

# Expressed sequence tags analysis of *Blattella germanica*

Hyang Suk CHUNG<sup>1</sup>), Tai Hyun YU<sup>1</sup>), Bong Jin KIM<sup>2</sup>), Sun Mi KIM<sup>2</sup>), Joo Yeong KIM<sup>2</sup>),  
Hak Sun YU<sup>3</sup>), Hae Jin Jeong<sup>3</sup>) and Mee Sun OCK<sup>2</sup>)\*

<sup>1</sup>)Department of Otorhinolaryngology and <sup>2</sup>)Department of Parasitology, Kosin University College of Medicine, Busan 602-703,  
<sup>3</sup>)Department of Parasitology, Pusan National University College of Medicine, Busan, 602-739, Korea

**Abstract:** Four hundred and sixty five randomly selected clones from a cDNA library of *Blattella germanica* were partially sequenced and searched using BLAST as a means of analyzing the transcribed sequences of its genome. A total of 363 expressed sequence tags (ESTs) were generated from 465 clones after editing and trimming the vector and ambiguous sequences. About 42% (154/363) of these clones showed significant homology with other data base registered genes. These new *B. germanica* genes constituted a broad range of transcripts distributed among ribosomal proteins, energy metabolism, allergens, proteases, protease inhibitors, enzymes, translation, cell signaling pathways, and proteins of unknown function. Eighty clones were not well-matched by database searches, and these represent new *B. germanica*-specific ESTs. Some genes which drew our attention are discussed. The information obtained increases our understanding of the *B. germanica* genome.

**Key words:** *Blattella germanica*, cDNA library, ESTs, BLAST search, novel genes

## INTRODUCTION

The cockroach is among the oldest winged insects known, and its habits are closely associated with those of humans. Over four thousand species of cockroach are known, and about thirty species are harmful to humans in various ways. The importance of the German cockroach has been emphasized because it is the most populous and has the widest distribution (Ross and Cochran, 1975). *Blattella germanica* is also a well known cause of allergic diseases, rather than acting as a vector of infectious diseases (Richman et al., 1984).

Previous genetic studies on *B. germanica* have been limited to the study for some of its genes, e.g., allergen

genes (Arruda et al 1995; Helm et al., 1996) and the cytochrome P450 gene, which is related with juvenile hormone or insecticide tolerance (Martinez-Gonzalez and Hegardt 1994; Scharf et al., 1998).

The generation and analysis of expressed sequence tags (ESTs) provides useful information on development, metabolism, virulence factors, drug targets, and pathogenesis in various organisms (el-Sayed et al., 1995; Wu et al., 1996; Manger et al., 1998; Bahl et al., 2003).

To understand more about the expression pattern of its genome, we generated ESTs from a cDNA library of *B. germanica*. The analysis of such data provides valuable insights into the metabolism and growth of German cockroach.

• Received 19 August 2005, accepted after revision 8 November 2005.

\*Corresponding author (e-mail: sunnyock@kosin.ac.kr)

## MATERIALS AND METHODS

### Cockroach breeding

Adult male and female German cockroaches were reared at 24°C on an artificial diet, and given free access to water

### cDNA library construction

A *B. germanica* cDNA library was constructed using Uni-ZAP™-XR expression vector (Stratagene, USA). In brief, total RNA was isolated from 3 g of the midgut of adult *B. germanica*. After phenol extraction and ethanol precipitation, poly(A<sup>+</sup>) RNA was purified using a Stratagene Poly(A) Quick mRNA Isolation Kit, in accordance with the manufacturer's instructions. First-strand cDNA synthesis of the isolated poly(A<sup>+</sup>) RNA was then conducted in 50 µl reaction volumes, using 50 units of MMLV-reverse transcriptase at 37°C for 60 minutes. cDNA synthesis was primed using 5 µg oligo dT18. Second-strand synthesis was then conducted using RNase H and DNA polymerase I. After blunting the cDNA termini, *EcoR* I adaptor ligation and *EcoR* I phosphorylation were performed. Gel regions containing DNA molecules of length <400bp were then removed by Sepharose CL-2B gel filtration. Purified cDNA was ligated using dephosphorylated *EcoR* I Uni-ZAP™-XR vector arms, according to the manufacturer's instructions (Stratagene, USA), and then incubated using in vitro packaging extracts (Stratagene, USA).

### Sequencing of randomly selected cDNA clones

Colonies from *E. coli* XL-1 Blue MRF cells harboring plasmid were obtained en masse by in vivo excision using assistant helper phage. Random recombinant clones were selected by blue-white color selection of colonies grown on LB-ampicillin agar plates. Plasmids containing the cDNA insert were extracted using a Wizard plasmid DNA purification system (Promega, Madison, WI, USA), and the existence of cDNA inserts was confirmed by gel electrophoresis after double digestion with *EcoR* I and *Xho* I. cDNA inserts were sequenced at DNA Sequencing Service (Macrogen, Seoul, Korea).

