



OPEN

Transcriptome analysis of the bivalve *Placuna placenta* mantle reveals potential biomineralization-related genes

Ningjing Song, Jiangfeng Li, Baosheng Li, Ercai Pan & Yurong Ma✉

The shells of window pane oyster *Placuna placenta* are very thin and exhibit excellent optical transparency and mechanical robustness. However, little is known about the biomineralization-related proteins of the shells of *P. placenta*. In this work, we report the comprehensive transcriptome of the mantle tissue of *P. placenta* for the first time. The unigenes of the mantle tissue of *P. placenta* were annotated by using the public databases such as nr, GO, KOG, KEGG, and Pfam. 24,343 unigenes were annotated according to Pfam database, accounting for 21.48% of the total unigenes. We find that half of the annotated unigenes of the mantle tissue of *P. placenta* are consistent to the annotated unigenes from pacific oyster *Crassostrea gigas* according to nr database. The unigene sequence analysis from the mantle tissue of *P. placenta* indicates that 465,392 potential single nucleotide polymorphisms (SNPs) and 62,103 potential indel markers were identified from 60,371 unigenes. 178 unigenes of the mantle tissue of *P. placenta* are found to be homologous to those reported proteins related to the biomineralization process of molluscan shells, while 18 of them are highly expressed unigenes in the mantle tissue. It is proposed that four unigenes with the highest expression levels in the mantle tissue are very often related to the biomineralization process, while another three unigenes are potentially related to the biomineralization process according to the Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) analysis. In summary, the transcriptome analysis of the mantle tissue of *P. Placenta* shows the potential biomineralization-related proteins and this work may shed light for the shell formation mechanism of bivalves.

Abbreviations

qRT-PCR	Quantitative real-time PCR
SSR	Simple sequence repeat
SNP	Single nucleotide polymorphisms
Indel	Insertion–deletion mutations
KOG	EuKaryotic Ortholog Groups
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
CDSs	Coding regions
nr	NCBI non-redundant protein sequences
Pfam	Protein family
SMPs	Shell matrix proteins
VWA	Von willebrand factor type A
FPKM	Fragments per kilobase of feature per million mapped reads

The biological exoskeletons are usually composed of highly ordered hierarchical micro- and nanostructures occluded with organic molecules and exhibit superior mechanical, optical, thermal, magnetic properties^{1–3}. Calcium carbonate is one of the most abundant biominerals in nature, and the biomineralization and bioinspired mineralization of calcium carbonate have been the focus of investigation for many decades. The molluscan shells have well defined micro- and nanostructures and excellent mechanical properties, composed of 95% CaCO₃

School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing 100081, China. ✉email: yurong.ma@bit.edu.cn

and less than 5 wt % of organic macromolecules (proteins, glycoproteins, polysaccharides and lipids)^{4,5}. Generally, the polymorphs of CaCO₃ include aragonite, calcite, vaterite, calcite monohydrate and calcium carbonate hemihydrate⁶. Biogenic CaCO₃ exhibits complex microstructures such as prismatic, nacreous, foliate, and cross-lamellar microstructures^{7,8}. Importantly, protein components in the molluscan shells are responsible for the nucleation and crystal growth of biogenic CaCO₃ and can enhance the mechanical properties of the shells⁹. The matrix proteins occluded in the molluscan shells were extracted, sequenced and identified as biomineralization-related proteins which were found to have functions including stabilization of amorphous calcium carbonate, inhibition or acceleration of nucleation of calcium carbonate^{10–13}. A few examples of biomineralization-related proteins include mantle protein N25¹⁴, perlucin, perlustrin¹⁵, aspein¹⁶, prisilkin-39¹⁷, Shematrixin-2¹⁸, Pfn44¹⁹, and enzymes such as carbonic anhydrase²⁰. Most of matrix proteins in molluscan shells are specifically secreted by the mantle tissue of molluscs. In another word, the proteins from the mantle tissues of molluscs can tune the formation of different structural layers of the shells^{12,21,22}. It was considered that the shell matrix proteins extracted from skeletons or shells generally have high expression in mantle tissues. Identification and characterization of the proteins that are involved in the biomineralization process are critically important for understanding formation mechanisms of shells. Biomineralization-related proteins have been reported from the bivalve species such as perlucin from abalone *Haliotis laevis*²³, nacrein from pearl oyster *Pinctada fucata*²⁴, and aspartic acid rich proteins from *Atrina rigida*²⁵.

P. placenta, also known as window pane oyster, has two highly transparent, disc-shaped flat mineralized valves. Local people used to apply the shells of *P. placenta* as a substitute for window glass in houses in the old days. *P. placenta* live at the coasts of East China Sea and South China Sea. *P. placenta* are often found in sandy or muddy substrates in shallow estuarine lagoons and small bays. The inner surface of the *P. placenta* shell has a V-shaped ligament. The right valve is almost flat while the left valve is slightly convex²⁶. The *P. placenta* shell is composed of foliated calcite single crystals with thickness of about 300 nm. The *c* axis of these calcific laths is tilted about 24.4° ± 3.5° from the normal direction of the shell surface. Thus, the surfaces of the laths are close to the {108} planes of calcite^{27,28}. In addition, the shell of the *P. placenta* possesses unique transparent optical property, which can transmit ~80% of visible light. The hardness of the shell of the *P. placenta* increases ~55% in comparison to that of single-crystal calcite^{29,30}.

It was considered that the shell matrix proteins extracted from skeletons or shells generally have high expression in mantle tissues. The transcriptome is the complete set of transcripts in the tissues of organisms, which contains important information on gene expression and contributes to understand the functional components of the transcriptome and molecular constituents of cells and tissues³¹. Therefore, the study of molluscan mantle transcriptome is important for the identification of biomineralization related proteins, which can provide further understanding for the shell formation mechanism. Mantle transcriptome studies from different molluscan shell systems such as trochidae *Tectus pyramis*³², freshwater pearl mussel *Cristaria plicata*³³, pearl oyster *Pteria penguin*³⁴, *Pinctada martensii*³⁵, scallop *Patinoptecten yessoensis*³⁶ and *Chlamys farreri*³⁷ have been investigated for the purpose of identifying biomineralization related genes. However, high-throughput sequencing and the transcriptome-based identification of potential biomineralization unigenes of the mantle tissue of *P. placenta* transcriptome have not been reported, as far as we know.

Recently, a new generation efficient sequencing technology has been developed for transcriptome study by using Illumina platform. In comparison to traditional methods, Illumina sequencing has advantages such as low cost, high-throughput, and rapid sequence generation. Thus it has been applied as an effective way to identify potential unigenes or pathways involved in the processes of biomineralization in molluscs^{37–39}.

Herein, we report the mantle transcriptome for *P. placenta* by Illumina-based RNA-Seq technology for the first time. Dataset of the transcripts were annotated by using five different public databases. Potential single nucleotide polymorphisms (SNP), indel markers and simple sequence repeat (SSR) markers were analyzed based on the transcriptome dataset of the mantle tissue of *P. placenta*. Unigene sequences, pfam analysis and qRT-PCR analysis were done for the highly expressed unigenes of the mantle tissue and several potential genes related to biomineralization were identified.

Materials and methods

Animals. Adult wild *P. placenta* were two years old and collected during July to August (2020) from Beibu Gulf (Fangchenggang city, Guangxi province, China). The shells were almost-flat and had diameters in between 50 and 70 mm. Different tissues such as mantle, adduction muscle, gonad, gill, hepatopancreas, mouthparts, and intestine were dissected from fresh adult wild *P. placenta* and were frozen immediately in liquid nitrogen, then stored at -80 °C for further analysis. All animal experiments were done in accordance with the guidelines and approval of the Animal Research and Ethics Committees of the Chinese Academy of Sciences.

Materials. RNA Keeper-ICE tissue transfer buffer was bought from vazyme company (Nanjing, China). Quick RNA Isolation kit was bought from Biotech Corporation (Beijing, China). AMPure XP beads were bought from Beckman (CA, USA). HiScript® II Q RT SuperMix reagent was bought from vazyme company (Nanjing, China). TRIzol reagent was bought from TransGen Biotech (Beijing, China). Divalent cation fragmentation buffer was bought from Illumina (CA, USA). RNaseH and DNA polymerase I were bought from TaKaRa (Kyoto, Japan). Elution buffer was bought from GENEWIZ (Suzhou, China). Agarose gel electrophoresis buffer was obtained by dissolving 1 g of agarose in 100 mL of Tris Acetate EDTA electrophoresis buffer. Tris Acetate EDTA electrophoresis buffer was bought from TSINGKE (Beijing, China). Agarose was bought from Sigma (Darmstadt, Germany).

Total RNA extraction, cDNA library preparation and Illumina sequencing. Total RNA was extracted from the mantle tissue of *P. placenta* by using the Quick RNA Isolation Kit (Biotech Corporation, Beijing, China). The concentrations of RNA were quantified with a Qubit 2.0 Fluorometer (Invitrogen, Life Technologies, CA, USA), and the RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The weight of the mantle tissue for RNA extraction was no less than 300 mg for each sample and the experiments were repeated for three times.

The cloned DNA (cDNA) library construction and sequencing analysis were performed at Beijing BioMarker Technologies (Beijing, China). The cDNA library was constructed as follows. In the case of eukaryotes, polyadenylation (poly (A)) tail messenger RNA (mRNA) was purified with oligonucleotides (dT) magnetic beads from total RNA, and then the mRNA-enriched RNA was randomly segmented into small fragments in a divalent cation fragmentation buffer (Illumina, Hayward, CA). These short fragments were used as templates to synthesize the first strand cDNA using random hexamer primers. The second strand cDNA was generated using RNaseH and DNA polymerase I. The double stranded cDNA was purified by beads (Beckman, CA, USA) and washed with elution buffer for end repairing and tailing A. Afterwards, these short fragments were ligated to sequencing adapters according to Illumina's protocol (San Diego, CA, USA) and were then separated by agarose gel electrophoresis.

In order to obtain the cDNA library, suitable fragments (300–500 bp) were selected as templates for PCR amplification. The effective concentration of the cDNA library (> 4nM) was accurately quantified by the quantitative real-time PCR (qRT-PCR) method. Finally, the cDNA libraries were sequenced on an Illumina HiSeq™ 4000.

De novo assembly and sequence annotation. Before de novo assembly, raw data were saved in fastq format and filtered for downstream analysis using in-house pear script, which mainly contains three filter steps: (1) discard sequencing adapter; (2) remove reads with more than 10% of unknown bases; (3) filter low-quality reads which contain more than 50% of low-quality bases ($Q \leq 20$). After the above filtering processes, de novo assembly was carried out using Trinity with `min_kmer_cov` set to 3 by default and all other parameters set to their default values⁴⁰.

The coding regions of the assembled unigenes also known as the coding DNA sequence (CDS), were predicted by two steps. First, the unigene sequences from the mantle tissues of *P. placenta* were successively translated and compared to the NCBI non-redundant protein sequences (nr database) and a manually annotated and reviewed protein sequences (Swiss-Prot) database. Then, the Basic Local Alignment Search Tool (BLAST) analysis was used to annotate the functions of the protein sequences based on nr and Swiss-Prot databases with a cut-off E-value of 10^{-5} . The protein sequences are actually amino acid sequences of proteins. These two proteins database have correct sequence to predict the coding regions of the unigenes. Second, The CDSs were determined using ESTScan with default settings if the unigenes could not be aligned to any entry of the above mentioned two databases⁴¹. ESTScan software was used to determine the open reading frame and obtain the nucleic acid and amino acid sequences from the unannotated unigenes. Finally, all annotated unigenes were aligned to the public databases, including the euKaryotic Ortholog Groups (KOG, <http://www.ncbi.nlm.nih.gov/COG/>), Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>), Gene Ontology (GO, <http://www.geneontology.org/>) and Protein family (Pfam, <http://pfam.sanger.ac.uk/>) databases.

