fishes

MDPI

Article

Comprehensive Transcriptome Sequencing and Analysis of *Euspira gilva*: Insights into Aquaculture and Conservation

Zhixing Su [1,†], Jiayuan Xu [1,†], Xiaokang Lv [1], Xuefeng Song [1], Yanming Sui [2], Benjian Wang [3], Xiaoshan Wang [1], Bianbian Zhang [1], Baojun Tang [1,*] and Liguo Yang [1,4,*]

1 East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China; suzhixing@ecsf.ac.cn (Z.S.); xujiayuan@ecsf.ac.cn (J.X.); lvxiaokang@ecsf.ac.cn (X.L.); songxuefeng@ecsf.ac.cn (X.S.); wangxiaoshan@ecsf.ac.cn (X.W.); zhangbianbian@ecsf.ac.cn (B.Z.)
2 College of Marine and Biological Engineering, Yancheng Institute of Technology, Yancheng 224051, China; suiyanming@ecsf.ac.cn
3 Ocean and Fisheries Bureau of Fuding City, Fuding 355299, China; wangbenjian@ecsf.ac.cn
4 College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China
* Correspondence: tangbaojun@ecsf.ac.cn (B.T.); yangliguo@ecsf.ac.cn (L.Y.)
† These authors contributed equally to this work.

Abstract: *Euspira gilva*, a member of the family Naticidae, is predominantly found in intertidal soft mud, sandy soil, and sandy seabeds along the coast of China, where it is valued for its nutritional richness and significant economic value. This study presents a comprehensive transcriptome sequencing and analysis of *E. gilva* specimens from the Lianyungang area, yielding 3385 high-quality isoform sequences and 3310 non-redundant transcripts. Annotation against various databases, including NR, Swiss-Prot, KEGG, KOG, eggNOG, GO, and Pfam, successfully annotated a significant number of transcripts. A total of 7929 simple sequence repeat (SSR) loci were identified, with single nucleotide repeats predominating at 85.0%. Predictive analysis of coding DNA sequences (CDS) resulted in 1340 BLAST comparisons, while ESTScan predicted 840. Further, 530 long non-coding RNAs (lncRNAs) were identified through the application of the CPC2, CNCI, Pfam, and PLEK algorithms. The highest overall sequence similarity in the NR database was observed with *Pomacea canaliculata*, a freshwater species, but with a similarity of only 36.6%, indicating a unique genetic makeup of *E. gilva*. The KEGG database annotation revealed a predominance of signal transduction pathways, particularly the PI3K-Akt signaling pathway, with 29 non-redundant transcripts encoding key genes such as *IGH* (immunoglobulin heavy chain), *PCK* (phosphoenolpyruvate carboxykinase), *COL2A* (collagen, type II, alpha), *ITGB1* (integrin beta 1), and *GNG7* (guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-7). These genes play crucial roles in cellular processes, including cell growth, transcription, translation, proliferation, movement, and glycogen metabolism. The findings of this research elucidate the full-length transcriptome profile of *E. gilva*, thereby establishing a foundational dataset and providing valuable insights for the species' aquaculture, health management, conservation efforts, and future molecular biological investigations.

Keywords: *Euspira gilva*; full-length transcriptome; SSR; transcription factor; functional annotation

Citation: Su, Z.; Xu, J.; Lv, X.; Song, X.; Sui, Y.; Wang, B.; Wang, X.; Zhang, B.; Tang, B.; Yang, L. Comprehensive Transcriptome Sequencing and Analysis of *Euspira gilva*: Insights into Aquaculture and Conservation. *Fishes* **2024**, *9*, 483. https://doi.org/10.3390/fishes9120483

Academic Editor: Eric Hallerman

Received: 18 September 2024
Revised: 26 November 2024
Accepted: 27 November 2024
Published: 28 November 2024

1. Introduction

Euspira gilva, known as *Lunatia gilva* or *Laguncula pulchella*, belongs to the Phylum Mollusca, Class Gastropoda, Superorder Caenogastropoda, Family Naticidae, and Genus *Euspira* [1,2]. Family Naticidae is believed to have originated in the Late Triassic period (approximately 227 million years ago), exhibiting rich species diversity and interspecific morphological variation, with an estimated 260–270 modern species inhabiting intertidal zones to depths of several meters [1]. This species is found in the intertidal zones, thriving in soft mud, sandy soil, and sandy seabeds, predominantly along the coastal regions of China, as well as the Korean Peninsula and Japan [3,4]. Nocturnal predators by nature, *E. gilva*

migrate from the sea bottom to forage at night, primarily consuming bivalve mollusks and occasionally other gastropods. During predation, they ensnare their prey with their foot and secrete acidic substances near the heart area to bore a hole, allowing access to the soft parts of the prey with their proboscis and radula for feeding [3,5]. *E. gilva* is recognized in Japan as an invasive species, exerting a significant adverse impact on Japanese clam fisheries [3].

The egg masses of *E. gilva*, referred to as "sand collars", are composed of sand particles and egg chambers bound by a mucilaginous matrix, each collar containing a substantial number of eggs [5,6]. Previous research on *E. gilva* has focused on the histological observation of the digestive system [4], changes in organelles during vitellogenesis [7], nutritional component analysis [8], influence of shell morphological traits on body mass traits [9], impact of salinity on embryonic development and growth [10], effects of environmental heavy metal accumulation [11], and taxonomic studies. However, there is a dearth of information on the molecular biology of *E. gilva*, particularly in the areas of genetics and genomics [1]. Previous literature retrieval reveals that only the mitochondrial genome of *E. gilva* was reported in 2020, with a complete sequence length of 16,119 base pairs [12]. Phylogenetic analysis suggested a close phylogenetic relationship between *E. gilva* and both *Neverita didyma* and *Glossaulax reiniana* [12]. Similar to *N. didyma*, the populations of *E. gilva* are decreasing due to over-harvest [13].

With the advancement of high-throughput sequencing technologies, transcriptome sequencing has become increasingly applied to the study of gastropods [14,15]. Second-generation sequencing technologies typically yield shorter read lengths (ranging from 100 to 250 base pairs), necessitating further bioinformatics analysis and subsequent assembly using software like Trinity to obtain transcript sequences. However, the presence of numerous repetitive elements in the genomic DNA of gastropods can affect the assembly results, such as suboptimal N50 lengths of unigenes and a majority of non-full-length transcript sequences [16]. Third-generation sequencing technologies, also known as single-molecule real-time sequencing, include technologies developed by Pacific Biosciences and Oxford Nanopore Technologies, offering several advantages: (a) longer sequencing reads; (b) high accuracy; (c) high throughput; (d) effective acquisition of high-quality full-length transcript sequences by directly reading the full-length cDNA; (e) direct sequencing without the need for fragmentation or post-sequencing assembly; (f) analysis of alternative splicing; (g) lower sequencing costs; (h) shorter cycles [17]. At present, this technology has been gradually applied to the study of gastropod genomics [18,19].

