin-like proteins, which indicate the involvement of multifunctional SMPs in molluscs[66]. Nevertheless, we discovered a large proportion of newly evolved proteins dominating their biomineralization. Our proteomic and transcriptomic results on *C. squamiferum* and *G. aegis* shellomes confirmed that this morphologically conserved hard part utilizes a considerably large proportion of lineage- or species-specific proteins (Supplementary Fig. 11).

By exploring genomic phylostratigraphy, we unveiled evolutionary milestone events via the assessment of the distinct expression patterns of transcripts distinguished by their phylogenetic origin[67] and uncovered the proportion of younger genes contributing to biomineraliazation. The TAI denotes the integration of gene age and expression level in diverse tissues[47,68] and reflects the boosted evolutionary events in the mantle. The mantle of *C. squamiferum* has a comparatively young transcriptomic age implicated by the augmented weighted mean of the latest phylostratum. With a certain degree of biomineralization gene co-option, the emergence of new genes during the branching of later phylostrata induced the elevated mantle TAI, which accounted for the emergence of novel proteins during shell construction.

On the other hand, transcriptomic evidence revealed that species-specific proteins lack considerable contribution to *C. squamiferum* sclerite secretion. Given that the foot serves as a tissue conserved throughout gastropod evolution, foot TAI reveals the prompt emergence and utilization of lineage-restricted genes. Altogether, we revealed an unexpected pattern

where the secretion of a novel biological structure is dominated by conserved genes, and the secretion of a notably older hard part uses a large number of novel lineage-restricted genes.

Conclusion

The two distinctive hard parts of the hydrothermal vent endemic scaly-foot snail *Chrysomallon squamiferum* employ an ancient "biomineralization tool-kit" of varying protein contents. Sclerite proteins showed a higher homology to previously characterized shell matrix proteins in molluscan lineages other than Peltospiridae, which represents the co-option of calcifying genes into the assembly of a largely proteinaceous structure. This process was further supported by the scattered gene expression patterns of sclerite matrix proteins, which represent the co-options of non-biomineralizing genes into exoskeleton formation through an adaptative process, across the examined soft tissues. The study of orthologues in the foot epidermis of *C. squamiferum* unveiled an unexpected expression pattern of sclerite-making orthologues in the shell-forming tissue in closely related species. This finding indicates a rapid co-option of biomineralization genes, which shifted their roles in biomineralization from the formation of calcium carbonate-based shells to that of chitinaceous/proteinaceous sclerites. On the other hand, shell proteomes and the corresponding mantle edge transcriptomes incorporate novel proteins that likely originate from younger genes, which further demonstrates the high plasticity of molluscan shellomes. Hence, we speculate that co-option is a milestone event contributing to the crucial development of dermal sclerites and some other hard parts in molluscs.

Materials and methods

Animal collection and processing

Specimens of *Chrysomallon squamiferum* were collected from Longqi "Tiamat Chimney" (49°38.97'E, 37°47.03'S, 2785 m deep) and Tiancheng (63°55.40'E, 27°51.05'S, 2682 m deep) vent fields via the remotely operated vehicle *Sea Dragon III* on-board R/V *Dayangyihao* during the China Ocean Mineral Resources Research & Development Association DY52nd cruise in April 2019[37]. *Gigantopelta aegis* snails were also collected from Longqi vent field, "Tiamat Chimney" (49°38.97'E, 37°47.03'S, 2785 m deep) during the same cruise[39]. The snails were immediately transferred to a −80 °C freezer upon arrival and transported to the laboratory for further processing.

Frozen *C. squamiferum* samples were thawed on ice prior to the manual separation of hard parts from soft tissues. The scales were pulled out from the foot of *C. squamiferum* using stainless steel tweezers. Visible soft tissues were initially removed using cotton swabs under a dissection microscope. For the removal of residual epidermal cell tissue on the detaching end of the scales, the scales were gently shaken at a rotation speed of 250 rpm in 0.2% commercial bleach (Clorox, United States) for 1 h at room temperature. This process was repeated with 0.1% and 0.05% commercial bleach for 1 h in the order of descending bleach concentration. The cleaned scales were repeatedly rinsed using Milli-Q water and lyophilized overnight at a chamber pressure of 0.014 mbar and a temperature of −99 °C. The shells were carefully removed from the soft body using stainless steel tweezers, and the periostracum of *C. squamiferum* shells was completely peeled off after overnight lyophilization. The remaining periostracum was eliminated with brushes and tweezers. The shells were sequentially bleached and rinsed as performed on the *C. squamiferum* sclerites.

