Fish & Shellfish Immunology 46 (2015) 366–377

Contents lists available at ScienceDirect

Fish & Shellfish Immunology

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

Full length article

Transcriptome profiling analysis of naked carp (*Gymnocypris przewalskii*) provides insights into the immune-related genes in highland fish

Chao Tong [a, b, c, 1], Cunfang Zhang [a, b, 1], Renyi Zhang [a, b, c], Kai Zhao [a, b, *]

[a] Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810001, China
[b] Laboratory of Plateau Fish Evolutionary and Functional Genomics, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810001, China
[c] University of Chinese Academy of Sciences, Beijing, 100049, China

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form
8 June 2015
Accepted 19 June 2015
Available online 24 June 2015

Keywords:
Gymnocypris przewalskii
Gill and kidney
Transcriptome
Immune-related gene
Gene family
SSR mark

ABSTRACT

The naked carp, *Gymnocypris przewalskii*, is one of the dominant aquaculture fish species in Qinghai Province, China. Its wild stocks have severely suffered from overfishing, and the farming species are vulnerable to various pathogens infections. Here we report the first immune-related tissues transcriptome of a wild naked carp using a deep sequencing approach. A total of 158,087 unigenes are generated, 2687 gill-specific gene and 3215 kidney-specific genes are identified, respectively. Gene ontology analysis shows that 51,671 unigenes are involved in three major functional categories: biological process, cellular component, and molecular function. Further analysis shows that numerous consensus sequences are homologous to known immune-related genes. Pathways mapping annotate 56,270 unigenes and identify a large number of immune-related pathways. In addition, we focus on the immune-related genes and gene family in Toll-like receptor signaling pathway involved in innate immunity, including toll-like receptors (TLRs), interferon regulatory factors (IRFs), interleukins (ILs) and tumor necrosis factors (TNFs). Eventually, we identify 5 TLRs, 4 IRFs, 3 ILs and 2 TNFs with a completed coding sequence though mining the transcriptome data. Phylogeny analysis shows these genes of naked carp are mostly close to zebrafish. Protein domain and selection pressure analyses together show that all these genes are highly conserved in gene sequence and protein domain structure with other species, and purifying selection underwent in these genes, implied functionally important features are conserved in the genes above. Intriguingly, we detect positive selection signals in naked carp *TLR4*, and significant divergence occurred among tested species *TLR4*, suggested that naked carp *TLR4* function may be affected. Finally, we identify 23,867 simple sequence repeat (SSR) marks in this transcriptome. Taken together, this study not only contributes a large number of candidate genes in naked carp immunity, and also helps improve current understanding of immunogenetics basis and evolutionary history of immune related genes and gene family in highland fish species.

1. Introduction

The naked carp, *Gymnocypris przewalskii*, is a cold-water and tetraploid fish that belongs to the order Schizothoracine and the family Cyprinidae [1]. It is mainly distributed in Lake Qinghai and its peripheral rivers at northeast of the Qinghai–Tibetan Plateau (QTP) at high altitude. Naked carp is one of the dominant aquaculture fish species in Qinghai Province, China due to its rich nutrients and trace elements [2,3]. Human activities, including overfishing and habitat destruction, have dramatically decreased the number of naked carp population, makes naked carp listed as an endangered species in the “China Species Red List” [4]. In recent years, the local government and environmentalists have tried to expand its populations through large-scale artificial reproduction and farming [5,6]. However, the nightmare has

* Corresponding author. Northwest Institute of Plateau Biology, Chinese Academy of Sciences, 23rd Xinning Road, Xining, Qinghai 810001, China.
E-mail address: zhaokai@nwipb.cas.cn (K. Zhao).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.fsi.2015.06.025
1050-4648/

come, rapid development of naked carp farming industry is threatened by infectious diseases, mainly known as the "white spot disease", caused by infection with a highly virulent ciliate parasite *Ichthyophthirius multifiliis* [2,3]. Moreover, saprolegniasis, a fungal disease has been treated as the second-severe threat facing its farming industry [2]. The number of severe outbreaks has increased and resulted in more than 65% of mortality and caused great economic losses in farmed naked carp [3]. Here is the severe problem that naked carp farming industry is facing today, we know little about molecular mechanisms underlying the immune response to pathogens in this highland fish species. Thus, characterization of the immunogenetic bases of naked carp contributes to establishment of effective measures in disease control and evolutionary adaptation to high altitude aquatic environment.

As the recent rapid development of aquaculture, fish diseases caused by pathogens are regarded as a significant constraint on the development of aquaculture resulting in great economic losses [7]. The ability to detect the presence of the invading pathogen is key to successful defense against pathogenic infection [8]. Under these circumstances, fish species heavily depend on innate or adaptive immune responses for the rapid elimination of pathogens [9]. There has been increasing interest in investigation of the molecular mechanisms of immune defense against pathogenic infection in fish species. Recent genome-wide studies were performed in several aquaculture fish species have provided insights into immunogenetic bases and the molecular and genetic mechanisms underlying their responses to environmental stressors [10–14]. However, the whole-genome sequencing (WGS) approach is limited by extremely high cost and species specificity [14]. Moreover, most teleost fish species underwent a fish-specific genome duplication (FSGD) and existed in the form of polyploidy [15]. Within Tibet fish species, a majority of fish species are tetraploid or hexaploid [1], this dramatically increases the difficulty of genome sequencing in Tibet fish.

