Article

Whole Genome Sequencing of the Blue Tilapia (*Oreochromis aureus*) Provides a Valuable Genetic Resource for Biomedical Research on Tilapias

Chao Bian [1,2,†], Jia Li [2,†], Xueqiang Lin [3,†], Xiyang Chen [2,4], Yunhai Yi [2,4], Xinxin You [2,4], Yiping Zhang [2,4], Yunyun Lv [2,4] and Qiong Shi [2,4,*]

1   Center of Reproduction, Development and Aging, Faculty of Health Sciences, University of Macau, Taipa, Macau 999078, China
2   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
3   BGI Marine-Hainan, BGI Marine, BGI, Wenchang 571327, China
4   BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China
*   Correspondence: shiqiong@genomics.cn; Tel.: +86-185-6627-9826
†   These authors contributed equally to this work.

check for updates

Abstract: Blue tilapia (*Oreochromis aureus*) has been an economically important fish in Asian countries. It can grow and reproduce in both freshwater and brackish water conditions, whereas it is also considered as a significant invasive species around the world. This species has been widely used as the hybridization parent(s) for tilapia breeding with a major aim to produce novel strains. However, available genomic resources are still limited for this important tilapia species. Here, we for the first time sequenced and assembled a draft genome for a seawater cultured blue tilapia (0.92 Gb), with 97.8% completeness and a scaffold N50 of 1.1 Mb, which suggests a relatively high quality of this genome assembly. We also predicted 23,117 protein-coding genes in the blue tilapia genome. Comparisons of predicted antimicrobial peptides between the blue tilapia and its close relative Nile tilapia proved that these immunological genes are highly similar with a genome-wide scattering distribution. As a valuable genetic resource, our blue tilapia genome assembly will benefit for biomedical researches and practical molecular breeding for high resistance to various diseases, which have been a critical problem in the aquaculture of tilapias.

Keywords: blue tilapia (*Oreochromis aureus*); whole genome sequencing; genome assembly; genome annotation; antimicrobial peptide

1. Introduction

Tilapias are world famous for their high yields, rapid growth rates, and powerful adaptivity to various environments. They were cultivated by human beings 2500 years ago, and now they have become the second most important aquaculture fish globally [1]. Tilapias have also been spread to many regions beyond their native ranges with a surprised worldwide distribution [2]. By far, they have been mainly classified into five genera, including Sarotherodon, Oreochromis, Tilapia, Tristromella and Danakilia [2].

Evolved from marine ancestors, most tilapia fishes are able to tolerate low-salinity of seawater [3]. Blue tilapia (*Oreochromis aureus*), with a blue skin profile, is native to the Northern and Western Africa and the Middle East [4]. It was introduced to the oasis of the Jordan River as well as to warm water areas of South and Central America and South East Asia. The blue tilapia has become widespread aquatic communities in both marine and estuarine waters [1,5], partially due to its rapid growth rates,

omnivorous feeding, as well as a relatively cold patience. For shortage of freshwater in many countries, blue tilapia has been gradually cultivated in brackish and sea waters [5]. With the advantages in feeding strategies, blue tilapia may modify interactions between introduced and native species [6]. It therefore becomes an invasive species in many countries (such as the USA and Mexico), where it has triggered remarkable changes in fish community structure of local waters [2].
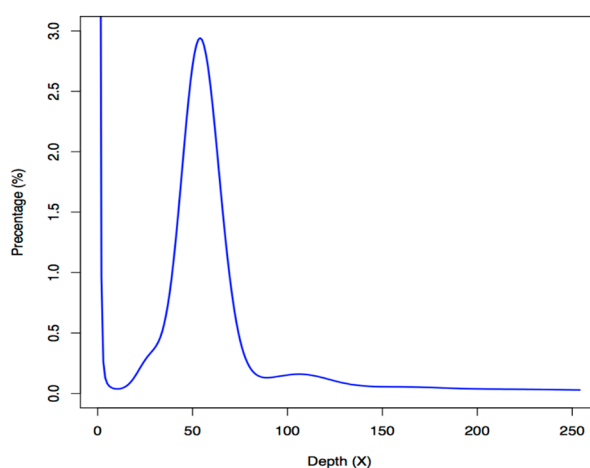
Studies on the genetic and molecular basis of sex determination in tilapias have been carried out for over 50 years. The hypothesized sex chromosome systems for tilapia species, such as XX-XY system for Nile tilapia and WZ-ZZ for blue tilapia, were reported over a half century ago [7]. The primary evidences of these hypotheses were obtained from analysis of sex-ratio of progeny from various experiments, such as inter-specific crosses [8], intra-specific crosses using sex-reversed individuals [9], chromosome set manipulations through gynogenesis [10] and androgenesis [11]. However, the sex of tilapias is controlled by an integration of genetic determination and environmental temperatures, although the details are not clearly determined yet [12,13]. The sex determination mechanisms among closely related tilapia species are diverse, and their ability to mate and produce fertile hybrids further complicates the elucidation of these sex determination systems [14,15].

Although the blue tilapia is a widely cultured tilapia species, and it is also commonly utilized in breeding for production of monosex tilapias, its genome resources are still limited. Meanwhile, more and more diseases have been developing in aquaculture areas, which requires more precise genetic supports to maintain high quality of tilapias. Therefore, we sequenced and assembled a draft genome of the blue tilapia for the first time, and we subsequently performed a series of genomic analyses related to biology and immunology (such as antimicrobial peptides (AMPs)) of the blue tilapia and genetic comparisons with its close relative, the Nile tilapia (*O. niloticus*).

2. Results

2.1. Statistics of Genome Assembly and Annotation

A total of 239.89 gigabases (Gb) of Illumina raw reads were sequenced; after removal of low-quality reads, adapter sequences and PCR-duplicates, we obtained 161.53 Gb of clean data for subsequent genome assembly (see more details in Table S1). We estimated the genome size of blue tilapia to be approximately 1.02 Gb using the routine K-mer approach [16] (Figure 1), which is slightly larger than that of the Nile tilapia (about 0.90 Gb; NCBI release date: 2016/10/31).

