Journal of *Marine Science and Engineering*

MDPI

Article

Metagenomic Profiling of Bacterial Communities and Functional Genes in *Penaeus monodon*

Juan Chen [1,2,†], Yundong Li [1,3,4,†], Song Jiang [1], Qibin Yang [1,3], Jianhua Huang [1,4], Lishi Yang [1], Jianzhi Shi [1], Zhibin Lu [1,5], Yan Zhang [1,5], Shigui Jiang [1] and Falin Zhou [1,3,4,*]

1 South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Key Laboratory of South China Sea Fishery Resources Exploitation and Utilization, Ministry of Agriculture and Rural Affairs, Guangzhou 510300, China; chenj29725@163.com (J.C.); liyd2019@163.com (Y.L.); tojiangsong@163.com (S.J.); yangqibin1208@163.com (Q.Y.); huangjianhua@scsfri.ac.cn (J.H.); yangls2016@163.com (L.Y.); shijianzhi1989@163.com (J.S.); lzb9502@163.com (Z.L.); zhangyan202205@126.com (Y.Z.); jiangsg@21cn.com (S.J.)
2 College of Fisheries and Life Sciences, Dalian Ocean University, Dalian 116023, China
3 Key Laboratory of Efficient Utilization and Processing of Marine Fishery Resources of Hainan Province, Sanya Tropical Fisheries Research Institute, Sanya 572018, China
4 Shenzhen Base of South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shenzhen 518108, China
5 Guangzhou Nansha Fishery Industry Park Co., Ltd., Guangzhou 510000, China
* Correspondence: zhoufalin@aliyun.com
† These authors contributed equally to this work.

Abstract: *Penaeus monodon* is one of the world's most important aquaculture species, with its host-associated microbial community playing a crucial role in its growth, metabolism, immune response, and adaptation. In our study, we utilized Illumina high-throughput sequencing to investigate the composition, structure, and function of the intestinal microbial communities of *P. monodon* from two different regions in Guangdong. Our results identified 176 phyla across both populations, with *Proteobacteria* and *Firmicutes* being predominant. Furthermore, we identified 3095 genera, with *Photobacterium*, *Vibrio*, and *Aliiroseovarius* being the most dominant. Functional gene analysis based on KEGG data indicated that the carbohydrate metabolism and amino acid metabolism were significant at the secondary metabolic pathway level. The eggNOG functional annotation revealed that the genes involved in replication, recombination, and repair are of paramount importance. The CAZy annotation results indicated that Glycoside Hydrolases (GH) have the highest abundance. The Pfam annotation analysis showed that the two most prevalent domains are P-loop NTPase and NADP Rossmann. Our investigation provides a reference for species-level and functional-level analyses of the intestinal microbiota of *P. monodon*, contributing valuable insights into its microbial ecology.

Keywords: metagenome; intestinal flora composition; functional genes

check for updates

Citation: Chen, J.; Li, Y.; Jiang, S.; Yang, Q.; Huang, J.; Yang, L.; Shi, J.; Lu, Z.; Zhang, Y.; Jiang, S.; et al. Metagenomic Profiling of Bacterial Communities and Functional Genes in *Penaeus monodon*. *J. Mar. Sci. Eng.* **2024**, *12*, 1481. https://doi.org/10.3390/jmse12091481

Academic Editor: Concetta Gugliandolo

Received: 14 July 2024
Revised: 11 August 2024
Accepted: 15 August 2024
Published: 26 August 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Penaeus monodon is a member of the phylum *Crustacea*, order *Decapoda*, suborder *Natantia*, family *Penaeidae*, and genus *Penaeus*. It is commonly known by various vernacular names, such as grass shrimp, black tiger shrimp, and bamboo shrimp. Designated as the giant tiger prawn by the Food and Agriculture Organization (FAO) of the United Nations, this species is widely distributed along the coastlines of the western Pacific Ocean and the Indian Ocean. It ranks globally among the top three cultivated shrimp species owing to its large size, rapid growth, high yield per unit area, and delectable taste [1].

During the growth process of animals, intestinal microbiota play a crucial role. Long-term coevolution has led to a symbiotic relationship between intestinal microbiota and hosts, where the intestinal microbiota are often referred to as the host's "external organ",

contributing significantly to the host's nutritional metabolism, immune regulation, and disease prevention. The gastrointestinal tract of multicellular organisms is typically complex, possessing various physiological functions that aid in maintaining a stable internal balance and regulating normal growth and development [2]. Intestinal microbiota are closely associated with the host's health and play pivotal roles in the metabolism, immunity, and physiological functions of the human body, thus garnering considerable attention [3–5]. Research indicates that intestinal microbiota form complex and close networks of relationships with hosts through cooperation or competition during different developmental stages, under varying nutritional levels, and in different environmental conditions. They participate in the host's nutrient acquisition and immune regulation, maintaining normal intestinal function and the ecological balance within the host's body to cope with changes in endogenous and exogenous factors [6–8].

Fanying et al. [9] utilized high-throughput sequencing technology (HiSeq) to analyze intestinal microbiota samples from different growth stages of *Hexagrammos otakii*. Their findings revealed a gradual adaptation of the intestinal microbiota's functionality to the organism and environmental demands during growth, accompanied by significant variations in the microbial community's structure across different growth stages. This highlights the influence of growth and development on the functional dynamics of intestinal microbiota in aquatic organisms. Francesco Cicala et al. [10] investigated the structural and functional transitions of microbial communities during the developmental stages of *P. monodon*. Their results demonstrated that each developmental stage of *P. monodon* may create specific conditions conducive to the proliferation of certain bacterial taxa. Haiqing Wang [11] conducted a study on the diversity of gut bacteria communities in eight economically important marine species. The study's findings revealed that the gut microbiota of *Litopenaeus vannamei* and *Lateolabrax japonicus* exhibited the highest richness. Specifically, the gut microbiota of *L. vannamei* consisted mainly of genera, such as *Vibrio*, *Pseudomonas*, *Photobacterium*, and *Alteromonas*, while those of *L. japonicus* were dominated by *Vibrio* and *Photobacterium* genera. Qiangzhuang [12] detected and compared the intestinal contents of three rare native fish species from Xinjiang, including *Diptychus maculatus*, *Schizothorax biddulphi*, and *Triplophysa yarkandensis*. They identified significant similarities and differences in the composition and potential functions of the intestinal microbiota among these species, suggesting a co-evolutionary relationship between the intestinal microbiota and their hosts. This finding provides valuable data and theoretical support for the conservation and artificial breeding of these native fish species in the wild.

While previous studies have provided valuable information for understanding the gut microbiota of *P. monodon*, metagenomic research within the species *P. monodon* is quite limited. Considering that the South China Sea region is an important aquaculture area for *P. monodon*, we believe that metagenomic research on *P. monodon* is necessary. Our research focuses on this field in the hope of providing some reference data for subsequent related studies. In this study, species and gene annotation analyses were conducted using macro-genomic techniques to annotate the species and genes of the intestinal microbiota of *P. monodon* from two different regions. This research serves as a platform for subsequent functional gene discovery by investigating the differences in the composition and functionality of the intestinal microbiota of *P. monodon* under different geographical and environmental conditions; understanding the impact of intestinal microbiota on *P. monodon;* exploring the genetic information of the intestinal microbiota of *P. monodon;* and revealing the key mechanisms of diversity, functionality, and interaction with the host. This study has enhanced our understanding of the microbial community in the *P. monodon* ecosystem and also provides a reference for developing more effective aquaculture management strategies and health maintenance methods.

