

Research article

Comparative transcriptome, digital gene expression and proteome profiling analyses provide insights into the brachyurization from the megalopa to the first juvenile in *Eriocheir sinensis*

Yunxia Yang^b, Fangcao Jin^{a,1}, Wanyi Liu^a, Guangming Huo^c, Feng Zhou^c, Jie Yan^a, Kaiya Zhou^a, Peng Li^{a,*}

^a Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, PR China

^b School of Fishery, Zhejiang Ocean University, Zhoushan 316022, PR China

^c School of Food Science, Nanjing Xiaozhuang University, Nanjing 211171, PR China



ARTICLE INFO

Keywords:

Eriocheir sinensis
Brachyurization
Transcriptome
Proteome
Digital gene expression

ABSTRACT

Eriocheir sinensis larva normally experiences 11 stages. The reduced abdomen folded beneath the thorax is the most prominent characteristic of morphological alteration from megalopa to juvenile crab. Up to date, the molecular mechanisms of brachyurization remain a mystery. Here, transcriptome library, digital gene expression (DGE) libraries and proteome libraries at two developmental stages [the megalopa stage of *E. sinensis* (stage M) and the first stage of juvenile crab (stage J1)] of the Chinese mitten crab larva were constructed for RNA sequencing and iTRAQ approaches followed by bioinformatics analysis, respectively. In total, 1106 genes and 871 proteins were differentially expressed between the stage M and stage J1. Moreover, several important pathways were identified, including biosynthesis of secondary metabolites, metabolic pathways, focal adhesion, and some disease pathways. Besides, muscle contraction, oxidative phosphorylation, calcium signaling, PI3K-Akt, DNA replication pathway, and integrin signaling pathway also had important functions in brachyurization process. Furthermore, the components, actin, actin-related protein, collagens, filamin-A/B, laminin, integrins, paxillin, and fibronectin had up-regulated expression levels in M stage compared to J1 stage.

1. Introduction

The brachyuran decapod crustaceans are characterized by a reduced abdomen folded beneath the cephalothorax, and inserted in a special cavity or between the pereopods in the early developmental process, called brachyurization metamorphosis [1]. The Chinese mitten crab (*Eriocheir sinensis*), a brachyuran decapod crab species, is distributed in eastern and northern China. Up to now, the molecular mechanisms of brachyurization development in brachyuran decapods are still unclarified.

The high-throughput sequencing approaches, including Illumina RNA-seq, iTRAQ and digital gene expression (DGE), are newly developed tools for omics perspective at the transcriptomic and posttranslational levels. Transcriptome and DGE can be used to identify candidate genes that regulate the developmental process of brachyurization [2], while proteome may provide strategy to

* Corresponding author. College of Life Sciences, Nanjing Normal University, No. 1 Wenyuan Road, Nanjing 210023, Jiangsu, PR China.
E-mail address: lipeng@njnu.edu.cn (P. Li).

¹ Present address: Hangzhou Normal University Division of Health Sciences.

<https://doi.org/10.1016/j.heliyon.2022.e12736>

Received 25 March 2022; Received in revised form 13 December 2022; Accepted 26 December 2022

Available online 7 January 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

examine specific physiological response at the protein levels [3]. Numerous studies have been performed to identify a certain class of molecules in *E. sinensis*, such as reproductive and embryo development [4,5,6], immunoreaction [7,8], eyestalk ablation [9], osmoregulation and stress adaption [10], brachyurization and adaptation to benthic lifestyle [2], metamorphosis and nutrition metabolism [11,12], immune-related miRNAs [13], etc. Although brachyurization has become perspicuous and intelligible in morphological changing processes, the molecular mechanisms of brachyurization in brachyuran crabs remain elusive and obscure. Hence, this study aimed to investigate brachyurization-associated genes and proteins in *E. sinensis* via transcriptome, proteome and DGE analyses.

Here, the unigenes and proteins identified from the larvae of *E. sinensis* through transcriptome, proteome, and DGE sequencing approaches were analyzed by bioinformatics approaches. Approximately 4 billion high-quality nucleotide sequences were generated by Illumina technology. Overall, 76,112 unigenes were obtained in a single run, and then assembled into 187,504 distinct contiguous sequences. Then, ten DGE libraries were established, and the differences in gene expression patterns between two developmental stages of *E. sinensis* were compared using a DGE system. Furthermore, the transcriptome sequences, proteome sequences and DGE profiles can be valuable resources for identifying the genes and proteins responsible for brachyurization in *E. sinensis*.

2. Materials and methods

2.1. Total RNA extraction, cDNA library construction and illumina sequencing and data analysis

RNeasy Mini Kit (Qiagen, Germany) was utilized to isolate total RNA from the whole Chinese mitten crab larvae at two developmental stages (stage M: the megalopa stage; stage J1: the first stage of juvenile crab). Briefly, total RNA was extracted from *E. sinensis* in accordance with the kit's instructions. mRNA was purified from total mixture RNA using the Oligo (dT) magnetic beads at an equal ratio of stage M and stage J1. Poly(A) mRNA was extracted with OligoTex mRNA mini kit (Qiagen). Then, mRNA was cut into short fragments (100–400 bp) with Fragmentation buffer. After 2% TAE-agarose gel electrophoresis, different fragments (200 ± 25 bp) were chosen and excised for PCR amplification as a template. cDNA was synthesized with SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen, USA). High-throughput sequencing was performed using the Illumina HiSeq™ 2000 at the Beijing Genomics Institute (BGI; Shenzhen, Shenzhen, China).