Figure 1. A k-mer analysis of the blue tilapia genome. The *x*-axis is the sequencing depth of each unique 19-mer, and the *y*-axis is the percentage of these unique 19-mers. The peak depth (K_depth) is at 54, and the corresponding k-mer number (N) is 55,102,309,616. We therefore calculated the genome size (G) to be ~1.02 Gb based on the following formula: G = N/K_depth [16].
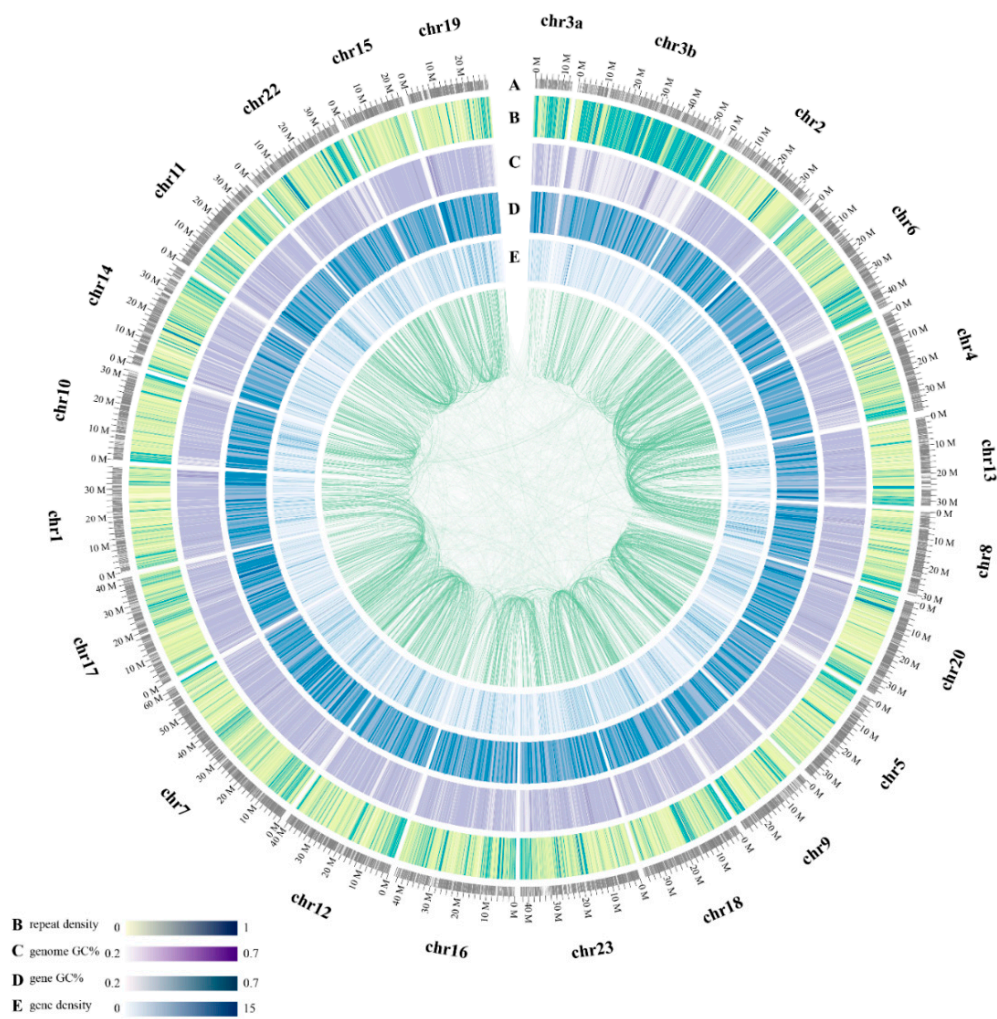
De novo assembling of the blue tilapia genome was performed using SOAPdenovo2 software [17]. A total of 53,082 scaffolds with a N50 value of 1.1 Mb, and 106,865 contigs with a N50 value of 53.2 kb, were assembled (Table 1 and Table S2). BUSCO [18] was used to estimate the completeness of our blue tilapia assembly. We determined that 4482 conserved vertebrate genes were covered, representing 97.8% completeness (in the total of 4584 genes).

Table 1. Statistics of the genome assembly and annotation of both blue and Nile tilapias.

Parameter	Blue Tilapia	Nile Tilapia [15]
Genome Assembly		
Contig N50 size (kb)	53.2	3.11
Scaffold N50 size (Mb)	1.10	-
Estimated genome size (Gb)	1.02	1.20
Assembled genome size (Gb)	0.92	1.01
Genome annotation		
Protein-coding gene number	23,117	29,249
Annotated functional gene number	22,573	-
Unannotated functional gene number	544	-