### Basic local alignment search tool (BLAST) search analysis

Sequence outputs were manually edited to remove vector and ambiguous sequences. Sequence outputs of <100 bases in length were also rejected. The sequence data of cDNA clones obtained by random partial sequencing were searched for using BLAST at the National Center for Biotechnology Information (NCBI) for similarities in nucleic acid and protein databases. The BLASTN algorithm was used in conjunction with a nucleotide sequence database with a probability (P) cut-off of 10<sup>-4</sup>. Matches of translational products versus nucleic acid sequences search for using the BLASTX algorithm with a probability (P) cut-off of 10<sup>-4</sup>. Scores >160 for BLASTN or >80 for BLASTX were considered significant.

## RESULTS

The submitted ESTs for BLAST searching comprised 363 ESTs from 465 randomly selected clones of the cDNA library of *B. germanica*. Clones of <100 bp in length or not successfully sequenced were excluded (102 clones). The average size of the 363 ESTs was 604 bp. These 363 sequenced clones were divided into three groups based on matches with public data sequences (Table 1). The matched with database group of 154 clones, showed high homology with the DNA sequence of *B. germanica* or other organisms. Of these clones, 10 ESTs corresponded to 3 previously identified *B. germanica* genes, including cytochrome c

**Table 1.** Composition and ESTs categories of *Blattella germanica* cDNA library

| Group  | No. of clones |
|--|---------------|
| Total clones sequenced   | 465           |
| ESTs submitted to dbEST database                                   | 363           |
| Match to database  | 154           |
| Clone with homology to <i>B. germanica</i><br>(redundant clones 7) | 10            |
| Clone with homology to other<br>organisms (redundant clones 34)    | 144           |
| Non-match to database  | 80            |
| Non-significant clones   | 129           |

