Developmental and Comparative Immunology 39 (2013) 388–398

Contents lists available at SciVerse ScienceDirect

Developmental and Comparative Immunology

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

A comprehensive analysis of the *Manduca sexta* immunotranscriptome

Ramesh T. Gunaratna, Haobo Jiang *

Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA

ARTICLE INFO

Article history:
Received 9 August 2012
Revised 12 October 2012
Accepted 17 October 2012
Available online 23 November 2012

Keywords:
Insect immunity
Hemolymph proteins
RNA-Seq
Gene discovery
Transcript profiling

ABSTRACT

As a biochemical model, *Manduca sexta* has substantially contributed to our knowledge on insect innate immunity. The RNA-Seq approach was implemented in three studies to examine tissue immunotranscriptomes of this species. With the latest and largest focusing on highly regulated process- and tissue-specific genes, we further analyzed the same set of data using BLAST2GO to explore functional aspects of the larval fat body (F) and hemocyte (H) transcriptomes with (I) or without (C) immune challenge. Using immunity-related sequences from other insects, we found 383 homologous contigs and compared them with those discovered based on relative abundance changes. The major overlap of the two lists validated our previous research designed for gene discovery and transcript profiling in organisms lacking sequenced genomes. By concatenating the contigs, we established a repertoire of 232 immunity-related genes encoding proteins for pathogen recognition (16%), signal transduction (53%), microbe killing (13%) and others (18%). We examined their transcript levels along with attribute classifications and detected prominent differences in nine of the 30 level 2 gene ontology (GO) categories. The increase in extracellular proteins (155%) was consistent with the highly induced synthesis of defense molecules (*e.g.*, antimicrobial peptides) in fat body after the immune challenge. We identified most members of the putative Toll, IMD, MAPK-JNK-p38 and JAK-STAT pathways and small changes in their mRNA levels. Together, these findings set the stage for on-going analysis of the *M. sexta* immunogenome.

1. Introduction

Insects possess a pristine form of the metazoan antimicrobial defense known as innate immunity (Hultmark, 1993), together with a facet of adaptive immunity via phagocyte-mediated immune memory (Pham et al., 2007). However, they lack the luxury of B and T cell-mediated adaptive immunity found in vertebrates (Agaisse, 2007). Insect immunity, comprising humoral and cellular responses, is rapid and effective in identifying and eliminating invading pathogens and parasites (Brey and Hultmark, 1998; Jiang et al., 2010; Lemaitre and Hoffmann, 2007). The general process of insect immunity, before deploying killing mechanisms, consists of pathogen recognition via specific binding molecules (Kurata et al., 2006; Sansonetti, 2006; Yu et al., 2002), signal transduction and modulation via plasma serine proteinases and serine proteinase inhibitors (Gillespie et al., 1997; Kanost, 1999; Kanost et al., 2001; Marmaras and Lampropoulou, 2009), and receptor-mediated intracellular signaling via Toll (Valanne et al., 2011), IMD

Abbreviations: CF and CH, control (C) fat body (F) and hemocytes (H) from naïve larvae; IF and IH, induced (I) fat body and hemocytes from larvae injected with bacteria; Alk, anaplastic lymphoma kinase; AMP, antimicrobial peptide; ANK, ankyrin; aPKC, atypical protein kinase C; ARN, adjusted read number; AtgX, autophagy-related protein X; BP, biological process; CC, cellular component; CRD, carbohydrate recognition domain; CTL, C-type lectin; Dscam, down syndrome cell adhesion molecule; ECSIT, evolutionarily conserved intermediate in Toll pathway; EGF, NIM and EMI, epidermal growth factor, nimrod and emilin; EST, expressed sequence tag; FN, fibronectin; GO, gene ontology; HAIP, hemocyte aggregation inhibitor protein; Hem, hemipterous; HP, hemolymph proteinase; IAP, inhibitor of apoptosis; Ig, immunoglobulin; IKK, IκB kinase; IMD, immune deficiency; IML, immulectin; JAK-STAT, Janus kinase-signal transducer and activator of transcription; JNK, Jun N-terminal kinase; Jra, Jun-related antigen; LNF, library normalization factor; LPS, lipopolysaccharide; LRR, leucine-rich repeat; MAPK, mitogen-activated protein kinase; MASK, multiple ankyrin repeats single KH domain; MEKK, MAP kinase kinase kinase; MF, molecular function; MLK, mixed-linage kinase; NFκB and IκB, nuclear factor-κB and its inhibitor; NO and NOS, nitric oxide and its synthase; NRN, normalized read number; NTF, nuclear translocation; ORF, open reading frame; PAP, proPO-activating proteinase; PDGF and VEGF, platelet-derived and vascular endothelial growth factors; PG and PGRP, peptidoglycan and its recognition protein; PIAS, protein inhibitor of activated STAT; PO and proPO, phenoloxidase and its precursor; POSH, plenty of SH3 domains; PPBP, paralytic peptide-binding protein; PSP, plasmatocyte spreading peptide; PVR, PDGF/VEGF receptor; RA, relative abundance; RISC, RNA interference silencing complex; ROS, reactive oxygen species; SH2/3, src homology 2/3 domain; SOCS, suppressor of cytokine signaling; SP and SPH, serine proteinase and its homolog; SUMO, small ubiquitin-like modifier; TAK, transforming growth factor β (TGFβ) activated kinase; TEP, thioester-containing protein; TIR, Toll/interleukin-1 receptor; TRAF, tumor necrosis factor (TNF) receptor-associated factor; UBC, ubiquitin-conjugating domain; VWD, Von Willebrand disease factor; WAP, whey acidic protein; ZnF, zinc finger; α2M, α2-macrogobulins; βGRP, β-1, 3-glucanas related protein.