Raw data were firstly processed by removing low-quality reads, reads containing ploy-N and reads containing adapter via in-house perl scripts. Meanwhile, Q20 (the proportion of bases with a quality score >20), sequence duplication level and GC-content of the clean reads were determined. The downstream analysis was conducted according to the high-quality clean data. Trinity was used to perform transcriptome assembly [14] by default parameter set. The contigs were the longest assembled sequences. To determine contigs from the same transcript and the distance between these contigs, the reads were mapped to the paired-end reads-containing contigs. Finally, the contigs without Ns and ability to extend on both ends were regarded as Unigenes. Gene annotation was performed using the following databases: Nr, Nt, KO/COG, PFAM, KO, GO, and Swiss-Prot.

RSEM was used to estimate gene expression levels [15]. Differential expression analysis was carried out using the DESeq R package v1.10.1 [16,17]. Differential expression was calculated from DGE data by DESeq using a negative binomial distribution-based model. To control false discovery rate (FDR), the obtained P-values were adjusted using the Benjamini and Hochberg's method. Genes (an adjusted $P < 0.05$) determined by DESeq were defined as differentially expressed genes (DEGs). RT-PCR was employed to examine the expression levels of genes in triplicate specimens.

2.2. Protein isolation, digestion and sequencing

Tissues from stage M and stage J1 (triplicate biological specimens per group) were ground to powder in liquid nitrogen. The methods of protein extraction and digestion were in accordance with the general instruction manual. Briefly, the powder was resuspended in lysis buffer and added with PMSF and EDTA to the final concentrations of 1 and 2 mM, respectively. After mixing, the specimens were added with dithiothreitol (DTT) to a final concentration of 10 mM, followed by sonication at 200 W for 15 min. The extracted proteins were digested with Trypsin Gold at the protein: trypsin ratio of 20:1 for 4 h at 37 °C. After trypsinization, the peptides were dried using a centrifugal vacuum. The peptides were then labeled using an iTRAQ Reagent 8-plex Kit. Fractionation of the iTRAQ-labeled peptides was conducted by SCX chromatography using a Shimadzu LC-20AB HPLC Pump system. Then, ProteinPilot V4.5 was used for peptide identification according to the paragon parameters. Further filtration was performed based on the unused score to verify the results of ProteinPilot. The peptides with unused score ≥ 1.3 were selected, in order to ensure a reliability level of over 95%. The filtrated proteins were subjected to further analyses.

2.3. DGE library construction and sequencing

Illumina Gene Expression Sample Prep Kit was utilized to construct the tag library of stage M and stage J1 samples. First, mRNA was extracted from the specimens at the two developmental stages. The cDNA was synthesized as described above. Then, *Nla* III was used to digest the double-stranded cDNAs to produce a CATG cohesive end. Illumina adapter I containing a *Mme* I restriction site was ligated to the digested cDNA with Dynabeads. All adapter I-containing cDNAs were purified, followed by digestion with *Mme* I. To produce a tag library, the adapter I containing 21 bp tags was ligated to the Illumina adapter II. Liner PCR was used to amplify these tag fragments for 15 cycles. Then, the 85-bp amplicons were separated on nucleic acid non-denatured pre-made glue 6% gel to generate single-strand molecules. After anchoring to Solexa sequencing array, these molecules were sequenced using the Illumina platform at BGI (Shenzhen, China).

Table 1
Summary for the transcriptome of *Eriocheir sinensis*.

Data items	Number/length
Total number of raw reads	58,292,820
Total number of clean reads	51,921,414
Total number of clean nucleotides (nt)	4,672,927,260
Total number of contigs	187,504
Mean length of contigs (nt)	292
Total number of unigenes	76,112
Mean length of unigenes (nt)	628

2.4. De novo assembly and analyses of the transcriptome and proteome

Before assembly, the high-quality clean reads were retrieved by eliminating low-quality reads containing >10% Q20 bases, reads containing >10% ambiguous bases, duplicate sequences and adaptor sequences. Then, the SOAP *de novo* program [18] was employed for assembly of short reads containing 21 bp K-mers. Smaller K-mers or larger K-mers all were not appropriate. Finally, non-redundant unigenes with the longest length were obtained.

Nr, KEGG pathway, GO, COG and Swiss-Prot databases were used for unigenes and protein alignment with E-value 10^{-5}. Then, the functional annotation and sequence direction of unigenes and proteins were determined based on the analysis of the alignment data. The sequences of unigenes were first aligned with BLASTx, which acquired proteins with the highest sequence similarity to the unigenes along with their protein functional annotation. The function of annotated unigenes was classified by GO and analyzed using the Blast2go software. WEGO software was employed to analyze the distribution of gene functions in a species. All unigenes were subjected to GO annotations analysis. Unigene with higher priority aligned to the databases could enter the next circle. At last, the amino acid and nucleotide (5'-3') sequences of the coding regions were retrieved. When the unigenes could not be aligned to any database, other data of the amino acid and nucleotide (5'-3') sequences of the coding regions were obtained through ESTScan.