2. Materials and Methods

2.1. Sample Collection

In this study, a total of six samples were collected from two locations in Guangdong, China—Shenzhen and Zhuhai—with three samples from each location. The samples from Shenzhen (SZ1-3) and Zhuhai (ZH1-3) were all obtained from high-level pond aquaculture locations. Both locations used the same commercial feed. During sampling, 21 randomly selected *P. monodon* specimens of similar size and matched weight were divided into three groups with seven per group in order to ensure the diversity and representativeness of the samples. Their intestinal contents were then evenly distributed into three cryovials on ice and promptly stored at −80 °C for preservation.

2.2. DNA Extraction

DNA was extracted from the samples using the E.Z.N.A.® Soil DNA Kit (OMEGA Bio-tek, Norcross, GA, USA) following the manufacturer's instructions, and the obtained DNA was stored at −80 °C. NanoDrop2000 (Thermo Fisher Scientific, Waltham, MA, USA) was used to detect DNA purity and concentration, and 1% agarose gel electrophoresis was performed to detect DNA integrity. The voltage was 5 V/cm, and the time was 20 min.

2.3. Metagenomic Sequencing

The extracted genomic DNA was submitted to a testing laboratory (Biomarker Technologies, Rohnert Park, CA, USA) for metagenomic sequencing. DNA fragments with an average size of approximately 400 bp were obtained using sonication. Paired-end sequencing was performed on an Illumina NovaSeq platform (Illumina, San Diego, CA, USA).

2.4. Data Processing

We employed the fastp software (0.23.2) for preprocessing our sequencing data, filtering the raw tags, and ensuring a high-quality standard (≥Q20) for the clean tags obtained. For the alignment process, we utilized bowtie2 (2.5.3) to map the sequences against the host genome, thereby effectively removing any host contamination. For the metagenome assembly, we employed the MEGAHIT [13] software (1.2.9), applying a filter to remove any contig sequences shorter than 300 base pairs to ensure the robustness of our assembly. The assembly outcomes were then assessed using the QUAST [14] software (5.2.0), which provided a comprehensive evaluation of the contiguity and quality of the assembled genome. Coding regions within the assembled genome were identified using the MetaGeneMark [15] software (http://exon.gatech.edu/meta_gmhmmp.cgi, Version 3.26, accessed on 5 June 2023). To refine our assembly by eliminating redundant sequences, we applied the MMseqs2 [16] software (https://github.com/soedinglab/mmseqs2, Version 12-113e3, accessed on 5 June 2023), setting a similarity threshold of 95% and a coverage threshold of 90% for this process. By comparing the protein sequences of non-redundant genes with those in the Nr database and the corresponding sequence annotation, we generated annotation information for the genes of the sequenced genome.

2.5. Species and Functional Annotations

The comparison of non-redundant gene protein sequences with the Nr database and existing sequence annotation information facilitated the development of annotation data for genes within the corresponding sequenced genome. During alignment, BLAST was employed to match non-redundant gene sequences with protein sequences recorded in the Kyoto Encyclopedia of Genes and Genomes (KEGG), eggNOG, CAZy, and Pfam databases. This allowed for the identification of corresponding functions and the calculation of functional category abundances.

3. Results

3.1. Statistics

The analysis of the raw sequencing data from each sample revealed very few ambiguous bases, with approximately 40% GC base content. The accuracy of the base calling exceeded 99.9% for over 90.0% of the bases and reached 99.0% for over 95.0% of the bases (Table 1). Therefore, the average quality of the raw sequencing data in this study met expectations and allowed for subsequent analysis. Afterward, analyses were performed using the KEGG, eggNOG, CAZy, and Pfam databases.

Table 1. Statistical evaluation of the sequenced sample data.

Samples	Clean Data Base (bp)	Number of Reads	GC (%)	≥Q20 (%)
SZ1	12,620,082,756	19,019,083	39.57	98.48
SZ2	20,778,893,486	63,925,983	44.63	98.97
SZ3	18,358,186,992	60,297,886	43.75	98.94
ZH1	19,302,573,370	61,799,557	36.03	99.26
ZH2	14,778,510,542	33,389,695	37.43	98.86
ZH3	19,819,879,072	63,647,616	39.12	99.0

Note: Samples: Samples; Raw data base (bp): Raw data amount: Clean data base (bp): Data amount after quality control; Number of Reads: Number of reads of the final effective data; Q20 (%): The percentage of the number of bases with a quality value greater than or equal to 20 in the total number of bases; GC (%): Sample GC content, the percentage of G or C type bases in all bases.

3.2. Species-Level Analysis

3.2.1. Community Composition Analysis

The structural composition and abundance distributions of the intestinal microbiota at the phylum and genus levels in two regions of *P. monodon* were obtained, as shown in Table 2 and Figure 1. At the phylum level, a total of 176 phyla were identified (Table 2), with *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Planctomycetes*, *Actinobacteria*, *Cyanobacteria*, *Verrucomicrobia*, *Candidatus-Gracilibacteria*, *Tenericutes*, and *Fusobacteria* being the most abundant. Among these, *Proteobacteria* dominated in both populations (averaging 43.58% in SZ and 29.36% in ZH), while *Firmicutes* ranked second in both SZ and ZH (5.39% and 3.05%, respectively). The abundance of all phyla was lower in ZH compared to SZ, with a majority of species in ZH being unassigned and unclassified.

At the genus level, a total of 3095 genera were identified (Table 2), with *Photobacterium*, *Vibrio*, *Aliiroseovarius*, *Clostridium*, *Shewanella*, *Klebsiella*, *Bacillus*, *Fusibacter*, *Aliivibrio*, and *Gimesia* being the most abundant. Except for *Photobacterium* and *Fusibacter*, the abundance of the remaining genera was lower in ZH compared to SZ. In SZ, the dominant genera were *Photobacterium* (15.99%), *Vibrio* (15.48%), *Aliiroseovarius* (1.73%), *Clostridium* (0.67%), *Shewanella* (0.63%), and *Bacillus* (0.57%). In ZH, the dominant genera were *Photobacterium* (16.41%), *Vibrio* (11.47%), *Klebsiella* (0.26%), *Shewanella* (0.18%), *Aliivibrio* (0.11%), and *Bacillus* (0.09%). *Photobacterium* and *Vibrio* were the two most abundant genera in both *P. monodon* populations. In PmZH, some genera were unassigned and unclassified.

Table 2. Statistical table of annotated species of each species classification level of sample.