In this study, we utilized the PacBio Sequel system to perform full-length transcriptome sequencing on *E. gilva*. This technology enables the discovery of new gene and transcript information in terms of gene structure, accurate identification of alternative splicing information, poly A site prediction, and gene fusion phenomena. Functional annotations are conducted by comparing various databases, and predictions and annotations are made for transcription factors (TFs), coding protein frames (CDS), long non-coding RNAs (LncRNAs), and simple sequence repeats (SSRs). This research not only enriches the full-length transcriptome database of Family Naticidae but also promotes the application of transcriptomics in *E. gilva*, laying a theoretical foundation for further research on this species.

2. Materials and Methods

2.1. Sample Collection

Adult specimens of *E. gilva*, exhibiting healthy appearance, were collected from Haizhou Bay in GanYu District, Lianyungang City, Jiangsu Province. The specimens were observed and measured at the East China Sea Fisheries Research Institute of the Chinese Academy of Fishery Sciences (Jiangsu GanYu Research Center, Shanghai, China). Tissue samples were dissected, rapidly frozen in liquid nitrogen, and stored at −80 °C in an ultra-low-temperature freezer for subsequent use.

2.2. Total RNA Extraction and Full-Length Transcriptome Sequencing

Total RNA was extracted from a mixed sample comprising tissues from three individuals using the RNAiso Plus reagent (TaKaRa, Dalian, China). This approach ensured a comprehensive representation of the transcriptome across different tissues and individuals. The integrity of the extracted RNA was assessed using 1% agarose gel electrophoresis to ensure the quality of the RNA for subsequent analyses. Once the RNA samples met the sequencing requirements, their concentration and purity were determined using a NanoDrop spectrophotometer (NanoDrop products, Thermo Scientific, Waltham, MA, USA). Full-length transcriptome sequencing of *E. gilva* was performed on this mixed RNA sample, aiming to capture a broad spectrum of transcripts, and was facilitated by Shanghai OE Bio-Tech Co., Ltd. (Shanghai, China).

2.3. Data Processing and Correction

For the PacBio SMRT sequencing raw data, adapters were removed and effective inserts (Subreads) were subjected to circular consensus sequence (CCS) analysis for mutual comparison and quality correction to obtain consensus sequences. Subsequently, the isoseq3 (https://isoseq.how/, accessed on 4 May 2023) tool was employed to call Lima and cluster for the removal of primers and poly A tails, as well as the identification and removal of chimera structures. Similar sequences were clustered to generate full-length non-chimera (FLNC) consensus sequences. The polish command of isoseq3 was then utilized to correct the FLNC sequences, resulting in polished consensus sequences that were further subjected to quality control to yield isoform sequences [17]. High-quality isoform sequences from all samples were subjected to redundancy removal and consensus sequence correction using the CD-Hit program (https://www.bioinformatics.org/cd-hit/, accessed on 4 May 2023), followed by clustering.

2.4. Non-Redundant Transcript Annotation Using Databases

The diamond software (version 0.9.9, https://www.crystalimpact.com/diamond/, accessed on 5 May 2023) was used for DNA sequence alignment with an e-value cutoff of $<1 \times 10^{-5}$ to select annotations with the highest sequence similarity to proteins. Functional annotation information was obtained by comparing against databases such as NR (https://www.ncbi.nlm.nih.gov/, accessed on 5 May 2023), COG (Clusters of Orthologous Groups)/KOG (https://www.ncbi.nlm.nih.gov/COG/, accessed on 5 May 2023), GO (https://www.geneontology.org/), Swissprot (http://www.expasy.ch/sprot, accessed on 5 May 2023), eggNOG (http://eggnog-mapper.embl.de/, accessed on 5 May 2023), and KEGG (https://www.genome.jp/kegg/, accessed on 6 May 2023). The highest-scoring families were identified through multiple sequence alignment and hidden Markov model validation in the Pfam (http://pfam-legacy.xfam.org/, accessed on 6 May 2023) database.

2.5. Transcription Factor (TF) Annotation

Non-redundant transcript sequences were aligned to the Animal Transcription Factor Database (AnimalTFDB) (http://bioinfo.life.hust.edu.cn/AnimalTFDB4/, accessed on 7 May 2023) using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed on 7 May 2023), with the best match being selected to obtain transcription factor family annotation information [20].

2.6. Coding DNA Sequence (CDS) Prediction

Non-redundant transcript sequences were aligned with the NR, COG/KOG, and Swissprot protein databases using BLAST with an e-value cutoff of $<1 \times 10^{-5}$. The highest-ranking protein in the alignment results was used to determine the coding region of the transcript, which was then translated into an amino acid sequence.

2.7. Long Non-Coding RNA (LncRNA) Prediction and Annotation

Transcripts longer than 200 bp and devoid of open reading frames (ORFs) greater than 300 were selected. Transcripts annotated in the coding database were removed from the annotation information. The remaining transcripts were subjected to prediction and analysis using software such as CPC2 (CPC2-beta) [21], CNCI (version 2) [22], Pfam (version 1.3) [23], and PLEK (version 2) [24] to eliminate transcripts with coding potential, thereby identifying the target LncRNAs.

2.8. SSR and Transposon Element Amplification Prediction

Dispersed repeat sequences, including DNA transposons and retrotransposons that transpose via a DNA–DNA mechanism, were predicted using the RepeatMasker software (version 4.1.4) [25]. Common types of retrotransposons include Long Terminal Repeat (LTR), Long Interspersed Nuclear Element (LINE), Short Interspersed Nuclear Element (SINE), and DNA transposons.

3. Results

3.1. Full-Length Transcriptome Data Statistics

The PacBio SMRT sequencing yielded a comprehensive transcriptome database for *E. gilva*, comprising 501,209 circular consensus sequence reads derived from 19,000,704 raw sequences and 3385 high-quality isoform sequences. Subsequent analyses were based on Unigene sequences identified at a 98.0% similarity threshold, resulting in 3310 non-redundant transcript sequences primarily ranging from 60 to 10,000 base pairs in length (Figure 1).

