Gene 576 (2016) 261–267

Contents lists available at ScienceDirect

Gene

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

Research paper

Genome-wide identification and characterization of microRNA genes and their targets in large yellow croaker (*Larimichthys crocea*)

Yong Huang *, Jia-Heng Cheng, Fu-Nong Luo, Hao Pan, Xiao-Juan Sun, Lan-Yu Diao, Xiao-Juan Qin

College of Animal Science and Technology, Henan University of Science and Technology, Luoyang, China

ARTICLE INFO

Article history:
Received 16 August 2015
Received in revised form 4 October 2015
Accepted 13 October 2015
Available online 30 October 2015

Keywords:
MicroRNA
Large yellow croaker
Computational identification
Targets

ABSTRACT

MicroRNAs (miRNAs or miRs) are a class of non-coding RNAs of 20–25 nucleotides (nt) in length, which regulates the expression of gene in eukaryotic organism. Studies has been confirmed that miRNA plays an important role in various biological and metabolic processes in both animals and plants. Predicting new miRNAs by computer based homology search analysis is an effective way to discover novel miRNAs. Though a large number of miRNAs have been reported in many fish species, reports of miRNAs in large yellow croaker (*L. crocea*) are limited especially via the computational-based approaches. In this paper, a method of comparative genomic approach by computational genomic homology based on the conservation of miRNA sequences and the stem–loop hairpin secondary structures of miRNAs was adopted. A total of 199 potential miRNAs were predicted representing 81 families. 12 of them were chose to be validated by real time RT-PCR, apart from miR-7132b-5p which was not detected. Results indicated that the prediction method that we used to identify the miRNAs was effective. Furthermore, 948 potential target genes were predicted. Gene ontology (GO) analysis revealed that 175, 287, and 486 target genes were involved in cellular components, biological processes and molecular functions, respectively. Overall, our findings provide a first computational identification and characterization of *L. crocea* miRNAs and their potential targets in functional analysis, and will be useful in laying the foundation for further characterization of their role in the regulation of diversity of physiological processes.

1. Introduction

MicroRNAs (miRNAs or miRs) are endogenous non-coding RNAs, 20 to 25 nts in length that not only have been described in the genomes of almost all animals and plants but also in viruses, as well as several clades of unicellular eukaryotes (Tarver et al., 2012; Tarver et al., 2015; van der Burgt et al., 2009; Wienholds and Plasterk, 2005; Yu et al., 2015; Zhou et al., 2012; Zhu et al., 2009). They regulate gene expression at the post-transcriptional level by repressing gene translation or targeted degradation of specific mRNAs, depending on the extent of complementarity between the miRNA and its target mRNAs (Bartel, 2009; Du and Zamore, 2007; Huang et al., 2010a). MiRNAs originate in the nucleus, where long primary transcripts (pri-miRNAs) are initially transcribed with a cap and poly-A tail, and the RNase III endonuclease Drosha cleaves pri-miRNA hairpins into pre-miRNAs. Next, pre-miRNA is transported to the cytoplasm by the exportin 5/Ran GTPase pathway, where it undergoes further processing by the Dicer endonuclease into a short double-stranded RNA duplex, the miRNA/miRNA* duplex. Finally, one strand of the duplex is incorporated into the RNA-induced silencing complex (RISC) (Ketting, 2010; Yang and Wang, 2011). Accumulating evidences proved that miRNAs play key roles in a broad range of biological processes including growth, development, metabolism, cellular differentiation, proliferation, viral infection, apoptosis, oncogenesis, cancers and diseases (Dokanehiifard et al., 2015; Huang et al., 2011; Macfarlane and Murphy, 2010; Shi et al., 2014; Wang et al., 2014; Yu et al., 2015; Zou et al., 2015a; Mao et al., 2015b).