Samples	Kingdom	Phylum	Class	Order	Family	Genus	Species
SZ1	6	83	103	216	460	1337	4186
SZ2	7	170	151	320	731	3003	19,798
SZ3	6	161	142	304	690	2674	13,848
ZH1	5	78	91	180	346	889	2823
ZH2	5	66	96	190	378	817	2010
ZH3	6	112	124	265	561	1743	6270
Total	7	176	158	331	749	3095	21,597

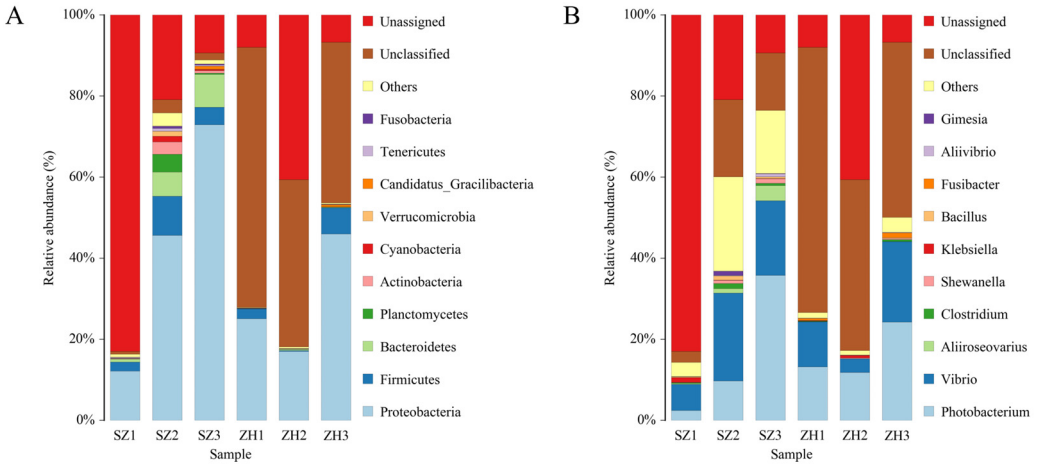


Figure 1. Distribution map of species composition. (**A**) Group-phylum distribution map of species composition; (**B**) Group-genus distribution map of species composition; the X-coordinate shows sample name; the Y-coordinate shows the percentage of relative abundance. Each color represents one species; a colored block's length represents the relative abundance proportion of the species. For the best view, the histogram only shows the species with the ten highest abundance levels, while other species are merged as "Others" in the graph; "Unassigned" represents species without a taxonomic annotation. "Unassigned" refers to sequences that have been identified but are not yet categorized under a specific taxonomic group; "Unclassified" indicates sequences that lack sufficient information for placement into any known taxonomic category.

3.2.2. Beta Diversity

A beta diversity analysis of the intestinal microbiota community was conducted (Figure 2). A principal coordinates analysis (PCoA) was performed to decompose the sample distance matrix and display the natural distribution of samples at a specific distance scale. The samples in each group were close to each other, indicating a small difference in species composition among the samples within the group. Specifically, samples from ZH were more concentrated, while those from SZ showed a slight dispersion. The contribution rates of axes 1 and 2 were 47.87% and 38.85%, respectively (Figure 2A).

To validate whether the observed differences on the PCoA plot were statistically significant, a PERMANOVA analysis was conducted on the distance matrix. The results showed an R2 value of 0.478 and a *p*-value of 0.001 (Figure 2B), indicating a significant difference between ZH and SZ.

3.3. Functional Level Analysis

3.3.1. Genes Prediction

The MetaGeneMark [15] software (http://exon.gatech.edu/meta_gmhmmp.cgi, Version 3.26, accessed on 28 July 2023) with default parameters was used to identify coding regions in the genomes. The predictions are presented in Table 3. The number of predicted genes in the SZ samples is greater than that of the ZH samples, and the total number of bases of all predicted genes is also higher in the SZ samples than in the ZH samples.

Table 3. Statistics of the gene prediction results.

Sample ID	Gene Number	Total Length (bp)	Average (bp)	Max Length (bp)	Min Length (bp)
SZ1	347,214	99,836,934	287.0	13,491	102
SZ2	832,574	508,767,852	611.0	24,186	102
SZ3	313,039	194,545,122	621.0	13,491	102
ZH1	37,635	25,640,667	681.0	21,687	102
ZH2	235,625	63,989,850	271.0	11,010	102
ZH3	102,771	65,502,831	637.0	14,070	102

Note: Samples: Sample ID; Genes Numbers: Number of predicted genes; Total Length: Sum of base numbers of all predicted genes; Average Length: The average base number of predicted genes.

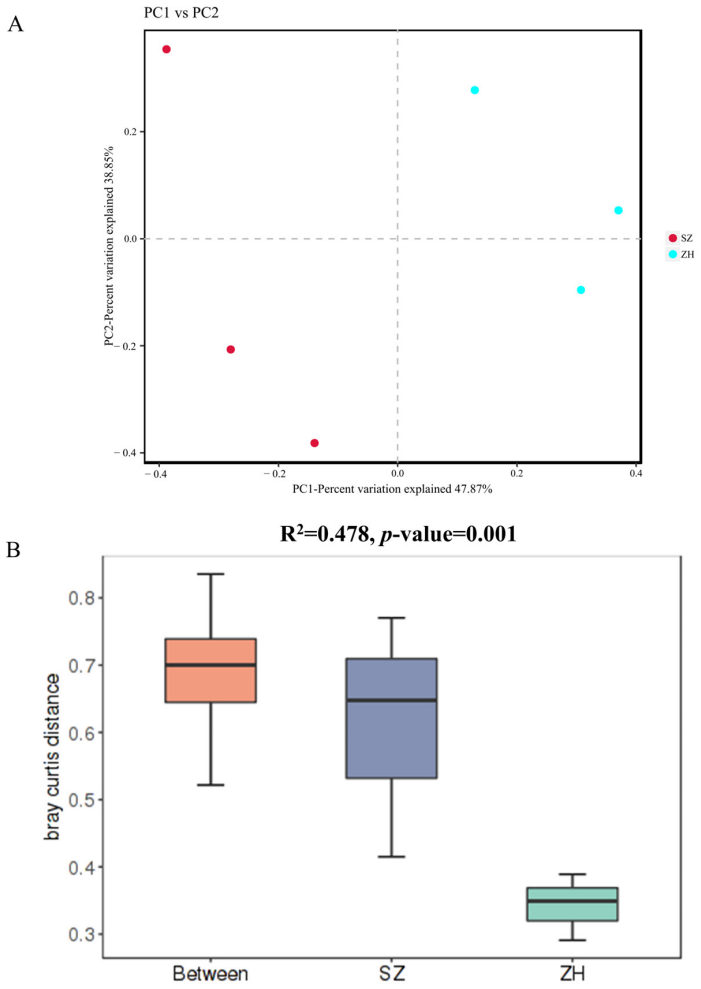


Figure 2. Beta diversity of different intestinal samples in *P. monodon*. (**A**) PCoA. (**B**) PERMANOVA.

3.3.2. Construction of Non-Redundant Genes Set

The MMseqs2 software (https://github.com/soedinglab/mmseqs2, Version 12-113e3, accessed on 28 July 2023) was used to remove redundancy utilizing a similarity threshold set at 95% and a coverage threshold set at 90%. The length distribution histogram of genes in the non-redundant gene set is shown in Figure 3. As the length intervals of the non-redundant gene set increase, the number of sequences within those intervals decreases. The interval with the most sequences is 200–400 bp.

