Construction of cDNA Library and EST Analysis of *Tenebriomolitor* larvae

JiEun Jeong¹, Se-Won Kang¹, Hee-Ju Hwang¹, Sung-Hwa Chae¹, Sang-Haeng Choi², Hong-Seog Park³, YeonSoo Han⁵, Bok-Reul Lee³, Dae-Hyun Seog⁶ and Yong Seok Lee⁷

**Abstract**—To further advance research on immune-related genes from *T. molitor*, we constructed an DNA library and analyzed expressed sequence tag (EST) sequences from 1,056 clones. After removing vector sequence and quality checking through the Phred program (trim alt 0.05 (P-score>20), 1039 sequences were generated. The average length of insert was 792 bp. In addition, we identified 162 clusters, 167 contigs and 391 contigs after clustering and assembling process using a TGICL package. EST sequences were searched against NCBI nr database by local BLAST (blastx, E<e−10). We identified 976 sequences that showed significant hits with known or unknown gene sequences. To predict the potential functions of the genes, KOG (clusters of orthologous groups for eukaryotic complete genomes) analysis was conducted (blastx, E<e−10). As a result, 621 genes were matched. Of these, most of the genes belonged to Z category (cytokines-related genes). This study will be helpful for screening the innate immune-related genes and signaling cascade.

**Keywords**—EST, Innate immunity, Tenebriomolitor

**I. INTRODUCTION**

In invertebrates, pattern recognition proteins (PRPs) are important components of the invertebrate immune system including hemolymph coagulation and melanization [1].

The insect Toll signaling pathway is activated by recognition of pathogen like bacteria and fungi that induce the expression of antimicrobial peptides in host defense system [2]. Antimicrobial peptides are an evolutionarily conserved component of the innate immune response and they are classified into four groups: cecropins [3], attacins [4], lysozymes [5], and defensins [6].

Recently, many researchers have attempted biochemical studies about antimicrobial defense systems using hemolymphs of infected large beetles [7]. When compared to the Tribolium model, the *Tenebrio* model has relatively large amounts of blood that could be isolated. Nevertheless, the sequence information of the *Tenebrio* genome is still not reported so the sequence information from the *Tribolium* genome is generally used. Here, we constructed a cDNA library and analyzed EST sequences of *Tenebriomolitor* larvae known as a good model of innate immunity research. This transcriptomic sequence data will be helpful for screening the innate immune-related genes and signaling cascade.

**II. MATERIAL AND METHOD**

**A. cDNA library construction**

The total RNA from *T. molitor* larvae was isolated by Trizol reagents (MRC) after being homogenized using TissueLyser (Qiagen) and mRNA was purified from the total RNA using Stratagene Absolutely mRNA Purification Kit (Stratagene, USA). The cDNA library was synthesized using ZAP Express cDNA Synthesis Kit [pBK-CMV vector].

After that, cDNA that has more than 500bp in length was ligated into a pBK-CMV vector and packaged using the GigapackIIIGold packing extract system. The library contained 5.0 X 10⁶ plaque forming units (pfu). Average insert size was determined by an agarose gel (0.8%) electrophoresis run. The average size of insert for the library was 1.8 kb. The entire processing is illustrated in Fig1.

**B. DNA sequencing**

cDNA library clones were cultured in TB medium containing kanamycin (50 mg / ml) at 37 ℃ for 15 hours. Plasmids DNAs were isolated by the mini-preparation method with MultiScreen 96-well Filter Plates and amplified by the dye-terminator cycling method (BigDye v3.1) using 5’ universal primers in the vector (Stratagen). Single pass sequencing was performed by ABI3730 XL capillary sequencer.

![Fig. 1 Schematic diagrams of cDNA library construction](image)

**C. Sequence analysis**

The nucleotide sequences of 1,056 clones were determined by 5’ end-single path sequencing from cDNA libraries. The base calling and quality assignment of the chromatogram files were conducted with the Phred program [8, 9]. In addition, the vector trimming was performed with cross-match software. The trace files were trimmed with trim-alt 0.05
These pre-treated sequences were analyzed by clustering (30 bp or more 94% homology) and assembly. Finally, the contigs and singletons were searched against the NCBI local BLAST.

D. EST Annotation

All sequences were prepared as a multi-fasta format for sequence analysis and the annotation was performed through local BLAST search [10]. The function of genes was analyzed by KOG (Clusters of orthologous groups)[11].

