





Article

Whole Genome Sequencing of the Giant Grouper (*Epinephelus lanceolatus*) and High-Throughput Screening of Putative Antimicrobial Peptide Genes

Dengdong Wang^{1,2,†}, Xiyang Chen^{3,4,†}, Xinhui Zhang⁴, Jia Li⁴, Yunhai Yi^{3,4}, Chao Bian⁴, Qiong Shi^{3,4,5}, Haoran Lin^{1,2}, Shuisheng Li^{1,2,*}, Yong Zhang^{1,2,*} and Xinxin You^{3,4,*}

¹ State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory for Aquatic Economic Animals and Guangdong Provincial Engineering Technology Research Center for Healthy Breeding of Important Economic Fish, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China

² Zhanjiang Bay Laboratory, Guangdong Research Center on Reproductive Control and Breeding Technology of Indigenous Valuable Fish Species, Fisheries College, Guangdong Ocean University, Zhanjiang 524088, China

³ BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China

⁴ Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen 518083, China

⁵ Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China

* Correspondence: lshuish@mail.sysu.edu.cn (S.L.); lsszy@mail.sysu.edu.cn (Y.Z.); youxinxin@genomics.cn (X.Y.)

† These authors contributed equally to this work.

Received: 30 July 2019; Accepted: 26 August 2019; Published: 28 August 2019



Abstract: Giant groupers, the largest grouper type in the world, are of economic importance in marine aquaculture for their rapid growth. At the same time, bacterial and viral diseases have become the main threats to the grouper industry. Here, we report a high-quality genome of a giant grouper sequenced by an Illumina HiSeq X-Ten and PacBio Bioscience Sequel platform. A total of 254 putative antimicrobial peptide (AMP) genes were identified, which can be divided into 34 classes according to the annotation of the Antimicrobial Peptides Database (APD3). Their locations in pseudochromosomes were also determined. Thrombin-, lectin-, and scolopendin-derived putative AMPs were the three largest parts. In addition, expressions of putative AMPs were measured by our transcriptome data. Two putative AMP genes (*gapdh1* and *gapdh2*) were involved in glycolysis, which had extremely high expression levels in giant grouper muscle. As it has been reported that AMPs inhibit the growth of a broad spectrum of microbes and participate in regulating innate and adaptive immune responses, genome sequencing of this study provides a comprehensive cataloging of putative AMPs of groupers, supporting antimicrobial research and aquaculture therapy. These genomic resources will be beneficial to further molecular breeding of this economically important fish.

Keywords: giant grouper; *Epinephelus lanceolatus*; genome sequencing; antimicrobial peptide; growth

1. Introduction

Groupers are coral reef fishes in the subfamily Epinephelinae of the family Serranidae (order Perciformes), which are known for their delicious taste, tender flesh, and rich nutrition [1]. As economically important fish species in marine aquaculture, groupers reached a worldwide production of 155,000 tons in 2015, with a total value of USD 630 million [2]. More specifically, mainland China is responsible for an estimated 65% of the total production [2]. There are at least

47 grouper species plus 15 grouper hybrids that have been trialed or are currently aquacultured [1]. The giant grouper, *Epinephelus lanceolatus*, is the largest grouper in the world and can grow to 2.3 m, weigh up to 400 kg [3], and it is popular for its rapid growth, reaching up to 3 kg in the first year [4]. *E. lanceolatus* itself is difficult to breed and rear; therefore, incorporating the rapid growth rate of giant groupers in the genome of hybrids has been the major focus of research on hybrid groupers [5]. Today, hybrid groupers account for a notable proportion of production. In mainland China, a commonly farmed hybrid is *E. fuscoguttatus* × *E. lanceolatus*, which is named Hulong grouper and probably accounts for more than 70% of grouper production of mainland China [2]. The molecular mechanisms underlying the superior growth of Hulong groupers have been explored based on RNA-seq, and the results showed that the upregulated expression of the upstream growth hormone and insulin-like growth factor (GH/IGF) axis related genes in the brain and liver, along with upregulated glycolytic genes as well as ryanodine receptors (RyRs) and troponins involved in the calcium signaling pathway in muscle, led to enhanced growth in the Hulong grouper [6,7].

The rapid development of the intensive culture of groupers has led to more and more severe incidences of infectious diseases, and the main impact of disease is economic loss due to reduced production [8]. It was reported that in the Asia-Pacific region, 365 diseases or disease syndrome occurrences were found in groupers, and bacterial and viral diseases constituted 40% and 26% of the problems reported, respectively [1]. Many kinds of antibiotics have been used to control bacterial diseases in grouper aquaculture; however, serious drawbacks have emerged, such as drug residues and the emergence of resistant bacterial strains [9]. Antimicrobial peptides (AMPs) are a diverse class of small cationic peptide molecules that are produced as evolutionarily ancient weapons by multicellular organisms [10]. Several kinds of AMPs from groupers, including epinecidin [11,12], hepcidin [13], defensin [14], and piscidin [15], have been cloned and studied. However, a systematic screening of AMP genes in groupers has not been reported yet.

Access to and utilization of an array of genetic resources within groupers will facilitate the analysis of AMP genes in groupers. The high-density genetic map of groupers [16] and several candidate genes and single nucleotide polymorphism (SNP) sites related to growth have been identified through quantitative trait loci (QTL) analysis [17] and genome-wide association study (GWAS) [18]. Genomic resources, especially the chromosomal-level genome assembly, would greatly help the studies in evolution, phylogeny, and biology of the groupers. In this study, we generated a high-quality genome assembly of the giant grouper by integrating the use of Illumina (San Diego, CA, USA) short reads and PacBio (Menlo Park, CA, USA) long reads. Then, the scaffolds of giant groupers were assembled based on the published high-resolution genetic map of orange-spotted grouper (*E. coioides*) [16], and they were defined as pseudochromosomes of giant groupers. Finally, glycolytic genes and putative AMP genes were identified and located in the pseudochromosomes, and the transcriptomic quantification of putative AMP genes was also determined using the available RNA-seq data [7]. These genomic recourses will be beneficial to further molecular breeding of this economically important fish.

2. Results

2.1. Genome Size Estimation

A total of 82.80 gigabases (Gb) Illumina clean data (approximately 71.07 times that of the estimated 1.165 Gb genome) and 31.15 Gb PacBio sequence data (approximately 26.74 times that of the estimated genome) were generated. A K-mer analysis was performed with clean reads of two short-insert libraries (500 and 800 bp). According to the k-mer number (20,970,039,103) and k-mer depth (18) (Figure 1), the genome size was estimated to be approximately 1.165 Gb.

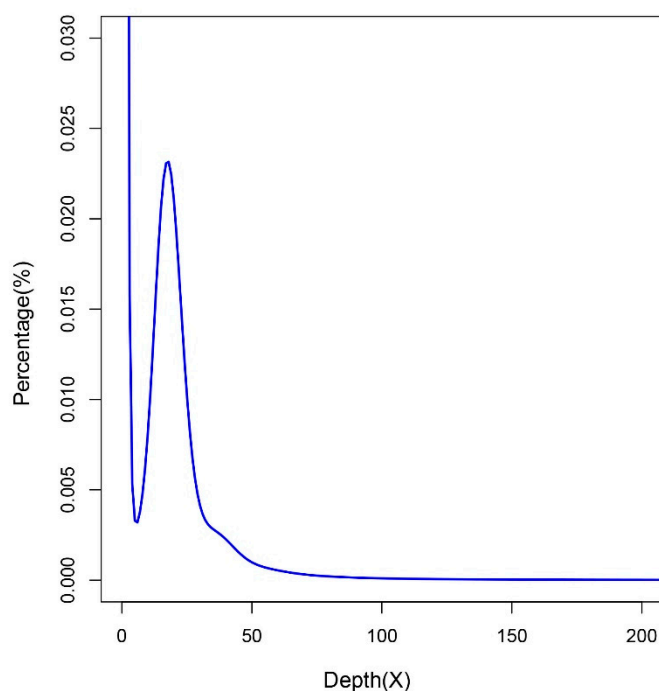


Figure 1. K-mer distribution of the giant grouper with a k-mer size of 17. The x-axis and y-axis are the sequencing depth and percentage of unique 17-mers, respectively.

2.2. De Novo Genome Assembly and Annotation

A total of 3,077,169 contigs were generated, and the length of contig N50 was determined to be 76,419 bp by using clean Illumina sequencing data. Subsequently, hybrid assembly involving PacBio reads that were error corrected by Illumina reads and the Illumina contigs was performed by DBG2OLC [19]. Finally, the achieved total contigs were reduced to 3207 (Table 1), which is much smaller than the assembly from Illumina data. The contig N50 and scaffold N50 of the final genome assembly were 1,469,414 and 1,505,601 bp, respectively (Table 1), reflecting the high quality of assembly. The total genome size reached up to 1.128 Gb, and the GC content was 41.4%. Our assembled genome is of high quality, as a genome completeness assessment by benchmarking universal single-copy orthologs (BUSCO) [20] proved that the assembly contained 93.1% complete gene models.