SNP, Indel, SSR markers identification and primer design. Single nucleotide polymorphisms (SNP) refers to genetic markers formed by mutations of the single nucleotide on the genome. There are a large number of SNPs with abundant polymorphisms. Generally, SNP is defined as single nucleotide mutations with a mutation frequency more than 1%. Indel refers to the insertion and deletion of small fragments in the sample genome relative to the control genome. The small fragments may contain one or more bases. We identified SNPs and indels by using samtools, picard-tools and GATK2 software⁴².

Simple sequence repeat (SSR), also known as short tandem repeats and microsatellite markers, which contains the number of repeat motifs with less than six nucleotides (di-, tri-, tetra-, penta-, and hexa-nucleotide). The repeat motifs exist often in the genome of eukaryotes. Unigenes sequences were used to search for potential SSR markers using MISA software (version 1.0, <http://pgrc.ipk-gatersleben.de/misa/misa.html>) with the default parameters. Afterwards, Primer 3 software (version 2.3.5, <http://fokker.wi.mit.edu/primer3/> default parameters) was applied for SSR primer design.

Identification of unigenes involved in biomineralization and comparison of transcripts. Biomineralization-related genes have been reported in the mantle tissues of molluscs such as prismaticin-14⁴³, pif⁴⁴, amorphous calcium carbonate binding protein⁴⁵, and asporin⁴⁶. Identification of the potential biomineralization-related unigenes were carried out by keyword searching according to the reported expressed unigenes in the nr database which were reported to be related to biomineralization. In general, three kinds of unigenes were considered as potential biomineralization-related unigenes, well-known unigenes that have been reported to play important roles for the shell formation, the reported unigenes which are highly homologous with the reported shell formation related unigenes, the proteins those were found in mollusk shells and had high expression in the mantle tissues of molluscan animals. All of the above-mentioned unigenes may probably play key roles for the shell formation. If several unigenes were assigned with the same reference gene, the unigene with the lowest E-value and highest score was assumed to be the homolog of the reference gene.

To compare the similarities and differences between the mineralization proteins of various molluscan animals, the transcriptome annotations of *P. placenta* were compared with the mantle transcriptome annotations from two other molluscs, *P. yessoensis*³⁶ and *C. farreri*³⁷.

Raw reads	29,996,311
Raw bases	4.49G
Clean reads	28,963,193
Clean bases	4.34G
Error rate	0.01%
Q20	98.54%
Q30	95.94%
GC Content	44.45%

Table 1. Statistics of Illumina sequencing in the mantle transcriptome of *P. placenta*.

Items	Total number	Total length (nt)	Mean length (nt)	Shortest length (nt)	Median length (nt)	Longest length (nt)	N50
Transcripts	138,384	106,972,212	773	201	378	39,605	1437
Unigenes	113,325	79,007,878	697	201	345	39,605	1249

Table 2. Statistics of de novo assembly for the mantle transcriptome of *P. placenta*.

Pfam database analysis. Pfam is one of the most comprehensive classification systems for protein domain annotation⁴⁷. Biom mineralization-related unigenes can be found from the Pfam database analysis of the *P. placenta* mantle according to the characteristics of domains found in molluscan shell matrix proteins (SMPs). The Hidden Markov Models (HMMs) of protein domains were obtained from Pfam. The HMMs were applied to study the protein domains and the unigene annotations by using HMMER3 program⁴⁸.

Quantitative real-time polymerase chain reaction (qRT-PCR) analysis. In order to determine the expression levels of the potential biom mineralization-related genes in different tissues of *P. placenta*, the unigenes from seven candidate tissues were characterized by using qRT-PCR. Seven sets of tissue samples were dissected from wild type of adult *P. placenta*, with shell sizes ranging from 50 to 70 mm. All the soft tissue including adduction muscle, gonad, gill, mantle, hepatopancreas, mouthparts, and intestine were dissected from fifteen fresh animals. For qRT-PCR measurement, the weight of each kind of tissue is no less than 300 mg. Each qRT-PCR analysis was replicated for three times. These tissues were kept in RNA Keeper-ICE tissue transfer buffer for RNA extraction. RNA Keeper-ICE tissue transfer buffer penetrates into the tissues during the defreezing process and the RNase of the tissues were inactivated. Total RNA from the different tissues were extracted with TRIzol Reagent at a ratio of 1 mL per 100 mg tissue. The concentration and purity of RNA were quantified using NanoDrop[®] ND-2000 (Thermo scientific). The RNA integrity was evaluated by Bioanalyzer 2200 (Agilent). The primer sequences used for qRT-PCR are listed in Table S1. 2 µg sample of isolated RNA was used to synthesize first-strand cDNA using HiScript[®] II Q RT SuperMix for qRT-PCR (Vazyme). PCR reactions were performed on an ABI 7900 RealTime PCR Detection System (Applied Biosystems, USA). The 10 µL qRT-PCR reaction mixture consisted of 5 µL of 2 × master mix, 1 µL of cDNA, and 0.4 µL primers. The actin gene was used as the reference. The qRT-PCRs were performed under the following conditions: heating at 95 °C 5 min, at 95 °C for 15 s for 40 cycles, at annealing temperature 60 °C for 1 min, and at 95 °C for 15 s, followed by melting curve detection at 60–95 °C. The comparative CT method was used to determine relative mRNA abundance⁴⁹.

Result

Sequence analysis and de novo assembly. A total of 29,996,311 raw reads were obtained after Illumina sequencing analysis. Considering that the raw reads contain a small amount of sequencing adapters and low-quality reads, the resulting clean reads were obtained after quality filtration and then subjected to de novo assembly. Firstly, the quality of the extracted RNA is in A grade (RIN > 7, for our data). Secondly, the Q20 percentage and Q30 percentage are 98.54% and 95.94%, respectively, and the error rate is 0.01%, which indicate that RNA sequencing results are qualified and reliable. The percentage of GC content for the clean reads is 44.45% (Table 1). These results show that the obtained unigenes are suitable for annotation.

Filtered reads are assembled into 138,384 transcripts, with a total length of 106,972,212 nucleotides. These transcripts are subsequently assembled into 113,325 unigenes and the mean length of transcripts is 697 nucleotides, ranging from 201 to 39,605 nucleotides. N50 length is defined as the unigenes length L, while 50% of all bases in the sequences are in contigs of length less than L. The mean length of unigenes is 773 nucleotides, with N50 of 1,437. In summary, the total length of all unigenes is 79,007,878 nucleotides or 73.9% of the length of all transcripts (Table 2). The length-frequency distribution for unigenes and transcripts shows that the numbers of transcripts and unigenes decline with the increase of length, while the ratios of unigenes to transcripts with the same length ranges remain almost the same. The number of the unigenes with length less than 500 bp is 76,237, accounting for 66.39% of all unigenes, while less than one-third of long reads (greater than 1000 bp) are assembled into unigenes, indicating that short reads are more likely to be assembled into unigenes (Fig. 1).

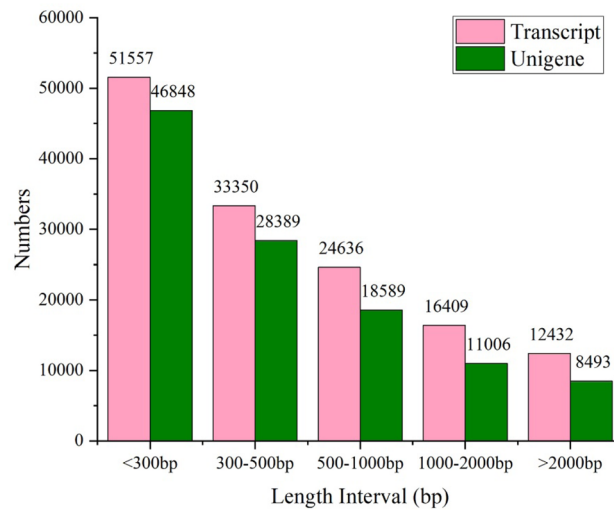


Figure 1. Sequence length distribution of transcripts and unigenes for the mantle transcriptome of *P. placenta* assembled from Illumina reads.

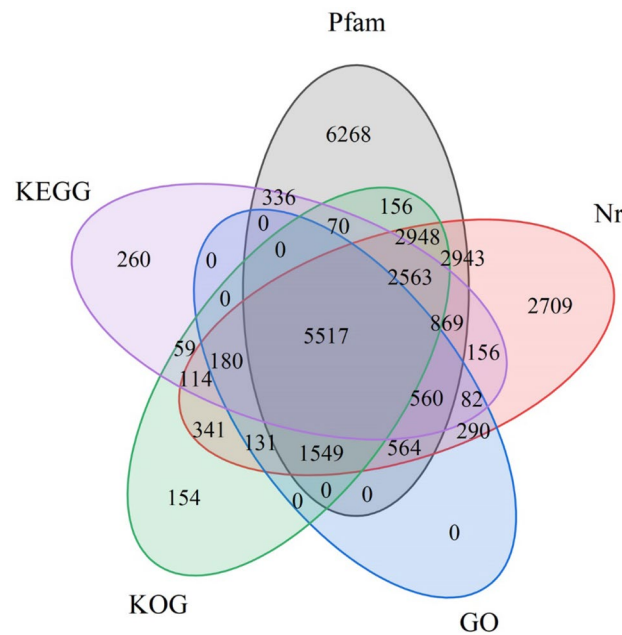


Figure 2. Venn diagram illustrating the numbers of unigenes expressed from public databases nr, Pfam, GO, KOG, and KEGG.

Functional annotations of unigenes. The assembled unigenes of mantle of *P. placenta*, 113,325 in total, were further investigated by using the five public databases (Pfam, nr, GO, KOG, and KEGG) with a cutoff E-value of 10^{-5} (Fig. 2). The unigenes annotated in the Pfam database contain about 21.48% of all unigenes, followed by 18.99% annotated unigenes in the nr database, 12.16% in the KOG database, 9.5% in the KEGG database, and 7.83% in the GO database. In the alignment with NCBI nr database, 21,516 unigenes were successfully annotated (Table 3). Venn diagram of the numbers of unigenes expressed from public databases nr, Pfam, GO, KOG, and KEGG were shown in Fig. 2, which show clearly that many of the unigenes were annotated from more than one database.

Among these successfully annotated unigenes, 46.67% of the annotated unigenes have strong homology with the aligned proteins (E-values less than $1.0E^{-45}$) in the nr database (Fig. 3a). Nearly half of the unigenes (45.43%) return 60–80% similarities to annotated unigenes, and 21.45% return 80–100% similarity to annotated unigenes (Fig. 3b). 55.51% of the annotated unigenes of mantle of *P. placenta* matches best with that extracted from pacific oyster *C. gigas* according to nr database (Fig. 3c). A possible reason for the above homologous genes of the two

The database to be used for annotation	Number of annotated unigenes	Percentage (%) of annotated unigenes
nr	21,516	18.99
KEGG	10,766	9.5
Pfam	24,343	21.48
GO	8873	7.83
KOG	13,782	12.16

Table 3. Statistics of the numbers and percentages of unigenes annotated in different databases and percentage of each database.

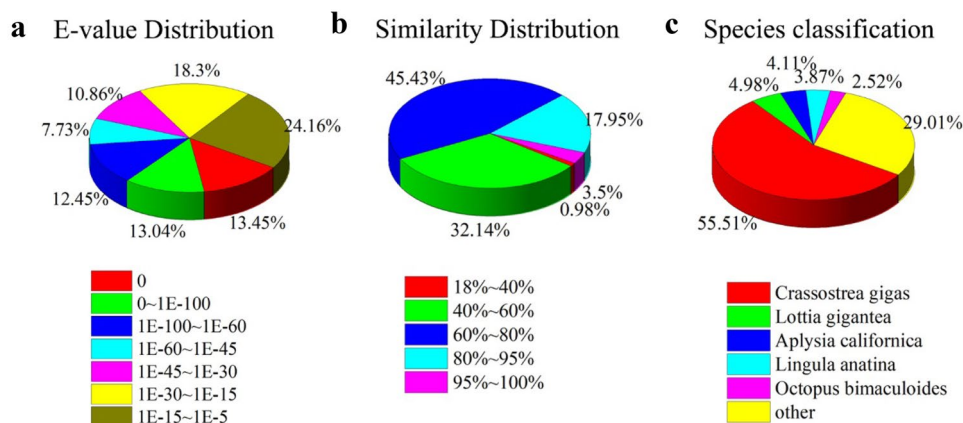


Figure 3. Distribution of E-values (a), and similarities (b) of annotated unigenes from the mantle tissue of *P. placenta*, and the homologous comparison of annotated unigenes from *P. placenta* and other species (c) according to nr database.

different animals might be that the *C. gigas* is one of the few molluscs whose genomes have been completely sequenced and the details of numerous genes of *C. gigas* have been released. The strong similarity revealed in our study further implies that the genome studies of *C. gigas* can be reliable references for the study of *P. placenta* even though the evolutionary relationship between *C. gigas* and *P. placenta* is not very close.

