Journal of Structural Biology 212 (2020) 107612

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

From sporadic single genes to a broader transcriptomic approach: Insights into the formation of the biomineralized exoskeleton in decapod crustaceans

Shai A. Shaked[a,1], Shai Abehsera[a], Tom Levy[a,2], Vered Chalifa-Caspi[b], Amir Sagi[a,*,3]

[a] Department of Life Sciences and the National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[b] Bioinformatics Core Facility, Ben-Gurion University of the Negev, Beer-Sheva, Israel

ARTICLE INFO

Keywords:
Biomineralization
Cuticle
Cherax quadricarinatus
Exoskeleton
Gastrolith
Transcriptomic library

ABSTRACT

One fundamental character common to pancrustaceans (Crustacea and Hexapoda) is a mineralized rigid exoskeleton whose principal organic components are chitin and proteins. In contrast to traditional research in the field that has been devoted to the structural and physicochemical aspects of biomineralization, the present study explores transcriptomic aspects of biomineralization as a first step towards adding a complementary molecular layer to this field. The rigidity of the exoskeleton in pancrustaceans dictates essential molt cycles enabling morphological changes and growth. Thus, formation and mineralization of the exoskeleton are concomitant to the timeline of the molt cycle. Skeletal proteinaceous toolkit elements have been discovered in previous studies using innovative molt-related binary gene expression patterns derived from transcriptomic libraries representing the major stages comprising the molt cycle of the decapod crustacean *Cherax quadricarinatus*. Here, we revisited some prominent exoskeleton-related structural proteins encoding and, using the above molt-related binary pattern methodology, enlarged the transcriptomic database of *C. quadricarinatus*. The latter was done by establishing a new transcriptomic library of the cuticle forming epithelium and molar tooth at four different molt stages (i.e., inter-molt, early pre-molt, late pre-molt and post-molt) and incorporating it to a previous transcriptome derived from the gastroliths and mandible. The wider multigenic approach facilitated by the newly expanded transcriptomic database not only revisited single genes of the molecular toolkit, but also provided both scattered and specific information that broaden the overview of proteins and gene clusters which are involved in the construction and biomineralization of the exoskeleton in decapod crustaceans.

1. Introduction

The biomineralized hardened skeleton is a feature common to the wide range of phylogenetic classes that inherited an organic scaffold and a 'biomineralization toolkit' from a common ancestor (Murdock, 2020). Known to be an ancient feature used by organisms for protection and muscle support, the exoskeleton has been extensively studied in crustacean species (Abehsera et al., 2018a; Bentov et al., 2016a), but the genes, proteins and protein expression patterns associated with the development and functioning of the crustacean exoskeleton are yet to be thoroughly investigated. The recent construction of transcriptomic libraries of the red-claw crayfish *Cherax quadricarinatus* (Abehsera et al., 2015), as a model animal, has provided the first step towards revealing the scope and the intricacies of the molecular toolkit responsible for exoskeleton formation and mineralization in crustaceans. To date, a small number of genes encoding the proteins associated with the formation and hardening of the *C. quadricarinatus* exoskeleton have been discovered and fully characterized (Bentov et al., 2016a; Glazer and Sagi, 2012; Glazer et al., 2015; Roer et al., 2015), but it is our belief that there are many more genes yet to be revealed. By significantly enlarging our *C. quadricarinatus* transcriptomic database and applying our previously developed binary expression pattern method (Abehsera et al., 2015), we set out here to characterize and better understand both the time frame and the scope of the molecular toolkit responsible for the formation and biomineralization of the crustacean exoskeleton.

The decapod crustacean exoskeleton consists of a chitin-protein scaffold that becomes hardened by the deposition of minerals, such as

* Corresponding author.
E-mail address: sagia@bgu.ac.il (A. Sagi).
[1] Orcid ID: 0000-0003-1995-6419.
[2] Orcid ID: 0000-0003-1484-0310.
[3] Orcid ID: 0000-0002-4229-1059.

https://doi.org/10.1016/j.jsb.2020.107612
Received 5 July 2020; Received in revised form 27 August 2020; Accepted 29 August 2020
Available online 05 September 2020

calcium carbonate and calcium phosphate, in both their amorphous and crystalline forms. The creation of such hardened scaffolds is enabled by a complex proteinaceous toolkit that has evolved in crustaceans. Some of the genes/proteins comprising this toolkit were discovered by applying the above-mentioned innovative molt-related binary expression pattern method to existing transcriptomic data (Abehsera et al., 2015). In this binary expression pattern method, each gene of interest is assigned a four-digit code that represents whether the gene exhibits relatively high or relatively low expression in each of the four major molt stages of *C. quadricarinatus*, namely, intermolt, early pre-molt, late pre-molt and post-molt, in that order in the code. This quantification method has made it possible to perform a reliable analysis of distinct molt-related gene expression patterning based on transcriptomic libraries representing the major stages of *C. quadricarinatus* molt cycle (Abehsera et al., 2015).

Most crayfish, and our study animal *C. quadricarinatus* in particular, live in freshwater habitats with low concentrations of calcium. To cope with the challenge of a limited calcium supply, these animals make use of a unique temporal calcium reservoir comprised largely of amorphous calcium carbonate (ACC). Two such reservoirs, known as gastroliths, are built and then resorbed during each molt cycle (Huxley, 1880). During pre-molt, the gastroliths are built in two dedicated anterior lateral areas of the cardiac stomach wall. After ecdysis (the molt event), biomineralization of the new cuticle is initiated simultaneously with the collapse of the gastroliths into the animal stomach. There, digesting enzymes act to release the mineral, which subsequently precipitates within the exoskeletal organic matrix of the new cuticle (Glazer and Sagi, 2012; Shechter et al., 2008a). Thus, the mineralization timelines of the gastroliths and other parts of the exoskeletal cuticle along the molt cycle differ, one from the other (Shechter et al., 2008a). Gastroliths are formed and mineralized in the early and late pre-molt stages, and thus the most common binary expression pattern of genes associated with the gastrolith-forming epithelium is 0110 (i.e., high expression in both pre-molt stages and low expression in intermolt and post-molt). The carapace cuticle is formed in the late pre-molt stage, mineralized in the post-molt stage, and maintained in the intermolt stage. Thus, it is likely that the most abundant binary expression patterns in the cuticle-forming epithelium – alongside the 0110 pattern – will be the 0100, 0010, 0001 and 0011 patterns (Abehsera et al., 2018a, 2015, 2018b; Shechter et al., 2008a).

In this study, we revisited the binary molt-related expression patterns of the few known genes related to exoskeleton formation and mineralization toolkits, with focus on *C. quadricarinatus*. These toolkits included sporadic cases of gene families involved in exoskeleton anabolism and catabolism processes—from genes encoding functional and structural proteins incorporated into chitin scaffolds to genes encoding proteins that build specific cuticular elements. To gain a deeper understanding of these toolkits, we applied the above-described binary expression pattern method to a new extensive transcriptomic database of two exoskeletal elements formed and mineralized at different molt stages, namely, the cuticle and the gastroliths.

2. Materials and methods

2.1. RNA preparation for the new transcriptomic library

To enable further characterization of genes related to the formation and biomineralization of the exoskeleton, a new *C. quadricarinatus* transcriptomic library was constructed. This library complemented the library previously established for this species with additional descriptive information in terms of both molt stages and tissues (Abehsera et al., 2015). As the first step towards preparing the new library, total RNA was extracted from two tissues, namely, the molar-tooth-forming epithelium and cuticle-forming epithelium, of three *C. quadricarinatus* individuals at each of the four molt stages (i.e., intermolt, early pre-molt, late pre-molt and post-molt) using EZ-RNA Total RNA Isolation Kit (Biological Industries, Beit Ha'emek, Israel), according to the manufacturer's protocol, as previously described (Levy et al., 2016). RNA samples (n = 24) were sent for sequencing using paired end 150-bp stretches on an Illumina NovaSeq 6000 system. Bioinformatic analysis was carried out by the BGU Bioinformatics Core Facility, using the NeatSeq-Flow platform (Sklarz et al., 2018) and in-house R scripts.

A super-transcriptome was then assembled using the 24 samples of the current study and samples from two previous projects (Abehsera et al., 2015), as follows: (1) 19 samples from the molar-forming epithelium and the gastrolith-forming epithelium from three molting stages in which gastrolith is present and intermolt in which gastrolith was not developed yet (50-bp single end reads; 208,604,795 reads in total); and (2) pooled RNA extracted from the molar-forming epithelium, and the gastrolith-forming epithelium (100-bp paired-end reads, 187,414,352 pairs in total). After filtering out the transcripts with very low expression, the resulting transcriptome contained 196,434 transcripts (> 200 bp) from 125,979 putative genes.

Due to the large transcriptome size and to the presence of fragmented transcripts for key genes associated with the molt and mineralization processes, additional reference transcriptomes were created for each tissue separately. A tissue-specific transcriptome for the gastrolith was not produced, since the 19 gastrolith samples were sequenced at 50-bp single-end reads, and therefore gene expression in the gastrolith was assessed using the super-transcriptome. Gene expression in the cuticle and molar tissues was assessed using the respective tissue-specific transcriptome.

2.2. Quality assessment of the raw reads

All reads from the current study and the previous projects were quality trimmed with TrimGalore! Version 0.4.2, which removed less than 1–8% of the reads. Ribosomal RNA was filtered out as follows: reads were aligned to a database of crustacean rRNA sequences downloaded from NCBI with bwa mem version 0.7.16 (Li, 2013) using default parameters. Reads that aligned to the rRNA database (11–14%) were discarded using samtools version 1.3 (Li et al., 2009). A total of 2,183,388,592 clean reads were retained for further analysis.

2.3. Transcriptome assembly and annotation

Reference transcriptomes were *de novo* assembled using Trinity version 2.5.1 (Grabherr et al., 2011), and further filtered to exclude transcripts with very low expression. To this end, the clean reads were aligned to the transcriptome, and only transcripts in which at least one of the experimental groups had at least two replicates with ≥3 transcripts per million (TPM) counts were retained. Due to its large size, a more stringent cutoff of ≥15TPM counts was used in the super-transcriptome. Each transcriptome was quality assessed using Quast (Gurevich et al., 2013) version 4.5 and BUSCO (Simao et al., 2015) vs. Metazoa_odb9 database. The super-transcriptome included 97.2% of the BUSCO proteins, whereas the tissue-specific cuticle and molar transcriptomes included 98.0 and 94.8% of the BUSCO proteins, respectively. For transcriptome annotation, the most highly expressed transcript per gene was selected using the filter_low_expr_transcripts.pl script from the Trinity software suite. These representative transcripts were annotated using Trinotate version 3.1.1 (Trinotate.github.io).

2.4. Differential expression analysis

Clean reads were aligned to the respective filtered reference transcriptome (see above) with bowtie2 version 2.3.5 (Langmead and Salzberg, 2012), and gene expression was estimated with RSEM version 1.3.0 (Li and Dewey, 2011). For quality assessment, counts were subjected to variance stabilizing transformation (DESeq2 (Love et al., 2014)), sample-wise correlation analysis and principal component analysis (PCA). One sample from the molar-forming epithelium at inter-

molt was found to be exceptional and was thus excluded from further analysis. Statistical testing for differential expression was carried out using DESeq2, a method specifically tailored for count data by the use of negative binomial generalized linear models. Genes with significant differences between any two molt stages, in each of the tissues separately, were identified using DESeq2 contrasts. Genes were considered to be differentially expressed in a certain contrast if they had a false discovery rate (FDR) adjusted p-value < 0.05 and linear fold change > 1.3 or < -1.3, where the minus sign denotes down-regulation. For each tissue, a unified list of differentially expressed genes in any of the comparisons was constructed.

2.5. Binary gene expression pattern analysis

Two separate unified lists of differentially expressed genes, one for the cuticle and one for the gastroliths, were submitted to binary gene expression pattern analysis. To this end, a list of all possible binary patterns ($n = 14$), not including "0000" and "1111", was constructed. For each gene, Pearson's correlation coefficient between the gene's variance stabilized counts across all samples was computed, and an artificial expression profile of each pattern was constructed. For example, in the cuticle (for which there were 3 replicate samples per molt stage), the pattern 0110 was represented by the artificial profile "000111111000". Note that high correlation of a gene to a pattern does not indicate a "presence" and "absence" status, but rather relatively "high" and "low" expression levels. Genes were assigned to the pattern with which they had the highest Pearson's correlation, as long as the correlation was larger than 0.8. Hierarchical clustering of the genes in each pattern, after z-scoring of their variance stabilized expression values, was carried out using R's pheatmap function (Kolde, 2012).