Table 1. Genome assembly statistics of the giant grouper.

Criteria	Contig	Scaffold
Number	3207	3187
Total length (bp)	1,128,030,970	1,128,030,990
Longest (bp)	9,533,321	9,533,321
N50 (bp)	1,469,414	1,505,601
N90 (bp)	209,611	210,944
>2 kb	3182	3162

Annotation of repeat sequences was performed by de novo and homology predictions based on the RepBase database [21]. The giant grouper genome comprised 45.1% repetitive sequences with a length of 508,638,319 bp, in which 42.4% were transposable elements (TEs) with a length of 478,176,524 bp. DNA transposons (24.6% of the genome) were the most abundant type of TE, followed by long interspersed elements (LINEs, 15.8% of the genome) and long terminal repeats (LTRs, 7.4% of the genome) (Table S1).

De novo predictions based on the repeat-masked genome, RNA-seq predictions based on transcriptomic data from our previous work [6,7], and homolog predictions generated a comprehensive

and nonredundant protein-coding gene set containing 24,794 genes. Annotation completeness, assessed by BUSCO analysis, indicated that gene sequences covered 85% complete single-copy orthologs. The gene set was annotated by InterPro, KEGG, Swiss-Prot, and TrEMBL databases, and approximately 93.37% (23,149 genes) of the gene set were supported by the above-mentioned databases.

2.3. Pseudochromosome Construction

Based on the genetic linkage map of the orange-spotted grouper, a total of 1256 scaffolds were anchored into 24 pseudochromosomes (Chr) of the giant grouper, and a total length of 999.69 Mb was assembled, which comprised 88.62% of the assembled genome sequences and 22,206 genes (from a total of 24,794 genes). The largest pseudochromosome was Chr13 with 54.06 Mb containing 56 scaffolds, and the smallest pseudochromosome was Chr3 with 20.61 Mb containing 18 scaffolds. The average pseudochromosome length was 41.65 Mb with 52 scaffolds (Table 2 and Figure 2). Figure 2 summarizes the distribution of genes, GC content in genomic intervals of 100 kb, and interchromosomal relationships of our assembled giant grouper pseudochromosomes.

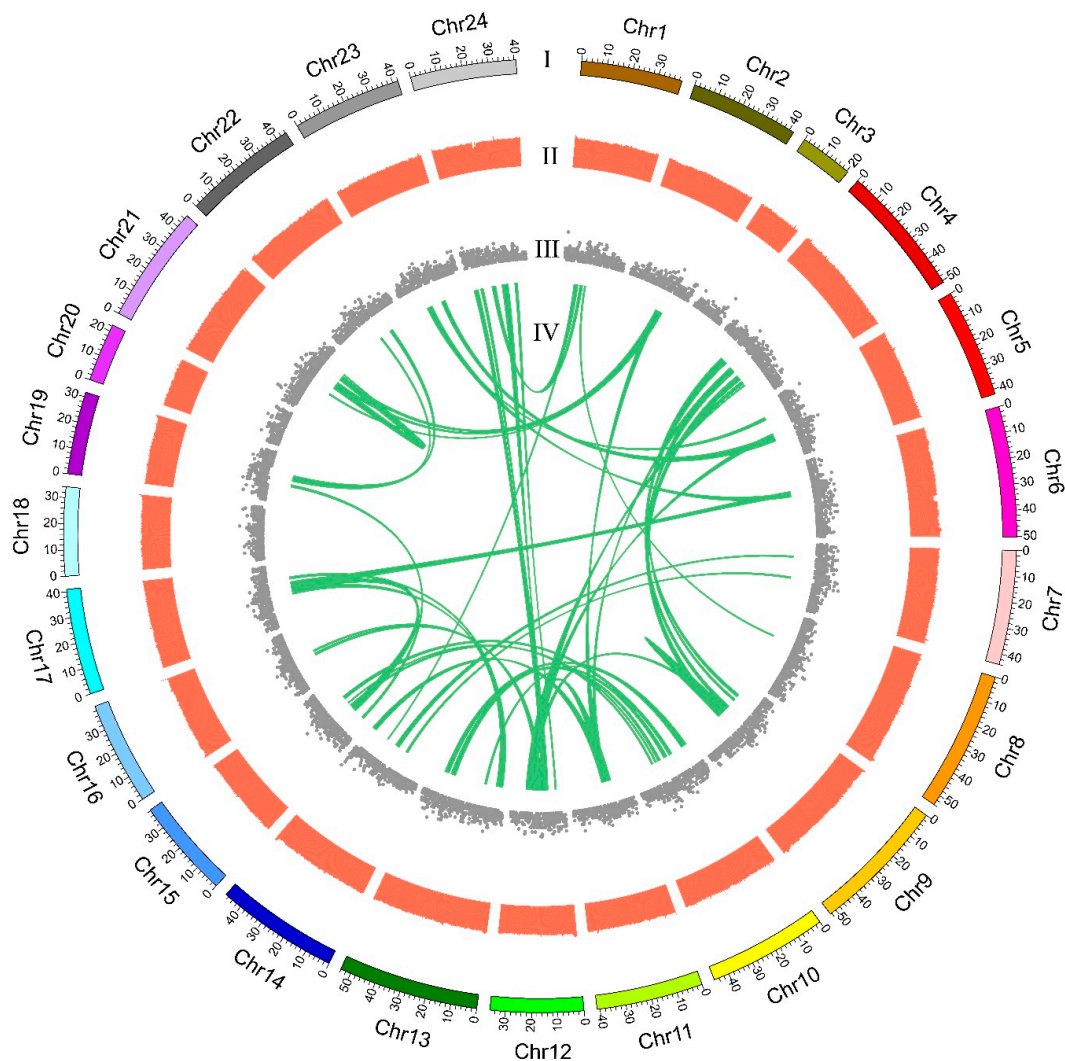


Figure 2. Circos atlas representation of pseudochromosome information. (I) The length of each pseudochromosome. (II) GC content of 100-kb genomic intervals (GC content from 0.25 to 0.51). (III) Density of gene distribution in each 100-kb genomic interval. (IV) Schematic presentation of major interchromosomal relationships in the giant grouper genome, which represents the collinearity of genes between two chromosomes.

Table 2. Characteristics of pseudochromosomes of *Epinephelus lanceolatus*.

Chr	Length (Mb)	Number of Genes	Number of Scaffolds
1	39.05	917	53
2	41.84	899	54
3	20.61	374	18
4	50.74	1198	78
5	41.95	912	53
6	50.39	978	72
7	44.01	1062	55
8	54.01	1051	56
9	52.56	1127	70
10	45.93	1209	49
11	41.12	820	41
12	35.95	918	45
13	54.06	1359	56
14	47.27	947	60
15	38.31	894	35
16	39.82	728	57
17	40.89	858	47
18	34.17	675	58
19	31.81	898	24
20	22.54	491	24
21	45.46	1007	68
22	44.25	944	80
23	42.07	1035	43
24	40.81	905	60
total	999.69	22,206	1256

2.4. Identification, Transcriptomic Quantification, and Annotation of Putative Antimicrobial Peptides (AMPs)

A total of 2927 AMP sequences were collected from the online Antimicrobial Peptides Database (APD3, <http://aps.unmc.edu/AP/main.php>) (Table S2), which were employed as query sequences for putative AMP identification by BLAST. A total of 254 putative AMP genes were obtained (Table S3), which can be divided into 34 classes according to annotation of AMPs in the APD3 (Figure 4a). Each putative AMP gene was renamed by class followed by a serial number. In addition, thrombin-derived C-terminal peptides (TCPs, 64) [22,23], lectin-derived (29), and scolopendin-derived (23) were the top three classes among them, which is consistent with our previous study [24].

We also downloaded another recently published gene set of a giant grouper (PRJNA516312) from the National Center for Biotechnology Information (NCBI) and identified AMPs with the same method. We obtained 326 putative AMPs that were classified into 36 groups (Table S4). TCPs (75), lectin-derived (46), and histone-derived (41) were the top three classes. Comparison between gene set from PRJNA516312 and present study revealed differences in the case of predicted putative AMPs. We speculated that it may be associated with differences in the annotation strategy, which resulted in divergence of the two gene sets.

