RESEARCH ARTICLE

De novo assembly and comparative transcriptome analysis of the foot from Chinese green mussel (*Perna viridis*) in response to cadmium stimulation

Xinhui Zhang[1,2☯], Zhiqiang Ruan[2☯], Xinxin You[2], Jintu Wang[3], Jieming Chen[2], Chao Peng[2], Qiong Shi[1,2]*

1 Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China, 2 Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Fisheries, BGI, Shenzhen, China, 3 BGI-Shenzhen, BGI, Shenzhen, China

☯ These authors contributed equally to this work.
* shiqiong@genomics.cn

OPEN ACCESS

Abstract

The Chinese green mussel, *Perna viridis*, is a marine bivalve with important economic values as well as biomonitoring roles for aquatic pollution. Byssus, secreted by the foot gland, has been proved to bind heavy metals effectively. In this study, using the RNA sequencing technology, we performed comparative transcriptomic analysis on the mussel feet with or without inducing by cadmium (Cd). Our current work is aiming at providing insights into the molecular mechanisms of byssus binding to heavy metal ions. The transcriptome sequencing generated a total of 26.13-Gb raw data. After a careful assembly of clean data, we obtained a primary set of 105,127 unigenes, in which 32,268 unigenes were annotated. Based on the expression profiles, we identified 9,048 differentially expressed genes (DEGs) between Cd treatment (50 or 100 μg/L) at 48 h and the control, suggesting an extensive transcriptome response of the mussels during the Cd stimulation. Moreover, we observed that the expression levels of 54 byssus protein coding genes increased significantly after the 48-h Cd stimulation. In addition, 16 critical byssus protein coding genes were picked for profiling by quantitative real-time PCR (qRT-PCR). Finally, we reached a primary conclusion that high content of tyrosine (Tyr), cysteine (Cys), histidine (His) residues or the special motif plays an important role in the accumulation of heavy metals in byssus. We also proposed an interesting model for the confirmed byssal Cd accumulation, in which biosynthesis of byssus proteins may play simultaneously critical roles since their transcription levels were significantly elevated.

Abbreviations: ALP, antistasin-like protein; BPTI/Kunitz, BPTI/Kunitz domain-containing protein; CBGO5433, Hypothetical protein CBG05433; Cd, Cadmium; COG, Cluster of Orthologous Groups of proteins; Cu, Copper; CYSRP, Sporozoite cysteine-rich protein; DEGs, Differentially expressed genes; EGF, Epidermal growth factor; FDR, False discovery rate; Fe, Iron; GO, Gene Ontology Consortium; KEGG, Kyoto Encyclopedia of Genes and Genomes; Mn, Manganese; Mucin-2-like, Mucin-2-like; Mfp, Mussel foot protein; NF70, Neurofilament protein NF70; Nr, Non-redundant protein database; Nt, Non-redundant nucleotide database; nt, nucleotide; Precol-D, Precollagen-D; Precol-NG, Precollagen-NG; PTMP1, Proximal thread matrix protein 1; PTMP1b, Proximal thread matrix protein 1b; Pvfp, *Perna viridis* foot protein; qRT-PCR, Quantitative real-time PCR; RIN, RNA Integrity Number; RNA-seq, RNA sequencing; RPKM, Reads per kb per Million reads; SCO-spondin, SCO-spondin-like; SHyTSR1, similar to HyTSR1; SHyTSR1-protein, Similar to HyTSR1 protein; STUM14, Suppressor of tumorigenicity 14 protein homolog; Swiss-prot, Swiss-protein database; SOD-like, Superoxide dismutase-like; STFF, similar to trefoil factor; THCP, Thioester-containing protein; THIH, Thrombin inhibitor haemalin; TYR, Tyrosinase; Zn, Zinc.

Introduction

Mussels are widely used for biomonitoring of coastal environments with respect to the impact of heavy metal contamination [1–4]. The Chinese green mussel, *Perna viridis*, has been applied as a biomonitor throughout the Indo-Pacific region for assessing a wide range of heavy metals [5, 6]. In laboratory conditions, the accumulation rates of cadmium (Cd) and zinc (Zn) in tissues were reported to be different, and this might be due to differential mechanisms of metal binding and regulation; compared with other tissues, byssus had been outstanding since it always owned the highest accumulation rates [7].

As we know, sessile mussels can survive in fluctuant and turbulent ocean environments, because they use byssal threads to attach themselves to hard surfaces in the sea [8]. Byssus, secreted from the foot gland, consists of fibre and adhesive coating proteins that cooperatively provide tenacity, although cyclic biological, mechanical and chemical stresses arise from changes in the salinity, temperature or other factors of seawater [9]. It has been reported that the byssus is composed of several important proteins, including foot proteins (*Fp-1~6*) [10], precollagens (*PreCol-D*, *PreCol-NG* and *PreCol-P*) [11–14], tyrosinases (*TYR1~5*) [15], and proximal thread matrix protein (*PTMP1*) [16, 17]. By far, we have obtained 252 protein sequences by liquid chromatography-tandem mass spectrometry (LC-MS/MS) from the *P. viridis* byssus (unpublished data), in which many are not reported. Current studies have focused on the formation of byssal proteins and adhesion mechanism from bivalvia. For example, Miao et al. [18] applied a integration of transcriptomic and proteomic approaches to explore a core set of genes from scallop attachment, in which 75 foot-specific and 7 unique scallop byssal proteins were identified. Lee et al. [19] found that there was the unusual amino acid 3,4-dihydroxy-L-phenylanine (dopa) in the mussel foot proteins, which is a critical substance for adhesion. However, metal enrichment in byssal proteins was rarely reported. Virtually nothing is known about the potential regulatory mechanism, and lack of molecular information blocked the investigations to reveal the critical regulatory genes.

RNA sequencing (RNA-Seq) has been achieved widespread as a revolutionary tool for transcriptomic studies [20, 21], since it can reveal molecular basis of functional responses to environmental changes [22, 23] and underlying molecular mechanisms of an organism against environmental stresses [24, 25]. Here, we performed an Illumina sequencing strategy and reported a *de novo* assembly of the foot transcriptomes from Chinese green mussel with or without inducing by Cd. The novel data significantly enrich the overall understanding of the mussel biology, with a focus on the byssus, a tissue known to accumulate heavy metal ions. The potential usefulness of the achieved reference transcriptomes has been explored by comparison of transcription profiles of byssus proteins, which supports an interesting model to reveal increased biosynthesis of byssus proteins during the Cd inducing process.

Materials and methods

Sample collection

P. viridis was purchased from Yantian seafood Market in Shenzhen, China (Latitude 22˚35' N and Longitude 114˚17' E), which individual size are not much difference, and the appearance are not damaged. All the mussels were collected from the same location and chosen with similar size and weight (mean shell length 7.32±0.19 cm, mean fresh weight 24.46±0.89 g). Before Cd treatment, these mussels were scrubbed to remove surface sediments, and acclimated in running seawater (24˚C and 28 ‰) for two weeks. All experiments were performed in accordance with the guidelines of the Animal Ethics Committee and approved by the Institutional Review Board on Bioethics and Biosafety of BGI.

Experimental design

The mussels were randomly divided into three groups and placed in three pre-cleaned tanks containing aerated artificial seawater (24°C and 28‰), and each tank contained 12 mussels. One group was set as control (without Cd stimulation), the second group was exposed to 50 μg/L of cadmium chloride ($CdCl_2$), and the third group was exposed to 100 μg/L of $CdCl_2$. The seawater with or without $CdCl_2$ was changed daily (9:00 AM). The mussels were dissected then at 24, 48 or 72 hours (three groups). Three samples including the controls were used for each time point and each treatment group.

Cd accumulation in the byssus samples

The byssus samples from the Control and treatment groups (see more details in **Table 1**) were harvested and stored at -20°C refrigerator for detection of Cd accumulation. In brief, the individual sample was placed in a digestion tank containing 3 ml of 67% nitric acid and 2 ml of 30% hydrogen peroxide, and then digested by a microwave digestion instrument (CEM, USA). After digestion, the samples were separately diluted with purified water to a final volume of 40 ml. Subsequently, the samples were measured by inductively coupled plasma mass spectrometry (ICP-MS; Agilent, USA) with the manufacturer's Mass-Hunter Workstation software (version B.07.00). The concentration of Cd in each byssus sample was quantified with the calibration curve by plotting the ratio of Cd between the treatment and the corresponding control.