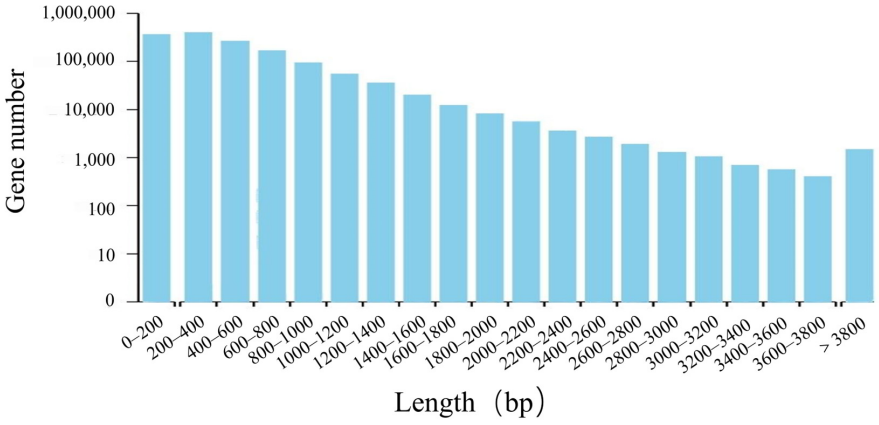


Figure 3. Length distribution of non-redundant genes. The X-coordinate is the length range of non-redundant genes set; the Y-coordinate is the number of sequences in the range.

3.3.3. Kegg Function Annotation

Based on KEGG data, functional genes annotated to metabolic pathways at the secondary level were statistically analyzed, as shown in Figure 4. The results indicate that global and overview maps are the most commonly used maps (Figure 4). In terms of metabolism, carbohydrate metabolism, amino acid metabolism, and the metabolism of cofactors and vitamins all play crucial roles. In genetic information processing, translation, replication, repairing, folding, sorting, and degradation are essential processes. In environmental information processing, membrane transport plays a significant role. In cellular processes, cellular community–prokaryotes are important, while cell growth and death have a minimal impact (Figure 4).

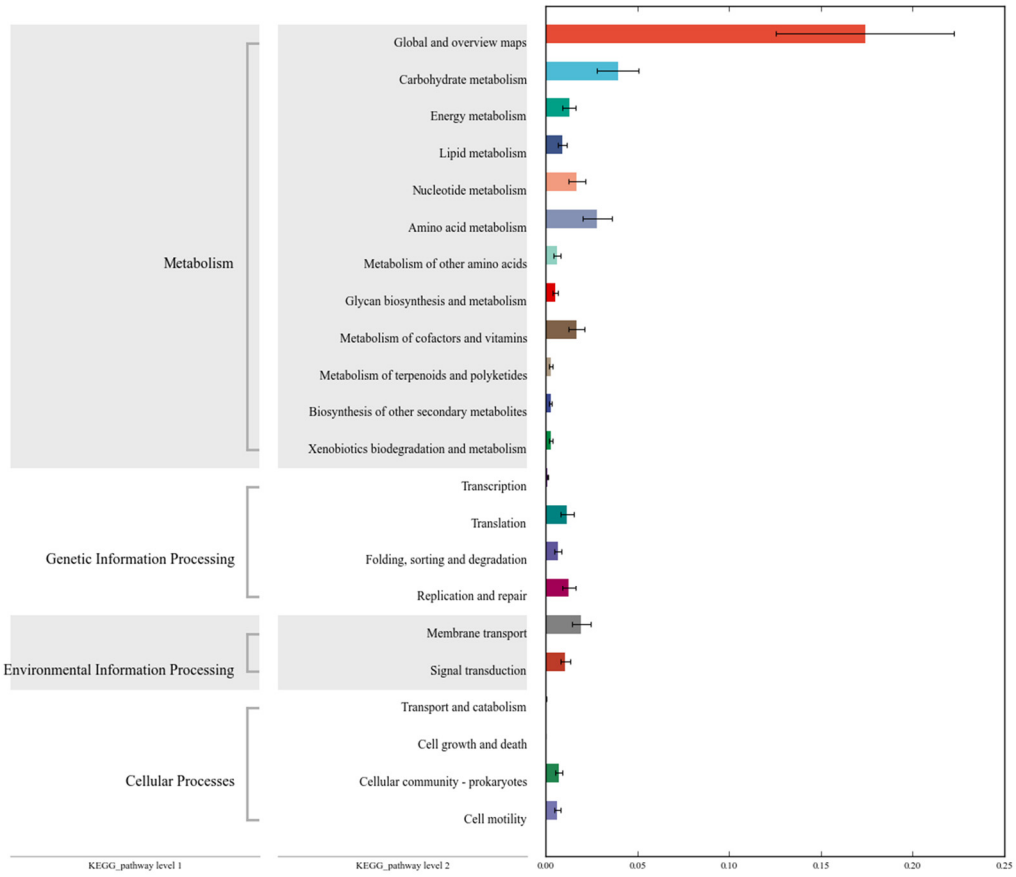


Figure 4. Statistical figure of KEGG metabolic pathways-related functional genes at the secondary level. The X-coordinate shows the relative content of number of corresponding functional genes; the Y-coordinate shows the KEGG secondary functional classification.

3.3.4. eggNOG Database Annotation

By performing BLAST comparisons between non-redundant protein sequences and the eggNOG database, the most similar sequences in the eggNOG database were identified (Figure 5). The annotation and classification information corresponding to these sequences was assigned to the respective sequenced genome genes, as shown in Figure 5. The results indicate that, apart from some annotations, such as "Function unknown" and "General function prediction only", replication, recombination, and repair play a crucial role in eggNOG. Amino acid transport and metabolism also play significant roles. However, their involvement of cell cycle control, cell division, chromosome partitioning is minimal.

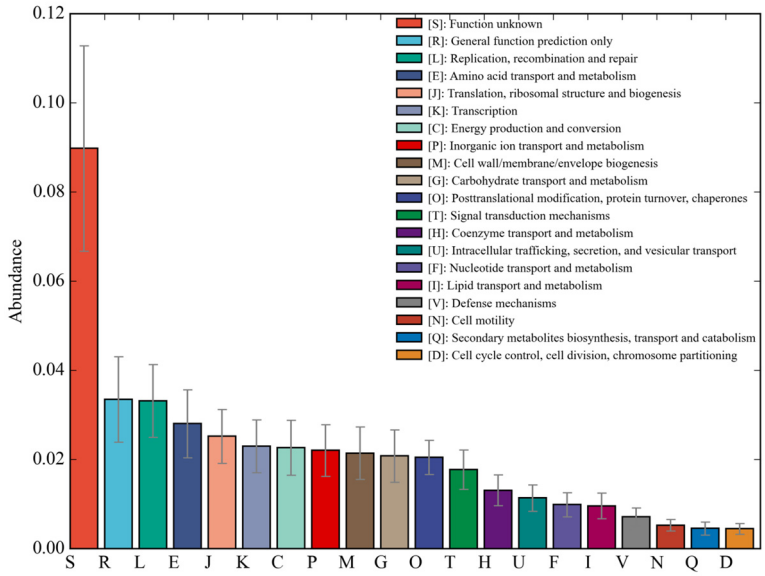


Figure 5. Figure of statistics of the function classifications of eggNOG functional genes. The X-coordinate shows the content of each classification of eggNOG; the Y-coordinate shows the relative content of the number of corresponding functional genes.

