Lifestyle and Host Defense Mechanisms of the Dung Beetle, Euoniticellus intermedius: The Toll Signaling Pathway

Authors: Hull, Rodney, Alaouna, Mohamed, Khanyile, Lucky, Byrne, Marcus, and Ntwasa, Monde

Lifestyle and host defense mechanisms of the dung beetle, *Euoniticellus intermedius*: The Toll signaling pathway

Rodney Hull[1a], Mohamed Alaouna[1b], Lucky Khanyile[1c], Marcus Byrne[2d], Monde Ntwasa[1e]

[1]School of Molecular & Cell Biology. University of the Witwatersrand, Johannesburg. Private Bag X3, Wits. 2050, South Africa
[2]School of Animal and Plant Sciences. University of the Witwatersrand, Johannesburg. Private Bag X3, Wits. 2050, South Africa

Abstract

The dung beetle, *Euoniticellus intermedius* (Reiche) (Coleoptera: Scarabaeidae) is an important ecological and agricultural agent. Their main activity, the burying of dung, improves quality of the soil and reduces pests that could cause illness in animals. *E. intermedius* are therefore important for agriculture and for good maintenance of the environment, and are regarded as effective biological control agents for parasites of the gastrointestinal tract in livestock. The ability of *E. intermedius* to co-exist comfortably with many microorganisms, some of which are important human pathogens, stimulated our interest in its host defense strategies. The aim of this study was to investigate the Toll signaling pathway, which is strongly activated by fungi. Gene expression associated with fungal infection was analyzed by using 2-D gel electrophoresis and mass spectroscopy. Furthermore, the partial adult transcriptome was investigated for the presence of known immune response genes by using high-throughput sequencing and bioinformatics. The results presented here suggest that *E. intermedius* responds to fungal challenge via the Toll signaling pathway.

Cite this paper as:
Hull R, Alaouna M, Khanyile L, Byrne M, Ntwasa M. 2013. Lifestyle and host defense mechanisms of the dung beetle, *Euoniticellus intermedius*: The Toll signaling pathway. *Journal of Insect Science* 13:108. Available online: http://www.insectscience.org/13.108

Introduction

The dung beetle, *Euoniticellus intermedius* (Reiche) (Coleoptera: Scarabaeidae), belongs to the most diverse order of insects, with more described species than any other in the animal and plant kingdoms (Farrell 1998). The two main branches of Class Insecta are the hemimetabolous insects, such as the grasshopper, which have incomplete metamorphosis, and the holometabolous insects, such as the beetles, which have complete metamorphosis. These branches split about 416 m.y.a. (Labandeira and Phillips 1996; Hoffmann 2004; Erezyilmaz 2006).

Drosophila melanogaster Meigen (Diptera: Drosophilidae) is another holometabolous insect and is regarded as the most genetically-tractable and widely-studied laboratory model of the holometabolous insects, but there is evidence that it is highly specialized and does not fully represent invertebrate evolutionary characteristics. For example, the beetle *Tribolium castaneum* (Herbst) (Coleoptera: Tenebrionidae) shares with vertebrates ancestral genes that are not present in *D. melanogaster*. For example, components of some signaling pathways that are conserved in coleopterans and vertebrates are not conserved in *D. melanogaster* (Van der Zee et al. 2008). Moreover, *T. castaneum* embryogenesis has retained the ancestral short-germ band mode, which resembles vertebrates (Tautz and Sommer 1995; Handel et al. 2000; Liu and Kaufman 2005). A recent study on protein evolution suggests that coleopterans have lower rates of divergence when compared to dipterans where a highly accelerated protein evolution is evident (Savard et al. 2006). Coleopterans are therefore likely to be more suitable for comparative studies against vertebrates than the dipterans.