**Table 2.** Database match of *Blattella germanica* EST to the genes of the other organisms

| Clone No                 | Length | Accession No | Putative homologue                     | Organism                         | Score | P(N)     |
|--------------------------|--------|--------------|--|----------------------------------|-------|----------|
| <b>Ribosomal protein</b> |        |              |  |                                  |       |          |
| Bg11019                  | 831    | NP_524726    | ribosomal protein L8                   | <i>Drosophila melanogaster</i>   | 83    | 4.00E-15 |
| Bg9035                   | 659    | AAL26575     | ribosomal protein L8                   | <i>Spodoptera frugiperda</i>     | 311   | 3.00E-84 |
| Bg9065                   | 644    | AAK76989     | ribosomal protein L9                   | <i>Spodoptera frugiperda</i>     | 327   | 4.00E-89 |
| Bg3103                   | 381    | AAK83857     | ribosomal protein L17/23               | <i>Spodoptera frugiperda</i>     | 154   | 2.00E-37 |
| Bg10031                  | 863    | AAL62470     | ribosomal protein L18A                 | <i>Spodoptera frugiperda</i>     | 259   | 3.00E-68 |
| Bg9046                   | 286    | AAL26577     | ribosomal protein L29                  | <i>Spodoptera frugiperda</i>     | 108   | 1.00E-23 |
| Bg9031                   | 443    | CAC19413     | ribosomal protein L31                  | <i>Heliothis virescens</i>       | 186   | 5.00E-47 |
| Bg11023                  | 504    | AAK92169     | ribosomal protein L35A                 | <i>Spodoptera frugiperda</i>     | 163   | 8.00E-40 |
| Bg9025                   | 354    | AAK92172     | ribosomal protein L37A                 | <i>Spodoptera frugiperda</i>     | 155   | 8.00E-38 |
| Bg12004                  | 520    | NP_476874    | ribosomal protein S2                   | <i>Drosophila melanogaster</i>   | 226   | 1.00E-58 |
| Bg11037                  | 829    | AAL26579     | ribosomal protein S3A                  | <i>Spodoptera frugiperda</i>     | 348   | 5.00E-95 |
| Bg9014                   | 562    | NP_524884    | ribosomal protein S14                  | <i>Drosophila melanogaster</i>   | 212   | 2.00E-54 |
| Bg11059                  | 951    | AAK92190     | ribosomal protein S21                  | <i>Spodoptera frugiperda</i>     | 57    | 2.00E-07 |
| Bg7009                   | 357    | P47991       | 60S ribosomal protein L6               | <i>Caenorhabditis elegans</i>    | 73    | 1.00E-13 |
| Bg9060                   | 295    | P32429       | 60S ribosomal protein L7A              | <i>Gallus gallus</i>             | 75    | 2.00E-13 |
| Bg8043                   | 937    | O96647       | 60S ribosomal protein L10              | <i>Bombyx mandarina</i>          | 131   | 2.00E-31 |
| Bg4022                   | 563    | P46222       | 60S ribosomal protein L11              | <i>Drosophila melanogaster</i>   | 303   | 2.00E-82 |
| Bg8048                   | 407    | P41126       | 60S ribosomal protein L13              | <i>Drosophila melanogaster</i>   | 119   | 1.00E-27 |
| Bg8031                   | 619    | P41093       | 60S ribosomal protein L18A             | <i>Drosophila melanogaster</i>   | 249   | 7.00E-72 |
| Bg6016                   | 832    | P36241       | 60S ribosomal protein L19              | <i>Drosophila melanogaster</i>   | 164   | 2.00E-40 |
| Bg4018                   | 504    | P23131       | 60S ribosomal protein L23              | <i>Homo Sapiens</i>              | 215   | 3.00E-56 |
| Bg8025                   | 550    | Q02877       | 60S ribosomal protein L26              | <i>Homo Sapiens</i>              | 133   | 2.00E-31 |
| Bg7021                   | 505    | P46615       | 60S ribosomal protein L32              | <i>Drosophila pseudoobscura</i>  | 234   | 7.00E-62 |
| Bg7041                   | 514    | P02433       | 60S ribosomal protein L32              | <i>Homo Sapiens</i>              | 167   | 8.00E-42 |
| Bg9006                   | 888    | AAK921       | 60S ribosomal protein L35              | <i>Spodoptera frugiperda</i>     | 114   | 1.00E-24 |
| Bg8001                   | 352    | Q96257       | 60S ribosomal protein L37              | <i>Spodoptera frugiperda</i>     | 114   | 4.00E-26 |
| Bg8015                   | 572    | P05389       | 60S acidic ribosomal protein P2        | <i>Drosophila melanogaster</i>   | 94    | 1.00E-19 |
| Bg5006                   | 819    | P52813       | 40S ribosomal protein S3A              | <i>Anopheles gambiae</i>         | 268   | 7.00E-81 |
| Bg9071                   | 221    | P55830       | 40S ribosomal protein S3A              | <i>Drosophila melanogaster</i>   | 64    | 3.00E-10 |
| Bg10029                  | 947    | P02350       | 40S ribosomal protein S3A              | <i>Xenopus laevis</i>            | 147   | 3.00E-36 |
| Bg11001                  | 890    | P47835       | 40S ribosomal protein S3B              | <i>Xenopus laevis</i>            | 285   | 7.00E-79 |
| <b>Energy metabolism</b> |        |              |  |                                  |       |          |
| Bg7011                   | 871    | P33502       | NADH-Ubiquinone oxidoreductase chain 1 | <i>Anopheles quadrimaculatus</i> | 273   | 4.00E-73 |
| Bg8035                   | 848    | P29867       | NADH-Ubiquinone oxidoreductase chain 2 | <i>Drosophila mauritiana</i>     | 119   | 6.00E-27 |
| Bg9045                   | 899    | Q34048       | NADH-Ubiquinone oxidoreductase chain 4 | <i>Ceratitidis capitata</i>      | 248   | 5.00E-65 |
| Bg7025                   | 510    | Q34050       | NADH-Ubiquinone oxidoreductase chain 6 | <i>Ceratitidis capitata</i>      | 112   | 4.00E-25 |
| Bg1003                   | 564    | P07704       | cytochrome b                           | <i>Drosophila yakuba</i>         | 241   | 4.00E-64 |
| Bg9056                   | 836    | AAG17094     | cytochrome b                           | <i>Bifiditermes improbus</i>     | 217   | 2.00E-81 |
| Bg10006                  | 942    | AAG17097     | cytochrome b                           | <i>Cryptotermes cynocephalus</i> | 230   | 1.00E-59 |
| Bg6013                   | 905    | P00400       | cytochrome c oxidase polypeptide I     | <i>Drosophila yakuba</i>         | 207   | 1.00E-53 |
| Bg7003                   | 219    | P00399       | cytochrome c oxidase polypeptide I     | <i>Drosophila melanogaster</i>   | 51    | 5.00E-07 |
| Bg8054                   | 123    | P50671       | cytochrome c oxidase polypeptide I     | <i>Choristoneura rosaceana</i>   | 42    | 2.00E-04 |
| Bg5016                   | 695    | P29877       | cytochrome c oxidase polypeptide II    | <i>Periplaneta americana</i>     | 269   | 3.00E-72 |