Extraction of total RNA

Feet of the mussels were collected, snap frozen in liquid nitrogen and then stored at -80°C before use. Total RNA was extracted from each sample using TRIzol Reagent (Invitrogen, Life Technologies, USA) in accordance with the manufacturer's instructions. The RNA samples were subsequently treated with RNase-Free DNase I (Qiagen, USA) to eliminate genomic DNA. Finally, high-quality RNA samples (RNA concentration≥1200 ng/μL, RNA Integrity Number≥9.0) were used for further construction of cDNA libraries.

Table 1. Summary of generated transcriptome data.

Samples	Raw reads (Number)	Raw data (Gb)	Clean reads (Number)	Clean data (Gb)
0ug/L 24h (Control-24)	18,096,389	0.85	17,960,666	0.84
0ug/L 48h (Control-48)	21,673,770	1.01	21,431,075	1.00
0ug/L 72h (Control-72)	19,222,198	0.90	19,133,838	0.89
50ug/L 24h (Cd_50–24)	17,853,671	0.83	17,683,479	0.83
50ug/L 48h (Cd_50–48)	21,088,268	0.99	20,964,735	0.98
50ug/L 72h (Cd_50–72)	19,051,903	0.89	18,969,849	0.89
100ug/L 24h (Cd_100–24)	19,617,933	0.92	19,467,801	0.91
100ug/L 48h (Cd_100–48)	18,309,288	0.86	18,197,897	0.85
100ug/L 72h (Cd_100–72)	18,814,565	0.88	18,677,331	0.87
All_sample[a]	105,090,368	18.03	102,109,334	17.52
Total	278,818,353	26.13	274,596,005	25.53

Note

[a]All_sample, the mixture of nine treated samples (at 24, 48 and 72 h) with equal amount of total RNA, was sequenced by a 90-bp pair-end sequencing, while other samples were applied by a 49-bp single-end sequencing.

https://doi.org/10.1371/journal.pone.0176677.t001

Transcriptome data analysis

Ten cDNA libraries were constructed (Table 1) and sequenced using Illumina HiSeq[TM] 2000 platform (San Diego, CA, USA). We filtered raw reads to remove adaptors, reads with more than 5% unknown nucleotides, and other low-quality reads. Short-reads assembly software SOAPdenovo-Trans [26] was applied for transcriptome *de novo* assembly. The clean reads were then mapped back to contigs, the longest assembled sequences. Finally, we obtained unigenes (more than 150 nucleotides), which were clustered using Tgicl [27] to reduce redundancy and used for downstream analysis.

BLASTx or BLASTn alignment (e-value$\leq 10^{-5}$) was performed to search those achieved unigenes from six public databases, including Non-redundant (Nr), Swiss-prot protein (Swiss-Prot), Kyoto Encyclopedia of Gene and Genomes (KEGG), Cluster of Orthologous Group of proteins (COG) and Non-redundant nucleotide database (Nt). For prediction of unigene functions, we used Blast2GO [28] program to annotate unigenes and obtain corresponding Gene Ontology Consortium (GO) annotation for each unigene. Based on the GO annotation, we applied WEGO [29] to classify GO functions for unigenes and calculate the overall distribution of gene functions. The pathway annotation for the achieved unigenes was also obtained under the KEGG annotation.

We then used the former blast results (in addition to the Nt blast results) to extract coding sequences (CDs) from the achieved unigene sequences and translated them into peptide sequences. CDs of unigenes without any hit in the blast were further predicted by ESTScan [30] and were also translated into peptide sequences.

The expression level of each unigene was calculated using RPKM (Reads per kb per million reads) method [31] with the formula RPKM = $(1000000*C)/(N*L*1000)$. RPKM is the expression level of a given gene; C is the number of reads that were uniquely aligned to the gene; N is the total number of reads that were uniquely aligned to all genes; and L is the number of nucleotide bases in the gene. The RPKM method is able to eliminate the influence of different gene length and sequencing discrepancy on the calculation of gene expression. Therefore, the calculated gene expression can be directly used for comparing the differences of gene expression among comparative samples. Differentially expressed genes (DEGs) were determined by DEseq2 [32] with defaulted filtering criteria that absolute $\log_2$ (ratio)≥ 1 and false discovery rate (FDR)≤ 0.05. Finally, these DEGs were subjected to GO enrichment and KEGG pathway analyses as reported before [33, 34].

Validation of the RNA-seq analyses by qRT-PCR

Quantitative real-time PCR (qRT-PCR) was frequently used to confirm the data generated by high-throughput sequencing [35, 36]. Here, tissue-specific expression patterns of 16 representative candidate byssus protein coding genes associated with stress responses were confirmed. The Chinese green mussel foot gland samples were collected from Cd treated mussels (i.e., Control 48-h, 50 μg/L $CdCl_2$ 48-h, 100 μg/L $CdCl_2$ 48-h). qRT-PCRs were carried out using gene-specific primers (S1 Table), which were designed using Primer3 version 4.0.0 [37]. Total RNAs from the foot gland samples were extracted and purified by RNeasy Animal Mini Kit (Qiagen, Valencia, CA). All samples were examined in triplicate with β-actin as the internal control and the $2^{-\Delta\Delta Ct}$ method [38] was used to calculate the relative expression.

Results

High accumulation of Cd in the byssus samples from *P. viridis*

High accumulation of Cd in the byssus of Chinese green mussel was confirmed. The highest byssal Cd accumulation was observed in the Cd_100–48 group (16,436 ng/g), which is around
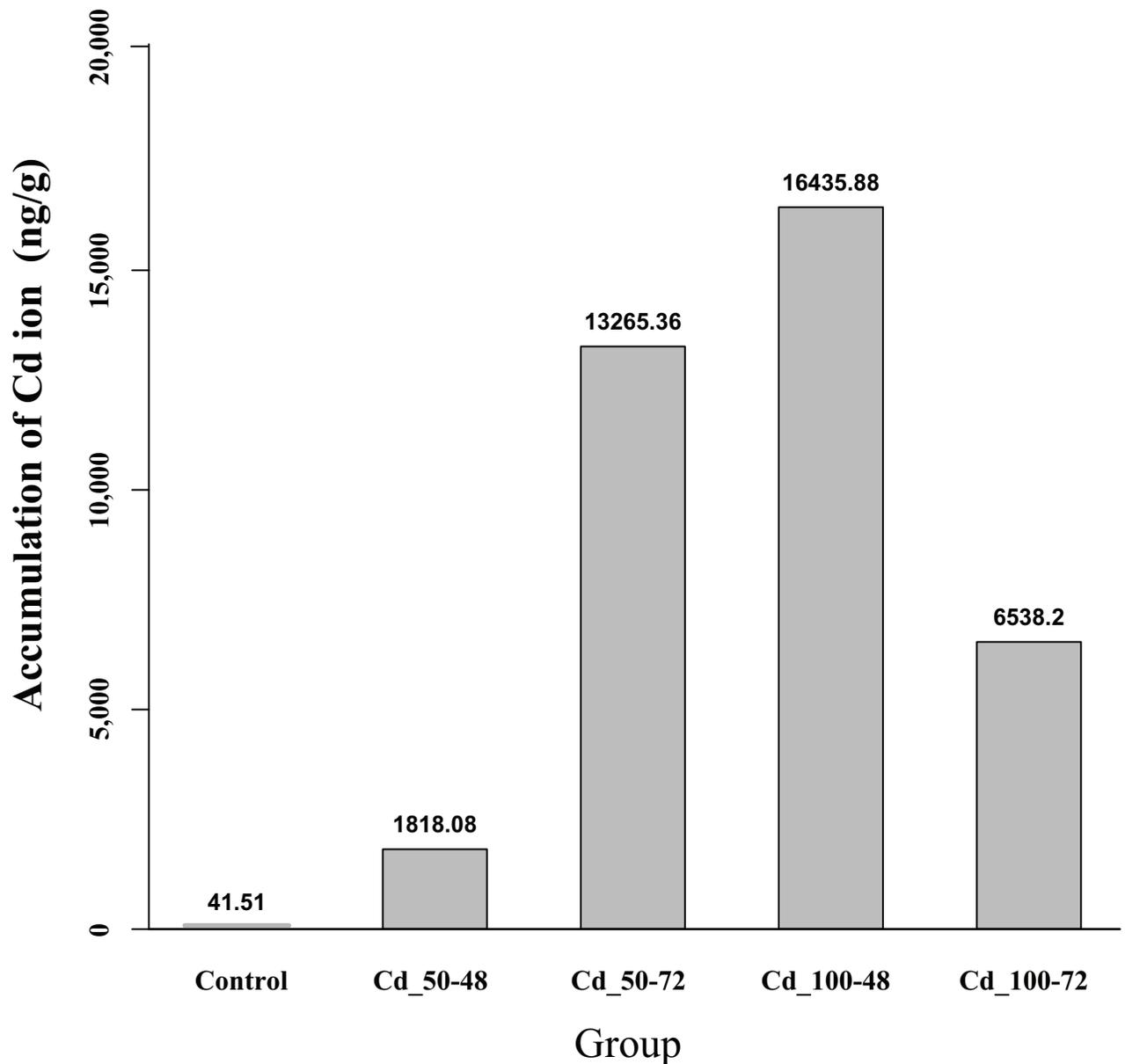
Fig 1. High accumulation of Cd in the *P. viridis* byssus.

https://doi.org/10.1371/journal.pone.0176677.g001

400 times of the Control (Fig 1). Although these results were dependent on the concentrations of Cd in the environmental water and exposure duration of Cd to the mussels, our primary data confirmed a remarkable ability of the byssus to accumulate Cd in a short period.