Due to their important agricultural benefits, *E. intermedius* are often introduced to control ecological damage. Some of the benefits include nutrient recycling, improvements to soil tilth, and pest control (Bertone et al. 2006). Tunneling beetles, such as *E. intermedius*, are beneficial to pasture health, as they enhance soil conditions by increasing percolation. Furthermore, dung beetles reduce the number of parasites acquired by cattle (Fincher 1975). They also reduce the population of pestiferous flies, such as the African buffalo fly (*Haematobia thirouxi potans* (Bezzi)), the Australian buffalo fly (*Haematobia irritans exigua* de Meijere), the bush fly (*Musca vetustissima* Walker), the face fly (*Musca autumnalis* De Geer), and the horn fly (*Haematobia irritans irritans* (L.)) (Lastro 2006). Consequently, the introduction of *E. intermedius* into new habitats has a high impact on the environment. *E. intermedius* was introduced in Texas, USA, to control dung accumulation, and the beetles spread quickly to Mexico, where they had a positive impact on soil fertility and productivity. *E. intermedius* also help control nematodes that are potential cattle parasites (Montes and Halffter 1998; Anduaga 2004).

Adult *E. intermedius* feed on the microbe-rich particulate portion of the dung and do not actually consume the dung fibers (Holter and Scholtz 2007). The larvae, on the other hand, consume most of the dung fibers in the brood balls. Development from egg to adult takes 5 to 6 weeks, and adults have a lifespan of about 2 months. The evolutionary history and lifestyle of dung beetles point to an interesting immune system that could be more comparable to vertebrate innate immunity than *D. melanogaster* in some respects.

The general trend in insect immune signaling is that recognition of pathogens depends upon the existence of molecular patterns, called

pathogen associated molecular patterns, on the cell wall of pathogens and the pattern recognition receptors in the immune system of the target organism. Lysine containing peptidoglycans found in Gram-positive bacteria and b-1,3-glucans in fungi stimulate the Toll signaling pathway, while diaminopimelic acid containing peptidoglycans found in Gram-negative bacteria and some Gram-positive bacilli stimulate the immune deficiency pathway (Hoffmann 2004; Zou et al. 2007; Altincicek et al. 2008; Valanne et al. 2011). In *D. melanogaster*, it has been shown that virulence factors and pathogen associated molecular patterns in fungi and bacteria activate the Toll pathway via two parallel protease cascades that culminate in the cleavage of pro-spaëtzle to produce the active physiological ligand of Toll. The virulence factors, such as cell wall components, in live organisms activate the Toll pathway via pattern recognition receptors and a serine protease cascade, which eventually cleaves the inactivate pro-spaëtzle to release an active C-terminal fragment (C-106), which then binds to Toll. The parallel pathway is induced by pathogen-secreted proteases and proceeds via a cascade involving Persephone (Psh) and activation of Toll by activated spaëtzle (Imler et al. 2004; Chamy et al. 2008; Kellenberger et al. 2011; Valanne et al. 2011). Intracellular signals transduced by the binding of spaëtzle and Toll are relayed by a phosphorylation cascade that culminates in the cytoplasmic release and nuclear localization of NFkB-like transcription factors. The Toll pathway leads to the release of the dorsal immune factor transcription factor, which activates the expression of various effector genes, primarily antimicrobial peptides.

In recent years, ground-breaking work in coleopteran immunity has been accumulating, and the nature of the signaling network is emerging. Studies using the larger beetles, *Tenebrio molitor* and *Holotrichia diomphalia*, show that beetles responded to b-1,3-glucans and to lysine containing peptidoglycan and polymeric diaminopimelic acid containing peptidoglycans by using the Toll signaling pathway. Both types of pathogen associated molecular patterns form complexes with Gram-negative binding protein 3(GNBP3) in fungi and peptidoglycan recognition protein in Gram-negative bacteria. These complexes activate the serine protease cascade via an apical protease known as the modular serine protease, followed by two types of CLIP domain serine proteases at penultimate and terminal positions before pro-spaëtzle (Park et al. 2010; Kellenberger et al. 2011; Ntwasa et al. 2012). The studies on beetles revealed important pieces of evidence indicating that beetle immunity might differ from that of flies. For example, studies conducted on the red flour beetle revealed that the peptidoglycan recognition proteins, used by invertebrates to sense pathogens, diversified in beetles before the radiation of holometabolous insects, while in *D. melanogaster* there has been sustained diversification through numerous duplications. Furthermore, many immune related genes, including the CLIP-domain serine proteases and their inhibitors, Toll related proteins, and antimicrobial peptides, diversified extensively through evolution most likely because of the diverse habitats of the species (Zou et al. 2007).

