

## Dung beetle database: comparison with other invertebrate transcriptomes

Lucky M Khanyile<sup>1</sup>, Rodney Hull<sup>1</sup> and Monde Ntwasa<sup>1,\*</sup>

<sup>1</sup>Department of Molecular and Cell Biology, Gatehouse 512, University of the Witwatersrand, Wits - 2050, South Africa; Monde Ntwasa\* - Email - monde.ntwasa@wits.ac.za; \* Corresponding author

received November 12, 2008; accepted December 01, 2008; published December 06, 2008

### Abstract:

The dung beetle *E. intermedius*, a member of the highly diverse order, Coleoptera has immense economic benefits. It was estimated that insect ecological services in the United States amounted to some \$60 billion in 2006 with dung beetles being major contributors. *E. intermedius* may be endowed with a robust immune system given its microbe-rich habitat. Dung beetles live on juice and microbes from the dung and are therefore, potential models for the study of infectious agents and ecological damage. The *E. intermedius* database is a web-based system for the genome and transcriptome of the dung beetle. The database will be expanded to include differentially expressed genes in response to various stresses especially infectious agents such as fungi, bacteria and viruses.

**Availability:** The dung beetle transcriptome database is freely available at <http://flylab.wits.ac.za/>

**Keywords:** *E. intermedius*; coleopteran; insect immunity; invertebrate transcriptomes; database

### Background:

The order, Coleoptera, is the most diverse on earth. The scarab beetle, *Euoniticellus intermedius*, is highly effective in new habitats and often more successful than native beetles. *E. intermedius* makes tunnels beneath dung pats where they feed and reproduce. Adults obtain nutrients from the microbe - rich liquid portion of manure feed and do not actually consume the dung. The dung beetle larvae consume most of the dung in brood balls where they are laid.

Dung beetles have important agricultural benefits and are often introduced in the environment to alleviate ecological damage. A recent meeting of some 1500 conservation biologists at Columbia University found that loss of dung beetles could hurt the ecosystem [1]. Some of the benefits include nutrient recycling, improvements to soil tilth and pest control. *E. intermedius* beetles are the most beneficial to pasture health and in agriculture as they enhance soil conditions by increasing percolation. Dung beetles also reduce the number of parasites acquired by cattle and greatly reduce the population of pestiferous flies (that are dangerous to livestock) such as African buffalo fly (*Haematobia thirouxii potans*), the Australian buffalo fly (*Haematobia irritans exigua*), the bush fly (*Musca vetustissima Walker*), the face fly (*Musca autumnalis*) and the horn fly (*Haematobia irritans irritans*). These dung beetles also help control nematodes that are potential cattle parasites [2].

The value of insect ecological services in the United States of America was estimated at \$60 billion in 2006 with the dung beetles being the most significant contributors. Scarab beetles contributed services such as dung burial, reducing forage fouling, nitrogen volatilization and control of cattle parasites and pest flies [3].

Mostly behavioural, ecological and taxonomic studies of *E. intermedius* have been conducted so far. This is the first significant molecular biology study of *E. intermedius*. We are interested in these beetles because of their microbe-rich habitat as they may have a potent immune system. Their immune system could provide useful strategies for treatment of infectious diseases affecting humans and animals. The study of beetles could also help in managing ecosystems and in enhancing agricultural practices.

### Methodology:

#### Data collection and sequencing

cDNA was synthesized from mRNA isolated from adult beetles sequenced by the GS(FLX) technology by a commercial facility (Inqaba Biotec. Pty (Ltd)).

#### Sequence processing and analysis

Sequence processing and analysis were performed by the use of the EST2Uni (EST analysis software to create an annotated UNIGene database) pipeline [4]. The EST2uni pipeline automates the pre-processing, clustering, annotation, databases creation, data mining and retrieval of the EST collection. To reach a fully functional EST2uni pipeline on a Unix-base server, the following

open-source tools were installed; HMMER module an implementation of profile hidden Markov models [5] GO-Perl, BLAST [6], modified Sputnik that is distributed with EST2uni source code [4], ESTScan, TGICL [7], RepeatMasker, LUCY [8], SeqClean, Exonerate, Phred and Cross-match [9], Apache, MySQL, PHP, Perl and Bioperl. In addition NCBI UniVec and RepBase [10] databases were used locally for the pre-processing stage. Sequences and results of the analyses are stored in a MySQL relational database that is created automatically by the EST2uni pipeline.