Approximately 234.40 Mb of repeat sequences, accounting for about 25.35% of the genome assembly, were predicted in the blue tilapia genome (Table S3). They included 101.27 Mb of DNA transposons, 114.47 Mb of long interspersed nuclear elements (LINEs), 11.60 Mb of short interspersed nuclear elements (LINEs), and 50.65 Mb of long terminal repeats (LTRs). We also performed a detailed comparison of repeat sequences between the Nile and blue tilapia genomes, and found that the hAT and L2 types of repeat sequences in the blue tilapia were remarkably longer than those in the Nile tilapia. On the other hand, unknown repeats in the Nile tilapia were about 8 folds as long as those in the blue tilapia, possibly due to the higher completeness of the Nile tilapia genome assembly with assistance of PacBio data (see more details about the comparisons in Table S3). A total of 23,117 protein-coding genes were predicted in the blue tilapia genome (Table S4), of which 22,573 genes can be annotated at least one function from four popular public databases, including Swiss-Prot [19], TrEMBL [19], Interpro [20], and KEGG [21] (Table S5).
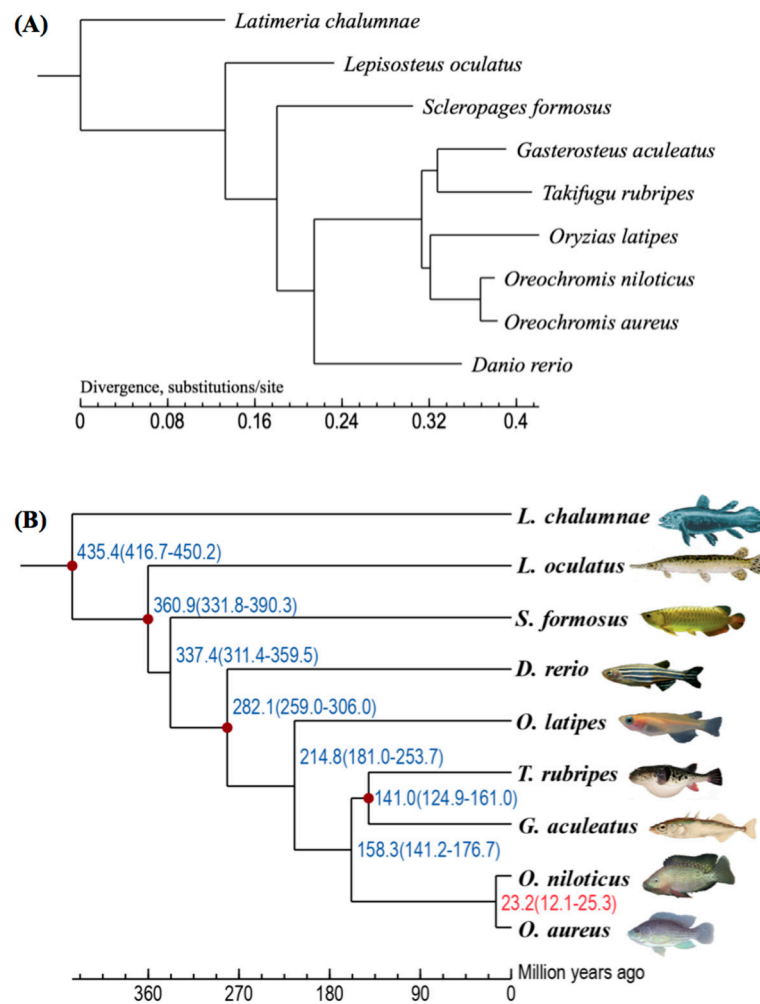
We also constructed pseudochromosomes (Chrs) for the blue tilapia genome, with assistance of the information of one-to-one syntenic blocks between the blue tilapia and the Nile tilapia (NCBI release date: 2016/10/31) [15]. A total of 91.3% (0.84 Gb/0.92 Gb) scaffolds from the latter were assigned onto 22 Chrs of the blue tilapia. Detailed distributions of gene density, GC content, repeat sequence content of each Chr, and the inner-chromosome syntenic blocks were summarized in Figure 2.

Figure 2. A circos view of the blue tilapia genome. From outside to the inside rings: (**A**) chromosome length (Mb) and numbers, (**B**) distribution of repeat density in 100 kb non-overlapping windows, (**C**) distribution of genome GC content, (**D**) distribution of gene GC content, and (**E**) distribution of gene density. Syntenic blocks are connected with green lines, and each line indicates one pair of paralog genes in the blue tilapia genome.

2.2. Summary of Gene Clustering and Phylogeny

We downloaded the protein sets of eight teleost species, including zebrafish (*Danio rerio*), Nile tilapia, three-spined stickleback (*Gasterosteus aculeatus*), Japanese puffer (*Takafugu rubripes*), medaka (*Oryzias latipes*), Asian arowana (*Scleropages formosus*), spotted gar (*Lepisosteus oculatus*) and coelacanth (*Latimeria chalumnae*) from the Ensembl database. A total of 173,955 proteins were collected from the blue tilapia and the above-mentioned eight fish species for building gene families. A Markov Chain Clustering (MCL) in the OrthoMCL software [22] with default parameters was utilized to identify gene families. All the 173,955 proteins were categorized into 18,096 gene families, of which only a family contained 9 proteins from above indicated 9 fish species (i.e., only one protein from each species in this family) was selected as single-copy gene family. We then extracted 3751 one-to-one single-copy gene families to construct a phylogenetic tree (Figure 3A). It was estimated that the divergence time between the blue tilapia and the Nile tilapia was about 23.2 million years ago (Figure 3B).

Figure 3. A phylogenetic tree of nine examined fish species. (**A**) The phylogenetic position of the blue tilapia was determined on the basis of one-to-one orthologues from the nine fish species. (**B**) The divergence times were predicted with references (red dots) from the TimeTree (http://www.timetree.org/).
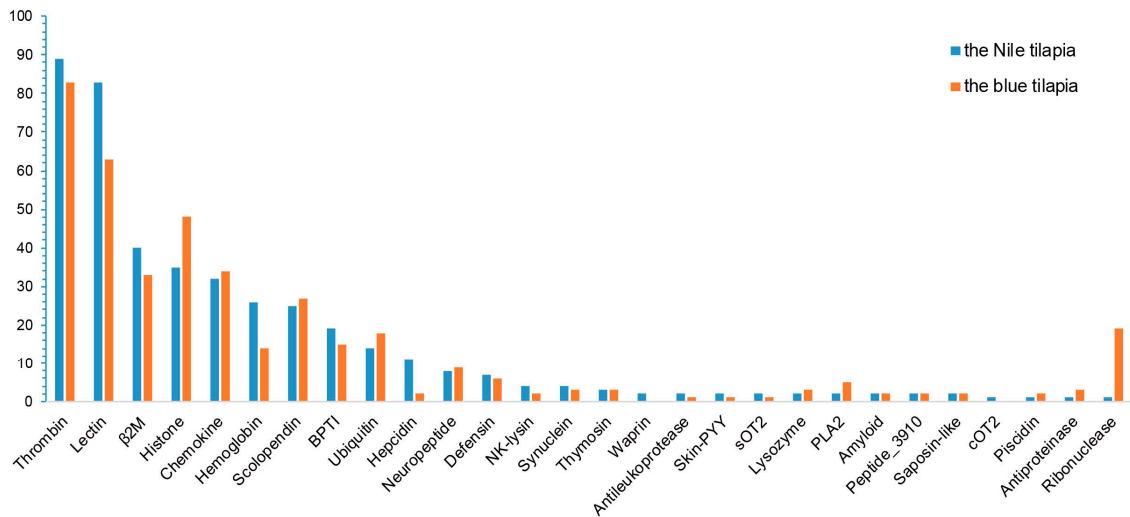
2.3. Whole-Genome Chromosomal Evolution