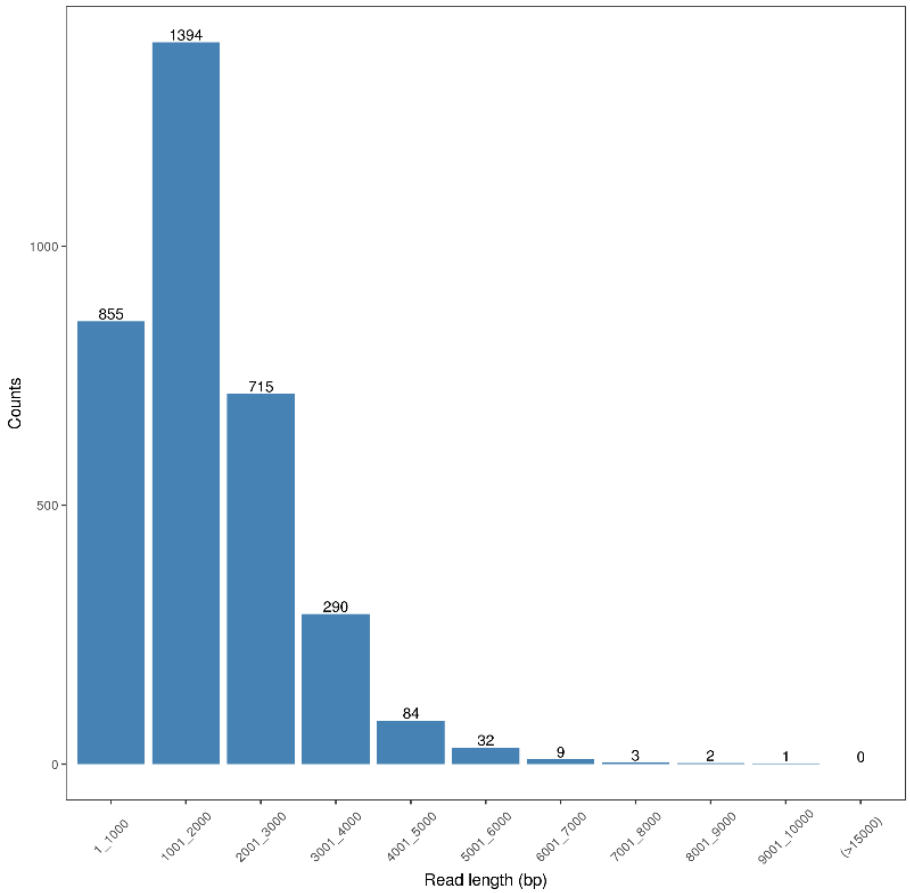

Figure 1. Distribution of sequence lengths of non-redundant transcripts in the transcriptome of *E. gilva*. The abscissa represents the sequence length distribution interval, and the ordinate represents the sequence counts.

3.2. Database Annotation

The full-length transcriptome of *E. gilva* was annotated using seven databases, resulting in 1296, 872, 636, 791, 1045, 827, and 1406 transcripts successfully annotated in the NR, Swiss-Prot, KEGG, KOG, eggNOG, GO, and Pfam databases, respectively. The results of the database annotation are presented in Table 1. Notably, the highest similarity in the NR database annotation was observed with *Pomacea canaliculata* (36.6%) (Figure 2). The KEGG database annotation indicated a predominance of signal transduction pathways (Figure 3). In the GO database annotation, the largest number of annotations was associated with cellular components, specifically 'cell' (711) and 'cell part' (708) (Figure 4).

Table 1. Functional annotation of *E. gilva* non-redundant transcripts.

Anno Database	Annotated Number	300 ≤ Length < 1000	Length ≥ 1000
NR	1296 (41.67%)	189 (6.08%)	1101 (35.40%)
Swissprot	872 (28.04%)	140 (4.50%)	729 (23.44%)
KEGG	636 (20.45%)	115 (3.70%)	520 (16.72%)
KOG	791 (25.43%)	140 (4.50%)	648 (20.84%)
eggNOG	1045 (33.60%)	185 (5.95%)	854 (27.46%)
GO	827 (26.59%)	134 (4.31%)	690 (22.19%)
Pfam	1406 (45.21%)	206 (6.62%)	1200 (38.59%)

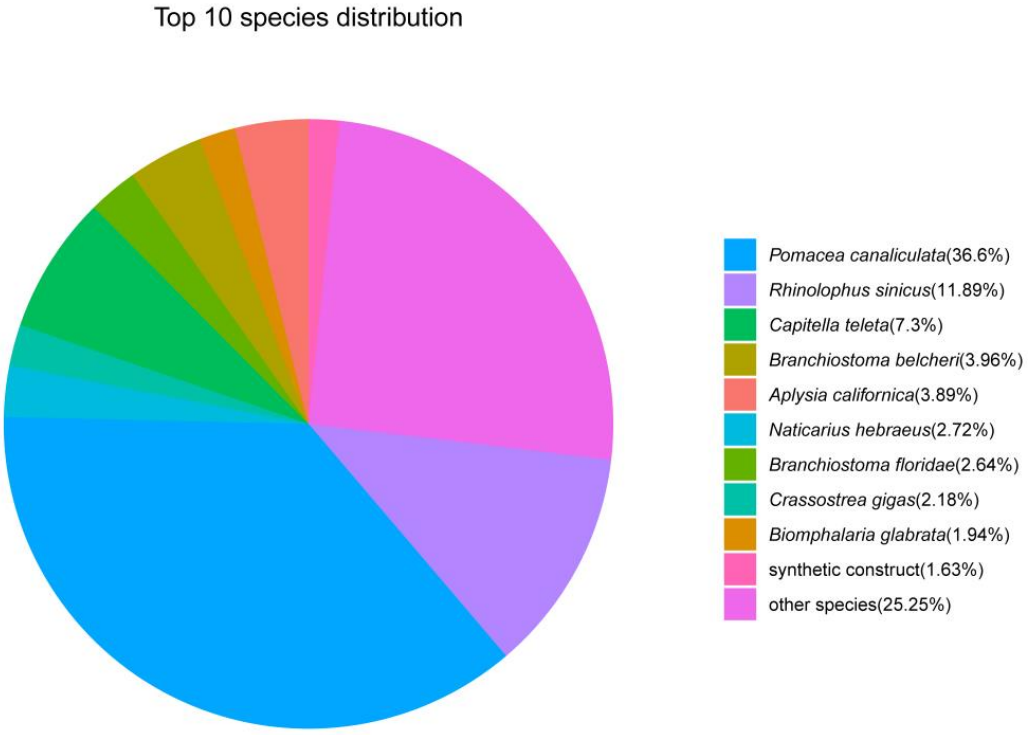

Figure 2. Top 10 homologous species distribution within the NR database. Pie slices in different colors represent the proportion of each species noted.