III. RESULT AND DISCUSSION

A total of 1,056 clones was randomly selected from the Tenebriomolitor cDNA library and determined 5’-end sequence by single-pass sequencing strategy. All the sequences were trimmed: the vector sequences and the low quality sequences. We identified 1,039 high quality ESTs of 792 bp on average as shown in Table 1.

As a result of the clustering and assembling the ESTs sequences using TGICL package[12], we identified 558 distinct sequences composed of 391 singletons and 167 contigs in 162 clusters.

All the sequences were compared against the NCBI Non-Redundant Database using the BLASTX algorithm. The E-value of < 1e-5 were used as a threshold against BLASTX. As a result, 981 of the sequences had significant matches in the database (Table 1).

In order to predict the gene function, KOG (Clusters of orthologous groups for eukaryotic complete genomes) analysis was conducted [11]. The gene function of the sequences were predicted through the local BLAST (blastx, E <e-10) search against the KOG database. 621 sequences had significant hits and were classified into the following 22 categories as represented in Figure 2.

The Z category that belongs to cytoskeleton-related genes showed the higher expression level among them. Total sequences were classified into as follows: Translation, ribosomal structure and biogenesis (7.09%), RNA processing and modification (2.9%), Transcription (2.58%), Replication, recombination and repair (0.48%), Chromatin structure and dynamics (0.64%), Cell cycle control, cell division, chromosome partitioning (0.97%), Defense mechanisms (0.48%), Signal transduction mechanisms (6.28%), Cell wall/membrane/envelope biogenesis (0.32%), Cell motility (0.32%), Cytoskeleton (21.42%), Extracellular structures (1.29%), Intracellular trafficking, secretion, and vesicular transport (3.38%), Posttranslational modification, protein turnover, chaperones (7.57%), Energy production and conversion (9.34%), Carbohydrate transport and metabolism (1.93%), Amino acid transport and metabolism (1.77%), Nucleotide transport and metabolism (0.32%), Coenzyme transport and metabolism (0.64%), Lipid transport and metabolism (2.25%), Inorganic ion transport and metabolism (2.25%), Secondary metabolites biosynthesis, transport and catabolism (0.64%), General function prediction only (10.79%), Function unknown (5.8%), 2 Category (more than double (8.53%).

| T. molitor |
|---|---|
| Number of clone that sequenced | 1056 |
| Number of clone that used for sequence analysis after sequencing | 1039 |
| (basecalling(Phred≥20)/vectormasking/ESTs≥100bp) | |

| Clustering & Assembling |
|---|---|
| Number of Clusters | 162 |
| Number of Contigs | 167 |
| Number of Singletons | 391 |
| Annotation Against NCBI nr Database | 976 |
| Against KOG Database | 621 |

Fig. 2 KOG analysis results of EST sequences in Tenebriomolitor

Code descriptions of KOG : J (Translation, ribosomal structure and biogenesis), A (RNA processing and modification), K (Transcription), L (Replication, recombination and repair), B (Chromatin structure and dynamics), D (Cell cycle control, cell division, chromosome partitioning), V (Defense mechanisms), T (Signal transduction mechanisms), M (Cell wall/membrane/envelope biogenesis), N (Cell motility), Z (Cytoskeleton), W (Extracellular structures), U (Intracellular trafficking, secretion, and vesicular transport), O (Posttranslational modification, protein turnover, chaperones), C (Energy production and conversion), G (Carbohydrate transport and metabolism), E (Amino acid transport and metabolism), F (Nucleotide transport and metabolism), H (Coenzyme transport and metabolism), I (Lipid transport and metabolism), P (Inorganic ion transport and metabolism), Q (Secondary metabolites biosynthesis, transport and catabolism), R (General function prediction only), S (Function unknown)

We constructed high quality cDNA library (library titer = 5.0 × 10^6 pfu/ml) and ESTs database as following site
http://blast.inje.ac.kr/~tenebrio/. Also, we identified several putative genes such as serpin [3] and tenecin 3 [13] involved in the regulation of Toll signaling pathway. This study will be very useful in screening for new and potential immune-related genes and allow for further investigation into the biological function of these genes in conjunction with insect immunity and signaling cascade.

IV. CONCLUSION

Tenebriomolitor has been intensively studied as a model insect to elucidate its melanization mechanism and signaling cascades involved in innate immunity. We identified several putative genes involved in insect innate immunity through cDNA Library construction and ESTs analysis of T. molitor larvae. Also, transcriptomic database construction will be useful to study signaling pathway of innate immunity.

ACKNOWLEDGMENT

This work was supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ008186012011), Rural Development Administration in Republic of Korea.

REFERENCES