High throughput sequencing of the foot and related *de novo* assembly

An overall profile of the transcriptomes of *P. viridis* was obtained through Illumina RNA-seq of ten libraries (three Cd concentrations: 0, 50 and 100 μg/L $CdCl_2$ × three time points: 24, 48 and 72 h, and a mixture of all the nine samples) of mRNA samples from pooled mussel feet. In total, we generated 26.13 Gb of raw data from the ten libraries (see more details for each library in Table 1), which were deposited at the NCBI Sequence Read Archive (SRA No. SRP077042).

Table 2. Summary of the RNA-seq based assembly, functional annotation and coding sequences (CDs) from the foot transcriptomes of *P. viridis*.

Assembly	Number of unigene	105,127
	Total length of unigene (nt)	71,677,450
	N50 (nt)	1,257
	Mean length of unigene (nt) Max length of unigene (nt) Min length of unigene (nt)	682 34,811 150
Annotation	Unigenes with Nr	30,089 (28.62%)
(E-value≤1e-5)	Unigenes with Nt	9,460 (9.00%)
	Unigenes with Swiss-prot	23,060 (21.94%)
	Unigenes with KEGG	20,085 (19.11%)
	Unigenes with COG	10,573 (10.06%)
	Unigenes with GO	12,123 (11.53%)
	Total number of annotated unigenes	32,268 (30.69%)
Coding sequences	With blast hit	30,491
(CDs)	ESTscan prediction	5,397
	Total number of cds	35,888

https://doi.org/10.1371/journal.pone.0176677.t002

After removal of adaptor sequences and low-quality reads, we obtained 25.53 Gb of clean data. Subsequently, these clean data were assembled into 105,127 unigenes by SOAPdenovo-Trans [26]. Lengths of the unigenes ranged from 150 to 34,811 nt, with an average length of 682 nt and the N50 size of 1,257 nt (**Table 2**).

Annotation and functional classification of *P. viridis* foot unigenes

To predict the function classification and annotation of the *P. viridis* unigenes, we aligned the unigenes against 6 public databases using BLASTx or BLASTn (e-value≤10^{-5}). Our results indicated that a total of 32,268 unigenes could be annotated with known biological functions, and 30,089 (28.62%) unigenes matched to known proteins in the NCBI Non-redundant (Nr) protein database (**Table 2** and **Fig 2A**). The species distribution against the Nr database was summarized in **Fig 2B**, in which the majority of unigenes (57.20%) matched to the Pacific oyster (*Crassostrea gigas*). On the other hand, a total of 23,060 (21.94%) and 9,460 (9.00%) unigenes matched to known proteins in the Swiss-prot and the Nt respectively (**Table 2**).

Databases for GO and COG were used to classify functions of unigenes. Based on the COG annotation, we finally predicted 10,573 (10.06%) unigenes and classified them into 25 functional categories (**S1 Fig**). Among them, the "General function prediction only" was the most popular group (19.38%), which was followed by "Translation, ribosomal structure and biogenesis" (8.12%) and "Replication, recombination and repair" (7.92%). Based on the GO annotation, we observed that 12,123 (11.53%) unigenes were each assigned to a gene ontology term and categorized into 59 subcategories belonging to 3 main categories (**S2 Fig**), including biological process (23), cellular component (18) and molecular function (18). Binding (43.8%) and catalytic activities (35.1%) were obviously more than others in the category of molecular function; in the category of biological process, cellular process and metabolic process were obviously dominant; in the cellular component, cell and cell part was larger than others. Through BLASTx searches against the KEGG database, we annotated 20,085 (19.11%) unigenes and assigned them into 259 KEGG pathways (**S2 Table**). The representative pathways include metabolic pathway (2,620 unigenes), focal adhesion (987 unigenes), regulation of actin cytoskeleton (896 unigenes) and pathway in cancer (839 unigenes).These genes are generally
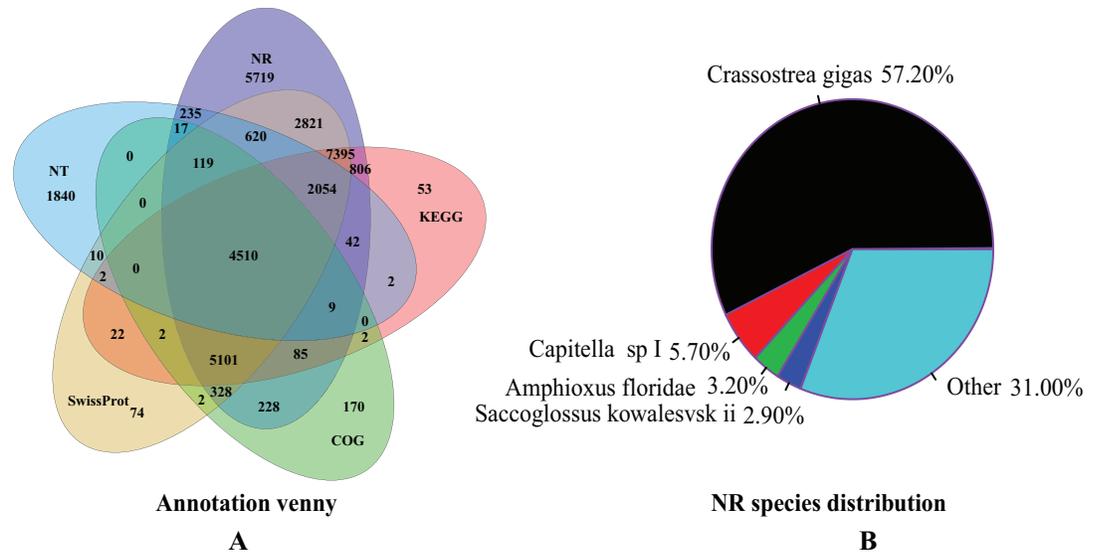
Fig 2. Annotation of unigenes from the *P. viridis* foot transcriptomes. Venn diagrams of annotated unigenes (A) and Nr species distribution of unigenes (B) are useful for assessment of the transcriptome assemblies.

involved in the binding, catalytic activities, focal adhesion and regulation of actin cytoskeleton, which are all related to byssus for playing many important physiological roles including attachment, movement and self-defense.

Coding sequences predicted for the assembled unigenes

Using the former blast results to extract coding sequences, we predicted a total of 30,491 coding sequences (CDs). Those CDs without any hit in the blast searching were subsequently predicted by ESTScan software, which generated 5,397 more CDs. Finally, a total of 35,888 CDs were predicted (**Table 2**), with lengths ranging from 100 to 14,292 nt.

Comparison of DEGs between *P. viridis* foot in response to Cd stimulation

We performed a comparative analysis to elucidate the molecular mechanisms that enable the Chinese green mussel, under heavy-metal stressed conditions, to activate the secretion of byssus as well as accumulation of heavy metals. To identify the differences among control, Cd_50 and Cd_100 at three time points (Cd treatment for 24, 48 and 72 h), we calculated gene expression levels (**S3 Table**) using the RPKM method [31]. Based on the RPKM values at the 24-h treatment, we identified 861 and 845 DEGs (with $p<0.05$, FDR<0.001, absolute log2Ratio ≥ 1) for Cd_50 and Cd_100 respectively. At the 48-h treatment, 13,405 and 12,452 DEGs were identified respectively. At the 72-h treatment, 619 and 1,440 DEGs were identified in the Cd_50 vs Control and Cd_100 vs Control respectively (**Fig 3A and 3B**). Interestingly, the number of DEGs at 48-h was more than those at 24-h and 72-h, suggesting existence of a response summit during the 48 h of Cd treatment.