2.6. Schematic representation of structural proteins amino acid sequences

Protein encoding genes analyzed in this paper were translated from nucleotide to protein sequence using ExPasy [https://web.expasy.org/translate/(Gasteiger et al., 2003)] and later analyzed using SMART [http://smart.embl-heidelberg.de/(Letunic and Bork, 2018)]. For each protein encoding gene, longest translated open reading frame was chosen for analysis in SMART. Properties for SMART analysis were as follows: protein sequences were submitted in FASTA format, and Outlier homologues, PFAM domains, signal peptides and internal repeats were searched for each protein sequence.

2.7. Phylogenetic trees of genes encoding for cuticular structural proteins

To assess sequence similarity and evolutionary distance between known cuticular structural proteins of crustaceans' species, multiple sequence alignments and phylogenetic trees of several protein encoding genes were made. Genes encoding for cuticular structural proteins discovered in *C. quadricarinatus* as well as in other crustaceans' species were searched in our new transcriptomic database. The sequences of these proteins transcripts derived from our new transcriptomic library were used to construct phylogenetic trees depicting evolutionary distance between our study animal and other crustacean species cuticular structural proteins involving in formation and mineralization of the exoskeleton. Genes analyzed are listed in supplemental table S1. The transcripts were then blasted (in blastn mode) against Transcriptome Shotgun Assembly (TSA) database of decapods: "organism: Decapoda (taxid: 6683)" or against Non-redundant protein sequences (nr) in blastx mode (https://blast.ncbi.nlm.nih.gov). All protein encoding genes found in more than 10 decapod crustacean species (including *C. quadricarinatus*) were analyzed as follows: Multiple sequence alignment (MSA) of the mRNA sequences was carried out using MAFFT program version 7 (Katoh et al., 2019). To test which is the most suitable model to select for phylogenetic analysis with likelihood-based criteria, Smart Model Selection (SMS) function with Bayesian information criterion (BIC) in PhyML (Lefort et al., 2017) was used. The model that was eventually chosen is K80 and model decoration (i.e. RAS and equilibrium frequency options) were +G (Lefort et al., 2017). Evolutionary phylogenetic analysis was visualized using iTOL (Letunic and Bork, 2019).

2.8. Comparison of differentially expressed genes between cuticle and gastrolith

Differentially expressed genes from the cuticle transcriptome that had a binary expression pattern (Pearson's correlation > 0.8) and whose normalized counts in all "positive" samples (i.e. considered as 1 in the pattern) were larger than 500 were used as query in a blastn search against all genes from the super-transcriptome. Up to 10 best hits per query, with e-value $<$ 1e-7, percent identity > 90 and query coverage ("qcovs") > 70, were retrieved, considering only the best HSP per hit. The expression and annotation details of the cuticle and their matched gastrolith genes were joined in one file. The list consisted of genes having different binary expression pattern in the cuticle and gastrolith, where in both tissues we requested that all positive samples have > 500 normalized counts.

The same analysis was repeated in the opposite direction, such that differentially expressed genes from the super-transcriptome having a binary expression pattern in the gastrolith and whose normalized counts in all positive samples was > 500 were blasted against the cuticle transcriptome, followed by merging the expression results according to the blast matches and preparation of the gene list described above.

3. Results and discussion

Molting in arthropods is an essential process involving a repetitive replacement of the exoskeleton to enable morphological changes and growth. The pancrustacean exoskeleton is made up of a chitin-protein scaffold hardened with calcium carbonate and calcium phosphate, both amorphous and crystalline (Bentov et al., 2016b). It has previously been shown that a number of proteins are involved in scaffold formation and in biomineral transportation and deposition in the chitinous pancrustacean scaffold (Kuballa et al., 2007; Watanabe et al., 2000). Out of these, a small number have been discovered in crustaceans and are revisited in the present article. Since the genes involved in skeletal formation and mineralization are expressed in patterns closely related to the molt cycle, we examined their expression patterns in the gastrolith- and cuticle-forming epithelia—tissues that differ in their mineralization and formation timelines over the molt cycle. Gastroliths have been extensively studied in two crayfish species, the red swamp crayfish *Procambrus clarkii* and *C. quadricarinatus*, with many of the studies focusing on proteins in the organic matrices in these two species (Glazer et al., 2015, 2010; Ishii et al., 1996, 1998; Shechter et al., 2008b; Tsutsui et al., 1999). Here, we used transcriptomic libraries to study the expression patterns of the few known molt-related genes responsible for formation and mineralization in the gastrolith- and exoskeleton-forming epithelium of *C. quadricarinatus*. Importantly, we then sought to broaden our overview of molt and mineralization related proteins by using a multigenic approach based on the above datasets to study gene clusters sharing similar molt-related expression patterns.

3.1. Transcriptomic library

The present study addresses the transcriptomics of the formation and mineralization of cuticular structures. The newly established transcriptomic database of *C. quadricarinatus* constructed for the study comprises 2,183,388,592 clean reads, which were retained for construction of the three transcriptomes described above. The above-described assemblies yielded filtered transcriptomes with 169,390, 135,476 and 196,434 contigs in the cuticle and molar transcriptomes

and the super-transcriptome, respectively. Additional data for the transcriptomes of the cuticle, molar tissue and the super-transcriptome, respectively, were: total contigs length 186,448,530, 168,339,245 and 232,078,712 bp; sequencing depth 840, 927 and 2743; and average contig length 1100, 1242 and 1181 bp, with N90 values of 466, 506 and 470 bp, respectively. In the following sections, we describe the use of the above transcriptomic datasets for revisiting single genes with respect to their molt-related expression patterns.

3.2. GAMP

P. clarkii was the first crustacean in which a structural protein was discovered and characterized in the gastroliths (Ishii et al., 1996, 1998; Tsutsui et al., 1999). Named GAMP (gastrolith matrix protein), this 50,500-Da insoluble acidic protein was found to be tightly bound to chitin, possibly through covalent bonds (Ishii et al., 1998). The high proportion of glutamic acid and glutamine residues in the protein may imply that it participates in calcification by calcium binding and inhibition of calcium carbonate crystallization in addition to its chitin binding characteristics (Ishii et al., 1998).

In the present study, we also found the *GAMP* transcript in *C. quadricarinatus* transcriptomes, with the transcript showing a binary expression pattern similar to that of other proteins involved in gastrolith formation and mineralization. Using the binary patterning method, expression values for the gastrolith-forming epithelium obtained in a previously assembled transcriptomic library (Abehsera et al., 2015) were compared to the cuticular data from the newly assembled transcriptomic library reported in the present study. While the binary expression pattern of *GAMP* in the gastrolith was found to be 0110, the cuticular pattern was 0011 (Fig. 1A). These findings are in keeping with the different mineralization timelines of the two organs (Shechter et al., 2008a) and provide supporting evidence for the involvement of GAMP in hardening and mineralization processes during the molt cycle. Phylogenetic analysis of *GAMP* in decapod crustacean species was carried out showing high similarity with *C. quadricarinatus GAMP* transcript being on the earlier branches among decapod crustaceans (Fig. 1B).

Since the identification of GAMP, additional gastrolith proteins (GAPs) have been identified. Two of these proteins, designated GAP65 and GAP10, were extracted from the gastrolith of *C. quadricarinatus* (Bentov et al., 2010; Glazer et al., 2010; Shechter et al., 2008b). The proteins were named according to the migration of a proteinaceous extract of gastroliths on SDS-PAGE at apparent approximate molecular masses of 65 and 10 kDa, respectively.

3.3. GAP65

In the *C. quadricarinatus* gastroliths, the negatively charged glycoprotein GAP65 was found to be the most abundant protein of the GAP family (Shechter et al., 2008b). Sequencing of this protein revealed a signal sequence and three other domains – two that bind chitin and one (a low-density lipoprotein receptor class A domain) that interacts with calcium ions – thus suggesting that GAP65 participates both in the formation and in the calcification of the gastrolith (Abehsera et al., 2018a; Shechter et al., 2008b), suggesting GAP65 to be a structural protein involved in formation and mineralization of the gastrolith forming epithelium. This conclusion is supported by a study of gastrolith matrix proteins in *C. quadricarinatus* showing three putative protein complexes (Glazer et al., 2015), each containing GAP65 and each believed to play a different role in gastrolith formation (through chitin metabolism), sclerotization (through chitin network remodeling) and

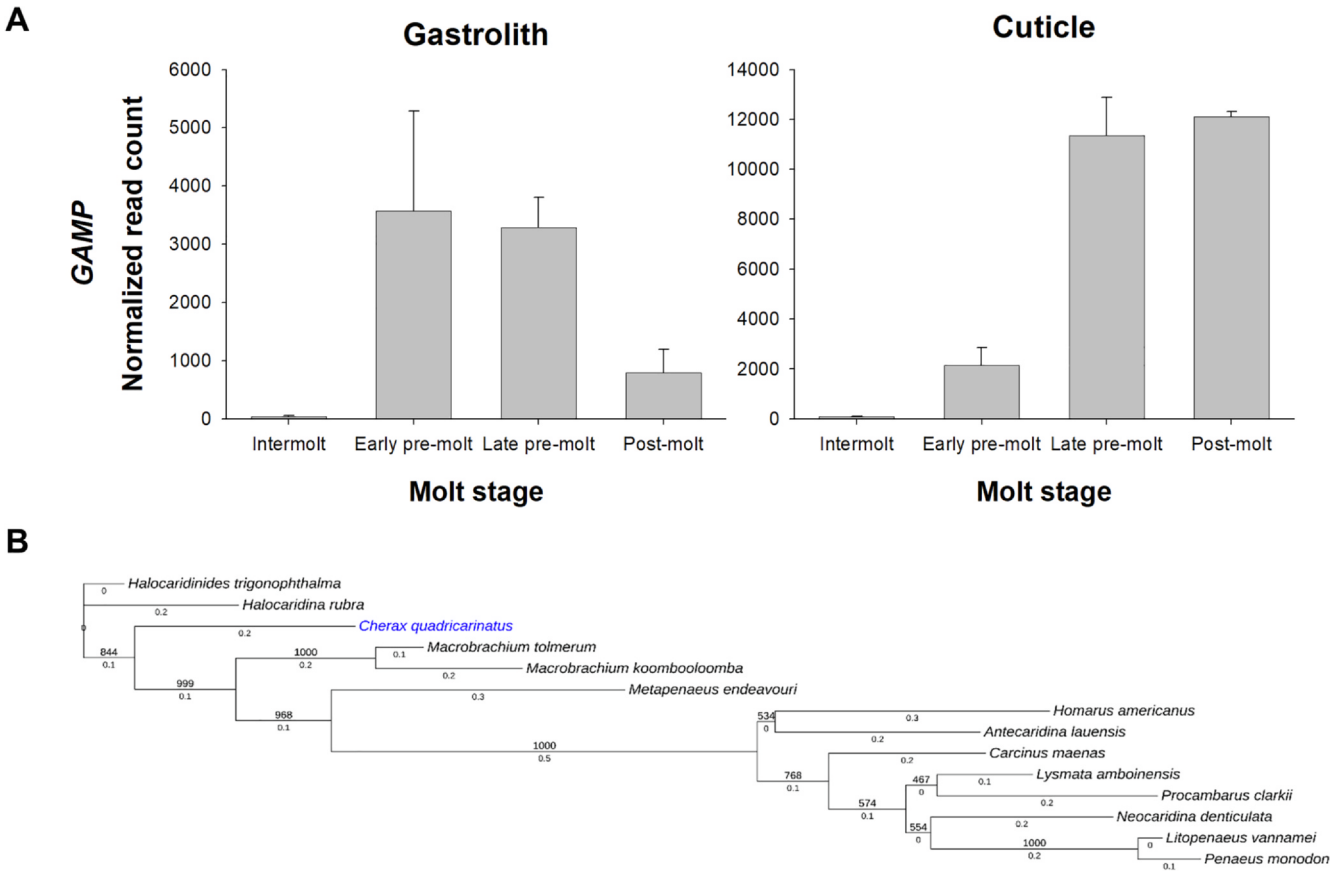


Fig. 1. (A) *GAMP* expression pattern in gastrolith- and cuticle-forming epithelium. Expression levels are presented as normalized read counts in four different molt stages of *C. quadricarinatus*. X-axis represents the four molt stages. Error bars represent standard error. (B) Phylogenetic tree of *GAMP* transcripts sequences in 14 decapod crustacean species. Supporting values on junctions and branch lengths are given. *C. quadricarinatus* is marked in blue.