Based on the transcriptomic data of brain, liver, and muscle from our previous work [6,7], transcripts per million (TPM) values of each putative AMP gene were calculated (Tables S5–S7). TPM values reflected the transcription level of putative AMP genes. Among 254 putative AMP genes, 209, 193, and 177 putative AMP genes with TPM values were detected in brain, liver, and muscle tissues, respectively. The top 20 putative AMPs with high TPM values in each tissue are presented in Table 3.

Table 3. The top 20 transcripts per million (TPM) rankings of antimicrobial peptides (AMPs) or putative AMP precursors in each of the three transcriptomic datasets.

Ranking	Muscle	Liver	Brain
1	GAPDH1 ¹ (49,379.59)	Hemoglobin1 (24,013.91)	Hemoglobin1 (11,531.97)
2	GAPDH2 (5017.39)	Hemoglobin12 (23,219.52)	Hemoglobin12 (7262.71)
3	Hemoglobin12 (2440.09)	sOT2 ² (14,410.88)	sOT2 (1416.06)
4	Hemoglobin1 (2295.59)	Antiproteinase1 (11,271.67)	β2-Microglobulin1 (1136.57)
5	Ap-s ³ (339.63)	Antiproteinase5 (7898.22)	Neuropeptide5 (858.07)
6	β2-Microglobulin1 (258.26)	Antiproteinase2 (6566.73)	BPTI4 ⁴ (654.21)
7	β2-Microglobulin4 (137.71)	Thrombin1 (5759.76)	Neuropeptide6 (631.31)
8	Ubiquicidin (86.92)	GAPDH1 (5378.38)	Sapoin2 (572.89)
9	BPTI16 (85.56)	BPTI12 (4143.13)	Lectin12 (535.93)
10	Sapoin2 (81.02)	Thrombin23 (2039.84)	Synuclein (350.56)
11	BPTI7 (57.05)	Antiproteinase3 (1806.51)	β2-Microglobulin4 (340.17)
12	Lectin25 (43.21)	Thrombin29 (1544.98)	Amyloid2 (250.76)
13	Thrombin53 (40.79)	Thrombin22 (1313.59)	Amyloid1 (249.58)
14	Thrombin31 (38.01)	Thrombin46 (1201.45)	Lysozyme2 (208.82)
15	BPTI15 (32.09)	Thrombin6 (1039.43)	Ubiquicidin (191.88)
16	CcAMP ⁵ (30.21)	β2-Microglobulin1 (662.14)	Thymosin2 (177.76)
17	BPTI14 (27.33)	Thrombin64 (621.69)	Thrombin45 (171.27)
18	Lectin3 (23.87)	Thrombin42 (590.54)	Ubiquitin10 (119.27)
19	Ubiquitin1 (23.61)	Thrombin47 (554.85)	LEAP-2_2 ⁶ (116.44)
20	Ubiquitin5 (21.93)	Thrombin18 (473.18)	Lectin19 (103.89)

¹ glyceraldehyde 3-phosphate dehydrogenase; ² an AMP derived from *Pelodiscus sinensis*; ³ an AMP purified from *Argopecten purpuratus*; ⁴ bovine pancreatic trypsin inhibitor; ⁵ an AMP from *Coridius chinensis*; ⁶ liver-expressed antimicrobial peptide 2. NOTE: TPM values of putative AMPs are exhibited in parentheses.

2.5. Location of Putative AMP Genes and Growth-Related Genes

Out of 254 putative AMPs, 228 were mapped to the 24 assembled pseudochromosomes, with an average number of 9 per chromosome (Figure 3). Chr10 and Chr15 had the highest hits of 22 and 19 genes, respectively. The chromosomes with the lowest counts were Chr20 and Chr3, both with 3 genes. Subsequently, putative AMPs were also confirmed by KEGG enrichment analysis, in which 167 genes were clustered into 35 KEGG items. The representative entries included the immune system (53 genes), signaling molecules and interactions (48 genes), signal transduction (46 genes), and cancers: overview (37 genes) (Figure 4b). Among them, glycogen synthesis may be related to the superior growth of groupers [6]. Thus, in this study, we found that 24 glycolytic- and Ca²⁺-regulating putative AMP genes were located in 18 chromosomes. We found that *gapdh2*, *eno2*, and *tpi1a* were located in Chr22 (Table 4 and Figure 3).

Interestingly, we found that 2 genes (*gapdh1* and *gapdh2*) involved in glycolysis were precursors of predicted AMPs (Figure 3). The expression level of *gapdh1* in the giant grouper muscle was extremely high (Table 5), which has a high identity (96.88%) and query alignment ratio (100%) with the skipjack tuna GAPDH-related antimicrobial peptide (SJGAP) [25]. While *gapdh2* matched to yellowfin tuna glyceraldehyde-3-phosphate dehydrogenase-related antimicrobial peptide (YFGAP) [25], with 87.50% identity and 100.00% query alignment ratio. SJGAP and YFGAP are AMPs from the skin of skipjack tuna (*Katsuwonus pelamis*) and yellowfin tuna (*Thunnus albacares*), respectively, and both have potent antimicrobial activities [25,26]. To investigate the alignments of SJGAP and YFGAP in *gapdh*, we performed multiple sequence alignments of *gapdh1* (Figure 5a) and *gapdh2* (Figure 5b) from zebrafish, yellowfin tuna, and giant groupers. As shown in Figure 5, *gapdh1* of zebrafish, yellowfin tuna, and giant groupers showed higher similarity with YFGAP and SJGAP than *gapdh2*, suggesting that *gapdh1* are more likely to play a role in the antimicrobial process in these fishes. Gene structures of *gapdh1* and *gapdh2* are also exhibited in Figure 5.

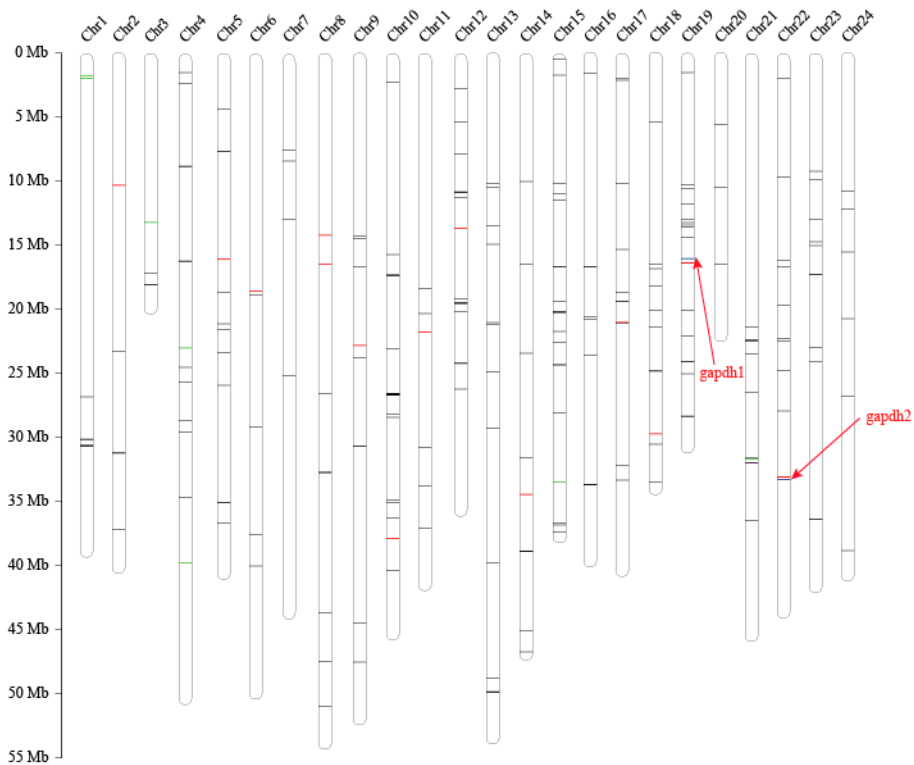


Figure 3. Pseudochromosome lengths, genes involved in AMPs and glycolysis, and Ca²⁺-regulating genes of the giant grouper. NOTE: Black bars represent AMPs; red bars represent genes involved glycolysis; green bars represent genes involved Ca²⁺ regulation; and blue bars represent the gapdhs, which are both AMPs and glycolysis.