* Corresponding author. Tel.: +1(405) 744 9400; fax: +1(405) 744 6039.
E-mail address: haobo.jiang@okstate.edu (H. Jiang).

0145-305X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.dci.2012.10.004

(Silverman and Maniatis, 2001), JNK (Ramet et al., 2002a), JAK-STAT (Baeg et al., 2005; Hou and Perrimon, 1997; Kisseleva et al., 2002), and MAPK-JNK-p38 (Han et al., 1998; Ragab et al., 2011) pathways. Signal transduction regulates both humoral and cellular immune responses. The former includes various antimicrobial peptides (AMPs) (Engstrom, 1999; Jiang, 2008), complement-like molecules (Aoun et al., 2011), and proteins involved in enzyme cascades that regulate melanin formation (Jang et al., 2008; Kanost and Gorman, 2008), which are synthesized and released into the plasma to entrap and kill invading pathogens or parasites (Gillespie et al., 1997; Hoffmann, 2003). In contrast, cellular immunity takes place in hemocytes and is comprised of phagocytosis, nodulation, and encapsulation (Fauvarque and Williams, 2011; Lavine and Strand, 2002; Strand, 2008; Zhuang et al., 2005).

Innate immunity plays a role in making insects the most diverse and abundant group of metazoans in the world (Chapman, 2006; Hultmark, 2003). This makes the immune system worth investigating in its own right. On the other hand, the common ancestry and similarities among insects and mammals make insects excellent model organisms (Hoffmann and Reichhart, 1997; Hultmark, 1993, 2003). These permit discovering evolutionary roots and features of animal immunity (Hoffmann et al., 1999; Khush and Lemaitre, 2000; Vilmos and Kurucz, 1998) and allow functional comparisons between diverse metazoan systems to identify shared and unique aspects of innate immunity (Khush and Lemaitre, 2000; Rolff and Reynolds, 2009; Wajant and Scheurich, 2004).

The advent of microarrays and next generation sequencing technologies coupled with bioinformatics tools has generated a large amount of immunotranscriptome data from insects with known genome sequence, such as *Drosophila* sp. (De Gregorio et al., 2001; Irving et al., 2001; Sackton et al., 2007), *Anopheles gambiae* (Christophides et al., 2002), *Apis mellifera* (Evans et al., 2006), *Aedes aegypti* (Waterhouse et al., 2007), *Tribolium castaneum* (Zou et al., 2007), *Bombyx mori* (Tanaka et al., 2008) and *Acyrthosiphon pisum* (Gerardo et al., 2010). Most of the immunotranscriptomic studies so far, for insects without sequenced genomes, lack quantitative levels of transcripts (Altincicek and Vilcinskas, 2007; Vogel et al., 2011; Zhang et al., 2010). As a member of economically important lepidopterans, *Manduca sexta* has been studied extensively in the field of insect physiology for decades (Jiang et al., 2010). Despite its prominent role, the *M. sexta* genome sequence is not yet published. Recently, transcriptomes of fat body, hemocytes, and midgut, in which many immunity-related genes are expressed, were determined using 454 pyrosequencing and Sanger sequencing technology (Pauchet et al., 2010; Zhang et al., 2011; Zou et al., 2008). The quantitative nature of the most recent study allowed us to analyze immune inducible and tissue specific gene expression. Although genome- and homology-independent discovery of new genes is possible, stringent thresholds set in the exploration hindered complete immunotranscriptomic analysis (Zhang et al., 2011). Therefore, the current work intended to extend the analysis by identifying most of the immunity-related genes in *M. sexta*, as a step towards the annotation of its immunogenome.

