International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:5, No:11, 2011

Construction of cDNALibrary and EST Analysis of *Tenebriomolitor*larvae

JiEun Jeong[1], Se-Won Kang[1], Hee-Ju Hwang[1], Sung-Hwa Chae[2], Sang-Haeng Choi[3], Hong-Seog Park[3]YeonSoo Han[4], Bok-Reul Lee[5], Dae-Hyun Seog[6*†]and Yong Seok Lee [1*†]

***Abstract*—**Tofurther advance research on immune-related genes from *T. molitor*, we constructed acDNA library and analyzed expressed sequence taq (EST) sequences from 1,056 clones. After removing vector sequence and quality checkingthrough thePhred 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 searchedagainst NCBI nr database by local BLAST (blastx, E<e-5). We identified 976 sequences that showed significant hits with known or unknown gene sequence. 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 (cytoskeleton-related genes). This studywill 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 thet 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. Thistranscriptomic sequence data will be helpful for screening the innate immune-related genes and signaling cascade.

J.E. Jeong is research professor in Inje University and UHRC, Busan, Korea
S.W. Kang and H.J. Hwang are researching in Inje University and UHRC, Busan, Korea
S.H. Chae is working in Research Institute, GnC BIO Co., LTD., Daejeon, Korea
S. H. Choi and H. S. Park are working at Genome Resource Center, KRIBB, Daejeon, Korea
Y. S. Han is a professor of the College of Agriculture and Life Science, Chonnam National University, Gwangju, Korea.
B. L. Lee is a professor of College of Pharmacy, Pusan National University
D. H. Seog, one of corresponding authors, is a professor of the Department of Biochemistry, College of Medicine, Inje University,Busan, Korea(e-mail: daehyun@inje.ac.kr)
Y.S. Lee, another of corresponding authors, is a professor of the Department of Parasitology, College of Medicine, Inje University, Busan 614-735, Korea (e-mail: yslee@inje.ac.kr)

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 10^5 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 (BigDyev3.1) using 5' universal primers in the vector (Stratagen). Single pass sequencing was performed by ABI3730 XL capillary sequencer.

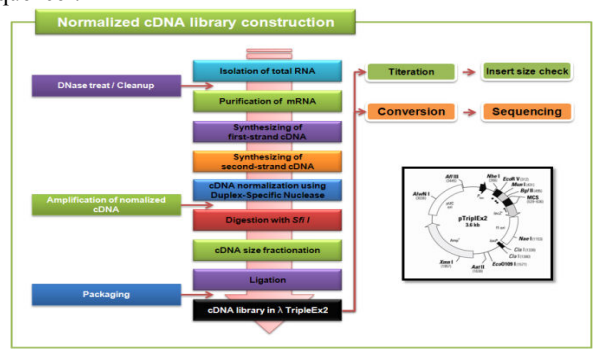

Fig. 1 Schematic diagrams of cDNAlibrary construction

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

(P-score>20).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 (Table1).

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%).

TABLE I
THE GENERAL INFORMATION OF EST SEQUENCES

	T. molitor
Sequence Analysis	
Number of clone that sequenced	1056
Number of clone that used for sequence analysis after sequencing (basecalling(Phred≥20)/vectormasking/ESTs≥100bp)	1039
Clustering & Assembling	
Number of Clusters	162
Number of Contigs	167
Number of Singletons	391
Annotation	
Against NCBI nr Database	976
Against KOG Database	621

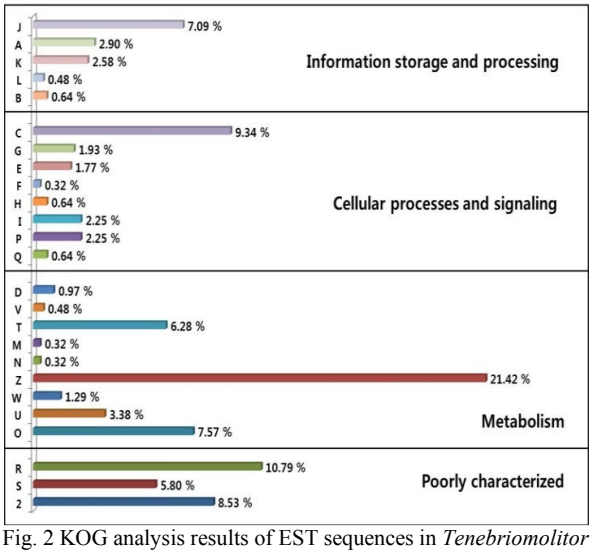

Fig. 2 KOG analysis results of EST sequences in *Tenebriomolitor*

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 × 10^6pfu/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 genesinvolved in insect innate immunity throughcDNA Library construction andESTs analysis of *T. molitor* larvae. Also, transcriptomicdatabase 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

[1] J. W. Park, et al., "A synthetic peptidoglycan fragment as a competitive inhibitor of the melanization cascade," J Biol Chem, vol. 281, pp. 7747-55, Mar 24 2006.
[2] K. B. Roh, et al., "Proteolytic cascade for the activation of the insect toll pathway induced by the fungal cell wall component," J Biol Chem, vol. 284, pp. 19474-81, Jul 17 2009.
[3] H. Steiner, et al., "Sequence and specificity of two antibacterial proteins involved in insect immunity," Nature, vol. 292, pp. 246-8, Jul 16 1981.
[4] D. Hultmark, et al., "Insect immunity. Attacins, a family of antibacterial proteins from Hyalophora cecropia," EMBO J, vol. 2, pp. 571-6, 1983.
[5] H. G. Boman, et al., "Insect immunity--a gene system for antibacterial proteins," Mem Inst Oswaldo Cruz, vol. 82 Suppl 3, pp. 115-24, 1987.
[6] J. Lambert, et al., "Insect immunity: isolation from immune blood of the dipteran Phormia terranovae of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides," Proc Natl Acad Sci U S A, vol. 86, pp. 262-6, Jan 1989.
[7] R. Jiang, et al., "Three pairs of protease-serpin complexes cooperatively regulate the insect innate immune responses," J Biol Chem, vol. 284, pp. 35652-8, Dec 18 2009.
[8] B. Ewing, et al., "Base-calling of automated sequencer traces using phred. I. Accuracy assessment," Genome Res, vol. 8, pp. 175-85, Mar 1998.
[9] B. Ewing and P. Green, "Base-calling of automated sequencer traces using phred. II. Error probabilities," Genome Res, vol. 8, pp. 186-94, Mar 1998.
[10] S. F. Altschul, et al., "Basic local alignment search tool," J Mol Biol, vol. 215, pp. 403-10, Oct 5 1990.
[11] R. L. Tatusov, et al., "The COG database: an updated version includes eukaryotes," BMC Bioinformatics, vol. 4, p. 41, Sep 11 2003.
[12] G. Pertea, et al., "TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets," Bioinformatics, vol. 19, pp. 651-2, Mar 22 2003.
[13] D. H. Kim, et al., "Internalization of tenecin 3 by a fungal cellular process is essential for its fungicidal effect on Candida albicans," Eur J Biochem, vol. 268, pp. 4449-58, Aug 2001.