To classify the gene functions, the mantle unigenes of *P. placenta* were investigated by using the GO database. This database is an international standardized gene functional classification system that offers a dynamically updated controlled vocabulary and strict definitions that completely describe the properties of genes and their products in any organism. There are 8,873 unigenes assigned to three main GO categories (Table 3), including biological process category, cellular component category, and molecular function category, and 44 terms in total as shown in Fig. 4. For GO categorization, the “metabolic process” and “cellular process” terms are dominant in the biological process category, the “cell” and “cell part” terms are dominant in the cellular component category, and the “binding” and “catalytic activity” terms were dominant in the molecular function category. The terms “extracellular region” and “extracellular matrix” associated to cellular component category are generally supposed to be related to SMPs. Most SMPs were up-regulated to a greater content at the juvenile stage than those at the adult stage of pearl oyster *P. fucata*⁵⁰. A total of 31 unigenes related to SMPs are recognized in the terms “extracellular region” from the mantle tissue of *P. placenta*, such as collagen, chitin binding protein, calcium ion binding protein (Excel S1).

The mantle unigenes of *P. placenta* were investigated by using the KOG database for better functional classification. 13,782 unigenes of mantle tissues of *P. placenta* were assigned to 26 categories according to KOG database (Table 3). About 41% of the unigenes are assigned to the three main categories, (R) category “general function prediction only” (2739, 17.62%), (T) category “signal transduction mechanisms” (2445, 15.73%), (O) category “posttranslational modification, protein turnover, chaperones” (1314; 8.5%). In this study, the numbers of unigenes of mantle tissue *P. placenta* assigned to the P (inorganic ion transport and metabolism) category and T category were 606 (3.9%) and 2445 (15.73%), respectively (Fig. 5, Excel S2). Many of the annotated unigenes assigned to T and P categories in Excel S2 are homologous with ferritin, calcium ion binding protein, calmodulin and other biomineralization-related proteins as reported in the literature. The P category (inorganic ion transport and metabolism) contains some Ca²⁺-related proteins such as Ca²⁺ binding protein, Ca²⁺ transporting ATPase, Voltage-gated Ca²⁺ channels (Excel S2). In contrast, there are less than 10% unigenes unevenly distributed in the remaining categories such as Y category “nuclear structure” and N category “cell motility”, accounting for 0.33% and 0.3%, respectively (Fig. 5).

To better understand biological pathways involved in the *P. placenta* mantle unigenes, all unigenes of mantle tissues were assigned according to the reference canonical pathways in the KEGG database. According to the

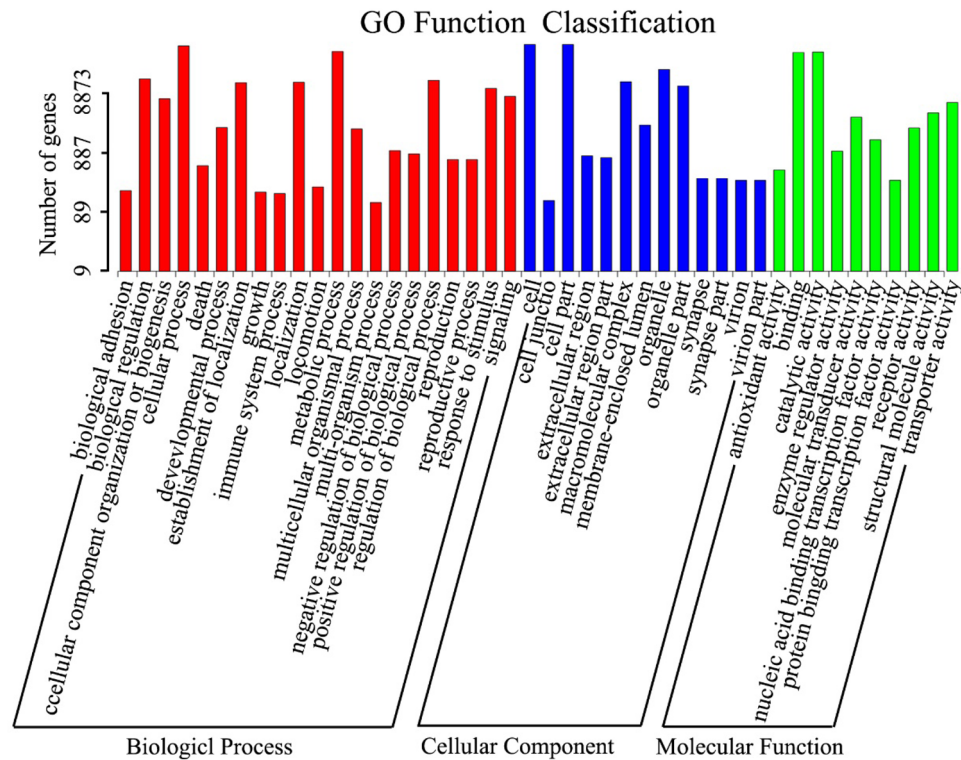


Figure 4. GO categorizations (Biological processes, Cellular components and Molecular functions) of unigenes in the mantle transcriptome of *P. placenta*.

classification results, 10,766 unigenes were assigned to five specific pathways (Table 3). 1,716 unigenes were assigned to metabolism, followed by genetic information processing (1,245 unigenes), organismal systems (741 unigenes), environmental information processing (518 unigenes), cellular processes (472 unigenes). According to the KEGG classification, the pathway of translation owns the largest number of unigenes (634 unigenes), representing for 50.92% of the total annotated unigenes in the category of genetic information processes, followed by pathway of global and overview maps (336 unigenes) in metabolism category (Fig. 6). It was reported that the calcium signaling pathway, Wnt (wingless/int1) signaling and tyrosine metabolism pathways probably play significant roles in the biomineralization process^{37,50}. However, no unigene was found in the calcium signaling pathway in the KEGG database from the unigenes of the mantle tissue of *P. placenta*. 23 unigenes were found to be involved in Wnt signaling and tyrosine metabolism pathways. Among them, the calcyclin binding-protein, frizzled-related protein and dopamine beta-monoxygenase proteins are supposed to be related to calcium signal transduction and the shell formation process (Excel S3). The chemokine signaling and leukocyte transendothelial migration pathways were significantly enriched after the shell damage treatment, indicating that granulocytic hemocytes containing matrix proteins, a type of leukocyte, may transport calcium ions or contribute to the biomineralization process of calcite crystals^{51–53}. There are four unigenes found in these two signaling pathways according to KEGG database and only C–C motif chemokine may involve the biomineralization process (Excel S2).

SNPs, Indels, SSR markers identification and primer design. SNPs are potentially useful for genetic linkage mapping and quantitative trait locus analysis. There is no reported literature related to the SNP identification of *P. placenta*. In this study, polymorphism analysis identified 465,392 potential SNPs from 60,371 unigenes with frequency of about 7.56 SNPs per unigene (Excel S5). The numbers of transition (261,292) and transversion (202,192) mutations accounted for 56.14% and 43.45% in SNP types, respectively. The most abundant type of base variations is C/T polymorphism (70,283; 15.1%), because C in CG base is often methylated and transfers to thymine after spontaneous deamination. Another abundant type of base variations is A/T polymorphism, such as G to A (70,176; 15.08%), A to G (61,051; 13.12%) (Fig. 7). 62,103 potential indels in total were also identified among 21,630 unigenes with frequency of about 2.87 indels per unigene (Excel S6). The numbers of unigenes decrease gradually with the increase of indel length from 1 to 10 bp (Fig. S1). The indel lengths are mainly 1, 2 and 3 bp. The numbers of deletion are larger than that of insertion when the indel length is more than 10 bp (Fig. S1).

In addition, 21,640 unigenes containing 26,048 potential SSR markers were identified (Fig. 8). Among these unigenes, the most abundant type of repeats is p1 (mono-nucleotide) (15,224; 58.4%), followed by p2 (di-nucleotide) (6227; 23.9%), p3 (tri-nucleotide) (2700; 10.3%), p4 (tetra-nucleotide) (361; 1.3%) repeats. Furthermore, there are some c- types (complex repeat motifs), accounting for 5.8% (1524), different from the above mentioned nucleotides in these SSR markers. The frequencies of SSRs with different numbers of repeat units were calculated.

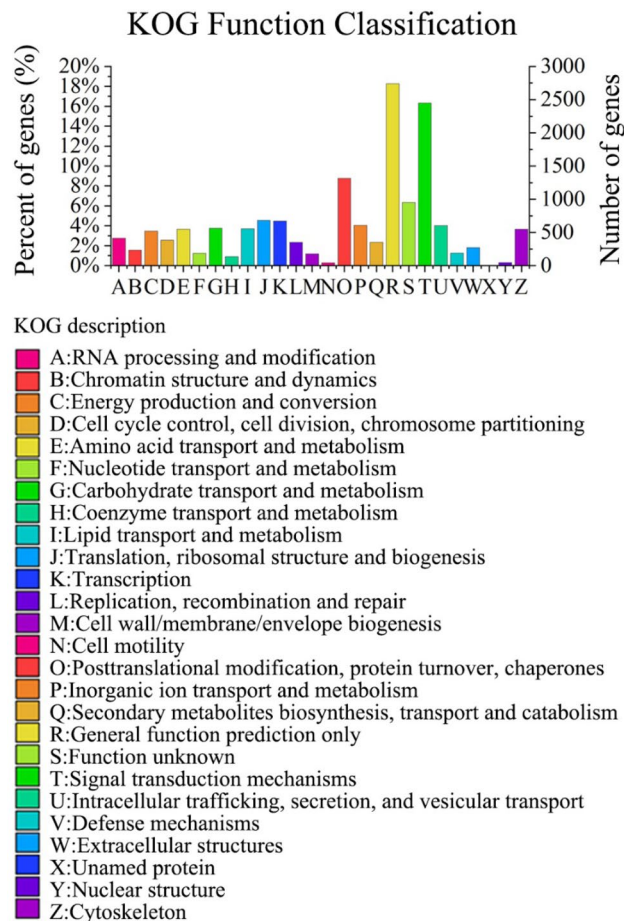


Figure 5. KOG categorizations of unigenes in the mantle transcriptome of *P. placenta*.

Among the p1 repeat units, the counts of the repeat units between 9 and 12 are dominant, accounting for more than 50%, followed by 13–16 repeat units (17.74%) and 17–20 repeat units (12.85%). More than half of 5–8 repeat units were assigned to the p2 repeat units, followed by 9–12 repeat units. However, only 5–8 repeat units are present in both p3 and p4 repeats (Fig. 8). The A/T repeat units of mono-nucleotide are dominant among these types, accounting for 55.57%. AT/TA (9.42%) was dominant in the p2 repeat units, followed by TG/CA (4.94%), GA/GT (4.16%) (Fig. S2). ATG/GAT (2.2%) was dominant in the p3 repeat units, followed by ATC/TGA (1.71%), CAT/TCA (1.2%). The remaining types of motifs have more complicated types but less numbers (c-type, complex repeat motifs), accounted for 7.28% in total. To further test the SSR markers, forward and reverse primer pairs were obtained using Primer 3.0 (Excel S7).

