



## OPEN ACCESS

## EDITED BY

Yunyan Deng,  
Institute of Oceanology (CAS), China

## REVIEWED BY

Heng Wang,  
Dalian Ocean University, China  
Tilman Schell,  
Centre for Translational Biodiversity  
Genomics (LOEWE-TBG), Germany

## \*CORRESPONDENCE

Kefeng Wu

✉ winokhere@sina.com

Hui Luo

✉ luohui@gdmu.edu.cn

†These authors have contributed  
equally to this work

## SPECIALTY SECTION

This article was submitted to  
Marine Biology,  
a section of the journal  
Frontiers in Marine Science

RECEIVED 13 September 2022

ACCEPTED 15 December 2022

PUBLISHED 12 January 2023

## CITATION

Qi Y, Chen L, Wu B, Tang X, Zhu X,  
Li R, Wu K and Luo H (2023)  
*Sipunculus nudus* genome provides  
insights into evolution of spiralian  
phyla and development.  
*Front. Mar. Sci.* 9:1043311.  
doi: 10.3389/fmars.2022.1043311

## COPYRIGHT

© 2023 Qi, Chen, Wu, Tang, Zhu, Li, Wu  
and Luo. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# *Sipunculus nudus* genome provides insights into evolution of spiralian phyla and development

Yi Qi<sup>1,2†</sup>, Liang Chen<sup>1†</sup>, Binhua Wu<sup>1</sup>, Xiaoning Tang<sup>3</sup>, Xiao Zhu<sup>1</sup>,  
Ru Li<sup>4</sup>, Kefeng Wu<sup>1,5\*</sup> and Hui Luo<sup>1,4\*</sup>

<sup>1</sup>Institute of Marine Medicine, Guangdong Medical University, Zhanjiang, China, <sup>2</sup>Pharmacy  
Department, Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China,

<sup>3</sup>Affiliated Hospital of Guangdong Medical University Pharmacy Department, Zhanjiang,  
Guangdong, China, <sup>4</sup>Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang),  
Zhanjiang, Guangdong, China, <sup>5</sup>The Marine Biomedical Research Institute of Guangdong Zhanjiang,  
Zhanjiang, Guangdong, China

**Introduction:** *Sipunculus nudus* is the most known species in the genus  
*Sipunculus*, distributed in tropical and subtropical coastal waters.

**Methods:** PacBio sequencing and Illumina sequencing were combined for  
whole-genome sequencing of *S. nudus*. LC-MS/MS analysis was performed for  
the metabolomics of *S. nudus*.

**Results:** Herein, we report a 1.75 Gb complete genome assembly with a contig  
N50 size of 450 kb of the *S. nudus* based on a strategy combining third-  
generation long-read sequencing and Illumina sequencing. A total of 80391  
protein-coding genes are annotated in this genome. Furthermore, gene family  
evolution analysis shows that *S. nudus* belonged to *Mollusca* or was close to  
*Mollusca*, but distinct with *Annelida*. Transcriptome analysis indicates the  
involvement of complex developmental events in larvae. KEGG pathway  
analysis of differentially expressed genes showed that these DEGs were  
mainly enriched in the pathway of amino acid metabolism, lipid metabolism  
and transport and catabolism. LC-MS/MS analysis shows that *S. nudus* was rich  
in a variety of nutritional and functional components, such as carnitine, free  
amino acids, unsaturated fatty acids, inosine and methionine sulfoxide.  
Combined transcriptome with LC-MS/MS analysis, the results showed that  
gene expression and metabolic level involved in the inosine, arginine and  
proline biosynthesis of *S. nudus* during different growth stages were  
significantly changed.

**Discussion:** Our genome assembly provides an important genome resource and new insight into the relationships of *Sipuncula* to other spiralian phyla. Meanwhile, transcriptome and LC-MS/MS analysis reveal the systematic gene expression profiles and metabolite components of *S. nudus* during different growth stages, which provide new insight into the exploration and development of bioactive molecules of *S. nudus*.

#### KEYWORDS

sipunculus nudus, genome, transcriptome, evolution, metabolite

## 1 Introduction

*Sipunculus nudus* (*S. nudus*) is a cosmopolitan species mainly distributed in tropical and subtropical coastal waters, except for polar waters. It is unsegmented wormlike animals and usually used as a model organism for various fields of research. In recent years, the wild populations of *S. nudus* suffered from predatory and unhindered exploration stimulated by the market price (Yahui et al., 2015). Furthermore, growing levels of oceans pollution have led to a sharp decrease in the availability of the wild resource. Thus, more and more attention has been paid to artificial cultivation of *S. nudus*. As the artificial cultivation of technology matures, the amount of *S. nudus* has increased rapidly and the development of reasonable utilization of the resource has become a major study topic.

In the last decades, *S. nudus* extract was reported to be rich in a variety of nutritional and functional components consisting of free amino acids, fatty acids, polysaccharides, mineral elements, nucleosides and nucleobases and so on. Various researches suggested that *S. nudus* had anti-oxidation, anti-bacterial, immune regulation, anti-inflammation and peripheral analgesia effects (Su et al., 2016; Lin et al., 2021). Sun et al. found that soluble polysaccharides extracted from *S. nudus* can scavenge hydroxyl radicals as a natural antioxidant (Sun et al., 2017). Other antioxidant components were also found in *S. nudus*, including unsaturated fatty acids, taurine and SOD, which can enhance the immune system. In addition, it is reported that *S. nudus* extract can treat the deformity caused by radiation loss and protect the organs damaged by radiation (Jiang et al., 2015). Furthermore, after the continuous intragastric administration of *S. nudus* extract for 7 days, the activity of antioxidant enzymes of mice in liver tissue increased and the mice showed excellent anti-fatigue activity (Liu et al., 2012). Lin et al. reported that collagen peptides derived from *S. nudus* accelerate wound healing and the underlying mechanisms were related to reducing inflammation and improving collagen deposition (Lin et al., 2021). However, the detailed components of *S. nudus* extract were still unclear.

On the other hand, the relationships within *Sipuncula* and the relationships of *Sipuncula* to other spiralian phyla have been strongly debated. According to morphological aspect, Scheltema thought that *S. nudus* was more close to *Mollusca* (Scheltema, 1993). However, the primary analysis of the mitochondrial gene arrangement between *Sipuncula* and *Annelida*, *Echiurida*, *Mollusca* indicated that the consistence of gene arrangement between *Sipuncula* and *Annelida*, *Echiurida* was higher, but lower with *Mollusca* (Ysa et al., 2021). Additionally, based on the complete mitochondrial (mt) genome sequence of *S. nudus*, the results showed that *Sipuncula* species belonged to *Annelida* or was close to *Annelida* (Thomas et al., 2009). However, Jennings and Halanych thought that it is less significance for the phylogenetic analysis using the mt-genome data compared with whole genome sequences (Jennings and Halanych, 2005). Unfortunately, none of *S. nudus* genomes, even in *Sipuncula*, has been reported. Here we describe the complete genome of *S. nudus* collected from the coast of south China, which may provide a reference genome in *Sipuncula* and provide evidences for the relationships of *Sipuncula* to other spiralian phyla.

## 2 Materials and methods

### 2.1 Samples collection and genomic DNA isolation

Samples were collected from North Bay (21°19'54"N, 109°48'57"E) in Zhanjiang of China at August, 2020. The temperature of seawater was 24–31°C and the salinity was 30–35‰. The length of larve was 0.6–1.2cm, and the length of one-quarter of adult, half of adult and adult were 2.5–4cm, 5–8cm, and 12–15cm. Total DNA was extracted from *S. nudus* using the following phenol/chloroform protocol. A whole *S. nudus* was homogenized to powder in liquid N<sub>2</sub>. The powder was transferred to a 1.5-mL microcentrifuge tube and homogenized in 0.5 mL lysis buffer containing 200mM Tris-HCl (pH 8.0), 20mM disodium EDTA,

1% SDS, 2 mg/ml RNase A (Qiagen, Hilden, Germany), and 20 mg/ml Proteinase K (Merck, Kenilworth, NJ, USA). After the incubation, the homogenate was centrifuged at 4°C and 12,000 rpm for 5 min. Then the supernatant was carefully removed and discarded, and 10mM Tris-HCl was added to dissolve the DNA. The sample was centrifuged at 4°C and 12,000 rpm for 10 min, and the supernatant was discarded after careful removal. The DNA precipitate was washed with 1 mL cold 70% ethanol, and then centrifuged at 4°C and 12,000 rpm for 5 min. The DNA pellet was air-dried at room temperature, dissolved in 50  $\mu$ L 10mM Tris-HCl, pH 8.0, and stored at  $-20^{\circ}\text{C}$ .

## 2.2 PacBio sequencing

PacBio CLR library was prepared using SMRTbell Template Prep Kit-SPv3 following manufacturer's recommendations. The library QC was performed using Qubit and Agilent 2100. The final library was sequenced on the Pacific Biosciences Sequel II system at Genedenovo Biotechnology Co., Ltd (Guangzhou, China) to produce third generation long read data.