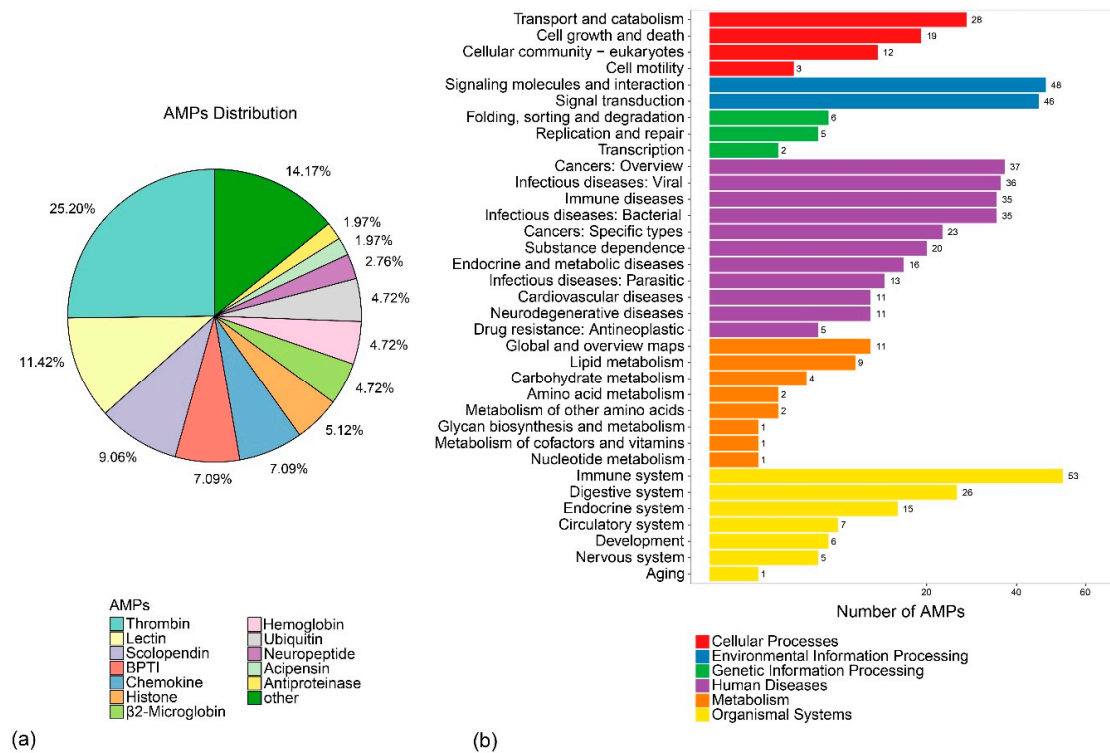


Figure 4. (a) AMP distribution and (b) KEGG metabolic pathway annotation of AMPs.

Table 4. Location of glycolytic- and Ca²⁺-regulating genes in the giant grouper pseudochromosome.

Gene Name	Chr	Gene ID	Function Type
<i>tnt-fast</i> ¹	Chr1	longdun_GLEAN_10010987	Ca ²⁺ regulating
<i>tnt-skeletal</i> ²	Chr1	longdun_GLEAN_10010985	Ca ²⁺ regulating
<i>pgk</i> ³	Chr2	longdun_GLEAN_10005384	Glycolytic
<i>tnt-slow</i>	Chr3	longdun_GLEAN_10021266	Ca ²⁺ regulating
<i>tnc</i> ⁴	Chr4	longdun_GLEAN_10018260	Ca ²⁺ regulating
<i>tnt-cardiac</i>	Chr4	longdun_GLEAN_10017734	Ca ²⁺ regulating
<i>pgam2</i> ⁵	Chr5	longdun_GLEAN_10022627	Glycolytic
<i>pgm2</i> ⁶	Chr6	longdun_GLEAN_10022325	Glycolytic
<i>gPi</i> ⁷	Chr8	longdun_GLEAN_10012880	Glycolytic
<i>pyk</i> ⁸	Chr8	longdun_GLEAN_10018335	Glycolytic
<i>pfk-muscle</i> ⁹	Chr9	longdun_GLEAN_10019679	Glycolytic
<i>ald</i> ¹⁰	Chr10	longdun_GLEAN_10019814	Glycolytic
<i>pfk-liver</i>	Chr11	longdun_GLEAN_10018530	Glycolytic
<i>pgm1</i>	Chr12	longdun_GLEAN_10020822	Glycolytic
<i>pgm3</i>	Chr14	longdun_GLEAN_10008616	Glycolytic
<i>ryr2</i> ¹¹	Chr15	longdun_GLEAN_10008973	Ca ²⁺ regulating
<i>pgam1a</i>	Chr17	longdun_GLEAN_10012515	Glycolytic
<i>eno1</i> ¹²	Chr18	longdun_GLEAN_10002289	Glycolytic
<i>gapdh1</i> ¹³	Chr19	longdun_GLEAN_10017174	Glycolytic
<i>tpi1b</i> ¹⁴	Chr19	longdun_GLEAN_10017191	Glycolytic
<i>ryr1</i>	Chr21	longdun_GLEAN_10010008	Ca ²⁺ regulating
<i>gapdh2</i>	Chr22	longdun_GLEAN_10014462	Glycolytic
<i>eno2</i>	Chr22	longdun_GLEAN_10014466	Glycolytic
<i>tpi1a</i>	Chr22	longdun_GLEAN_10014467	Glycolytic

¹ troponin I; ² troponin T; ³ phosphoglycerate kinase; ⁴ troponin C; ⁵ phosphoglycerate mutase; ⁶ phosphoglucomutase; ⁷ phosphoglucose isomerase; ⁸ pyruvate kinase; ⁹ phosphofructokinases; ¹⁰ fructose-bisphosphate aldolase; ¹¹ ryanodine receptor; ¹² enolases; ¹³ glyceraldehyde phosphate dehydrogenase; and ¹⁴ triosephosphate isomerase.

Table 5. Congruent relationship and transcription levels of the 2 growth-related putative AMP genes.

Putative AMP Gene Name	Gene ID	Query AMP (AMP ID in APD3 Database)	TPM		
			Brain	Liver	Muscle
<i>gapdh1</i>	longdun_GLEAN_10017174	Skipjack tuna GAPDH-related antimicrobial peptide (SJGAP) (2680) Yellowfin tuna	4.77	5378.38	49379.59
<i>gapdh2</i>	longdun_GLEAN_10014462	glyceraldehyde-3-phosphate dehydrogenase-related antimicrobial peptide (YFGAP) (2012)	12.19	0.17	5017.39

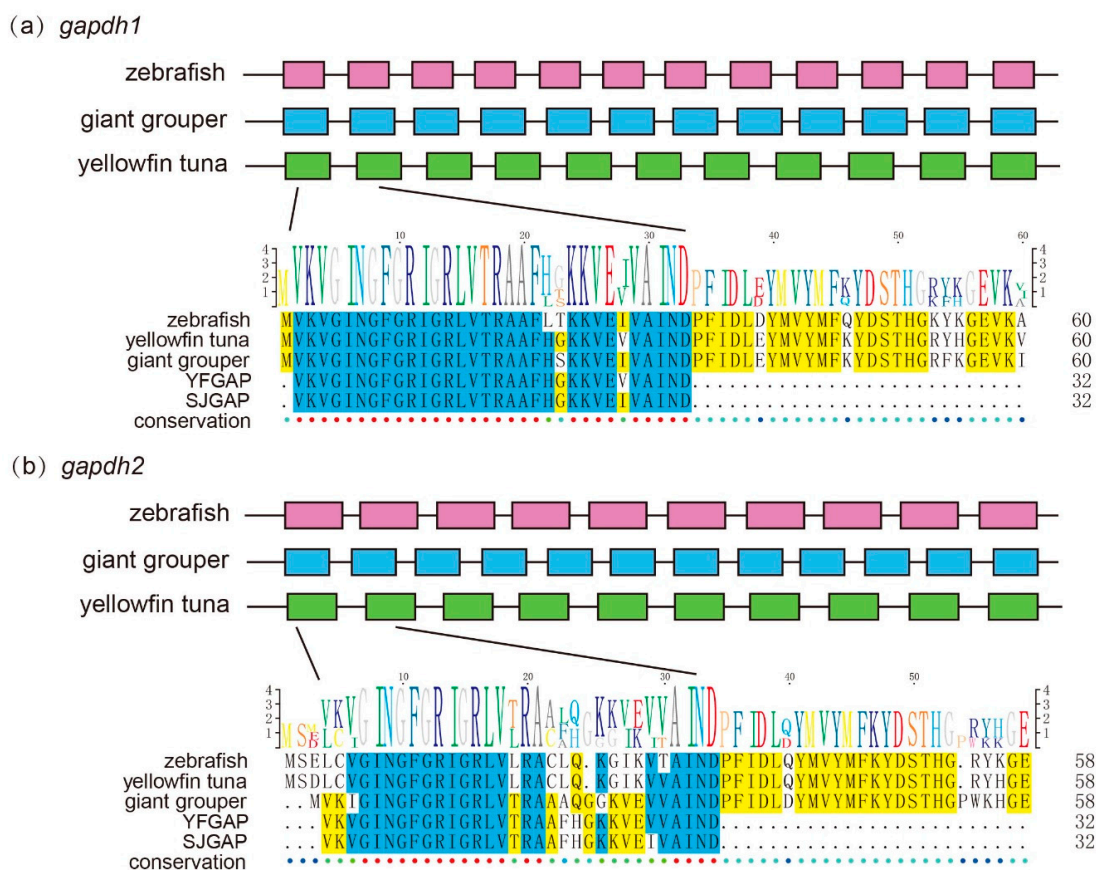


Figure 5. Structure of the 2 growth-related putative AMP genes (*gapdh1* and *gapdh2*) in zebrafish, giant groupers, and yellowfin tuna. The pink, blue, and green boxes represent coding sequence (CDS) of genes in zebrafish, giant groupers, and yellowfin tuna, respectively. Multiple sequence alignments of partial *gapdh1* (a) and *gapdh2* (b) that match to YFGAP and SJGAP from zebrafish, yellowfin tuna, and giant grouper were displayed. Blue and yellow marks represent >80% and >50% identity, respectively.