3.3.5. CAZy Carbohydrate-Active Enzymes Annotation

A systematic identification and quantitative analysis of the CAZyme families revealed a relative abundance of different CAZy categories in the samples (Figure 6). Among the six major functional categories of proteins in this database, GH (Glycoside Hydrolases) is the most dominant at 26.9%, followed by CBM (Carbohydrate-Binding Modules) at 23.9%, CE (Carbohydrate Esterases) at 22.8%, and GT (Glycosyl Transferases) at 21.5%. The categories with the lowest abundance are AA (Auxiliary Activities) and PL (Polysaccharide Lyases) at 3.1% and 1.8%, respectively, which are significantly lower than the other four functional protein categories.

3.3.6. Pfam Database Annotation

During annotation, the Linclust algorithm was used to identify the top 10 most similar sequences in the Pfam database by aligning the protein sequences of non-redundant genes with the Pfam database, applying an e-value cutoff of 10^{-5}. (Table 4). The annotation and classification information of the sequences was derived from this alignment. The most similar sequence was P-loop NTPase, primarily functioning as a P-loop containing the nucleoside triphosphate hydrolase superfamily. The next most similar sequences were NADP Rossmann, HTH, TIM barrel, PBP, RNase H, HUP, PLP aminotransferase, Nucleot cyclase, and Thioredoxin.

Table 4. Pfam database clan abundance.

Clan Accession	Clan ID	Clan Description
CL0023.34	P-loop NTPase	P-loop containing nucleoside triphosphate hydrolase superfamily
CL0063.25	NADP Rossmann	FAD/NAD(P)-binding Rossmann fold superfamily
CL0123.18	HTH	Helix-turn-helix clan
CL0036.24	TIM barrel	Common phosphate binding-site TIM barrel superfamily
CL0177.16	PBP	Periplasmic binding protein clan
CL0219.14	RNase H	Ribonuclease H-like superfamily
CL0039.12	HUP	HUP–HIGH-signature proteins, UspA, and PP-ATPase.
CL0061.13	PLP aminotran	PLP dependent aminotransferase superfamily
CL0276.8	Nucleot cyclase	Nucleotide cyclase superfamily

Note: Clan accession: the number of each Pfam protein clan annotated by the gene; Clan ID: ID of each Pfam protein clan annotated by the gene; Clan description: the functional description of each Pfam protein clan annotated by the gene.

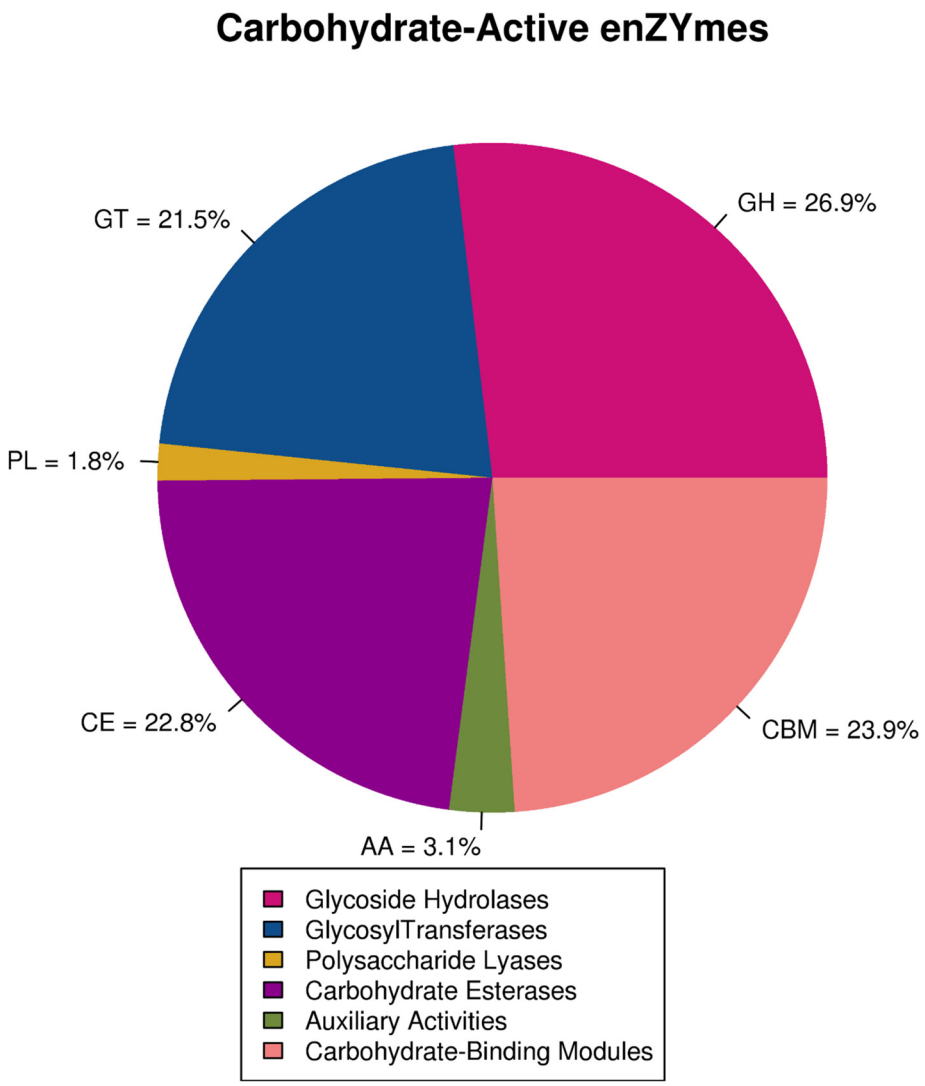


Figure 6. Carbohydrate enzyme distribution scale diagram. Different colors are used to represent each type of carbohydrate-active enzyme, and the size of each sector area is used to represent the proportion of the relative content of the functional genes in a given type of active enzyme.

4. Discussion

4.1. Species-Level Analysis

The intestine is the primary site for digestion and absorption in shrimp, while also serving three major functions: acting as a mechanical barrier, an immune barrier, and a biological barrier [17,18]. Initially, intestinal microbiota research relied on traditional culture techniques. However, the limitations of these techniques, such as restrictive culture conditions and the ability to only provide general phenotypic descriptions, resulted in a significant proportion of uncultivable micro-organisms. Consequently, these methods failed to fully reflect the microbial community structures in analyzed samples [19]. In recent years, the rapid development of high-throughput sequencing technology, coupled with big data software, has enabled accurate representation of overall microbial community structures. Even microbial populations present in low abundance can be detected, making this approach widely applicable in current research [20–22]. In this study, Illumina high-throughput sequencing technology was employed to analyze the intestinal bacterial composition of two *P. monodon* populations. The results revealed the identification of 176 phyla at the phylum level. *Proteobacteria*, followed by *Firmicutes*, was the most prevalent phylum at the phylum level in both regions, representing the highest abundance in the two populations. At the genus level, 3095 genera were identified, with *Photobacterium*,

Vibrio, and *Aliiroseovarius* being the most abundant. *Photobacterium* exhibited a higher abundance in PmZH compared to PmSZ, while *Vibrio* and *Aliiroseovarius* showed a higher abundance in PmSZ than in PmZH. Significant differences in the abundance of intestinal microbiota compositions were observed between the two populations.