Genome and transcriptome databases have been widely used for protein identification, especially for tandem mass spectral data. NCBI nr, SwissProt and UniProt were used to annotate the proteins. UniProt contains a massive amount of protein sequences and functional data. SwissProt and NCBI nr have been recognized as the good annotated databases. However, SwissProt database is non-redundant, and the NCBI nr database includes non-identical sequences from GenBank CDS translations.

2.5. Mapping of DGE tags to reference transcriptome databases

All low-quality tags, such as tags with only one copy (sequencing errors), low complexity tags, empty tags (only adaptor sequences) and tags with unknown sequences 'N', were eliminated before read mapping to transcriptome databases. For tag mapping, a virtual library was constructed with our transcriptome database. All clean tags were mapped to the library and only one nucleotide mismatch was acceptable. The clean tag mapped to multiple genes was filtered, and the remainder was designed as an unambiguous clean tag. To determine gene expression levels, the number of unambiguous clean tags for each gene was measured and then normalized to the number of transcripts per million clean tags (TPM) [19,20].

2.6. Analyses of differentially expressed genes and proteins

The frequency of tags in each DGE library was determined based on the DGE significance analysis method [21]. The thresholds of FDR<0.001 and \log_2 ratio>1 were deemed as significant differences in gene expression. Differentially expressed proteins (DEPs) were further quantified by iTRAQ analysis. For GO enrichment analysis, DEGs/DEPs were mapped to the GO database. For pathway enrichment analysis, the DEGs/DEPs were mapped to the KEGG database. The hypergeometric test was used to calculate the enriched P-values [22]. The bonferroni correction's P-value (less than 0.05) and FDR less than 0.01 was chosen as the thresholds for detecting the gene and/or protein sets with significant enrichment in the GO and KEGG analyses, respectively.

3. Results

3.1. The transcriptome assembly and analysis in *E. sinensis*

The RNA mixtures with J1 and M stages of *E. sinensis* were sequenced using the Solexa/Illumina RNA-Seq platform. About 58,292,820 raw reads and 51,921,414 clean reads were assembled and analyzed (Table 1).

In total, 187,504 contigs were assembled with an average size of 292 nt. GC percentage and Q20% were 48.4% and 97.13%, respectively. Moreover, 76,112 unigenes with an average size of 628 nt were identified. BLASTn (cut-off E-value = 10^{-10}) was used to search these cDNA sequences. The results showed that the assembled 30,987 unigenes were highly reliable and covered most of the transcriptome sequences with query coverage of 76.5% and mean identity of 98.5%. All unigenes were deposited in the NCBI Short Read Archive database (accession number SRR16996440/PRJNA782386).

Then, all of these unigenes were annotated to match sequences against the Nr, Swiss-Prot, PFAM, GO and KEGG databases by using BLAST (E-value 10^{-5}) [23,24]. The numbers of best matched unigenes from the BLAST search are demonstrated in Table 2.

Table 2
Statistical data for the unigene annotation.

Annotated database	Number of unigenes	Percentage (%)
Nr	19157	21.24
Nt	3370	3.73
KO	8321	9.22
SwissProt	15277	16.94
PFAM	20198	22.4
GO	21185	23.49
KOG	10930	12.12
All databases	1885	2.09
At least one database	26768	29.68
Total unigenes	90161	100

Table 3
Statistical data of DGE library sequencing and tag mapping.

Summary		Es-M ^a	Es-J1 ^a
Raw data	Total	5,854,677	5,966,011
	Distinct tags	244,413	229,772
Clean tags	Total number	5,686,189	5,801,179
	Distinct tag number	94,791	85,366
All tags mapped to genes	Total number	4,536,136	4,987,715
	Total % of clean tags	79.77%	85.98%
	Distinct tag number	53,940	51,854
	Distinct tag % of clean tags	56.90%	60.74%
Unambiguous tags mapped to genes	Total number	4,009,476	4,286,175
	Total % of clean tags	70.51%	73.88%
	Distinct tag number	45,676	437,23
	Distinct tag % of clean tags	48.19%	51.22%
All tag-mapped genes	Total number	22,137	20,814
	Total % of reference genes	29.08%	27.35%
Unambiguous tag-mapped genes	Total number	17,192	16,014
	Total % of reference genes	22.59%	21.04%

^a Es-M: the megalopa stage of *E. sinensis*; Es-J1: the first stage of juvenile crab.