Transcriptome profiling analysis is a rapid and effective approach for genome survey, massive functional gene and molecular maker identification [16]. Recent rapid advances in high-throughput deep sequencing technologies have offered the opportunity to generate transcriptome in almost any non-model organisms of interest. Recent years have witnessed an increase in studies of understanding the transcripts from various aquaculture fish species using deep transcriptome sequencing technology, such as *Megalobrama amblycephala* [17], *Cynoglossus semilaevis* [18], *Cyprinus carpio* [19], *Lateolabrax japonicas* [20]. In this study, we performed the transcriptome analysis on two immune-related tissues gill and kidney of Tibet fish, *G. przewalskii*. We determined gene expression at the transcriptome level, identified immune-related genes and main gene families in innate immunity, tested the selection pressure on genes and detected molecular markers. Our study provides abundant genomic resource for future research into the immunogenetics and evolutionary history of immune-related genes of naked carp.

2. Materials and methods

2.1. Fish sampling and animals and ethics

Healthy naked carp (*G. przewalskii*) with an average body weight of 25.2 g (n = 3), used here were collected from Lake Qinghai, China. After collection, the fish were dissected, and both gill and kidney were collected and immediately stored in liquid nitrogen at −80 °C. All animal experiments were approved by Agriculture Department of Qinghai Province, China.

2.2. RNA extraction, library construction and illumina sequencing

Total RNA was extracted from the frozen gill and kidney tissues using Trizol Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions and then treated with RNase-free DNase I (Thermo Scientific, USA) to remove genomic DNA contamination. RNA quality and quantity were examined using agarose gel electrophoresis and Agilent 2100 Bioanalyzer (Agilent, USA), respectively. Equal amounts of RNA from both tissues of a same individual were pooled firstly. To eliminate the individual differences, equal quantities of high-quality mixed RNA of each tissue from three different individuals were mixed, and 5 μg total RNA was collected for each mixed group, respectively.

Two cDNA libraries (gill and kidney) were constructed using the Illumina TruSeq RNA Sample Preparation Kit (Illumina, USA) following manufacturer's instructions. Subsequently, then beads coated with oligo $(dT)_{16}$ were used to isolate poly (A) mRNA. Fragmentation buffer (Ambion, Austin, TX, USA) was added to generate short mRNA fragments through digestion. Random hexamers primers were used for first strand cDNA synthesis, followed by synthesis of the second strand. After several steps of purification, adapter addition and cDNA length selection, two libraries were prepared and sequenced independently using an Illumina HiSeq™ 2000 platform (Shanghai Personal Biotechnology Cp., Ltd. Shanghai, China).

2.3. De novo assembly of sequencing reads and functional annotation

RNA-Seq raw reads were processed to obtain high-quality reads by removing the adapter sequences and low-quality bases at the 3′ end, trimming low-quality bases (Q < 20) from the 5′ and 3′ ends of the remaining reads. Reads filtering out reads containing 'N' (unknown bases in read) and greater than 25 bp were considered for analysis. Transcriptome assembly was carried out with short reads assembling Trinity program (http://trinityrnaseq.sf.net) with default parameters. In order to count how many base pairs per reads (single reads) generated by Hiseq 2000, this result can be calculated by total clean nucleotides (nt) divided by total clean reads [21]. Contigs from each sample's assembly (gill and kidney) were performed by TGICL software [22], generated non-redundant unigenes as long as possible, with a minimum overlap length of 100 bp.

All unigene were aligned by BLASTX search (cut-off E-value of $1e^{-6}$) in known protein databases to identify the coding sequence (CDS), including NCBI non-redundant database (Nr), Swiss-Prot protein database, Kyoto Encyclopedia of Genes and Genomes database (KEGG), and Clusters of Orthologous Groups of proteins database (COG). That cannot be aligned to any protein database were also scanned using ESTScan [23], producing predict coding region and direction. Blast2GO software [24] was used to obtain Gene Ontology (GO) (http://www.geneontology.org/) annotation of the unigenes based on BLASTX hits against the Nr database (E-value < 10^{-5}). Each annotated sequence was assigned to detailed GO terms and calculated under the categories of biological process, cellular component, and molecular function. The unigene sequences were also aligned to the COG database (http://www.ncbi.nlm.nih.gov/COG/) to predict and classify functions. Pathway assignments were generated using the KEGG database (http://www.genome.jp/kegg) and the BLASTX algorithm with an E-value threshold of 10^{-5}.

2.4. Tissue-specific genes and comparative expression analysis

Gene expression values were calculated as reads aligned to gene