### Database content

The database contains 6064 ESTs further processed into 744 CONTIGS and 1918 singletons. Sequence comparisons between the *E. intermedius* transcriptome and other species were performed by TBLASTX and BLASTX at a cutoff E-value  $<1 \times 10^{-10}$ . The compared species fall in the two main insect branches; holometabolous (*Apis*

*melifera*, *Bombyx mori*, *Tribolium castaneum*, *Drosophila melanogaster*) and hemimetabolous (*Acyrtosiphon pisum*). The comparative results show that *E. intermedius* has closer sequence homology to *Tribolium* than to the other species. Functional comparison was based on *Drosophila melanogaster* gene ontology. Motifs and domains were identified by using the Pfam and HMMER at a cut-off E-value  $<1 \times 10^{-10}$ . Information collected from these sources shows that there are 472 (17%) unigenes in the *E. intermedius* transcriptome database that have identifiable motifs and domains. The database can be searched by using BLAST and by unigene ID (Figure 1).

### Future development:

The database will be expanded to include genome sequences of *E. intermedius*. In addition differentially expressed genes in response to a variety of stresses on *D. melanogaster* and *E. intermedius* will be included. To this end the website is called the Flylab Genomebase to reflect its broader purpose.

The screenshot shows a web interface titled "Queries" with a "Search unigenes:" section. It contains several search criteria, each with a checkbox and a form field:

- Name: begins with [dropdown] [text input] (e.g. CL505Contig1)
- Upload file with list: [text input] [Browse...] [Help]
- Type: [dropdown] (contig)
- Blast result: Database: [dropdown] (Acyrtosiphon\_pisum) Search Pattern: [text input] (e.g. rubisco) [Help]
- HMMER result: Database: pfam Search Pattern: [text input] (e.g. ehand) [Help]
- cDNA Libraries: [dropdown] (some) ESTs from libraries: [text input] (All...)
- pSSR presence: pSSR name: [dropdown] (begins with) [text input] pSSR type: [dropdown] (ALL)
- pSNP name: pSNP name: [dropdown] (begins with) [text input]
- pSNP type: confidence: [dropdown] (ALL) snp/indel: [dropdown] (BOTH)
- GO annotation: Database: Drosophila\_melanogas Aspect: [dropdown] (Molecular Function) GO: [text input] (nucleotide binding)
- Complete CDS:

**Figure 1:** A snapshot of the database. In addition to the BLAST page for comparing user sequences with the database. This page allows the user to view the full sequences in the dataset and search entries using the sequence ID.

### References:

- [01] E. Stokstad, *Science*, 305: 1230a (2004) [PMID: 15333817] [02] S. Anduaga, *Environ. Entomol.*, 33: 1306 (2004) [PMID: 17540064]

- 
- [03] J. E. Losey and Vaughan, M. *BioScience*, 56: 311 (2006)
- [04] J. Forment *et al.*, *BMC Bioinformatics*, 9: 5 (2008) [PMID: 18179701]
- [05] A. Krogh *et al.*, *J. Mol. Biol.*, 235, 1501 (1994) [PMID: 8107089]
- [06] S. F. Altschul *et al.*, *J. Mol. Biol.*, 215 (1990) [PMID: 2231712]
- [07] G. Pertea *et al.*, *Bioinformatics*, 19: 651 (2003) [PMID: 12651724]
- [08] H. H. Chou and Holmes, M.H. *Bioinformatics*, 17: 1093 (2001) [PMID: 11751217]
- [09] B. Ewing *et al.*, *Genome Research*, 8: 175 (1998) [PMID: 9521922]
- [10] J. Jurka *et al.*, *Cytogenetic and Genome Research*, 110: 462 (2005) [PMID: 16093699]

Edited by P. Kanguene

Citation: Khanyile *et al.*, *Bioinformatics* 3(4): 159-161 (2008)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.