3. Discussion

Along with the publication of genomes of model organisms and species with specific evolutionary characteristics, an increasing number of important crops and economic animals have been sequenced, as in the case of rice [27], wheat [28], barley [29], maize [30,31], soybeans [32], and cotton [33] as well as cows [34], pigs [35,36], sheep [37], goats [38,39], cod [40], sea basses [41,42], mudskippers [43], salmonids [44,45], carps [46,47], and tongue sole [48]. In this study, we sequenced the genome of giant groupers with the purpose of systematically gaining its genetic information and providing opportunities to accelerate breeding improvement.

Most of the genomes of economic crops or animals have focused on growth traits. While some fish genome projects found immune genes that were lost (elephant shark [49], Atlantic cod [40]) or expanded (large yellow croaker [50]), we performed this project to screen putative AMPs with an attempt to explore immune resources for bacterial and viral disease therapy. Especially, some AMPs have been widely used in agriculture as potential alternatives to antibiotics [51]. This work may help those who make efforts to develop drugs for groupers and reduce the usage of antibiotics and other chemically poisoning drugs.

A total of 254 putative AMP genes of the giant grouper were classified into 34 groups in the present study. Among them, thrombin (64 AMPs), lectin (29 AMPs) and scolopendin (23 AMPs) were the top three groups. Thrombin was also the largest grouper in our previous works, including the blue tilapia (*Oreochromis aureus*), Nile tilapia (*Oreochromis niloticus*) [52], blue-spotted mudskipper (*Boleophthalmus pectinirostris*), and giant-fin mudskipper (*Periophthalmus magnuspinnatus*) [24]. Lectin

also accounted for a major part in the above-mentioned four fishes. It seems that thrombin may play a vital role in fish.

It has been reported that several kinds of AMPs, including epinecidin [11,12], hepcidin [13], defensin [14], and piscidin [15], have been cloned and studied from groupers. We identified EC-hepcidin1 (query ID 1701 in Table S2), a hepcidin AMP derived from the liver and stomach of orange-spotted grouper [13], from the giant grouper gene set. Enap-1 (query ID 294 in Table S2), a defensin from horse (*Equus caballus*) [53], was also identified in giant groupers. However, the other two AMPs that have been reported in groupers were not found in our annotated gene set.

Cumulative evidence is showing that AMPs not only inhibit the growth of a broad spectrum of microbes through membrane disruption [10], but they also participate in regulating innate and adaptive immune responses [54,55]. High expression levels of the *gapdh1* gene in the liver and muscle of giant groupers may imply its active glycolysis activity along with antimicrobial activity. Previous studies have reported that strong antibacterial activities against both Gram-positive and Gram-negative bacteria of N-terminal segments of this protein were shown in tuna fish [25,26]. Antifungal efficacies of GAPDH-derived peptide were also demonstrated in many studies [56,57]. However, GAPDH has not been shown to be cleaved in groupers, and it is possible that cleavage might be tissue specific. Even if it has a high level in muscle, it would be predicted to be at full length and enzymatically active for glycolysis only, whereas it might be cleaved to AMPs in skin or other tissues. Antimicrobial functions of this peptide (if produced) are worthy of further investigation in giant groupers.

Moreover, the great majority of giant groupers are born to be female, barely male, and some, but mostly one of them in the population, would change its gender to male after the first or second maturation (all females have the ability to change sex). This kind of protogynous hermaphrodite is mostly shared by groupers [58]. Materials in this work could provide opportunities to explore the mechanisms of sex change in groupers.

4. Materials and Methods

4.1. Sample Preparation and Sequencing

Genomic DNA was extracted from the muscle of a wild giant grouper cultured in the Guangdong Marine Fishery Experimental Center, Huizhou, China. Six libraries, including three short-insert libraries (270, 500, and 800 bp) and three long-insert libraries (2, 5, and 10 kb), were constructed for sequencing by an Illumina HiSeq X-Ten platform (Illumina, San Diego, CA, USA), except for the 800 bp insert-size library, which was sequenced by an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). For high-quality assembly, we also constructed a 20 kb insert library for sequencing with the PacBio Bioscience Sequel platform (Pacific Biosciences, Menlo Park, CA, USA). All experiments were carried out according to the guidelines of the Animal Ethics Committee and were approved by the Institutional Review Board on Bioethics and Biosafety of BGI (No. FT14015).

All sequenced data from *E. lanceolatus* are available in the NCBI database at BioProject ID, PRJNA533524. All Illumina reads are available under accession numbers SRR8926032, SRR8926031, SRR8926030, SRR8926029, SRR8926035, SRR8926034, and SRR8926033 in the NCBI database.

4.2. Genome Assembly

The Illumina raw sequences with adapter contamination, low quality, and replicated PCR were filtered out by a SOAPfilter (version 2.2, BGI-Shenzhen, Shenzhen, China). PacBio raw data were corrected by Illumina clean reads through LoRDEC (version 0.4.1, <http://www.atgc-montpellier.fr/lordec/>) [59]. The Illumina clean reads were assembled by Platanus (version 1.2.4, Tokyo Institute of Technology, Tokyo, Japan) [60] to construct contigs. Subsequently, contigs were aligned to PacBio reads by DBG2OLC [19] to construct consensus contigs. Finally, the assembled genome was polished by Pilon (version 1.22, Broad Institute of MIT and Harvard, Cambridge, MA, USA) [61] with short-insert library reads.

4.3. Pseudochromosome Construction

SNP-containing reads in the genetic linkage map from the orange-spotted grouper [16] were mapped to the giant grouper assembled genome sequence, and only the matching reads were selected. Linkage groups (LGs) of the giant grouper were assigned using JoinMap4.1 software (Kyazma, Wageningen, Netherlands) [62]. Subsequently, a genetic linkage map was constructed. Single nucleotide polymorphisms in the genetic linkage map of the giant grouper were used for assembling the pseudochromosomes. To increase the accuracy of pseudochromosome assembly, we chose at least two SNPs in each scaffold [63]. Based on genetic distances between SNP markers, we determined the position and orientation of each scaffold and anchored these scaffolds to construct pseudochromosomes.

4.4. Repeat Annotation

Repeat elements were predicted by de novo and homology methods. De novo predictions were performed by LTR_FINDER (version 1.0.6, Fudan University, Shanghai, China) [64] and RepeatModeler (version 1.08, <http://www.repeatmasker.org/RepeatModeler/>) [65]. The merged repeat library was aligned to the assembled genome sequence by RepeatMasker (version 4.06, Institute for Systems Biology, Seattle, WA, USA) to produce repeat elements [65]. The homology prediction based on RepBase (version 21.01, Genetic Information Research Institute, Sunnyvale, CA, USA) was performed by RepeatMasker and RepeatProteinMask (version 4.06, Institute for Systems Biology, Seattle, WA, USA) [65]. Subsequently, nonredundant repeat elements were obtained by integrating de novo and homology data.