## 2.3 Illumina sequencing

Illumina library was prepared with an insert size of 500 bp using Paired-End DNA Sample Prep kit (Illumina Inc., San Diego, CA, USA). These libraries were sequenced using NovaSeq 6000 (Illumina Inc., San Diego, CA, USA) NGS platform at Genedenovo company (Guangzhou, China) to produce second generation short reads data. Raw reads from illumina platform were processed to get high quality clean reads according to four stringent filtering standards: 1) removing reads with  $\geq 10\%$  unidentified nucleotides (N); 2) removing reads with  $> 50\%$  bases having phred quality scores of  $\leq 20$ ; 3) removing reads aligned to the barcode adapter. Then, jellyfish (version 2.2.6) was used to count kmer and GenomeScope (version 1.0.0) was relatively accurate for estimating the size of the genome, repetitive sequences, and heterozygosity (Marçais and Kingsford, 2011; Vurture et al., 2017).

## 2.4 De novo assembly

Samtools (Version 1.9) was used to extract the reads from subreads.bam file for assembly (Li et al., 2009). *De novo* assembly was performed with PacBio sequencing data using MECAT (Xiao et al., 2017). The longest subread per polymerase read was used for assembly. The parameter for mecat2pw was “-n 50” and parameter for mecat2canu was “Overlapper=meat2asmpw”. Considering the high error rate of third generation long reads data, pilon was used to correct potential sequence error in initial assembly with Illumina short reads data (Walker et al., 2014). To remove the potential

contamination of bacterial sequences in the current assembly, contigs in the primary assembly were subject to a decontamination pipeline. Firstly, contigs were blasted against NT database and MEGAN 6 was used to analyze the distribution of the aligned species (Huson et al., 2016). And then, we cut each scaffold into 100bp overlapping 1Kb windows and blasted them against the NT database using blastn. The blast results were further analyzed using MEGAN. If  $>60\%$  of windows in a contig had best hits as bacterial sequences with identity  $>70\%$ , this contig was picked out as possible contamination. The 1Kb windows were blasted against NR database using blastx and the blast results were further analyzed using MEGAN.

## 2.5 Completeness assessment for the assembled genome

Firstly, we applied the mapping rates of PacBio SMRT long reads and the Illumina short reads to the assembled genome as an indicator of the completeness of the assembled genome. Long reads were mapped using minimap2 (version 2.17) with the default parameters, while short reads were mapped using BWA-MEM (version 0.7.15) with the default parameters (Li and Durbin, 2009; Li, 2017). Next, we used BUSCO (version 5.2.2) with the parameters “-l metazoa\_odb10 -sp fly” to assess the completeness of the assembled genome based on the Metazoan dataset (Manni et al., 2021).

## 2.6 Repeat sequences annotation

The genome sequence was compared with the existing repeat sequence database, and the repeat sequence in the genome was found through homology. After alignment by mmseq2 (release 12-113e3), sequences with coverage greater than or equal to 50% and library created by RepeatModeler (version 2.0.2) were seemed as library file (Mirdita et al., 2019). We used RepeatMasker software (version 4.0.9) to carry out homologous alignment between genome sequence and repeat sequence in Repbase library (<http://www.girinst.org/repbase>, 20181026), so as to search and annotate repeat sequence in genome sequence (Daren et al., 2014; Bao et al., 2015).

## 2.7 Annotation of coding genes' structure

Coding genes annotation were integrated different methods to ensure maximum recognition of coding genes in the genome according to the following methods:

- (1) *Denovo* prediction: Augustus (Version: 3.3.3) and GlimmerHMM (Version: 3.0.1) software to predict the

coding genes of the whole genome (Majoros et al., 2003; Stanke and Morgenstern, 2005; Stifanic and Batel, 2007). Using the features included in the gene model: splicing signal model, exon length distribution, promoter and poly-A signal, differences in gene density and structure of different CG component regions, etc., to determine the location of coding exons and predict the number of genes in the sequence for both complete and incomplete genes, as well as for genes on both positive and negative chains were included.

- (2) Homology prediction: we compare the coding protein sequence of the known homologous species with the genome sequence of the new species, and then find the corresponding gene region in the new species through clustering, so as to achieve the purpose of homology prediction.
- (3) Integration: gene regions in the genome can be predicted by different methods, but the gene sets obtained by their respective methods all have their own defects. Using software MAKER (Version: 2.31.10), gene sets predicted by various methods can be integrated into a non-redundant, more complete gene set, and the final reliable gene set can be obtained through manual integration (Campbell et al., 2014). In our study, MAKER was run twice. The first round was used to the integration of RNA-seq and homologous proteins. After the prediction by Snap and Augustus, MAKER was used to integrate all the results of prediction, including RNA-seq, homologous proteins, snap, augustus, and genemark. The parameters were set as following: softmask=1, est2genome=1, protein2genome=1, cpus=5, pred\_flank=200, AED\_threshold=1.

## 2.8 Functional annotation of coding gene

In order to understand the function of each gene, we annotated the gene using protein database including NR, GO, SwissProt, KEGG, KOG and CAZy. Functional annotation was mainly based on the principle that functionally homologous sequences often have sequence similarity. The predicted protein sequence of the gene is aligned with different protein function databases (BLAST 2.10.0+), and then the function of the aligned sequence is used as the function of the target sequence (Camacho et al., 2009). In the alignment process, since there may be many alignment results for each sequence, in order to ensure the biological significance of subsequent analysis, the

threshold value  $evalue \leq 1e-5$  is used for filtering, and then the 20 sequences with the highest score value were selected from the alignment results of each sequence as the alignment results of the sequence.

## 2.9 Non-coding RNA annotation

According to the structural characteristics of tRNA, tRNA sequence in genome was searched by tRNAscan-SE (Version: 2.0.7) software with the default parameters (Chan and Lowe, 2019). Because rRNA is highly conservative, rRNA sequences of related species can be selected as reference sequences, and rRNA in genomes can be found through BLASTN alignment. In addition, using the covariance model of RFAM 11.0 family and INFERNAL (Version: 1.1) software with the default parameters built in RFAM, we can predict miRNA and snRNA sequence information on the genome (Nawrocki, 2014).

## 2.10 Protein family analysis

The software Diamond (Version: 2.0.6) and OrthoMCL (Version: release-5) with the default parameters were used to identify orthologous unigenes among species by default (Li, 2003; Buchfink et al., 2015). Briefly, all of the unigenes were blasted with each other,  $E\text{-value} < 1e-5$  and  $query\ coverage > 30\%$  were thought as orthologous genes. The orthologous genes between species were classified into one protein family, and other species specific genes were also classified into different families by OrthoMCL.

## 2.11 Function annotation of species specific genes

The KEGG and GO function annotation and enrichment analysis of obtained specific genes of each species were then conducted. GO enrichment analysis provides all GO terms that significantly enriched in species specific genes. All species specific genes were mapped to GO terms in the Gene Ontology database (<http://www.geneontology.org/>). Significantly enriched GO terms in species specific genes were defined by hypergeometric test. The calculated p-value were subjected to FDR correction, with the threshold of  $FDR \leq 0.05$ . KEGG is the major public pathway-related database. Pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in species specific genes comparing with the whole genome background. The calculating formula is the same as that in the GO analysis.

## 2.12 Evolution analysis

According to protein sequences of single copy orthologous genes, the phylogenetic tree of each species was constructed as follows:

- 1) The different protein sequences that belong to the same single copy orthologous gene family were gone through multiple alignment using software MUSCLE with the default parameters (Edgar, 2004).
- 2) The phylogenetic tree was constructed using NJ and ML methods. The NJ tree was constructed with MEGA, and the ML tree was constructed by igtree.
- 3) The Bootstrap method was used to test 1000 times to construct the final evolution trees.

## 2.13 RNA Extraction, library construction and sequencing

Total RNA of larve, one-quarter of adult, half of adult and adult was extracted using Trizol reagent kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Three replications were performed in above experiments. RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and checked using RNase free agarose gel electrophoresis. After total RNA was extracted, eukaryotic mRNA was enriched by Oligo(dT) beads. Then the enriched mRNA was fragmented into short fragments using fragmentation buffer and reversibly transcribed into cDNA by using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB #7530, New England Biolabs, Ipswich, MA, USA). The purified double-stranded cDNA fragments were end repaired and ligated to Illumina sequencing adapters. The ligation reaction was purified with the AMPure XP Beads. Ligated fragments were subjected to size selection by agarose gel electrophoresis and polymerase chain reaction (PCR) amplified. The resulting cDNA library was sequenced using Illumina Novaseq6000 by Gene Denovo Biotechnology Co. (Guangzhou, China).