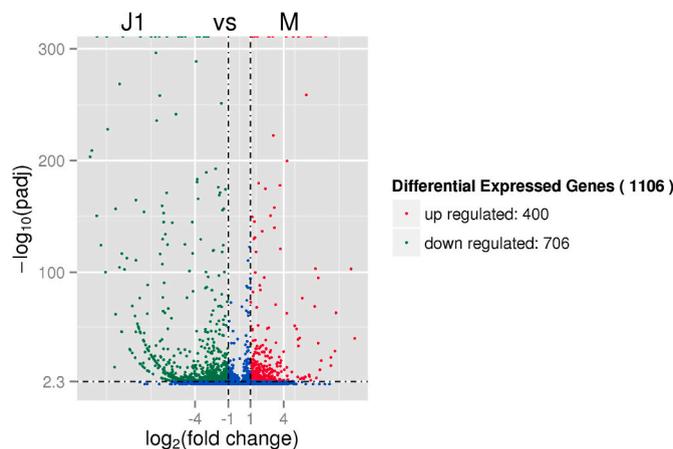


Fig. 1. DEGs between M and J1 stages of *Eriocheir sinensis*. M: the megalopa stage of *E. sinensis*; J1: the first stage of juvenile crab.

3.2. Mapping of DGE sequences to the reference transcriptome database

DGE could generate absolute instead of gene expression data and further avoid the limitations of microarray assays [25]. Therefore, the variations in gene expression during brachyurization were analyzed using the DGE method. The significant features of these libraries are demonstrated in Table 3. Each library could generate nearly 6 million raw tags and 5.6–6.1 million clean tags (Table 3) after filtering the low-quality tags.

The DGE tags were mapped to the transcriptome database to elucidate the molecular mechanism during brachyurization development from J1 stage to M stage. Tags mapped to a unique sequence can be used to identify a transcript, which represent the most

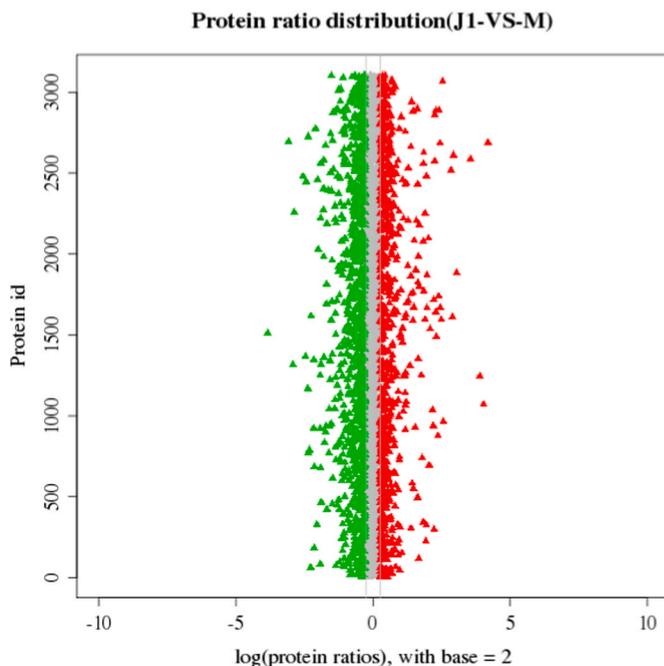


Fig. 2. The results from real-time quantitative PCR detection. Seven genes were determined by qPCR in triplicate samples. M: the megalopa stage of *E. sinensis*; J1: the first stage of juvenile crab.

Table 4
Overview of the proteomics sequencing results in *Eriocheir sinensis*.

Item	Number
Total spectra	270349
Spectra	52702
Unique spectra	48539
Peptides	17207
Unique peptides	16549
Proteins	4300

critical important of DGE library. Approximately 20,814 unique tags in the stage J1 (27.35%) and 22,137 in the stage M (29.08%) were unequivocally identified in our reference transcriptome database (Table 3).

3.3. Analysis for DEGs/DEPs between stage M and stage J1

After mapping the DGE library sequences to the transcriptome sequences, 1106 DEGs between M and J1 stages were identified from the mapped tags (Fig. 1), with 400 up-regulated genes and 706 down-regulated genes (stage J1 vs stage M). As demonstrated in Fig. 2, the expression levels of all candidate genes were detected by real-time quantitative PCR (qPCR), and differential expression patterns are in agreement with the sequencing results of transcriptome.

The high-throughput proteomics analysis of *E. sinensis* resulted in the initial identification of 4300 proteins, of which all were maintained for statistical analysis after quality control steps (Table 4). According to the abundance of the proteins, those with fold-change > 1.2 and $P < 0.05$ were defined as DEPs. Comparative analysis of two samples (Stage J1 vs Stage M) revealed that there were 871 regulated proteins, including 353 up-regulated proteins and 518 down-regulated proteins (Fig. 3).

The information about the correlation between genes and proteins are shown in Table 5.

Combining proteome and transcriptome analyses, the correlation between DEPs/genes was analyzed. Fig. 4 shows the abundance ratios of DEPs/genes.