4.5. Gene Annotation

We applied three different strategies to predict the protein-coding genes. For the de novo prediction strategy, AUGUSTUS (version 2.5, Institute of Microbiology and Genetics, University of Göttingen, Göttingen, Germany) [66] and GENSCAN (version 1.0, Stanford University, Stanford, CA, USA) [67] were employed to predict genes from the repeat-masked genome. The second strategy was homology-based annotation. Protein sequences of three-spined stickleback (*Gasterosteus aculeatus*), spotted gar (*Lepisosteus oculatus*), Nile tilapia (*Oreochromis niloticus*), medaka (*Oryzias latipes*), Japanese puffer (*Takifugu rubripes*), spotted green pufferfish (*Tetraodon nigroviridis*), platyfish (*Xiphophorus maculatus*), and zebrafish (*Danio rerio*) were downloaded from Ensembl database (release version 90, <https://asia.ensembl.org/index.html>). Protein sequences of Asian seabass (*Lates calcarifer*) were downloaded from <http://seabass.sanbi.ac.za>. Downloaded protein sequences were aligned to the assembled genome of the giant grouper by TBLASTN (e-value: $1e-5$) [68]. GeneWise (version 2.2.0, The European Bioinformatics Institute, Cambridge, UK) [69] was used to predict the gene structure of each BLAST hit. For the third strategy, transcriptome-based prediction, raw data were downloaded from the National Center for Biotechnology Information (NCBI), including liver, brain, and muscle transcriptomic data. Subsequently, we employed TopHat (version 2.1.1, Johns Hopkins University, Baltimore, MD, USA) [70] and cufflinks (version 2.2.1, <http://cufflinks.cbc.umd.edu/>) [71] with raw reads to predict protein-coding genes. As a result, integrated and nonredundant gene sets were obtained by GLEAN [72] with the above-mentioned three results.

The predicted gene sets were aligned to InterPro [73], KEGG [74], TrEMBL, and Swiss-Prot [75] databases to accomplish functional annotation.

4.6. Identification and Transcriptomic Quantification of AMPs

A total of 2927 AMP sequences that have been reported to exhibit antimicrobial activity were collected from the APD3 database as a query sequence (Table S2). An index database of annotated gene sets was built for alignment by makeblastdb command. Collected active AMPs sequences were aligned to gene set sequences to identify potential AMPs based on sequence similarity by TBLASTN (e-value: $1e-5$). Alignment hits were dealt with by in-house scripts. Those hits with

a query alignment ratio less than 0.5 were filtered out, and redundant data were also removed. To calculate the TPM value of putative AMP genes, we performed referring sequence-based transcript quantification. Raw reads were filtered by SOAPnuke filter tools (version 1.5.6, BGI-Shenzhen, Shenzhen, China) [76]. Clean reads were mapped to the assembled genome by HISAT2 (version 2.0.4, <https://github.com/DaehwanKimLab/hisat2>) [77]. Subsequently, TPM values of each transcriptome were calculated by RSEM (version 1.2.12, <https://deweylab.github.io/RSEM/>) [78].

5. Conclusions

We report a high-quality genome of the giant grouper. The assembly reached up to 1.128 Gb, accounting for 96.8% of the estimated genome size. A total of 24,794 protein-coding genes were annotated through de novo prediction, transcriptomic data, and homolog prediction. Then, 254 putative AMP genes were identified, located in pseudochromosomes, and expressions were measured. Two putative AMP genes were connected to glycolysis, of which *gapdh1* was highly expressed in muscle. Genome sequencing let us identify AMPs systematically in groupers so as to support antimicrobial research and possibly provide suggestions for therapy. These genomics resources will be beneficial for further molecular breeding of this economically important fish. This work shall aid in the effort against infectious diseases in the giant grouper industry.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1660-3397/17/9/503/s1>, Table S1: Annotation of repeat sequences of giant grouper, Table S2: AMP sequences collected from APD3, Table S3: putative AMP genes identified in the giant grouper genome, Table S4: Summary of the 326 identified AMPs from another giant grouper genome (PRJNA516312) and statistics of its classification, Table S5: TPM of AMPs in the brain, Table S6: TPM of AMPs in the liver, and Table S7: TPM of AMPs in the muscle.

Author Contributions: Q.S., S.L., H.L., Y.Z., and X.Y. conceived and designed this project. D.W., X.C., X.Z., J.L., Y.Y., and C.B. performed data analyses. D.W., X.C., X.Z., and X.Y. wrote the paper.

Funding: This research was funded by the Independent Project of Guangdong Province Laboratory (ZJW-2019-06), National Natural Science Foundation of China (31672631, 31872572), Guangdong Provincial Natural Science Foundation (2018B030311026, 2018A030313890), Guangdong Provincial Science and Technology Program (2017B0202450001, 2017B090904022), Guangdong Provincial Special Fund For Modern Agriculture Industry Technology Innovation Teams, the Special Fund for Fisheries-Scientific Research of Guangdong Province (SDYY-2018-04), Special Fund of State Key Laboratory of Developmental Biology of Freshwater (2018KF001) and Shenzhen Dapeng Special Program for Industrial Development (KT20170205).

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

References

1. Rimmer, M.A.; Glamuzina, B. A review of grouper (Family Serranidae: Subfamily Epinephelinae) aquaculture from a sustainability science perspective. *Rev. Aquac.* **2019**, *11*, 58–87. [[CrossRef](#)]
2. FAO. *FishStatJ, a Tool for Fishery Statistics Analysis*; FAO Fisheries and Aquaculture Department, FIPS–Statistics and information: Rome, Italy, 2017.
3. Bright, D.; Reynolds, A.; Nguyen, N.H.; Knuckey, R.; Knibb, W.; Elizur, A. A study into parental assignment of the communal spawning protogynous hermaphrodite, giant grouper (*Epinephelus lanceolatus*). *Aquaculture* **2016**, *459*, 19–25. [[CrossRef](#)]
4. Sadvoy, Y.J.; Donaldson, T.J.; Graham, T.R.; McGilvray, F.; Muldoon, G.J.; Philipps, M.J.; Rimmer, M.A.; Smith, A.; Yeeting, B. While Stocks Last: The Live Reef Food Fish Trade. *Asian Dev. Bank* **2003**, *1632*, 169.
5. Fan, B.; Liu, X.C.; Meng, Z.N.; Tan, B.H.; Wang, L.; Zhang, H.F.; Zhang, Y.; Wang, Y.X.; Lin, H.R. Cryopreservation of giant grouper *Epinephelus lanceolatus* (Bloch, 1790) sperm. *J. Appl. Ichthyol.* **2014**, *30*, 334–339. [[CrossRef](#)]
6. Sun, Y.; Guo, C.-Y.; Wang, D.-D.; Li, X.F.; Xiao, L.; Zhang, X.; You, X.; Shi, Q.; Hu, G.-J.; Fang, C.; et al. Transcriptome analysis reveals the molecular mechanisms underlying growth superiority in a novel grouper hybrid (*Epinephelus fuscogutatus*♀ × *E. lanceolatus*♂). *BMC Genet.* **2016**, *17*, 175. [[CrossRef](#)] [[PubMed](#)]
7. Sun, Y.; Huang, Y.; Hu, G.; Zhang, X.; Ruan, Z.; Zhao, X.; Guo, C.; Tang, Z.; Li, X.; You, X.; et al. Comparative Transcriptomic Study of Muscle Provides New Insights into the Growth Superiority of a Novel Grouper Hybrid. *PLoS ONE* **2016**, *11*, e0168802. [[CrossRef](#)]