At the 2nd time point (48-h) of Cd treatment, a total of 9,048 DEGs with the same trend (6,828 up- or 2,171 down-regulated genes) were common between the two comparisons (Cd_50 vs control and Cd_100 vs control) (**Fig 3C**). All these DEGs were subjected to KEGG metabolic pathway analysis, and the results indicate that 2,461 up-regulated and 305 down-regulated genes were mapped to KEGG ID [39], respectively. These genes were enriched in 42
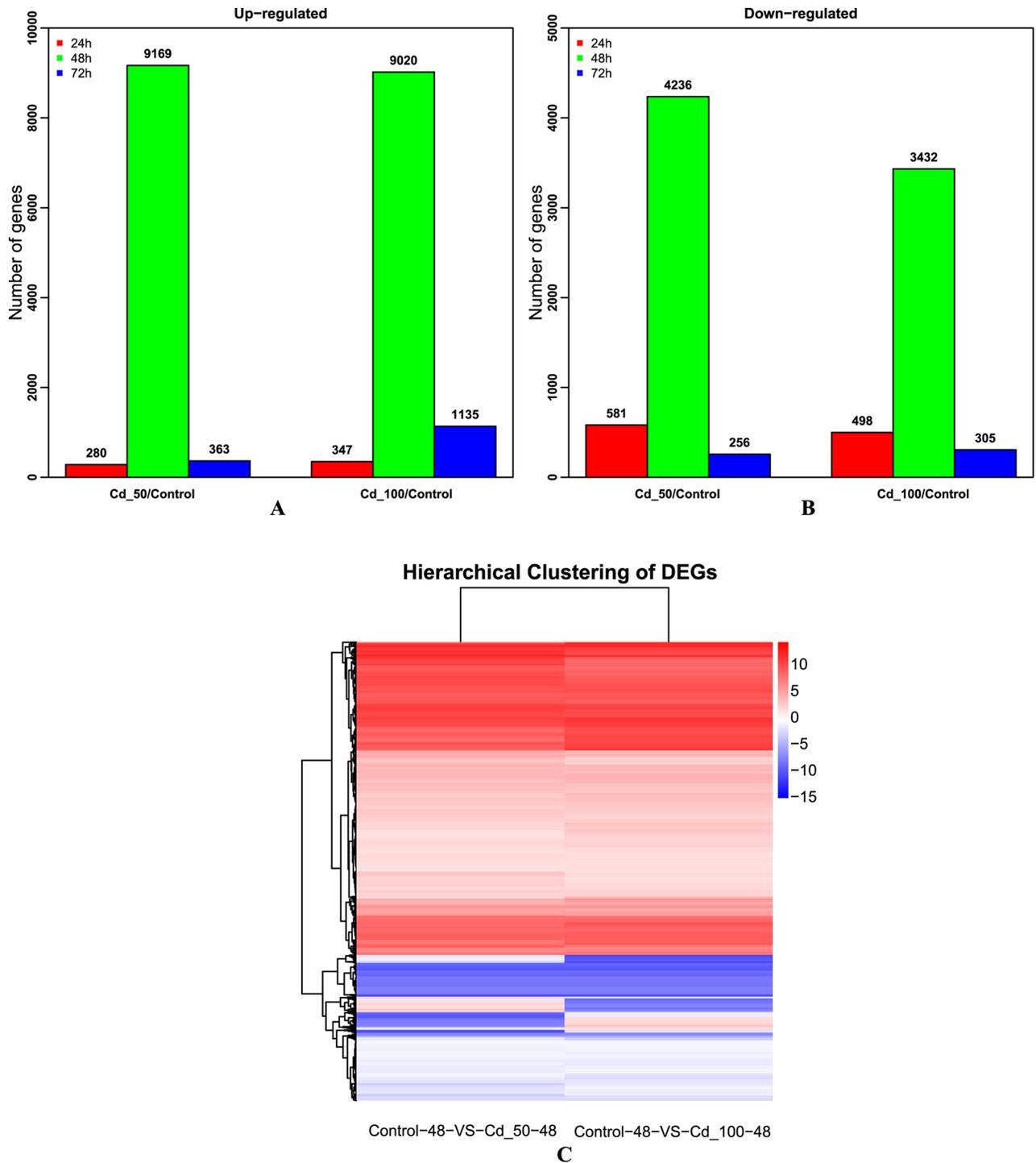
Fig 3. DEG number for Cd_50 vs Control and Cd_100 vs Control. (A) Up-regulated gene numbers at the three time points; (B) Down-regulated gene numbers at the three time points; (C) Heatmap representation of cluster analysis for the expression patterns of over 9,000 DEGs between the control and different concentrations of Cd stimulation (48-h treatment). The RPKM change patterns of the two concentration of Cd stimulation (Cd_50 and Cd_100) were shown to be clustered together.

Fig 4. Function classification of DEGs in the *P. viridis* transcriptomes (48-h treatment). (A) KEGG metabolic pathway annotation of DEGs; (B) Histogram presentations of the COG classification.
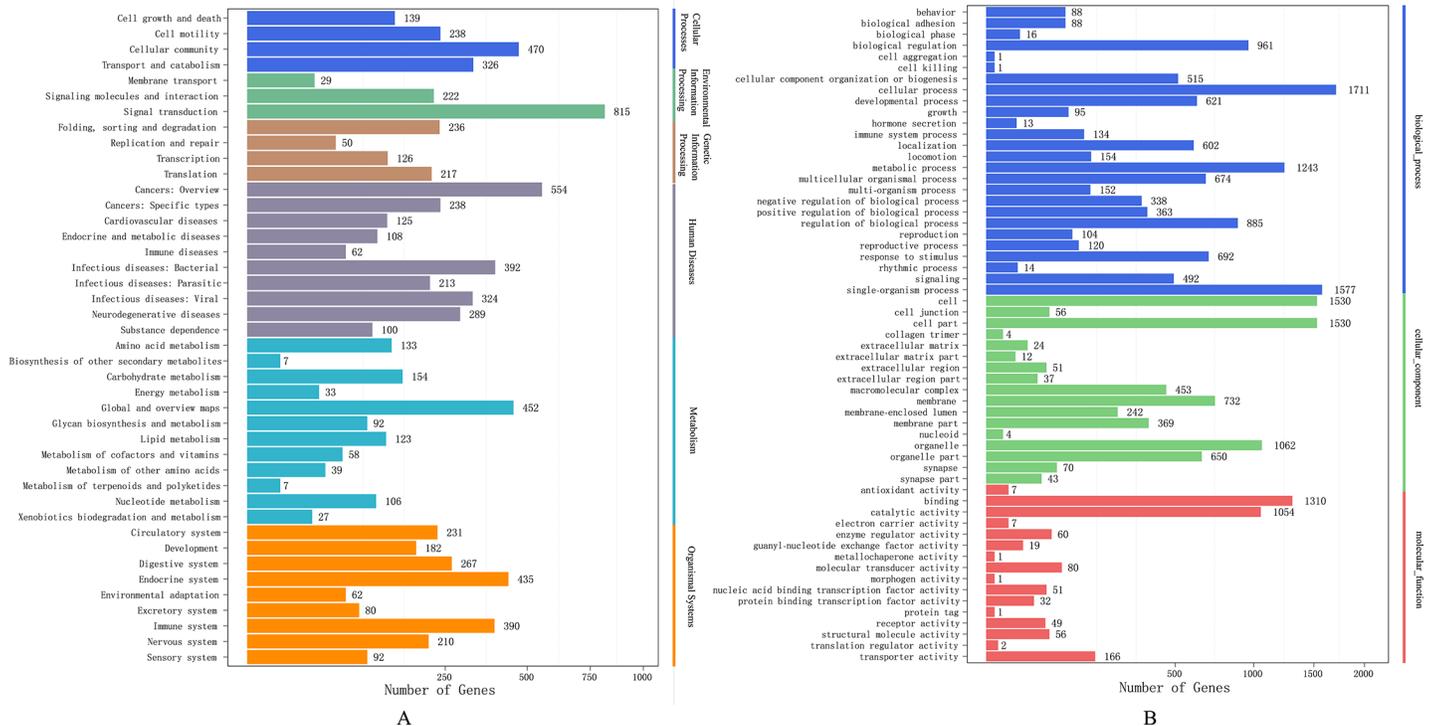
https://doi.org/10.1371/journal.pone.0176677.g004

KEGG pathways under the following main groups (**Fig 4A**): Signal transduction (815 genes), Cancers: Overview (554 genes), Cellular community (470 genes), Global and overview maps (452 genes) and Immune system (390 genes). DEGs were also confirmed by GO enrichment analysis, in which 2,287 up-regulated and 242 down-regulated genes were enriched in GO terms respectively (**Fig 4B**). They include cellular process (1,171 genes), single-organism process, cell and cell part, binding and metabolic process. The DEGs functional enriched terms provided a helpful resource for finding important cellular structures, pathways, processes and protein functions. The DEGs potentially involved in stress-associated responses and detoxification of xenobiotics are interest to environmental protection of researchers.