Table 2. continued

| Clone No                            | Length | Accession No | Putative homologue                                  | Organism                        | Score | P(N)     |
|-------------------------------------|--------|--------------|---|---------------------------------|-------|----------|
| Bg8064                              | 150    | P98048       | cytochrome c oxidase polypeptide II                 | <i>Yponomeuta malinellus</i>    | 62    | 2.00E-10 |
| Bg8068                              | 379    | P29877       | cytochrome c oxidase polypeptide II                 | <i>American cockroach</i>       | 173   | 5.00E-44 |
| Bg4013                              | 783    | P14574       | cytochrome c oxidase polypeptide III                | <i>Locusta migratoria</i>       | 223   | 4.00E-58 |
| Bg8045                              | 883    | P00417       | cytochrome c oxidase polypeptide III                | <i>Drosophila melanogaster</i>  | 221   | 1.00E-57 |
| Bg12010                             | 983    | AAB31450     | cytochrome c oxidase subunit I                      | <i>Blattella germanica</i>      | 203   | 3.00E-51 |
| Bg3202                              | 387    | AAF89137     | cytochrome oxidase subunit III                      | <i>Cicindela belfragei</i>      | 152   | 9.00E-37 |
| Bg10013                             | 740    | AAG01168     | cytochrome oxidase subunit III                      | <i>Samia cynthia ricini</i>     | 229   | 2.00E-59 |
| Bg3209                              | 430    | BAA32127     | cytochrome oxidase II                               | <i>Blattella germanica</i>      | 244   | 9.00E-65 |
| Bg7024                              | 795    | Q9V4U9       | cytochrome P450 6a13                                | <i>Drosophila melanogaster</i>  | 120   | 3.00E-27 |
| <b>Allergen</b>                     |        |              |   |                                 |       |          |
| Bg8050                              | 287    | AAB82404     | Cr-PII  | <i>Periplaneta americana</i>    | 58    | 2.00E-08 |
| Bg10001                             | 294    | AAC34737     | Cr-PII allergen                                     | <i>Periplaneta americana</i>    | 58    | 2.00E-08 |
| Bg1010                              | 340    | AAD13530     | major allergen Blag1.0101                           | <i>Blattella germanica</i>      | 121   | 2.00E-27 |
| Bg7008                              | 365    | AAD13532     | major allergen Blag1.0101                           | <i>Blattella germanica</i>      | 135   | 1.00E-31 |
| <b>Protease</b>                     |        |              |   |                                 |       |          |
| Bg6009                              | 408    | P35035       | Trypsin 1 precursor                                 | <i>Anopheles gambiae</i>        | 123   | 1.00E-28 |
| Bg3106                              | 241    | P35036       | Trypsin 2 precursor                                 | <i>Anopheles gambiae</i>        | 87    | 3.00E-17 |
| Bg4101                              | 439    | S35339       | trypsin (EC 3.4.21.4) 1 precursor                   | <i>Anopheles gambiae</i>        | 123   | 4.00E-28 |
| Bg11002                             | 384    | AAD31269     | trypsinogen Rdo T3 precursor                        | <i>Rhyzopertha dominica</i>     | 130   | 2.00E-30 |
| Bg3109                              | 382    | P04069       | Carboxypeptidase B                                  | <i>Astacus astacus</i>          | 84    | 3.00E-16 |
| Bg11072                             | 430    | 1EQ9A        | Chain A, Crystal Structure Of Fire Ant Chymotrypsin | <i>Solenopsis invicta</i>       | 86    | 7.00E-17 |
| Bg11049                             | 534    | AAA97479     | Astryp1   | <i>Anopheles stephensi</i>      | 129   | 2.00E-29 |
| <b>Enzyme related to metabolism</b> |        |              |   |                                 |       |          |
| Bg4008                              | 607    | Q9Y600       | Cysteine sulfinic acid decarboxylase                | <i>Homo sapiens</i>             | 57    | 3.00E-08 |
| Bg6014                              | 838    | Q59296       | Catalase  | <i>Campylobacter jejuni</i>     | 67    | 3.00E-11 |
| Bg8004                              | 757    | P26221       | Endoglucanase E-4 precursor                         | <i>Thermobi fidafusca</i>       | 134   | 1.00E-31 |
| Bg9010                              | 847    | S41881       | alpha-amylase (EC 3.2.1.1) 1 precursor              | <i>Litopenaeus vannamei</i>     | 192   | 4.00E-48 |
| Bg9051                              | 607    | BAB91145     | beta-glucosidase                                    | <i>Neotermes koshunensis</i>    | 82    | 3.00E-15 |
| Bg6015                              | 442    | P49010       | beta-N-acetylglucosaminidase precursor              | <i>Bombyx mori</i>              | 120   | 8.00E-28 |
| Bg9015                              | 949    | P18173       | Glucosedehydrogenase                                | <i>Drosophila melanogaster</i>  | 100   | 2.00E-20 |
| Bg9043                              | 589    | JC4081       | surcease/fructanase precursor                       | <i>Actinomyces naeslundii</i>   | 52    | 3.00E-06 |
| Bg11053                             | 924    | AAC79122     | alpha-amylase                                       | <i>Drosophila ananassae</i>     | 195   | 6.00E-49 |
| Bg10043                             | 869    | A34406       | aldehydereductase (EC 1.1.1.21)                     | <i>Oryctolagus cuniculus</i>    | 122   | 6.00E-27 |
| Bg9033                              | 835    | AAB61345     | lysozyme  | <i>Anopheles darlingi</i>       | 69    | 6.00E-11 |
| Bg10002                             | 453    | BAB33297     | Esterase-like protein (ESR-LP)                      | <i>Bombyx mori</i>              | 67    | 8.00E-11 |
| <b>Protease inhibitor</b>           |        |              |   |                                 |       |          |
| Bg5047                              | 340    | Q06684       | Rhodniin (Thrombin inhibitor)                       | <i>Rhodnius prolixus</i>        | 58    | 3.00E-09 |
| Bg9028                              | 404    | S45677       | proteinase inhibitor                                | <i>Pacifastacus leniusculus</i> | 44    | 4.00E-04 |
| Bg10033                             | 1026   | AAK57342     | thrombin inhibitor infestin precursor               | <i>Triatoma infestans</i>       | 64    | 2.00E-09 |
| <b>Translation</b>                  |        |              |   |                                 |       |          |
| Bg9029                              | 888    | NP_524611    | elongation factor 1 alpha 100E                      | <i>Drosophila melanogaster</i>  | 233   | 3.00E-60 |
| Bg9007                              | 377    | P29522       | elongation factor 1-β                               | <i>Bombyx mori</i>              | 66    | 5.00E-41 |
| Bg12018                             | 758    | BAB21109     | elongation factor 1 delta                           | <i>Bombyx mori</i>              | 130   | 1.00E-29 |
| Bg5012                              | 486    | Q9VL18       | elongation factor 1-delta                           | <i>Drosophila melanogaster</i>  | 67    | 2.00E-11 |