During recent years the identification and characterization of miRNA and their target genes from animals have been extensively studied (Collins, 2011; Kozomara and Griffiths-Jones, 2014; Unver and Budak, 2009), which is essential to understand the regulation network of miRNAs in gene expression. Currently, two approaches have been reported for identification of miRNAs: (1) cloning and miRNA high-throughput sequencing from a small RNA library (Creighton et al., 2009; Lee et al., 1993; Samols et al., 2005; Sunkar et al., 2005; Wang et al., 2015; Xiao, 2011), (2) computational prediction from genomic sequences and ESTs and GSS (Akter et al., 2014; Baev et al., 2009; Frazier et al., 2010; Lai et al., 2003; Lin et al., 2010; Luo and Zhang, 2009; Panda et al., 2014; Wang et al., 2013; Wei et al., 2012; Wei et al., 2014). Computational approaches based on the secondary structure characteristics (Zou et al., 2009), phylogenetic conservation of both sequence and structure, sequence alignment (Quan et al., 2015) and

Abbreviations: RISC, RNA-induced silencing complex; Nt, Nucleotides; *L. crocea*, *Larimichthys crocea*; SVM, Support Vector Machine; GO, Gene ontology; NCBI, National Center for Biotechnology Information; UTRs, Untranslated region; MFE, Minimum free energy.

* Corresponding author.
E-mail address: huangyong1979111@163.com (Y. Huang).

http://dx.doi.org/10.1016/j.gene.2015.10.044
0378-1119/

thermodynamic stability of hairpins have been developed as alternative methods to discover miRNAs that usually cannot be detected by the direct cloning, particularly those miRNAs were at low abundance. Because this approach is inexpensive, rapid, accurate, and affordable with experimental methods, a great number of miRNAs were identified in vertebrates (Barozai, 2012; Gong et al., 2010; Huang et al., 2008; Long and Chen, 2009), insect (Chatterjee and Chaudhuri, 2006; Jia et al., 2010; Tong et al., 2006) and plant (Han et al., 2014; Patanun et al., 2013; Zhang et al., 2009). The large yellow croaker is one of the most economically important marine fish in China and East Asian countries due to its rich nutrients and trace elements, especially selenium (Tang et al., 2009; Zhao et al., 2012). Thus far, only 1637 fishes miRNAs from 8 different fish species have been annotated according to the miRBase database release 21 in June 2014 (http://www.mirbase.org/), but no one *L. crocea* miRNAs are listed in the miRBase. More recently, a research have been performed experimentally to identify *L. crocea* miRNAs (Qi et al., 2014) and no studies about miRNAs have been reported via the computational approaches considering *L. crocea* available genomic sequences in the public NCBI database.

In the present study, we focused on computational approaches by employing a well-defined comparative genomic based homologue search to identify novel miRNA homologs in *L. crocea*. Because of considering all of the previously reported animal miRNAs in the miRBase for the analysis, the computational approach employed here is sensitive enough to predict the novel miRNA homologues from distantly related species as well. A total of 199 *L. crocea* miRNAs were identified involving 81 miRNA families. And then their characteristics, the conservation of miRNAs and their precursors were also investigated. Furthermore, we random selected 12 predicted *L. crocea* miRNAs to confirm exist by stem–loop quantitative RT-PCR. Subsequently, we predicted 1147 potential target genes of these predicted miRNAs on the 3′UTR of *L. crocea* genes. Most of the target genes were found to encode transcription factors. Finally, the functional annotation of miRNA targets based on the GO database was conducted to provide additional information regarding the biological functions of *L. crocea* miRNAs. Our results significantly expand the set of known miRNAs in *L. crocea* and provide theoretical significance and practical value to further study of *L. crocea* functional genomics.

2. Materials and methods

2.1 Sequences data

To search potential *L. crocea* miRNAs, a total of 8695 previously known animal miRNAs were defined as a reference set of miRNA sequences. These animal mature miRNAs and their precursor sequences were downloaded from the miRBase (Release 21.0; June 2014). Most of these were identified or verified by experiments, and others were computationally predicted as their close homolog. To avoid the redundant or overlapping miRNAs, the repeated sequences of miRNAs within the above species were removed and the remaining sequences were used as query sequences for BLAST search. The genomic sequences of *L. crocea* and mRNA sequences were downloaded from NCBI database.