Differential expression of byssus protein coding genes

In order to further examine the molecular mechanisms related to accumulation of heavy metals in the byssus, 252 byssus protein sequences from the LC-MS/MS data (our unpublished data) were used to screen the examined DEGs. In Cd stressed conditions, we observed that a total of 54 byssus protein coding genes with the same trend (up- or down-regulation) are common for the two comparisons (Cd_50–48 vs control and Cd_100–48 vs control) (**S4 Table**), and 40 byssus protein coding genes were up-regulated more than twofold in Cd_50 and Cd_100 conditions compared with the corresponding controls.

Experimental validation of the RNA-seq analysis by qRT-PCR

To validate the reliability of the differential expression of byssus protein coding genes obtained from the RNA-Seq analyses, qRT-PCR was used to dissect dynamic changes in gene expression levels of 16 byssus protein coding genes. As shown in **Fig 5**, the data from both methods
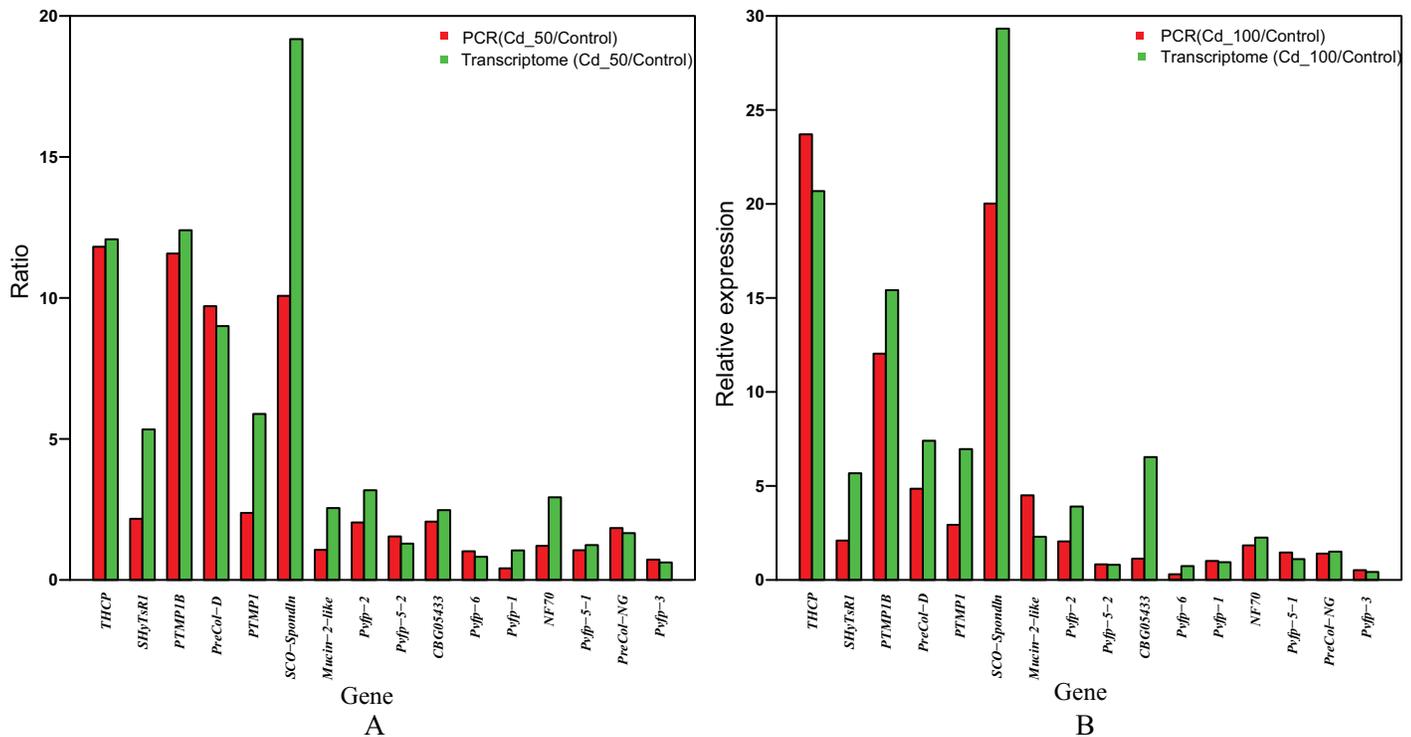
Fig 5. Validation of byssus protein coding genes by qRT-PCR. β-actin was used as an internal control and each value represents average of 3 separate biological replicates. Real-time PCR validates the byssus protein coding genes (A) between the Cd_50 and Control and (B) between the Cd_100 and Control (48-h).

(transcriptome and qRT-PCR) were relatively consistent, suggesting that the RNAseq results are credible.

Discussion

Byssus, secreted from the mussel foot gland, is composed of at least 6 foot proteins (*Fp1~6*) and 3 collagenous proteins [13, 40, 41]. Byssus can effectively accumulate heavy metals [42–44] (our study also confirmed this conclusion; **Fig 1**), and always with the highest accumulation rates (such as for Cd and Zn) among examined tissues under laboratory conditions [7]. Previous reports have proved that byssus proteins and metal ions could form metal complexes (chelating reaction), which can increase cohesive strength and abrasion resistance of the mussel byssus [40], in favor of resistance to the wave forces associated with the intertidal habitat. Swann et al. [45] reported that byssus accumulated hundred-fold concentrations of metal ions (including Manganese (Mn), Zn, Copper (Cu) and Iron (Fe)), from the surrounding seawater. Here, we performed comparative transcriptome analysis of foot gland with and without Cd stimulation so as to dig out and profile the candidate byssus protein coding genes.

Under laboratory conditions, Yap et al. [7] reported that byssus had the highest accumulation rates of Cd and Zn at the 48-h time point. In our present study, we also observed that the number of DEGs at 48 h was more than those at 24 h and 72 h (**Fig 3A and 3B**), which is similar to Yap's report and revealed a remarkable ability of byssus to accumulate Cd within a short period of treatment. Individual differences may have a significant effect on the number of DEGs, but several DEGs were verified by qRT-PCR and the results are consistent with our transcriptome data, suggesting little effect on differential expression by individual differences.

Our data demonstrated that 40 byssus protein coding genes were up-regulated more than two-fold along with promotion of the byssus protein synthesis. In total, 10 up-regulated byssus protein coding genes were assigned to one or more GO terms. These coding genes are involved in cellular process, binding, response to stimulus, metabolic process, catalytic activity and biological regulation, which probably provide protection function for *P. viridis* to live in the Cd stimulated condition.

Several coding genes for byssus proteins presented significant differences between the two regimes (Cd_50–48 VS Control-48, Cd_100–48 VS Control-48; **Fig 6A**). For instance, *PreCol-D*, *PTMP1*, similar to *HyTSR1* ptotein, similar to trefoil factor (*STFF*) and antistasin-like protein (*ALP*) have more than 5-fold transcriptional changes in Cd_50 and Cd_100 relative to the controls. Based on the sequence analysis, we observed that *P. viridis ALP* has a high content of cysteine (about 20%) and tyrosine residues, which contains 8 internal repeats of a 30-amino-acid sequence [46] with a highly conserved pattern of 6 cysteine (Cys) and 2 glycine (Gly) residues (**Fig 6B**). As we know, cysteine has remarkable metal-binding and redox activities [47–49], and its interaction to metal ions is primarily through the thiol group of Cys residues [50]. Therefore, we infer that our new *ALP* protein may be able to accumulate heavy metals to a high degree. *P. viridis PreCol-D* has both of the histidine-rich domains, and four residues of Dopa in the N-terminal flanking domain of *PreCol-D*. Significance of these structures may support the mussel to utilize more metal chelate cross-links, which have been implicated for byssal thread stability [12]. The *P. viridis PTMP1*, localized exclusively in the proximal byssus section [16], has two von willebgand factor type A (VWA) domains (A1 and A2); both of them present a metal ion dependent adhesion sites (MIDAS) [51] and very similar folds as shown by an overlay of their structures (**Fig 6C**). Surprisingly, however, zinc ion was reported to bind the A1 MIDAS motif but not the A2 motif [16]. In our study, comparison of *P. viridis PTMP1* with its homologous sequence reported from *Mytilidae galloprovincialis* revealed their remarkable conservation.