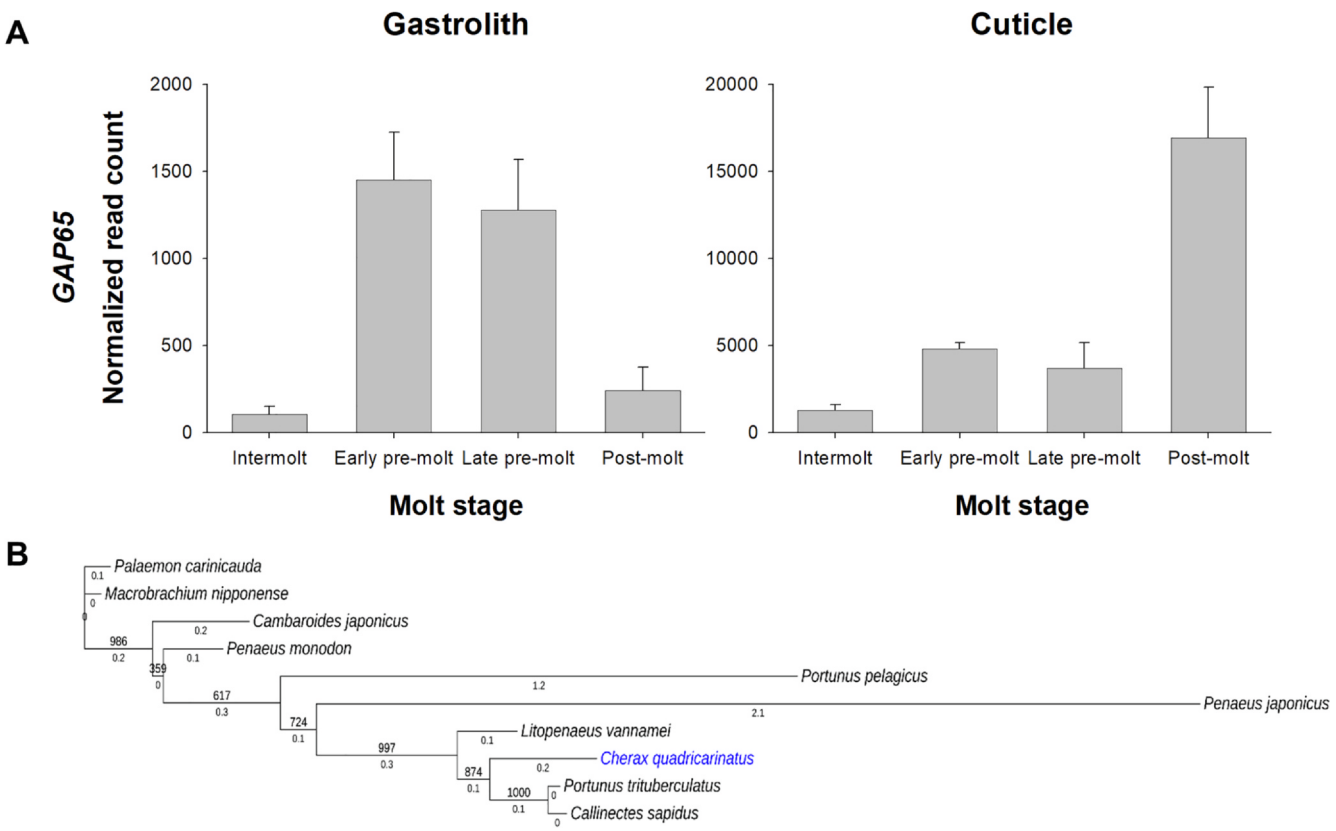


Fig. 2. (A) *GAP65* expression pattern in gastrolith- and cuticle-forming epithelium. Expression levels are presented as normalized read counts in four different molt stages of *C. quadricarinatus*. X-axis represents the four molt stages. Error bars represent standard error. (B) Phylogenetic tree of *GAP65* transcripts sequences in 10 decapod crustacean species. Supporting values on junctions and branch lengths are given. *C. quadricarinatus* is marked in blue.

hardening (through $CaCO_3$ precipitation. Determination of GAP65 protein function was carried out through knock-down experiments using RNAi, which resulted in a clear structural phenotypic change, showing deformed, less dense gastroliths with more loosely packed columnar chitin structures and significantly larger ACC spherules (Shechter et al., 2008b). When the calcium binding activity of GAP65 was investigated *in vitro*, it was found that synthetic ACC spherules were produced from calcium carbonate and that these spherules remained stable for up to two months (Shechter et al., 2008b).

Binary gene expression patterning of *GAP65* revealed a molt-related expression pattern of 0110 in the gastrolith forming tissue (Fig. 2A), similar to the expression pattern of *GAMP* (Fig. 1). The expression pattern in the cuticle tissue was 0001 (Fig. 2A), as expected, since GAP65 is involved in biomineralization and ACC stabilization: GAP65 is known to have both calcium- and chitin-binding properties and is highly expressed in both the gastrolith- and cuticle-forming tissue during the biomineralization phases of these tissues (Abehsera et al., 2018a). The findings obtained by our new transcriptomic library are thus in keeping with the notion that GAP65 is essential for the formation and hardening of cuticular structures. *C. quadricarinatus GAP65* transcript was compared to orthologs from other decapod crustaceans (Fig. 2B). This finding suggests that *GAP65* is a conserved protein encoding gene which was developed in different evolutionary events.

3.4. GAP10

Following the characterization of GAP65, a second GAP protein – GAP10, a small, 10-kDa acidic molecule – was discovered in the *C. quadricarinatus* gastrolith (Glazer et al., 2010). That study showed that in addition to its expression in the gastrolith, GAP10 is also expressed in the carapace cuticle-forming epithelium of *C. quadricarinatus*. GAP10 contains a signal peptide and two known consensus sequences previously identified in arthropod cuticular proteins, namely, the AAP (A/V) repeat and four copies of glycine rich residues (GGX). The AAP (A/V) residue is a small hydrophobic tetrapeptide that is found mainly in proteins of the hard cuticle and is therefore believed to stabilize the three-dimensional protein structure through hydrophobic interactions (Glazer et al., 2010). The glycine-rich residues form β-turns stabilized by hydrogen bonds. Both the glycine-rich regions and the AAP(A/V) consensus sequence endow matrix proteins with elasticity and high tensile strength (Glazer et al., 2010). The amino acid composition of GAP10 was found to be similar to that of other cuticular proteins associated with crustacean calcification, all which have calcium-binding domains inhibiting the precipitation of calcium carbonate. These serves as proof that GAP10, similar to GAP65, is another key structural protein in formation and mineralization of exoskeletal tissues of *C. quadricarinatus*. Microarray experiments on the GAP10 encoding transcript showed that *GAP10* is highly upregulated in the gastrolith-forming tissue during early and late pre-molt (Glazer et al., 2010). According to the new transcriptomic library established in the current study, *GAP10* is significantly expressed in both the gastrolith- and cuticle-forming tissues in all four major molting stages. Taken together, the calcium-binding ability of GAP10 and its high expression levels concomitant with calcification-related molt stages in the gastrolith and the carapace cuticle indicate that GAP10 plays an essential role in the hardening of these two cuticular structures (Fig. 3). Loss-of-function experiments resulted in a prolonged pre-molt stage and led to the development of gastroliths with irregularities appearing exclusively on the gastrolith surface, i.e., in the most recently deposited gastrolith layers (Glazer and Sagi, 2012). These phenotypical changes suggest that this protein is intimately involved in the structural formation of the chitin–protein–mineral complex of the gastrolith-forming tissue (Glazer et al., 2010).

The binary expression pattern of *GAP10* in the gastrolith-forming

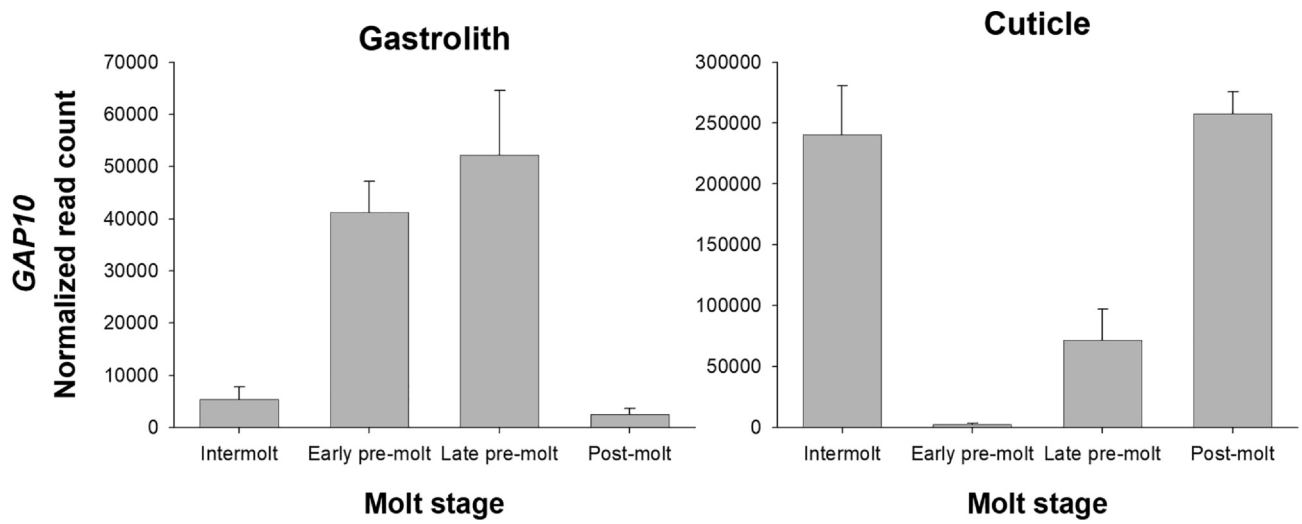


Fig. 3. *GAP10* expression pattern in gastrolith- and cuticle-forming epithelium. Expression levels are presented as normalized read counts in four different molt stages of *C. quadricarinatus*. X-axis represents the four molt stages. Error bars represent standard error.

epithelium was 0110, a pattern common to gastrolith-forming and mineralizing proteins (Fig. 3). In the carapace cuticle-forming epithelium, the *GAP10* expression pattern was 1001 (Fig. 3), which may indicate that GAP10 is also involved in carapace cuticle biomineralization and maintenance. Further *in vivo* and *in vitro* studies are required to confirm this premise.

The current study provides, for the first time, an overview of the possible roles of the proteins first discovered in gastroliths or in the carapace cuticle. Unlike their uniform pattern in gastroliths, *GAMP, GAP65* and *GAP10* differ from each other in their binary expression patterns in the carapace cuticle-forming epithelium, namely, 0011 for *GAMP*, 0001 for *GAP65*, and 1001 for *GAP10*. These patterns indicate that the GAMP, GAP65 and GAP10 proteins previously believed to be involved only in gastrolith formation also play different roles in the formation, mineralization, hardening and maintenance of the carapace cuticle.

3.5. Cuticular proteins involved in biomineralization processes

Following the molt event, gastrolith digestion and hardening of the new cuticle occur simultaneously, with genes related to gastrolith development being downregulated and those related to exoskeleton hardening being upregulated (Glazer and Sagi, 2012). Among the *C. quadricarinatus* cuticular proteins related directly to cuticle hardening via the sedimentation of biominerals, the MARS (mandible rich alanine structural) and CPAP3 (cuticular protein analogous to peritrophins 3) families have recently been discovered (using the binary expression pattern method) and characterized (Abehsera et al., 2015, 2017, 2018b). The new insights acquired by integrating the findings of those studies with the understanding acquired from the new transcriptomic library are presented below.

3.6. The MARS gene family

MARS proteins were found to be involved in the vertical organization of skeletal elements in the incisor tooth of *C. quadricarinatus* (Abehsera et al., 2017). The molar and incisor teeth are cuticular structures that, similar to the carapace cuticle, are replaced during every molt cycle. However, unlike other cuticular elements, the molar tooth is partially mineralized in the pre-molt stage (Tynyakov et al., 2015a) so as to enable the animal to process food immediately after ecdysis.