Numbers of orthologous genes in the blocks between the blue tilapia and pre-3-round whole genome duplicated (pre-3R WGD) species were relatively low, such as 11,221 with human, 9966 with chicken, and 12,157 with spotted gar. However, the numbers were increased when compared with the 3R WGD fishes, such as 13,600 with zebrafish, 12,755 with medaka, 15,033 with half-smooth sole, and 17,532 with Nile tilapia (the largest number due to the closest relationship). Finally, we inferred a chromosome model of teleost ancestor using conserved syntenic blocks from the human genome based on the method of a previous report [23].

Detailed evolutionary relationships of chromosomal blocks from an ancestral vertebrate genome to representative fish genomes were deduced and provided in Figure S1. It seems that the chromosomal evolution of fishes is differentially complicated, involving various chromosomal losses, translocations, fissions and fusions, and fragmental or whole-genome duplications. Please read a detailed discussion in our previous report of arowana genomes [24].

2.4. Antimicrobial Peptides in Both the Blue and Nile Tilapias

For high-throughput identification of antimicrobial peptides in the two tilapia species, we collected available active AMPs as a local reference database (Table S6) and employed BLAST to search against the annotated gene sets of both tilapias. A total of 407 putative AMP genes were identified from the blue tilapia, covering 32 classes; while 428 putative AMP genes were identified from the Nile tilapia with a division of 34 classes (Figure 4).

Figure 4. Statistics of different AMPs from the Nile tilapia and the blue tilapia. Those classes with only one AMP in both tilapias were not shown, such as Amylin, Ap-s, CcAMP1, GAPDH, hGlyrichin and LEAP-2.

After a KEGG clustering analysis, we predicted that the 407 putative AMP genes in the blue tilapia were enriched to 198 pathways, with representative relations with "immune disease", "immune system", and "infectious disease: Viral and signaling molecules and interaction" (Figure 5a). Similarly, the putative AMP genes in the Nile tilapia were clustered into 161 pathways, with the major classes related to immune and diseases (similar to the blue tilapia; Figure 5b).
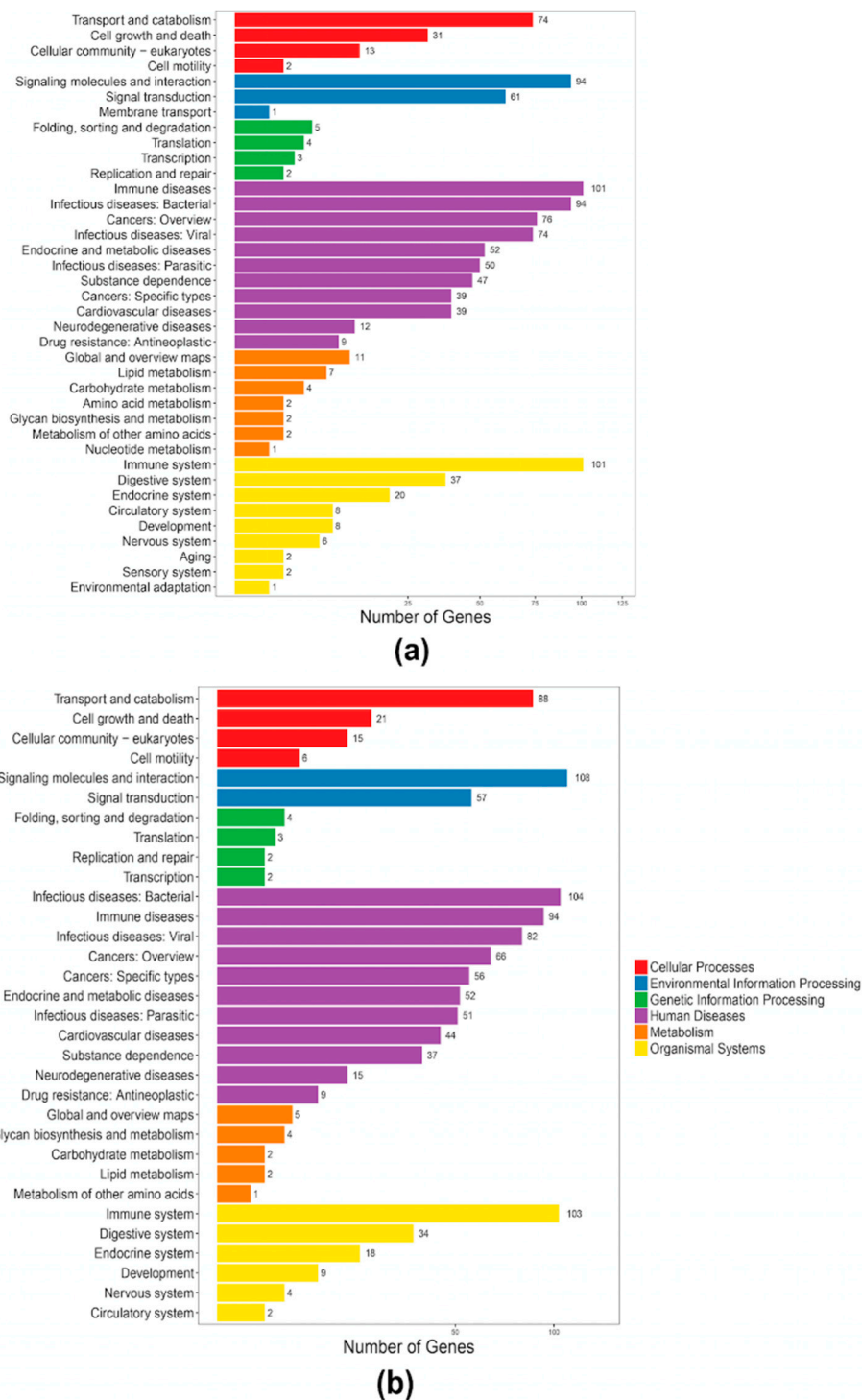
Figure 5. KEGG annotation of the putative AMP genes in the blue tilapia (**a**) and the Nile tilapia (**b**).