In our present work, we observed that transcription levels of foot proteins (*Pvfp-1* to *Pvfp-6*) and mussel byssus protein tyrosinase (*TYR1* to *TYR5*) increased significantly in the Cd stimulated conditions. *TYR* is a copper-containing enzyme [52], which can convert tyrosine into adhesive DOPA residues [53]. It has been recognized as a key component of byssus adhesion proteins [54–56]. *TYR* can directly bind copper, and the Cu(A) and Cu(B) sites are both required for copper binding for catalytic activity [57]. Mussel foot proteins (*Mfps*) own the posttranslationally modified Dopa. In particular, Dopa is recognized as a key component of mussel adhesion proteins that enables mussels to adhere to different surfaces [58]. A lot of previous reports have proved that Dopa and metal ions could form metal complexes (chelating reaction), which can increase cohesive strength and abrasion resistance of the mussel byssus. Metal ions can enable adhesion of non-adhering *Mfps*, especially *Mfp-1* and *Mfp-2*, to a surprising degree [58]; *Mfp-2* cohesiveness depends on complexation of metal ions, particularly Fe^{3+}. It consists of 11 tandem repeats of an epidermal growth factor (*EGF*) domain that has a consensus motif for Ca^{2+} binding [59, 60]. Comparison of *P. viridis Pvfp-2* with its homologous sequence reported from *Mytilidae edulis* revealed their remarkable Cys conservation (**Fig 6D**). Pvfp-4 riches in histidine and binds copper with high capacity, which was proposed to interact with the histidine-rich ends of *PreCols* via copper complexes [61]. *Pvfp-3*, *5–1* and *6* have high contents of Cys and Tyr residues (**Fig 6E**), while Cys has remarkable metal-binding and redox activities and the interaction to metal ions is primarily through the thiol group of Cys residues. Therefore, we infer that *Pvfp-2*, *3*, *5* and *6* may play the critical roles for accumulate of Cd and other metal ions.

A common feature of *Pvfps* and *ALP* proteins are existence of a high content of Tyr or Cys, and other proteins of DEGs have high contents of His residues or existence the special motif of
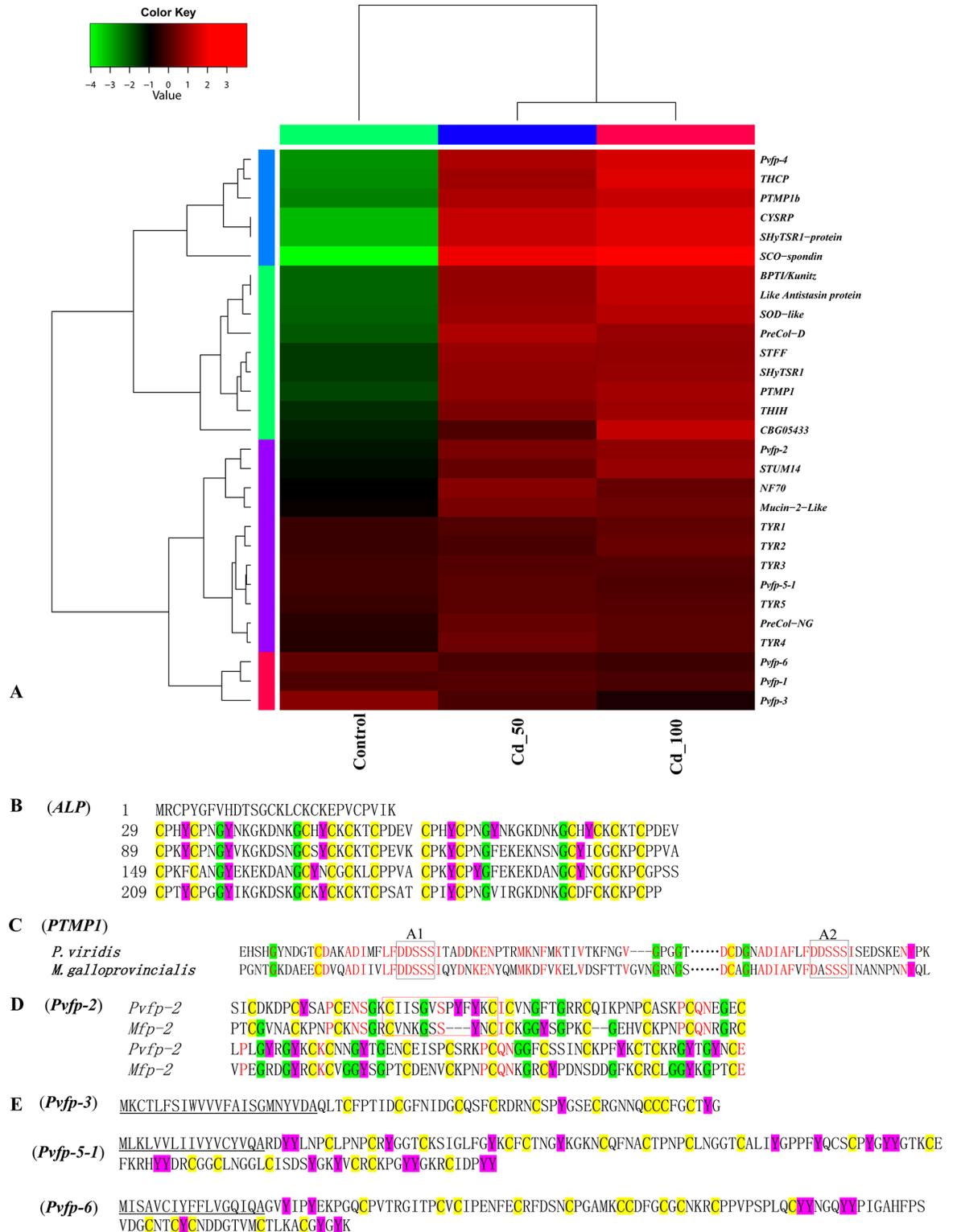
Fig 6. Hierarchical cluster and structure analysis of DEGs for the byssus metal accumulation. The cysteine (Cys, C), tyrosine (Tyr, Y) and glycine (Gly, G) are highlighted in yellow, pink and green colors respectively. (A) The color key represents RPKM normalized log$_2$ transformed counts in control, Cd_50 and Cd_100. (B) Amino acid sequence of the *P. viridis* antistasin-like protein (*ALP*) with 8 highly conserved patterns of CPXXCXXGXXXXXXXXXGCXXCXCXXCXX; (C) Comparison of *P. viridis* PTMP1 with its homologus sequence reported from *M. galloprovincialis*. The amino acids of the putative MIDAS motif are

boxed; (D) Comparison of *P. viridis Pvfp-2* with its homologus sequence from *M. edulis*. Note that the red box is a calcium binding motif. (E) Amino acid sequence of *P. viridis* foot proteins (*Pvfp-3*, *5–1*, *6*), which are rich of Cys and Tyr residues. The underlined regions are signal sequences.

https://doi.org/10.1371/journal.pone.0176677.g006

binding metal. These residues could form metal complexes with many heavy metal ions, which may be one important explanation for the high metal accumulation and remarkable adhesive strength of the byssus. We therefore proposed an interesting model for the byssus Cd accumulation, in which biosynthesis of byssus proteins may play simultaneously critical roles since their transcription levels were significantly elevated.

Conclusion

The mussel is a marine bivalve with an economic importance as well as biomonitoring roles for aquatic pollution. In the present study, we performed transcriptomic analysis on the mussel feet and examined their gene transcriptional changes during cadmium stimulation using the RNA sequencing technology. Based on differential expression analysis, we observed that several coding genes for byssus proteins were altered remarkably in response to the Cd stimulation. These proteins have a high content of Tyr, Cys or His, which could form metal complexes with many heavy metal ions. Therefore, we proposed that they may play an important role in the confirmed high Cd accumulation in the byssus samples. An interesting model was proposed, at the first time, in which biosynthesis of byssus proteins may play simultaneously critical roles since their transcription levels were significantly elevated. Our data will also provide a valuable genetic resource for identification of more important proteins to guide further recombinant protein engineering and biomimetic material processing for removal of heavy metal pollution from mine factories or nuclear power plants.

