Biochemical Systematics and Ecology 64 (2016) 111–121

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

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

Comparative genomics analysis of decapod shrimps in the Pancrustacea clade

Jianbo Yuan, Xiaojun Zhang*, Chengzhang Liu, Xiaoqing Sun, Elayaraja Sivaramasamy, Fuhua Li, Jianhai Xiang**

Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

ARTICLE INFO

Article history:
Received 25 October 2015
Received in revised form 23 November 2015
Accepted 28 November 2015
Available online 22 December 2015

Keywords:
Shrimp
Pancrustacea
Crustacea
Divergence time
Gene family

ABSTRACT

A reference genome is important for genomic and phylogenomic analysis, but except for a brachiopod *Daphnia pulex*, other crustacean genomes have not been completely sequenced. As one of the most economically important crustaceans, decapod shrimps have attracted great attentions worldwide. But the highly complexity block us from completing shrimp genomes. To explore the decapod shrimp genomes, we performed phylogenomic and comparative genomics analyses of pancrustaceans using transcriptome data. A group of 85 single-copy genes were collected from two shrimps and seven genome-available arthropods for phylogenetic analysis. It was found that the two major members of Pancrustacea, Hexapoda and Crustacea, fall into respective monophyletic clades. *D. pulex* and shrimps were estimated divergent at about 395 Myr and the gene family compositions were significantly different between them. The gene families involved in ion binding, peroxidase activity and cytoskeletal part were significantly expanded in shrimps, which were considered tightly associated with the seawater environment adaptation. When comparing the two shrimps, gene families involved in synapse in *Litopenaeus vannamei* and cellular component in *Penaeus monodon* showed significant expansion, which implies the nervous system of *L. vannamei* and cellular structure of *P. monodon* might be adaptively evolved respectively.

1. Introduction

There are four major groups of arthropods: Hexapoda (mainly insects), Crustacea (such as shrimps and crabs), Myriapoda (such as millipedes and centipedes) and Chelicerata (such as spiders and horseshoe crab) (Boore et al., 1998). Most previous phylogenetic analyses suggested that hexapods and crustaceans were phylogenetically close with each other and formed a clade of Pancrustacea (Giribet et al., 2001; Regier et al., 2005). However, the relationships among the constituent lineages of Pancrustacea remain controversial. One immediate question is which group of crustaceans is the closest relative of hexapods. Many results supported Branchiopoda was the closest group of Hexapoda, while Malacostraca (the largest class of Crustacea) occupied a sister position to Hexapoda + Branchiopoda group (Aleshin et al., 2009; Regier et al., 2005). However, some of other evidences did not support Branchiopoda close to Hexapoda, but close to Malacostraca in turn (Giribet et al., 2001; Rota-

* Corresponding author. Institute of Oceanology, Chinese Academy of Sciences, 7 Nanhai Road, Qingdao 266071, China.
** Corresponding author. Institute of Oceanology, Chinese Academy of Sciences, 7 Nanhai Road, Qingdao 266071, China.
E-mail addresses: xjzhang@qdio.ac.cn (X. Zhang), jhxiang@qdio.ac.cn (J. Xiang).

http://dx.doi.org/10.1016/j.bse.2015.11.019
0305-1978/

Stabelli et al., 2010). Therefore, a more consensus phylogenetic tree of Pancrustacea is needed to clarify the relationship of Branchiopoda, Hexapoda and Malacostraca.

With the rapid development of genome sequencing technology, whole-genome approaches become popular for phylogeny inference. Among the approaches, the selection of the conserved single-copy genes was widely used in phylogenomic analysis (Torruella et al., 2012), because single-copy genes are less likely to be confused by paralogous relationships. Besides, as described by De Smet et al. (2013), single-copy genes are not random fractions of the genome, but instead often highly conserved and mostly involved in essential housekeeping functions across all eukaryotes (De Smet et al., 2013). Therefore, single-copy genes are fairly appropriate for phylogenetic analysis.