## 2.14 Quantitative real-time PCR Validation

To validate the transcriptional level results from RNA-Seq data analysis, differentially expressed genes (DEGs) which were involved in the inosine biosynthesis in *S. nudus* were selected for quantitative real-time PCR validation. Real-time RT-PCR was performed using Real-Time PCR System (Bio-Rad). GAPDH was served as the internal control (reference gene) for normalization of the target gene expression and to correct for variation between samples. The thermal cycle for RT-PCR was as

follows: 95°C for 2 mins, followed by 40 cycles of 95°C for 10 s, 58°C for 15 s and 72°C for 20 s. Melting curve analyses of the amplification products were performed at the end of each PCR reaction to ensure that only specific products were amplified. Primers used for the candidate genes are designed using Primer Premier 5 and listed in Table S1. The comparative  $2^{-\Delta\Delta CT}$  method was employed to calculate relative expression levels between the target genes. Three replications were performed in above experiments.

## 2.15 Metabolites extraction

50 mg of sample was weighted to an EP tube. After the addition of 1000  $\mu$ L of extract solvent (acetonitrile-methanol-water, 2:2:1, containing internal standard), the samples were vortexed for 30 s, homogenized at 45 Hz for 4 min, and sonicated for 5 min in ice-water bath. The homogenate and sonicate circle was repeated for 3 times, followed by incubation at -20 °C for 1 h and centrifugation at 12000 rpm and 4 °C for 15 min. The resulting supernatants were transferred to LC-MS vials and stored at -80 °C until the UHPLC-QE Orbitrap/MS analysis. The quality control (QC) sample was prepared by mixing an equal aliquot of the supernatants from all of the samples. Six replications were performed in above experiments.

## 2.16 LC-MS/MS analysis

LC-MS/MS analysis were performed using an UHPLC system (1290, Agilent Technologies) with a UPLC HSS T3 column (2.1 mm  $\times$  100 mm, 1.8  $\mu$ m) coupled to Q Exactive (Orbitrap MS, Thermo). To obtain higher metabolite coverage and better detection effect, positive ion mode (POS) and negative ion mode (NEG) were used. The mobile phase A was 0.1% formic acid in water for positive, and 5 mmol/L ammonium acetate in water for negative, and the mobile phase B was acetonitrile. The elution gradient was set as follows: 0 min, 1% B; 1 min, 1% B; 8 min, 99% B; 10 min, 99% B; 10.1 min, 1% B; 12 min, 1% B. The flow rate was 0.5 mL/min. The injection volume was 2  $\mu$ L. The QE mass spectrometer was used for its ability to acquire MS/MS spectra on an information-dependent basis (IDA) during an LC/MS experiment. In this mode, the acquisition software (Xcalibur 4.0.27, Thermo) continuously evaluates the full scan survey MS data as it collects and triggers the acquisition of MS/MS spectra depending on preselected criteria.

For a preliminary visualization of differences between different groups of samples, the unsupervised dimensionality reduction method principal component analysis (PCA) was applied in all samples using R package models. In addition, orthogonal partial least-squares discriminant analysis (OPLS-DA) was applied in comparison groups using R package models

as well. The OPLS-DA model was further validated by cross-validation and 200 permutation test. For cross-validation, the data was partitioned into seven subsets, where each of the subsets was then used as a validation set (Wang et al., 2020).

## 3 Results

### 3.1 Quality and quantity of the sequencing and genome survey

Genomes of *S. nudus* are particularly challenging to be sequenced and assembled with short next-generation sequencing reads due to repetitive content. To solve this issue, PacBio sequencing and Illumina sequencing were combined for whole-genome sequencing of *S. nudus*. Based on PacBio sequencing, 246.36 Gb of read bases were generated and N50 of the PacBio reads was 27,456. The mean length of the PacBio reads was 18502.5. Through Illumina sequencing, 655 Mb clean reads were obtained and high quality of clean reads (removing reads with adapters and low quality reads) account for 98.26%. The Q20 of reads was 97.76% at least and the base composition was balanced in all samples, which showed the high quality of clean reads as well. Based on k-mer and mapping coverage, we estimated the genome size. The k-mer distribution is shown in Figure S1 and the k-mer depth was 61.86. The results showed that the genome size of *S. nudus* was estimated 1095 Mb and heterozygous ratio was 1.31%.

### 3.2 Summary of genome assembly and annotation for *S. nudus*

For genome assembly, considering the high error rate of third generation long reads data, pilon was used to correct potential sequence error in initial assembly with illumina short reads data. The final genome assembly was 1759 Mb, with a contig N50 size of 450 kb (Table 1). The GC content was 36.34%. After polishing, the quality and integrity of the assembly were demonstrated by the mapping of 97.05% paired-end reads. We also tested for the presence of 978 conserved Benchmarking Universal Single-Copy Orthologs (BUSCO) genes and found that 849 of the BUSCOs were completely captured in the assembly, including 819 of complete and single-copy BUSCOs, 30 of complete and duplicated BUSCOs, 45 of fragmented BUSCOs and 84 of missing BUSCOs (C: 86.8% [S: 83.7%, D: 3.1%], F: 4.6%, M: 8.6%, n: 978). These results indicate that the high integrity and accuracy of our assembled genome. Genome annotation was performed by a series of methods, including *de novo*, homology-based and EST/cDNA-based prediction. A total of 80391 protein-coding genes were predicted in *S. nudus* genome, and the average length of the genes was 10548. A total of 10119 single exon genes were predicted and the number

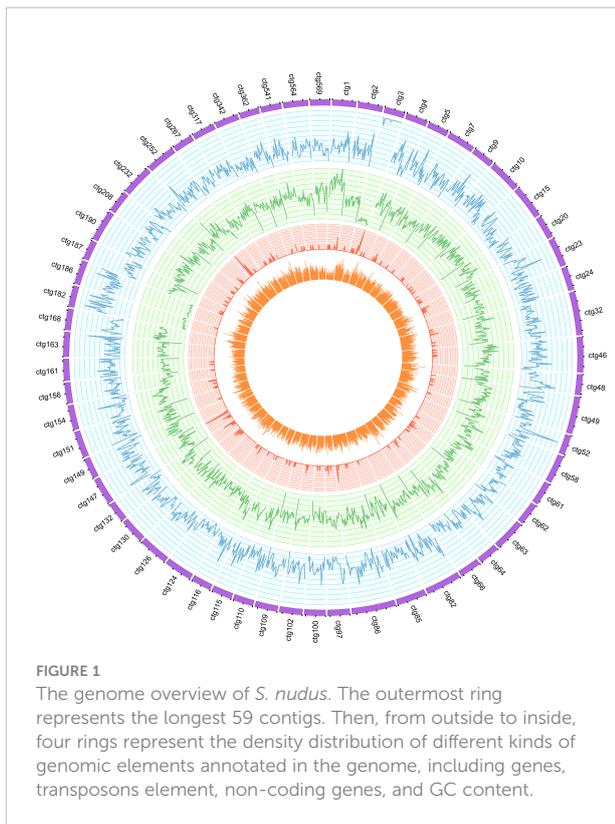
of exons per mRNA was 5.32 (Table 1). The results of repeat sequences annotation showed that the repeat content accounts for 56.19% (988 Mb) of the genome (Figure 1), dominated by transposons element, which are usually considered active modulators of genome evolution. The major type of repeats is long terminal repeat (LTR) and followed by helitron (Figure S2). BUSCO estimate was performed to check the quality of annotation and found that 783 of complete BUSCOs were annotated, including 490 of complete and single-copy BUSCOs, 293 complete and duplicated BUSCOs, 134 fragmented BUSCOs and 61 missing BUSCOs (C: 80.1% [S: 50.1%, D: 30%], F: 13.7%, M: 6.2%, n: 978). Furthermore, protein-coding genes were annotated in NR, Swissprot, GO, COG and KEGG database. The results showed that 40694 (51%) of genes were annotated in NR database, and 22769 (28%), 15135 (19%), 19729 (25%), 34137 (42%) were annotated in Swissprot, GO, COG and KEGG database, respectively.

### 3.3 Genome comparison and Phylogenetic analysis

Gene family analysis of *S. nudus* and other eight species, including *Batillaria attramentaria*, *Biomphalaria glabrata*, *Capitella teleta*, *Crassostrea gigas*, *Dimorphilus gyrocolliatus*, *Eisenia fetida*, *Eisenia andrei* and *Helobdella robusta*, identifies a core set of 1795 gene families (50299 genes, Figure 2A). Gene family comparison between *S. nudus* and other nine species showed that the most number of common gene families were shared between *S. nudus* and *C. teleta* (7438

TABLE 1 Summary for genome sequencing, assembly and annotation.