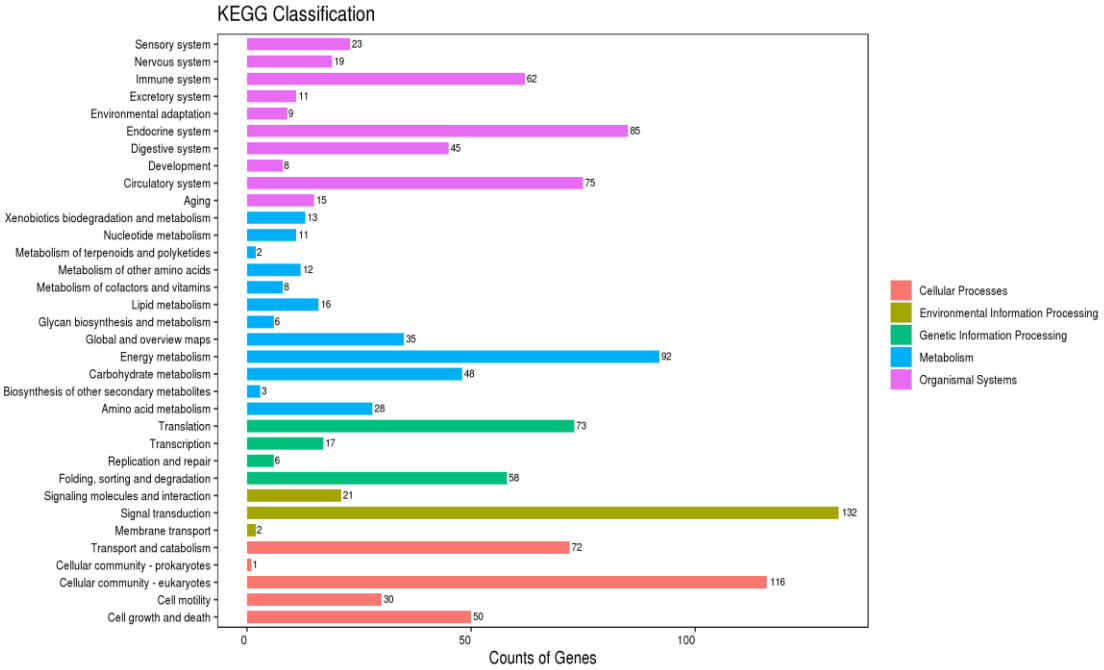

Figure 3. KEGG annotation statistics chart at Level 2. The horizontal axis represents the number of genes, the vertical axis represents the name of the Level 2 pathway, and the number on the right side of the column represents the number of genes annotated to the Level 2 pathway.

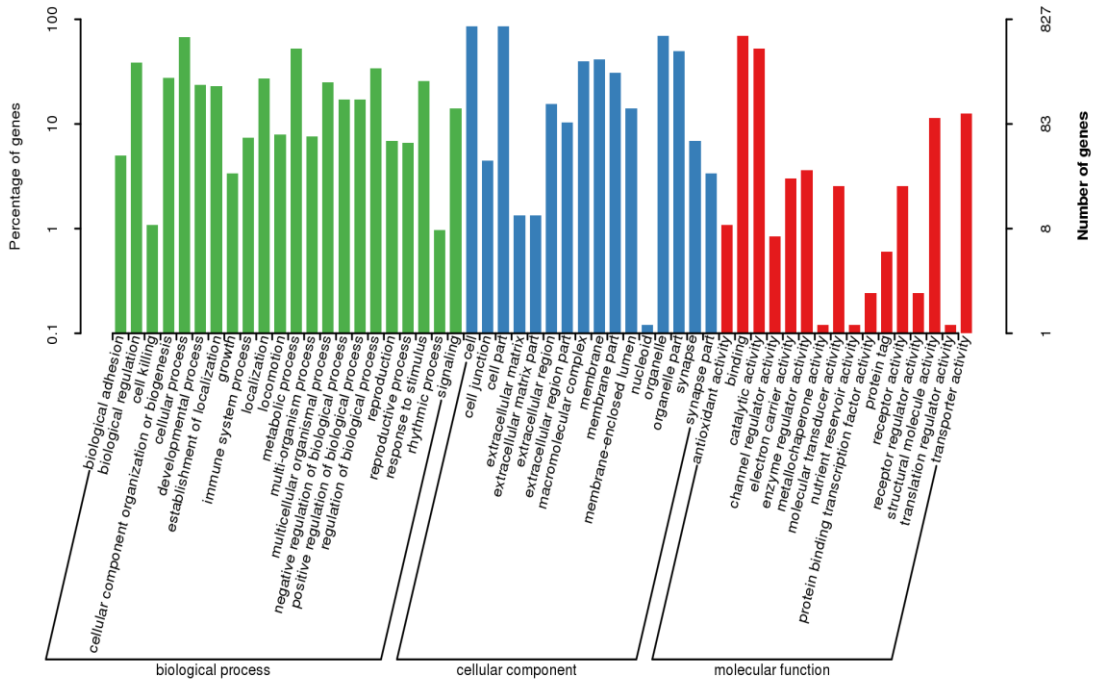

Figure 4. GO function classification statistics map. The horizontal axis represents the GO function classification, the left vertical axis represents the proportion of genes annotated to this category, and the right vertical axis represents the number of genes annotated to this category.

3.3. Transcription Factor (TF) Annotation Analysis

The AnimalTFDB, which encompasses sequences from 75 species and 70 animal transcription factor families, was utilized for annotation. A total of 52 transcription factors distributed across 19 families were annotated from the *E. gilva* transcriptome, with the

Zf-C2H2 family (10 members) and the *bHLH* family (8 members) being the most represented (Figure 5).