Table 2. continued

| Clone No                      | Length | Accession No | Putative homologue                              | Organism                       | Score | P(N)     |
|-------------------------------|--------|--------------|---|--------------------------------|-------|----------|
| <b>Cell signaling pathway</b> |        |              |   |                                |       |          |
| Bg11067                       | 384    | Q09966       | Putative G protein-coupled receptor B0244.7     | <i>Caenorhabditis elegans</i>  | 30    | 3.9      |
| <b>Others</b>                 |        |              |   |                                |       |          |
| Bg8028                        | 901    | P14792       | Ubiquitin                                       | <i>Caenorhabditis elegans</i>  | 109   | 5.00E-14 |
| Bg10044                       | 322    | NP_476776    | Ubiquitin fusion 52                             | <i>Drosophila melanogaster</i> | 140   | 3.00E-33 |
| Bg5014                        | 484    | P22943       | 12kDa heat shock protein                        | <i>Saccharomy cerevisiae</i>   | 94    | 1.00E-19 |
| Bg7040                        | 829    | P41822       | ferritin subunit precursor                      | <i>Aedes aegypti</i>           | 82    | 8.00E-16 |
| Bg9050                        | 583    | NP_523683    | Peroxiredoxin2540                               | <i>Drosophila melanogaster</i> | 84    | 2.00E-19 |
| Bg7005                        | 492    | O43653       | Prostate stem cell antigen precursor            | <i>Homo Sapiens</i>            | 42    | 5.00E-04 |
| Bg4010                        | 359    | P40618       | High mobility group protein 4 (HMG-4)           | <i>Gallus gallus</i>           | 55    | 3.00E-08 |
| Bg3208                        | 433    | AAH10444     | matrilin2                                       | <i>Homo Sapiens</i>            | 49    | 1.00E-05 |
| Bg8012                        | 395    | AAM21357     | mucin-like protein 1                            | <i>Ctenocephalides felis</i>   | 49    | 8.00E-06 |
| Bg8016                        | 842    | O76767       | ER lumen protein retaining receptor             | <i>Drosophila melanogaster</i> | 140   | 3.00E-33 |
| Bg8039                        | 447    | Q27377       | odorant-binding protein A10 precursor           | <i>Drosophila melanogaster</i> | 5     | 1.00E-06 |
| Bg9016                        | 880    | AAA51540     | 4F2 antigen heavy chain                         | <i>Homo sapiens</i>            | 71    | 1.00E-11 |
| <b>Not classified</b>         |        |              |   |                                |       |          |
| Bg9041                        | 645    | AAF45949     | CG3556 gene product                             | <i>Drosophila melanogaster</i> | 85    | 5.00E-16 |
| Bg9040                        | 401    | NP_611703    | CG4250 gene product                             | <i>Drosophila melanogaster</i> | 63    | 7.00E-10 |
| Bg11018                       | 927    | AAF55754     | CG4362 gene product                             | <i>Drosophila melanogaster</i> | 84    | 3.00E-15 |
| Bg9020                        | 902    | NP_611243    | CG6459 gene product                             | <i>Drosophila melanogaster</i> | 133   | 3.00E-30 |
| Bg9037                        | 304    | AAF50709     | CG6592 gene product                             | <i>Drosophila melanogaster</i> | 65    | 1.00E-10 |
| Bg10012                       | 854    | NP_612081    | CG9119 gene product                             | <i>Drosophila melanogaster</i> | 90    | 2.00E-19 |
| Bg9069                        | 425    | AAF48872     | CG6696 gene product                             | <i>Drosophila melanogaster</i> | 77    | 3.00E-14 |
| Bg9044                        | 833    | AAF56428     | CG10423 gene product                            | <i>Drosophila melanogaster</i> | 118   | 9.00E-26 |
| Bg7001                        | 899    | AAF58797     | CG12405 gene product                            | <i>Drosophila melanogaster</i> | 100   | 3.00E-20 |
| Bg8018                        | 285    | P30652       | 23.7KD protein ZK6436 in chromosome III         | <i>Caenorhabditis elegans</i>  | 51    | 4.00E-07 |
| Bg9003                        | 484    | A45835       | Ly6 homolog RK10 precursor                      | <i>Norway rat</i>              | 47    | 8.00E-05 |
| Bg9055                        | 877    | AAF91388     | SocE  | <i>Myxococcus xanthus</i>      | 99    | 4.00E-20 |
| Bg11070                       | 856    | NP_523610    | clumy   | <i>Drosophila melanogaster</i> | 123   | 2.00E-34 |
| Bg11054                       | 864    | NP_476631    | RpL19-P1;Enhancer of Delta KP135                | <i>Drosophila melanogaster</i> | 143   | 2.00E-33 |
| Bg12020                       | 226    | E81737       | hypothetical protein TC0128                     | <i>Chlamydia muridarum</i>     | 44    | 4.00E-04 |
| Bg4020                        | 841    | P34472       | 136.3kD a protein F58A4.5 in chromosome III     | <i>Caenorhabditis elegans</i>  | 76    | 6.00E-14 |
| Bg9061                        | 245    | AAL49280     | RE74144p  | <i>Drosophila melanogaster</i> | 44    | 5.00E-04 |
| Bg10005                       | 634    | NP_502360    | Arabidopsis pathogenesis-related protein 5 like | <i>Caenorhabditis elegans</i>  | 109   | 2.00E-23 |
| Bg11011                       | 547    | AAL31950     | CDH1-D  | <i>Gallus gallus</i>           | 51    | 7.00E-14 |