	Values
Total length (bp)	1759002260
Length of the longest sequence (bp)	4464325
N50 length (bp)	450734
GC content (%)	36.3421
Reads mapping to genome	97.05%
Sequencing coverage depth of PacBio sequencing	100X
Sequencing coverage depth of Illumina sequencing	50X
Number of predicted genes	80391
Gene Average Length	10547.96
mRNA Average Length	1925.96
CDS Average Length	912.91
Exon Average Length	228.21
Single exon genes	10119
Number of exons per mRNA	5.32



gene families). Referring to specific gene families for each species, GO and KEGG enrichment were performed. The top 20 of pathway enrichment for specific genes in *S. nudus* showed that most of these pathways were classified into metabolism (Figure 2B). The highest Q-value of enriched pathway was O-glycan biosynthesis (ko00514), followed by phagosome (ko04145) and tryptophan metabolism (ko00380). In addition, many pathways involved in amino acid metabolism were enriched, such as lysine degradation (ko00310), phenylalanine metabolism (ko00360), histidine metabolism (ko00340), arginine and proline metabolism (ko00330). These results suggested that *S. nudus* might be rich in amino acids.

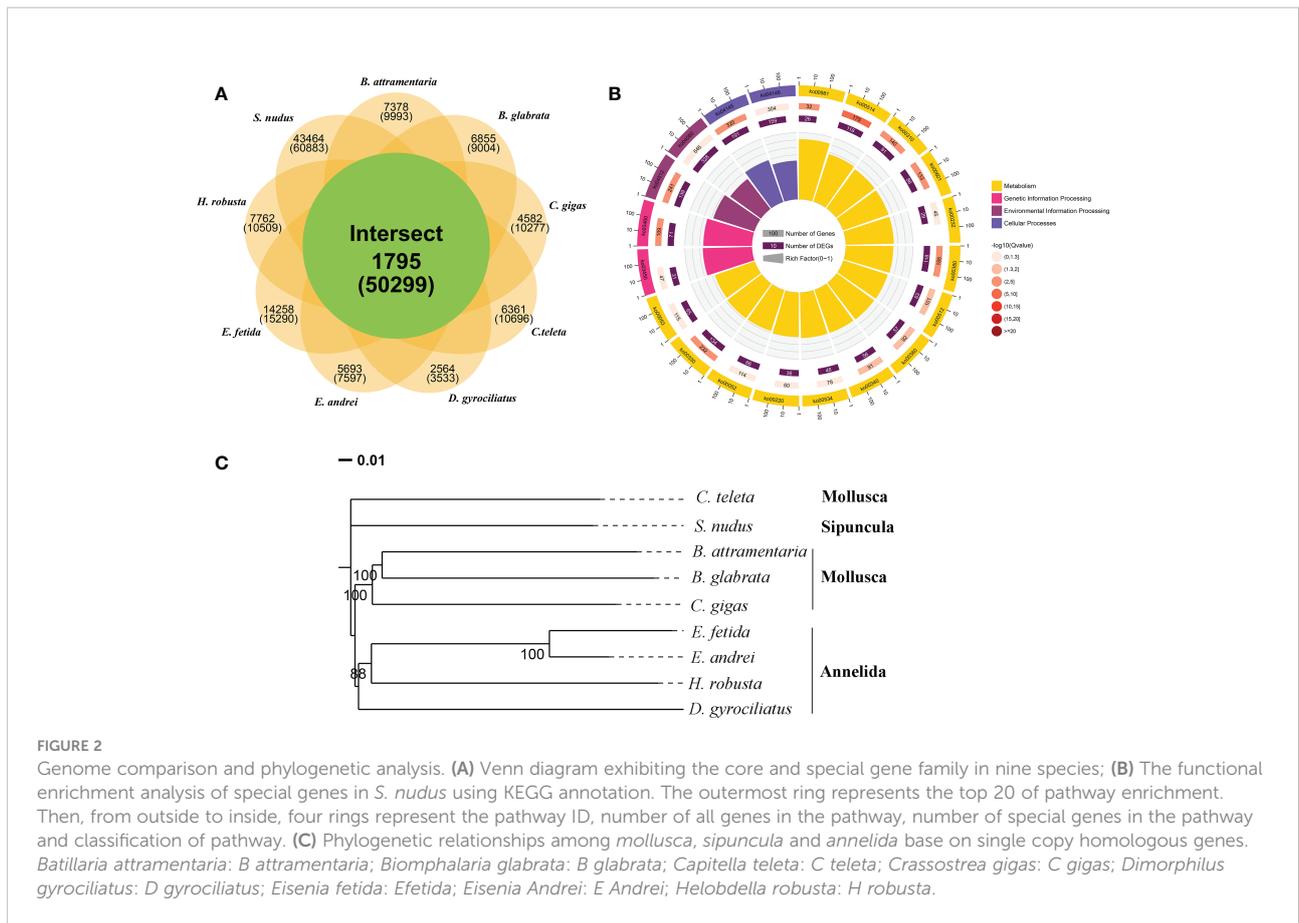
In the present study, we conducted phylogenetic tree among *Mollusca*, *Sipuncula* and *Annelida* base on single copy homologous genes (Figure 2C). Strikingly, our phylogenetic analysis revealed that *S. nudus* was clustered into *Mollusca*, suggesting that *S. nudus* belonged to *Mollusca* or was close to *Mollusca*, but distinct with *Annelida*. These results were consistent with the analysis of shared gene families between *S. nudus* and other nine species.

### 3.4 Gene expression analysis of *S. nudus* during growth

To characterize the patterns of gene expression during growth, four libraries were constructed using samples from various growth stages of *S. nudus*, with each sample consisting of three biological replicates (AR: larve samples; BR: one-quarter of adult; CR: half of adult; DR: adult). BUSCO estimate was performed to check the quality of the assembled transcriptome that was used for annotation and found that 712 of complete BUSCOs were annotated, including 516 of complete and single-copy BUSCOs, 196 complete and duplicated BUSCOs, 187 fragmented BUSCOs and 79 missing BUSCOs (C: 72.8% [S: 52.8%, D: 20%], F: 19.1%, M: 8.1%, n: 978). Differentially expressed genes (DEGs) across different growth stages were identified based on FPKM by applying cutoff of p-value < 0.05. The most number of DEGs were identified between one-quarter of adult and adult, comprising of 316 up-regulated genes (accounting for 73% of all significant differentially expressed genes) and 120 down-regulated genes (accounting for 27%). In addition, a total of 414 genes were differentially expressed between larve and adult, which comprised of 192 up-regulated genes and 222 down-regulated genes (Figure 3A). Only 16 DEGs were identified between one-quarter of adult and half of adult, with 11 up-regulated and 5 down-regulated. A Venn diagram of the distribution of DEGs is shown in Figure 3B. The results showed that 2793 DEGs were shared among the four groups. The number of specific DEGs in larve (1140) was remarkably greater than those in one-quarter of adult (179, Figure 3B). These results indicated the involvement of complex developmental events in larve.

### 3.5 Functional classification of differentially expressed genes

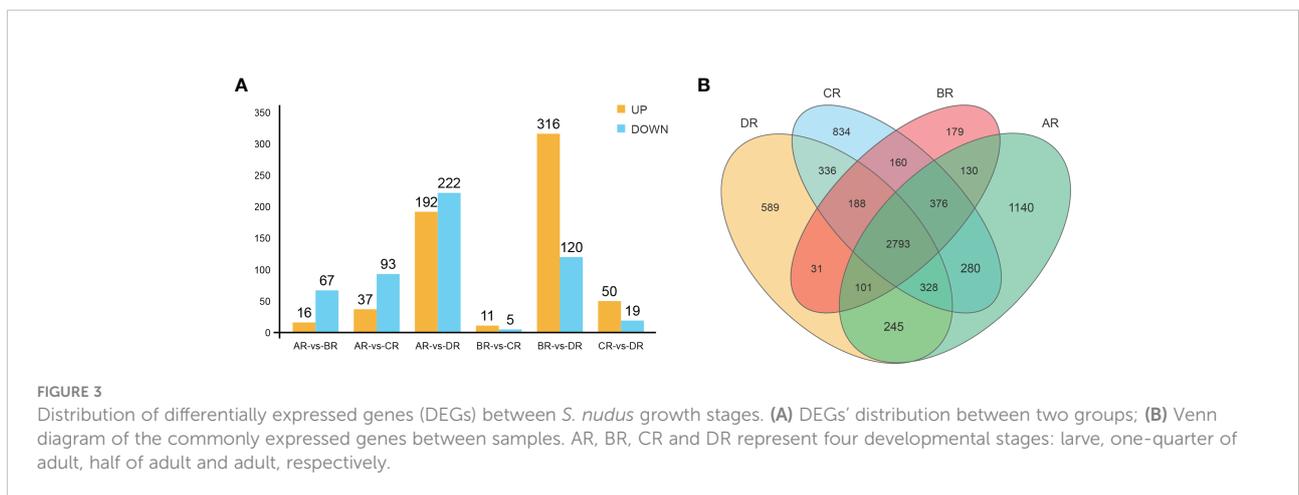
To characterize the functional differences between various growth stages of *S. nudus*, DEGs were analyzed by GO and KEGG enrichment analyses to explore the relevant biological functions. GO functional enrichment analysis revealed that the most gene number was enriched in groups of AR-VS-DR and BR-VS-DR, which was consistent with DEGs analysis (Figure S3). These significantly regulated genes were mainly related to cellular process, single-organism process, metabolic process and catalytic activity. To further investigate the function of these DEGs, we mapped the identified DEGs to specific biochemical pathways. The results showed that the most gene number was enriched in groups of AR-VS-DR and BR-VS-DR, which was in



agreement with GO functional enrichment. The results suggested that specific regulators were required for the transition from a larve to an adult. KEGG pathway analysis showed that these DEGs were mainly enriched in the pathways of amino acid metabolism, lipid metabolism and transport and catabolism (Figure 4).