3. Discussion

3.1. High-Throughput Screening of AMPs from Our High-Quality Genome Assembly

Due to the high completeness and the long scaffold N50 (Section 2.1), our blue tilapia genome assembly is of high quality. As we reported previously [25], the genome-derived gene set has been valuable for a high-throughput screening of AMPs. Over 400 putative AMP sequences were identified for both the blue and the Nile tilapias (Section 2.4), which provides a genetic resource for comparisons of immunology between the two closed tilapia species.

3.2. Comparisons of AMPs between the Blue and Nile Tilapias

A previous work [26] compared the resistance of the Nile tilapia and the blue tilapia to the diseases caused by *Aeromonas sobria*, a pathogenic bacterium that has produced large losses in tilapia aquaculture. Related data demonstrated that the Nile tilapia has a higher resistance to *A. sobria*-related diseases than the blue tilapia. In our present study, we observed that the Nile tilapia has more putative AMP genes and two extra classes than the blue tilapia (Figure 5), named Waprin and cOT2. Waprin (query ID 1589 in Table S6) has been reported to present an antimicrobial activity against Gram-positive bacteria [27], and previous works have proved that cOT2 (query ID 2797 in Table S6) could cause morphological changes to bacterial cells [28].

As shown in Figure 6, the numbers of lectin, hemoglobin and hepcidin in the Nile tilapia were remarkably more than those in the blue tilapia. Lectin has antimicrobial and antiparasitic activities [29]; in addition, previous works have proved that MCL-4, a novel isoform of lectin from Manila clam (*Ruditapes philippinarum*), facilitated the phagocytic ability of hemocytes for *Vibrio tubiashii* and suppressed the growth of *Alteromonas haloplanktis* [30]. Another work has demonstrated that HcLec4, a lectin with 4 carbohydrate recognition domains from *Hyriopsis cumingii*, up-regulated expression of AMPs at the early stage of bacterial infection [31]. The number of hemoglobin in the Nile tilapia is approximately twice as many as that in the blue tilapia, and hemoglobin was proven to have antiparasitic and antimicrobial activities [32,33].
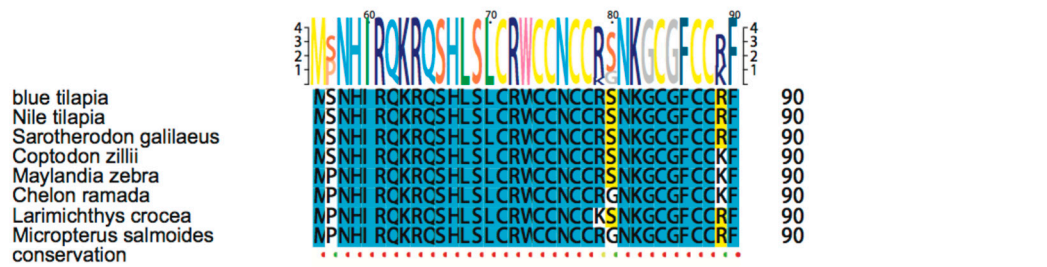
Hepcidin, one of the most important and common AMPs in fishes, also showed remarkable differences between the Nile tilapia and the blue tilapia. There existed 11 hepcidins (query IDs 1701 and 809 in Table S6) in the Nile tilapia, while the blue tilapia only had two hepcidins (query ID 1701). A Swiss-Prot annotation provided a strong evidence that 12 out of the 13 putative AMP genes were predicted to be hepcidin. APD1701, a novel hepcidin from Orange-spotted grouper (*Epinephelus coioides*), was proven to have antimicrobial activities against *Vibrio vulnificus* and *Staphylococcus aureus* [34]. APD809 is a cDNA sequence of hepcidin-like AMPs in Mozambique tilapia (*Oreochromis mossambicus*), whose synthetic peptide was active against gram-positive bacteria, such as *Listeria monocytogenes*, *Enterococcus faecium* and *Staphylococcus aureus* [35]. To validate the identification of putative hepcidin genes, we preformed multiple sequence alignment. As shown in Figure 6, representative putative and known hepcidin genes exhibited a high similarity. Interestingly, we found that these putative hepcidin genes in both the Nile and the blue tilapias can also be divided into two categories, hepcidin-1 (Figure 6a) and hepcidin-2 (Figure 6b) based on the sequence similarity to those known hepcidin sequences from other fish species.

Although it seems that the Nile tilapia presents more antimicrobial activities, the practical aquaculture of the blue tilapia in Asian countries, especially in Southern China, needs more hybridization strains from the blue tilapia due to its high tolerance to cold temperature and high salinity [1].

Figure 6. Multiple sequence alignment of putative hepcidin genes in fishes. (**a**) hepcidin-1; (**b**) hepcidin-2. Yellow and blue marks represent identity > 50% and > 80%, respectively.

4. Materials and Methods

4.1. Sample Preparation and Sequencing

A female blue tilapia was collected from a local pond (water salinity of 5~8‰) of the BGI-Marine tilapia aquaculture base in Fengpo Town, Wengchang City, Hainan Province, China. The aquaculture water was a mixture of local rain and seawater, since we have built a water gate on the beach to collect seawater when tide rises. Genomic DNA from muscle tissue was extracted using Qiagen GenomicTip100 (Qiagen, Germantown, MD, USA). All animal experiments were performed in accordance with the

guidelines of the Animal Ethics Committee and were approved by the Institutional Review Board on Bioethics and Biosafety of BGI (approval ID: FT18134).

The isolated genomic DNA was subsequently applied to construct three short-insert libraries (250, 500 and 800 bp) and four long-insert libraries (2, 5, 10 and 20 kb) with the standard protocol provided by Illumina (San Diego, CA, USA). The paired-end sequencing for 125-bp reads with a routine whole genome shotgun sequencing strategy was performed on an Illumina HiSeq 2500 platform as previously reported [36,37]. We further trimmed 5 bases in both ends of the raw reads, discarded those duplicated reads, and removed reads with 10 or more Ns and low-quality bases to improve the quality of sequenced reads [24].