oxidase subunit I, cytochrome oxidase II, and major allergen Bla g 1. The non-matched to database group contained 80 clones. One hundred twenty nine clones weren't significant.

ESTs were classified into putative function cate-

gories based on BLAST search results with associated predicted or known functions (Table 2). The most frequently found gene was that of ribosomal protein as 31 clones (27%). Nineteen genes (18%) were of uncertain function. Twenty, 12 and 12 ESTs corresponded

to energy metabolism, enzymes related to metabolism and others, and 4 ESTs to allergens, 7 to proteases, 3 to protease inhibitors, 4 to translation factors, and 1 to cell signaling pathway.

## DISCUSSION

The ESTs of 363 clones from a randomly selected 464 clones of *B. germanica* cDNA library were submitted for BLAST search and analyzed to determine the transcribed genome sequences. One hundred fifty four matched ESTs showed high homology with genes of various organisms including *B. germanica*. The ESTs of 41 clones showed the redundancy to other ESTs. Ten clones that had shared exact homology with genes previously characterized in *B. germanica* corresponded to 3 genes, cytochrome oxidase subunit I, cytochrome oxidase II, and major allergen Bla g 1. Sixty-five ESTs were confirmed from among the 80 not-matched clones.

The most abundant group of ESTs in this study belonged to the ribosomal proteins. This result was expected because ribosomal protein genes are expressed ubiquitously at all stages of development. Moreover, the ribosomal protein family is generally well conserved and contains about 55 proteins in prokaryotes and 88 in eukaryotes (Doudna and Rath, 2002). An increasing number of studies have reported that numerous ribosomal proteins have extra-ribosomal functions, such as, involvements with several human genetic disorders (Wool, 1996). A recent study reported that ribosomal protein promotes DNA base excision repair in mammals such as the human and the mouse. This protein gene was expected to be used to repair 8-oxoguanine in man (Cappelli et al., 2003). In addition, the ribosomal protein family provides valuable comparative genomic and phylogenetic data on insecta (Landais et al., 2003).