### 3.6 Qualitative and quantitative analysis of metabolites

To obtain higher metabolite coverage and better detection effect, POS and NEG mode were used. Differential metabolites among different growth stages of *S. nudus* were identified. PCA



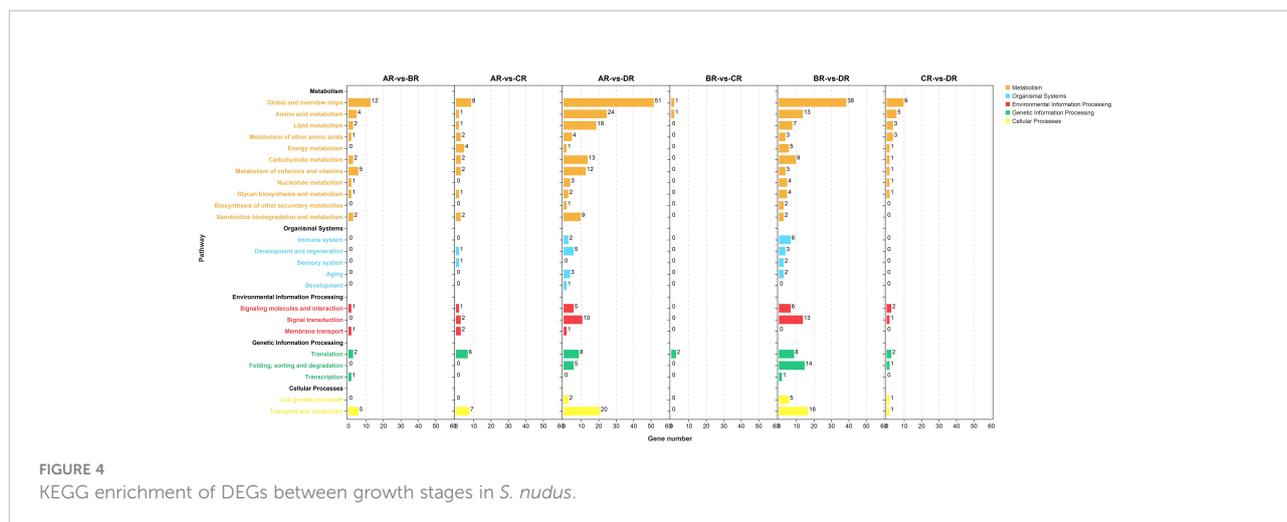


FIGURE 4 KEGG enrichment of DEGs between growth stages in *S. nudus*.

and OPLS-DA showed clear separation and clustering of sample groups (Figures S4, S5). Under the POS mode, the most number of differential metabolites were identified between larve and adult (313), comprising of 62 up-regulated metabolites (accounting for 20%) and 251 down-regulated metabolites (accounting for 80%, Figure 5). The top 10 metabolites in four developmental stages of *S. nudus* were identified and we found that the highest content of metabolites was similar, mainly including carnitine, proline, 3-dehydroxycarnitine, asparagine, inosine and methionine sulfoxide (Table S2). When under the NEG mode, the most number of differential metabolites were identified between larve and adult (516) as well as between one-quarter of adult and adult (562). The top 10 highest content of metabolites in four developmental stages of *S. nudus* were also similar, mainly including 2-hydroxyethanesulfonate, glycine, arginine, proline, 16-methylheptadecanoic acid, oleic acid and linoleic acid (Table S3). These results suggested that *S. nudus* was rich in a variety of nutritional and functional components, such as free amino acids, fatty acids and carnitine.

### 3.7 Gene expression and metabolic level involved in the inosine biosynthesis of *S. nudus* during different growth stages

Based on the qualitative and quantitative analyses of metabolites in *S. nudus*, the content of inosine was very high in whole growth stages. Therefore, we analyzed the level of gene expression and metabolites involved in the inosine biosynthesis of *S. nudus* during different growth stages. Based on the KEGG enrichment analysis, a total of 10 DEGs encoding 5 enzymes were identified in the inosine biosynthesis (Figure 6). In this pathway, there were two branches for the synthesis of the inosine from AMP. In the pathway from AMP to IMP for the inosine biosynthesis, genes encoding AMP deaminase (EC: 3.5.4.6) and nucleotidase (3.1.3.5) were high expressed in half of adult stage

and low expressed in adult stage. While in the other branch for the synthesis of the inosine from AMP, genes encoding purine-nucleoside phosphorylase (EC: 2.4.2.1), which catalyze the synthesis of adenine and inosine, had significantly increased expressions in the adult stage. To further confirm the expression profiles, DEGs in this pathway were selected for qRT-PCR analysis. The results of qRT-PCR had strong consistency with those of transcriptome analysis (Figure 7). According to the qualitative and quantitative analysis of metabolites in *S. nudus*, the content of adenine and inosine was highest in the adult stage, which was consistent with the gene expression level. The results may suggest that the inosine biosynthesis of *S. nudus* was mainly from AMP to inosine.

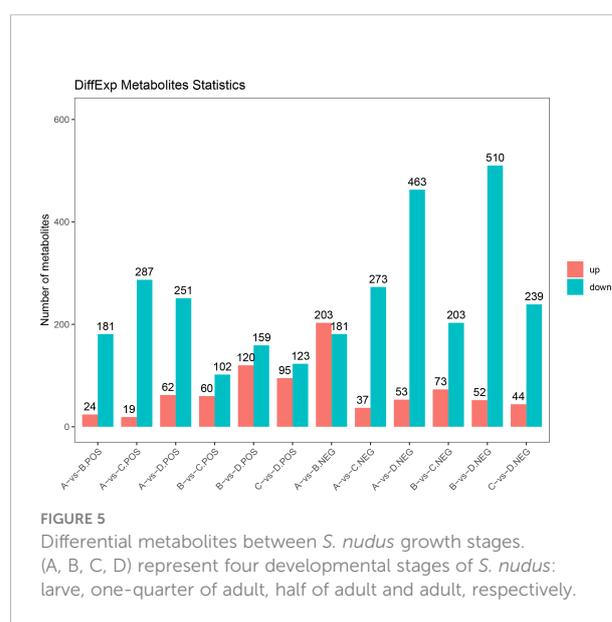
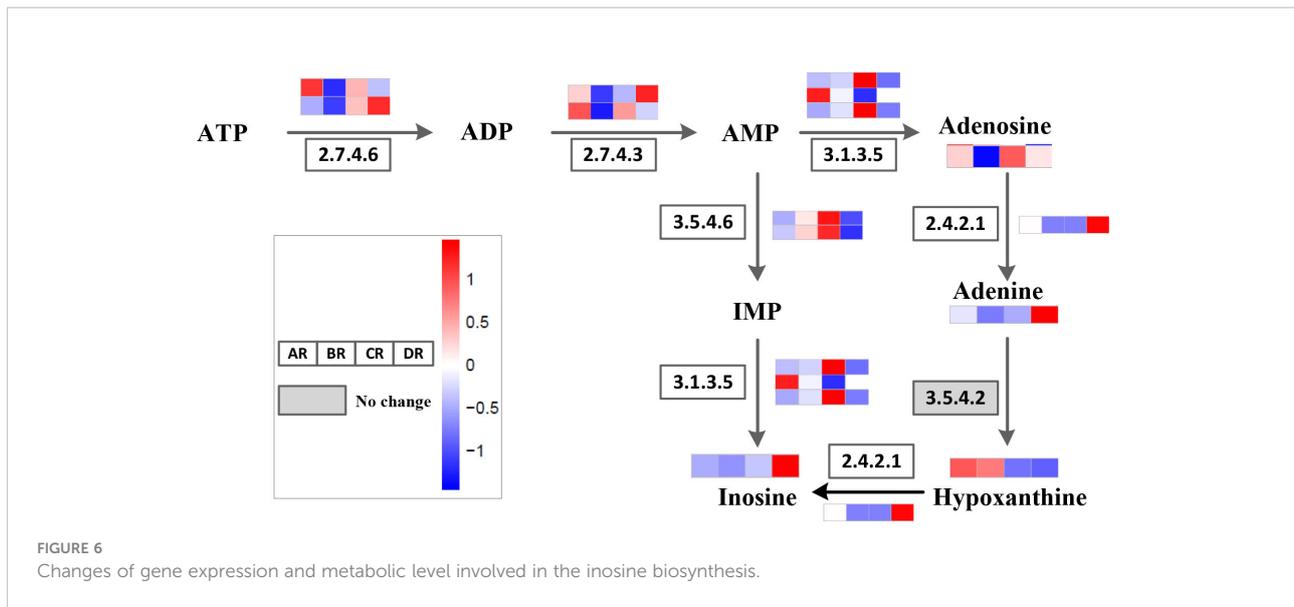


FIGURE 5 Differential metabolites between *S. nudus* growth stages. (A, B, C, D) represent four developmental stages of *S. nudus*: larve, one-quarter of adult, half of adult and adult, respectively.



