Developmental and Comparative Immunology 48 (2015) 210–220

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Lectin-like molecules in transcriptome of *Littorina littorea* hemocytes

Alexander M. Gorbushin *, Elena A. Borisova

Institute of Evolutionary Physiology and Biochemistry of the Russian Academy of Sciences (IEPhB RAS), St-Petersburg, Russia

ARTICLE INFO

Article history:
Received 27 July 2014
Revised 12 October 2014
Accepted 13 October 2014
Available online 22 October 2014

Keywords:
Innate immunity
Lectins
Hemocytes
Transcriptome
Mollusca

ABSTRACT

The common periwinkle *Littorina littorea* was introduced in the list of models for comparative immunobiology as a representative of phylogenetically important taxon Caenogastropoda. Using Illumina sequencing technology, we *de novo* assembled the transcriptome of *Littorina littorea* hemocytes from 182 million mRNA-Seq pair-end 100 bp reads into a total of 15,526 contigs clustered in 4472 unigenes. The transcriptome profile was analyzed for presence of carbohydrate-binding molecules in a variety of architectural contexts. Hemocytes' repertoire of lectin-like proteins bearing conserved carbohydrate-recognition domains (CRDs) is highly diversified, including 11 of 15 lectin families earlier described in animals, as well as the novel members of lectin family found for the first time in mollusc species. The new molluscan lineage-specific domain combinations were confirmed by cloning and sequencing, including the fuco-lectin related molecules (FLReMs) composed of N-terminal region with no sequence homology to any known protein, a middle Fucolectin Tachylectin-4 Pentaxrin (FTP) domain, and a C-terminal epidermal growth factor (EGF) repeat region. The repertoire of lectin-like molecules is discussed in terms of their potential participation in the receptor phase of immune response. In total, immune-associated functions may be attributed to 70 transcripts belonging to 6 lectin families. These lectin-like genes show low overlap between species of invertebrates, suggesting relatively rapid evolution of immune-associated genes in the group. The repertoire provides valuable candidates for further characterization of the gene functions in mollusc immunity.

1. Introduction

A fundamental problem of comparative immunobiology is the understanding of the diversity of receptor systems involved in the immune pattern recognition and response to xenogenic invaders. Invertebrates' "simple" immune systems only possess innate immunity and an evolutionary archaic mechanism of molecular "self-nonself" discrimination, typically based on lectins binding to specific carbohydrate moieties on the pathogen's surface (Vasta et al., 2007). The carbohydrate-binding activity of proteins usually resides in discrete carbohydrate-recognition domains (CRDs) of lectins, and classification of lectins is based on structural similarities between these domains (Gerardo and Hafiz, 2008; Varki et al., 2009). The recognition functions of the lectins are closely associated with CRDs, while other domains or non-conserved regions in these polypeptides mediate functions in response to the recognition events. The diversity of CRDs in lectins reflects the different functions that the proteins perform.

Lectins of invertebrates are understudied. To date, few published papers partially describe a diversity of carbohydrate-recognizing molecules in Protostomata (Kopacek et al., 2010; Loker, 2011; Wang and Wang, 2013). This is due to insufficient data for adequate comparative analysis, the lack of traceable homology of molluscan lectins to carbohydrate-recognizing molecules of well-studied vertebrates, and the lack of clear homology of structure and functions within invertebrate lectins. In this regard, it is urgent to study and characterize lectins of Mollusca – animals occupying important phylogenetic position within Protostomata.

