Biochemical Systematics and Ecology 66 (2016) 188–195

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

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

Characterization of pikeperch (*Sander lucioperca*) transcriptome and development of SSR markers

Xiaofei Han [a], Qufei Ling [a, *], Caijuan Li [a, **], Guocheng Wang [a], Zhengchao Xu [a], Guoqing Lu [b]

[a] School of Biology and Basic Medical Sciences, Soochow University, 199, Renai Road, Suzhou, Jiangsu 215123, PR China
[b] Department of Biology & School of Interdisciplinary Informatics, University of Nebraska at Omaha, Omaha, NE 68182-0040, United States

ARTICLE INFO

Article history:
Received 4 November 2015
Received in revised form 26 March 2016
Accepted 8 April 2016
Available online 19 May 2016

Keywords:
Sander lucioperca
Illumina RNA-sequencing
SSR markers

ABSTRACT

The pikeperch (*Sander lucioperca* L.) is one of the most economically important freshwater species and has been recently explored as a potential candidate for aquaculture. To facilitate pikeperch research, we sequenced its transcriptome and developed a set of microsatellite markers. We conducted Illumina RNA-sequencing and obtained over 50 million reads from a pooled cDNA library of different tissues. The clean reads were *de novo* assembled into 56,746 transcripts, with an average length of 1474 bp. The annotation analysis demonstrated 37,386 transcripts (65.9%) with homologous sequences in the NCBI Nr protein database. Of these annotated transcripts, 18,576 sequences were successfully assigned into Gene Ontology (GO) terms, 23,566 transcripts into the Cluster of Orthologous Groups (COG), and 12,081 transcripts to 322 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Totally, 16,368 SSRs (≥10 bp) were detected from 11,921 unigenes. The validation of randomly selected 300 SSR markers demonstrated 87.0% of the markers can be successfully amplified, suggesting RNA-Seq is an efficient tool for the development of molecular markers. This study provides not only a valuable transcriptomic resource, but also a set of SSR markers for basic as well as applied research in pikeperch.

1. Introduction

The pikeperch (*Sander lucioperca* L.), native to Black, Caspian and Baltic Sea drainages, is an economically important freshwater species of Percidae family (Deelder and Willemsen, 1964; Svärdson and Molin, 1973; Sonesten, 1991). In China, it inhabits in the Ili River and the Irtysh River of the Xinjiang Uigur Autonomous Region. The pikeperch has been introduced into seven provinces in China since 1993 (when artificial propagation and culture were successfully practiced in Xinjiang) mainly due to its unique characteristics such as strong disease resistance, fast growth, tender meat, less intermuscular bones, and high protein content (Liu and Li, 1995). The pikeperch is being developed as an important freshwater aquaculture species in China.

* Corresponding author.
** Corresponding author.
E-mail addresses: lingqf@suda.edu.cn (Q. Ling), cjli@suda.edu.cn (C. Li).

http://dx.doi.org/10.1016/j.bse.2016.04.001
0305-1978/

Molecular markers are essential tools used for brood stock management and selective breeding. Microsatellites, simple sequence repeats (SSR), are widely used markers due to many of their advantages such as hypervariability, reproducibility, codominant inheritance, and extensive genomic coverage (Powell et al., 1996). SSRs have been used to assess genetic diversity, create linkage maps, identify subsequent quantitative trait loci (Sun, 2010; Duran et al., 2009). In the Percidae family, SSRs have been isolated from several species, including *Zingel asper* (Dubut et al., 2010), *Perca flavescens* (Leclerc et al., 2000), *Stizostedion vitreum* (Wirth et al., 1999). These SSRs could be cross-amplified in the pikeperch. However, only nine SSR markers have been isolated from pikeperch (Kohlmann and Kersten, 2008). It is thus essential to develop more SSR markers for pikeperch genetic research and aquaculture practice.

The transcriptome consists of information on genes and their expression and regulation. Sequencing and assembly of the transcriptome (the complete set of expressed genes in an organism) is an effective way to obtain functional genomic data in species whose genome sequence is available. The next-generation sequencing has been widely applied to characterize the transcriptome and develop a large number of genetic markers (SSRs and SNPs), especially for non-model organisms, such as the Blunt Snout Bream (*Megalobrama amblycephala*) (Gao et al., 2012), and recent reports in Blackfin tuna (*Thunnus atlanticus*), Abalone (*Haliotis laevigata*), Red snapper (*Lutjanus campechanus*), and Gazami crab (*Portunus trituberculatus*) (Antoni et al., 2014; Shiel et al., 2014; Norrell et al., 2014; Lv et al., 2014). In this study, we presented the pikeperch transcriptome using Illumina paired-end sequencing technology, described its functional annotation, and discussed the development of SSR markers for future marker-assisted selective breeding and brood stock management.

2. Materials and methods

2.1. Sample material

Two mature pikeperch individuals(one male, 1.4 kg and one female, 1.3 kg)were sampled from the Irtysh River in Xinjiang, China. Tissues of muscle, fin, brain, spleen, liver, bone, kidneys, heart and gonads were collected and stored in RNA sample protector (D311A, TaKaRa Biotechnology Co., Ltd.).

2.2. RNA extraction and library preparation

Total RNA was extracted with TRIzol® reagent according to the manufacturer's instructions (Invitrogen). RNA quality and quantity were measured with the NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA), 1.0% agarose gels and Agilent 2100 Bioanalyzer. Equal amount of RNA from each tissue was mixed to form the cDNA library, using a TruSeq™ RNA sample prep kit (Illumina San Diego, CA). The library was size-selected for target fragments of 300–500 bp on a 2.0% agarose gel followed by PCR amplification using Phusion DNA polymerase (NEB) for 15 cycles. After quantification by TBS380, paired-end sequencing was conducted using the Illumina Hiseq 2000 platform.

2.3. Data filtering and de novo assembly

SeqPrep (https://github.com/jstjohn/SeqPrep) and sickle (https://github.com/najoshi/sickle) software were used to perform a stringent filtering process. Adapter sequences, low quality reads (quality score<20), ambiguous sequences represented as “N” and reads shorter than 20 nt were removed. *De novo* assembly of the high-quality clean reads was performed using Trinity (http://trinityrnaseq.sourceforge.net/) with default settings (Grabherr et al., 2011).

2.4. Gene annotation and analysis

The assembled transcripts were blasted against the NCBI non-redundant (Nr) protein database with an E-value threshold of $1e^{-5}$ (Altschul et al., 1997). Gene Ontology (GO) Classification with blast2go application was performed based on the blast result (Ashburner et al., 2000). Clusters of Orthologous Groups (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were respectively performed using BLAST 2.2.28+software (Kanehisa et al., 2006, 2008).

2.5. SSR detection and primer design

Perfect SSRs with repeat motif longer than 10 bp were detected using msacommander software and SSRs primer pairs were designed by Primer 3.0 (Faircloth, 2008). 300 SSRs with repeat motifs more than 15 bp (mononucleotide repeats excluded) were randomly selected for primer synthesis and validation. Of those successfully amplified loci, 100 SSR loci were randomly selected for polymorphism evaluation.

2.6. SSR polymorphism validation

Thirty pikeperch individuals from Suzhou Donghu Fish Farm, Jiangsu, China were used for the validation of SSR markers. Genomic DNA was extracted from fin tissue using a standard proteinase K digestion, phenol-chloroform procedure. PCR mixture containing 100 ng template DNA, 10 μL 2 × Es Taq Master Mix (CW2296, Shiji Kangwei, China), 0.4 μM of each primer