### 3.8 Analysis of gene expression and metabolic level in the arginine and proline biosynthesis in *S. nudus* during different growth stages

The results of qualitative and quantitative analysis of metabolites revealed that the content of arginine and proline was very high in whole growth stages both under POS and NEG mode. Therefore, we investigated the expression profiles of genes and metabolite levels involved in the arginine and proline biosynthesis. In total, we identified 7 differentially expressed genes encoding 5 enzymes for the different steps in the arginine and proline biosynthesis (Figure 8). In the pathway from L-glutamate 5-carboxylate to arginine, genes were all up-regulated in half of adult stage corresponding to level of arginine production. From L-glutamate 5-carboxylate to proline, the level of L-glutamate 5-carboxylate and 1-pyrroline-5-carboxylate were highest in the larve stage, while the gene encoding pyrroline-5-carboxylate reductase (EC 1.5.1.2) show different expression pattern during different growth stages.

## 4 Discussion

Currently, sequencing projects for non-model organisms have revolutionized the field of biology and medical research. However, these new sequencing technologies also brought tremendous challenges to *de novo* assembly tools and strategies, because the quality of the assembled genome seriously affects the subsequent studies. Therefore, several novel *de novo* assembly tools and strategies have been developed, such as Abyss, Soapdenovo and so on. In this

study, PacBio sequencing and Illumina sequencing were combined for whole-genome sequencing of *S. nudus*. In this procedure, Illumina sequencing data were used to correct the PacBio sequencing data. The results showed a difference between the estimated genome size and the genome assembly. The reason of this difference we inferred was the heterozygosity of the *S. nudus* genome and these sequences which could not be recognized by Illumina sequencing but PacBio sequencing. In this study, heterozygous ratio was up to 1.31% which would be difficult to simplify the de Bruijn figure in the process of genome assembly and lead to the difference between the estimated genome size and the genome assembly. Furthermore, results of genome annotation showed that a total of 80391 protein-coding genes were predicted in *S. nudus* genome, which was more than some *Molluscs* and *Annelids*, such as *Lottia gigantea* (23800), *Capitella teleta* (32389) and *Helobdella robusta* (23400). We inferred that this is most likely influenced by high fragmentation of the assembly, which was confirmed by the BUSCO estimate (13.7% of fragmented BUSCOs were annotated and 19.1% fragmented BUSCOs in the assembled transcriptome). The other reason may be the repetitive content. There need to be further improved for the *S. nudus* genome while some other related species were sequenced.

*S. nudus* is a cosmopolitan species mainly distributed in tropical and subtropical coastal waters, which is usually used as a model organism for various fields of science. However, the classification status of this species is controversial. Although a subdivision into a posterior trunk and an anterior introvert that can be fully retracted into the trunk is present, the body of *S. nudus* shows no segmentation. The fossil records suggest that *Sipunculans* have undergone little changes over the past 520 million years. Early in 1767, Linnaeus placed *S. nudus* within

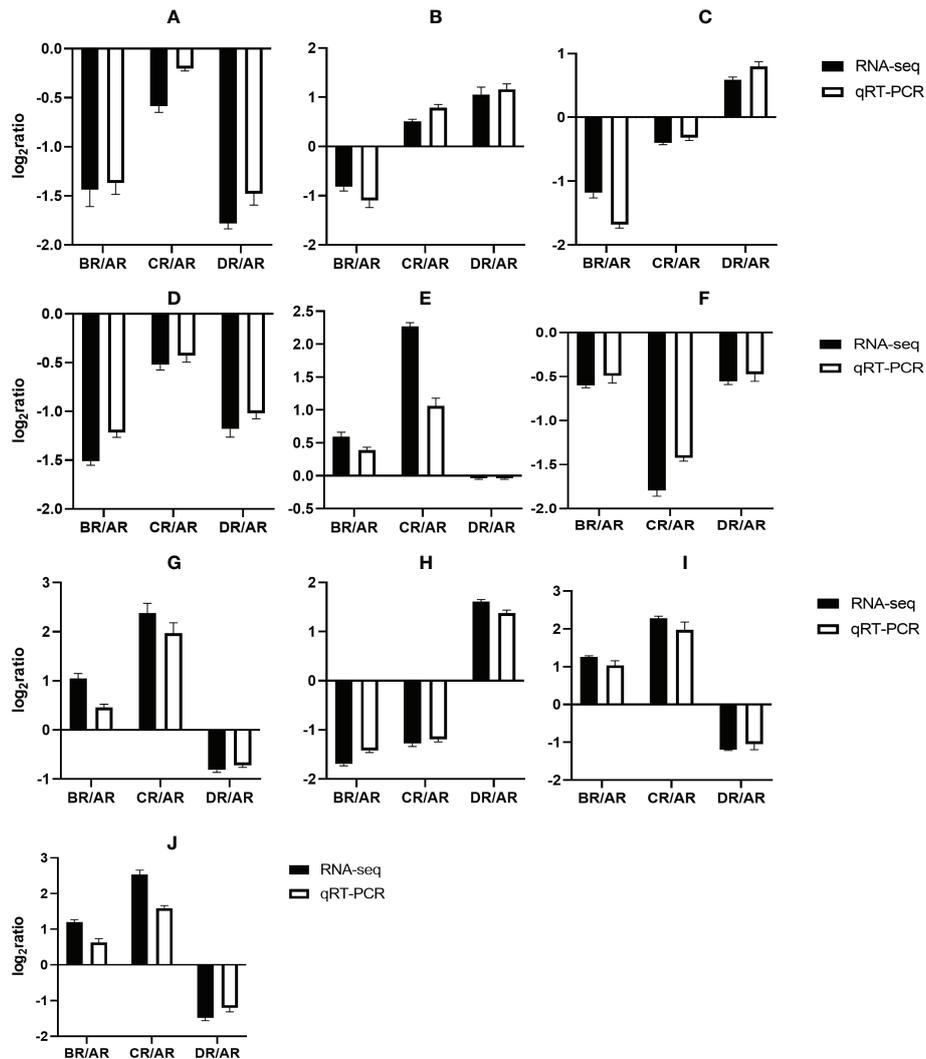
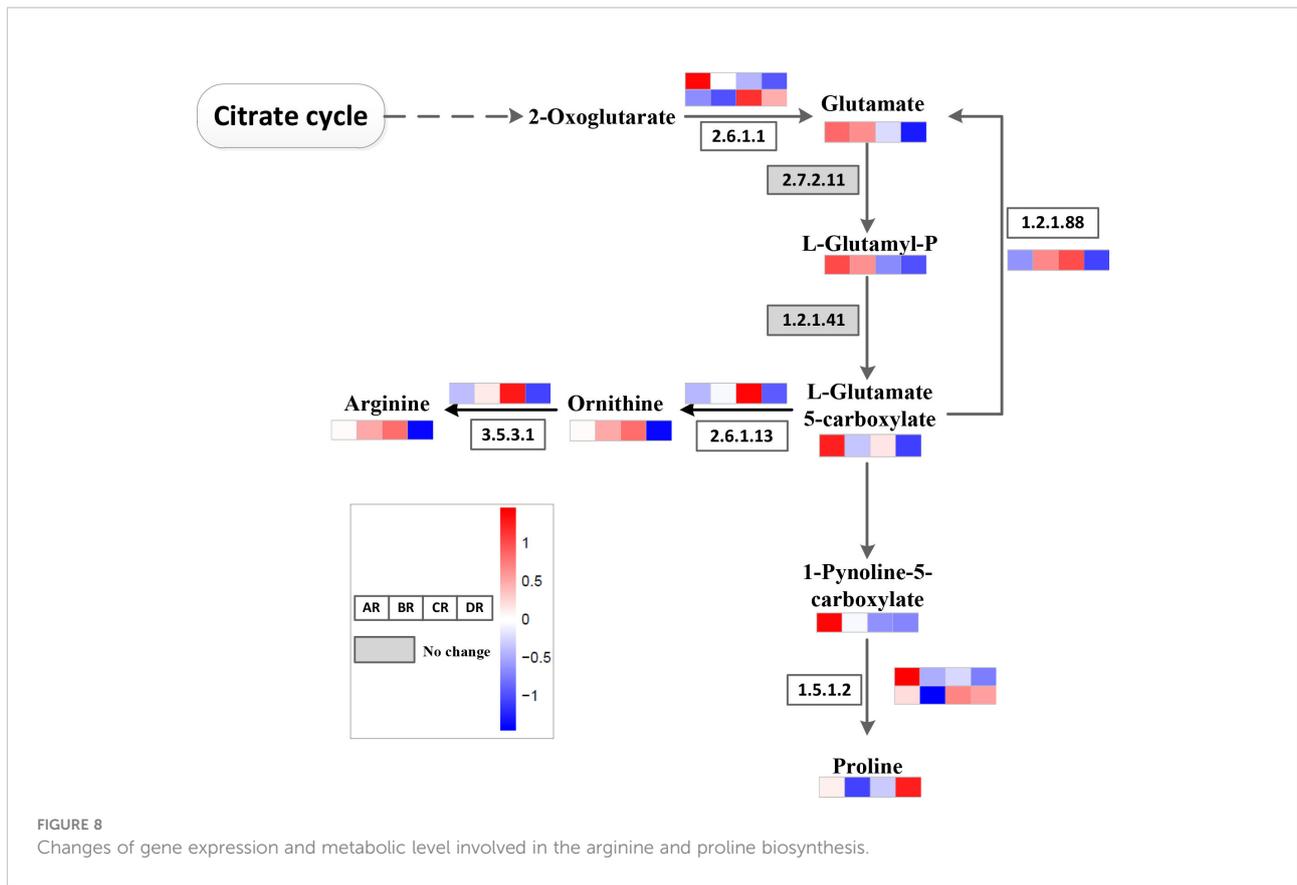