Frozen *G. aegis* samples were processed in the same manner as *C. squamiferum*. Two individuals were combined as one experimental replicate. Six individuals were used in this experiment. The collection, cleansing, and subsequent processing of *G. aegis* shells were performed in accordance with the same methods as those used for *C. squamiferum* shells.

Shell and sclerite protein extraction

Freeze-dried shells and scales from individuals were crushed in the presence of liquid nitrogen using ceramic pestle and mortar to powder the biomineralized parts. For the extraction of the matrix proteins from the scales, an extraction buffer (2 M hydroxylamine hydrochloride (Sigma-Aldrich, ≥99%), 2 M guanidine hydrochloride (Sigma-Aldrich, ≥98%), and 0.2 M potassium carbonate (Sigma-Aldrich, ≥99%), which was adjusted to pH 9 as described[69], were added to the powder, and the resulting mixture was incubated at 45 °C with shaking at 250 rpm for 12 h. The supernatant was collected after centrifugation at 13,000 rpm (Eppendorf, Centrifuge 5424 R, German) for 5 min at room temperature, and the reaction was terminated through the addition of 3 µL 2% trifluoroacetic acid (Sigma-Aldrich, ReagentPlus, ≥99%, United States) solution per 1 mL supernatant. The extraction procedure was repeated thrice using the same batch of scale powder, and the resultant protein solutions were concentrated using a 3 kDa molecular weight cut-off (MWCO) centrifugal concentrator (Millipore, Amicon Ultra-15, United States) at 25 °C and 4000 g RCF (Himac, Centrifuge CT 18 R, Japan).

SMP extraction was performed in accordance with a previously published protocol[70,71] with adjustments. Shell powder (~1.0 g) lyophilized at a chamber pressure of 0.014 mbar and a temperature of −99 °C overnight was suspended in 20 mL 10% v/v acetic acid with mild rotation at 4 °C overnight for decalcification. The supernatant was collected after 30 min of centrifugation at 4 °C and 4000 g RCF and concentrated to 5 mL for an extra 20 min with a 3 kDa MWCO centrifugal concentrator. Continuous dialyzed shell protein solution was against 10× volume of Milli-Q water under the same centrifugation conditions.

Methanol/chloroform precipitation was applied to precipitate the extracted scale and shell protein. The pellet acquired was dissolved with 30 µL lysis buffer (8 M urea, 40 mM HEPES, Sigma-Alrich, United States), and the reconstituted protein was further purified using a ReadyPrepTM 2-D cleanup kit (Bio-Rad, United States).

LC-MS/MS

Longqi scale and shell proteins were analyzed by Dionex UltiMate 3000 RSLCnano accompanying Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher, United States)[1,2]. Chromatographic separation was executed with a 120 min gradient at a flow rate of 300 nL/min, in which started with 10 min of 0% (100% mobile phase A: 0.1% formic acid in H_2O) to 2% buffer B (0.1% formic acid in ACN), followed by 2 min of 2% buffer B to 6%, then 70 min linear gradient of 6% buffer B to 20%, and then 10 min of 20% buffer B to 30%, and finally ended with 8 min of 30% buffer B to 100%. Orbitrap was set to positive ion mode with scan range of 400–1500 m/z, where the dynamic exclusion duration was set to 40 s. Resolution was set to 60,000 for MS detection at FT mode, and HCD collision energy was set to 30% for MS/MS detection at ion trap mode. Isolation window was set to 1.6 m/z with a fixed first mass of 110 m/z.

Tiancheng *C. squamiferum* scale proteins and shell proteins, and *G. aegis* shell proteins were anaylzed with nanoElute® separation system coupled with timsTOF Pro Mass-spectrometer (Bruker Headquarters Billerica, United States). Digested proteins were injected into the nanoElute® system, chromatographic separation was performed on a C18 column (ionopticks Aurora UPLC column, Part No. AUR2-25075C18A-CSI) with a 30 min gradient at a flow rate of 300 nL/min. Mobile phase A was 98% H_2O with 2% ACN and 0.1% formic acid, and phase B was 0.1% formic acid in ACN. Liquid chromatographic separation was achieved with 0.5 min of 2% buffer B to 5%, 26.5 min of 5% buffer B to 30%, and 0.5 min of 30% buffer B to 95%, then 0.1 min of 95% buffer B to 2%, and 2% buffer was maintained for 1.9 min to reach 30 min gradient. Mass spectrometry was set to positive ion mode with scan range of 100-1700 m/z, where the ramp time was set to 100 ms and the spectra rate was set to 9.52 Hz. The electrospray ionization (ESI) source was set with capillary voltage of 1400 V, 3.0 L/min dry gas at a temperature of 180 °C. Ion mobility was scanned from 0.85 to 1.30 Vs/cm^2 ($1/k_0$), and one TIMS-MS scan together with four Parallel Accumulation-Serial Fragmentation (PASEF) MS/MS scans composed the one acquisition cycle (0.53 s).