The MARS family comprises four proteins, each containing a signal peptide, followed by a predicted cleavage site. Within each putative active protein, three motifs consisting of a repetitive set of amino acids were detected. Loss-of-function experiments of two *MARS* family genes, *MARS3* and *MARS4*, showed clear phenotypic differences in dsRNA-injected crayfish compared to control animals. Animals in which these two genes had been silenced developed teeth with surface irregularities and reduced tooth sharpness (Abehsera et al., 2017). The incisors of the silenced animals (injected with *MARS3* and *MARS4*-dsRNA) also presented an altered mineral composition in that the calcium concentration along the incisor's central axis was significantly reduced (Abehsera et al., 2017). These phenotypic changes upon administration of dsRNA suggests that MARS are key cuticular structural proteins required for proper formation of *C. quadricarinatus* incisor tooth.

Binary expression patterns of the *MARS* gene family revealed that the expression pattern in the molar-forming epithelium (0110) differed markedly from that in the carapace cuticle-forming epithelium (0001), which may imply that MARS proteins are also involved in the mineralization of the carapace cuticle (Fig. 4A). However, it should be noted that expression levels in the carapace cuticle were low, so it is possible that the four *MARS* genes do not play a significant role in the carapace cuticle hardening, but this possibility remains to be verified by carrying out further *in vitro* and *in vivo* experiments using functional genomics. An important conclusion that may be drawn from the expression pattern uniformity of the four *MARS* genes in the molar-forming epithelium is that the four constitute a family of genes that act exclusively in the teeth. According to *in silico* data from our new transcriptomic library, the *MARS* gene family is not expressed in the gastrolith-forming epithelium (data not shown), thereby providing support for the premise that these proteins are unique to the teeth of *C. quadricarinatus*. The MARS gene family, first discovered in our study animal is partially conserved from an evolutionary point of view. Three out of four of the genes comprising this family (MARS2-4) were found to be similar to these genes from the red king crab, *Paralithodes camtschaticus* (Fig. 4B). *MARS1* transcripts were found to be most similar to transcripts in *Macrobrachium nipposense* and *Macrobrachium tolmerum*.

3.7. CPAP3 family

The obstructor protein family was discovered first in the fruit fly *Drosophila melanogaster* (Behr and Hoch, 2005) and later in other invertebrates (Colbourne et al., 2005, 2011; Jasrapuria et al., 2012, 2010). Based on the existence of three peritrophinA chitin binding domains in these proteins, the obstructor protein family was subsequently renamed the cuticular protein analogous to peritrophins 3 (CPAP3) family (Jasrapuria et al., 2012). Members of this family were

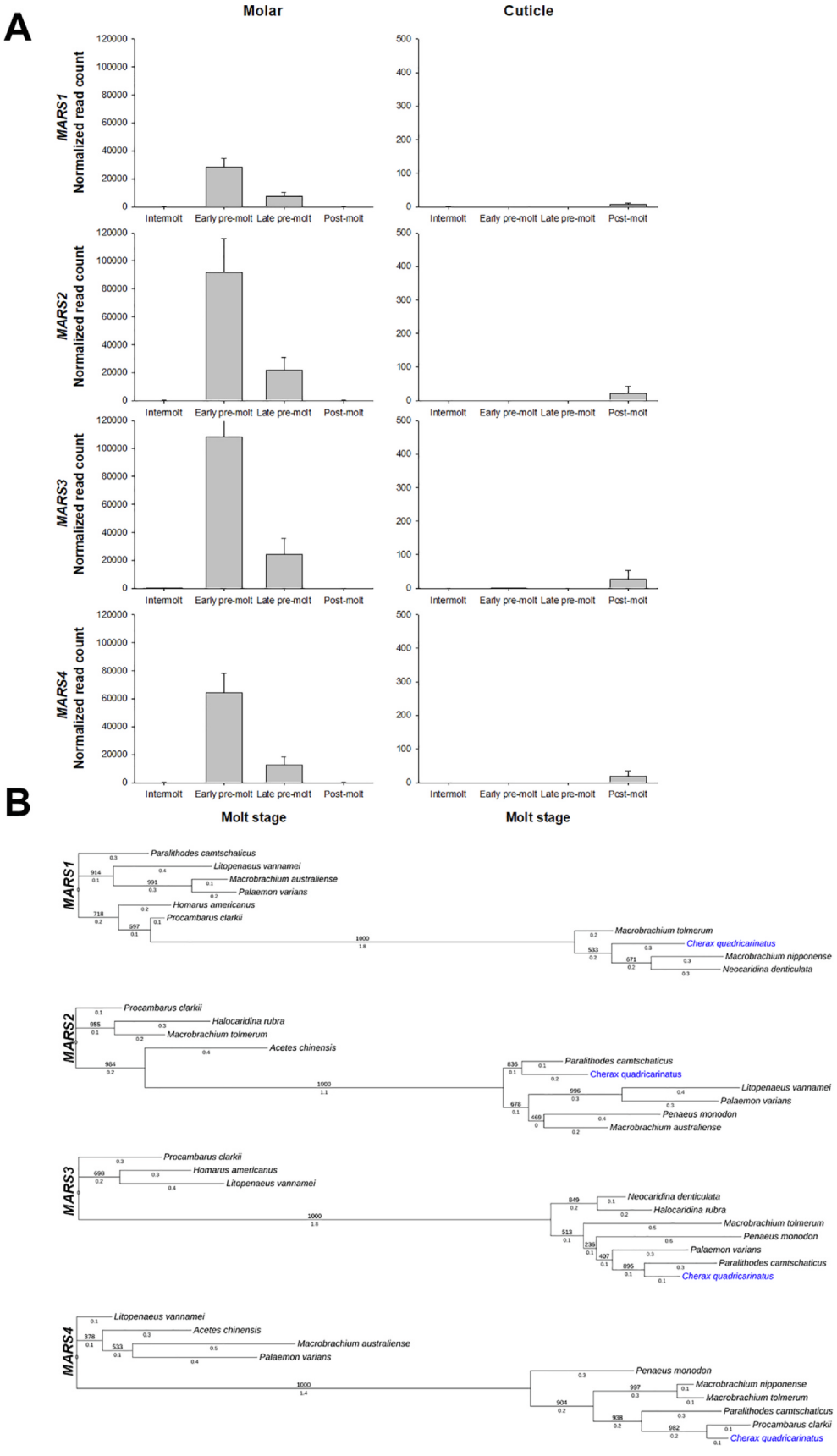


Fig. 4. (A) Expression patterns of four *MARS* genes in molar- and cuticle-forming epithelium. Expression levels are presented as normalized read counts in four different molt stages of *C. quadricarinatus*. X-axis represents the four molt stages. Error bars represent standard error. (B) Phylogenetic trees of four *MARS* gene family transcripts sequences in 10 decapod crustacean species. Supporting values on junctions and branch lengths are given. *C. quadricarinatus* is marked in blue.

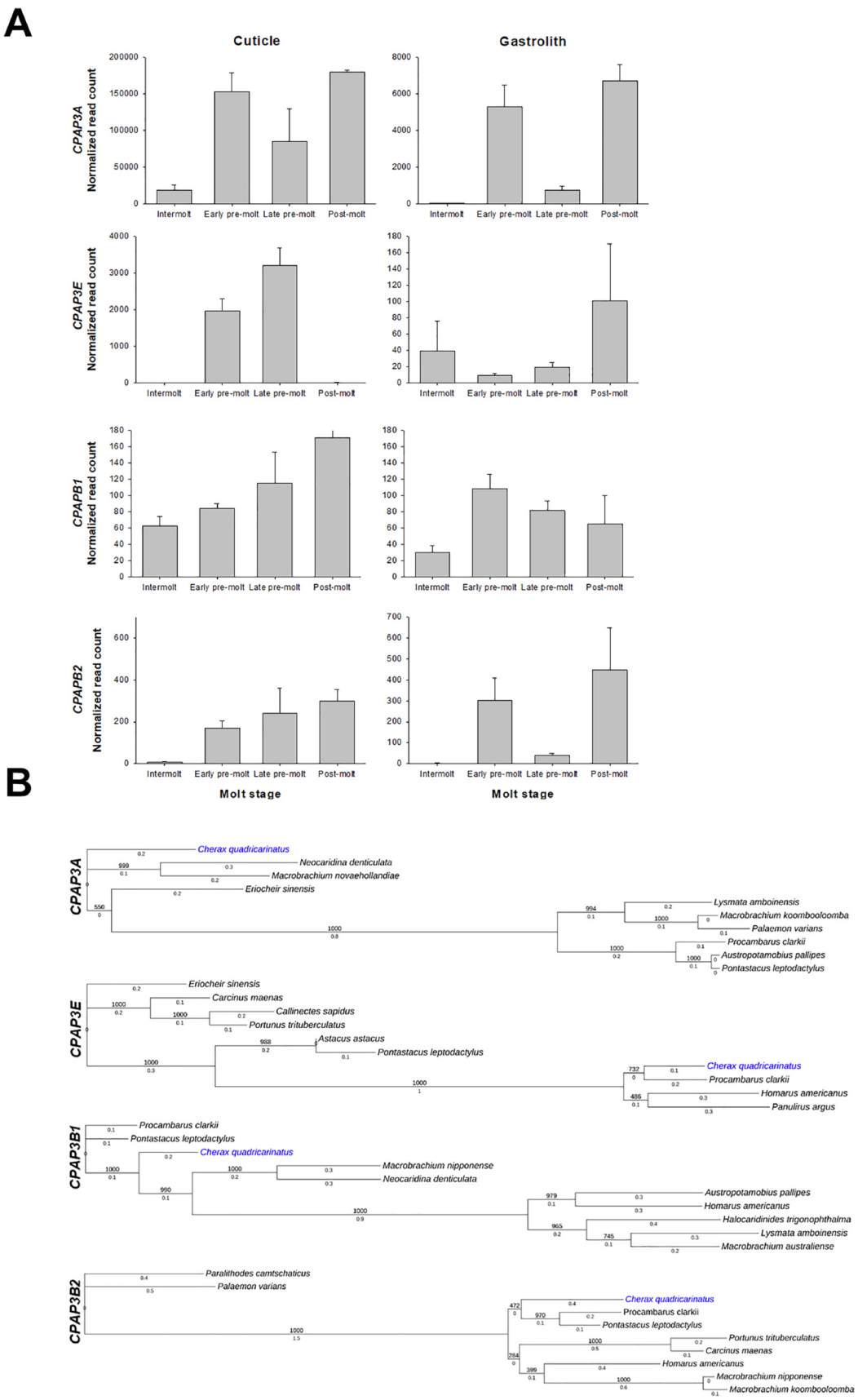


Fig. 5. (A) Expression patterns of *CPAP3* genes in cuticle- and gastrolith-forming epithelium. Expression levels are presented as normalized read counts in four different molt stages of *C. quadricarinatus*. X-axis represents the four molt stages. Error bars represent standard error. (B) Phylogenetic trees of four *CPAP3* gene family transcripts sequences in 10 decapod crustacean species. Supporting values on junctions and branch lengths are given. *C. quadricarinatus* is marked in blue.

found in *C. quadricarinatus* by mining cuticular protein-encoding genes from a transcriptomic library using the binary expression pattern method (Abehsera et al., 2015). The seven CPAP3 proteins found in *C. quadricarinatus* were named according to their similarities in domain organization and to the proteins' relationship with known CPAPs of other organisms (Abehsera et al., 2018b).

Loss-of-function experiments on *CPAP3* genes at post-molt resulted mainly in mortality (Abehsera et al., 2018b), whereas pre-molt silencing led to a significant reduction (by about 40%) in cuticle thickness. This latter result supports the notion that *CPAP3* genes play a major role in the formation of the new cuticle during pre-molt. However, after silencing of these genes, there were no marked changes in the quantity of chitin lamellas or in the length of the carapace, thereby indicating that silencing of *CPAP3* genes does not affect these cuticular structures (Abehsera et al., 2018b). In the current study, expression patterns of four members of the *CPAP3* gene family in the cuticle- and gastrolith-forming epithelia were generated (Fig. 5A). Unlike other gene families described above, the expression patterns of the *CPAP3* genes differed markedly, one from the other. Nonetheless, since RNAi experiments silencing the *CPAP3* genes resulted in lethality and since each gene was upregulated in at least one of the four molt stages, it seems likely that *CPAP3* genes have different roles in chitin-calcium regulation and maintenance. According to our library, *CPAP3A* was the most abundantly expressed of the genes in this family, with normalized read counts of ≈200,000 during post molt, the stage during which the cuticle hardens, and this finding is thus in keeping with the mortality observed in post-molt-silenced animals.