8. Harikrishnan, R.; Balasundaram, C.; Heo, M. Fish health aspects of grouper aquaculture Fish health aspects in grouper aquaculture. *Aquaculture* **2017**, *320*, 1–21. [[CrossRef](#)]
9. Chuang, S.-C.; Huang, W.-L.; Kau, S.-W.; Yang, Y.-P.; Yang, C.-D. Pleurocidin Peptide Enhances Grouper Anti-Vibrio harveyi Immunity Elicited by Poly(lactide-co-glycolide)-Encapsulated Recombinant Glyceraldehyde-3-phosphate Dehydrogenase. *Vaccines* **2014**, *2*, 380–396. [[CrossRef](#)]
10. Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **2002**, *415*, 389–395. [[CrossRef](#)]
11. Yin, Z.-X.; He, W.; Chen, W.-J.; Yan, J.-H.; Yang, J.-N.; Chan, S.-M.; He, J.-G. Cloning, expression and antimicrobial activity of an antimicrobial peptide, epinecidin-1, from the orange-spotted grouper, *Epinephelus coioides*. *Aquaculture* **2006**, *253*, 204–211. [[CrossRef](#)]
12. Pan, C.-Y.; Chen, J.-Y.; Cheng, Y.-S.E.; Chen, C.-Y.; Ni, I.-H.; Sheen, J.-F.; Pan, Y.-L.; Kuo, C.-M. Gene Expression and Localization of the Epinecidin-1 Antimicrobial Peptide in the Grouper (*Epinephelus coioides*), and Its Role in Protecting Fish against Pathogenic Infection. *DNA Cell Biol.* **2007**, *26*, 403–413. [[CrossRef](#)] [[PubMed](#)]
13. Zhou, J.-G.; Wei, J.-G.; Xu, D.; Cui, H.-C.; Yan, Y.; Ou-Yang, Z.-L.; Huang, X.-H.; Huang, Y.-H.; Qin, Q.-W. Molecular cloning and characterization of two novel hepcidins from orange-spotted grouper, *Epinephelus coioides*. *Fish Shellfish Immunol.* **2011**, *30*, 559–568. [[CrossRef](#)] [[PubMed](#)]
14. Guo, M.; Wei, J.; Huang, X.; Huang, Y.; Qin, Q. Antiviral effects of β -defensin derived from orange-spotted grouper (*Epinephelus coioides*). *Fish Shellfish Immunol.* **2012**, *32*, 828–838. [[CrossRef](#)] [[PubMed](#)]
15. Li, Z.P.; Chen, D.W.; Pan, Y.Q.; Deng, L. Two isoforms of piscidin from Malabar grouper, *Epinephelus malabaricus*: Expression and functional characterization. *Fish Shellfish Immunol.* **2016**, *57*, 222–235. [[CrossRef](#)] [[PubMed](#)]
16. You, X.; Shu, L.; Li, S.; Chen, J.; Luo, J.; Lu, J.; Mu, Q.; Bai, J.; Xia, Q.; Chen, Q.; et al. Construction of high-density genetic linkage maps for orange-spotted grouper *Epinephelus coioides* using multiplexed shotgun genotyping. *BMC Genet.* **2013**, *14*, 113. [[CrossRef](#)]
17. Yu, H.; You, X.; Li, J.; Liu, H.; Meng, Z.; Xiao, L.; Zhang, H.; Lin, H.R.; Zhang, Y.; Shi, Q. Genome-wide mapping of growth-related quantitative trait loci in orange-spotted grouper (*Epinephelus coioides*) using double digest restriction-site associated DNA sequencing (ddRADseq). *Int. J. Mol. Sci.* **2016**, *17*, 501. [[CrossRef](#)] [[PubMed](#)]
18. Yu, H.; You, X.; Li, J.; Zhang, X.; Zhang, S.; Jiang, S.; Lin, X.; Lin, H.R.; Meng, Z.; Shi, Q. A genome-wide association study on growth traits in orange-spotted grouper (*Epinephelus coioides*) with RAD-seq genotyping. *Sci. China Life Sci.* **2018**, *61*, 934–946. [[CrossRef](#)] [[PubMed](#)]
19. Ye, C.; Hill, C.M.; Wu, S.; Ruan, J.; Ma, Z.S. DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Sci. Rep.* **2016**, *6*, 31900. [[CrossRef](#)] [[PubMed](#)]
20. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
21. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [[CrossRef](#)]
22. Pählman, L.I.; Mörgelin, M.; Kasetty, G.; Olin, A.I.; Schmidtchen, A.; Herwald, H. Antimicrobial activity of fibrinogen and fibrinogen-derived peptides—a novel link between coagulation and innate immunity. *Thromb. Haemost.* **2013**, *109*, 930–939. [[PubMed](#)]
23. Papareddy, P.; Rydengård, V.; Pasupuleti, M.; Walse, B.; Mörgelin, M.; Chalupka, A.; Malmsten, M.; Schmidtchen, A. Proteolysis of human thrombin generates novel host defense peptides. *PLoS Pathog.* **2010**, *6*, e1000857. [[CrossRef](#)] [[PubMed](#)]
24. Yi, Y.; You, X.; Bian, C.; Chen, S.; Lv, Z.; Qiu, L.; Shi, Q. High-Throughput Identification of Antimicrobial Peptides from Amphibious Mudskippers. *Mar. Drugs* **2017**, *15*, 364. [[CrossRef](#)] [[PubMed](#)]
25. Seo, J.K.; Lee, M.J.; Go, H.J.; Kim, Y.J.; Park, N.G. Antimicrobial function of the GAPDH-related antimicrobial peptide in the skin of skipjack tuna, *Katsuwonus pelamis*. *Fish Shellfish Immunol.* **2014**, *36*, 571–581. [[CrossRef](#)] [[PubMed](#)]
26. Seo, J.-K.; Lee, M.J.; Go, H.-J.; Park, T.H.; Park, N.G. Purification and characterization of YFGAP, a GAPDH-related novel antimicrobial peptide, from the skin of yellowfin tuna, *Thunnus albacares*. *Fish Shellfish Immunol.* **2012**, *33*, 743–752. [[CrossRef](#)] [[PubMed](#)]

27. Du, H.; Yu, Y.; Ma, Y.; Gao, Q.; Cao, Y.; Chen, Z.; Ma, B.; Qi, M.; Li, Y.; Zhao, X.; et al. Sequencing and de novo assembly of a near complete indica rice genome. *Nat. Commun.* **2017**, *8*, 15324. [[CrossRef](#)]
28. Luo, M.C.; Gu, Y.Q.; Puiu, D.; Wang, H.; Twardziok, S.O.; Deal, K.R.; Huo, N.; Zhu, T.; Wang, L.; Wang, Y.; et al. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* **2017**, *551*, 498–502. [[CrossRef](#)]
29. Mascher, M.; Gundlach, H.; Himmelbach, A.; Beier, S.; Twardziok, S.O.; Wicker, T.; Radchuk, V.; Dockter, C.; Hedley, P.E.; Russell, J.; et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **2017**, *544*, 427. [[CrossRef](#)]
30. Yang, N.; Xu, X.W.; Wang, R.R.; Peng, W.L.; Cai, L.; Song, J.M.; Li, W.; Luo, X.; Niu, L.; Wang, Y.; et al. Contributions of *Zea mays* subspecies mexicana haplotypes to modern maize. *Nat. Commun.* **2017**, *8*, 1874. [[CrossRef](#)]
31. Jiao, Y.; Peluso, P.; Shi, J.; Liang, T.; Stitzer, M.C.; Wang, B.; Campbell, M.S.; Stein, J.C.; Wei, X.; Chin, C.S.; et al. Improved maize reference genome with single-molecule technologies. *Nature* **2017**, *546*, 524–527. [[CrossRef](#)]
32. Fang, C.; Ma, Y.; Wu, S.; Liu, Z.; Wang, Z.; Yang, R.; Hu, G.; Zhou, Z.; Yu, H.; Zhang, M.; et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **2017**, *18*, 161. [[CrossRef](#)] [[PubMed](#)]
33. Fang, L.; Wang, Q.; Hu, Y.; Jia, Y.; Chen, J.; Liu, B.; Zhang, Z.; Guan, X.; Chen, S.; Zhou, B.; et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* **2017**, *49*, 1089–1098. [[CrossRef](#)] [[PubMed](#)]
34. Zimin, A.V.; Delcher, A.L.; Florea, L.; Kelley, D.R.; Schatz, M.C.; Puiu, D.; Hanrahan, F.; Pertea, G.; Van Tassell, C.P.; Sonstegard, T.S.; et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **2009**, *10*, R42. [[CrossRef](#)] [[PubMed](#)]
35. Zhu, H.; Shuai, S.; Liu, Y.; Lou, P.; Li, R.; Shen, X.; Zhang, M.; Zhou, C.; Li, M.; Zhang, Y.; et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* **2013**, *45*, 1431–1438.
36. Rubin, C.-J.; Megens, H.-J.; Martinez Barrio, A.; Maqbool, K.; Sayyab, S.; Schwochow, D.; Wang, C.; Carlborg, Ö.; Jern, P.; Jørgensen, C.B.; et al. Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 19529–19536. [[CrossRef](#)]
37. Jiang, Y.; Xie, M.; Chen, W.; Talbot, R.; Maddox, J.F.; Faraut, T.; Wu, C.; Muzny, D.M.; Li, Y.; Zhang, W.; et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **2014**, *344*, 1168–1173. [[CrossRef](#)]
38. Dong, Y.; Xie, M.; Jiang, Y.; Xiao, N.; Du, X.; Zhang, W.; Tosser-Klopp, G.; Wang, J.; Yang, S.; Liang, J.; et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* **2013**, *31*, 135–141. [[CrossRef](#)]
39. Hutchison, J.L.; Nystrom, J.C.; Schwartz, J.C.; Hastie, A.R.; Smith, T.P.L.; Liachko, I.; Kelley, C.M.; Lam, E.T.; Van Tassell, C.P.; Phillippy, A.M.; et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **2017**, *49*, 643–650.
40. Star, B.; Nederbragt, A.J.; Jentoft, S.; Grimholt, U.; Malmstrøm, M.; Gregers, T.F.; Rounge, T.B.; Paulsen, J.; Solbakken, M.H.; Sharma, A.; et al. The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **2011**, *477*, 207–210. [[CrossRef](#)]
41. Tine, M.; Kuhl, H.; Gagnaire, P.A.; Louro, B.; Desmarais, E.; Martins, R.S.T.; Hecht, J.; Knaust, F.; Belkhir, K.; Klages, S.; et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat. Commun.* **2014**, *5*, 5770. [[CrossRef](#)]
42. Vij, S.; Kuhl, H.; Kuznetsova, I.S.; Komissarov, A.; Yurchenko, A.A.; Van Heusden, P.; Singh, S.; Thevasagayam, N.M.; Prakki, S.R.S.; Purushothaman, K.; et al. Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-layered Scaffolding. *PLoS Genet.* **2016**, *12*, 1–35.
43. You, X.; Bian, C.; Zan, Q.; Xu, X.; Liu, X.; Chen, J.; Wang, J.; Qiu, Y.; Li, W.; Zhang, X.; et al. Mudskipper genomes provide insights into terrestrial adaptation of amphibious fishes. *Nat. Commun.* **2014**, *5*, 5594. [[CrossRef](#)] [[PubMed](#)]
44. Davidson, W.S.; Koop, B.F.; Jones, S.J.M.; Iturra, P.; Vidal, R.; Maass, A.; Jonassen, I.; Lien, S.; Omholt, S.W. Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol.* **2010**, *11*, 403. [[PubMed](#)]