Protein identifications

The collected proteins were desalted with PierceTM C18 spin columns (Thermo Fisher, United States) and digested with PierceTM trypsin

protease (Thermo Fisher, MS Grade, United States). Processed Longqi *C. squamiferum* sclerite proteins were analyzed on a Dionex UltiMate 3000 RSLCnano equipped with an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher, United States; LC-MS/MS (1), Supplementary Data), and Tiancheng *C. squamiferum* sclerite proteins, shell proteins from both vent sites, and *G. aegis* shell proteins were analyzed on a nanoElute separation system coupled with timsTOF Pro mass-spectrometer (Bruker Headquarters Billerica, United States; LC-MS/MS (2); Supplementary Data). The obtained scaly-foot snail peptide fragments were matched against the translated gene models of *C. squamiferum* constructed in a previous study on this species[35] with Protein Discoverer 3.0[72], and the reversed sequences were applied as decoys. The obtained *G. aegis* peptide fragments were matched by Proteome Discoverer 3.0 against the translated gene models of *G. aegis* constructed in a prior work[39]. The search parameters of Proteome Discoverer were set at 0.02 Da and 10 ppm for fragment-ion mass tolerance parent ion tolerance, respectively. Fix modification was selected with carbamidomethyl (cysteine) and variable modification with oxidation (methionine), where the missed trypsin cleavage value was set at 2. Proteins with at least one identified unique sequence and the expected confident identification level exceeding 0.95 were accepted.

Bioinformatic analyses

Published transcriptome data on five tissues, including scale-secreting epithelium, esophageal gland, gill, foot, and mantle, from three scaly-foot snail individuals obtained from Solitaire vent field sequenced from in situ fixed specimens using an RNA-stabilizing solution[35] were used in analyses. The transcriptome data on five tissues (mantle, foot, gill, esophageal gland, and digestive gland) of four *G. aegis* individuals from Longqi vent field, on the other hand, were directly obtained from published data by Lan et al.[39]. Gene expression quantification and differential expression analysis were performed using our previously procedure[35] and described as follow.

Differentially expressed genes of *C. squamiferum* were established using DESeq2 with paired test ($n = 3$) among sclerite-secreting epithelium, mantle, and foot ($n = 2$) versus other tissues. The differentially expressed genes of *G. aegis* were verified with EdgeR utilizing the calculated CPM value of transcripts from a past study (Supplementary Data 27–30)[39]. Transcripts with FDR < 0.05 and fold change > 2 in both analyses were then determined as differentially expressed genes. For proteomic analysis, signal peptides were predicted using the online version of SignalP v6.0[73], and transmembrane helices were predicted using TMHMM v2.0[74] to identify the secreted macromolecules. DeepLoc 2.0[75] was introduced to predict the cellular location of SMPs and possible signals. RLCDs were predicted using online XSTREAM v1.73 (https://amnewmanlab.stanford.edu/xstream/) under default settings, and ID region prediction was performed with Mobid-Lite v3.9.0[76] at default settings.

Heatmaps and binary plots were established using ChiPlot (https://www.chiplot.online/), and GO enrichment was established using TBtools[77]. Average linkage clustering and Euclidean distance measuring methods were performed in heatmap construction. Conserved functional domains were predicted using InterProScan v. 5.52–86.0[78], and domains identified against SMART, Pfam, and CDD databases with e-values < 0.01 were obtained. Interspecies protein comparison was performed using blastp v2.13.0 + [79], with a cutoff e-value < 10e-6, whereas the SMP database utilized in the search comprised 497 SMP sequences obtained from nine lophotrochozoans (Supplementary Data 14). Orthofinder[80], which shows close phylogenetic relationships, was used to uncover the orthologous proteins related to biomineralization in *C. squamiferum* and *G. aegis*[39]. Blastp results indicate interspecies SMP similarity, and Orthofinder results revealed the orthologs. These findings were visualized with a circus plot by OmicsStudio (https://www.omicstudio.cn/tool). Genes of *C. squamiferum* and *G. aegis* were assigned to 12 phylostrata independently with Diamond v2.1.4[81]. TAI was then estimated using myTAI R package[68] with default parameters, in which the square root transformation of *G. aegis* expression data was used to calculate the effective *p*-value. TAI results were visualized using TBtools[77] and further adjusted with Adobe Illustrator.