Although the other three *CPAP3* genes exhibit somewhat different patterns with lower expression levels compared to *CPAP3A* (Fig. 5), they also appeared to be crucial players in the molt stages of *C. quadricarinatus*, since their expression levels were significantly higher than the basal level. In the gastrolith-forming epithelium, only *CPAP3A* was differentially expressed. In the carapace cuticle-forming epithelium, the binary expression patterns for each of the genes were different, one from the other, a finding that appears to indicate that each of these proteins bearing a chitin-binding domain plays a distinctive role in carapace cuticle formation and maintenance. Phylogenetic trees of four transcripts from the *CPAP3* gene family described is this article were made (Fig. 5B). *CPAP3B1* and *CPAP3B2* were found to be similarly close to the transcripts of *Pontastcus leptodactylus* and *Procambarus clarkii.* Interestingly, although being part of the lately developed branches in the tree, *CPAP3E* transcript of *C. quadricarinatus* was also found to be similar to the transcript of *P. clarkii.*

3.8. M15 and M13 genes

Two additional protein-encoding genes, designated *Cq-M13* and *Cq-M15*, have been found and characterized in the molar tooth of *C. quadricarinatus* (Tynyakov et al., 2015a, 2015b). These genes belong to a gene family that contains chitin-binding domains, such as chitin-binding domain 2 (CBD2) and Rebers-Riddiford (R-R) consensus sequences. Cq-M13 was first found in a 2D gel protein profile extracted from a *C. quadricarinatus* molar tooth. Later, the full gene transcript was obtained, and a putative signal peptide, followed by an R-R domain and phosphorylation sites, were found in the protein sequence. Cq-M13 was found to be a small protein with a molecular mass of 13-kDa and a structure similar to that of the pro-resilin proteins in arthropod cuticles. Loss-of-function experiments resulted in a molt cycle that was prolonged by almost twofold, with no phenotypic effects (Tynyakov et al., 2015a). Recombinant Cq-M13 was found to bind tightly to chitin, as indicated by the harsh measures required to detach it from chitin (Tynyakov et al., 2015a). The expression pattern of *M13* confirmed the findings described above in that a binary 0110 pattern was found for both the molar- and the carapace cuticle-forming epithelium (Fig. 6A), concomitant with molar tooth formation and hardening. This expression pattern in the carapace cuticle may suggest that M13 plays a role in chitin metabolism in the carapace cuticle, but its relatively low expression level contradicts this notion, and this dichotomy requires further examination. *M13* expression in the gastrolith-forming epithelium was exclusive to late pre-molt stage (data not shown), but its low normalized read count suggests that M13 may not be involved in the formation or hardening of *C. quadricarinatus* gastroliths.

Cq-M15 was also discovered and isolated from the *C. quadricarinatus* molar tooth and was found to be expressed in several exoskeletal tissues (Tynyakov et al., 2015b). Similarly to Cq-M13 and to other arthropod cuticular proteins, this 17-kDa acidic protein contains a signal peptide, the R-R chitin-binding domain, and amino acids that correspond to phosphorylation sites (Tynyakov et al., 2015b). A recombinant Cq-M15 protein was found to bind chitin and to interact with calcium ions in a concentration-dependent manner, enhancing calcium carbonate precipitation *in vitro* over time (Tynyakov et al., 2015b). Functional genomics experiments using RNAi injections to knock down *Cq-M15* expression resulted in mortality. In both the molar- and carapace cuticle-forming epithelium, the *M15* binary expression pattern was 0110, with the highest expression in late pre-molt (Fig. 6B). High expression levels (hundreds of thousands of reads) may imply that M15 is pivotal in the formation and mineralization of cuticular tissues. Moreover, recombinant Cq-M15 interacted with calcium phosphate and calcium carbonate *in vitro*, suggesting that M15 does indeed interact with biominerals (Tynyakov et al., 2015b). Thus, the lethal effect of silencing the *M15* gene in *C. quadricarinatus* can be attributed to the consequent absence of the M15 protein that is a key structural protein required for molting and building of the chitin-protein–mineral complex.

Phylogenetic trees of *M13* and *M15* transcripts from 10 decapod crustacean's species were constructed. While *M13* transcript of *C. quadricarinatus* seems to be one of the first to be developed in decapod crustacean's species (Fig. 6C), *M15* transcript in our study animal seems to have been developed in a late time point in an evolutionary point of view (Fig. 6D). Despite those different time points of development, both transcripts were found to be highly similar to the transcripts of the white leg shrimp, *Litopenaeus vannamei*, suggesting that the sequence similarities between these two protein encoding genes are conserved in other crustacean's species as well.

3.9. Schematic representation of amino acid sequences and predicted proteins structures

Protein sequences of the genes described above were analyzed using SMART (http://smart.embl-heidelberg.de/) for prediction of their structures. This representation shows domains of known function, such as chitin binding domains, which are typical for structural proteins in arthropods. By comparing the schematic images of amino acid sequence for each protein, the similarities between proteins from the same gene families (CPAP3 or MARS) was made clearly visible (Fig. 7). All cuticular structural proteins analyzed in this paper were found to have a signal peptide, and most of them had chitin binding domains, typical of cuticular proteins. GAP65, known to be a structural protein involved in formation and mineralization of cuticular elements, had also chitin binding domain type-2 and low density apolipoprotein domain, related to binding of biominerals (Glazer et al., 2010). In MARS gene family, all genes analyzed had low complexity regions with no known chitin binding domain, which is reasonable since MARS genes are responsible for mineralization of the molar and incisor. However, all proteins from obstructor gene family had chitin binding domains, most of them were chitin binding domain type-2. This strengthens the findings from functional genomics experiments on these gene family which resulted in less thick chitinous scaffolds at the cuticle of *C. quadricarinatus*. M13 and M15 had a signal peptide, and while M13 had no additional known domains, M15 is predicted to have a chitin binding domain type 4. This finding is in keeping with the fact that M15 is highly expressed in the cuticle forming epithelium at early and late premolt stages, when chitin scaffolds which serves as the building blocks of the cuticle are formed

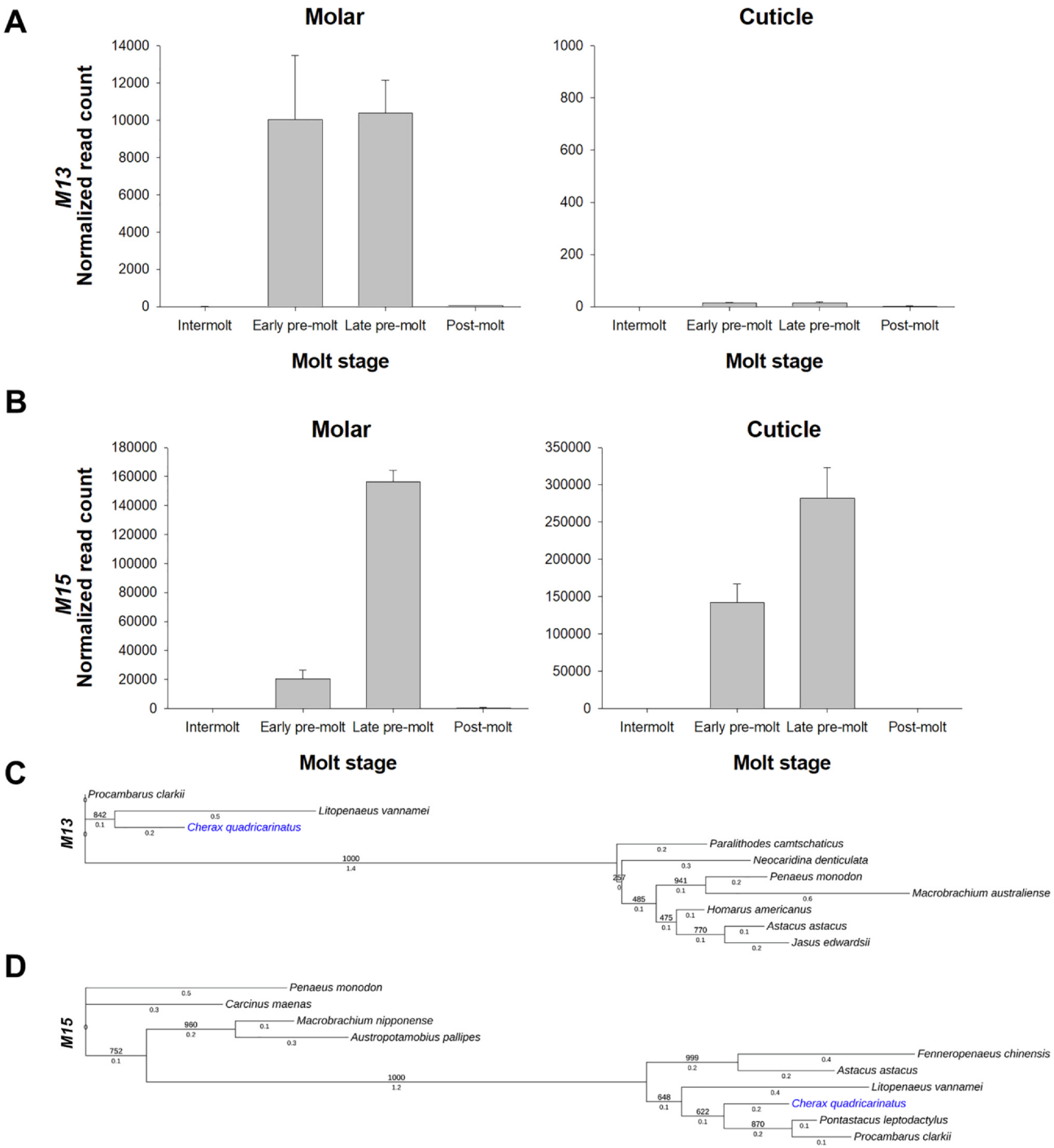


Fig. 6. Expression patterns of M13 (A) and M15 (B) in molar- and cuticle-forming epithelium. Expression levels are presented as normalized read counts in four different molt stages of *C. quadricarinatus*. X-axis represents the four molt stages. Error bars represent standard error. Phylogenetic trees of *M13* (C) and *M15* (D) transcripts sequences in 10 decapod crustacean species. Supporting values on junctions and branch lengths are given.

and organized.