3.4. GO and KEGG analyses for DEPs and DEGs

The differential expression genes and proteins could be grouped into diverse classes associated with binding (40.05%), catalytic activity (41.94%), cell and cell parts (21.68%, respectively), cellular processes (17.05%), metabolic processes (15.94%), organelle

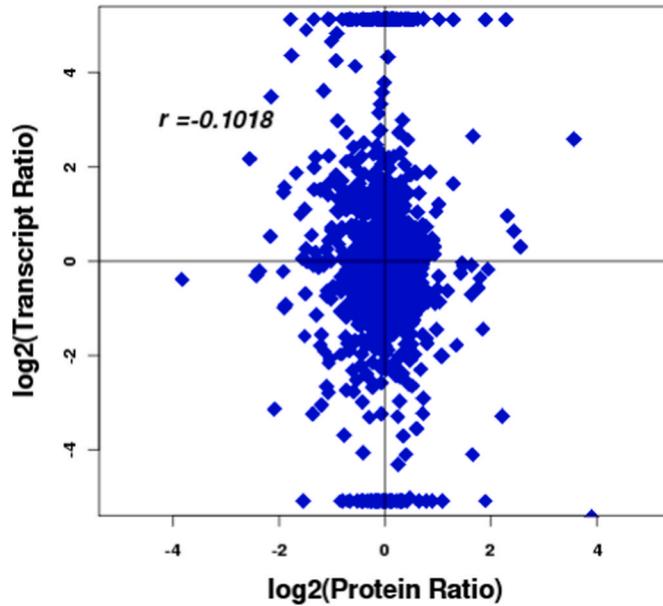


Fig. 3. The fold-change distribution of all proteins. The X-axis represents protein ratio [log (protein ratio), with base = 2]. The protein ratios above and below 0 indicate the up-regulated and down-regulated expression, respectively. Red dots mean the potential DEGs with up-regulation, while green dots mean the potential DEGs with down-regulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 5

The numbers of proteins and genes, and the correlation between them in *Eriocheir sinensis*.

Group name	Type	Number of proteins	Number of genes	Number of correlations
J1 vs M ^a	Identification	4300	14573	1703
	Quantitation	3108	14573	1398
	Differential Expression	871	1106	83

^a M: the megalopa stage of *E. sinensis*; J1: the first stage of juvenile crab.

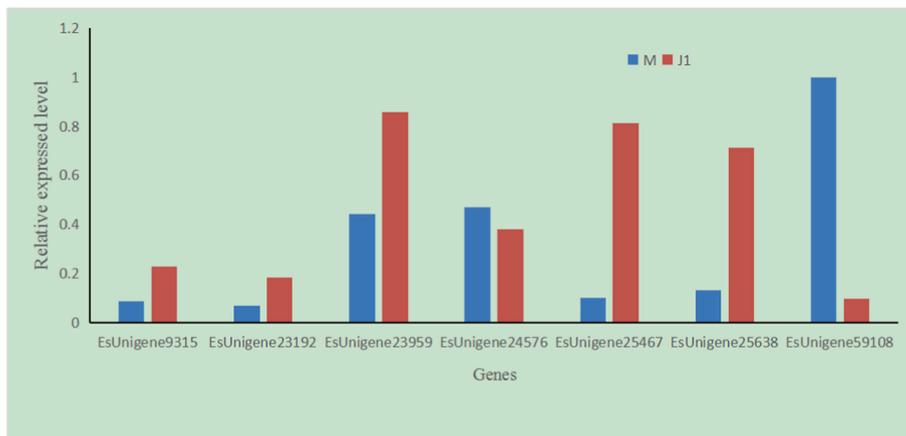


Fig. 4. Gene abundance ratio versus protein abundance ratio. X- and Y-axis indicate the expression levels of DEGs/DEPs, respectively.

(16.08%) and macromolecular complexes (11.67%) (Fig. 5). The KEGG database was used to obtain more information for predicting the functions, whereby 1140 DEGs/DEPs were classified into 235 KEGG pathways. The top 30 pathways of DEGs/DEPs are displayed in Table 6.