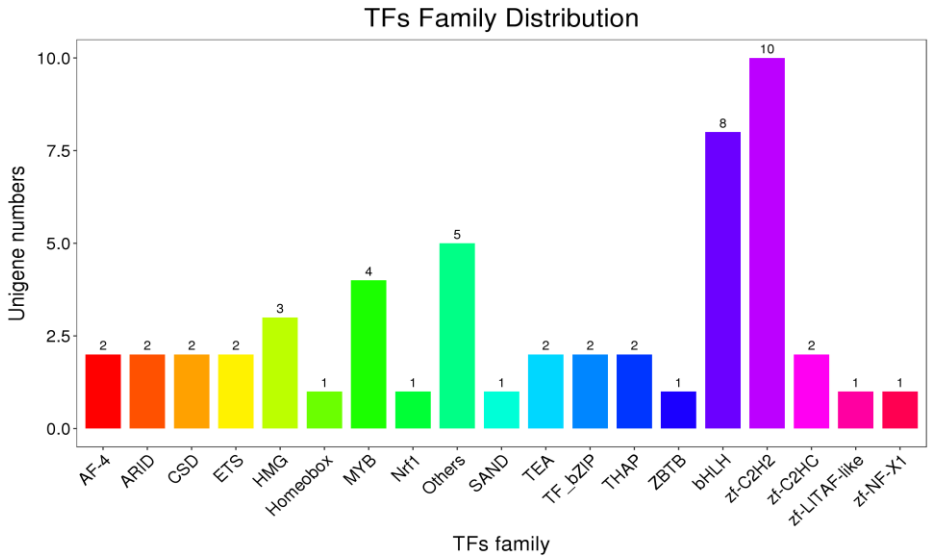

Figure 5. TF family Unigene distribution map. The abscissa represents the transcription factor family, and the ordinate represents the number of transcription factors contained in the transcription factor family.

3.4. Coding DNA Sequence (CDS) Prediction Analysis

CDS prediction analysis is instrumental for understanding gene length distribution and serves as a foundation for subsequent protein and key component analyses. In the *E. gilva* transcriptome, 1340 transcripts were annotated through BLAST comparison, and 840 were predicted by ESTScan. The most prevalent CDS length range was 401–600 base pairs, with 252 transcripts falling within this range (Figure 6).

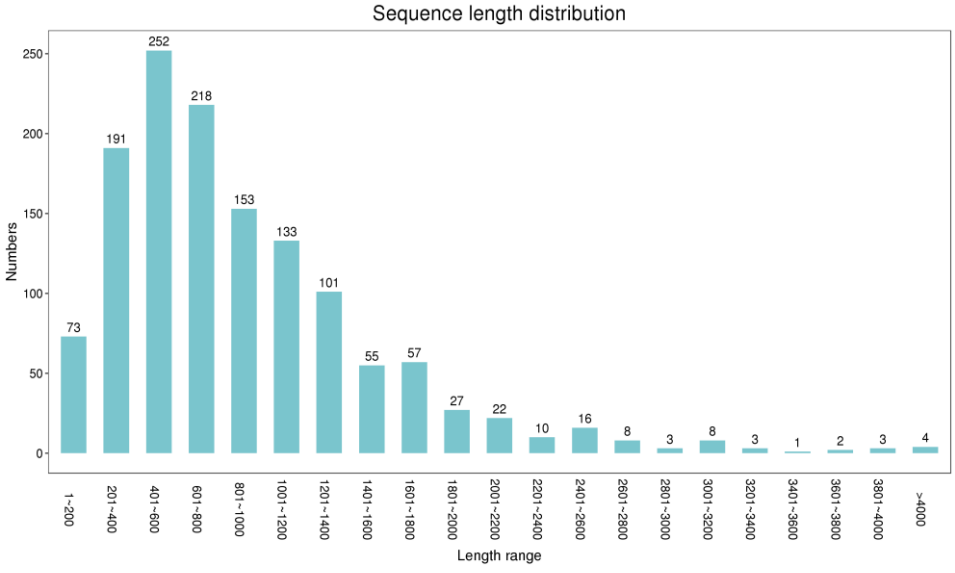

Figure 6. CDS sequence length distribution map. The abscissa represents the CDS length distribution interval, and the ordinate represents the number of CDS sequences.

3.5. Long Non-Coding RNA (LncRNA) Analysis

LncRNAs, which are RNA molecules longer than 200 base pairs and lacking protein-coding capacity, were identified based on their characteristics. A total of 530 LncRNAs

were filtered using CPC2, CNCI, Pfam, and PLEK, with an average length of 948 base pairs and the most common length range being 1701–2200 base pairs (Figure 7).

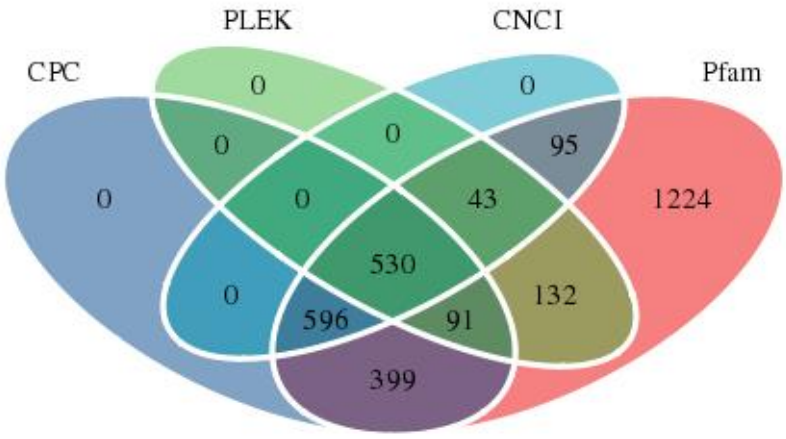

Figure 7. Venn diagram of lncRNA transcripts identified from CNCI, Pfam, PLEK, and CPC.

3.6. SSR and Transposon Element Analysis

SSR analysis using the MISA software (version 2.1) identified 7929 SSR loci, with single-nucleotide repeats being the most abundant (85.0%), followed by dinucleotide repeats (11.4%), trinucleotide repeats (3.2%), and tetra-, penta-, and hexanucleotide repeats (0.30%, 0.06%, and 0.04%, respectively) (Figure 8). The Repeat Masker software (version 4.1.4) was employed for the prediction of interspersed repeat elements, revealing that the most common were long interspersed repeat elements, with 26 instances detected.