3.10. Reviewing known exoskeleton matrix proteins from other crustacean species

Apart from the proteins discussed up to this point and discovered in our study animal, skeletal proteins were discovered in other crustacean species. Some of these proteins were found to be differentially expressed in the cuticle forming epithelia of *C. quadricarinatus*. This comparison suggests for the first time that the molecular toolkits involved in formation and hardening of chitin scaffolds is evolutionary conserved. In *C. quadricarinatus*, two such matrix protein encoding genes, *CAP-1* and *CASP-2*, were expressed in pattern 0110 related to formation of the chitin scaffolds comprising the cuticle. One protein encoding gene, *CAP-2*, is expressed in pattern 0001 related to mineralization of the cuticle (Fig. 8A, left panel). Those three genes are expressed in low read counts in the gastrolith forming epithelium, suggesting they are probably not pivotal to gastrolith formation or mineralization (Fig. 8A, right panel). Our phylogenetic analysis suggested evolutionary insights behind the above mentioned matrix structural proteins. According to our phylogenetic analysis, *CAP- 1* gene in *C. quadricarinatus* was found to be mostly similar to that of *P.*

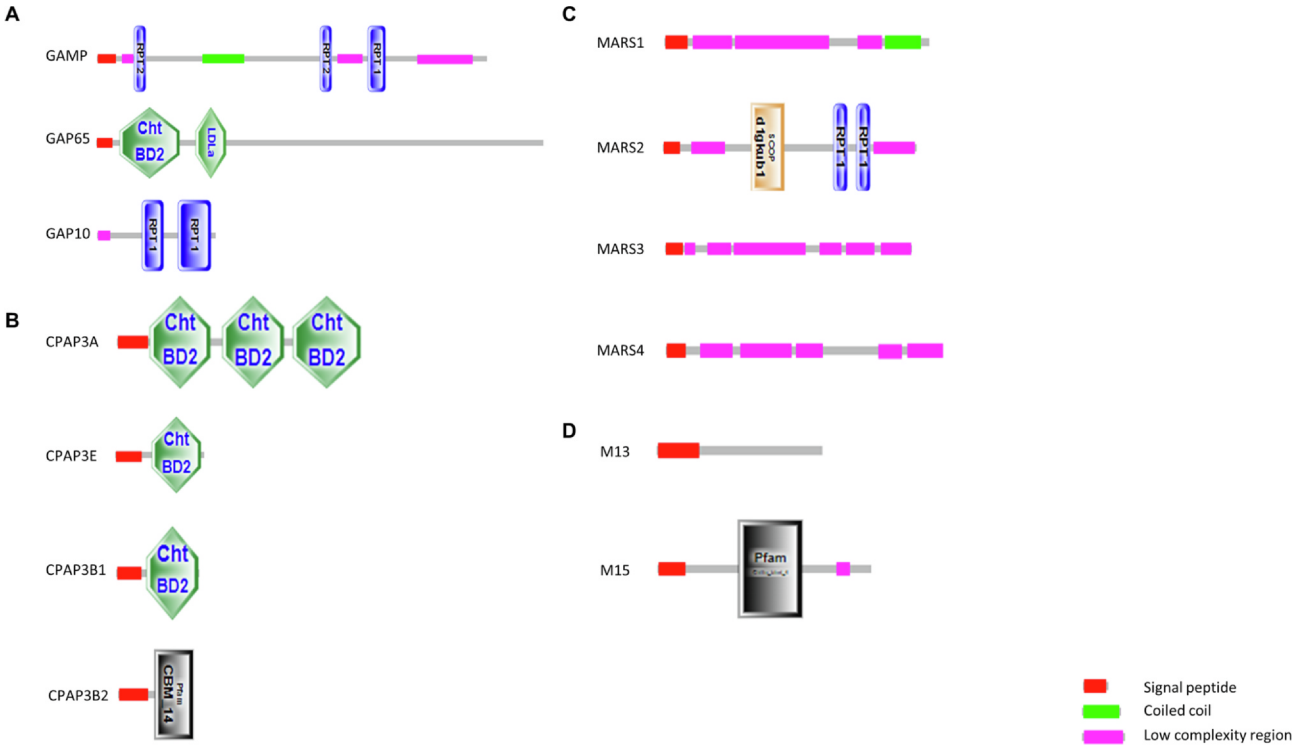


Fig. 7. Schematic representation of amino acid sequences of all structural genes reported in this paper, created using SMART. (A) GAMP, GAP65 and GAP10 schematic images. (B) CPAP3 gene family schematic images. (C) MARS gene family schematic images. (D) M13 and M15 schematic images. Red represents signal peptide, pink represents low complexity region, green represents coiled coil.

camtschaticus (Fig. 8B). Interestingly, the same gene in other crayfish, *P. clarkii*, which is the closest relative to *C. quadricarinatus* in this analysis is found within a distant clade. Similar to *CAP-1*, *C. quadricarinatus CAP-2* was also found to be distant from this gene in *P. clarkii.* On the contrary, *CASP-2* gene in *C. quadricarinatus* was found to be close to its ortholog in *P. clarkii.*

3.11. The broader picture: a multigenic approach discovering differentially expressed genes in the gastrolith and cuticle

Although numerous proteins are involved in the processes of molt and mineralization in decapod crustaceans, and a few genes have already been mined from transcriptomic libraries of several crustacean species (Glazer and Sagi, 2012; Roer et al., 2015), the molecular toolkit that enables these precise and complex processes to take place is far from being fully characterized. Thus, it is reasonable to assume that, other than the single genes and the gene families described above, hundreds to thousands of protein-encoding genes are associated with the different processes of the molt cycle. Comprising relative expression levels in the four major molt stages of *C. quadricarinatus,* our transcriptomic libraries have extended the knowledge base regarding the types and numbers of genes participating in molt-related processes in the gastrolith and carapace cuticle. Five specific binary expression patterns were chosen due to their hypothesized relation to the formation, biomineralization and hardening timelines of cuticular structures. The timelines of the formation and mineralization of the gastrolith and carapace cuticle in *C. quadricarinatus* (Shechter et al., 2008a) were found to be inversely related, and thus these two tissues were chosen for the broad analysis aimed at shedding light upon the transcriptomic background of their formation and mineralization.

Traditionally, the bulk of the fundamental research on biomineralization (Addadi et al., 2003; Weiner et al., 2009) has addressed the structural, physical and chemical properties of biominerals, leaving the molecular background almost unstudied. Towards addressing this lacuna, the current study uses transcript heat maps for the gastroliths and the carapace cuticle of *C. quadricarinatus* to examine the molt cycle from a transcriptomic point of view. Heat maps are a commonly used tool for describing gene expression patterns in cyclic processes, such as cell cycle processes [see, e.g., (Coller et al., 2006; Ly et al., 2014)] and the *C. quadricarinatus* molt cycle. To decipher the transcriptomic background responsible for the formation and mineralization of cuticular structures in *C. quadricarinatus,* heat maps of differentially expressed genes in five molt-mineralization related binary expression patterns of *C. quadricarinatus* were constructed. These heat maps revealed significant numbers of genes, ranging from hundreds to thousands, as hypothesized, with each exhibiting expression patterns typical of the formation or mineralization of cuticular structures in both the cuticle-forming (Fig. 9A) and gastrolith-forming (Fig. 9B) epithelia. The five chosen binary expression patterns (i.e., 0100, 0010, 0110, 0011 and 0001) used to generate the heat maps are directly related to the timeline of exoskeletal element formation and mineralization. The fact that heat maps of the five expression patterns chosen in the carapace cuticle and in the gastroliths include hundreds and sometimes thousands of genes implies that the extent of the molecular toolkit is broader than the single genes revisited above and calls for further study. We note that the multigenic analysis performed in this study does not differentiate between expressed genes related to biomineralization of cuticular tissues and those associated with the formation of these tissues. Such a differentiation would require a thorough study of the identity of genes believed to play a role either in the formation or the mineralization of the gastroliths and cuticle.

In silico data from the molt cycle transcriptomic library was used to calculate the percentage of differentially expressed genes in each of the five binary expression patterns (i.e., 0100, 0010, 0110, 0011 and 0001) out of the total number of all the differentially expressed genes in the cuticle- and gastrolith-forming tissue (Fig. 8C, D, respectively). Out of all the differentially expressed genes in a given type of epithelium (i.e., cuticle or gastrolith), between ~ 1 and ~ 12% were significantly

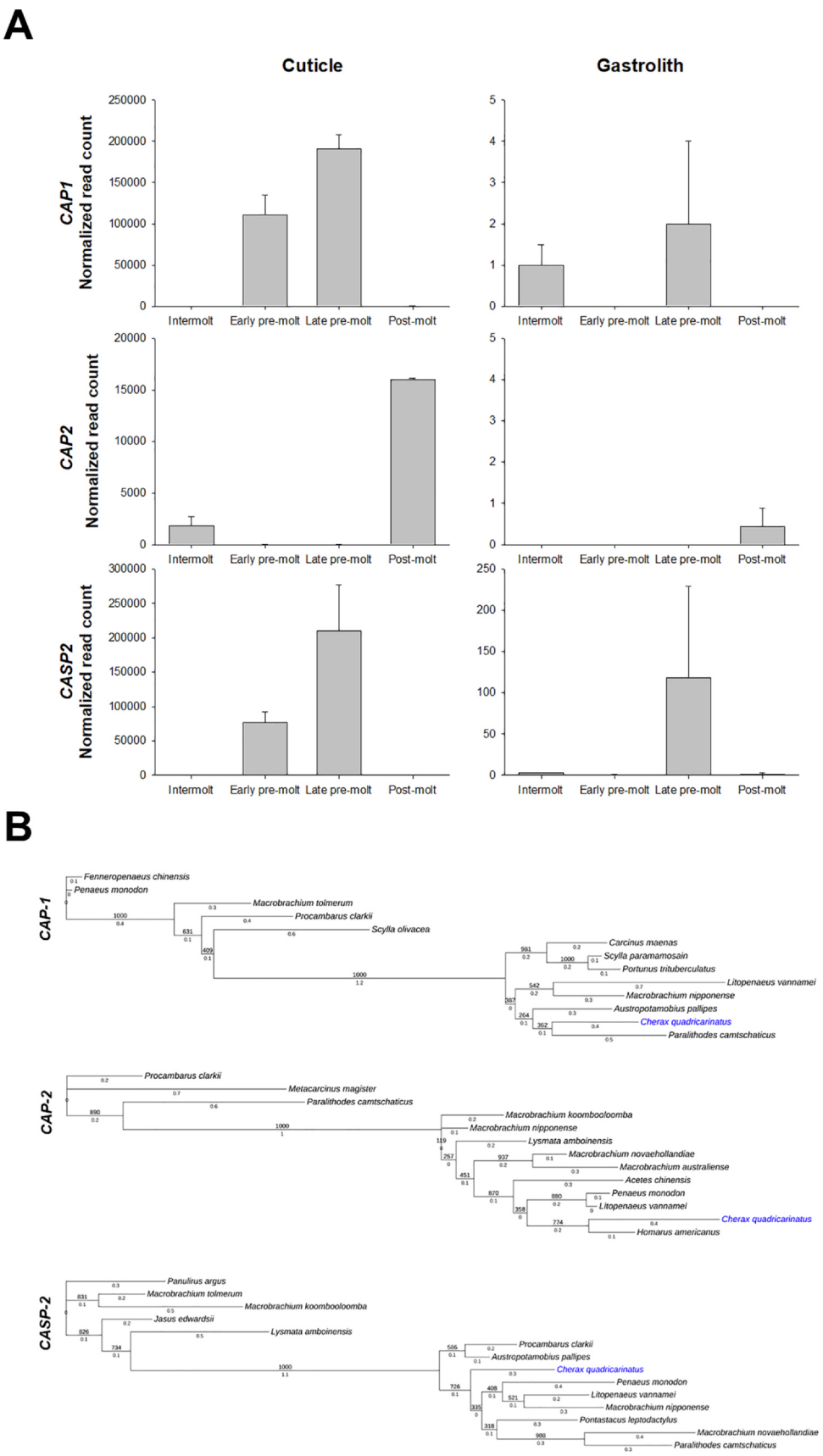


Fig. 8. (A) *CAP-1*, *CAP-2* and *CASP2* expressions pattern in gastrolith- and cuticle-forming epithelium. Expression levels are presented as normalized read counts in four different molt stages of *C. quadricarinatus*. X-axis represents the four molt stages. Error bars represent standard error. (B) Phylogenetic trees of three matrix proteins transcripts: *CAP-1*, *CAP-2* and *CASP-2*. Supporting values on junctions and branch lengths are given. *C. quadricarinatus* is marked in blue.