Supporting information

S1 Fig. COG classification of all unigenes in the *P. viridis* transcriptomes.
(PDF)

S2 Fig. GO annotation of all unigenes in the *P. viridis* transcriptomes.
(PDF)

S1 Table. Sequences of the primers used for qRT-PCR.
(XLSX)

S2 Table. The KEGG pathway annotation of unigenes from *P. viridis* foot transcriptomes.
(XLSX)

S3 Table. RPKM values of all genes detected in the Control, Cd_50 and Cd_100 at 3 time points.
(XLSX)

S4 Table. Differential expression of byssus protein coding genes at the 48-h Cd stimulation.
(XLSX)

Author Contributions

Conceptualization: QS XZ.

Data curation: XZ JW.

Formal analysis: XZ JW ZR.

Funding acquisition: QS.

Investigation: XY JC CP.

Methodology: QS XZ.

Project administration: QS.

Resources: QS.

Software: XZ JW.

Supervision: QS.

Validation: XZ ZR.

Visualization: QS.

Writing – original draft: XZ QS.

Writing – review & editing: QS XZ XY.

References

1. Acosta V, Lodeiros C. Metales pesados en la almeja Tivela mactroides Born, 1778 (Bivalvia: Veneridae) en localidades costeras con diferentes grados de contaminación en Venezuela Heavy metals in the clam Tivela mactroides Born, 1778 (Bivalvia: Veneridae) from coastal localities with different degrees of contamination in Venezuela. Ciencias marinas. 2004; 30(2):323–33.

2. Ferreira A, Machado A, Zalmon I. Temporal and spatial variation on heavy metal concentrations in the oyster Ostrea equestris on the northern coast of Rio de Janeiro State, Brazil. Brazilian Journal of Biology. 2005; 65(1):67–76.

3. Goswami P, Hariharan G, Godhantaraman N, Munuswamy N. An integrated use of multiple biomarkers to investigate the individual and combined effect of copper and cadmium on the marine green mussel (Perna viridis). J Environ Sci Heal A, Toxic/hazardous substances & environmental engineering. 2014; 49(13):1564–77. Epub 2014/08/20.

4. Huang L, Zou Y, Weng HW, Li HY, Liu JS, Yang WD. Proteomic profile in Perna viridis after exposed to Prorocentrum lima, a dinoflagellate producing DSP toxins. Environ Pollut. 2015; 196:350–7. Epub 2014/12/03. https://doi.org/10.1016/j.envpol.2014.10.019 PMID: 25463732

5. Nicholson S, Lam PK. Pollution monitoring in Southeast Asia using biomarkers in the mytilid mussel Perna viridis (Mytilidae: Bivalvia). Environ Int. 2005; 31(1):121–32. https://doi.org/10.1016/j.envint.2004.05.007 PMID: 15607786

6. Pinto R, Acosta V, Segnini MI, Brito L, Martinez G. Temporal variations of heavy metals levels in Perna viridis, on the Chacopata-Bocaripo lagoon axis, Sucre State, Venezuela. Mar Pollut Bull. 2015; 91(2):418–23. Epub 2014/12/03. https://doi.org/10.1016/j.marpolbul.2014.09.059 PMID: 25444616

7. Yap CK, Ismail A, Tan SG, Omar H. Accumulation, depuration and distribution of cadmium and zinc in the green-lipped mussel Perna viridis (Linnaeus) under laboratory conditions. Hydrobiologia. 2003; 498 (1–3):151–60.

8. Waite JH. Nature's underwater adhesive specialist. Int J Adhes. 1987; 7(1):9–14.

9. Lu Q, Danner E, Waite JH, Israelachvili JN, Zeng H, Hwang DS. Adhesion of mussel foot proteins to different substrate surfaces. J R Soc Interface, Interface / the Royal Society. 2013; 10(79):20120759. Epub 2012/11/23. PubMed Central PMCID: PMC3565691.

10. Lin Q, Gourdon D, Sun C, Holten-Andersen N, Anderson TH, Waite JH, et al. Adhesion mechanisms of the mussel foot proteins mfp-1 and mfp-3. Proc Natl Acad Sci USA. 2007; 104(10):3782–6. Epub 2007/03/16. PubMed Central PMCID: PMC1820661. https://doi.org/10.1073/pnas.0607852104 PMID: 17360430

11. Waite JH, Qin X-X, Coyne KJ. The peculiar collagens of mussel byssus. Matrix Biology. 1998; 17 (2):93–106. PMID: 9694590

12. Lucas JM, Vaccaro E, Waite JH. A molecular, morphometric and mechanical comparison of the structural elements of byssus from Mytilus edulis and Mytilus galloprovincialis. J Exp Biol. 2002; 205 (12):1807–17.

13. Qin X-X, Coyne KJ, Waite JH. Tough Tendons Mussel Byssus Has Collagen with Silk-Like Domains. J Biol Chem. 1997; 272(51):32623–7. PMID: 9405478

14. Coyne KJ, Qin X-X, Waite JH. Extensible collagen in mussel byssus: a natural block copolymer. Science. 1997; 277(5333):1830–2. PMID: 9295275

15. Guerette PA, Hoon S, Seow Y, Raida M, Masic A, Wong FT, et al. Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. Nat Biotechnol. 2013; 31 (10):908–15. Epub 2013/09/10. https://doi.org/10.1038/nbt.2671 PMID: 24013196

16. Sun C, Lucas JM, Waite JH. Collagen-binding matrix proteins from elastomeric extraorganismic byssal fibers. Biomacromolecules. 2002; 3(6):1240–8. Epub 2002/11/12. PMID: 12425661

17. Suhre MH, Gertz M, Steegborn C, Scheibel T. Structural and functional features of a collagen-binding matrix protein from the mussel byssus. Nat Commun. 2014; 5:3392. Epub 2014/02/27. https://doi.org/10.1038/ncomms4392 PMID: 24569701

18. Miao Y, Zhang L, Sun Y, Jiao W, Li Y, Sun J, et al. Integration of Transcriptomic and Proteomic Approaches Provides a Core Set of Genes for Understanding of Scallop Attachment. Mar Biotechnol (NY). 2015; 17(5):523–32. Epub 2015/05/29.

19. Lee H, Scherer NF, Messersmith PB. Single-molecule mechanics of mussel adhesion. Proc Natl Acad Sci USA. 2006; 103(35):12999–3003. Epub 2006/08/22. PubMed Central PMCID: PMC1559742. https://doi.org/10.1073/pnas.0605552103 PMID: 16920796

20. Yu J, Fedorova ND, Montalbano BG, Bhatnagar D, Cleveland TE, Bennett JW, et al. Tight control of mycotoxin biosynthesis gene expression in Aspergillus flavus by temperature as revealed by RNA-Seq. FEMS microbiology letters. 2011; 322(2):145–9. Epub 2011/06/29. https://doi.org/10.1111/j.1574-6968.2011.02345.x PMID: 21707733

21. Lin JQ, Zhao XX, Zhi QQ, Zhao M, He ZM. Transcriptomic profiling of Aspergillus flavus in response to 5-azacytidine. Fungal Genet Biol: FG & B. 2013; 56:78–86. Epub 2013/05/07.

22. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1):57–63. Epub 2008/11/19. PubMed Central PMCID: PMC2949280. https://doi.org/10.1038/nrg2484 PMID: 19015660

23. Suarez-Ulloa V, Fernandez-Tajes J, Manfrin C, Gerdol M, Venier P, Eirin-Lopez JM. Bivalve omics: state of the art and potential applications for the biomonitoring of harmful marine compounds. Mar Drugs. 2013; 11(11):4370–89. Epub 2013/11/06. PubMed Central PMCID: PMC3853733. https://doi.org/10.3390/md11114370 PMID: 24189277

24. Leung PT, Ip JC, Mak SS, Qiu JW, Lam PK, Wong CK, et al. De novo transcriptome analysis of Perna viridis highlights tissue-specific patterns for environmental studies. BMC Genomics. 2014; 15:804. Epub 2014/09/23. PubMed Central PMCID: PMC4190305. https://doi.org/10.1186/1471-2164-15-804 PMID: 25239240

25. Rebl A, Korytář T, Köbis JM, Verleih M, Krasnov A, Jaros J, et al. Transcriptome profiling reveals insight into distinct immune responses to Aeromonas salmonicida in gill of two rainbow trout strains. Mar Biotechnol. 2014; 16(3):333–48. https://doi.org/10.1007/s10126-013-9552-x PMID: 24122123

26. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014; 30(12):1660–6. Epub 2014/02/18. https://doi.org/10.1093/bioinformatics/btu077 PMID: 24532719

27. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics. 2003; 19(5):651–2. PMID: 12651724

28. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21(18):3674–6. Epub 2005/08/06. https://doi.org/10.1093/bioinformatics/bti610 PMID: 16081474

29. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. Nucleic acids Res. 2006; 34(Web Server issue):W293–7. Epub 2006/07/18. PubMed Central PMCID: PMC1538768. https://doi.org/10.1093/nar/gkl031 PMID: 16845012

30. Iseli C, Jongeneel CV, Bucher P, editors. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. ISMB; 1999.

31. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5(7):621–8. Epub 2008/06/03. https://doi.org/10.1038/nmeth.1226 PMID: 18516045

32. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11 (10):R106. Epub 2010/10/29. PubMed Central PMCID: PMC3218662. https://doi.org/10.1186/gb-2010-11-10-r106 PMID: 20979621

33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25(1):25–9. Epub 2000/05/10. PubMed Central PMCID: PMC3037419. https://doi.org/10.1038/75556 PMID: 10802651

34. Altermann E, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics. 2005; 6:60. Epub 2005/05/05. PubMed Central PMCID: PMC1112592. https://doi.org/10.1186/1471-2164-6-60 PMID: 15869710

35. Chen S, Yang P, Jiang F, Wei Y, Ma Z, Kang L. De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. PloS One. 2010; 5(12):e15633. Epub 2011/01/07. PubMed Central PMCID: PMC3012706. https://doi.org/10.1371/journal.pone.0015633 PMID: 21209894

36. Kalavacharla V, Liu Z, Meyers BC, Thimmapuram J, Melmaiee K. Identification and analysis of common bean (Phaseolus vulgaris L.) transcriptomes by massively parallel pyrosequencing. BMC Plant Biol. 2011; 11:135. Epub 2011/10/12. PubMed Central PMCID: PMC3209450. https://doi.org/10.1186/1471-2229-11-135 PMID: 21985325

37. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012; 40(15):e115. Epub 2012/06/26. PubMed Central PMCID: PMC3424584. https://doi.org/10.1093/nar/gks596 PMID: 22730293

38. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods. 2001; 25(4):402–8. https://doi.org/10.1006/meth.2001.1262 PMID: 11846609

39. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007; 35(Web Server):W182–W5. https://doi.org/10.1093/nar/gkm321 PMID: 17526522

40. Hedlund J, Andersson M, Fant C, Bitton R, Bianco-Peled H, Elwing H, et al. Change of colloidal and surface properties of Mytilus edulis foot protein 1 in the presence of an oxidation (NaIO$_4$) or a complex-binding (Cu2+) agent. Biomacromolecules. 2009; 10(4):845–9. Epub 2009/02/13. https://doi.org/10.1021/bm801325j PMID: 19209903

41. Corne K, Qin X, Waite J. Extensible collagen in mussel byssus: a natural block polymer. Science. 1997; 277:1830–2. PMID: 9295275

42. Chen SHS Q.; Zhao X.K.; Meng F.; Huang G.G.; Huang P.W. Accumulation and distribution of I125 in Perna Viridis. Tropic Oceanology. 1994;(03):62–6.

43. Yap CK, Ismail A, Cheng WH, Tan SG. Crystalline style and tissue redistribution in Perna viridis as indicators of Cu and Pb bioavailabilities and contamination in coastal waters. Ecotox Environ Safe. 2006; 63(3):413–23. Epub 2006/01/13.

44. Szefer P, Ikuta K, Kushiyama S, Szefer K, Frelek K, Geldon J. Distribution and association of trace metals in soft tissue and byssus of Mytilus edulis from the east coast of Kyushu Island, Japan. Archives of Environmental Contamination and Toxicology. 1997; 32(2):184–90. PMID: 9069195

45. Swann CP, Adewole T, Waite JH. Preferential manganese accumulation in dreissenid byssal threads. Com Biochem Phys B: Biochemistry and Molecular Biology. 1998; 119(4):755–9.

46. Holt GD, Krivan HC, Gasic GJ, Ginsburg V. Antistasin, an inhibitor of coagulation and metastasis, binds to sulfatide (Gal(3-SO4) beta 1-1Cer) and has a sequence homology with other proteins that bind sulfated glycoconjugates. J Biol Chem. 1989; 264(21):12138–40. Epub 1989/07/25. PMID: 2745433

47. Mejare M, Bulow L. Metal-binding proteins and peptides in bioremediation and phytoremediation of heavy metals. Trends Biotechnol. 2001; 19(2):67–73. Epub 2001/02/13. PMID: 11164556

48. Quig D. Cysteine metabolism and metal toxicity. Alternative medicine review: Journal of Clinical Therapeutic. 1998; 3(4):262–70. Epub 1998/09/03.

49. Giles NM, Watts AB, Giles GI, Fry FH, Littlechild JA, Jacob C. Metal and redox modulation of cysteine protein function. Chem Biol. 2003; 10(8):677–93. Epub 2003/09/05. PMID: 12954327

50. Baker DH, Czarnecki-Maulden GL. Pharmacologic role of cysteine in ameliorating or exacerbating mineral toxicities. J Nutr. 1987; 117(6):1003–10. Epub 1987/06/01. PMID: 3298579

51. Lee JO, Rieu P, Arnaout MA, Liddington R. Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). Cell. 1995; 80(4):631–8. Epub 1995/02/24. PMID: 7867070

52. Aguilera F, McDougall C, Degnan BM. Evolution of the tyrosinase gene family in bivalve molluscs: independent expansion of the mantle gene repertoire. Acta Biomater. 2014; 10(9):3855–65. Epub 2014/04/08. https://doi.org/10.1016/j.actbio.2014.03.031 PMID: 24704693

53. Sanchez-Ferrer A, Rodriguez-Lopez JN, Garcia-Canovas F, Garcia-Carmona F. Tyrosinase: a comprehensive review of its mechanism. BBA. 1995; 1247(1):1–11. Epub 1995/02/22. PMID: 7873577

54. Choi YS, Yang YJ, Yang B, Cha HJ. In vivo modification of tyrosine residues in recombinant mussel adhesive protein by tyrosinase co-expression in Escherichia coli. Microb Cell Fact. 2012; 11(1):1.

55. Numata K, Baker PJ. Synthesis of adhesive peptides similar to those found in blue mussel (Mytilus edulis) using papain and tyrosinase. Biomacromolecules. 2014; 15(8):3206–12. https://doi.org/10.1021/bm5009052 PMID: 25030620

56. Hwang DS, Yoo HJ, Jun JH, Moon WK, Cha HJ. Expression of functional recombinant mussel adhesive protein Mgfp-5 in Escherichia coli. Appl Environ Microb. 2004; 70(6):3352–9. Epub 2004/06/09. PubMed Central PMCID: PMC427802.

57. Spritz RA, Ho L, Furumura M, Hearing VJ Jr. Mutational analysis of copper binding by human tyrosinase. J Invest Dermatol. 1997; 109(2):207–12. Epub 1997/08/01. PMID: 9242509

58. Lee BP, Messersmith PB, Israelachvili JN, Waite JH. Mussel-Inspired Adhesives and Coatings. Annual Rev Mater Res. 2011; 41:99–132. Epub 2011/11/08. PubMed Central PMCID: PMC3207216.

59. Hwang DS, Zeng H, Masic A, Harrington MJ, Israelachvili JN, Waite JH. Protein- and metal-dependent interactions of a prominent protein in mussel adhesive plaques. J Biol Chem. 2010; 285(33):25850–8. Epub 2010/06/23. PubMed Central PMCID: PMC2919147. https://doi.org/10.1074/jbc.M110.133157 PMID: 20566644

60. Inoue K, Takeuchi Y, Miki D, Odo S. Mussel adhesive plaque protein gene is a novel member of epidermal growth factor-like gene family. J Biol Chem. 1995; 270(12):6698–701. Epub 1995/03/24. PMID: 7896812

61. Zhao H, Waite JH. Proteins in load-bearing junctions: the histidine-rich metal-binding protein of mussel byssus. Biochemistry. 2006; 45(47):14223–31. Epub 2006/11/23. PubMed Central PMCID: PMC1892233. https://doi.org/10.1021/bi061677n PMID: 17115717