Cytochrome ESTs followed ribosomal proteins in number. These included cytochrome b, cytochrome oxidase polypeptides I, II, and III, cytochrome oxidase subunits II, and III, and cytochrome P450. These abundances could be explained by the high mRNA expression levels of cytochrome c oxidase subunit 1 in gut

and fat bodies (Martinez-Gonzalez and Hegardt, 1994). The cytochrome c oxidase subunit I and the cytochrome P450 genes are over-expressed in pyrethroid-resistant strains of *B. germanica* (Pridgeon and Liu, 2003). Moreover, the cytochrome oxidase and P450 genes are good targets for the control of insecticide resistant German cockroach. The complete nucleotide sequences of the mitochondrial genome of several insects were recently identified for several purposes, such as, medicinal, sanitational, and forensic (Bae et al., 2004; Kim et al., 2005). The complete cockroach mitochondrial genome will be a useful source of information for molecular and evolutionary studies and for cockroach control.

*B. germanica* have been reported to have n=11 or n=12 chromosomes (Cochran and Ross, 1967; Ock and Kim, 1989). Although no specific information is available on the genome size of the German cockroach, it has been estimated to be ca.  $1 \times 10^{10}$  bp and CV=2.0 (haploid c-value in pg) (Ussery and Hallin, 2004), which is three times as large as the human genome. Wen et al (2001) reported that the *B. germanica* P450 gene is related to five pseudogenes compared to two pseudogenes in *Drosophila*. These pseudogenes, especially nuclear mitochondrial pseudogenes, have recently been viewed as tools for clarifying the relationship between DNA loss and genome size. Bensasson et al., (2001) reported that rates of DNA loss in pseudogenes are slow in the mountain grasshopper. However, in *Drosophila*, rates were high enough to contribute to the paucity of pseudogene sequences in the genome. The presence of many copies of pseudogenes is likely to explain the large genome size of *B. germanica*.

Of the protease genes identified in this study, trypsin influences growth and metamorphosis. Aalberse (2000) classified about 40 allergens into 4 structural families and other structures and designated trypsin-like serine proteases as one group of the antiparallel  $\beta$ -strands family. Moreover, there are reports that proteases extracted from *B. germanica* may have allergenic properties (Iraneta et al., 1999; Wongtim et al., 1993). We confirmed in a previous study that the trypsin of *B. germanica* reacts with the

sera of allergic patients (Ock et al., 2005). A further characterization of trypsin in this respect would provide information of allergy, since trypsin plays an important part in the activation of PAR-2 (protease-activated receptor-2).

The clone Bg9033 was identified as a lysozyme. The secretion of lysozymes is known to be increased in the gastrointestinal tract of *B. germanica* during metamorphosis and food ingestion (Aigaki et al., 2002). Thus genetic information on lysosome would be helpful in studies of cockroach metamorphosis and digestion. The clone Bg6014, a catalase is known to affect defense mechanism and to expand life span by blocking free hydroxyl radical production in *Drosophila* (Hotokezaka et al., 2002; Missirlis et al., 2001). Further studies on catalase, cytochrome oxidase, and P450 would provide information useful for cockroach control.

In addition, elongation factors, ubiquitin, iron storage protein ferritin, and G protein-coupled receptor were all confirmed to be present in the German cockroach.

In the present study, we found 363 cDNA clones in the German cockroach genome, and 360 of these were identified for the first time in the German cockroach. These ESTs should provide valuable information on the development and metabolism of *B. germanica* and lead to the discovery of control targets.