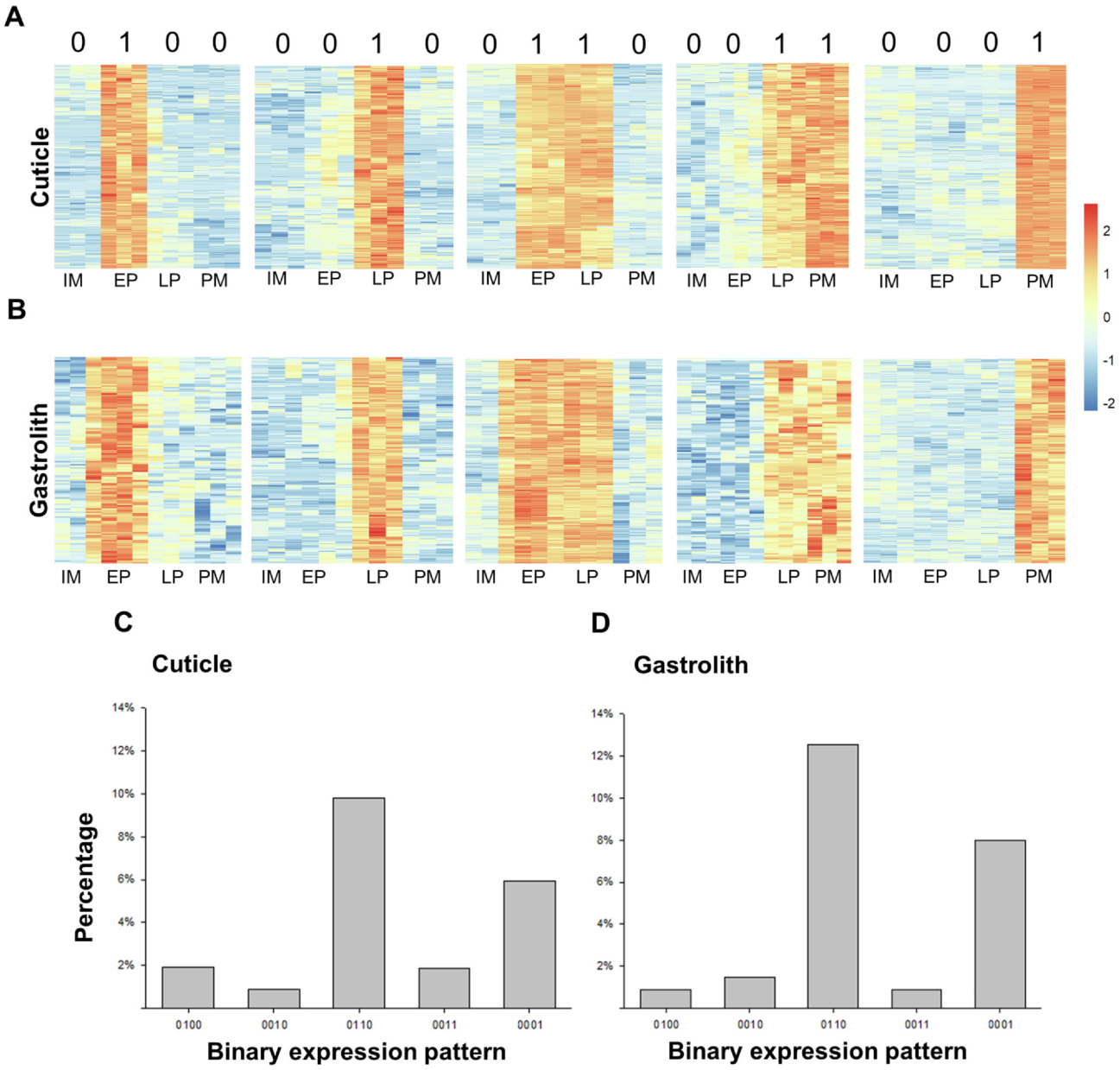


Fig. 9. (A&B) Heat maps of differentially expressed genes in four different molt stages – intermolt (IM), early pre-molt (EP), late pre-molt (LP) and post-molt (PM) – and in different molt-related binary expression patterns in the carapace cuticle-forming epithelium (A) and the gastrolith-forming epithelium (B). Relative expression levels are indicated by red for high values and blue for low values. Scale bar denotes two-fold change in expression level, minus sign denotes down-regulation. (C&D) Percentage of differentially expressed genes out of all differentially expressed genes in (C) the carapace cuticle-forming epithelium (47,631 DE genes in total) and (D) the gastrolith-forming epithelium (14,962 DE genes in total) in specific molt-related binary expression patterns.

expressed in a specific molt-related binary expression pattern in both the carapace cuticle (Fig. 9C) and gastrolith (Fig. 9D). In both types of epithelium, the binary expression pattern 0110, which characterizes genes upregulated throughout the entire pre-molt phase, made up the largest number of differentially expressed genes in each tissue separately (Fig. 9C, D). Indeed, during the two pre-molt stages the gastroliths are formed and mineralized and the new carapace cuticle is formed (Shechter et al., 2008a). These processes involve a massive part of the toolkit, as exemplified by the findings. It could be speculated that formation of the scaffold might involve a similar 'formation toolkit' in both tissues, while a group of biomineralization related genes is uniquely found in the gastrolith-forming epithelium.

The second most abundant binary expression pattern in the heat maps was 0001 (genes upregulated during post-molt), comprising ~ 6% and ~ 8% of the differentially expressed genes in the cuticle and gastrolith, respectively. This expression pattern is expected to be related to exclusively different toolkits in each tissue, based on their different roles during the molt cycle. While the carapace cuticle is mineralized during post molt, the old gastroliths will have already collapsed and undergone digestion in the same period (Shechter et al., 2008a). This synchronization may imply that while differentially expressed genes in the cuticle are involved in biomineralization and hardening during the post molt stage, the upregulated genes in the post molt gastrolith are responsible for organization of the forming epithelium of a new gastrolith for the subsequent molt cycle.

The other three binary expression patterns (0100, 0010 and 0011) were obtained for a relatively small percentage of differentially expressed genes, ranging between ~0.9% and ~2% in the cuticle- and gastrolith-forming epithelium (Fig. 9C, D). Importantly, it should be stressed that this small percentage actually comprises hundreds of genes

that are highly upregulated during a single pre-molt stage or during both late pre-molt and post molt; we note here that the function of these genes is yet to be studied. It should also be noted that in the present study we did not compare the identities of differentially expressed genes in the gastrolith with those in the cuticle-forming epithelium to determine whether these two cuticular elements share genes related to their formation and mineralization albeit at different times in the molt cycle. Such a comparison would be an important first step in shedding light on the complex mechanism of biomineralization in *C. quadricarinatus* in particular and in decapod crustaceans in general.

3.12. Mining candidates with mineralization related expression patterns

To mine proteins related to mineralization of skeletal tissues of *C. quadricarinatus*, a comparison between genes differentially expressed in the cuticle and gastrolith forming epithelia was made. Thus, for the cuticle forming epithelium, mineralization related expression pattern 0001 was chosen. For the gastrolith forming epithelium, expression patterns 0110 and 0010 were chosen. The gastrolith is formed and mineralized concomitantly during the molt cycle of *C. quadricarinatus*, meaning genes expressed in these patterns could be responsible either to formation or mineralization. However, crossing candidate transcripts expressed in such patterns in the gastrolith forming epithelium with transcripts expressed in pattern 0001 in the cuticle narrowed the list down exclusively to protein encoding genes responsible for mineralization in *C. quadricarinatus*.

Candidates were mined from a list of differentially expressed genes which had different binary expression patterns in the cuticle and gastrolith transcriptome (see materials and methods). Venn diagrams of these candidates were made, as following: (1) candidates transcripts expressed in binary expression patterns 0001 and 0110 in the cuticle and gastrolith forming epithelia, respectively, and (2) candidates transcripts expressed in binary expression patterns 0001 and 0010 in the cuticle and gastrolith forming epithelia, respectively. Eleven transcripts were found to be expressed in the three patterns analyzed in the cuticle and gastrolith forming epithelium (Fig. 10). These 11 transcripts are putative candidates for future functional genomics work to characterize their function in mineralization of *C. quadricarinatus* cuticular elements. Candidates are presented in table 1 (Fig. 10) with their binary expression pattern in cuticle and gastrolith forming epithelium and their annotation (if exists) from blastp.

Proteomics characterizations of candidate genes mined from transcriptomic libraries:

Although transcriptomic libraries constructed via RNA-seq are a major tool in understanding fluctuations in gene expression upon developmental stages in vast animal species, it is not the only or ultimate tool for proteins functions assessment. Translation and post-translational modifications may grant proteins with characteristics which cannot be deduced from their sequence only. To cope with this issue, future combination of the binary expression pattern method and laboratory methods to characterize proteins post-translational modification is preferred. Combining these two approaches will allow transcript functions characterization through deciphering the function of the proteins translated. By doing so, new and un-annotated protein encoding genes from the transcriptomic library could be characterized by their function in formation and mineralization of cuticular elements of *C. quadricarinatus*.

3.13. Endocrine regulation of the molt cycle in decapod crustaceans

All ecdysozoans exhibit molt cycles and cuticular growth under strict endocrine regulation (Aguinaldo et al., 1997). It seems that most ecdysone cassette related-genes expressed within the molt cycle are not necessarily proteins found in the cuticular extracellular matrix which serve structural functions (Hyde et al., 2019; Shyamal et al., 2018). Using information from insect species, mostly *Drosophila melanogaster*, an endocrine model for molting in crustaceans was suggested (Hyde et al., 2019) and several studies exploring the endocrine system regulating the molt cycle in crustaceans and in our study animal *C. quadricarinatus* were carried out (Hyde et al., 2019; Pamuru et al., 2012; Shechter et al., 2007). As of today, there is no specific address of direct regulation of ecdysone on transcription of genes encoding cuticular structural proteins. However, it is hypothesized that secretion of ecdysone activates transcription of key genes expressed in cuticular forming epithelia. Thus, the present study enlarging the known transcriptomic database of *C. quadricarinatus*, will allow further research of transcripts related to the ecdysone cassette in decapod crustaceans and discovery of new structural proteins responsible for hardening and mineralization of exoskeletal elements.

4. Summary and concluding remarks

The formation and biomineralization of the crustacean exoskeleton are supported by an extremely complex toolkit, involving thousands of different proteins of distinctive types and with specific roles. The abundance of genes upregulated and highly expressed in a molt-stage-related manner in terms of exoskeleton formation and mineralization implies that a great deal of work is still to be done to fully characterize the entire toolkit that makes biomineralization in crustaceans in general and *C. quadricarinatus* in particular a highly specialized process (Lowenstam and Weiner, 1989). To date, the complex biomineralization mechanism has been studied mainly from a traditional standpoint focusing on the physical and chemical properties of biominerals (Addadi et al., 2003; Weiner et al., 2009), but a comprehensive understanding of the processes involved can be reached only by unraveling the details of the molecular toolkit.

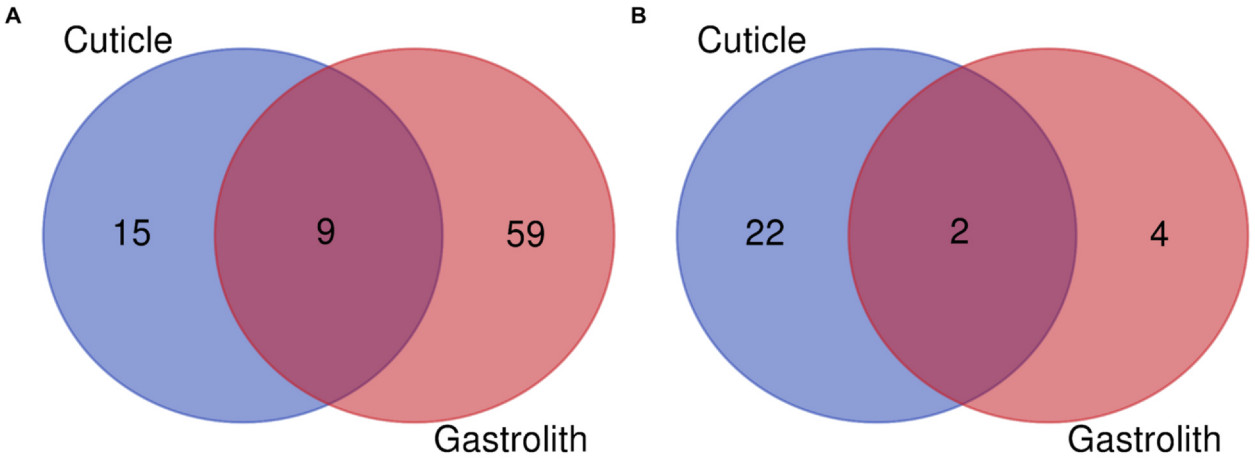
Towards addressing the above goal, this article presents, for the first time, a broad overview of decapod crustacean biomineralization from a transcriptomic point of view. Transcriptomic libraries such as those described in this article are prerequisite for further understanding these complex processes. We started the article by revisiting the few known protein-encoding genes related to exocuticular formation and hardening in crustaceans and thereafter explored a much broader transcriptomic approach confirming that the biomineralization toolkit in *C. quadricarinatus* involves hundreds or even thousands of genes with similar molt-related expression patterns. Based on their molt-related expression patterns, these genes could be sorted into functional groups that distinguish between genes involved in forming the chitinous scaffold and those involved in the biomineralization and hardening of the scaffold. Within these two groups, proteins that play a structural role in the scaffold could be identified and distinguished from others that play an enzymatic or regulatory role. In order to fully understand biomineralization processes, *in vitro* and *in silico* studies will not be sufficient, and functional genomics will be needed. Success in characterizing and deciphering the molt and mineralization mechanism of *C. quadricarinatus* will greatly contribute not only to the field of structural biology, but also to diverse biomimetic efforts for future industrial and biomedical uses.

