Comparative Biochemistry and Physiology, Part D 10 (2014) 44–51

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

Transcriptome analysis of the gill of *Takifugu rubripes* using Illumina sequencing for discovery of SNPs

Jun Cui [a], Hongdi Wang [a], Shikai Liu [b], Xuemei Qiu [a,*], Zhiqiang Jiang [a], Xiuli Wang [a]

[a] Key Laboratory of Mariculture & Stock Enhancement in North China's Sea, Ministry of Agriculture, Dalian Ocean University, Dalian 116023, China
[b] The Fish Molecular Genetics and Biotechnology Laboratory, Aquatic Genomics Unit, School of Fisheries, Aquaculture and Aquatic Sciences and Program of Cell and Molecular Biosciences, Auburn University, Auburn, AL 36849, USA

ARTICLE INFO

Article history:
Received 17 January 2014
Received in revised form 18 March 2014
Accepted 20 March 2014
Available online 27 March 2014

Keywords:
SNPs
Gill
RNA-Seq
Takifugu rubripes

ABSTRACT

Single nucleotide polymorphisms (SNPs) have become the marker of choice for genome-wide association studies in many species. High-throughput sequencing of RNA was developed primarily to analyze global gene expression, while it is an efficient way to discover SNPs from the expressed genes. In this study, we conducted transcriptome sequencing of the gill samples of *Takifugu rubripes* analyzed by using Illumina HiSeq 2000 platform to identify gene-associated SNPs from the transcriptome of *T. rubripes* gill. A total of 27,085,235 unique-mapped-reads from 55,061,524 raw data reads were generated. A total of 56,972 putative SNPs were discovered, which were located in 11,327 genes. 35,839 SNPs were transitions (Ts), 21,074 SNPs were transversions (Tv) and 88.1% of 56,972 SNPs were assigned to the 22 chromosomes. The average minor allele frequency (MAF) of the SNPs was 0.26. GO and KEGG pathway analyses were conducted to analyze the genes containing SNPs. Validation of selected SNPs revealed that 63.4% of SNPs (34/52) were true SNPs. RNA-Seq is a cost-effective way to discover gene-associated SNPs. In this study, a large number of SNPs were identified and these data will be useful resources for population genetic study, evolution analysis, resource assessment, genetic linkage analysis and genome-wide association studies. The results of our study can also offer some useful information as molecular makers to help select and cultivate *T. rubripes*.

1. Introduction

The fugu (*Takifugu rubripes*) is widely used as a model system, whose genome project was initiated by Sydney Brenner, Greg Elgar, Sam Aparicio and Byrappa Venkatesh (Brenner et al., 1993). Fugu genome is among the smallest vertebrate genomes and has proved to be a useful 'reference' genome for identifying genes and other functional elements such as regulatory elements in human and other vertebrate genomes, and for understanding the structure and evolution of vertebrate genomes. In the Fugu Genome Project (http://www.fugu-sg.org/), the v5 assembly of the genome comprises 7119 scaffolds covering 393 Mb. About 72% of the assembly (281,557,002 bp) is organized into 22 chromosomes. Another 14% of the assembly (56,157,414 bp) is assigned to chromosomes but the orientation and order of the scaffolds are not known. The remaining 14% of the assembly (54,661,818 bp) is concatenated into a single sequence (Kai et al., 2011). In addition to its importance as a model system, *T. rubripes* plays an important role in aquaculture. *T. rubripes*, widely distributed in Asia (China, Japan and Korea), is one of the most important aquaculture species in China.

T. rubripes is loved as a delicious food by Chinese, Japanese and Korean because of its tender flesh, delicious tasty and high abundance of protein (Wang et al., 2008, 2010). In recent years, the consumption demand for *T. rubripes* is increasing and the price soars in China. However, with the expansion of artificial cultivation, the development of *T. rubripes* industry has been seriously hindered by numerous infectious diseases, which caused the major economic losses. To provide insights into mechanisms underlying disease resistance in *T. rubripes*, identify the molecular markers associated with disease-resistant traits, and develop disease-resistant strains, it is of interest to identify immune related genes and their associated SNPs. Mucosal epithelial surfaces act as dynamic interfaces between the external environment and internal milieu (Cerutti et al., 2011). Gill, as one of the mucosal barriers, plays an important role in mucosal immune responses (Dickerson and Clark, 1998). Several studies on identification of immune-related genes in the gill have been conducted in fish species. A number of differentially expressed immune-related genes including TNFα1, TNFα2, IL-1β2, TGFβ, iNOSa and iNOSb were identified by quantitative real-time PCR in goldfish gill after the *Dactylogyrus intermedius* infection (Lu et al., 2013). In addition, many immune-related genes were characterized,

* Corresponding author. Tel.: +86 411 84763220; fax: +86 411 84763517.
E-mail address: xuemeiqiu417@hotmail.com (X. Qiu).

http://dx.doi.org/10.1016/j.cbd.2014.03.001
1744-117X/

such as hepcidin-like and TLR9 in the gill of *Lateolabrax japonicus* and *Ctenopharyngodon idellus*, respectively (Yang et al., 2006, 2011).