4.2. Estimation of Genome Size

The sequenced k-mers were confirmed to be at a Poisson distribution [38]. Therefore, we calculated the genome size of the blue tilapia by employing the following equation: G = k-mer_number/k-mer_depth [16]. In this equation, the G represents the estimated genome size, the k-mer_number stands for the total number of k-mers, and the k-mer_depth is the core peak of k-mer accumulation.

4.3. Genome Assembly and Annotation

We employed SOAPdenovo2 (version 2.04.4) software [17] with core parameters (pregraph −K 27 −d 1; scaff −F −b 1.5 −p 16) to construct contigs and original scaffolds by using clean reads. We then employed the paired-end reads of long-insert libraries (2, 5, 10 and 20 kb) to align onto the contigs for building scaffolds. Gaps in scaffolds were filled up with the paired-end reads of three short-inset libraries (250, 500 and 800 bp) using the GapCloser software (v1.12- r6, default parameters and −p set to 25). Raw reads and the genome assembly have been deposited in the NCBI under the project ID PRJNA539829.

Repeat sequences in the blue tilapia assembly were predicted by an integration of three routine approaches, including *de novo*, homology and tandem repeat predictions [36]. For the *de novo* prediction, RepeatModeller v1.04 (Institute for Systems Biology, Seattle, WA, USA) and LTR_FINDER v1.0.6 [39] were utilized to construct a repeat reference library. The genome sequences were then mapped onto the reference library to predict the *de novo* repeat sequences using RepeatMasker v3.2.9 [40]. For the homology annotation, our genome sequences were mapped onto the RepBase v21.01 database [41] using RepeatMasker v4.06 and RepeatProteinMask v4.06. The tandem repeats were subsequently predicted using Tandem Repeat Finder [42] (version 4.04). These repeat data from above three approaches were integrated to generate a non-redundant repeat set.

Three combined approaches were used to annotate the gene set of the blue tilapia, including *de novo*, homology and transcriptome-based annotations. At first, we masked the repeat sequencing in the assembled genome as "N". For the *de novo* annotation, the AUGUSTUS v2.5 [43] and GENSCAN v1.0 [44] were employed to annotate gene models from the repeat masked genome. For the homology annotation, protein sequences of zebrafish, Japanese puffer, green spotted puffer (*Tetraodon nigroviridis*), Nile tilapia and three-spined stickleback were downloaded from the Ensembl database (release 75). These sequences were aligned onto the blue tilapia assembly to generate alignments using TblastN [45] with an e-value $< 1.0 \times 10^{-5}$. Subsequently, GeneWise v2.2.0 [46] was employed to predict the potential gene structures on these alignments. For the transcriptome-based annotation, we employed Tophat v2.1.1 [47] to align muscle transcriptome reads onto the blue tilapia genome to obtain alignments, and then Cufflink v2.2.1 [48] was utilized to predict the potential gene structures on these alignments. Finally, we applied GLEAN [49] to generate the integrated results from the three approaches into a final gene set. This gene set were searched against four public functional databases, including Swiss-Prot [19], TrEMBL [19], Interpro [20], and KEGG [21], to predict potential functions of each gene using BLASTp [45].

4.4. Constructions of the Phylogenetic and Divergence Time Trees

The protein sequences of each single-copy gene family were aligned each other using MUSCLE (v. 3.8.31) [50] with default parameters. The protein alignments were then converted to their corresponding coding sequences using an in-house Perl script. These nucleotide sequences were linked into a continuous sequence for each species. Nondegenerated sites, obtained from the continuous sequence of each species, were then joined into a new sequence of each species to build a phylogenetic tree using MrBayes [51] (Version 3.2, with the GTR + gamma model). The Mcmctree software in the PAML package [52] was employed to estimate divergence times among the blue tilapia and eight other fish species.

4.5. Chromosomal Localization of the Blue Tilapia Sequences

Based on the genomic conservation between the blue tilapia and the Nile tilapia, we used the Nile tilapia genome as the reference to assemble the blue tilapia pseudo-chromosomes. Firstly, we downloaded the newest released version of chromosome data of the Nile tilapia (NCBI release date: 2016/10/31). The assembled scaffolds of the blue tilapia was aligned to the chromosomal sequences of the Nile tilapia using the Blastz program [53] with optimized parameters of "T = 2 C = 2 H = 2000 Y = 3400 L = 6000 K = 2200". Finally, we chose the best hits of syntenic blocks with local Perl scripts.

4.6. Reconstruction of the Ancestral Genome for Examination of Whole-Genome Chromosomal Evolution

At first, we downloaded the protein sequences of seven vertebrate species (including human, chicken, spotted gar, zebrafish, and medaka) from Ensemble (release 87), and those of the half-smooth tongue sole and Nile tilapia from the NCBI genome database. We then conducted protein alignments between the blue tilapia and other species by performing BLASTP with an e-value $< 1.0 \times 10^{-5}$ to find conserved gene-level syntenic blocks.

4.7. High-Throughput Identification of Antimicrobial Peptides

Query sequences were collected from the online Antimicrobial Peptides Database (APD3) [25]. The subject sequences were the annotated gene sets of the Nile tilapia and the blue tilapia. We built an index for the subject sequences by using makeblastdb, and the identification step was performed by TBLASTN (e-value: 1.0×10^{-5}). The alignment hits with aligned ratio less than 0.5 were filtered out, and those redundant results were also removed. Classification was referred to the detail information from the APD3 database. Furthermore, known hepcidin protein sequences were downloaded from NCBI, and multiple sequence alignment of hepcidin was performed by BioEdit [54]. The alignment results of hepcidin were further analyzed and visualized by TEXshade [55] (Version: 2.9.2).