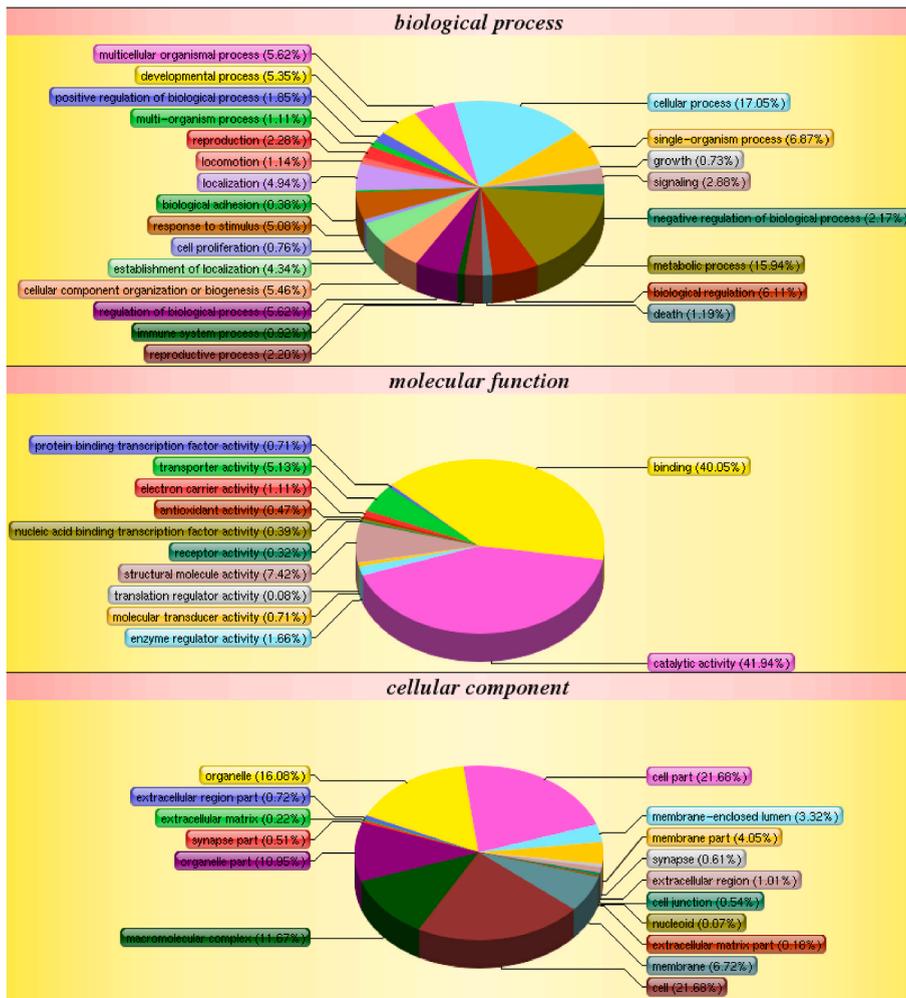


Fig. 5. GO analysis of DEGs/DEPs. These proteins were classified into diverse group: biological process, molecular function and cellular component.

4. Discussion

The larva of *E. sinensis* experience various developmental stages, including 5 zoeal stages and 1 megalopal stage, followed by the first to third juvenile crab stages [26,27]. The reduced abdomen folded beneath the thorax is the most prominent characteristic of morphological alteration from megalopa to juvenile crab, which is known as brachyurization [28]. Various taxonomical studies have indicated that the brachyuran organization can lead to greater biological functions with wider distribution and better adaptation to the littoral habitats. This organization has a phenotypic plasticity and undergone secondary modifications in response to habit and habitat alterations. To date, paleontological evidence has been proven invaluable to understanding the origins and evolution of brachyurans. However, molecular mechanism of brachyurization in brachyurans remains unknown, even though there are some researches about this question [29,2]. In this work, the transcriptome, proteome and DGE of *E. sinensis* could significantly enrich the databases and provide better understanding of the mechanism of brachyurization. [30] indicated that various genes in the pathways of oxidative phosphorylation, muscle contraction, calcium signaling, and lipid metabolism were significantly down-regulated in the abdomen of *E. sinensis* during brachyurization metamorphosis, while DNA replication and PI3K-Akt pathway genes demonstrated up-regulation. These results are also observed in this study.

At the same time, we found that the components of integrin signaling pathway and extracellular matrix also play the important functions in brachyurization metamorphosis. The results of proteome indicated that actin, actin-related protein, collagens, filamin-A/B, laminin, integrins, paxillin, and fibronectin were highly expressed in M stage than in J1 stage. These proteins are the components of the extracellular matrix and involved in integrin signaling pathway. The extracellular matrix can function as an important source for angiogenic, motility, survival and growth factors, which remarkably affects cell biology and development. Furthermore, cell adhesion to the extracellular matrix can trigger intracellular signaling pathways and interact with integrins and other cell surface receptors to modulate cell differentiation, migration, and cell cycle progression [31,32,33,34]. These proteins, as the main components of integrin signaling pathway and extracellular matrix, showed lower expression levels in the stage J1. These findings demonstrate that the

Table 6The top 30 pathways of DEGs/DEPs in *Eriocheir sinensis*.

No.	Pathway	The different expressed proteins and genes	Pathway ID
1	Metabolic pathways	291 (25.44%)	ko01100
2	Ribosome	71 (6.21%)	ko03010
3	Huntington's disease	66 (5.77%)	ko05016
4	Alzheimer's disease	63 (5.51%)	ko05010
5	Parkinson's disease	59 (5.16%)	ko05012
6	Oxidative phosphorylation	58 (5.07%)	ko00190
7	Protein processing in endoplasmic reticulum	48 (4.2%)	ko04141
8	Phagosome	47 (4.11%)	ko04145
9	Focal adhesion	45 (3.93%)	ko04510
10	Regulation of actin cytoskeleton	44 (3.85%)	ko04810
11	RNA transport	37 (3.23%)	ko03013
12	Tight junction	36 (3.15%)	ko04530
13	Spliceosome	35 (3.06%)	ko03040
14	Pathways in cancer	35 (3.06%)	ko05200
15	Lysosome	34 (2.97%)	ko04142
16	Influenza A	31 (2.71%)	ko05164
17	Valine, leucine and isoleucine degradation	31 (2.71%)	ko00280
18	Pancreatic secretion	31 (2.71%)	ko04972
19	Salmonella infection	31 (2.71%)	ko05132
20	Epstein-Barr virus infection	30 (2.62%)	ko05169
21	Pathogenic <i>Escherichia coli</i> infection	30 (2.62%)	ko05130
22	Cardiac muscle contraction	29 (2.53%)	ko04260
23	Dilated cardiomyopathy	28 (2.45%)	ko05414
24	Proteasome	28 (2.45%)	ko03050
25	Glutathione metabolism	27 (2.36%)	ko00480
26	Hypertrophic cardiomyopathy	27 (2.36%)	ko05410
27	Tuberculosis	26 (2.27%)	ko05152
28	Shigellosis	26 (2.27%)	ko05131
29	Leukocyte transendothelial migration	25 (2.19%)	ko04670
30	Glycolysis/Gluconeogenesis	25 (2.19%)	ko00010

brachyurization development may be related with the dysregulation of extracellular matrix or integrin signaling pathway.