Next-generation sequencing-based RNA-Seq analyses have dramatically changed the way to investigate the functional complexity of transcriptome in many organisms (Anisimov, 2008; Wang et al., 2009). RNA-Seq is a relatively new technology for transcriptomic study across the whole genome in aquaculture species (Liu et al., 2011b, 2012, 2013). RNA-seq based expression profiling has allowed identification of a large number of immune-related genes in the gill. For instance, the arhamnose-binding lectin (RBL) was dramatically upregulated in the gill of catfish infected with *Flavobacterium columnare* (Beck et al., 2012). Illumina-based RNA-Seq was utilized to examine transcriptome profiling in catfish gill after bath immersion infection and identified a large number of important immune-related genes such as IκBs, BCL-3, TAX1BP1, olfactomedin 4, iNOS2b, IFI44, and VHSV (Sun et al., 2012). In addition, transcriptome analysis has been widely reported in many species, including several aquaculture species such as catfish (Liu et al., 2012; Li et al., 2012; Sun et al., 2012), Atlantic cod (Hubert et al., 2010), silver carp (Zheng et al., 2011), pearl oyster (Zhao et al., 2012), carp (Ji et al., 2012), and Amur ide (Xu et al., 2013). This technique has also been used as an efficient and cost-effective method to systematically identify SNPs from transcribed regions in the genomes of several different fish species. By sequencing of the pooled RNA samples from multiple individuals of channel catfish and blue catfish, a set of quality SNPs were identified including 342,104 intra-specific SNPs for channel catfish, 366,269 intra-specific SNPs for blue catfish, and 420,727 inter-specific SNPs between channel catfish and blue catfish (Liu et al., 2011b). A total of 712,042 intra-stain SNPs were discovered in four strains, mirror carp (483,276 SNPs), purse red carp (486, 629 SNPs), Xingguo red carp (478,028 SNPs) and Yellow River carp (488,281 SNPs) (Xu et al., 2012). The large set of SNPs had been also developed in some other aquaculture species, including oyster (Quilang et al., 2007), Altantic salmon (Moen et al., 2008), Atlantic cod (Hubert et al., 2010) and rainbow trout (Salem et al., 2012). Some SNP makers associated with growth traits have been identified from the growth-related genes including *Leptin*, *MC4R*, *IGF*, *Myf5*, *GHRH*, and *Myf6* (Zhang et al., 2012) in our laboratory. Other genetic and genomic studies were also conducted with the focus on identification and characterization of microsatellite markers (Hao et al., 2006; Gu et al., 2010), construction of bacterial artificial chromosome (BAC) and expressed sequence tag (EST) library (Hao et al., 2007).

In this study, we sequenced the transcriptome of the gill of *T. rubripes* using Illumina HiSeq 2000 platform to identify gene-associated SNPs. A total of 56,972 putative SNPs were discovered, which were located in 11,327 genes. These SNPs should provide useful resources for population genetic study, resource assessment, genetic linkage analysis, and genome-wide association studies, and facilitate the development of marker-assisted selection in *T. rubripes* in the future.

2. Methods and methods

2.1. Ethics statement

This study was approved by the Animal Care and Use Committee of the Key Laboratory of Mariculture & Stock Enhancement in North China's Sea at Dalian Ocean University. All surgery was performed under sodium pentobarbital anesthesia, and all efforts were made to minimize suffering.

2.2. Sample collection and RNA isolation

A total of 45 *Takifugu rubripes* (length 20 cm) were sampled from Dalian Tianzheng Industrial Co., Ltd (Dalian, China). The gills of these fish were collected and pooled. Tissues were placed into RNAlater (Ambion), stored at room temperature for 24 h, and then moved to −80 °C for storage until RNA isolation. Total RNA was extracted from the pooled gill using the TRIzol R Reagent (Invitrogen, CA, USA) by following the manufacturer's protocol. The quantity and quality of total RNA were measured using an Agilent 2100 Bioanalyzer.

2.3. cDNA library construction and sequencing

Total RNA was sent out for next generation sequencing provided by the Beijing Institute of Genomics, Chinese Academy of Sciences. cDNA libraries were constructed from mRNA from gill. cDNA libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit (Illumina) according to the TruSeq protocol. After KAPA quantitation and dilution, the libraries were clustered 3 per lane and sequenced on an Illumina HiSeq 2000 instrument with 100 bp paired-end reads.

The reads were mapped to the fugu *T. rubripes* fifth genome assembly by the BWA program. During the mapping phase, up to five mismatches were allowed. The expression levels (RPKM, Reads Per Kilobase of exon model per Million mapped reads) for each gene were calculated using uniquely mapped reads by an in-house built Perl module according to the equation:

$$RPKM = \frac{exon_{reads} \times 10^9}{unique_{reads} \times gene_{length}}$$

The cutoff value of gene expression was calculated for each sequencing sample; genes with RPKM greater than cutoff value were defined as expressed genes (Ramskold et al., 2009).

2.4. SNP identification

BWA and SAMtools (Tools for alignments in the SAM format) software were used to align reads to the fugu genome assembly (version 5.0) for detecting SNPs. Filtering thresholds were set as: consensus quality is no less than 20 and coverage is no less than 10.

2.5. Gene Ontology

Gene Ontology (GO) was conducted to the genes containing SNPs. GO annotation analysis was performed using Blast2GO, an automated tool for the assignment of GO terms. The annotation result was categorized with respect to biological process, molecular function, and cellular component at level 2.

2.6. KEGG pathway analysis

In order to gain an overview of gene pathway networks, KEGG analysis was performed using the online KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/kegg/kass/). The bi-directional best hit (BBH) method was used to obtain KEGG orthology assignments.

2.7. SNP validation

To evaluate the validation rate of the SNPs identified by bioinformatics analysis, we randomly selected 52 SNPs and validated them by PCR amplification and direct sequencing (Yang et al., 2012). PCR primers were designed according to the assembled transcript sequences, and DNA of 10 individuals was used as the PCR template (Table 1).

3. Results and discussion

3.1. Transcriptome sequencing

Illumina sequencing was conducted to generate short sequence reads from the gill of *T. rubripes*. A total of 27,085,235 unique-mapped-reads were obtained from 55,061,524 raw reads after being