Insects and crustaceans are two abundant groups of organisms living on the earth (Maia et al., 2013). As for insects, many insect genomes have been sequenced and analyzed extensively in the last decades (Cao et al., 2013; Wang et al., 2014). However, none of crustaceans' genome has been completely sequenced except *Daphnia pulex*, which is a keystone species in Branchiopoda (Colbourne et al., 2011). Genome size of *D. pulex* is about 200 Mb, but it contains 30,970 genes and only 9.4% repetitive sequences. The low complexity of the genome provides a good reference for genome analysis of other crustaceans, such as decapod shrimps, whose genomes are big and complicated for sequencing and assembly (Kenny et al., 2014; Yu et al., 2015). Decapod shrimps are among the most economically important crustaceans, which provide an important global food source (FAO Fishery Statistics, 2011). During the past few years, many researchers have made many efforts on genome sequencing of decapod shrimps. However, none of the shrimp genomes has been ultimately completed because of the large genome size and highly duplication (Yu et al., 2015). Therefore, it is of interest to investigate the genome composition of decapod shrimps. Gene gain and loss were thought to be an important source of molecular evolution, and they are important for the analysis of genome composition (Georgiades et al., 2011). Thus, to understand the genome evolution of decapod shrimps, it is of interest to characterize the gene gain and loss of decapod shrimps.

Although shrimp genomes have not been completed, some published shrimp transcriptomes provide alternative resources of the protein-coding genes. Since most encoded transcripts (98%) are expressed in measured conditions (e.g. various developmental stages and various tissues), the number of contigs assembled by RNA-Seq can cover more than 91% annotated genes of the full genome (Grabherr et al., 2013; Wang et al., 2010). Furthermore, de novo assembly of RNA-Seq data can reconstruct 86% of the annotated genes at full length (Grabherr et al., 2013). Thus, acquiring transcriptomes with higher sequencing depth under more biological conditions can promote the collection of the full genes of a genome (Wang et al., 2010). In this way, some transcriptome data have been successfully used for gene families expansion analysis in previous researches (Christie, 2014), showing the potential of RNA-Seq as valuable resources for the analysis of single-copy genes and gene family expansion.

In this study, we performed a phylogenetic analysis of the pancrustaceans based on single-copy genes collected from complete genomes and transcriptomes. Then, the divergence time of each species was also estimated. Our phylogenetic analysis enabled us to track the gene gain and loss events among each branches. Finally, we investigated the genome evolution of shrimps through analyzing gene gain and loss events between two shrimps and *D. pulex*.

2. Materials and methods

2.1. Sources of genome and transcriptome dataset

The complete genes and genomes of seven arthropods (*Tetranychus urticae*, *Ixodes scapularis*, *D. pulex*, *Anopheles gambiae*, *Drosophila melanogaster*, *Apis mellifera* and *Tribolium castaneum*) were collected from National Center for Biotechnology Information NCBI (www.ncbi.nm.nih.gov), Online resource for community Annotation of Eukaryotes OrcAE (http://bioinformatics.psb.ugent.be/orcae/), Joint Genome Institute JGI (genome.jgi.doe.gov/), and VecterBase (https://www.vectorbase.org/) (Table 1). All the Illumina paired-end transcriptome data of two decapod shrimps, *Litopenaeus vannamei* and *Penaeus monodon*, were collected from our group (SRR1460493, SRR1460494, SRR1460495, SRR1460504, SRR1460505) and the SRA database of NCBI (SRR1039534, SRR1951370, SRR1951373, SRR1951372, SRR1951371, SRR653437, SRR346404,

Table 1
Whole genes of seven arthropods and the transcriptome data of two shrimps.

Species	Class	Genome size	Gene number	Database
Tetranychus urticae	Chelicerata	90 Mb	18,286	OrcAE
Ixodes scapularis	Chelicerata	1.76 Gb	20,486	VecterBase
Anopheles gambiae	Hexapoda	278 Mb	12,659	NCBI
Drosophila melanogaster	Hexapoda	180 Mb	23,948	NCBI
Apis mellifera	Hexapoda	236 Mb	11,736	NCBI
Tribolium castaneum	Hexapoda	160 Mb	9927	NCBI
Daphnia pulex	Crustacea	200 Mb	30,810	JGI
Litopenaeus vannamei	Crustacea	2.50 Gb[b]	17,040[a]	NCBI
Penaeus monodon	Crustacea	2.17 Gb[b]	16,959[a]	NCBI

[a] The gene number of two shrimps that are statistic from the protein-coding unigenes.
[b] The genome size of two shrimps was estimated by flow cytometry (Zhao et al., 2012).