Previous studies have indicated that *Proteobacteria* and *Firmicutes* play dominant roles in the intestinal microbiota of shrimp. The intestinal bacterial community of *L. vannamei* was consistently dominated by three bacterial phyla—namely, *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria*—across all stages [23]. In intensively managed commercial ponds in Ecuador, *Proteobacteria* were found to dominate the intestinal microbiota of *L. vannamei* during both the nursery and harvest stages [24]. Additionally, Wang Xiaolu et al. [25] explored the characteristics of the intestinal microbiota of South African *P. monodon* in factory farming models using HE staining of tissue sections, high-throughput sequencing, and Biolog ECO technology. Their results revealed that the microbial community structure at the phylum level primarily was comprised of *Proteobacteria*, *Planctomycetes*, *Actinobacteria*, and *Verrucomicrobia*. Lanfen Fan [26] investigated the role of intestinal microbiota in *L. vannamei* in relation to varying growth performances. The sequencing data indicated statistically significant diversity at the phylum and genus levels. The most abundant phyla observed were *Proteobacteria*, *Cyanobacteria*, *Tenericutes*, *Fusobacteria*, *Firmicutes*, *Verrucomicrobia*, *Bacteroidetes*, *Planctomycetes*, *Actinobacteria*, and *Chloroflexi*, which is consistent with the findings of this study. Notably, at the genus level, the diversity of intestinal bacteria in ZH was lower than that in SZ, which can be explained by changes in the relative abundance of certain bacteria.

Photobacterium is a highly abundant genus in both populations. Previous studies have shown that *Photobacterium* in the intestines of shrimp is sensitive to environmental stressors, such as salinity [27], pH [28], and temperature [29]. Significant changes in these factors may lead to pathogenicity through the expression of virulence factors by *Photobacterium*. Research by Aseer Manilal et al. [30] suggests that the overaccumulation of *Photobacterium* may alter the health status of shrimp. Therefore, careful monitoring of the abundance of *Photobacterium* should be conducted during the cultivation of shrimp in order to avoid potential risks.

4.2. Analysis of Gut Microbiota Composition Based on Functional Abundance

4.2.1. KEGG Database Annotation Analysis

The KEGG functional annotation results show that at the Level 2 hierarchy, the metabolic pathways with the most annotations are carbohydrate metabolism and amino acid metabolism. In previous studies, the supplementation of carbohydrates reduced nitrogenous metabolites and significantly enhanced the growth performance of juvenile and sub-adult *P. monodon* [31]. During nitrogen metabolism, ammonia is produced [32], and when ammonia levels exceed safe limits in the water, it can enter a shrimp's bloodstream through its respiratory organs (gills) [33]. After entering the bloodstream, ammonia binds with hemoglobin, hindering the normal transport of oxygen and causing hypoxia symptoms in shrimp. In severe cases, this can even lead to death [34]. Prolonged exposure to high ammonia concentrations can also damage the liver and pancreas functions of *P. monodon*, affecting their immune system, reducing their disease resistance, and making them more susceptible to pathogen infections. Therefore, carbohydrate metabolism plays a crucial role in the immune capacity of *P. monodon*. Additionally, research by Pradyut Biswas et al. [35] demonstrated that the supplementation of both lysine and phytase in a soybean-based diet not only reduces the nitrogen and phosphorus load of the culture system, but also significantly changes the fatty acid profile of *P. monodon* tissues. This indicates that the intake of carbohydrates and amino acids can promote the growth and development of *P. monodon*. Microbes obtain energy from carbohydrates and amino acids, and the high abundance of carbohydrate metabolism and amino acid metabolism in the gut of *P. monodon* is essential for maintaining gut health and homeostasis, as well as promoting shrimp growth and development.

4.2.2. eggNOG Database Annotation Analysis

Through eggNOG functional annotation, it was found that replication, recombination, and repair play crucial roles. Previous studies have shown that the processes of DNA replication, repair, and recombination are increasingly recognized as integrated events in cellular life [36]. The recombination and replication machinery cooperate to maintain genomic integrity [37]. This indicates that DNA replication, repair, and recombination are key mechanisms for maintaining genomic stability in the intestines of *P. monodon*. They repair errors, damages, and mutations in DNA in a timely manner, adapt to environmental changes, survive in changing environments, and resist external pressures, thereby exerting a positive influence on the health of *P. monodon*.

4.2.3. CAZy Database Annotation Analysis

In this study, we conducted a comprehensive analysis and functional annotation of carbohydrate-active enzymes in the samples using the CAZy database, covering six major functional protein categories. Among these, Glycoside Hydrolases (GH) had the highest abundance, followed by Carbohydrate-Binding Modules (CBM). GHs are a widely distributed class of enzymes capable of hydrolyzing glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety [38]. In the gut of *P. monodon*, these enzymes help break down ingested carbohydrates into smaller molecules, thereby promoting nutrient absorption and utilization. CBMs are non-catalytic domains that can fold into specific three-dimensional structures, enabling them to bind carbohydrates [39]. This helps precisely direct GHs to their substrates, enhancing the enzymes' catalytic efficiency and specificity. Through this synergistic interaction, GHs and CBMs combined maintain the stability of gut microbiota, promote effective nutrient uptake by the host, and potentially have a positive impact on gut health.

4.2.4. Pfam Database Annotation Analysis

In the Pfam annotation analysis of the non-redundant genes in the samples, the top two hits were P-loop NTPase and NADP Rossmann. In the past, extensive research has been conducted on many P-loop NTPases, revealing their key roles in a variety of complex biological processes, such as programmed cell death, disease, and stress responses in plants and animals [40]. Some scholars have identified a gene in plants that encodes an NADP-binding Rossmann fold protein in genes with reduced peroxidase activity. This indicates that NADP Rossmann fold proteins are usually related to redox reactions, which play key roles in the immune system, such as producing reactive oxygen species (ROS) to combat pathogens [41,42]. Through the key roles of P-loop NTPase and NADP Rossmann in intracellular signal transduction processes, they regulate the stress response and redox reactions in aquatic animals, helping their bodies resist the invasion of pathogens.

4.2.5. Scientific Significance and Practical Value

Our study on *P. monodon*, a very economically valuable aquaculture species, aims to address the significant underrepresentation of metagenomic research in this area. Because of the infancy of metagenomic studies specific to *P. monodon*, possibly due to its genomic complexity and a lack of research focus, our work is essential in filling these gaps and providing a theoretical and data-driven foundation for future studies. By selecting pivotal cultivation sites in Shenzhen and Zhuhai, China, we are offering novel insights into the biological characteristics of *P. monodon*, with our research holding substantial regional representativeness and practical application value, especially in Guangdong's key farming regions. Moreover, our focus on gut metagenomics is crucial for understanding the intricate relationship between gut microbes and host health, which is vital for optimizing farming efficiency and disease management. We anticipate that our comprehensive approach will not only enhance our knowledge of *P. monodon*'s biology, but also contribute significantly to the advancement of aquaculture practices and research methodologies.