In the last decade, some Mollusca species have become model objects in comparative immunobiology. The common periwinkle, *Littorina littorea*, was introduced as a model for a phylogenetically important taxon, Caenogastropoda (Loker, 2011). Many parameters of humoral and cell-mediated immunity of the species were studied including plasma hemagglutinating activity (Yakovleva et al., 2001), surface lectins on hemocytes (Iakovleva and Gorbushin, 2005), hemocyte morphotypes and hemopoiesis (Gorbushin and Iakovleva, 2006), the cells' ferment activity, cytotoxic properties and generation of reactive oxygen intermediates (Gorbushin and Iakovleva, 2007), modulation of phagocytosis (Gorbushin and Iakovleva, 2007, 2008; Iakovleva et al., 2006c), receptors expression and intracellular signal transduction (Gorbushin et al., 2009; Iakovleva et al., 2006a,

Abbreviations: EDEM, endoplasmic reticulum degradation-enhancing alpha-mannosidase-like protein; FLReP, fuco-lectin related protein; FLReM, fuco-lectin related molecule; FTP, eel-Fucolectin Tachylectin-4 Pentaxrin-1 Domain; Siglecs, sialic acid binding lectins.

* Corresponding author. Institute of Evolutionary Physiology and Biochemistry of the Russian Academy of Sciences (IEPhB RAS), St-Petersburg, Russia. Tel.: +7 9052717374; fax: +7 812 5523012.
E-mail address: agorbushin@gmail.com (A.M. Gorbushin).

http://dx.doi.org/10.1016/j.dci.2014.10.007
0145-305X/

2006b). Some of these immunological studies focused on interactions between the gastropod and intramolluscan stages of digenetic trematode *Himastla elonagata* (Echinostomatidae) to provide insights on how host responses are subverted by the parasite (Gorbushin and Iakovleva, 2008; Gorbushin et al., 2009; Iakovleva and Gorbushin, 2005; Iakovleva et al., 2006a, 2006c). However, the molecular mechanisms responsible for the observed phenomena have not been characterized due to limited access to molecular biology data on molluscs in general, and in particular on *L. littorea*.

Next generation sequencing (NGS) is emerging as an essential tool to fill such gaps. The comprehensive nucleotide sequence sampling, with indication of abundance levels, makes NGS an effective approach for sequencing transcriptomes and large scale analysis of immune transcriptome data (Dheilly et al., 2014). We have applied the method to reveal the diversity of immune-associated genes expressed in *L. littorea*.

In this paper we describe repertoire of carbohydrate-binding molecules expressed in the common periwinkle hemocytes, the cells that mount both receptor and effector phases of an immune response and exist in the snail circulation system as a single type of hemolymph cells, hyalinocytes (Gorbushin and Iakovleva, 2006). We analyze these lectins in a variety of architectural contexts and suggest the most promising candidates for further immunological studies.

2. Materials and methods

2.1. Hemocyte cDNA preparation and transcriptome sequencing

Hemolymph samples were obtained aseptically from head-foot sinus in the buccal region of untreated snails using a 0.8 mm diameter needle as described previously (Gorbushin and Iakovleva, 2006) and shown in video (http://www.youtube.com/watch?v=u4fFjaBMnOc). Total RNA was extracted from 1.5 ml of hemolymph (appr. 3×10^6 hemocytes) using LS-Trizol Reagent (Invitrogen). RNA quality and concentration were determined by 1% agarose gel electrophoresis and a NanoDrop spectrophotometer. The hemocytes' nonnormalized, full length-enriched cDNA library was prepared from 5 μg of the total RNA from a single *L. littorea* individual using the SMART method (Zhu et al., 2001). The first cDNA chain was synthesized with primers «SMART Oligo II oligonucleotide», «CDS-T22 primer» using Mint-2 cDNA synthesis kit (Evrogen). Amplification of the library was performed with «SMART PCR primer» and Encyclo polymerase (Evrogen) in MJ Research PTC-200 DNA Thermal Cycler (17 cycles). Amplified cDNA was then purified using Evrogen PCR Purification Kit and subjected to quality control with Agilent Bioanalyzer. The cDNA was fragmented by sonication (Covaris) and then prepared for sequencing using Kapa Library Preparation Kit (Kapa Biosystems).

