Aquaculture 497 (2018) 56–63

Contents lists available at ScienceDirect

Aquaculture

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

Identification of SNPs associated with residual feed intake from the muscle of *Litopenaeus vannamei* using bulk segregant RNA-seq

Ping Dai[a,b,1], Jie Kong[a,b,1], Sen Wang[c], Xia Lu[a,b], Kun Luo[a,b], Baoxiang Cao[a,b], Xianhong Meng[a,b], Sheng Luan[a,b,*]

[a] Key Laboratory for Sustainable Utilization of Marine Fisheries Resources, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China
[b] Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266235, China
[c] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Feed efficiency
Litopenaeus vannamei
Residual feed intake
RNA sequencing
Single nucleotide polymorphism

ABSTRACT

Feed efficiency is one of the most important economic traits of *Litopenaeus vannamei*. Molecular breeding such as marker-assisted selection or genomic selection is becoming a promising method for genetic improvement of this trait, considering the difficulty in collecting individual feed efficiency data. In this study, we used residual feed intake (RFI) as the measure of feed efficiency, aiming to develop RFI-associated single nucleotide polymorphism (SNP) markers by an integrated method of pooled RNA-seq along with bulked segregant analysis. Based on a population consisted of multiple families, specifically, we constructed a high-efficiency bulk and a low-efficiency bulk for pooled RNA-seq of muscle samples. > 200,000 SNPs were called from pooled RNA-seq data of each bulk, with 123,219 SNPs being shared. Among them, 11,026 SNPs showed significant allelic imbalance between bulks (adjusted P value < 0.05). Differential expression analysis between bulks identified 4231 differentially expressed unigenes (DEGs), and 568 of the 11,026 significant SNPs were located in the DEGs. As shown by functional analysis, many DEGs harboring significant SNPs were involved in some basic metabolic processes and important biological pathways. The results of this study will provide some new resources for implementing molecular breeding of feed efficiency in L. *vannamei* and some new knowledge for elucidating the molecular mechanisms underlying this trait.

1. Introduction

The Pacific white shrimp (*Litopenaeus vannamei*) is one of the primary cultivated aquatic species in the world because of the high market value, and its global aquaculture production reached 4,155,827 ton in 2016 (FAO, 2018). This species also accounts for > 80% of the farmed shrimp production (Xiong et al., 2011). Feed cost is a major factor influencing the profitability of shrimp production system, considering that feed may represent more than half the costs in intensive shrimp culture (Tan and Dominy, 1997). A high feed efficiency will increase the profitability of shrimp production, and so feed utilization efficiency is becoming a worthwhile variable for genetic improvement of this species. Residual feed intake (RFI) has been gradually recognized as a more suitable measure of feed efficiency (Herd et al., 2003; Kennedy et al., 1993), which is defined as the difference between an animal's actual feed intake and its expected feed intake requirement predicted based on the animal's growth and maintenance (Koch et al., 1963). A low RFI value stands for high feed efficiency while a high RFI value indicates low efficiency. Notably, RFI has been reported to be heritable in growing L. *vannamei*, and to be genetically independent from growth (Dai et al., 2017a). Therefore, it is an ideal subject for studying the determinants of feed efficiency.

Since shrimp are generally reared in large groups, the difficulty inherent to obtain accurate measurement of feed intake for individual animals has been a major limitation to improve feed efficiency in animal breeding. In such a situation, genetic improvement could be facilitated if robust genetic markers are available. Specifically, identification of markers associated with RFI will be crucial to applying marker-assisted selection (MAS) or genomic selection (GS) for feed efficiency at an early age with lower cost than can be achieved measuring

* Corresponding author: Key Laboratory for Sustainable Utilization of Marine Fisheries Resources, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China.
E-mail address: luansheng@ysfri.ac.cn (S. Luan).
[1] Ping Dai and Jie Kong contributed equally to this work.

https://doi.org/10.1016/j.aquaculture.2018.07.045
Received 6 November 2017; Received in revised form 4 July 2018; Accepted 26 July 2018
Available online 27 July 2018
0044-8486/

individual feed intake. The well-developed next-generation sequencing is revolutionizing molecular breeding by facilitating the development of large numbers of single nucleotide polymorphism (SNP) markers. On the basis of whole genome sequencing, considerable researches have been undertaken for genomic selection using high-density SNP chips in cattle, pig and chicken (Fan et al., 2010; Onteru et al., 2013; Pryce et al., 2012; Santana et al., 2014; Yuan et al., 2015). However, *L. vannamei* genomics remains a mostly unexplored area of research because of the complex and large genome (≥2000 Mb), and to date the number of available SNP markers is far unable to implement MAS or GS for this species. In this case, RNA sequencing (RNA-seq) analysis is an ideal alternative method that can reduce data complexity and allow marker discovery in expressed sequences that directly deliver candidates for the trait of interest. This technology has been widely used for SNP discovery in aquatic species (Hubert et al., 2010; Lu et al., 2018; Zhao et al., 2012; Zheng et al., 2011).