## REFERENCES

- Aalberse RC (2000) Structural biology of allergens. *J Allergy Clin Immunol* **106**: 228-238.
- Aigaki T, Seong KH, Matsuo T (2002) Longevity determination genes in *Drosophila melanogaster*. *Mech Ageing Dev* **123**: 1531-1541.
- Arruda LK, Vailes LD, Mann BJ, Shannon J, Fox JW, Vedvick TS, Hayden ML, Chapman MD (1995) Molecular cloning of a major cockroach (*Blattella germanica*) allergen, Bla g 2. *J Biol Chem* **270**: 19563-19568.
- Bae JS, Kim I, Sohn HD, Jin BR (2004) The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. *Mol Phylogenet Evol* **32**: 978-85.
- Bahl A, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, Ginsburg H, Gupta D, Kissinger JC, Labo P, Li L, Mailman MD, Milgram AJ, Pearson DS, Roos DS, Schug J, Stoeckert CJ Jr, Whetzel P. Plasmo DB (2003) the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res* **31**: 212-215.
- Bensasson D, Petrov DA, Zhang DX, Hartl DL, Hewitt GM (2001) Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol Biol Evol* **18**: 246-253.
- Cappelli E, D'Osualdo A, Bogliolo M, Kelley MR, Frosina G (2003) *Drosophila* S3 ribosomal protein accelerates repair of 8-oxoguanine performed by human and mouse cell extracts. *Environ Mol Mutagen* **42**: 50-58.
- Cochran DG, Ross MH (1967) Preliminary studies of chromosome of twelve cockroach species (Blattaria: Blattellidae, Blattellidae, Blaberidae). *Ann Ent Soc Amer* **60**: 1265-1267.
- Doudna JA, Rath VL (2002) Structure and function of the eukaryotic ribosome; the next frontier. *Cell* **109**: 153-156.
- el-Sayed NM, Alarcon CM, Beck JC, Sheffield VC, Donelson JE (1995) DNA expressed sequence tags of *Trypanosoma burcei rhodesiense* provide new insights into the biology of the parasite. *Mol Biochem Parasitol* **73**: 75-90.
- Helm R, Cockrell G, Stanley JS, Brenner RJ, Burks W, Bannon GA (1996) Isolation and characterization of a clone encoding a major allergen (Bla g Bd90K) involved in IgE-mediated cockroach hypersensitivity. *J Allergy Clin Immunol* **98**: 172-180.
- Hotokezaka Y, Tobben U, Hotokezaka H, Van Leyen K, Beatrix B, Smith DH, Nakamura T, Wiedmann M (2002) Interaction of the eukaryotic elongation factor 1A with newly synthesized polypeptides. *J Biol Chem* **277**: 18545-18551.
- Iraneta SG, Duschak VG, Rodriguez SM, Seoane MA, Albonico JF, Alonso A (1999) Proteinase and gelatinolytic activities of house dust mite and cockroach extracts. *J Investig Allergol Clin Immunol* **9**: 235-240.
- Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM, Sohn HD, Jin BR (2005) The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). *Gene* **353**: 155-68.
- Landais I, Ogliaastro M, Mita K, Nohata J, Lopez-Ferber M, Dounor-Cerutti M, Shimada T, Fournier P, Devauchelle G (2003) Annotation pattern of ESTs from *Spodoptera frugiperda* Sf9 cells and analysis of ribosomal protein genes reveal insect-specific features and unexpectedly low codon usage bias. *Bioinformatics* **19**: 2343-2350.

- Manger ID, Hehl A, Parmley S, Sibley LD, Marra M, Hillier L, Waterston R, Boothroyd JC (1998) Expressed sequence tags analysis of the bradyzoite stage of *Toxoplasma gondii*: Identification of developmentally regulated genes. *Infect Immun* **66**: 1632-1637.
- Martinez-Gonzalez J, Hegardt FG (1994) Cytochrome c oxidase subunit I from the cockroach *Blattella germanica*: cloning, developmental pattern and tissue expression. *Insect Biochem Mol Biol* **24**: 619-626.
- Missirlis F, Phillips JP, Jackle H (2001) Cooperative action of antioxidant defense systems in *Drosophila*. *Curr Biol* **11**: 1272-1277.
- Ock MS, Kim BJ, Kim SM, Byun KH (2005) Cloning and expression of trypsin-encoding cDNA from *Blattella germanica* and its possibility as an allergen. *Korean J Parasitol* **43**: 101-110.
- Ock MS, Kim KR (1989) Karyotype analysis of *Blattella germanica* L. and *Blattella nipponica* Asahina in Korea. *J Kosin Med Col* **5**: 21-25.
- Pridgeon JW, Liu N (2003) Overexpression of the cytochrome c oxidase subunit I gene associated with a pyrethroid resistant strain of German cockroaches, *Blattella germanica* (L.). *Insect Biochem Mol Biol* **33**: 1043-1048.
- Richman PG, Picone FJ, Strunk RS, So J, Colten HR (1984) The important sources of German cockroach allergens as determined by RAST analyses. *J Allergy Clin Immunol* **73**: 590-595.
- Ross MH, Cochran DG (1975) The German cockroach, *Blattella germanica*. In: King RC editor. Handbook of genetics Vol. 3 New York: Plenum. p 35-62.
- Scharf ME, Neal JJ, Marcus CB, Bennett GW (1998) Cytochrome P450 purification and immunological detection in an insecticide resistant strain of German cockroach (*Blattella germanica* L.). *Insect Biochem Mol Biol* **28**: 1-9.
- Ussery DW, Hallin PF (2004) Genome update: length distributions of sequenced prokaryotic genomes. *Microbiol Comment* **150**: 513-516.
- Wen Z, Horak CE, Scott J (2001) CYP9E2, CYP4C21 and related pseudogenes from German cockroaches, *Blattella germanica*: implications for molecular evolution, expression studies and nomenclature of P450s. *Gene* **272**: 257-266.
- Wongtim S, Lehrer SB, Salvaggio JE, Horner WE (1993) Protease activity in cockroach and basidiomycete allergen extracts. *Allergy Proc* **14**: 263-268.
- Wool IG (1996) Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci* **21**: 164-165.
- Wu CH, Lee MF, Liao SC, Luo SF (1996) Sequencing analysis of cDNA clones encoding the American cockroach Cr-PI allergens. Homology with insect hemolymph proteins. *J Biol Chem* **271**: 17937-17943.