Identification of genes involved in biomineralization process. To better understand the proteins related to the shell formation process, the annotated unigenes of mantle tissues of *P. placenta* were compared with the sequences of proteins known to be associated with biomineralization process in molluscan shells using nr database. The annotated unigene with lowest E-value was selected as the representative unigene while several annotated unigenes were assigned to the same reference unigene. 178 homologous unigenes of 51 shell matrix proteins, such as calmodulin, perlucin, ferritin and carbonic anhydrase were found to be related to biomineralization process in the transcriptome of mantle tissue of *P. placenta* (Excel S10). This is the first time to report the potential biomineralization-related unigenes in the mantle tissue of *P. placenta*, as far as we know. A lot of researchers have focused on the identification of genes related to the shell formation of molluscan animals, and an increasing number of genes have been identified^{55–58}. 259 proteins were identified from oyster *C. gigas* shells by proteomic analysis⁵⁶. In comparison with the proteomic data of shell matrix of *C. gigas*, we identified a set of 158 unigenes that are probably related to shell formation, including house-keeping protein elongation factor 1a, and extracellular matrix protein collagen (Excel S9). Many of the shell-formation related proteins are enzymes such as glutathione peroxidase, hemicentin and tyrosinase, that may be involved in matrix construction or modification^{57,58}. In this study, only one enzyme tyrosinase (three unigenes) was found to be related to the shell formation process in the transcriptome of *P. placenta* (Excel S9). Twenty-one proteins with 66 homologous unigenes were identified to be related to calcitic shell formation of *P. placenta* (Table 4). Furthermore, eighteen of the above 66 unigenes were identified to be highly expressed in mantle tissue of *P. placenta*. A few highly

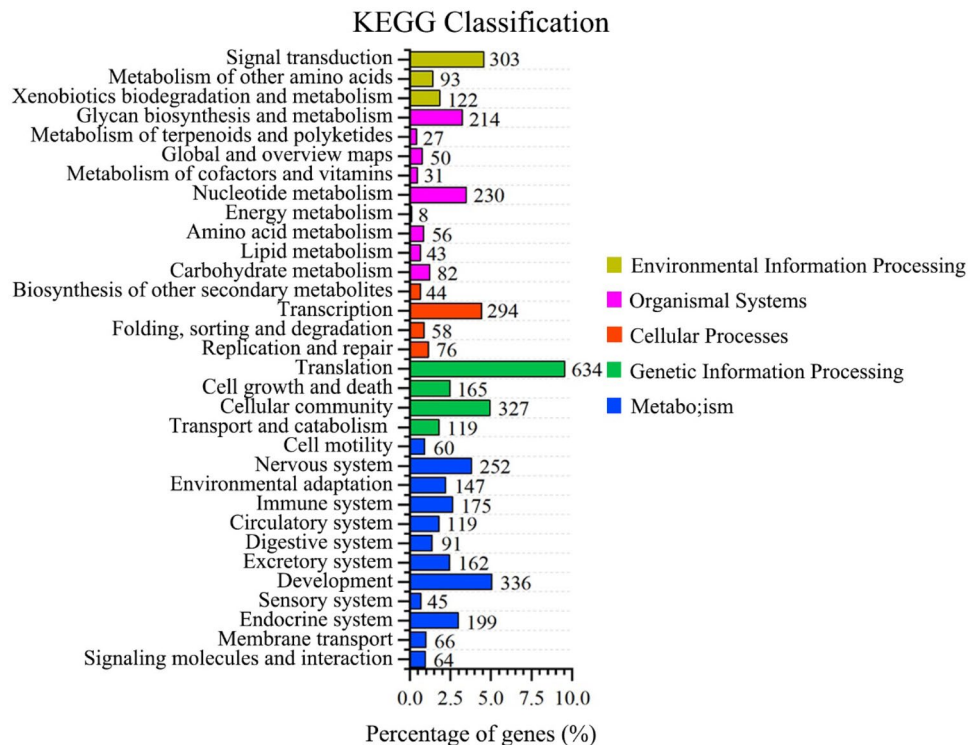


Figure 6. KEGG assignment of unigenes in the mantle transcriptome of *P. placenta*.

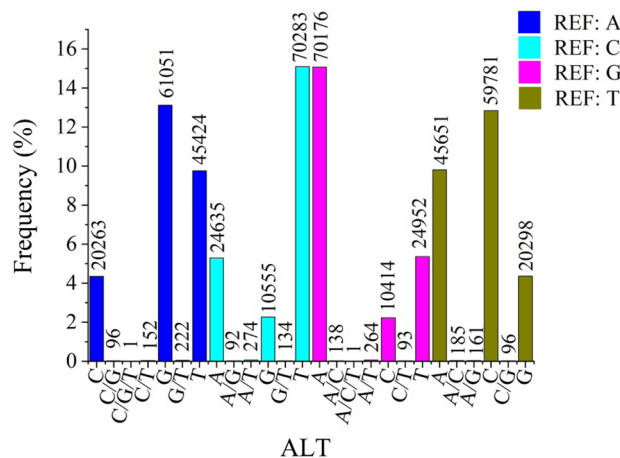


Figure 7. SNP types and frequencies from *P. placenta* mantle tissue. REF: The genotypes of the reference sequences at the defined sites, ALT: Other genotypes at the reference sites.

expressed unigenes with fragments per kilobase of feature per million mapped reads (FPKM) values of > 15 are carbonic anhydrase, calreticulin, ferritin, perlucin, gigasin-2, and tyrosinase-like proteins.

Comparison of the mantle transcriptomes of different molluscs. The shells of scallops *P. yessoensis*, *C. farreri* and *P. placenta* are composed of foliated calcite minerals. The mantle transcriptomes of scallop *P. yessoensis*, *C. farreri* and *P. placenta* were compared to find the similarities and differences of the biomineralization related proteins of these molluscan organisms^{36,37}. We find 117 biomineralization-related unigenes in the mantle of *P. placenta*, much less than those in *P. yessoensis* (162 unigenes), but much more than those in *C. farreri* (42 unigenes) (Fig. 9). There are six biomineralization-related unigenes expressed in the mantles of the three species, including sarcoplasmic calcium-binding protein, calcineurin a, calmodulin-like protein, perlucin, alkaline phosphatase, tyrosinase-like protein tyr-3. In comparison to *C. farreri*, *P. placenta* and *P. yessoensis* have more similar homologous biomineralization-related unigenes, 31 unigenes in total, about 27% of the biomineralization related unigenes of *P. placenta*. For example, collagen, chitin synthase 1/2, carbonic anhydrase-like, heat

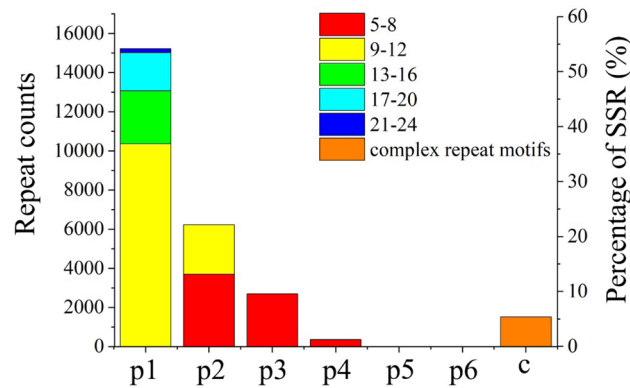


Figure 8. Distribution of SSRs based on the number of repeat units. p1: mono-nucleotide, p2: di-nucleotide, p3: tri-nucleotide, p4: tetra-nucleotide, p5: penta-nucleotide, p6: hexa-nucleotide, c: complex repeat motifs.

shock protein 70 and calmodulin were found in both *P. yessoensis* and *P. placenta*, potentially indicating their functional similarities for their biomineralization processes. There are probably more mineralization proteins to be discovered in the unigenes of the mantle tissue of *P. placenta* (Fig. 9, Excel S10).

Quantitative Real-Time PCR (qRT-PCR) analysis. Ten selected potential biomineralization-related unigenes in the different tissues such as adduction muscle, gonad, gill, mantle, hepatopancreas, mouthparts, and intestine of *P. placenta* were examined by qRT-PCR (Fig. 10). In general, four of ten unigenes have much higher expression in mantle tissue than those in the other tissues, which include *c76266_g1* (gigas-in-2), *c73086_g1* (tyrosinase-like), *c66761_g1* (pif-like) and *c59513_g1* (teneurin-2) unigenes (Fig. 10a–d). Pif is an important macromolecule for in vivo shell formation of nacre⁴⁴. In addition, it was found that pif can induce the formation of aragonite and vaterite crystals in the in vitro system⁷⁷. Tyrosinases are abundant in shells and their high expression in mantle of the pacific oyster *C. gigas* indicates that their functions are probably related to shell formation⁵⁶. Gigasin-2 and teneurin-2 were identified for the first time from the shell of *C. gigas* via shell matrix proteome characterization^{78,79}. According to the Pfam database analysis, teneurin-2 in *P. placenta* is predicted to be epidermal growth factor (EGF) domain and gigasin-2 has zona pellucida (ZP) domain (Excel S11). EGF domains are mostly found in the SMPs as tandem repeats and only present in the prismatic (calclitic) layers but not in the nacreous layer of *Pinctada*⁸⁰. The EGF domain is a calcium-binding motif composed of 45 amino acids arranged in two small β -sheets with six conserved cysteine residues⁸¹. Both EGF-like and ZP domains have been reported in the shells of *Lottia gigantea*⁸² and *C. gigas*⁷⁸. ZP domains are present in a range of extracellular filament or matrix proteins from a wide variety of eukaryotic organisms, and are characterized by eight conserved cysteine residues, which are involved in protein polymerization processes⁸³. We propose that the above mentioned four unigenes are primarily biomineralization-related unigenes. However, the expression levels of the other six unigenes in the mantle tissue are not as high as their expression levels in the other tissues, indicating that they probably have other functions except for the biomineralization process (Fig. 10e–j). We would discuss more about the functions of these ten unigenes in the discussion part.

Pfam database analysis. Pfam database search is important for understanding the possible biomineralization-related functions of the shell matrix proteins of molluscan animals. In molluscan animals, shell matrix proteins are very often repetitive, highly conserved, low complex domains. The functions of these protein domains involved in biomineralization have been studied by many groups^{13,84}. For example, the nacre protein perlucin contains a C-type lectin domain and has a broad carbohydrate-binding feature, which was supposed to facilitate calcium-dependent glycoprotein-protein interactions within the skeletal matrix²³. In molluscan animals, the pif-like proteins contain von willebrand factor type A (VWA), chitin-binding and laminin G domains, which can bind chitin framework and accelerate CaCO_3 precipitation inside the chitin membrane, and then regulate their vertical alignment^{85,86}. In bivalves, most EF-hand proteins from the mantle tissue of bivalves are Ca^{2+} sensors or signal modulators, which may induce conformational change by binding with Ca^{2+} , such as calmodulin, troponin C and myosin light chains⁸⁷.

In this study, the sequences of potential biomineralization genes were obtained from the transcriptomics analysis of *P. placenta* mantle. Based on these amino acid sequences, the information of their domain was obtained by using Pfam. Finally, we speculate on the function of potential biomineralization-related proteins of *P. placenta* based on the characteristics of the domain found in biomineralization proteins of molluscan. Identification of the possible biomineralization-related functions of the unigenes expressed in the mantle tissue of *P. placenta* were carried out by keyword searching according to the domain which were reported to be related to biomineralization in the Pfam database. Many potential biomineralization-related unigenes are proposed to be involved in the shell formation from the transcriptomes of the mantle tissue of *P. placenta* according to the Pfam database analysis. (Excel S11).