5. Conclusions

In summary, we for the first time provide a valuable genome assembly of the blue tilapia (0.92 Gb), with 97% completeness and 23,117 annotated protein-coding genes. The divergence time between the blue and the Nile tilapias was predicted to be 23.2 million years ago. Comparisons of antimicrobial peptides between the two tilapia species demonstrated that these AMP genes are remarkably genome-wide scattered. Given that the blue tilapia has an important economic value, its genome resource will build a valuable platform for further biomedical research and practical molecular breeding of tilapias.

Supplementary Materials: The following are available online at http://www.mdpi.com/1660-3397/17/7/386/s1. Table S1: Statistics of the clean reads. Table S2: Statistics of the blue tilapia genome assembly. Table S3: Statistics of repeat sequences in blue and Nile tilapia genomes. Table S4: Gene annotation of the assembled blue tilapia genome. Table S5: Functional assignments of the final gene set 'of the blue tilapia genome. Table S6: The collected reference AMP sequences. Figure S1: A schematic chart for the composition of ancestral chromosomes among various species.

Abbreviations

AMP	antimicrobial peptide
BUSCO	Benchmarking Universal Single-copy Orthologs
Gb	gigabase
KEGG	Kyoto Encyclopedia of Genes and Genomes
N50	50% of the genome is in fragments of this length or longer
WGD	whole genome duplication

References

1. Gupta, M.V.; Acosta, B.O. A review of global tilapia farming practices. *Aquac. Asia* **2004**, *9*, 7–16.
2. Chapman, F.A. Culture of Hybrid Tilapia: A Reference Profile. University of Florida IFAS Extension. Available online: http://edis.ifas.ufl.edu/pdffiles/FA/FA01200.pdf (accessed on 29 April 2019).
3. Kirk, R.G. A review of recent developments in *tilapia* culture, with special reference to fish farming in the heated effluents of power stations. *Aquaculture* **1972**, *1*, 45–60. [CrossRef]
4. Schramm, H.L.; Zale, A.V. Effects of cover and prey size on preferences of juvenile largemouth bass for blue tilapias and bluegills in tanks. *T. Am. Fish. Soc.* **1985**, *114*, 725–731. [CrossRef]
5. Suresh, A.V.; Lin, C.K. Tilapia culture in saline waters: A review. *Aquaculture* **1992**, *106*, 201–226. [CrossRef]
6. Peterson, M.S.; Slack, W.T.; Waggy, G.L.; Finley, J.; Woodley, C.M.; Partyka, M.L. Foraging in non-native environments: Comparison of Nile tilapia and three co-occurring native centrarchids in invaded coastal Mississippi watersheds. *Environ. Biol. Fish* **2006**, *76*, 283–301. [CrossRef]
7. Hickling, C.F. The Malacca tilapia hybrid. *J. Genet.* **1960**, *57*, 1–10. [CrossRef]
8. Pruginin, Y.; Rothbard, S.; Wohlfarth, G.; Halevy, A.; Moav, R.; Hulata, G. All-male broods of Tilapia nilotica × T. aurea hybrids. *Aquaculture* **1975**, *6*, 11–21. [CrossRef]
9. Mair, G.C.; Scott, A.G.; Penman, D.J.; Beardmore, J.A.; Skibinski, D.O. Sex determination in the genus *Oreochromis*: 1. Sex reversal, gynogenesis and triploidy in *O. niloticus* (L.). *Theor. Appl. Genet.* **1991**, *82*, 144–152. [CrossRef] [PubMed]
10. Müller-Belecke, A.; Hörstgen-Schwark, G. Sex determination in tilapia (*Oreochromis niloticus*) sex ratios in homozygous gynogenetic progeny and their offspring. *Aquaculture* **1995**, *137*, 57–65.
11. Myers, J.M.; Penman, D.J.; Basavaraju, Y.; Powell, S.F.; Baoprasertkul, P.; Rana, K.J.; Bromage, N.; McAndrew, B.J. Induction of diploid androgenetic and mitotic gynogenetic Nile tilapia (*Oreochromis niloticus* L.). *Theor. Appl. Genet.* **1995**, *90*, 205–210. [CrossRef]
12. Baroiller, J.F.; Chourrout, D.; Fostier, A.; Jalabert, B. Temperature and sex chromosomes govern sex ratios of the mouthbrooding Cichlid fish *Oreochromis niloticus*. *J. Exp. Zool.* **1995**, *273*, 216–223. [CrossRef]
13. Baroiller, J.F.; D'Cotta, H.; Bezault, E.; Wessels, S.; Hoerstgen-Schwark, G. Tilapia sex determination: Where temperature and genetics meet. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **2009**, *153*, 30–38. [CrossRef] [PubMed]
14. Cnaani, A.; Lee, B.Y.; Zilberman, N.; Ozouf-Costaz, C.; Hulata, G.; Ron, M.; D'Hont, A.; Baroiller, J.F.; D'Cotta, H.; Penman, D.J.; et al. Genetics of sex determination in tilapiine species. *Sex. Devel.* **2008**, *2*, 43–54. [CrossRef] [PubMed]
15. Conte, M.A.; Gammerdinger, W.J.; Bartie, K.L.; Penman, D.J.; Kocher, T.D. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genom.* **2017**, *18*, 341. [CrossRef] [PubMed]
16. Li, R.; Fan, W.; Tian, G.; Zhu, H.; He, L.; Cai, J.; Huang, Q.; Cai, Q.; Li, B.; Bai, Y.; et al. The sequence and *de novo* assembly of the giant panda genome. *Nature* **2010**, *463*, 311–317. [CrossRef] [PubMed]

17. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. Erratum: SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **2015**, *4*, 30. [CrossRef] [PubMed]
18. Simao, 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] [PubMed]
19. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370. [CrossRef] [PubMed]
20. Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L. InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2009**, *37*, D211–D215. [CrossRef] [PubMed]
21. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]
22. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [CrossRef] [PubMed]
23. Jaillon, O.; Aury, J.M.; Brunet, F.; Petit, J.L.; Stange-Thomann, N.; Mauceli, E.; Bouneau, L.; Fischer, C.; Ozouf-Costaz, C.; Bernot, A.; et al. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **2004**, *431*, 946–957. [CrossRef] [PubMed]
24. Bian, C.; Hu, Y.; Ravi, V.; Kuznetsova, I.S.; Shen, X.; Mu, X.; Sun, Y.; You, X.; Li, J.; Li, X.; et al. The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Sci. Rep.* **2016**, *6*, 24501. [CrossRef] [PubMed]
25. 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]
26. Cai, W.-Q.; Li, S.-F.; Ma, J.-Y. Diseases resistance of Nile tilapia (*Oreochromis niloticus*), blue tilapia (*Oreochromis aureus*) and their hybrid (female Nile tilapia×male blue tilapia) to *Aeromonas sobria*. *Aquaculture* **2004**, *229*, 79–87. [CrossRef]
27. Nair, D.G.; Fry, B.G.; Alewood, P.; Kumar, P.P.; Kini, R.M. Antimicrobial activity of omwaprin, a new member of the waprin family of snake venom proteins. *Biochem. J.* **2007**, *402*, 93–104. [CrossRef] [PubMed]
28. Prajanban, B.O.; Jangpromma, N.; Araki, T.; Klaynongsruang, S. Antimicrobial effects of novel peptides cOT2 and sOT2 derived from Crocodylus siamensis and Pelodiscus sinensis ovotransferrins. *Biochim. Biophys. Acta Biomembr.* **2017**, *1859*, 860–869. [CrossRef]
29. Iordache, F.; Ionita, M.; Mitrea, L.I.; Fafaneata, C.; Pop, A. Antimicrobial and antiparasitic activity of lectins. *Curr. Pharm. Biotechnol.* **2015**, *16*, 152–161. [CrossRef]
30. Takahashi, K.G.; Kuroda, T.; Muroga, K. Purification and antibacterial characterization of a novel isoform of the Manila clam lectin (MCL-4) from the plasma of the Manila clam, Ruditapes philippinarum. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2008**, *150*, 45–52. [CrossRef]
31. Zhao, L.L.; Wang, Y.Q.; Dai, Y.J.; Zhao, L.J.; Qin, Q.; Lin, L.; Ren, Q.; Lan, J.F. A novel C-type lectin with four CRDs is involved in the regulation of antimicrobial peptide gene expression in *Hyriopsis cumingii*. *Fish Shellfish Immunol.* **2016**, *55*, 339–347. [CrossRef]
32. Ullal, A.J.; Noga, E.J. Antiparasitic activity of the antimicrobial peptide HbbetaP-1, a member of the beta-haemoglobin peptide family. *J. Fish Dis.* **2010**, *33*, 657–664. [CrossRef] [PubMed]
33. Seo, J.K.; Lee, M.J.; Jung, H.G.; Go, H.J.; Kim, Y.J.; Park, N.G. Antimicrobial function of SHbetaAP, a novel hemoglobin beta chain-related antimicrobial peptide, isolated from the liver of skipjack tuna, *Katsuwonus pelamis*. *Fish Shellfish Immunol.* **2014**, *37*, 173–183. [CrossRef] [PubMed]
34. 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]
35. Huang, P.H.; Chen, J.Y.; Kuo, C.M. Three different hepcidins from tilapia, *Oreochromis mossambicus*: Analysis of their expressions and biological functions. *Mol. Immunol.* **2007**, *44*, 1922–1934. [CrossRef] [PubMed]
36. Song, L.; Bian, C.; Luo, Y.; Wang, L.; You, X.; Li, J.; Qiu, Y.; Ma, X.; Zhu, Z.; Ma, L.; et al. Draft genome of the Chinese mitten crab, *Eriocheir sinensis*. *GigaScience* **2016**, *5*, 1–5. [CrossRef] [PubMed]

37. Zhang, Z.; Zhang, K.; Chen, S.; Zhang, Z.; Zhang, J.; You, X.; Bian, C.; Xu, J.; Jia, C.; Qiang, J.; et al. Draft genome of the protandrous Chinese black porgy, *Acanthopagrus schlegelii*. *Gigascience* **2018**, *7*, 1–7. [CrossRef] [PubMed]

38. Liu, B.; Shi, Y.; Yuan, J.; Hu, X.; Zhang, H.; Li, N.; Li, Z.; Chen, Y.; Mu, D.; Fan, W. Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *Quant. Biol.* **2013**, *35*, 62–67.

39. Xu, Z.; Wang, H. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **2007**, *35*, W265–W268. [CrossRef]

40. Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinform.* **2004**, *5*, 4–10.

41. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [CrossRef]

42. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef] [PubMed]

43. 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]

44. Burge, C.; Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **1997**, *268*, 78–94. [CrossRef] [PubMed]

45. McGinnis, S.; Madden, T.L. BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **2004**, *32*, W20–W25. [CrossRef] [PubMed]

46. Birney, E.; Clamp, M.; Durbin, R. GeneWise and Genomewise. *Genome Res.* **2004**, *14*, 988–995. [CrossRef] [PubMed]

47. Trapnell, C.; Pachter, L.; Salzberg, S.L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **2009**, *25*, 1105–1111. [CrossRef] [PubMed]

48. Trapnell, C.; Hendrickson, D.G.; Sauvageau, M.; Goff, L.; Rinn, J.L.; Pachter, L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **2013**, *31*, 46–53. [CrossRef] [PubMed]

49. Elsik, C.G.; Mackey, A.J.; Reese, J.T.; Milshina, N.V.; Roos, D.S.; Weinstock, G.M. Creating a honey bee consensus gene set. *Genome Biol.* **2007**, *8*, R13. [CrossRef] [PubMed]

50. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]

51. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayres, D.L.; Darling, A.; Hohna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [CrossRef]

52. Yang, Z.; Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **2006**, *23*, 212–226. [CrossRef] [PubMed]

53. Schwartz, S.; Kent, W.J.; Smit, A.; Zhang, Z.; Baertsch, R.; Hardison, R.C.; Haussler, D.; Miller, W. Human-mouse alignments with BLASTZ. *Genome Res.* **2003**, *13*, 103–107. [CrossRef] [PubMed]

54. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.

55. Beitz, E. TEXshade: Shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics* **2000**, *16*, 135–139. [CrossRef] [PubMed]