In addition, we found that myosin had a high proportion and gene expression in M stage compared to J1 stage. Previous research has indicated that the fates of tail and trunk muscles are mainly influenced by the distributions of adult myoblasts in muscle tissues [35]. The tail absorption can be influenced by cAMP in a direct way [36], which is related to integrin signaling pathway. On the other hand, calcification in crustaceans plays vital roles in forming solid exoskeletons and attenuating physical damage [37]. Compared to the stage M, cuticle protein genes were up-regulated in the stage J1. These data are in agreement with a previous study [2].

Based on the analysis results of DEGs/DEPs between J1 and M stages, most genes and proteins belonged to the metabolite interconversion enzymes, including ligases, isomerases, lyases, hydrolases, transferases, oxidoreductases, translocases, and so on. Most of them participated in metabolic pathways. The results indicated that metabolism was activated in both M and J1 stages. Metabolic enzymes are required to catalyze the nutrients (e.g., carbohydrate, protein and lipid) and even nucleic acid to complete the developmental process of *E. sinensis*.

5. Conclusion

In summary, this study found that integrin signaling pathway and extracellular matrix remodeling exhibited important functions in brachyurization metamorphosis. In addition, the metabolite interconversion enzymes were activated during the morphological changes from stage M to stage J1. Altogether, our results can help explain the brachyurization patterns of *E. sinensis* and suggest comparative omics patterns with other brachyuran species.

Contribution statement

PL and KYZ conceived and designed this research. FCJ, YXY and PL conducted the experiments and data analysis. WYL, GMH, FZ, JY and KYZ analyzed and interpreted the data. YXY, FCJ and PL wrote the article. All authors discussed the results and approved the final revision for submission.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the National Natural Science Foundation of China (31000954), National Key Research and Development Project (No. 2019YFC1605800), Specialized Research Fund for the Doctoral Program of Higher Education of China (20103207120007), Priority Academic Program Development of Jiangsu Higher Education Institutions, and Top-notch Academic Programs Project of Jiangsu Higher Education Institutions (PPZY2015B117).