Calmodulin, calponin and mucin proteins are supposed to be associated with molluscan shell formation. Among those proteins, calponin was highly expressed in the mantle of *P. placenta*, with FPKM values > 551, but

Accession	Gene_name	Gene_ID	nr_Evalue	nr_Description	Low FPKM	High FPKM
EKC34657	Peroxidase1 ⁵⁶	c84192_g1	0	Chorion peroxidase [<i>Crassostrea gigas</i>]	27.46	63.27
EKC26108	Peroxidase2 ⁵⁶					
AAV69062.1	Alkaline phosphatase ^{56,59}	c84312_g1	3.00E-156	PREDICTED: alkaline phosphatase-like [<i>Lingula anatina</i>]	6.41	13.58
		c84796_g1	1.00E-156	Alkaline phosphatase, tissue-nonspecific isozyme [<i>Crassostrea gigas</i>]	4.31	8
BAG68618.1	BMP2/4 ^{60,61}	c349_g1	5.00E-31	Bone morphogenetic protein 2/4 [<i>Lymnaea stagnalis</i>]	0.26	2.13
		c123147_g1	5.00E-06	Bone morphogenetic protein 2/4 [<i>Septifer virgatus</i>]	0	10.27
BAJ52887.1	Carbonic anhydrase(CA), similar to Nacrein ⁵⁶	c84941_g5	1.00E-45	PREDICTED: carbonic anhydrase-like [<i>Crassostrea gigas</i>]	126.8	244.22
		c85950_g1	2.00E-39	PREDICTED: carbonic anhydrase 1-like [<i>Crassostrea gigas</i>]	103.95	139.29
		c71108_g1	8.00E-77	Carbonic anhydrase 2 [<i>Crassostrea gigas</i>]	67.5	124.48
		c81423_g3	1.00E-69	PREDICTED: carbonic anhydrase 1-like [<i>Crassostrea gigas</i>]	27.26	45.21
ACI22622.1	Calmodulin ⁵⁶	c114357_g1	8.00E-11	PREDICTED: calmodulin-like [<i>Crassostrea gigas</i>]	2.41	7.74
		c10600_g1	3.00E-09	RecName: Full = Calmodulin; Short = CaM	0.81	3.51
		c78814_g1	7.00E-29	RecName: Full = Calmodulin; Short = CaM	0.69	2.77
		c65518_g1	2.00E-12	RecName: Full = Calmodulin; Short = CaM	0.15	2.93
		c31509_g1	9.00E-28	Calmodulin [<i>Crassostrea gigas</i>]	0	1.12
		c102707_g1	2.00E-12	RecName: Full = Calmodulin; Short = CaM	0	0
		c52942_g1	4.00E-06	RecName: Full = Calmodulin-2; Short = CaM 2	0	0.38
		c60826_g1	1.00E-94	RecName: Full = Calmodulin; Short = CaM	0	1.38
		c62972_g1	8.00E-12	Calmodulin [<i>Crassostrea gigas</i>]	0	0
ABR68546.1	Calreticulin ^{62,63}	c84621_g1	2.00E-135	Ccalreticulin-like protein [<i>Littorina littorea</i>]	114.56	172.26
		c112169_g1	6.00E-06	PREDICTED: calreticulin-like [<i>Biomphalaria glabrata</i>]	0	0
BAF73720.1	Chitin synthase/PfCHS1 ^{56,64}	c86966_g4	3.00E-164	Chitin synthase, partial [<i>Macandrevia cranium</i>]	5.4	19.33
		c24142_g1	4.00E-08	Chitin synthase 1 [<i>Crassostrea gigas</i>]	0	52.05
P86734.1	Ependymin related protein 1 (EDPR) ^{56,59,65}	c104043_g1	3.00E-06	Mammalian ependymin-related protein 1-like [<i>Lingula anatina</i>]	0	0
		c107140_g1	1.00E-18	Mammalian ependymin-related protein 1-like [<i>Crassostrea gigas</i>]	0	0
		c110802_g1	4.00E-06	Mammalian ependymin-related protein 1-like [<i>Biomphalaria glabrata</i>]	0	3.51
AAQ12076.1	Ferritin ⁶⁶⁻⁶⁸	c71166_g1	2.00E-105	Ferritin [<i>Haliotis diversicolor supertexta</i>]	2598	5972.88
		c54191_g1	2.00E-41	Ferritin [<i>Scapharca broughtonii</i>]	0	0
		c90808_g1	3.00E-17	Ferritin [<i>Conus novaehollandiae</i>]	0	0
P86785	Gigasins-2 ⁵⁶	c76266_g1	5.00E-10	PREDICTED: gigasin-2 isoform X1 [<i>Crassostrea gigas</i>]	61.13	124.78
ADD16957.1	Perlucin ^{56,69}	c67461_g2	3.00E-23	Perlucin [<i>Haliotis diversicolor</i>]	65.55	100.36
		c74134_g1	4.00E-06	Perlucin [<i>Hyriopsis cumingii</i>]	4.71	8.62
		c100512_g1	4.00E-18	Perlucin [<i>Haliotis diversicolor</i>]	0.71	1.2
		c66124_g1	3.00E-29	Perlucin [<i>Crassostrea gigas</i>]	0.28	1.45
		c15372_g1	6.00E-16	Pperlucin [<i>Hyriopsis cumingii</i>]	0	0
		c71177_g1	1.00E-08	Perlucin [<i>Crassostrea gigas</i>]	0	2.11
BAH97338.1	Protein PIF ⁵⁶	c71325_g1	4.00E-20	PREDICTED: protein PIF-like [<i>Crassostrea gigas</i>]	31.79	44.89
		c66761_g1	5.00E-11	PREDICTED: protein PIF-like [<i>Crassostrea gigas</i>]	27.22	37.84
		c64812_g1	2.00E-06	PREDICTED: protein PIF-like [<i>Crassostrea gigas</i>]	0.45	1.42
BAA89420.1	Sarcoplasmic calcium-binding protein ^{70,71}	c102115_g1	1.00E-11	Sarcoplasmic calcium-binding protein [<i>Meretrix lusoria</i>]	0	0
		c107796_g1	4.00E-30	Sarcoplasmic calcium binding protein [<i>Ruditapes philippinarum</i>]	0	1.13

Continued

Accession	Gene_name	Gene_ID	nr_Evalue	nr_Description	Low FPKM	High FPKM
AAZ66340.1	Tyrosinase-like ^{56,72}	c73086_g1	1.00E-82	PREDICTED: tyrosinase-like protein [<i>Crassostrea gigas</i>]	303.51	430.78
		c73399_g1	6.00E-88	PREDICTED: tyrosinase-like protein 1 isoform X2 [<i>Crassostrea gigas</i>]	147.25	202.78
		c73399_g1	6.00E-88	PREDICTED: tyrosinase-like protein 1 isoform X2 [<i>Crassostrea gigas</i>]	147.25	202.78
		c84724_g1	1.00E-60	Tyrosinase [<i>Mizuhopecten yessoensis</i>]	4.59	15.55
		c84724_g1	1.00E-60	Tyrosinase [<i>Mizuhopecten yessoensis</i>]	4.59	15.55
		c74294_g1	6.00E-42	Putative tyrosinase-like protein tyr-3 [<i>Crassostrea gigas</i>]	4.47	9.96
		c111859_g1	2.00E-27	PREDICTED: tyrosinase-like protein [<i>Crassostrea gigas</i>]	0	0.62
		c63653_g1	3.00E-08	Putative tyrosinase-like protein tyr-3 [<i>Crassostrea gigas</i>]	0	3.21
		c6923_g2	7.00E-13	Putative tyrosinase-like protein tyr-3 [<i>Crassostrea gigas</i>]	0	0.24
		c94853_g1	3.00E-25	Tyrosinase A2 [<i>Pinctada maxima</i>]	0	2.86
P86987	Insoluble matrix shell proteins 6(IMSP6) ^{56,73}	c59878_g1	1.00E-08	Insoluble matrix shell protein 6-like, partial [<i>Crassostrea gigas</i>]	41.5	55.56
		c69385_g1	8.00E-08	PREDICTED: insoluble matrix shell protein 6-like [<i>Crassostrea gigas</i>]	32.11	104.85
		c85881_g6	9.00E-11	Putative insoluble matrix shell protein 5-like protein [<i>Pinctada fucata</i>]	0.44	5.13
EKC41461	Tyrosine-protein phosphatase Lar ⁵⁶	c85401_g1	0	Tyrosine-protein phosphatase Lar-like isoform X1 [<i>Crassostrea gigas</i>]	8.91	37.66
		c56636_g1	1.00E-05	Tyrosine-protein phosphatase Lar [<i>Crassostrea gigas</i>]	0	0
EKC31553	Amine oxidase ^{56,74,75}	c85095_g1	9.00E-27	Putative amine oxidase [copper-containing] [<i>Biomphalaria glabrata</i>]	2.9	4.99
		c45366_g1	3.00E-48	Putative amine oxidase [copper-containing] [<i>Biomphalaria glabrata</i>]	1.42	5.44
		c83183_g2	1.00E-65	Putative amine oxidase [copper-containing] [<i>Crassostrea gigas</i>]	0.68	3.06
		c19032_g1	4.00E-53	Putative amine oxidase [copper-containing] [<i>Biomphalaria glabrata</i>]	0.66	1.43
		c61900_g1	6.00E-16	Putative amine oxidase [copper-containing] [<i>Biomphalaria glabrata</i>]	0.36	2.72
		c33247_g1	1.00E-44	Putative amine oxidase [copper-containing] [<i>Crassostrea gigas</i>]	0.35	0.55
		c73931_g1	3.00E-10	Putative amine oxidase [copper-containing] [<i>Aplysia californica</i>]	0	2.02
CGI_10024867	Chitotriosidase1 ^{59,75}	c51351_g1	8.00E-146	PREDICTED: chitotriosidase-1-like [<i>Lingula anatina</i>]	44.12	51.43
		c79603_g1	6.00E-125	PREDICTED: chitotriosidase-1-like, partial [<i>Aplysia californica</i>]	0.33	1.73
H2A0L6	Beta-hexosaminidase ⁵⁶	c75081_g1	4.00E-46	RecName: Full = Putative beta-hexosaminidase; AltName: Full = Beta-N-acetylhexosaminidase [<i>Crassostrea gigas</i>]	35.32	73.48
K1QBN6	Caltractin ⁷⁶	c70548_g1	1E-48	PREDICTED: caltractin-like [<i>Acropora digitifera</i>]	13.11	24.88
K1RCP3	Teneurin-2 ⁵⁶	c59513_g1	4.00E-06	Teneurin-2 [<i>Crassostrea gigas</i>]	324.17	489.75

Table 4. Identification of genes involved in the calcitic shell formation of *P. placenta*^a. ^aThe first column is the accession number of the protein, which could be indexed in the Uniprot. Expression level is shown as FPKM. RecName is the protein name recommended by the UniProt database. AltName represents synonyms found in the literature or in other databases.

the expression values of most of the other two proteins are relatively low. Only c81494_g1 unigene (calmodulin-like protein) had a high expression level (FPKM values in between 38.17 and 76.44) (Excel S11), but it expressed a higher quantity in mouth parts than that in the other tissues (Fig. 10f), suggesting that calmodulin may play other roles in *P. placenta*. The insoluble matrix shell protein 6 shows a higher expression in the tissues of hepatopancreas and intestine than that of the other tissues in *P. placenta* (Fig. 10j).