5. Conclusions

This study conducted metagenomic analyses of the gut microbiota of *P. monodon* from two aquaculture sites, yielding insights at both the species and functional levels. At the phylum level, a total of 176 phyla were identified, with Proteobacteria and Firmicutes being the most prevalent in both locations. At the genus level, 3095 genera were identified, with Photobacterium, Vibrio, and Aliiroseovarius being the most abundant. Functional annotations were performed using the KEGG, eggNOG, CAZy, and Pfam databases. A KEGG functional annotation at Level 2 revealed that the most annotated metabolic pathways were carbohydrate metabolism and amino acid metabolism. The eggNOG functional annotation indicated that replication, recombination, and repair were the most prominent functions. The CAZy annotation results showed that Glycoside Hydrolases (GH) had the highest abundance, followed by Carbohydrate-Binding Modules (CBM). The Pfam annotation analysis identified P-loop NTPase and NADP Rossmann as the two most prevalent functions. The results suggest that the gut microbiota of *P. monodon* play crucial roles in immune response, genomic stability, and carbohydrate absorption. Despite our study being only an initial foray into the exploration of the microbial community of the region's *P. monodon*, it offers a reference for future experiments and research.

Author Contributions: S.J. (Shigui Jiang) and F.Z. conceived the project and supervised the work. J.C. performed the bioinformatics analysis and prepared the manuscript, the table and figures. J.C., Y.L., Z.L. and Y.Z. conducted the experiment. J.H., L.Y., S.J. (Song Jiang), Q.Y. and J.S. collected the samples and performed sequencing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key R & D Program of China (2022YFD2401900, 2022YFD2400104); Central Public-interest Scientific Institution Basal Research Fund, CAFS (2023TD34); China Agriculture Research System (CARS-48); Guangdong Basic and Applied Basic Research Foundation (2023A1515012410); Hainan Provincial Naturl Science Foundation of China (323MS127); Earmarked fund for HNARS (HNARS-10-ZJ01); Project funded by China Postdoctoral Science Foundation (2023T160178); Hainan Province postdoctoral research funding project (2022-BH-13, 2022-BHMS-03).

Institutional Review Board Statement: The animal study was reviewed and approved by the Animal Care and Use Committee of the South China Sea Fisheries Research Institute.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw FASTQ files for this study can be found in the NCBI Sequence Read Archive (SRA) [https://www.ncbi.nlm.nih.gov/sra/PRJNA1147588] with BioProject accession number PRJNA1147588.

Conflicts of Interest: Author Zhibin Lu and author Yan Zhang were employed by the company Guangzhou Nansha Fishery Industry Park Co., Ltd. And the remaining 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.

References

1. Yundong, L. Molecular Cloning of Growth Trait-Related Genes in Penaeus Monodon and Screening and Verification of Their SNPs Makers. Master's Thesis, Shanghai Ocean University, Shanghai, China, 2017.
2. Ahearn, G.A. Nutrient Transport by the Invertebrate Gut. In *Advances in Comparative and Environmental Physiology*; Springer: Berlin/Heidelberg, Germany, 1988; Volume 2, pp. 91–129, ISBN 978-3-642-73377-2.
3. Turnbaugh, P.J.; Ley, R.E.; Mahowald, M.A.; Magrini, V.; Mardis, E.R.; Gordon, J.I. An Obesity-Associated Gut Microbiome with Increased Capacity for Energy Harvest. *Nature* **2006**, *444*, 1027–1031. [CrossRef] [PubMed]
4. Kamada, N.; Kim, Y.-G.; Sham, H.P.; Vallance, B.A.; Puente, J.L.; Martens, E.C.; Núñez, G. Regulated Virulence Controls the Ability of a Pathogen to Compete with the Gut Microbiota. *Science* **2012**, *336*, 1325–1329.
5. Osbelt, L.; Thiemann, S.; Smit, N.; Lesker, T.R.; Schröter, M.; Gálvez, E.J.C.; Schmidt-Hohagen, K.; Pils, M.C.; Mühlen, S.; Dersch, P.; et al. Variations in Microbiota Composition of Laboratory Mice Influence Citrobacter Rodentium Infection via Variable Short-Chain Fatty Acid Production. *PLoS Pathog.* **2020**, *16*, e1008448.
6. Pan, L.; Fang, H.; Zhang, S.; Wang, X.; Jia, Y.; Hu, F.; Gao, F.; Guo, W. Changes in Digestive Enzyme Activities during the Larval, Juvenile, and Early Juvenile Stages of Large Yellow Croaker (*Larimichthys crocea*). *Adv. Fish. Sci.* **2013**, *34*, 54–60.