References

- [1] D. Guinot, J.M. Bouchard, Evolution of the abdominal holding system of brachyuran crabs (Crustacea, Decapoda, Brachyura), *Zoosystema* 20 (1998) 613–694.
- [2] C. Song, Z. Cui, M. Hui, et al., Comparative transcriptomic analysis provides insights into the molecular basis of brachyurization and adaptation to benthic lifestyle in *Eriocheir sinensis*, *Gene* 558 (1) (2015) 88–98.
- [3] Q. Meng, L. Hou, Y. Zhao, et al., iTRAQ-based proteomic study of the effects of *Spiroplasma eriocheiris* on Chinese mitten crab *Eriocheir sinensis* hemocytes, *Fish Shellfish Immunol.* 40 (1) (2014) 182–189.
- [4] W. Zhang, H. Wan, H. Jiang, et al., A transcriptome analysis of mitten crab testes (*Eriocheir sinensis*), *Genet. Mol. Biol.* 34 (1) (2011) 136–141.
- [5] L. He, H. Jiang, D. Cao, et al., Comparative transcriptome analysis of the accessory sex gland and testis from the Chinese mitten crab (*Eriocheir sinensis*), *PLoS ONE* 8 (1) (2013), e53915.
- [6] M. Hui, Z. Cui, Y. Liu, C. Song, Transcriptome profiles of embryos before and after cleavage in *Eriocheir sinensis*: identification of developmental genes at the earliest stages, *Chinese Journal of Oceanology and Limnology* 35 (4) (2017) 770–781.
- [7] X. Li, Z. Cui, Y. Liu, et al., Transcriptome analysis and discovery of genes involved in immune pathways from hepatopancreas of microbial challenged mitten crab *Eriocheir sinensis*, *PLoS ONE* 8 (7) (2013), e68233.
- [8] Z. Cui, X. Li, Y. Liu, et al., Transcriptome profiling analysis on whole bodies of microbial challenged *Eriocheir sinensis* larvae for immune gene identification and SNP development, *PLoS ONE* 8 (12) (2013), e82156.
- [9] Y. Sun, Y. Zhang, Y. Liu, et al., Changes in the organics metabolism in the hepatopancreas induced by eyestalk ablation of the Chinese mitten crab *Eriocheir sinensis* determined via transcriptome and DGE analysis, *PLoS ONE* 9 (4) (2014), e95827.
- [10] M. Hui, Y. Liu, C. Song, et al., Transcriptome changes in *Eriocheir sinensis* megalopae after desalination provide insights into osmoregulation and stress adaptation in larvae, *PLoS ONE* 9 (12) (2014), e114187.
- [11] Y. Li, M. Hui, Z. Cui, et al., Comparative transcriptomic analysis provides insights into the molecular basis of the metamorphosis and nutrition metabolism change from zoeae to megalopae in *Eriocheir sinensis*, *Comp. Biochem. Phys. D* 13 (2015) 1–9.
- [12] B. Wei, Z. Yang, J. Wang, et al., Effects of dietary lipids on the hepatopancreas transcriptome of Chinese mitten crab (*Eriocheir sinensis*), *PLoS ONE* 12 (7) (2017), e0182087.
- [13] J. Ou, Q. Meng, Y. Li, et al., Identification and comparative analysis of the *Eriocheir sinensis* microRNA transcriptome response to *Spiroplasma eriocheiris* infection using a deep sequencing approach, *Fish Shellfish Immunol.* 32 (2) (2012) 345–352.
- [14] M.G. Grabherr, B.J. Haas, M. Yassour, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nature Biotechnol.* 29 (7) (2011) 644–652.
- [15] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinf.* 12 (2011) 323.
- [16] S. Anders, W. Huber, Differential expression analysis for sequence count data, *Genome Biol.* 11 (2010) R106.
- [17] L.K. Wang, Z.X. Feng, X. Wang, et al., DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, *Bioinformatics* 26 (2010) 136–138.
- [18] R. Luo, B. Liu, Y. Xie, et al., SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, *Gigascience* 1 (1) (2012) 18.
- [19] P.A. 't Hoen, Y. Ariyurek, H.H. Thygesen, et al., Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms, *Nucleic Acids Res.* 36 (2008) 141.
- [20] A.S. Morrissy, R.D. Morin, A. Delaney, et al., Next-generation tag sequencing for cancer gene expression profiling, *Genome Res.* 19 (10) (2009) 1825–1835.
- [21] S. Audic, J.M. Claverie, The significance of digital gene expression profiles, *Genome Res.* 7 (1997) 986–995.
- [22] J. Cao, S. Zhang, A Bayesian extension of the hypergeometric test for functional enrichment analysis, *Biometr. J. Biometr. Soc. An Int. Soc. Devoted to the Math. Statist. Aspect. Biol.* 70 (1) (2014) 84–94.
- [23] J. Ye, L. Fang, H. Zheng, et al., WEGO: a web tool for plotting GO annotations, *Nucleic Acids Res.* 34 (2006) 293–297.
- [24] M. Kanehisa, S. Goto, S. Kawashima, et al., The KEGG resource for deciphering the genome, *Nucleic Acids Res.* 32 (2004) 277–280.
- [25] Y.F. Qin, H.M. Fang, Q.N. Tian, et al., Transcriptome profiling and digital gene expression by deep-sequencing in normal/regenerative tissues of planarian *Dugesia japonica*, *Genomics* 97 (2011) 364–371.
- [26] K. Anger, Effects of temperature and salinity on the larval development of the Chinese mitten crab *Eriocheir sinensis* (Decapoda, Grapsidae), *Mar. Ecol. Prog. Ser.* 72 (1991) 103–110.
- [27] R.B. Forward, R.A. Tankersley, D. Rittschof, Cues for metamorphosis of brachyuran crabs: an overview, *Am. Zool.* 41 (2001) 1108–1122.
- [28] Z. Števcic, The Main Features of Brachyuran Evolution, *Syst. Biol.* 20 (3) (1971) 331–340.
- [29] P. Li, J. Zha, H. Sun, et al., Identification of differentially expressed genes during the brachyurization of the Chinese mitten crab *Eriocheir japonica sinensis*, *Biochem. Genet.* 49 (2011) 645–655.
- [30] Z. Cui, Y. Liu, J. Yuan, et al., The Chinese mitten crab genome provides insights into adaptive plasticity and developmental regulation, *Nat. Commun.* 12 (1) (2021) 2395.
- [31] R.O. Hynes, Integrins: A family of cell surface receptors, *Cell* 48 (4) (1987) 549–554.
- [32] P. Gassmann, A. Enns, J. Haier, Role of tumor cell adhesion and migration in organ-specific metastasis formation, *Onkologie* 27 (6) (2004) 577–582.
- [33] C.M. Nelson, M.J. Bissell, Modeling dynamic reciprocity: engineering three-dimensional culture models of breast architecture, function, and neoplastic transformation, *Semin. Cancer Biol.* 15 (5) (2005) 342–352.
- [34] D.E. White, J.H. Rayment, W.J. Muller, Addressing the role of cell adhesion in tumor cell dormancy, *Cell Cycle* 5 (16) (2006) 1756–1759.
- [35] K. Shimizu-Nishikawa, Y. Shibota, A. Takei, et al., Regulation of specific developmental fates of larval- and adult-type muscles during metamorphosis of the frog *Xenopus*, *Develop. Biol.* 251 (1) (2002) 91–104.
- [36] G. Csaba, A. Szalkó, E. Kapa, Morphogenetic effect of cAMP during the metamorphosis of the frog larva, *Acta Biol. Acad. Scient. Hungar.* 32 (2) (1981) 99.
- [37] K. Itou, Y. Akahane, Post-molt processes of cuticle formation and calcification in the Japanese mitten crab *Eriocheir japonicus*, *Fish. Sci.* 75 (2009) 91–98.