Cadherin and collagen proteins contain enriched amount of von willebrand factor type A and epidermal growth factor domains, indicating that they were derived from the extracellular matrix⁸⁸. However, most of them show a low expression level in the mantle tissue of *P. placenta*. Only c83310_g1 (collagen alpha-1) shows a relatively high expression level (FPKM values in between 11.18 and 32.32) (Excel S11). Perlecan extracted from abalone nacre contains a functional C-type lectin domain which can increase the precipitation rate of calcium carbonate from a saturated solution, indicating that it may promote the nucleation and/or the growth of calcium

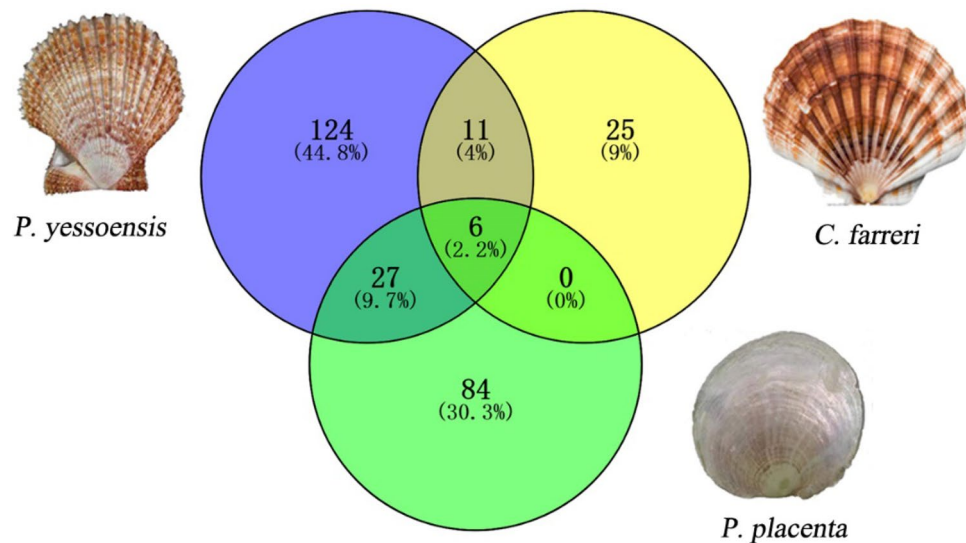


Figure 9. Comparison of mantle transcriptomes of three scallop shells composed of foliated calcite crystals. Blue circle: 168 biomining-related unigenes (124 exclusive) discovered from *P. yessoensis* mantle³⁶. Yellow circle: 42 biomining-related unigenes (25 exclusive) discovered from *C. farreri* mantle³⁷. Green circle: 117 biomining-related unigenes (84 exclusive) discovered from *P. placenta* mantle.

carbonate crystals^{15,23}. Among the proteins containing perlucin domains, only the c67461_g2 unigene (perlucin) shows a high expression level (FPKM values: 65.55–100.36) in the mantle tissue of *P. placenta* (Excel S11). However, c67461_g2 unigene shows higher expression in tissues of hepatopancreas than that of the other tissues by qRT-PCR (Fig. 10e). Carbonic_anhydrase domain was reported to be involved in the formation of both nacreous and prismatic layers^{20,24,89}. The expression values of all the carbonic anhydrase unigenes are relatively high, with FPKM values > 27, such as c71108_g1, c85950_g1, c81423_g3 and c84941_g5 unigenes (Excel S11). At the mean time, the expression of c84941_g5 was higher in gill than that in the other tissues (Fig. 10g). On the other hand, some extracellular enzymes or inorganic ion-binding proteins such as chitinase, hemocentin and peroxidase with relatively high expression levels in the mantle tissue of *P. placenta* are probably involved in the shell formation (Excel S11). Chitinase in shell matrix may reconstruct the chitinaceous scaffold and promote the interaction between chitin and chitin binding protein⁸². In comparison to the other ion-binding proteins, the c71155_g1 unigene (chitinase 1) has a relatively high expression level (FPKM values in between 96.85 and 125.15) (Excel S11), but we didn't know yet how specific this gene is in each tissue of *P. placenta*.

Discussion

Shell formation is a very complicated process that involves a series of proteins and genes, while living organisms produce biominerals with superior mechanical properties under biological control¹. The main objective of this study is to identify unigenes involved in biomineralization. Our study indicates that a high-coverage expression profile can be produced by using short-read Illumina sequencing technology and some effective sequence assembly tools such as Trinity and ESTScan softwares. A total of 113,325 unigenes with an average length of 697 bp was generated from the mantle tissue of *P. placenta* by using Illumina HiSeq™ 4000 sequencing technology, while 66.39% of the above unigenes (76,237) have lengths less than 500 bp (Fig. 1). The lengths of the unigene sequences from the mantle tissue of *P. placenta* are larger than those generated in the reported unigene sequences of *P. penguin*³⁴. These unigene sequences are similar to the reports for the yesso scallop *P. yessoensis* (93,204 unigenes; 733 bp)³⁶ and the freshwater pearl mussel *C. plicata* (98,501 unigenes; 689 bp)³³. However, the number and mean length of the unigenes from the mantle tissue of *P. placenta* are larger than that from the zhuikong scallop *C. farreri* (77,975 unigenes; 538 bp)³⁷ and the pearl oyster *P. maxima* (108,704 unigenes; 407 bp)⁹⁰. This difference may be attributed to the use of different sequencing platforms. A large quantity of genomic data is available for many bivalve species, but only 18.99% of the mantle tissue of *P. placenta* were annotated in nr database. This means that more than half of the unigenes of the mantle tissue of *P. placenta* have no known homologous unigenes. The low rate of annotated unigenes of the mantle tissue of *P. placenta* could be a result of limitations in the genomic information available for *P. placenta*, which is the case in many of other bivalve species^{33,36}.

The SNPs of the mantle tissue of *P. placenta* were obtained from the sequencing errors as well as the true SNPs. In our data, the obtained SNP density was 0.0059 SNPs/bp (0.59%) in the mantle tissue of *P. placenta* (Excel S5, Table 2), which was significantly larger than the sequencing error (0.01%) (Table 1). This result confirmed the reliability of the obtained SNPs data. The SNPs are potentially useful for genetic linkage mapping and for the analysis of quantitative traits of the *P. placenta*. The transcriptome we present here provides the most comprehensive polymorphism for the *P. placenta* to date, as far as we know. The SNP density of the eastern oyster *C. virginica* is 0.042 SNPs/bp⁵⁴, which was more polymorphic than that of the mantle of *P. placenta* (0.0059 SNPs/

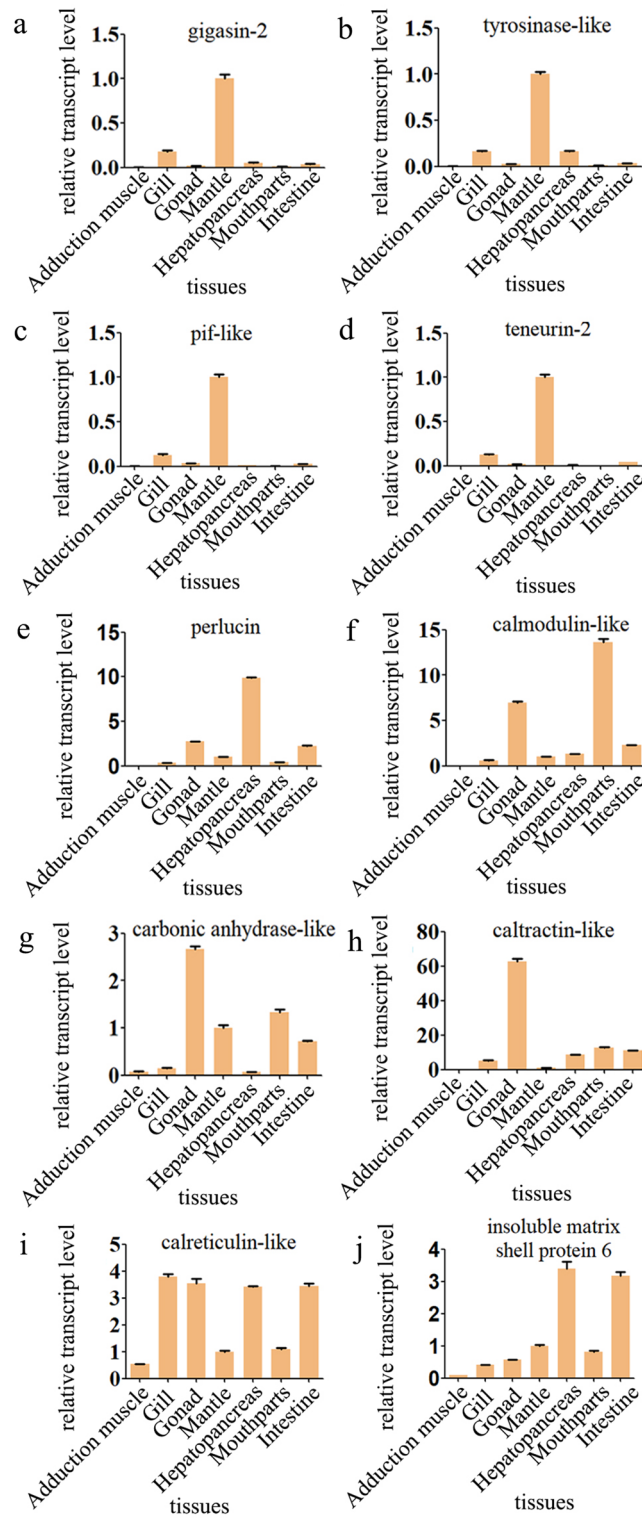


Figure 10. Differential expression of gigasin-2 (c76266_g1) (a), tyrosinase-like (c73086_g1) (b), pif-like(c66761_g1) (c), teneurin-2 (c59513_g1) (d), perlucin (c67461_g2) (e), calmodulin-like, (c81494_g1) (f), carbonic anhydrase-like (c84941_g5) (g), caltractin-like (c70548_g1) (h), calreticulin-like (c84621_g1) (i), and insoluble matrix shell protein 6 (c69385_g1) (j) in adduction muscle, gonad, gill, mantle, hepatopancreas, mouthparts, and intestine tissues of *P. placenta*, were determined using real-time PCR, The error bars represent the standard error of three biological replicates, statistical significance was considered at $p < 0.05$.

bp) (Excel S5, Table 2). 465,392 potential SNPs were constructed from 60,371 unigenes, with frequency of about 7.56 SNPs per unigene in the mantle tissue of *P. placenta*, which was consistent with the reported result of the mantle tissue of pearl oyster *P. martensii*³⁵. The indel density of the mantle tissue of *P. placenta* was 2.87 indels per unigene (Excel S6), which was much lower than that indel density of the mantle tissue of *P. martensii*³⁵. As is well known, the Illumina sequencing is considered to be robust against homopolymer errors and therefore it may be suited well for identification of indels⁹¹.

The variations in unigene expression between different tissues have been shown to be correlated with shell formation in molluscs such as *C. gigas*²², *P. penguin*³⁴, and *T. pyramis*³². The exact biomineralization functions of proteins such as perlucin¹⁵ and pif⁸⁵ have been investigated by in vitro and in vivo mineralization studies. According to the nr database, 178 potential biomineralization-related unigenes were identified in the mantle transcriptome of *P. placenta* in this work. Among these unigenes, ten selected potential biomineralization-related unigenes (FPKM values > 15) were examined in the different tissues of *P. placenta* by qRT-PCR. In the present study, four of ten unigenes (gigas-in-2, tyrosinase-like, pif-like and teneurin-2) have much higher expression in the mantle tissue than those in the other tissues, indicating that they are very often related with the biomineralization process of *P. placenta* shell (Fig. 10a–d). Three gigasin-2 isoforms were identified in water soluble matrix of *C. gigas* shell, which are proposed to be involved in bone remodeling processes and could be responsible for the biocompatibility between bone and nacre grafts⁷⁹. Meanwhile, gigasin-2 was highly expressed in the mantle tissue of *C. gigas*⁵⁶. However, homologous proteins have not been identified in other species since gigasin-2 was reported in 2012⁵⁶. It is known that tyrosinase family are potentially involved in melanin biosynthetic pathway in various organisms. Moreover, it was reported that tyrosinases from molluscs are secreted from the mantle and transported to the prismatic layer of the shell, while they contribute to melanin biosynthesis and shell pigmentation^{56,92,93}. In this study, there is no pigmentation in the mantle tissue and the transparent shell of *P. placenta*. Therefore, the high expression of tyrosinase in the mantle tissue indicates that their functions are not only related to melanin biosynthesis, but also related to the shell formation. After injection of pif dsRNA, both of calcite laths of the *C. gigas* shell and nacreous layer of the *P. fucata* shell grew to disordered structure in vivo, indicating that pif protein might be essential for the normal growth of the prismatic and nacreous layer^{44,77}. The teneurin-2 was first identified from diverse shell matrix proteome and had signal peptides in *C. gigas*, and it was proposed that it was secreted from the mantle into the shell⁷⁸. According to the above discussion about the expressions of the four unigenes and their functions for biomineralization process in the other molluscan species in the literatures, we propose that gigasin-2, tyrosinase-like, pif-like and teneurin-2 may play important roles for the biomineralization process. It would be important to extract these four proteins and investigate their functions for biomineralization via in-vitro crystallization process of CaCO₃ in the future. We consider to study the full-length cDNA sequences, gene expression and recombinant proteins of these four unigenes to understand their functions for in vivo and in vitro crystallization of CaCO₃.