The macrosynteny between *C. squamiferum* and *G. aegis* was constructed via JCVI[82] using coding sequences and genomic coordinate files. A total of 17 species were used to construct the gene families, including those of 14 gastropods: *Alviniconcha marisindica*, *Bellamya purificata*, *Lanistes nyassanus*, *Marisa cornuarietis*, *Pomacea canaliculata*, *Aplysia californica*, *Biomphalaria glabrata*, *Candidula unlfasclata*, *Elysia chlorotica*, *Chrysomallon squamiferum*, *Gigantopelta aegis*, *Bathyacmaea lactea*, *Eulepetopsis crystallina*, and *Lottia gigantea* as well as three outgroup species: *Acanthopleura granulata*, *Perna viridis*, and *Nautilus pompilius*. Orthologues without paralogues among all 17 species were generated by Orthofinder[80] phylopyprunner v1.2.4. The calibrated phylogenetic tree was constructed in accordance with a published pipeline[83–85]. CAFE5 v5.0.0[86] was used to detect the significant expansion of gene families with a lineage-specific "Viterbi P-value" less than 0.01. A list of sclerite-related proteins (Supplementary Data 24) was generated for co-option orthologue analysis. Based on the former Orthofinder result, the best-hit proteins between *C. squamiferum* and *G. aegis* were generated using blastp v2.13.0 + [79].

Shell microstructural identification

SEM was performed to observe microstructures of fractured shells and polished sclerites. Analyses of fractured shells from *C. squamiferum* and *G. aegis* were performed on a JSM-6390 (JEOL, United States) at the Material Characterization and Preparation Facility (MCPF) in Clear Water Bay, Hong Kong. Samples were fractured with tweezers, mounted on 25 mm brass-pin stub with Cu tape, and sputter coated with gold in K575xd dual-head coater (Emitech Ltd., United States) for SEM examination. The fractured shells were imaged under a high vacuum mode at an accelerating voltage ranging from 10 kV to 20 kV. The specimens were fractured along the growth line, whereas the microstructures of the inner shell surface in contact with the mantle and surfaces between the periostracum and prismatic layers were analyzed. The porosity of *C. squamiferum* and *G. aegis* shells was measured using ImageJ software[87], and the average percentage of porous area within the 10 μm × 10 μm space on the inner shell surface was calculated.

For the samples characterized through transmission electron microscopy (TEM), the thin samples of *C. squamiferum* were prepared through milling and fabrication processes with a grinding machine (Buehler Metaserv 2000) and Dual Station Ion Milling System (Gatan 600 CTMP), which enable cross-sectional milling to obtain electron-transparent specimen. The *C. squamiferum* samples were initially cut into small pieces (2 mm × 2 mm) and ultrasonically cleaned in ethanol for 10 min. The samples were first clamped using two wafers with epoxy in the middle for adhesion and then thinned and polished together on a grinder. The polished samples were mounted on a Cu aperture with 800 μm diameters for ion milling ahead of TEM imaging. The TEM images that exhibited shell microstructures were captured using a JEM-ARM200F cold field-emission gun transmission electron microscope operated at 200 kV with a point resolution of 1.1 Å.

Relatively large-sized *C. squamiferum* sclerites were collected for sample preparation for sclerite observations. The cross-sections of sclerites were mounted in epoxy resin and polished using an abrasive paper. The treated specimens were first analyzed with an e-LiNE Electron Beam Lithography system (Raith, Germany) at an accelerating voltage 8 kV under high vacuum mode at the MCPF. A light microscope was used to depict complete cross-sections. Terminologies from a previous paper were adapted to evaluate the biomineral layers and features[88].

Statistics and reproducibility

The sample size of the *C. squamiferum* sclerites consists of six individuals, including three specimens from the Tiancheng vent field and three from the Longqi vent field. Each individual was treated as a separate experimental replicate with all sclerites from one single specimen included. The sample size of the *C. squamiferum* shell consists of five individuals, including three individuals from the Tiancheng vent field and two from the Longqi vent field and each individual served as one experimental replicate. For *G. aegis* samples collected from Longqi vent field, two individuals were combined to

treat as one experimental replicate, and there were six individuals used in this experiment in total.