2.2 Computational identification of potential miRNAs

The alignment tool BLAST version 2.2.27 was used to identify the potentially conserved miRNAs and was downloaded from the NCBI website. BLASTN parameter settings are as follows: an expect value cut-off of 10; the window size 7; a low-complexity sequence filter; number of descriptions and alignments were 1000. All BLAST results were saved and used for further analysis. The prediction process was shown in Fig. . 1. Following five criteria were raised to identify the potential miRNAs: (1) predicted mature miRNAs need to have no more than 2 nt mismatches as compared with the known miRNAs, (2) the precursor of miRNA sequence could fold into a marked stem–loop hairpin secondary structure, (3) the mature miRNA containing the ~22 nt sequence should be located in either arm of the hairpin structure, (4) there are no loops in the miRNA/miRNA*duplex, (5) the minimal folding free energy (MFE) of predicted pre-miRNA secondary structures has to be lower than −20 kcal/mol, while the minimal folding free energy index (MEFI) of it usually must be over 0.8 (Zhang et al., 2006), and (6) the A + U content of pre-miRNAs within 30% to 70% was considered since the unstable structures of pre-miRNAs are needed to produce mature single-stranded miRNAs. Besides, we submitted the new predicted miRNA precursor sequences to program called SVM that was improved by our lab using the 'Support Vector Machine' model which identifies pre-miRNA-like hairpin sequences classifying them as real or pseudo-pre-miRNA. The model has been successfully applied to identify pre-miRNAs and mature miRNAs form (Huang et al., 2015; Huang et al., 2010b). Finally, some possible false sequences of pre-miRNAs should be deleted by manual detection.

2.3 Conservation analysis of L. crocea miRNAs

The conserved nature of miRNAs makes them a computational resource for homologues discovery (Sun et al., 2013). Consequently we also analyzed the *L. crocea* miRNA conservation with their homologues. The novel identified miR-155-5p, miR-223, miR-460a-3p and the well-known other animal miRNA homologues were selected for conservation and were done with the help of publically available weblogo: a sequence logo generator (Salvi et al., 2010). The results were saved.

2.4 Experimental verification of predicted miRNAs

To verify computational predictions, 12 miRNAs were randomly selected from the novel predicted miRNA to verify their expression by real time RT-PCR experiment method. Small RNA from maricultured large yellow croaker liver tissue sample was extracted using RNeasy Mini Kit Reagent (Qiagen, China) according to the manufacturer's protocol. Small RNA sample was quantified and assessed for quality using IMPLEN Nanophotometer (Pearl) and then stored at −80 °C for further use. Reverse transcription was carried out in 20 μL reaction mixture. Briefly, 1 μg of total RNA of each sample was used to generate a single-stranded miRNA cDNA using miRNA specific stem–loop RT primers according to criteria described previously (Chen et al., 2005; Hurley et al., 2012; Salone and Rederstorff, 2015). The specific stem–loop RT-PCR primers and gene specific primers were listed in **Table S1**. Real-time PCR was performed using Power SYBR-Green PCR master mix kit (Applied Biosystems) on the BIORAD iQ5 Real-time PCR Detection System. The 15 μL PCR included 1 μL RT product, 2 × SYBR-Green PCR master mix, 2 μM forward primer and 2 μM reverse primer. The reactions were incubated at 95 °C for 5 min followed by 40 cycles of 95 °C for 25 s, 60 °C for 30 s and 72 °C for 30 s. Normalization was performed with 5S rRNAs. The data was analyzed using $^{2-\Delta\Delta CT}$ method, in which $\Delta Ct = C_{T\ miRNA\text{-}CT5S}$. Ct represents the threshold cycle. The PCR products were cloned into pMD19-T Vector (Takara, China), and were sequenced by using M13 sequencing primers.

2.5 Target prediction for L. crocea miRNAs and their functions

The mRNA database of *L. crocea* downloaded from NCBI database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene) and their 3′ UTR sequences which are ≥20 nt in length were extracted using a perl script and used for target prediction. Potential targets of the predicted miRNAs were identified using RNAhybrid program. The parameters employed were described as follows: (1) P-value cutoff of 0.05, target duplex free energy $\Delta G \leq -24$ kcal/mol; (2) no mismatches in the seed region (5′ region of mature miRNA, from second to eighth nt position); (3) only one G:U pairing in the seed region; and (4) the miRNA sequences and potential mRNAs targets were no more than four gaps