Six of the remaining unigenes don't exhibit high expression in the mantle tissue, in comparison to the other tissues of *P. placenta*. Thus, it is hard to tell whether they participate in biomineralization or not based on the qRT-PCR analysis. These findings are somewhat similar from those of previous studies. For example, the researchers identified six types of perlucin and discovered their different expression levels in different tissues of *T. pyramis*. Some of the perlucin proteins were expressed at the highest levels in the digestive gland, while the others were expressed at high levels in the mantle or the gonad of *T. pyramis*³². The perlucin was isolated from the nacreous layer of the marine snail *Haliotis laevis* and it could promote the nucleation of CaCO₃ crystals on the calcite surface in the in vitro experiments⁹⁴. In addition, calmodulin-like protein can induce the nucleation of aragonite through binding with the 16-kDa protein and regulates the growth of calcite in the prismatic layer of pearl oyster *P. fucata*⁹⁵. This expression pattern and the in vitro crystallization experiments suggest that perlucin family may play important roles in both of the biomineralization process and digestive process^{32,94}. Similar to the perlucin, two types of calmodulin were expressed at the lowest level in the mantle than the other tissues in *T. pyramis*³². Calmodulin-like protein was expressed with the highest level in the mantle tissue of *P. fucata* species and has a potentially high affinity for calcium⁹⁶. The carbonic anhydrase family were expressed in the mantle and associated with the shell formation in the european abalone *Haliotis tuberculata*⁹⁷. Nacrein containing carbonic anhydrase domain was expressed in both the nacreous layer and the prismatic layer of *P. fucata*²⁴. Meanwhile, it was also highly expressed in the mantle of *T. pyramis* and *P. penguin*^{32,34}. However, the researchers analyzed the expression levels of the five studied carbonic anhydrase isoforms in different tissues and found that four of them were more highly expressed in the hemocytes than in the gills or the mantle in *C. gigas*²². In *P. placenta*, one of the perlucin family was identified by qRT-PCR and showed a highest expression in the digestive gland than that in the other tissues (Fig. 10e). The expression level of c81494_g1 unigene (calmodulin-like) was the highest in the mouthpart, moderate in the gills, very low in the mantle tissue (Fig. 10f). The expression of carbonic anhydrase-like was the highest level in the gill tissue, medium in the mantle in *P. placenta* (Fig. 10g). Based on the qRT-PCR and in vitro crystallization results in the reported literature, we conclude that the above mentioned three unigenes, perlucin, calmodulin-like and carbonic anhydrase-like unigenes are potentially related to the biomineralization process of *P. placenta*.

As a member of the calmodulin subfamily of EF-hand Ca²⁺-binding proteins, Caltractin was first identified in *C. gigas*^{56,98}. Another unigene calreticulin is also a calcium-binding protein, it was primarily involved in the unfolded protein response to cellular stress (temperature, salinity, air exposure and heavy metals) in the endoplasmic reticulum^{14,56,99}. Both of the two calcium-binding proteins, calreticulin-like and caltractin-like exhibit relatively low expression in the mantle tissue according to the qRT-PCR results (Fig. 10h,i). However, we consider that they may probably have some kind of function for biomineralization process since they are calcium-binding proteins^{100–102}.

Conclusions

In conclusion, the transcript dataset of the mantle tissue of *P. placenta* was investigated in details by using Illumina HiSeq™ 4000 platform and public unigene databases. The identified and annotated unigenes provide valuable genomic resources for the understanding of the biomineralization mechanism. More than half of the annotated unigenes of the mantle tissue of *P. placenta* are consistent with those proteins from the pacific oyster *C. gigas* according to nr database. The transcripts of mantle tissue of *P. placenta* were identified with SNP, SSR and indel markers. These SNP markers, SSR markers and primers may be used in the construction of a genetic linkage map and gene-based association studies. 66 homologous unigenes of 21 shell matrix proteins in the transcriptome of mantle tissue of *P. placenta* were found to be related to the calcitic shell formation, while eighteen of the above unigenes are highly expressed with FPKM larger than 15 in the mantle tissue. Furthermore, qRT-PCR analysis for ten of highly expressed homologous unigenes (FPKM > 50) related to biomineralization from six different tissues of *P. placenta* indicate that seven of them are potentially related to the biomineralization process of the calcitic shells of *P. placenta*. Especially, the qRT-PCR analysis shows that four of ten examined unigenes including teneurin-2, gigasin-2, pif-like, tyrosinase-like unigenes have the highest expression levels in the mantle tissue than the levels in the other tissues, indicating their primary functions for biomineralization process. This study can contribute to the understanding of the molecular mechanisms and the functional components of the proteins that involve the biomineralization process of the calcite foliated plates of *P. placenta*. The transcriptomic data generated in this study provide a basis for further studies of *P. placenta* genome. Moreover, the comparison of potential biomineralization genes also reveals the similarities and differences between shell formation matrix of different molluscan animals.

Data availability

All relevant data reported here are included in the main section of the manuscript or in the supplementary materials. All described materials are available upon request.

Received: 15 December 2021; Accepted: 4 March 2022

Published online: 18 March 2022

References

1. Talham, D. R. Biomineralization: Principles and concepts in bioinorganic materials chemistry Stephen Mann. *Cryst. Growth Des.* **2**, 675–675 (2002).
2. Zeichner-David, M. *et al.* Biosynthesis and characterization of rabbit tooth enamel extracellular-matrix proteins. *Biochem. J.* **251**, 631–641 (1988).
3. Gize, A. On biomineralization. *Mineral. Mag.* **54**, 344–345 (1990).
4. Jackson, D. *et al.* A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol.* **4**, 40 (2006).
5. Marin, F. & Luquet, G. Molluscan shell proteins. *C. R. Palevol.* **3**, 469–492 (2004).
6. Zou, Z. *et al.* A hydrated crystalline calcium carbonate phase: Calcium carbonate hemihydrate. *Science* **363**, 396–400 (2019).
7. Furuhashi, T., Schwarzing, C., Miksik, I., Smrz, M. & Beran, A. Molluscan shell evolution with review of shell calcification hypothesis. *Comp. Biochem. Phys. B.* **154**, 351–371 (2009).
8. Chateigner, D., Hedegaard, C. & Wenk, H.-R. Mollusc shell microstructures and crystallographic textures. *J. Struct. Geol.* **22**, 1723–1735 (2000).
9. Belcher, A. *et al.* Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature* **381**, 56–58 (1996).
10. Li, S. *et al.* Transcriptome and biomineralization responses of the pearl oyster *Pinctada fucata* to elevated CO₂ and temperature. *Sci. Rep.* **6**, 18943 (2016).
11. Gardner, L. D., Mills, D., Wiegand, A., Leavesley, D. & Elizur, A. Spatial analysis of biomineralization associated gene expression from the mantle organ of the pearl oyster *Pinctada maxima*. *BMC Genomics* **12**, 455 (2011).
12. Zhang, C. & Zhang, R. Matrix proteins in the outer shells of molluscs. *Mar. Biotechnol.* **8**, 572–586 (2006).
13. Kocot, K., Aguilera, F., McDougall, C., Jackson, D. & Degnan, B. Sea shell diversity and rapidly evolving secretomes: Insights into the evolution of biomineralization. *Front. Zool.* **13**, 23 (2016).
14. Yang, Y. *et al.* The ER-localized Ca²⁺-binding protein calreticulin couples ER stress to autophagy by associating with microtubule-associated protein 1A/1B light chain 3. *J. Biol. Chem.* **294**, 772–782 (2019).
15. Weiss, I. M., Kaufmann, S., Mann, K. & Fritz, M. Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the Mollusc *Haliotis laevigata*. *Biochem. Biophys. Res. Commun.* **267**, 17–21 (2000).
16. Tsukamoto, D., Sarashina, I. & Endo, K. Structure and expression of an unusually acidic matrix protein of pearl oyster shells. *Biochem. Biophys. Res. Commun.* **320**, 1175–1180 (2004).
17. Kong, Y. *et al.* Cloning and characterization of Prsilkin-39, a novel matrix protein serving a dual role in the prismatic layer formation from the oyster *Pinctada fucata*. *J. Biol. Chem.* **284**, 10841–10854 (2009).
18. Chen, Y. *et al.* Transcriptional regulation of the matrix protein Shematin-2 during shell formation in pearl oyster. *J. Biol. Chem.* **293**, 17803–17816 (2018).
19. Pan, C. *et al.* A novel acidic matrix protein, PfN44, stabilizes magnesium calcite to inhibit the crystallization of aragonite. *J. Biol. Chem.* **289**, 2776–2787 (2014).
20. Miyamoto, H. *et al.* A carbonic anhydrase from the nacreous layer in oyster pearls. *Proc. Natl. Acad. Sci.* **93**, 9657–9660 (1996).
21. Kong, W., Li, S., Xiang, L., Xie, L. & Zhang, R. Calcium carbonate mineralization mediated by in vitro cultured mantle cells from *Pinctada fucata*. *Biochem. Biophys. Res. Commun.* **463**, 1053–1058 (2015).
22. Ivanina, A. V., Falfushynska, H. I., Beniash, E., Piontkivska, H. & Sokolova, I. M. Biomineralization-related specialization of hemocytes and mantle tissues of the Pacific oyster *Crassostrea gigas*. *J. Exp. Biol.* **220**, 3209–3221 (2017).
23. Mann, K., Weiss, I. M., Andre, S., Gabius, H. J. & Fritz, M. The amino-acid sequence of the abalone (*Haliotis laevigata*) nacre protein perlucin—Detection of a functional C-type lectin domain with galactose/mannose specificity. *Eur. J. Biochem.* **267**, 5257–5264 (2000).
24. Miyamoto, H., Miyoshi, F. & Kohno, J. The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*. *Zool. Sci.* **22**, 311–315 (2005).
25. Gotliv, B. A. *et al.* Asprich: A novel aspartic acid-rich protein family from the prismatic shell matrix of the bivalve *Atrina rigida*. *ChemBioChem* **6**, 304–314 (2005).
26. Carpenter, K. E. & Niem, V. H. The living marine resources of the Western Central Pacific. *F. A. O.* **19**, 359–363 (1998).
27. Runnegar, B. Crystallography of the foliated calcite shell layers of bivalve molluscs. *Alcheringa* **8**, 273–290 (1984).