For protein identification from LC-MS/MS, decoy sequences were applied to reduce the likelihood of false-positive detections. The proteomics search was conducted using a species-specific database. Proteins identified in three or more experimental replicates were included in the sclerite/shell proteomes.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The proteomics data underlying this article are available a ProteomeXchange at https://proteomecentral.proteomexchange.org/ui and can be accessed with accession number PXD051659 under the project name, *Proteomic analyses reveal the key role of gene co-option in the evolution of the Scaly-foot Snail scleritome*. Numerical source data for the graphs presented in Fig. 5 can be found in a standalone GitHub repository (github.com/wcwongay/SFGscleriteproteomics) (https://doi.org/10.5281/zenodo.14886348). All other data are available from the first author on reasonable request.

Code availability

Detailed commands were documented in the Supplementary Code file in a standalone GitHub repository (github.com/wcwongay/SFGscleriteproteomics) (DOI: 10.5281/zenodo.14886348) to reproduce the results of this manuscript. Software packages used in this manuscript include Protein Discoverer 3.0[72] (https://doi.org/10.3390/proteomes9010015), Mobid-Lite v3.9.0[76] (http://old.protein.bio.unipd.it/mobidblite/), TBtools[77] (https://github.com/CJ-Chen/TBtools-II), InterProScan v.5.52-86.0[78] (https://www.ebi.ac.uk/interpro/about/interproscan/), blastp v2.13.0 + [79] (https://www.uniprot.org/blast), Orthofinder[80] (https://github.com/davidemms/OrthoFinder), Diamond v2.1.4[81] (https://github.com/bbuchfink/diamond), myTAI R package[68] (https://github.com/drostlab/myTAI), JCVI[82] (https://github.com/tanghaibao/jcvi), phylopyprunner v1.2.4 (https://github.com/fethalen/phylopypruner), and CAFE5 v5.0.0[86] (https://github.com/hahnlab/CAFE5). Online analysis tools applied in this manuscript include SignalP v6.0[73] (https://services.healthtech.dtu.dk/services/SignalP-6.0/), TMHMM v2.0[74] (https://services.healthtech.dtu.dk/services/TMHMM-2.0/), DeepLoc 2.0[75] (https://services.healthtech.dtu.dk/services/DeepLoc-2.0/), XSTREAM v1.73 (https://x-stream.github.io/), and OmicsStudio (https://www.omicstudio.cn/tool).

Received: 30 May 2024; Accepted: 19 February 2025;
Published online: 28 February 2025