Ligated products were PCR-amplified and sequenced from both 5′ and 3′ ends on an Illumina HiSeq 2000 platform. Raw data of Illumina sequencing were obtained after base calling and stored in fastq format. Cleaning steps of the raw reads were as follows: (1) trimming adapter and SMART-primer sequences; (2) removing the reads that contain ambiguous 'N' nucleotides over 10%; (3) filtering the reads with more than 50% bases having a quality score lower than 5. All subsequent analyses were based on the remaining clean paired and unpaired reads.

2.2. De novo *assembly and transcript abundance estimation*

De novo assembly of full-length transcripts was performed with Trinity software pipeline (http://trinityrnaseq.sourceforge.net/), according to the strategy by Grabherr et al. (2011). In short, Trinity was combined from three independent software modules: Inchworm, Chrysalis, and Butterfly. It segmented sequence data to many individual de Bruijn graphs (each represented transcriptional complexity for a given gene) and processed every graph independently to extract full-length splicing isoforms and to output transcripts from paralogous genes. The k-mer value was set by default to 25. A graphical component from which the collection of transcripts (isoforms) was derived has been named 'unigene'.

To compute the transcript's abundance estimates, the original reads were aligned to the Trinity transcripts database using script included in Trinity kit and Bowtie v1 (Langmead et al., 2009). Then, RSEM (Li and Dewey, 2011) was executed to estimate expression values (FPKM) based on the resulting alignments. 'FPKM' is the number of RNA-seq fragments per kilobase of transcript effective length per million fragments mapped to all transcripts. All assembly and computing operations were performed on a Bio-Linux 7 workstation with Intel i7 8-core processor and 64 GB memory.

2.3. Transcripts data-base filtration and analysis

In order to reduce noise, the most lowly expressed isoforms within a given unigene (FPKM < 1) were removed from the analysis. To eliminate redundancy, only the longest transcript of any single unigene was selected for further analysis to represent the assembled components. Putative coding regions were extracted from Trinity transcripts using TransDecoder (http://transdecoder.sourceforge.net/), which is included in the Trinity software distribution. The transcriptome functional annotation and analysis was performed with Trinotate pipeline (http://trinotate.sourceforge.net/). In short, Trinotate makes use of a number of well referenced methods for functional annotation including a homology search to known sequence data (NCBI-BLAST), protein domain identification (HMMER/PFAM), protein signal prediction (signalP/tmHMM), and comparison to currently curated annotation databases (EMBL Uniprot eggNOG/GO Pathways databases). All functional annotation data derived from the analysis of transcripts were integrated into a SQLite database.

2.4. Search for transcripts coding lectin-like proteins

The lectin prototype set of amino-acid sequences was collected (Table S1) and used as a query to search (tblastn and blastp) in the local database. Additionally, all sequences with conserved amino-acid motifs related to PFAM domains (E-value threshold of 1E–5) and referred as carbohydrate-recognition domains (CRDs) were analyzed manually. Conserved protein domains were confirmed with a CDD (Marchler-Bauer et al., 2011), PFAM (Finn et al., 2014) and SMART (Letunic et al., 2012) domain annotation resources.

2.5. Cloning and sequencing of selected transcripts

The total RNA was extracted from hemocytes with Trizol-reagent (Invitrogen) and double-stranded cDNA was synthesized with RevertAid RT kit (Thermo Scientific). For all RT-PCR applications the DNase I (Thermo Scientific)-treated RNA was used as template to avoid genomic DNA contamination. The lectin transcripts were amplified with gene-specific primers, cloned into pTZ57R/T vector (Thermo Scientific), and sequenced using standard M13 primers on CEQ8000 DNA-sequencer (Beckman Coulter).

3. Results

Using the nonnormalized cDNA library prepared from the *Littorina littorea* hemocytes, 181,572,134 Illumina 100 bp paired-end reads were generated. From these reads, 23,587,168 were first filtered out before assembly due to being low quality sequences or potential contaminations. 137,630,310 paired and 20,354,656 unpaired clean reads were then used to perform *de novo* transcriptome assembly