2. Methods and materials

2.1. Construction, sequencing, and assembling of cDNA libraries

Insect rearing, bacterial injection, RNA isolation, cDNA synthesis, and library sequencing were described previously (Zhang et al., 2011). Briefly, fat body (F) and hemocytes (H) were prepared as controls (C) from sixty naïve larvae (5th instar, day 3) for total RNA isolation and mRNA purification. Similarly, the same tissues were obtained from sixty induced (I) larvae (5th instar, day 3, injected with a mixture of bacteria 24 h before) for mRNA isolation and cDNA synthesis. After the CF, CH, IF and IH cDNA libraries were separately run on a 454 GS-FLX pyrosequencer, reads were assembled to 19,020 CIFH[1] contigs. For each contig, numbers of the CF, CH, IF and IH reads incorporated were extracted from the Newbler Assembler output and tabulated using Microsoft Excel. As the tissues were pooled from sixty insects, read numbers are expected to faithfully represent the naïve and induced states of fat body and hemocytes.

2.2. Homologous sequence search, GO mapping, annotation and InterProScan search

The contigs were analyzed using the BLAST2GO software (Conesa et al., 2005; Gotz et al., 2008). In search for homologous sequences, the non-redundant protein database at NCBI was searched using BLASTX (Altschul et al., 1990) with a cutoff *E*-value of 10^{-15}. The BLAST hits were mapped to their corresponding GO annotations using the gene ontology database and several additional data files (Gotz et al., 2008). Subsequent annotation of contigs, to link information on cellular component (CC), molecular function (MF), and biological process (BP), was done by applying the annotation rule to all the GO terms. However, certain evidence code weights were changed from their default values to: EXP = IDA = IPI = 5, IMP = IGI = 4, and IEP = 3. Annotations were examined to remove broad or level 1 annotation. Additionally, the GO term known as auxin biosynthesis process was removed from the list of GO terms as the process does not exist in insects. Annex-based GO term augmentation was performed afterwards to, firstly, obtain extra annotations and, secondly, further validate annotations (Gotz et al., 2008; Myhre et al., 2006). Protein domain and signal peptide were predicted using InterProScan (Quevillon et al., 2005), which enabled further sequence annotation (Gotz et al., 2008). In order to obtain more refined annotations, level 1 annotation removal and annex-based GO term augmentation were repeated.

2.3. Local BLASTX, domain search, and multiple sequence alignment

We downloaded immunity-related genes from *Drosophila melanogaster* (462 genes from FlyBase using the keyword "immunity"), *B. mori* (205 genes from Tanaka et al. (2008)), and *A. mellifera* (184 genes from Evans et al., 2006). Amino acid sequences of these genes were incorporated into a sequence database for local BLASTX analysis of the CIFH contigs. Domain prediction was performed in parallel search runs using batchwise domain search web utilities of web CD-search tool, Pfam and InterProScan. Sequence alignments and manual curation of the alignments were performed using MUSCLE (Edgar, 2004) implemented in MEGA 5 (Tamura et al., 2011).

2.4. Calculation of relative abundance of transcripts under immune challenge

Since each contig was assembled from reads in the four libraries, normalized read numbers (NRNs) were calculated as: actual reads number in library $X \times (LNF_{CF} + LNF_{IF} + LNF_{CH} + LNF_{IH})/LNF_x$, where X is CF, IF, CH or IH. Library normalization factors (LNFs) for CF (825), CH (3980), IF (1618) and IH (3352) are the sums of read numbers for rpS2-rpS5, rpL4 and rpL8 in the

[1] To get the CIFH contigs, readers can search the database "CIFH_contigs (March, 2010)" (http://www.darwin.biochem.okstate.edu/blast/blast.html) and retrieve sequences from a text file named "MsextaCIFHcontigs.fna" (http://www.ento-plp.okstate.edu/profiles/jiang.htm). According to GenBank policy, contigs from more than one tissue or treatment are not allowed to be deposited there. Only CF, IF, CH and IH contigs can be found in Transcriptome Shotgun Assembly Sequence Database at GenBank.