FIGURE 7

The expression pattern of DEGs involved in the inosine biosynthesis. (A, B) indicates genes encoding nucleoside-diphosphate kinase, (C, D) indicates genes encoding adenylate kinase, (E–G) indicates genes encoding nucleotidase, (H) indicates genes encoding purine nucleoside phosphorylase, (I, J) indicates genes encoding AMP deaminase. AR, BR, CR and DR represent four developmental stages of *S. nudus*: larve, one-quarter of adult, half of adult and adult, respectively.

the *Vermes Intestina*, a group containing truly “internal worms” which were later considered to as the derived group of *Annelids* (Linnaeus, 1767). In 1959, Hyman suggested the elevation of *Sipunculans* to phylum status and Stephen proposed the name *Sipuncula* for the phylum in 1965, which has been widely adopted (Hyman, 1959). Scheltema (1993) maintained the presence of a *Mollusca* during cleavage as an indication to place *Sipuncula* as the sister taxon to the *Mollusca* (Scheltema, 1993). However, cell lineage studies have shown that the concept of the *Mollus* vs. the *Annelida* is oversimplified and of limited phylogenetic significance (Nielsen and Meier, 2002). *Sipunculans* and *Echiurans* are often grouped together due to superficial body plan similarity (Dean,

2001). However, prominent differences including anal position and proboscis form suggest that the similar body plans are a result of convergence due to parallel burrowing lifestyles, rather than common ancestry. Recently, *Echiura* has been considered into the polychaete group that may have lost segmentation, leading to a more confusing placement of *Sipunculans* (Hessling and Westheide, 2002). Previous cladistic analyses based on morphological and complete mitochondrial genome data, have generated a great variety of hypotheses relating *Sipuncula*, including close relationship with *Annelida*, sister group to *Mollusca* (Eernisse et al., 1992; Song et al., 2014). However, Jennings and Halanych thought that it is



less significance for the phylogenetic analysis using the mt-genome data compared with whole genome sequences (Jennings and Halanych, 2005). In the present study, we conducted phylogenetic tree among *Mollusca*, *Sipuncula* and *Annelida* base on single copy homologous genes according to the whole genome sequences (Figure 2C). The results showed that *S. nudus* was a sister group to *Mollusca* or close to *mollusca*, but distinct with *Annelida*, which was supported by the absence of segmentation in *Sipuncula*.

In addition, genome comparison analysis of *S. nudus* and other nine species indicated that many KEGG pathways enriched for specific genes in *S. nudus* were involved in amino acid metabolism, such as tryptophan metabolism (ko00380). In addition, many pathways involved in amino acid metabolism were enriched, such as lysine degradation (ko00310), phenylalanine metabolism (ko00360), histidine metabolism (ko00340), arginine and proline metabolism (ko00330). These results were in good agreement with the metabolomic analysis. These results suggested that *S. nudus* was rich in amino acids. Liu et al. found that the content of amino acids of *S. nudus* in the Beihai Gulf run up to 68.02% and the content of glutamic acid was highest (Liu et al., 2016). Dong et al. analyzed the amino acids composition of *S. nudus* at larval, juvenile and adult stages and found that the crude protein contents at those 3 stages were

56.46%, 66.40% and 74.84%, respectively (Dong et al., 2012). Furthermore, Zhang et al. indicated that the sweetness amino acids, such as glycine and alanine, contribute to the unique flavour of the dried *S. nudus* (Zhang and Suzuki, 2000). These reports were further confirmed the results in the present study.

The previous research reported that *S. nudus* had multiple effects including anti-oxidation, anti-radiation, immune regulation, anti-bacterial, anti-inflammation and peripheral analgesia (Cui et al., 2014; Ge et al., 2018; Zhong et al., 2019). Some functional components were extracted from *S. nudus*. For example, Sun et al. found that soluble polysaccharides extracted from *S. nudus* can scavenge hydroxyl radicals and it is a natural antioxidant (Sun et al., 2017). In addition, some other antioxidant components, including unsaturated fatty acids, taurine and SOD, were also found in *S. nudus* (Wen et al., 2018). Furthermore, polysaccharide from *S. nudus* was reported to contribute to anti-fatigue activities (Liu et al., 2012). Collagen peptides derived from *S. nudus* was reported to contribute to accelerate wound healing (Lin et al., 2021). However, the detailed and systematic components of *S. nudus* extract were still not reported. In this study, we found that carnitine, free amino acids (such as proline and arginine), unsaturated fatty acids (oleic acid and linoleic acid), inosine, methionine sulfoxide were rich in *S. nudus*. Carnitine was the one of the *S. nudus* metabolites with

the highest content, which plays a critical role in energy production. Carnitine transports long-chain fatty acids into the mitochondria for oxidation and energy production. Therefore, carnitine was reported with the anti-fatigue effect (Ringseis et al., 2013). Additionally, carnitine transports the toxic compounds generated out of this cellular organelle to prevent their accumulation (Gupta et al., 2018). Inosine was another abundant metabolites in *S. nudus*, which was reported to have potent immunomodulatory and neuroprotective effects. Inosine enhances mast-cell degranulation, attenuates the production of pro-inflammatory mediators by macrophages, lymphocytes and neutrophils, and is protective in animal models of sepsis, ischemia-reperfusion and autoimmunity (Hagberg et al., 2010). Moreover, inosine preserves the viability of glial cells and neuronal cells during hypoxia, and stimulates axonal regrowth after injury. Furthermore, inosine has been used sporadically in clinical practice for various cardiovascular disorders, such as ischemic events (Haskó et al., 2004). We believe that these high levels of metabolites endow *S. nudus* anti-oxidation, anti-radiation, immune regulation, anti-bacterial, anti-inflammation and peripheral analgesia functions.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, SRP347907 and SRP344659.

## Author contributions

YQ and HL conceived and designed the experiments. LC, RL and BW performed the experiments. YQ and XT analyzed the data. XZ, YQ and KW contributed reagents/materials/analysis

tools. HL and LC wrote the paper. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Science and technology program of Guangdong Province (2019B090905011), Special Science and Technology Innovation Project of Guangdong Province, China (2019A01005, 2019A03023), Discipline construction project of Guangdong Medical University (4SG21009G), Special Support Project for Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang) [grant number: ZJW-2019-007].

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2022.1043311/full#supplementary-material>

## References

- Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6, 11. doi: 10.1186/s13100-015-0041-9
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12 (1), 59–60. doi: 10.1038/nmeth.3176
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421
- Campbell, M. S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-p. *Curr. Protoc. Bioinf.* 484, 11 11–11 39. doi: 10.1002/0471250953.bi0411s48
- Chan, P. P., and Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* 1962, 1–14. doi: 10.1007/978-1-4939-9173-0\_1
- Cui, F., Li, M., Chen, Y., Liu, Y., He, Y., Jiang, D., et al. (2014). Protective effects of polysaccharides from sipunculus nudus on beagle dogs exposed to  $\gamma$ -radiation. *PLoS One* 9 (8), e104299. doi: 10.1371/journal.pone.0104299
- Daren, C. C., Drew, R. S., Jacobo, R. V., Matthew, K. F., Audra, L. A., Sara, O. M., et al. (2014). Percent of the genome identified as repetitive elements by RepeatMasker. *PLoS One* 9, e106649.
- Dean, H. K. (2001). Marine biodiversity of Costa Rica: the phyla sipuncula and echiura. *Rev. Biol. Trop.* 49 Suppl 2 (Supl.2), 85–90.