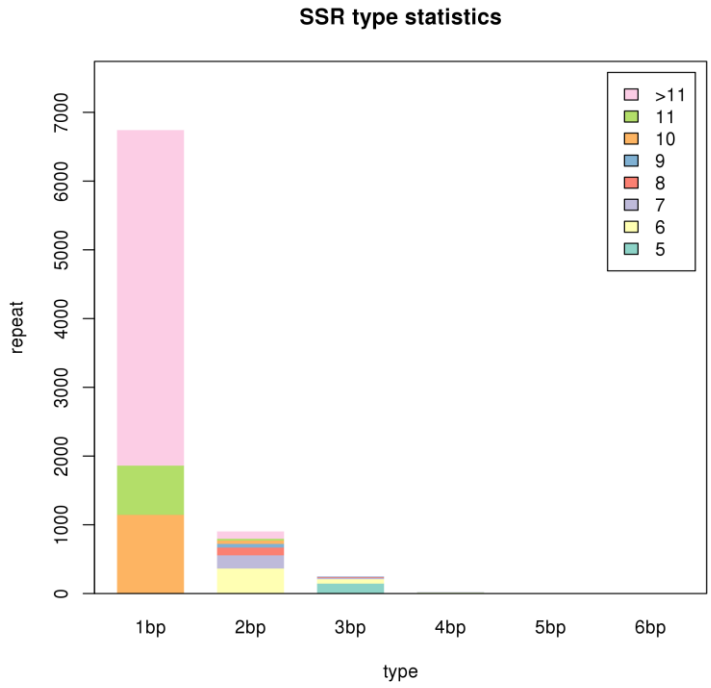

Figure 8. SSR type statistical results. The abscissa is the type of repeat, and the ordinate is the number of repeats of each type. (For example, 1 bp is a single-base repeat, 2 bp is a two-base repeat, the number of repetitions range from 5 to >11, and the ordinate is the sum of their corresponding frequencies).

4. Discussion

Euspira gilva, a significant mollusk species in China, is distributed along the coastal intertidal zones from north to south, primarily feeding on estuarine shellfish, and is a

natural enemy of aquacultured shellfish species [3]. It is also valued as a high-protein, low-fat, and low-sugar seafood delicacy with certain medicinal properties [8]. To date, while full-length transcriptome sequencing has been progressively applied to aquatic organisms, its application to the study of the Family Naticidae, to which *E. gilva* belongs, remains scarce [1]. To establish a full-length transcriptome database for *E. gilva* and to delve into its genomic data, providing support for research on this species and Family Naticidae, this study employed high-throughput PacBio SMRT sequencing technology to sequence the full-length transcriptome of *E. gilva* and conducted bioinformatics analysis on the obtained transcriptome data.

We obtained 3310 non-redundant transcripts, which is lower than the results from high-throughput Illumina sequencing of *N. didyma* [15]. This discrepancy may be attributed to the short read lengths of Illumina second-generation sequencing, incomplete splicing of transcripts, and susceptibility to errors in functional annotation and gene prediction, as well as possible species-specific factors [26]. The application of Pacific SMRT technology in this study yielded 3385 high-quality isoform sequences, and the annotation efficiency of the non-redundant transcripts obtained using this technology was also high. The full-length transcripts identified in this study will facilitate further research on *E. gilva*.

In terms of annotation, the highest similarity in the NR database annotation of *E. gilva* was with *P. canaliculata*, a freshwater species with a predominantly herbivorous diet, with a similarity of only 36.6%. The intrinsic connection of this similarity within Order Mesogastropoda warrants further investigation [27]. KEGG database annotation revealed a predominance of signal transduction pathways, specifically the PI3K-Akt signaling pathway, with 29 non-redundant transcripts accurately annotated to encode key genes such as *IGH* (immunoglobulin heavy chain), *PCK* (phosphoenolpyruvate carboxykinase), *COL2A* (collagen, type II, alpha), *ITGB1* (integrin beta 1), and *GNG7* (guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-7). The PI3K-Akt signaling pathway, activated by various cellular stimuli and toxins, is involved in the regulation of many fundamental cellular processes, including cell growth, transcription, translation, proliferation, movement, and glycogen metabolism [28]. The acquisition of key genes in these pathways will provide sequences to support molecular biological studies of *E. gilva* and advance research into the regulatory mechanisms of this species.

Long non-coding RNAs (lncRNAs), a class of non-coding RNAs with lengths exceeding 200 nucleotides, are increasingly recognized for their pivotal roles in a multitude of physiological processes. The significance of lncRNAs in invertebrate biology is becoming more evident, particularly with research on *Drosophila demonstrating* their integral involvement in sex determination [29]. Analogous findings have been documented in various gastropod species, including *Haliotis diversicolorata*, *T. atra*, and *P. canaliculata*, underscoring the extensive influence of lncRNAs on gastropod physiology and behavior [27,30,31]. In the ecologically and economically significant species *E. gilva*, we employed three distinct bioinformatics tools to identify 530 prevalent lncRNAs, among which are *lincRNA-p21*, *MALAT1*, and *FMR1-AS1*. *LincRNA-p21* has been implicated in the biological behavior of tumor cells, notably in the suppression of tumor cell invasion and metastasis [32]. *MALAT1* exerts a regulatory influence on transcription, pre-mRNA splicing, microRNA sequestration, and other cellular processes [33]. *FMR1-AS1*, through its interaction with TLR7, is capable of activating the NF-κB signaling pathway [34].

Despite the limited research on lncRNAs in gastropods, studies in aquatic animals have demonstrated that shell and body coloration are regulated by both miRNAs and lncRNAs [35,36]. *E. gilva*, predominantly characterized by a black hue, has exhibited occasional instances of black degeneration, with a mutation rate of approximately 1.5% based on unpublished data from our research group. Melanin synthesis, a complex biological process, is governed by a suite of genes, with tyrosinase being a critical rate-limiting enzyme that directly influences melanin formation [37]. The specific regulatory mechanisms underlying this variation in *E. gilva* remain to be elucidated. We hypothesize that such variations in shell color may be orchestrated by a complex interplay of regulatory factors, where

lncRNAs, miRNAs, and mRNAs interact within a ceRNA network, potentially modulating cellular activity and function, and thereby impacting physiological processes [38].

The SSR molecular marker technique has been widely applied in population genetic mapping, molecular marker-assisted breeding, and other research areas. For instance, SSR development and analysis in the yellow-mouthed snail indicated a relatively low level of genetic diversity in the population of the Yu Shan Islands [39]. High-frequency and diverse types of SSR sites were also found in *N. didyma* [15]. The discovery of 7929 SSR sites in this experiment will provide further basis for molecular markers, breeding, and genetic diversity analysis in *E. gilva*. The majority of SSRs in *E. gilva* are single nucleotides, while in *N. didyma* (63.4%), single nucleotides are slightly lower than in *E. gilva* (85.0%), with dinucleotide and trinucleotide types being more abundant; the mono-nucleotide repeat motifs of A/T was also the predominant repeat type [15]. This may be related to species differences but requires further validation. Furthermore, TF analysis will provide data supporting the study of the immune and transcriptional regulation of *E. gilva*.