References

1. Williams, R. J. P. An introduction to biominerals and the role of organic molecules in their formation. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **304**, 411–424 (1984).
2. Lowenstam, H. A. & Weiner, S. *On biomineralization*. (Oxford University Press on Demand, 1989).
3. Lowenstam, H. & Weiner, S. in *Biomineralization and biological metal accumulation* 191-203 (Springer, 1983).
4. Weiner, S., Traub, W. & Lowenstam, H. in *Biomineralization and biological metal accumulation* 205-224 (Springer, 1983).
5. Aguilera, F., McDougall, C. & Degnan, B. M. Evolution of the tyrosinase gene family in bivalve molluscs: independent expansion of the mantle gene repertoire. *Acta Biomater.* **10**, 3855–3865 (2014).
6. Tambutté, S., Tambutté, E., Zoccola, D. & Allemand, D. in *Handbook of Biomineralization* 243-259 (2007).
7. Simkiss, K. & Wilbur, K. M. *Biomineralization*. (Elsevier, 2012).
8. Smith, M. P. & Harper, D. A. Causes of the Cambrian explosion. *Science* **341**, 1355–1356 (2013).
9. Vendrasco, M. J., Kouchinsky, A. V., Porter, S. M. & Fernandez, C. Z. Phylogeny and escalation in Mellopegma and other Cambrian molluscs. *Palaeontol. Electron.* **14**, 1–44 (2011).
10. Murdock, D. J. & Donoghue, P. C. Evolutionary origins of animal skeletal biomineralization. *Cells Tissues Organs* **194**, 98–102 (2011).
11. Faivre, D. & Godec, T. U. From bacteria to mollusks: the principles underlying the biomineralization of iron oxide materials. *Angew. Chem. Int. Ed.* **54**, 4728–4747 (2015).
12. Weaver, J. C. et al. Analysis of an ultra hard magnetic biomineral in chiton radular teeth. *Mater. Today* **13**, 42–52 (2010).
13. Saunders, M., Kong, C., Shaw, J. A., Macey, D. J. & Clode, P. L. Characterization of biominerals in the radula teeth of the chiton, *Acanthopleura hirtosa*. *J. Struct. Biol.* **167**, 55–61 (2009).
14. Saunders, M., Kong, C., Shaw, J. A. & Clode, P. L. Matrix-mediated biomineralization in marine mollusks: a combined transmission electron microscopy and focused ion beam approach. *Microsc. Microanal.* **17**, 220–225 (2011).
15. Peters, S. E. & Gaines, R. R. Formation of the 'Great Unconformity' as a trigger for the Cambrian explosion. *Nature* **484**, 363–366 (2012).
16. Petrychenko, Y., Peryt, T. M. & Chechel, E. I. Early Cambrian seawater chemistry from fluid inclusions in halite from Siberian evaporites. *Chem. Geol.* **219**, 149–161 (2005).
17. Knoll, A. H. Biomineralization and evolutionary history. *Rev. Mineral. Geochem.* **54**, 329–356 (2003).
18. Vendrasco, M. J., Porter, S. M., Kouchinsky, A., Li, G. & Fernandez, C. Z. New data on molluscs and their shell microstructures from the Middle Cambrian Gowers Formation, Australia. *Palaeontology* **53**, 97–135 (2010).
19. Jackson, D. J. et al. Parallel evolution of nacre building gene sets in molluscs. *Mol. Biol. Evol.* **27**, 591–608 (2010).
20. Westbroek, P. & Marin, F. A marriage of bone and nacre. *Nature* **392**, 861–862 (1998).
21. Wood, R. A. Paleoecology of the earliest skeletal metazoan communities: implications for early biomineralization. *Earth-Sci. Rev.* **106**, 184–190 (2011).
22. Schiemann, S. M. et al. Clustered brachiopod Hox genes are not expressed collinearly and are associated with lophotrochozoan novelties. *Proc. Natl. Acad. Sci.* **114**, E1913–E1922 (2017).
23. Hohagen, J. & Jackson, D. J. An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. *BMC Dev. Biol.* **13**, 1–14 (2013).
24. McDougall, C. & Degnan, B. M. The evolution of mollusc shells. *Wiley Interdiscip. Rev. Dev. Biol.* **7**, e313 (2018).
25. Jackson, D. J. et al. A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol.* **4**, 1–10 (2006).
26. Marin, F., Luquet, G., Marie, B. & Medakovic, D. Molluscan shell proteins: primary structure, origin, and evolution. *Curr. Top. Dev. Biol.* **80**, 209–276 (2007).
27. Aguilera, F., McDougall, C. & Degnan, B. M. Co-option and de novo gene evolution underlie molluscan shell diversity. *Mol. Biol. Evol.* **34**, 779–792 (2017).
28. McMullin, E. R., Bergquist, D. C. & Fisher, C. R. Metazoans in extreme environments: adaptations of hydrothermal vent and hydrocarbon seep fauna. *Gravit. Space Res.* **13**, 13-23 (2007).
29. Fisher, C. R., Takai, K. & Le Bris, N. Hydrothermal vent ecosystems. *Oceanography* **20**, 14–23 (2007).
30. Isobe, N. et al. Uniaxial orientation of β-chitin nanofibres used as an organic framework in the scales of a hot vent snail. *J. R. Soc. Interface* **19**, 20220120 (2022).
31. Warén, A., Bengtson, S., Goffredi, S. K. & Van Dover, C. L. A hot-vent gastropod with iron sulfide dermal sclerites. *Science* **302**, 1007–1007 (2003).
32. Yao, H. et al. Protection mechanisms of the iron-plated armor of a deep-sea hydrothermal vent gastropod. *Proc. Natl. Acad. Sci.* **107**, 987–992 (2010).