45. Lien, S.; Koop, B.F.; Sandve, S.R.; Miller, J.R.; Kent, M.P.; Nome, T.; Hvidsten, T.R.; Leong, J.S.; Minkley, D.R.; Zimin, A.; et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **2016**, *533*, 200–205. [[CrossRef](#)] [[PubMed](#)]
46. Wang, Y.; Lu, Y.; Zhang, Y.; Ning, Z.; Li, Y.; Zhao, Q.; Lu, H.; Huang, R.; Xia, X.; Feng, Q.; et al. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat. Genet.* **2015**, *47*, 625–631. [[CrossRef](#)] [[PubMed](#)]
47. Xu, P.; Zhang, X.; Wang, X.; Li, J.; Liu, G.; Kuang, Y.; Xu, J.; Zheng, X.; Ren, L.; Wang, G.; et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat. Genet.* **2014**, *46*, 1212–1219. [[CrossRef](#)] [[PubMed](#)]
48. Chen, S.; Zhang, G.; Shao, C.; Huang, Q.; Liu, G.; Zhang, P.; Song, W.; An, N.; Chalopin, D.; Volff, J.N.; et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* **2014**, *46*, 253–260. [[CrossRef](#)] [[PubMed](#)]
49. Venkatesh, B.; Lee, A.P.; Ravi, V.; Maurya, A.K.; Lian, M.M.; Swann, J.B.; Ohta, Y.; Flajnik, M.F.; Sutoh, Y.; Kasahara, M.; et al. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **2014**, *505*, 174–179. [[CrossRef](#)] [[PubMed](#)]
50. Wu, C.; Zhang, D.; Kan, M.; Lv, Z.; Zhu, A.; Su, Y.; Zhou, D.; Zhang, J.; Zhang, Z.; Xu, M.; et al. The draft genome of the large yellow croaker reveals well-developed innate immunity. *Nat. Commun.* **2014**, *5*, 1–7. [[CrossRef](#)] [[PubMed](#)]
51. Wang, S.; Zeng, X.; Yang, Q.; Qiao, S. Antimicrobial Peptides as Potential Alternatives to Antibiotics in Food Animal Industry. *Int. J. Mol. Sci.* **2016**, *17*, 603. [[CrossRef](#)] [[PubMed](#)]
52. Bian, C.; Li, J.; Lin, X.; Chen, X.; Yi, Y.; You, X.; Zhang, Y.; Lv, Y.; Shi, Q. Whole Genome Sequencing of the Blue Tilapia (*Oreochromis aureus*) Provides a Valuable Genetic Resource for Biomedical Research on Tilapias. *Mar. Drugs* **2019**, *17*, 386. [[CrossRef](#)] [[PubMed](#)]
53. Couto, M.A.; Harwig, S.S.; Cullor, J.S.; Hughes, J.P.; Lehrer, R.I. Identification of eNAP-1, an antimicrobial peptide from equine neutrophils. *Infect. Immun.* **1992**, *60*, 3065–3071. [[PubMed](#)]
54. Lai, Y.; Gallo, R.L. AMPed up immunity: How antimicrobial peptides have multiple roles in immune defense. *Trends Immunol.* **2011**, *30*, 131–141. [[CrossRef](#)] [[PubMed](#)]
55. Hilchie, A.L.; Wuerth, K.; Hancock, R.E. Immune modulation by multifaceted cationic host. *Nat. Chem. Biol.* **2013**, *9*, 761–768. [[CrossRef](#)] [[PubMed](#)]
56. Wagener, J.; Schneider, J.J.; Baxmann, S.; Kalbacher, H.; Borelli, C.; Nuding, S.; Kuchler, R.; Wehkamp, J.; Kaeser, M.D.; Mailänder-Sanchez, D.; et al. A peptide derived from the highly conserved protein gapdh is involved in tissue protection by different antifungal strategies and epithelial immunomodulation. *J. Investig. Dermatol.* **2013**, *133*, 144–153. [[CrossRef](#)] [[PubMed](#)]
57. Branco, P.; Francisco, D.; Chambon, C.; Hébraud, M.; Arneborg, N.; Almeida, M.G.; Caldeira, J.; Albergaria, H. Identification of novel GAPDH-derived antimicrobial peptides secreted by *Saccharomyces cerevisiae* and involved in wine microbial interactions. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 843–853. [[CrossRef](#)]
58. Thompson, R.; Munro, J.L. Aspects of the biology and ecology of Caribbean reef fishes: Serranidae (hinds and groupers). *J. Fish Biol.* **1978**, *12*, 115–146. [[CrossRef](#)]
59. Rivals, E.; Salmela, L. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* **2014**, *30*, 3506–3514.
60. Fujiyama, A.; Harada, M.; Okuno, M.; Toyoda, A.; Maruyama, H.; Kajitani, R.; Kohara, Y.; Toshimoto, K.; Noguchi, H.; Itoh, T.; et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **2014**, *24*, 1384–1395.
61. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]
62. Van Ooijen, J. *Software for the Calculation of Genetic Linkage Maps in Experimental Populations*; Kyazma BV: Wageningen, The Netherlands, 2004.
63. Zhang, S.; Zhang, X.; Chen, X.; Xu, T.; Wang, M.; Qin, Q.; Zhong, L.; Jiang, H.; Zhu, X.; Liu, H.; et al. Construction of a High-Density Linkage Map and QTL Fine Mapping for Growth- and Sex-Related Traits in Channel Catfish (*Ictalurus punctatus*). *Front. Genet.* **2019**, *10*, 251. [[CrossRef](#)] [[PubMed](#)]
64. Zhao, X.; Hao, W. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **2007**, *35*, W265–W268.

65. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* **2009**, *25*, 4–10.
66. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: *Ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* **2006**, *34*, W435–W439. [[CrossRef](#)] [[PubMed](#)]
67. Burge, C.; Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **1997**, *268*, 78–94. [[CrossRef](#)] [[PubMed](#)]
68. Altschul, S.F.; Gish, W.; Miller, W.; Meyers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 8. [[CrossRef](#)]
69. Birney, E.; Clamp, M.; Durbin, R. GeneWise and Genomewise. *Genome Res.* **2004**, *14*, 988–995. [[CrossRef](#)] [[PubMed](#)]
70. Trapnell, C.; Pachter, L.; Salzberg, S.L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **2009**, *25*, 1105–1111. [[CrossRef](#)]
71. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)]
72. Lewis, S.; Searle, S.; Harris, N.; Gibson, M.; Lyer, V.; Richter, J.; Wiel, C.; Bayraktaroglu, L.; Birney, E.; Crosby, M.; et al. Creating a honey bee consensus gene set. *Genome Biol.* **2002**, *3*, R13. [[CrossRef](#)]
73. Bateman, A.; Mitchell, A.; Bairoch, A.; Quinn, A.F.; Laugraud, A.; Wu, C.H.; Sigrist, C.J.A.; Orengo, C.; Yeats, C.; McAnulla, C.; et al. InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2009**, *37*, D211–D215.
74. Kanehisa, M.; Goto, S. Kanehisa Laboratories Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2018**, *28*, 27–30.
75. Bairoch, A. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **2002**, *28*, 45–48. [[CrossRef](#)] [[PubMed](#)]
76. Chen, Y.; Chen, Y.; Shi, C.; Huang, Z.; Zhang, Y.; Li, S.; Li, Y.; Ye, J.; Yu, C.; Li, Z.; et al. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **2018**, *7*, 1–6. [[CrossRef](#)] [[PubMed](#)]
77. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)]
78. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).