5. Conclusions

Our study successfully employed PacBio Single Molecule Real-Time (SMRT) sequencing technology to achieve a comprehensive acquisition and functional annotation of the full-length transcriptome of *Euspira gilva*. The transcriptomic data obtained are characterized by a higher number of unigenes and an increased average length, as well as a significantly larger number of complete Open Reading Frames identified, than data obtained through Illumina sequencing methods. This accomplishment not only substantially enriches the gene repository for Family Naticidae but also provides a solid foundation for future research on behavioral ecology, selective breeding, and the development of molecular markers specific to *E. gilva*. The identification of 3385 high-quality isoform sequences and 3310 non-redundant transcripts, along with the annotation of key genes involved in signal transduction pathways such as the PI3K-Akt signaling pathway, underscores the depth and breadth of our transcriptomic analysis. The low sequence similarity of 36.6% with *Pomacea canaliculata* in the NR database highlights the unique genetic composition of *E. gilva*, emphasizing the importance of our work in uncovering the genetic underpinnings of this ecologically and economically important species. Furthermore, the identification of 7929 SSR loci and 530 lncRNAs through advanced bioinformatics algorithms contributes to the understanding of genetic diversity and regulatory mechanisms in *E. gilva*. The comprehensive transcriptomic profile elucidated in this study is expected to serve as a catalyst for further discoveries and applications, particularly in the fields of aquaculture, health management, and conservation of this species. In summary, our work has laid the groundwork for subsequent investigations into the molecular biology and genetic improvement of *E. gilva*, offering valuable insights that will inform ecological studies and sustainable management practices for this species.

Author Contributions: Conceptualization, L.Y.; methodology, Z.S. and J.X.; formal analysis, X.L. and X.S.; investigation, B.W., X.W. and B.Z.; data curation, Z.S., J.X. and L.Y.; writing—original draft preparation, B.T. and L.Y.; writing—review and editing, B.T. and Y.S.; funding acquisition, B.T. and L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported Ningbo City Key Research and Development Plan and the 'Open Solicitation of Project Leaders' Project (Grant 2023Z113).

Institutional Review Board Statement: The study was approved by East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences (approval code: 2024018 and approval date: 30 January 2024).

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, H.; Yang, Y.; Sun, S.; Kong, L.; Li, Q. Mitogenomic phylogeny of the Naticidae (Gastropoda: Littorinimorpha) reveals monophyly of the Polinicinae. *Zool. Scr.* **2020**, *49*, 295–306. [CrossRef]
2. Feng, J.; Guo, Y.; Yan, C.; Ye, Y.; Li, J.; Guo, B.; Lü, Z. Sequence comparison of the mitochondrial genomes in two species of the genus *Nerita* (Gastropoda: Neritimorpha: Neritidae): Phylogenetic implications and divergence time estimation for *Neritimorpha*. *Mol. Biol. Rep.* **2020**, *47*, 7903–7916. [CrossRef] [PubMed]
3. Yoshida, K.; Setogawa, T.; Sato, T.; Yamada, M.; Sato, T.; Narita, K.; Matsumoto, A.; Tomiyama, T. Male-biased sex ratio in the crawling individuals of an invasive naticid snail during summer: Implications for population management. *Sci. Rep.* **2022**, *12*, 7911. [CrossRef] [PubMed]
4. Ying, X.P.; Zhang, Y.P.; Ni, Q.Y.; Yang, W.X. Histological and histochemical studies on the digestive system of *Lunatia gilva*. *J. Zhejiang Univ. Sci.* **2003**, *30*, 75–78.
5. Huelsken, T.; Marek, C.; Schreiber, S.; Schmidt, I.; Hollmann, M. The naticidae (mollusca: Gastropoda) of giglio island (tuscany, italy): Shell characters, live animals, and a molecular analysis of egg masses. *Zootaxa* **2008**, *1770*, 1–40. [CrossRef]
6. Giglioli, M.E.C. The egg masses of the naticidae (gastropoda). *J. Fish. Res. Board Can.* **1955**, *12*, 287–327. [CrossRef]
7. Ying, X.P.; Yang, W.X. Organelle variations in oocytes with vitellogenesis of *Lunatia gilva*. *J. Zhejiang Univ. Sci.* **2003**, *30*, 204–209.
8. Zhang, P.; Zhong, W.; Ji, D.W.; Ma, J.Z.; Wang, T.G.; Lin, S.Z.; Huang, X.K.; Zhang, L.N. Analysis and evaluation of nutritional composition of *Lunatia gilva*. *J. Zhejiang Univ. Sci.* **2013**, *32*, 398–402.
9. Huang, X.K.; Xiao, G.Q.; Zhang, P.; Zhang, Y.P.; Wang, T.G.; Zhong, W.; Zhang, L.N.; Lin, S.Z. The relationship between morphometric traits and body weight of *Lunatica gilva*. *J. Aquacult.* **2017**, *38*, 25–29.
10. Wang, Y.N.; You, Z.J.; Chen, S.; Ye, X.H. Effects of various sea water specific gravity on embryonic development, larva growth and surviving of moon snall, *Lunatica gilva*. *J. Zhejiang Coll. Fish.* **1995**, 231–237.
11. Zhao, W.H.; Chen, X.G.; Fei, Z.H.; Qian, B.J. Preliminary study on ability of accumulation of heavy metal in different Mollusca and their different tissues and organs from Yancheng seashore bottomland. *J. Shanghai Jiaotong Univ.* **2009**, *27*, 76–78.
12. Wang, Z.; Xu, Z.; Wang, B.; Zheng, F.; Chen, W.; Zhou, Q. Characterization of the complete mitochondrial genome of *Euspira gilva* with phylogenetic analysis. *Mitochondrial DNA Part B* **2020**, *5*, 1479–1480. [CrossRef]
13. Zhao, D.; Kong, L.; Yu, H.; Li, Q. Cryptic genetic diversity of *Neverita didyma* in the coast of China revealed by phylogeographic analysis: Implications for management and conservation. *Conserv Genet.* **2018**, *19*, 275–282. [CrossRef]
14. Albertin, C.B.; Bonnaud, L.; Brown, C.T.; Crookes-Goodson, W.J.; da Fonseca, R.R.; Ragsdale, C.W. Cephalopod genomics: A plan of strategies and organization. *Stand. Genom. Sci.* **2012**, *7*, 175–188. [CrossRef] [PubMed]
15. Lu, W.X.; Chen, Y.X.; Qi, P.Z. Analysis of ssr information in the transcriptional group of *Neverita didyma*. *J. Zhejiang Ocean Univ.* **2018**, *37*, 215–220.
16. Farhat, S.; Modica, M.V.; Puillandre, N. Whole Genome Duplication and Gene Evolution in the Hyperdiverse Venomous Gastropods. *Mol. Biol. Evol.* **2023**, *40*, msad171. [CrossRef]
17. Abdel-Ghany, S.E.; Hamilton, M.; Jacobi, J.L.; Ngam, P.; Devitt, N.; Schilkey, F.; Ben-Hur, A.; Reddy, A.S. A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **2016**, *7*, 11706. [CrossRef]
18. Pardos-Blas, J.R.; Irisarri, I.; Abalde, S.; Afonso, C.M.L.; Tenorio, M.J.; Zardoya, R. The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *Gigascience* **2021**, *10*, giab037. [CrossRef]
19. Goodheart, J.A.; Rio, R.A.; Taraporevala, N.F.; Fiorenza, R.A.; Barnes, S.R.; Morrill, K. A chromosome-level genome for the nudibranch gastropod *Berghia stephanieae* helps parse clade-specific gene expression in novel and conserved phenotypes. *bioRxiv* **2023**, *22*, 9. [CrossRef]
20. Hu, H.; Miao, Y.R.; Jia, L.H.; Yu, Q.Y.; Zhang, Q.; Guo, A.Y. AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **2019**, *47*, D33–D38. [CrossRef]
21. Kang, Y.J.; Yang, D.C.; Kong, L.; Hou, M.; Meng, Y.Q.; Wei, L.; Gao, G. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [CrossRef] [PubMed]
22. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. [CrossRef]
23. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [CrossRef] [PubMed]
24. Li, A.; Zhang, J.; Zhou, Z. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinf.* **2014**, *15*, 311. [CrossRef]
25. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* **2009**, *25*, 4–10. [CrossRef]
26. Huang, A.C.; Su, J.Y.; Hung, Y.J.; Chiang, H.L.; Chen, Y.T.; Huang, Y.T.; Yu, C.A.; Lin, H.N.; Lin, C.L. SpliceAPP: An interactive web server to predict splicing errors arising from human mutations. *BMC Genom.* **2024**, *25*, 600. [CrossRef]
27. Xiao, Q.; Lin, Y.; Li, H.; Chen, Y.; Wei, W.; Li, P.; Chen, L. Transcriptome sequencing reveals the differentially expressed lncRNAs and mRNAs in response to cold acclimation and cold stress in *Pomacea canaliculata*. *BMC Genom.* **2022**, *23*, 382. [CrossRef]
28. Fruman, D.A.; Chiu, H.; Hopkins, B.D.; Bagrodia, S.; Cantley, L.C.; Abraham, R.T. The PI3K Pathway in Human Disease. *Cell* **2017**, *170*, 605–635. [CrossRef] [PubMed]