Author contributions

SAH and AS conceived the study. SA and TL constructed the RNA library. VC performed the bioinformatics. The manuscript was written by SAH and approved by all co-authors.

Funding

This research was supported by the Israel Science Foundation (grant No. 354/18).



Gene (cuticle)	blastp_RecName	Binary expression pattern (cuticle)	Binary expression pattern (gastrolith)
DN124641_c1_g1		0001	0110
DN125631_c3_g1	Bumetanide-sensitive sodium-(potassium)-chloride cotransporter;	0001	0110
DN126347_c4_g1	Sodium-dependent phosphate transporter 1;	0001	0110
DN132678_c2_g1	Protein draper ;	0001	0110
DN135284_c2_g2	Anion exchange protein 2;	0001	0110
DN136084_c6_g1		0001	0110
DN136353_c3_g1	5'-AMP-activated protein kinase subunit gamma-2;	0001	0110
DN136574_c4_g2		0001	0110
DN136782_c3_g2		0001	0110
DN122046_c13_g1	Transcriptional regulator ovo ;	0001	0010
DN124582_c6_g1		0001	0010

Fig. 10. Venn diagrams and list of differentially expressed genes in cuticle and gastrolith forming epithelia. (A) Genes expressed in binary expression patterns 0001 and 0110 in cuticle and gastrolith, respectively. (B) Genes expressed in binary expression patterns 0001 and 0010 in cuticle and gastrolith, respectively. Table: list of genes from Venn diagrams expressed in pattern 0001 in the cuticle forming epithelium and in patterns 0110 or 0010 in the gastrolith forming epithelium.

CRediT authorship contribution statement

Shai A. Shaked: Conceptualization, Writing - original draft, Writing - review & editing, Visualization, Data curation. **Shai Abehsera:** Investigation, Resources. **Tom Levy:** Investigation, Resources, Visualization. **Vered Chalifa-Caspi:** Formal analysis, Data curation. **Amir Sagi:** Conceptualization, Writing - original draft, Writing - review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr. Menachem Sklarz and Dr. Olabiyi Obayomi from Ben-Gurion University Bioinformatics Core Facility for bioinformatics analysis. The Dor aquaculture station team are thanked for the animal's supply that made this research possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2020.107612.

References

Abehsera, S., Weil, S., Manor, R., Sagi, A., 2018a. The search for proteins involved in the formation of crustacean cuticular structures. Hydrobiologia 825, 29–45.

Abehsera, S., Glazer, L., Tynyakov, J., Plaschkes, I., Chalifa-Caspi, V., Khalaila, I., Aflalo, E.D., Sagi, A., 2015. Binary gene expression patterning of the Molt cycle: the case of Chitin metabolism, Plos One 10.

Abehsera, S., Peles, S., Tynyakov, J., Bentov, S., Aflalo, E.D., Li, S.H., Li, F.H., Xiang, J.H., Sagi, A., 2017. MARS: a protein family involved in the formation of vertical skeletal elements. J. Struct. Biol. 198, 92–102.

Abehsera, S., Zaccai, S., Mittelman, B., Glazer, L., Weil, S., Khalaila, I., Davidov, G., Bitton, R., Zarivach, R., Li, S., Li, F., Xiang, J., Manor, R., Aflalo, E.D., Sagi, A., 2018b. CPAP3 proteins in the mineralized cuticle of a decapod crustacean. Sci. Rep. 8.

Addadi, L., Raz, S., Weiner, S., 2003. Taking advantage of disorder: amorphous calcium carbonate and its roles in biomineralization. Adv. Mater. 15, 959–970.

Aguinaldo, A.M.A., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., Lake, J.A., 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387, 489–493.

Behr, M., Hoch, M., 2005. Identification of the novel evolutionary conserved obstructor multigene family in invertebrates. FEBS Lett. 579, 6827–6833.

Bentov, S., Abehsera, S., Sagi, A., 2016a. The mineralized exoskeletons of crustaceans. In: Extracellular Composite Matrices in Arthropods. Springer, pp. 137–163.

Bentov, S., Weil, S., Glazer, L., Sagi, A., Berman, A., 2010. Stabilization of amorphous calcium carbonate by phosphate rich organic matrix proteins and by single phosphoamino acids. J. Struct. Biol. 171, 207–215.

Bentov, S., Aflalo, E.D., Tynyakov, J., Glazer, L., Sagi, A., 2016b. Calcium phosphate mineralization is widely applied in crustacean mandibles. Sci Rep-Uk 6.

Colbourne, J.K., Singan, V.R., Gilbert, D.G., 2005. wFleaBase: The Daphnia genome database. BMC Bioinf. 6.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Caceres, C.E., Carmel, L., Casola, C., Choi, J.H., Detter, J.C., Dong, Q.F., Dusheyko, S., Eads, B.D., Frohlich, T., Geiler-Samerotte, K.A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E.V., Kultz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S., Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z.J., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y.I., Yamagata, H., Yamada, T., Ye, Y.Z., Shaw, J.R., Andrews, J., Crease, T.J., Tang, H.X., Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E.V., Zdobnov, E.M., Grigoriev, I.V., Lynch, M., Boore, J.L., 2011. The Ecoresponsive Genome of Daphnia pulex. Science 331, 555–561.

Coller, H.A., Sang, L., Roberts, J.M., 2006. A new description of cellular quiescence. PLoS Biol. 4, e83.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucl. Acids Res. 31, 3784–3788.

Glazer, L., Sagi, A., 2012. On the involvement of proteins in the assembly of the crayfish gastrolith extracellular matrix. Invertebr Reprod. Dev. 56, 57–65.

Glazer, L., Roth, Z., Weil, S., Aflalo, E.D., Khalaila, I., Sagi, A., 2015. Proteomic analysis of the crayfish gastrolith chitinous extracellular matrix reveals putative protein

complexes and a central role for GAP 65. J. Proteomics 128, 333–343.
Glazer, L., Shechter, A., Tom, M., Yudkovski, Y., Weil, S., Aflalo, E.D., Pamuru, R.R., Khalaila, I., Bentov, S., Berman, A., Sagi, A., 2010. A protein involved in the assembly of an extracellular calcium storage matrix. J. Biol. Chem. 285, 12831–12839.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.
Huxley, T., 1880. The crayfish, 371 pp. London: Kregan Paul.
Hyde, C.J., Elizur, A., Ventura, T., 2019. The crustacean ecdysone cassette: a gatekeeper for molt and metamorphosis. J. Steroid Biochem. 185, 172–183.
Ishii, K., Yanagisawa, T., Nagasawa, H., 1996. Characterization of a matrix protein in the gastroliths of the crayfish Procambarus clarkii. Biosci. Biotech. Bioch. 60, 1479–1482.
Ishii, K., Tsutsui, N., Watanabe, T., Yanagisawa, T., Nagasawa, H., 1998. Solubilization and chemical characterization of an insoluble matrix protein in the gastroliths of a crayfish, Procambarus clarkii. Biosci. Biotech. Bioch. 62, 291–296.
Jasrapuria, S., Specht, C.A., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., 2012. Gene families of cuticular proteins analogous to peritrophins (CPAPs) in Tribolium castaneum have diverse functions. PLoS ONE 7.
Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., 2010. Genes encoding proteins with peritrophin A-type chitin-binding domains in Tribolium castaneum are grouped into three distinct families based on phylogeny, expression and function. Insect Biochem Molec 40, 214–227.
Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20, 1160–1166.
Kolde, R.J.R.p.v., 2012. Pheatmap: pretty heatmaps 61, 617.
Kuballa, A.V., Merritt, D.J., Elizur, A., 2007. Gene expression profiling of cuticular proteins across the moult cycle of the crab Portunus pelagicus. BMC Biol. 5.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–U354.
Lefort, V., Longueville, J.E., Gascuel, O., 2017. SMS: smart model selection in PhyML. Mol. Biol. Evol. 34, 2422–2424.
Letunic, I., Bork, P., 2018. 20 years of the SMART protein domain annotation resource. Nucl. Acids Res. 46, D493–D496.
Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucl. Acids Res. 47, W256–W259.
Levy, T., Rosen, O., Eilam, B., Azulay, D., Aflalo, E.D., Manor, R., Shechter, A., Sagi, A., 2016. A single injection of Hypertrophied Androgenic gland cells produces All-Female aquaculture. Mar. Biotechnol. 18, 554–563.
Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12, 323.
Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv arXiv: 1303.3997.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.
Lowenstam, H.A., Weiner, S., 1989. On biomineralization Oxford University Press on Demand.
Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M.J., Stratton, M.R., Lamond, A.I., 2014. A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. Elife 3, e01630.
Murdock, D.J.E., 2020. The 'biomineralization toolkit' and the origin of animal skeletons. Biol. Rev. Camb. Philos. Soc.
Pamuru, R.R., Rosen, O., Manor, R., Chung, J.S., Zmora, N., Glazer, L., Aflalo, E.D., Weil, S., Tamone, S.L., Sagi, A., 2012. Stimulation of molt by RNA interference of the molt-inhibiting hormone in the crayfish Cherax quadricarinatus. Gen Comp Endocr 178, 227–236.
Roer, R., Abehsera, S., Sagi, A., 2015. Exoskeletons across the Pancrustacea: comparative morphology, physiology, biochemistry and genetics. Integr. Comp. Biol. 55, 771–791.
Shechter, A., Berman, A., Singer, A., Freiman, A., Grinstein, M., Erez, J., Aflalo, E.D., Sagi, A., 2008a. Reciprocal changes in calcification of the gastrolith and cuticle during the molt cycle of the red claw crayfish Cherax quadricarinatus. Biol. Bull.-Us 214, 122–134.
Shechter, A., Tom, M., Yudkovski, Y., Weil, S., Chang, S.A., Chang, E.S., Chalifa-Caspi, V., Berman, A., Sagi, A., 2007. Search for hepatopancreatic ecdysteroid-responsive genes during the crayfish molt cycle: from a single gene to multigenicity. J. Exp. Biol. 210, 3525–3537.
Shechter, A., Glazer, L., Cheled, S., Mor, E., Weil, S., Berman, A., Bentov, S., Aflalo, E.D., Khalaila, I., Sagi, A., 2008b. A gastrolith protein serving a dual role in the formation of an amorphous mineral containing extracellular matrix. Proc. Natl. Acad. Sci. U.S.A. 105, 7129–7134.
Shyamal, S., Das, S., Guruacharya, A., Mykles, D.L., Durica, D.S., 2018. Transcriptomic analysis of crustacean molting gland (Y-organ) regulation via the mTOR signaling pathway. Sci. Rep.-Uk 8.
Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.
Sklarz, M., Levin, L., Gordon, M., Chalifa-Caspi, V., 2018. NeatSeq-Flow: A Lightweight High-Throughput Sequencing Workflow Platform for Non-Programmers and Programmers Alike. bioRxiv https://www.biorxiv.org/content/10.1101/173005v3.
Tsutsui, N., Ishii, K., Takagi, Y., Watanabe, T., Nagasawa, H., 1999. Cloning and expression of a cDNA encoding an insoluble matrix protein in the gastroliths of a crayfish, Procambarus clarkii. Zool. Sci. 16, 619–628.
Tynyakov, J., Bentov, S., Abehsera, S., Khalaila, I., Manor, R., Abilevich, L.K., Weil, S., Aflalo, E.D., Sagi, A., 2015a. A novel chitin binding crayfish molar tooth protein with elasticity properties. PLoS ONE 10.
Tynyakov, J., Bentov, S., Abehsera, S., Yehezkel, G., Roth, Z., Khalaila, I., Weil, S., Berman, A., Plaschkes, I., Tom, M., Aflalo, E.D., Sagi, A., 2015b. A crayfish molar tooth protein with putative mineralized exoskeletal chitinous matrix properties. J. Exp. Biol. 218, 3487–3498.
Watanabe, T., Persson, P., Endo, H., Kono, M., 2000. Molecular analysis of two genes, DD9A and B, which are expressed during the postmolt stage in the decapod crustacean Penaeus japonicus. Comp. Biochem. Phys. B 125, 127–136.
Weiner, S., Mahamid, J., Politi, Y., Ma, Y., Addadi, L., 2009. Overview of the amorphous precursor phase strategy in biomineralization. Front. Mater. Sci. Chin. 3, 104.