- Dong, L., Zhang, Q., Tong, T., Mingzhu, X. U., and Chen, J. (2012). Amino acid composition of peanut worm sipunculus nudus at different growth stages. *South China Fisheries Sci.* 8 (05), 60–65.
- Eernisse, D. J., Albert, J. S., and Anderson, F. E. (1992). Annelida And Arthropoda are not sister taxa: A phylogenetic analysis of spiralian metazoan morphology. *Systematic Biol* 41 (3), 305–330.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32 (5), 1792–1797. doi: 10.1093/nar/gkh340
- Ge, Y. H., Chen, Y. Y., Zhou, G. S., Liu, X., Tang, Y. P., Liu, R., et al. (2018). A novel antithrombotic protease from marine worm sipunculus nudus. *Int. J. Mol. Sci.* 19 (10), 3023. doi: 10.3390/ijms19103023
- Gupta, A., Rawat, S., Gupta, P., and Agrahari, R. (2018). Clinical research and therapeutic importance of dietary supplement LCarnitine: Review. *Asian J. Pharm. Res.* 8 (1), 47–58. doi: 10.5958/2231-5691.2018.00010.2
- Hagberg, H., Andersson, P., Lacarewicz, J., Jacobson, I., Butcher, S., and Sandberg, M. (2010). Extracellular adenosine, inosine, hypoxanthine, and xanthine in relation to tissue nucleotides and purines in rat striatum during transient ischemia. *J. Neurochemistry* 49 (1), 227–231.
- Haskó, G., Sitkovsky, M. V., and Szabó, C. (2004). Immunomodulatory and neuroprotective effects of inosine. *Trends Pharmacol. Sci.* 25 (3), 152–157. doi: 10.1016/j.tips.2004.01.006
- Hessling, R., and Westheide, W. (2002). Are echiura derived from a segmented ancestor? immunohistochemical analysis of the nervous system in developmental stages of bonellia viridis. *J. Morphology* 252 (2), 100–113.
- Huson, D. H., Beier, S., Flade, I., Görska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition - interactive exploration and analysis of Large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12 (6), e1004957. doi: 10.1371/journal.pcbi.1004957
- Hyman, L. H. (1959). Phylum sipunculida. *Afr. Invertebrates* 5, 610–696.
- Jennings, R. M., and Halanych, K. M. (2005) Mitochondrial genomes of clymenella torquata (Maldanidae) and riftia pachyptila (Siboglinidae): Evidence for conserved gene order in Annelida. *Mol. Biol. Evol.* 2, 210. doi: 10.1093/molbev/msi008
- Jiang, S., Shen, X., Liu, Y., He, Y., Jiang, D., and Chen, W. (2015). Radioprotective effects of sipunculus nudus l. polysaccharide combined with WR-2721, rhIL-11 and rhG-CSF on radiation-injured mice. *J. Radiat. Res.* 56 (3), 515–522.
- Li, L. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13 (9), 2178–2189. doi: 10.1101/gr.1224503
- Li, H. (2017). Minimap2: fast pairwise alignment for long nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352
- Linnaeus, C. (1767). *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species cum characteribus et differentiis*. 896, Tomus II, [Regnum vegetabile Holmiae: Impensis Laurentii Salvii.
- Lin, H., Zheng, Z., Yuan, J., Zhang, C., and Qin, X. (2021). Collagen peptides derived from sipunculus nudus accelerate wound healing. *Molecules* 26 (5), 1385. doi: 10.3390/molecules26051385
- Liu, X. J., Peng, Y. H., and Huang, G. Q. (2016). Composition and nutritional evaluation of amino acids of peanut worm sipunculus nudus dry body from five sea areas of beihai, guangxi. *Chin. J. Ecology* 35 (3), 741–746.
- Liu, Y. M., Qian, T. T., Lin-Fang, M. O., Yin, H. E., Jiang, S. Q., and Shen, X. R. (2012). Study on the anti-fatigue effects of polysaccharides from sipunculus nudus in mice. *Chin. J. Mar. Drugs* 31 (3), 41–44.
- Majoros, W. H., Mihaela, P., Corina, A., and Salzberg, S. L. (2003). GlimmerM, economy and unveil: three ab initio eukaryotic gene finders. *Nucleic Acids Res.* 31 (13), 3601–3604. doi: 10.1093/nar/gkg527
- Manni, M., Berkeley, M. R., Seppely, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38 (10), 4647–4654. doi: 10.1093/molbev/msab199
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27 (6), 764–770. doi: 10.1093/bioinformatics/btr011
- Mirdita, M., Steinegger, M., and Söding, J. (2019). MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* 35 (16), 2856–2858. doi: 10.1093/bioinformatics/bty1057
- Nawrocki, E. P. (2014). Annotating functional RNAs in genomes using infernal. *Methods Mol. Biol.* 1097, 163–197. doi: 10.1007/978-1-62703-709-9\_9
- Nielsen, C., and Meier, R. (2002). What cell lineages tells us about the evolution of spiralia remains to be seen. *Evolution.* 56 (12), 2554–7
- Ringseis, R., Keller, J., and Eder, K. (2013). Mechanisms underlying the anti-wasting effect of l -carnitine supplementation under pathologic conditions: evidence from experimental and clinical studies. *Eur. J. Nutr.* 52 (5), 1421–1442. doi: 10.1007/s00394-013-0511-0
- Scheltema, A. H. (1993). Aplacophora as progenetic aculiferans and the coelomate origin of mollusks as the sister taxon of sipuncula. *Biol. Bull.* 184 (1), 57–78. doi: 10.2307/1542380
- Song, S. X., Ding, S. X., Yan, Q. P., and Qin, Y. X. (2014). Complete mitochondrial genome of sipunculus nudus (Sipuncula, sipunculidae). *Mitochondrial DNA* 27 (2), 1–2.
- Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33 (suppl\_2), W465–W467. doi: 10.1093/nar/gki458
- Stifanic, M., and Batel, R. (2007). Genscan for arabidopsis is a valuable tool for predicting sponge coding sequences. *Biologia* 62 (2), 124–127. doi: 10.2478/s11756-007-0037-0
- Su, J., Jiang, L., Wu, J., Liu, Z., and Wu, Y. (2016). Anti-tumor and anti-virus activity of polysaccharides extracted from sipunculus nudus(SNP) on Hepg2.2.15. *Int. J. Biol. Macromolecules* 87, 597–602. doi: 10.1016/j.ijbiomac.2016.03.022
- Sun, X., Wang, M., Liu, B., and Sun, Z. (2017). Purification and characterization of angiotensin I converting enzyme inhibition peptides from sandworm sipunculus nudus. *J. Ocean Univ. China* 16 (5), 911–915. doi: 10.1007/s11802-017-3293-9
- Thomas, B., Bernhard, L., Christoph, B., Achim, M., Adina, M., and Lars, P. (2009). Mitochondrial genome sequence and gene order of sipunculus nudus give additional support for an inclusion of sipuncula into Annelida. *BMC Genomics* 10 (1), 27–27.
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33 (14), 2202–2204. doi: 10.1093/bioinformatics/btx153
- Walker, B. J., Abeel, T., Shea, T., Priest, M., and Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9 (11), e112963. doi: 10.1371/journal.pone.0112963
- Wang, H., Ding, J., Ding, S. Y., and Chang, Y. Q. (2020). Integrated metabolomic and transcriptomic analyses identify critical genes in eicosapentaenoic acid biosynthesis and metabolism in the sea urchin stronglylucrotus intermedius. *Sci. Rep.* 10 (1), 1697. doi: 10.1038/s41598-020-58643-x
- Wen, C. H. E. N., Xiangjun, W. A. N. G., Dongxu, S. H. A. O., et al. (2018). Study on extraction process of sipunculus nudus polysaccharide and its antioxidant activity. *Agric. Biotechnol.*
- Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., et al. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14 (11), 1072–1074. doi: 10.1038/nmeth.4432
- Yahui, G., Yuping, T., Sheng, G., Xin, L., Zhenhua, Z., Lili, Z., et al. (2015). Simultaneous quantitation of free amino acids, nucleosides and nucleobases in sipunculus nudus by ultra-high performance liquid chromatography with triple quadrupole mass spectrometry. *Molecules* 21 (4), 408.
- Ysa, B., Gdb, C., Yz, A., Jp, D., Jwq, A., and Ekkb, E. (2021). Another blow to the conserved gene order in Annelida: evidence from mitochondrial genomes of the calcareous tubeworm genus hydroides. *Mol. Phylogenet. Evolution.* 160, 107124.
- Zhang, C., and Suzuki, T. (2000). Study on tasty substance and functional components of the dried sha chong sipunculus nudus. *J. Zhanjiang Ocean Univ.* 20 (2), 24–27.
- Zhong, Q., Wei, B., Wang, S., Ke, S., Chen, J., Zhang, H., et al. (2019). The antioxidant activity of polysaccharides derived from marine organisms: An overview. *Mar Drugs* 1712. doi: 10.3390/md17120674