Markers associated with complex target traits have generally been obtained through linkage and QTL analyses that allow identification of genomic regions containing genes, but such association analyses need to genotype an entire population. Besides, a simple and efficient method, bulked segregant analysis (BSA), has been widely used to study complex traits such as growth and resistance in plants and animals (Asnaghi et al., 2004; Becker et al., 2011; Dussle et al., 2003; Lee et al., 2003; Suo et al., 2001). This method is based on the notion that phenotypic extremes should have a strong contrast in their genotypes at the genomic location linked to the trait. It will significantly reduce the scale and cost by selection of extremes or representative samples from a population showing wide ranges of phenotypic variation for a target trait. Importantly, BSA can be directly combined with RNA-seq for the development of genetic markers associated with specific target trait (Wang et al., 2013; Yao et al., 2017). This so called bulked segregant RNA-seq (BSR-seq) possesses the advantages of both BSA and RNA-seq, showing strong ability to detect genetic differences underlying a trait.

In a previous study, we have identified hundreds of RFI-associated genes in L. *vannamei* by comparative transcriptome analysis (Dai et al., 2017b). However, there has been no study on the identification of RFI-associated SNPs in this species. The present study combined both BSA and pooled RNA-Seq to analyze the whole transcriptome for SNPs associated with RFI and obtain insight into the potential functional effects. In addition, this work will provide new transcriptomic and SNP-derived resources, contributing to the development of molecular breeding for feed efficiency of L. *vannamei* and to a better understanding of molecular mechanisms of RFI.

2. Methods and methods

2.1. Animals and phenotypic data

A breeding population of L. *vannamei* was established using several commercial strains in 2012, and since then the closed and discrete generations were produced yearly based on a nested mating design (Tan et al., 2017). In 2015, 34 full-sib families produced by 31 sires and 31 dams were chosen for feeding test by housing animals in isolation. Specifically, 18 individuals were randomly collected from each family when the smallest shrimp reached 4 cm. The feeding test continued for 42 days, which was divided into two successive stages (days 1–21 and days 22–42). Trait records including individual feed intake and weight gain for the two stages were collected as previously described (Dai et al., 2017a). By definition, RFI is equal to the difference between expected and observed daily feed intake. Expected daily feed intake was calculated as a multiple regression with observed daily feed intake as the dependent variable, and the model was also described by Dai et al. (2017a).

Table 1
The composition of the high-efficiency bulk and the low-efficiency bulk.

High-efficiency bulk		Low-efficiency bulk	
Family ID	The number of individuals	Family ID	The number of individuals
7002	3	7002	1
7029	1	7023	3
7047	4	7029	1
7084	3	7082	2
7142	3	7137	2
7143	1	7143	1
7146	3	7146	1
7152	1	7152	2
7157	1	7154	2
7162	2	7158	1
7172	2	7160	1
7173	1	7162	1
7185	2	7163	2
7186	1	7166	4
7189	2	7172	2
7194	3	7173	1
7195	2	7178	3
7202	2	7184	1
7207	1	7189	1
7223	2	7202	1
		7207	1
		7218	1
		7220	3
		7224	2

2.2. Sampling and Illumina sequencing

After the feeding test, RFI phenotypes of 519 animals were collected across the two stages (Dai et al., 2017a), and 12 animals from two families have been used for individual RNA-seq in our previous study (Dai et al., 2017b). According to RFI ranking of the remaining 507 animals during days 22–42, 40 animals with the lowest RFI values and 40 animals with the highest RFI values were chosen to comprise a high-efficiency bulk (HB) and a low-efficiency bulk (LB), respectively. The composition of HB and LB is present in Table 1. Animals of HB and LB originated from 20 and 24 families, respectively, and 11 families were shared by both bulks.

The day after trait records being collected, muscle tissue of the third abdominal segment was excised from each of the 80 animals at half an hour after a feed and frozen in liquid nitrogen. The animals were about four months old at sampling time. Total RNA was extracted from each muscle sample using the TRIzol Reagent (Invitrogen, UK) and then treated with DNase I. RNA integrity and concentration were assessed using a Bioanalyzer 2100 (Agilent Technologies, USA). Equal masses of total RNA from 40 samples of each bulk were pooled for RNA-Seq. Sequencing libraries of two pooled samples were constructed using TruSeq RNA Sample Peparation Kit (Illumina, USA), as dictated by the TruSeq protocol, and sequenced on the Illumina HiSeq 4000 platform (Illumina, USA) with 150 bp paired-end reads. Raw reads were subjected to quality control using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Trimmomatic (Bolger et al., 2014) to obtain clean reads. In detail, adapter sequences (reads containing more than five adapter-polluted bases), low quality reads (low quality bases with phred Quality value < 19 accounting for > 15% of total bases) and reads with poly-N (the number of N bases accounting for > 5% of total bases) were filtered.

2.3. De novo assembly

The clean reads from the two pooled samples were *de nove* assembled using Trinity software package (Grabherr et al., 2011) with default parameters (k-mer length = 25). The transcripts with significant similarities (> 90%) were clustered using the BLAT with