7. Sun, X.; Gao, Y.; Yang, Y. Recent Advances in Metagenomics Research of Environmental Microbes. *Biodivers. Sci.* **2013**, *21*, 393–401.
8. Shi, X.; Zhang, Z.; Wang, Y.; Yu, Y.; Deng, W.; Li, H. Analysis of Bacterial Community Characteristics of Culturable Bacteria in the Digestive Tract of Larval and Juvenile Turbot (*Scophthalmus maximus*). *Adv. Fish. Sci.* **2015**, *36*, 73–82.
9. Fan, Y.; Wang, Y.; Ji, G.; Xu, L.; Wang, S.; Hu, F.; Diao, J.; Jiang, J.; Ye, H. Revealing the Intestinal Microbial Characteristics of Large Yellow Croaker (*Larimichthys crocea*) Based on Metagenomic Sequencing Technology. *Acta Microbiol. Sin.* **2023**, *63*, 357–375.
10. Cicala, F.; Lago-Lestón, A.; Gomez-Gil, B.; Gollas-Galván, T.; Chong-Robles, J.; Cortés-Jacinto, E.; Martínez-Porchas, M. Gut Microbiota Shifts in the Giant Tiger Shrimp, *Penaeus Monodon*, during the Postlarvae, Juvenile, and Adult Stages. *Aquacult. Int.* **2020**, *28*, 1421–1433. [CrossRef]
11. Wang, H. Population Molecular Diversity Analysis of Intestinal Bacterial Communities in Eight Marine Economic Animals. Master's Thesis, Graduate University of Chinese Academy of Sciences (Institute of Oceanology), Qingdao, China, 2010.
12. Qiang, Z. Metagenomic Analysis of Intestinal Microbiota in Three Indigenous Fish Species from the Kizil River in Xinjiang. Master's Thesis, Tarim University, Xinjiang, China, 2024.
13. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef] [PubMed]
14. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075.
15. Zhu, W.; Lomsadze, A.; Borodovsky, M. Ab Initio Gene Identification in Metagenomic Sequences. *Nucleic Acids Res.* **2010**, *38*, 132. [CrossRef]
16. Steinegger, M.; Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [CrossRef]
17. Vaughan, E.E.; Schut, F.; Heilig, H.; Zoetendal, E.G.; de Vos, W.M.; Akkermans, A.D. A molecular view of the intestinal ecosystem. *Curr. Issues Intest. Microbiol.* **2000**, *1*, 1–12. [PubMed]
18. Li, X.; Zhou, L.; Yu, Y.; Ni, J.; Xu, W.; Yan, Q. Composition of Gut Microbiota in the Gibel Carp (Carassius Auratus Gibelio) Varies with Host Development. *Microb. Ecol.* **2017**, *74*, 239–249. [CrossRef] [PubMed]
19. Amann, R.I.; Ludwig, W.; Schleifer, K.H. Phylogenetic Identification and in Situ Detection of Individual Microbial Cells without Cultivation. *Microbiol. Rev.* **1995**, *59*, 143–169. [CrossRef]
20. Li, Z.; He, L.; Wu, J.; Chen, J. Analysis of Dominant Bacterial Composition in Shrimp by PCR-DGGE Gene Fingerprinting. *Microbiol. Bull.* **2005**, 82–86.
21. Wen, C.; He, Y.; Xue, M.; Liang, H.; Dong, J. High-Throughput Sequencing Analysis of Deviations in Intestinal Microbiota Structure Caused by DNA Extraction in Shrimp. *Acta Microbiol. Sin.* **2016**, *56*, 130–142.
22. Zhou, Z.; Peng, S.; Bo, X.; Li, F. Analysis Methods for Microbial Detection Data Based on High-Throughput Sequencing Technology. *Prog. Biochem. Biophys.* **2017**, *44*, 58–69.
23. Huang, Z.; Li, X.; Wang, L.; Shao, Z. Changes in the Intestinal Bacterial Community during the Growth of White Shrimp, Litopenaeus Vannamei. *Aquac. Res.* **2016**, *47*, 1737–1746. [CrossRef]
24. Gainza, O.; Ramírez, C.; Ramos, A.S.; Romero, J. Intestinal Microbiota of White Shrimp Penaeus Vannamei Under Intensive Cultivation Conditions in Ecuador. *Microb. Ecol.* **2018**, *75*, 562–568. [CrossRef]
25. Wang, X.; Li, L.; Wang, Y.; Gai, C.; Ye, H.; Liu, H.; Fan, Y.; Guo, P.; Mu, C.; Ma, X. Study on Intestinal Tissue Structure, Microbial Characteristics, and Metabolic Activity of *Litopenaeus vannamei* in South Africa. *J. Guangxi Acad. Sci.* **2019**, *35*, 325–331.
26. Fan, L.; Li, Q.X. Characteristics of Intestinal Microbiota in the Pacific White Shrimp Litopenaeus Vannamei Differing Growth Performances in the Marine Cultured Environment. *Aquaculture* **2019**, *505*, 450–461. [CrossRef]
27. Wang, F.-I.; Chen, J.-C. Effect of Salinity on the Immune Response of Tiger Shrimp Penaeus Monodon and Its Susceptibility to Photobacterium Damselae Subsp. *Damselae. Fish Shellfish Immunol.* **2006**, *20*, 671–681. [CrossRef] [PubMed]
28. Prayitno, S.B.; Latchford, J.W. Experimental Infections of Crustaceans with Luminous Bacteria Related to Photobacterium and Vibrio. Effect of Salinity and pH on Infectiosity. *Aquaculture* **1995**, *132*, 105–112. [CrossRef]
29. Wang, F.-I.; Chen, J.-C. The Immune Response of Tiger Shrimp Penaeus Monodon and Its Susceptibility to Photobacterium Damselae Subsp. Damselae under Temperature Stress. *Aquaculture* **2006**, *258*, 34–41. [CrossRef]
30. Manilal, A.; Sujith, S.; Selvin, J.; Shakir, C.; Gandhimathi, R.; Kiran, G.S. Virulence of Vibrios Isolated from Diseased Black Tiger Shrimp, Penaeus Monodon, Fabricius. *J. World Aquac. Soc.* **2010**, *41*, 332–343. [CrossRef]
31. Kumar, S.; Shyne Anand, P.S.; De, D.; Sundaray, J.K.; Ananda Raja, R.; Biswas, G.; Ponniah, A.G.; Ghoshal, T.K.; Deo, A.D.; Panigrahi, A.; et al. Effects of Carbohydrate Supplementation on Water Quality, Microbial Dynamics and Growth Performance of Giant Tiger Prawn (*Penaeus monodon*). *Aquacult. Int.* **2014**, *22*, 901–912. [CrossRef]
32. Gaudinier, A.; Rodriguez-Medina, J.; Zhang, L.; Olson, A.; Liseron-Monfils, C.; Bågman, A.-M.; Foret, J.; Abbitt, S.; Tang, M.; Li, B.; et al. Transcriptional Regulation of Nitrogen-Associated Metabolism and Growth. *Nature* **2018**, *563*, 259–264. [CrossRef]
33. Zhang, X.; Pan, L.; Wei, C.; Tong, R.; Li, Y.; Ding, M.; Wang, H. Crustacean Hyperglycemic Hormone (CHH) Regulates the Ammonia Excretion and Metabolism in White Shrimp, Litopenaeus Vannamei under Ammonia-N Stress. *Sci. Total Environ.* **2020**, *723*, 138128. [CrossRef] [PubMed]
34. Zhang, Y.J.; Wang, C.; Liu, C.Q.; Wang, X.J.; Wang, J.G. Effects of Ammonium Nitrogen and Nitrite Nitrogen on Phenoloxidase Activity and Hemocyanin Content of Japanese Swamp Shrimp. *Fish. Sci.* **2010**, *29*, 31–34.

35. Biswas, P.; Pal, A.K.; Sahu, N.P.; Reddy, A.K.; Prusty, A.K.; Misra, S. Lysine and/or Phytase Supplementation in the Diet of Penaeus Monodon (Fabricius) Juveniles: Effect on Growth, Body Composition and Lipid Profile. *Aquaculture* **2007**, *265*, 253–260. [CrossRef]
36. Stauffer, M.E.; Chazin, W.J. Structural Mechanisms of DNA Replication, Repair, and Recombination*. *J. Biol. Chem.* **2004**, *279*, 30915–30918. [CrossRef] [PubMed]
37. Marians, K.J. Replication and Recombination Intersect. *Curr. Opin. Genet. Dev.* **2000**, *10*, 151–156. [CrossRef] [PubMed]
38. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes Database (CAZy): An Expert Resource for Glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238. [CrossRef] [PubMed]
39. Shoseyov, O.; Shani, Z.; Levy, I. Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 283–295. [CrossRef]
40. Leipe, D.D.; Koonin, E.V.; Aravind, L. STAND, a Class of P-Loop NTPases Including Animal and Plant Regulators of Programmed Cell Death: Multiple, Complex Domain Architectures, Unusual Phyletic Patterns, and Evolution by Horizontal Gene Transfer. *J. Mol. Biol.* **2004**, *343*, 1–28. [CrossRef] [PubMed]
41. Marttinen, E.M.; Decker, E.L.; Heinonen, P.; Reski, R.; Valkonen, J.P.T. Putative NAD(P)-Binding Rossmann Fold Protein Is Involved in Chitosan-Induced Peroxidase Activity and Lipoxygenase Expression in Physcomitrium Patens. *Mol. Plant-Microbe Interact.* **2023**, *36*, 682–692. [CrossRef]
42. Lovelace, A.H. ROS, Moss, and a Rossmann Fold Protein: Identification of a Key Signaling Component for Plant Immunity. *Mol. Plant-Microbe Interact.* **2023**, *36*, 675–676. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.