29. Mulvey, B.B.; Olcese, U.; Cabrera, J.R.; Horabin, J.I. An interactive network of long non-coding RNAs facilitates the Drosophila sex determination decision. *Biochim. Biophys. Acta* **2014**, *1839*, 773–784. [CrossRef] [PubMed]
30. Lu, G.; Yao, C.; Zhang, X.; Sun, Y.; Wang, Y.; Zhang, Z. Differentially expressed lncRNAs involved in immune responses of *Haliotis diversicolor* and *H. discus hannai* challenged with *Vibrio parahaemolyticus*. *Comp. Biochem. Physiol. D Genom. Proteom.* **2021**, *40*, 100873. [CrossRef]
31. Détrée, C.; Núñez-Acuña, G.; Tapia, F.; Gallardo-Escárate, C. Long non-coding RNAs are associated with spatiotemporal gene expression profiles in the marine gastropod *Tegula atra*. *Mar. Genom.* **2017**, *33*, 39–45. [CrossRef] [PubMed]
32. Huang, Y.; Yi, Q.; Feng, J.; Xie, W.; Sun, W.; Sun, W. The role of lincRNA-p21 in regulating the biology of cancer cells. *Hum. Cell* **2022**, *35*, 1640–1649. [CrossRef] [PubMed]
33. Xu, W.W.; Jin, J.; Wu, X.Y.; Ren, Q.L.; Farzaneh, M. MALAT1-related signaling pathways in colorectal cancer. *Cancer Cell Int.* **2022**, *22*, 126. [CrossRef] [PubMed]
34. Li, W.; Zhang, L.; Guo, B.; Deng, J.; Wu, S.; Li, F.; Wang, Y.; Lu, J.; Zhou, Y. Exosomal FMR1-AS1 facilitates maintaining cancer stem-like cell dynamic equilibrium via TLR7/NFκB/c-Myc signaling in female esophageal carcinoma. *Mol. Cancer* **2019**, *18*, 22. [CrossRef] [PubMed]
35. Luo, M.; Wang, L.; Yin, H.; Zhu, W.; Fu, J.; Dong, Z. Integrated analysis of long non-coding RNA and mRNA expression in different colored skin of koi carp. *BMC Genom.* **2019**, *20*, 515. [CrossRef]
36. Chen, S.; Nie, H.; Huo, Z.; Yan, X. Comprehensive analysis of differentially expressed mRNA, lncRNA and miRNA, and their ceRNA networks in the regulation of shell color in the Manila clam (*Ruditapes philippinarum*). *Int. J. Biol. Macromol.* **2024**, *256*, 128404. [CrossRef]
37. Ando, H.; Kondoh, H.; Ichihashi, M.; Hearing, V.J. Approaches to identify inhibitors of melanin biosynthesis via the quality control of tyrosinase. *J. Investig. Dermatol.* **2007**, *127*, 751–761. [CrossRef]
38. Ren, Y.; Chen, J.; Wang, Y.; Fu, S.; Bu, W.; Xue, H. The lncRNA-mediated ceRNA network of *Altica viridicyanea* is involved in the regulation of the Toll/Imd signaling pathway under antibiotic treatment. *Front. Physiol.* **2023**, *14*, 1244190. [CrossRef] [PubMed]
39. Li, W.; Zhao, S.; Jiao, H.F.; Lin, Z.H.; Bao, Y.B. Characterization and analysis of microsatellite markers in thais luteostoma using next generation sequencing. *Mar. Sci.* **2015**, *39*, 61–67.