28. Checa, A. G., Esteban-Delgado, F. J. & Rodriguez-Navarro, A. B. Crystallographic structure of the foliated calcite of bivalves. *J. Struct. Biol.* **157**, 393–402 (2007).
29. Li, L. & Ortiz, C. Biological design for simultaneous optical transparency and mechanical robustness in the shell of *Placuna placenta*. *Adv. Mater.* **25**, 2344–2350 (2013).
30. Li, L. & Ortiz, C. Pervasive nanoscale deformation twinning as a catalyst for efficient energy dissipation in a bioceramic armour. *Nat. Mater.* **13**, 501–507 (2014).
31. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
32. Shi, Y. *et al.* Transcriptome analysis of mantle tissues reveals potential biomineralization-related genes in *Tectus pyramis* Born. *Comp. Biochem. Phys. D.* **29**, 131–144 (2019).
33. Wang, X., Liu, Z. & Wu, W. Transcriptome analysis of the freshwater pearl mussel (*Cristaria plicata*) mantle unravels genes involved in the formation of shell and pearl. *Mol. Genet. Genomics.* **292**, 343–352 (2017).
34. Li, H. *et al.* Characterization of transcriptome and identification of biomineralization genes in winged pearl oyster (*Pteria penguin*) mantle tissue. *Comp. Biochem. Phys. D.* **21**, 67–76 (2017).
35. Shi, Y. *et al.* Characterization of the pearl oyster (*Pinctada martensii*) mantle transcriptome unravels biomineralization genes. *Mar. Biotechnol.* **15**, 175–187 (2013).
36. Sun, X., Yang, A., Wu, B., Zhou, L. & Liu, Z. Characterization of the mantle transcriptome of yesso scallop (*Patinopecten yes-soensis*): identification of genes potentially involved in biomineralization and pigmentation. *PLoS ONE* **10**, 4 (2015).
37. Shi, M. *et al.* Characterization of the Zhikong Scallop (*Chlamys farreri*) mantle transcriptome and identification of biomineralization-related genes. *Mar. Biotechnol.* **15**, 706–715 (2013).
38. Freer, A., Bridgett, S., Jiang, J. & Cusack, M. Biomineral proteins from *Mytilus edulis* mantle tissue transcriptome. *Mar. Biotechnol.* **16**, 34–45 (2014).
39. McGinty, E., Zenger, K., Jones, D. & Jerry, D. Transcriptome analysis of biomineralisation-related genes within the pearl sac: Host and donor oyster contribution. *Mar. Genomics.* **5**, 27–33 (2012).
40. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
41. Iseli, C., Jongeneel, C. & Bucher, P. ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Int. Conf. Intell. Syst. Mol. Biol.* **99**, 138–148 (1999).
42. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome. Res.* **20**, 1297–1303 (2010).
43. Suzuki, M. *et al.* Characterization of Prismaticin-14, a novel matrix protein from the prismatic layer of the Japanese pearl oyster (*Pinctada fucata*). *Biochem. J.* **382**, 205–213 (2004).
44. Suzuki, M. *et al.* An acidic matrix protein, pif, is a key macromolecule for nacre formation. *Science* **325**, 1388–1390 (2009).
45. Su, J. *et al.* Structural characterization of amorphous calcium carbonate-binding protein: An insight into the mechanism of amorphous calcium carbonate formation. *Biochem. J.* **453**, 179–186 (2013).
46. Kalamajski, S., Aspberg, A., Lindblom, K., Heinegård, D. & Oldberg, Å. Asporin competes with decorin for collagen binding, binds calcium and promotes osteoblast collagen mineralization. *Biochem. J.* **423**, 53–59 (2009).
47. Punta, M. *et al.* The Pfam protein families database. *Nucleic. Acids. Res.* **40**, D290–D301 (2012).
48. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic. Acids. Res.* **39**, W29–W37 (2011).
49. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT method. *Methods* **25**, 402–408 (2001).
50. Liu, J. *et al.* Microarray: a global analysis of biomineralization-related gene expression profiles during larval development in the pearl oyster, *Pinctada fucata*. *BMC Genomics* **16**, 15 (2015).
51. Wang, X. T. *et al.* Oyster shell proteins originate from multiple organs and their probable transport pathway to the shell formation front. *PLoS ONE* **8**, 6 (2013).
52. Mount, A. & Pickering, C. M. Testing the capacity of clothing to act as a vector for non-native seed in protected areas. *J. Environ. Manag.* **91**, 168–179 (2009).
53. Mount, A. S., Wheeler, A. P., Paraskar, R. P. & Snider, D. Hemocyte-mediated shell mineralization in the eastern oyster. *Science* **304**, 297–300 (2004).
54. Zhang, L. & Guo, X. Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. *Aquaculture* **302**, 124–129 (2010).
55. Song, X. R., Liu, Z. Q., Wang, L. L. & Song, L. S. Recent advances of shell matrix proteins and cellular orchestration in marine molluscan shell biomineralization. *Front. Mar. Sci.* **6**, 16 (2019).
56. Zhang, G. F. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
57. 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. Evol.* **30**, 2099–2112 (2013).
58. Drake, J. L. *et al.* Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*. *Proc. Natl. Acad. Sci.* **110**, 3788–3793 (2013).
59. Luo, Y.-J. *et al.* The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat. Commun.* **6**, 8301 (2015).
60. Nong, W. *et al.* Genome of the ramshorn snail *Biomphalaria straminea*—an obligate intermediate host of schistosomiasis. *GigaScience* **11** (2022).
61. Sarashina, I. Skeletal matrix proteins of invertebrate animals: Comparative analysis of their amino acid sequences. *Paleont. Res.* **10**, 311–336 (2009).
62. Gorbushin, A. M. & Borisova, E. A. Lectin-like molecules in transcriptome of *Littorina littorea* hemocytes. *Dev. Comp. Immunol.* **48**, 210–220 (2015).
63. McGinty, E. L., Zenger, K. R., Jones, D. B. & Jerry, D. R. Transcriptome analysis of biomineralisation-related genes within the pearl sac: Host and donor oyster contribution. *Mar. Genom.* **5**, 27–33 (2012).
64. Zakrzewski, A. C. *et al.* Early divergence, broad distribution, and high diversity of animal chitin synthases. *Genome Biol. Evol.* **6**, 316–325 (2014).
65. Zhang, S.-M. & Loker, E. Representation of an immune responsive gene family encoding fibrinogen-related proteins in the freshwater mollusc *Biomphalaria glabrata*, an intermediate host for *Schistosoma mansoni*. *Gene* **341**, 255–266 (2004).
66. Zheng, L. *et al.* Ferritin has an important immune function in the ark shell *Scapharca broughtonii*. *Dev. Comp. Immunol.* **59**, 15–24 (2016).
67. Xie, J. *et al.* Molecular and functional characterization of ferritin in abalone *Haliotis diversicolor supertexta*. *Acta. Oceanologica Sinica.* **31** (2012).
68. Munro, H. *et al.* The ferritin genes: structure, expression, and regulation. *Ann. N. Y. Acad. Sci.* **526**, 113–123 (2006).
69. Lin, J. Y., Ma, K. Y., Bai, Z. Y. & Li, J. L. Molecular cloning and characterization of perlucin from the freshwater pearl mussel, *Hyriopsis cumingii*. *Gene* **526**, 210–216 (2013).
70. Asada, T., Ashiba, G. & Watanabe, S. Calcium sensitivity of foot muscle myosin from clam (*Meretrix lusoria*). *J. Biochem.* **85**, 1543–1546 (1979).

71. Collins, J., Johnson, J. & Szent-Györgyi, A. Purification and characterization of a scallop sarcoplasmic calcium-binding protein. *Biochemistry* **22**, 341–345 (1983).
72. Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* **1**, 120 (2017).
73. Yang, J. *et al.* Transcriptome analysis of growth heterosis in pearl oyster *Pinctada fucata martensii*. *FEBS Open Bio* **8**, 1794–1803 (2018).
74. Megrabian, Z. B. & Nalbandian, R. M. Copper-containing amine oxidases of blood vessels (review of the literature). *Vopr. Med. Khim.* **33**, 13–18 (1987).
75. Piomelli, D., Shapiro, E., Feinmark, S. J. & Schwartz, J. H. Metabolites of arachidonic acid in the nervous system of *Aplysia*: possible mediators of synaptic modulation. *J. Neurosci.* **7**, 3675–3686 (1987).
76. Woolstra, C. R. *et al.* Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals. *Sci. Rep.* **7**, 17583 (2017).
77. Wang, X. *et al.* Evolution and functional analysis of the Pif97 gene of the Pacific oyster *Crassostrea gigas*. *Curr. Zool.* **59**, 109–115 (2013).
78. Feng, D., Li, Q., Yu, H., Kong, L. & Du, S. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. *Sci. Rep.* **7**, 45754 (2017).
79. Oliveira, D. V., Silva, T. S., Cordeiro, O. D., Cavaco, S. I. & Simes, D. C. Identification of proteins with potential osteogenic activity present in the water-soluble matrix proteins from *Crassostrea gigas* nacre using a proteomic approach. *Sci. World J.* **2012**, 765909–765909 (2012).
80. Marie, B., Le Roy, N., Zanella-Cléon, I., Becchi, M. & Marin, F. Molecular evolution of mollusc shell proteins: Insights from proteomic analysis of the edible mussel *Mytilus*. *J. Mol. Evol.* **72**, 531–546 (2011).
81. Stenflo, J., Stenberg, Y. & Muranyi, A. Calcium-binding EGF-like modules in coagulation proteinases: Function of the calcium ion in module interactions. *Biochim. Biophys. Acta.* **1477**, 51–63 (2000).
82. Marie, B. *et al.* The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage specific novelties. *Febs. J.* **280**, 214–232 (2012).
83. Jovine, L., Darie, C., Litscher, E. & Wassarman, P. Zona pellucida domain proteins. *Annu. Rev. Biochem.* **74**, 83–114 (2005).
84. Bjarnmark, N. A. *et al.* Transcriptomics provides insight into *Mytilus galloprovincialis* (Mollusca: Bivalvia) mantle function and its role in biomineralisation. *Mar. Genomics.* **27**, 37–45 (2016).
85. 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).
86. Bahn, S. Y., Jo, B. H., Choi, Y. S. & Cha, H. J. Control of nacre biomineralization by Pif80 in pearl oyster. *Sci. Adv.* **3**, e1700765 (2017).
87. Nelson, M. R. & Chazin, W. J. Structures of EF-hand Ca²⁺-binding proteins: Diversity in the organization, packing and response to Ca²⁺ binding. *Biometals* **11**, 297–318 (1998).
88. Özbek, S., Balasubramanian, P., Chiquet-Ehrismann, R., Tucker, R. & Adams, J. The evolution of extracellular matrix. *Mol. Biol. Cell.* **21**, 4300–4305 (2010).
89. Miyashita, T., Takagi, R., Miyamoto, H. & Matsushiro, A. Identical carbonic anhydrase contributes to nacreous or prismatic layer formation in *Pinctada fucata* (Mollusca: Bivalvia). *Veliger* **45**, 250–255 (2002).
90. Deng, Y. *et al.* De novo assembly, gene annotation, and simple sequence repeat marker development using Illumina paired-end transcriptome sequences in the pearl oyster *Pinctada maxima*. *Biosci. Biotechnol. Biochem.* **78**, 1685–1692 (2014).
91. Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on illumina HiSeq and genome analyzer systems. *Genome Biol.* **12**, R112 (2011).
92. Nagai, K., Yano, M., Morimoto, K. & Miyamoto, H. Tyrosinase localization in mollusc shells. *Comp. Biochem. Phys. B.* **146**, 207–214 (2007).
93. 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).
94. Blank, S. *et al.* The nacre protein perlucin nucleates growth of calcium carbonate crystals. *J. Microsc.* **212**, 280–291 (2003).
95. Yan, Z. *et al.* Biomineralization: functions of calmodulin-like protein in the shell formation of pearl oyster. *Biochim. Biophys. Acta.* **1770**, 1338–1344 (2007).
96. 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).
97. Roy, N. L. E. *et al.* Identification of two carbonic anhydrases in the mantle of the European Abalone *Haliotis tuberculata* (Gastropoda, Haliotidae): phylogenetic implications. *J. Exp. Zool. B* **318**, 353–367 (2012).
98. Veeraraghavan, S. *et al.* Structural independence of the two EF-hand domains of caltractin *. *J. Biol. Chem.* **277**, 28564–28571 (2002).
99. Fan, W., Hu, Y., Li, C., Xie, L. & Zhang, R. Cloning, Characterization, and expression analysis of Calreticulin from pearl oyster *Pinctada fucata*. *Tsinghua. Sci. Technol.* **13**, 466–473 (2008).
100. Hattan, S. J., Laue, T. M. & Chasteen, N. D. Purification and characterization of a novel calcium-binding protein from the extrapallial fluid of the mollusc, *Mytilus edulis*. *J. Biol. Chem.* **276**, 4461–4468 (2001).
101. Duplat, D. *et al.* Identification of calconectin, a calcium-binding protein specifically expressed by the mantle of *Pinctada margaritifera*. *Febs. Lett.* **580**, 2435–2441 (2006).
102. Liu, H. L. *et al.* Identification and characterization of a biomineralization related gene PFMG1 highly expressed in the mantle of *Pinctada fucata*. *ACS. Biochem.* **46**, 844–851 (2007).

Acknowledgements

This research was funded by National Natural Science Foundation of China (Grant No. 21877009), and Beijing Institute of Technology Research Fund Program for Young Scholars. This work was also supported by the facilities at the Analysis & Testing Center, Beijing Institute of Technology.

Author contributions

N.S. performed the main experiments, data analysis, and wrote the paper. Y.M. designed the experiments and revised the manuscript. All authors read and approved the final manuscript.

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/s41598-022-08610-5>.

Correspondence and requests for materials should be addressed to Y.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022