33. Chen, C., Copley, J. T., Linse, K., Rogers, A. D. & Sigwart, J. How the mollusc got its scales: convergent evolution of the molluscan scleritome. *Biol. J. Linn. Soc.* **114**, 949–954 (2015).
34. Okada, S. et al. The making of natural iron sulfide nanoparticles in a hot vent snail. *Proc. Natl. Acad. Sci.* **116**, 20376–20381 (2019).
35. Sun, J, et al. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nat. Commun.* **11**, 1–12 (2020).
36. Wernström, J. V., Gąsiorowski, L. & Hejnol, A. Brachiopod and mollusc biomineralization is a conserved process that was lost in the phoronid–bryozoan stem lineage. *Evodevo* **13**, 17 (2022).
37. Sun, J. et al. Nearest vent, dearest friend: biodiversity of Tiancheng vent field reveals cross-ridge similarities in the Indian Ocean. *R. Soc. Open Sci.* **7**, 200110 (2020).
38. Chen, C., Linse, K., Roterman, C. N., Copley, J. T. & Rogers, A. D. A new genus of large hydrothermal vent-endemic gastropod (Neomphalina: Peltospiridae). *Zool. J. Linn. Soc.* **175**, 319–335 (2015).
39. Lan, Y. et al. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat. Commun.* **12**, 1–15 (2021).
40. Carini, A. et al. Proteomic investigation of the blue mussel larval shell organic matrix. *J. Struct. Biol.* **208**, 107385 (2019).
41. Oudot, M. et al. The shell matrix and microstructure of the Ram's Horn squid: Molecular and structural characterization. *J. Struct. Biol.* **211**, 107507 (2020).
42. Song, N. et al. In vitro crystallization of calcium carbonate mediated by proteins extracted from *P. placenta* shells. *CrystEngComm* **24**, 7200–7215 (2022).
43. Weiner, S. Mollusk shell formation: isolation of two organic matrix proteins associated with calcite deposition in the bivalve *Mytilus californianus*. *Biochemistry* **22**, 4139–4145 (1983).
44. Zhang, C. & Zhang, R. Matrix proteins in the outer shells of molluscs. *Mar. Biotechnol.* **8**, 572–586 (2006).
45. Chen, C., Uematsu, K., Linse, K. & Sigwart, J. D. By more ways than one: rapid convergence at hydrothermal vents shown by 3D anatomical reconstruction of *Gigantopelta* (Mollusca: Neomphalina). *BMC Evolut. Biol.* **17**, 1–19 (2017).
46. Suzuki, Y. et al. Sclerite formation in the hydrothermal-vent "scaly-foot" gastropod—possible control of iron sulfide biomineralization by the animal. *Earth Planet. Sci. Lett.* **242**, 39–50 (2006).
47. Domazet-Lošo, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).
48. Li, S., Xie, L., Ma, Z. & Zhang, R. cDNA cloning and characterization of a novel calmodulin-like protein from pearl oyster *Pinctada fucata*. *FEBS J.* **272**, 4899–4910 (2005).
49. Shimizu, K., Kintsu, H., Awaji, M., Matumoto, T. & Suzuki, M. Evolution of biomineralization genes in the prismatic layer of the pen shell *Atrina pectinata*. *J. Mol. Evolut.* **88**, 742–758 (2020).
50. Shimizu, K. et al. Insights into the evolution of shells and love darts of land snails revealed from their matrix proteins. *Genome Biol. Evolut.***11**, 380–397 (2019).
51. Zhang, C., Xie, L., Huang, J., Chen, L. & Zhang, R. A novel putative tyrosinase involved in periostracum formation from the pearl oyster (*Pinctada fucata*). *Biochem. Biophys. Res. Commun.* **342**, 632–639 (2006).
52. Liu, C. et al. Proteomics of shell matrix proteins from the cuttlefish bone reveals unique evolution for cephalopod biomineralization. *ACS Biomater. Sci. Eng.* **9**, 1796-1807 (2021).
53. Chen, C., Copley, J. T., Linse, K., Rogers, A. D. & Sigwart, J. D. The heart of a dragon: 3D anatomical reconstruction of the 'scaly-foot gastropod'(Mollusca: Gastropoda: Neomphalina) reveals its extraordinary circulatory system. *Front. Zool.* **12**, 1–16 (2015).
54. Jackson, D. J., Macis, L., Reitner, J., Degnan, B. M. & Wörheide, G. Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. *Science* **316**, 1893–1895 (2007).
55. Sambamoorthy, G. & Raman, K. Understanding the evolution of functional redundancy in metabolic networks. *Bioinformatics* **34**, i981–i987 (2018).
56. Suzuki, M., Iwashima, A., Kimura, M., Kogure, T. & Nagasawa, H. The molecular evolution of the pif family proteins in various species of mollusks. *Mar. Biotechnol.* **15**, 145–158 (2013).
57. Jackson, D. J. et al. The Magellania venosa biomineralizing proteome: a window into brachiopod shell evolution. *Genome Biol. Evolut.* **7**, 1349–1362 (2015).
58. Upadhyay, A., Vengatesen, T. & Ying, T. Proteomic characterization of oyster shell organic matrix proteins (OMP). *Bioinformation* **12**, 266 (2016).
59. Gao, P. et al. Layer-by-layer proteomic analysis of *Mytilus galloprovincialis* shell. *PLoS One* **10**, e0133913 (2015).
60. Zhang, G. et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
61. Kocot, K. M., Aguilera, F., McDougall, C., Jackson, D. J. & Degnan, B. M. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. *Front. Zool.* **13**, 1–10 (2016).
62. Moya-Costa, R., Bauluz, B. & Cuenca-Bescós, G. Structure and composition of the incisor enamel of extant and fossil mammals with tooth pigmentation. *Lethaia* **52**, 370–388 (2019).
63. Murdock, D. J. The 'biomineralization toolkit'and the origin of animal skeletons. *Biol. Rev.* **95**, 1372–1392 (2020).
64. Smith, V. J. Phylogeny of whey acidic protein (WAP) four-disulfide core proteins and their role in lower vertebrates and invertebrates. *Biochem. Soc. Trans.* **39**, 1403–1408 (2011).
65. Liu, X. et al. Hyriopsis cumingii Hic52—a novel nacreous layer matrix protein with a collagen-like structure. *Int. J. Biol. Macromol.***102**, 667–673 (2017).
66. Arivalagan, J. et al. Shell matrix proteins of the clam, *Mya truncata*: roles beyond shell formation through proteomic study. *Mar. Genomics* **27**, 69–74 (2016).
67. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
68. Drost, H.-G., Gabel, A., Liu, J., Quint, M. & Grosse, I. myTAI: evolutionary transcriptomics with R. *Bioinformatics* **34**, 1589–1590 (2018).
69. Tan, Y. et al. Infiltration of chitin by protein coacervates defines the squid beak mechanical gradient. *Nat. Chem. Biol.* **11**, 488–495 (2015).
70. Marin, F., Pereira, L. & Westbroek, P. Large-scale fractionation of molluscan shell matrix. *Protein Expr. Purif.* **23**, 175–179 (2001).
71. Ramos-Silva, P. et al. The skeletal proteome of the coral Acropora millepora: the evolution of calcification by co-option and domain shuffling. *Mol. Biol. Evolut.* **30**, 2099–2112 (2013).
72. Orsburn, B. C. Proteome discoverer—a community enhanced data processing suite for protein informatics. *Proteomes* **9**, 15 (2021).
73. Teufel, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
74. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
75. Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H. & Winther, O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* **50**, W228-W234 (2022).
76. Necci, M., Piovesan, D., Dosztányi, Z. & Tosatto, S. C. MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* **33**, 1402–1404 (2017).
77. Chen, C. et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).
78. Quevillon, E. et al. InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116–W120 (2005).

79. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
80. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
81. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
82. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
83. He, X. et al. Genomic analysis of a scale worm provides insights into its adaptation to deep-sea hydrothermal vents. *Genome Biol. Evolut.* **15**, evad125 (2023).
84. Liu, X., Sigwart, J. & Sun, J. Phylogenomic analyses shed light on the relationships of chiton superfamilies and shell-eye evolution. *Mar. Life Sci. Technol.* **5**, 525–537 (2023).
85. Sun, Y. et al. Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm *Paraescarpia echinospica*. *Mol. Biol. Evolut.* **38**, 4116–4134 (2021).
86. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2020).
87. Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42 (2004).
88. Carter, J. G. & Clark, G. R. Classification and phylogenetic significance of molluscan shell microstructure. *Ser. Geol. Notes Short. Course* **13**, 50–71 (1985).

Acknowledgements

This work was supported by principal investigator projects of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (2021HJ01), Major Basic and Applied Research Projects of Guangdong Province (2019B030302004-04), Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC01), Hong Kong Special Administrative Region government (16101822, C2013-22GF), National Natural Science Foundation of China (No. 42176110), the Fundamental Research Funds for the Central Universities (202172002 and 202241002), and the Young Taishan Scholars Program of Shandong Province (tsqn202103036). We thank the support of the High-Performance Biological Supercomputing Center at the Ocean University of China, and the Biosciences Central Research Facility (BioCRF) at the Hong Kong University of Science and Technology, for this research.

Author contributions

P.-Y.Q., J.S., and W.C.W. conceived and designed the project. C.C. and J.S. collected and fixed the samples. C.C. provided the images of the specimen used in the manuscript. W.C.W., J.S., and Y.H.K. dissected the samples. W.C.W. and Y.H.K. performed protein extraction, protein sequencing, and identification. K.G. and X.H. conducted homology analyses. X.H. performed synteny and gene family analyses. W.C.W. and S.X. executed scanning electron microscopy analyses. W.C.W. performed bioinformatics analyses. W.C.W. and J.S. interpreted the data and drafted the paper. Y.X., L.L., N.W., C.C., and L.W. contributed to the critical review and editing of the manuscript. All authors contributed to the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07785-7.

Correspondence and requests for materials should be addressed to Pei-Yuan Qian or Jin Sun.

Peer review information *Communications Biology* thanks Ahmed Saadi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: David Favero. A peer review file is available.