There have been no significant molecular studies conducted on *E. intermedius* thus far, except for those involving the relationship between prophenoloxidase and immunity (Pomfret and Knell 2006). The aim of this study was to assess the immune defense pathways that are activated when *E. intermedius* is infected by fungi. It is part of a broader investigation of host defense mechanisms adopted by the dung beetle. The life cycle of *E. inter-*

medius is presented to highlight that it is dominated by microbes, some of which are human pathogens. The larval and adult stages are particularly dependent on these microbes for nutrition. Proteomic and bioinformatics analysis of the adult transcriptome of the dung beetle indicates that the Toll signaling pathway may be conserved and probably activated by fungi.

Materials and Methods

Rearing beetles and microscopy

Beetles were bred in 160 × 130 × 130 mm plastic containers, which were half full with soil. Cow dung was placed on the soil, and 3 breeding pairs of beetles (3 males and 3 females) were added to the containers. Every 3 to 4 days, frozen and then thawed cow dung was placed in the container. Once a week, the soil was sieved, and brood balls were removed. If the breeding pair had survived, they were moved to a new container with fresh soil, and dung and the brood balls were placed in a large (400 × 300 × 200 mm) plastic container covered with compacted soil. A wet sponge was then placed on the soil to keep it moist. Once beetles began to emerge, small, plastic dishes baited with dung were used as traps to capture them within the bins. These beetles were used as new breeding pairs or in experimental work.

Treatment of beetles

Adult male and female beetles were infected with the entomopathogenic fungus *Beauveria bassiana* (strain 80.2), and samples of haemolymph were collected before and after infection. Ten beetles were placed on a 90 mm plate containing *B. bassiana* (i.e., 1 beetle per 6.36 cm^2). The plate was then placed on a shaker for 5 minutes, after which the beetles were transferred to a small jar containing sterile soil and dung and then left overnight to

allow an immune response to occur.

Collection and preparation of haemolymph samples

To collect haemolymph from adult beetles, the exoskeleton was pierced below the hind limb using a heated tungsten spike. A 10 ml microsyringe needle was then pushed through this hole. The plunger was slowly withdrawn, and between 2 and 5 μL of haemolymph was collected and pooled from 5 beetles without discriminating between genders. The extracted haemolymph was diluted 1:4 in extraction buffer (0.1% trifluoroacetic acid (to obtain pH3), 10 μg/mL aprotinin, 1 μM leupeptin, 15 μM phenylmethylsulfonyl fluoride, and 20 μM phenylthiourea). The samples were not centrifuged to remove hemocytes, because both extracellular and intracellular events at this stage were of interest. Protein concentration was measured by the Bradford method (Bradford 1976).

2-dimensional gel electrophoresis

Haemolymph proteins were re-suspended in 2-D PAGE re-hydration buffer (8 M urea, 2% CHAPS, 50 mM DTT, 0.2% Bio-lyte, 0.5 % Bromophenol Blue) and applied onto 7 cm immobilized pH 3-10 gradient strips (Bio-Rad, www.bio-rad.com). Passive rehydration was allowed to occur for 12 hours at 20° C. Isoelectric focusing was performed in the Protean IEF system (Bio-Rad) according to the following program: an initial low voltage (250 V) 20-minute linear ramping step, followed by a high voltage (4000 V) 2-hour linear ramping step. The final step was a rapid ramping step for 10,000 volt hours. The strip was equilibrated in a new rehydration tray and covered with equilibration solution one (6 M urea, 2% SDS, 0.375 M Tris HCl pH 8.8, 20% glycerol, 2% DTT). After shaking for 10 minutes, equilibration solution one was replaced with equilibration solution two (6 M urea, 2% SDS,

0.375 M Tris HCl pH 8.8, 20% glycerol, 2.5% iodoacetamide) and shaken for 10 more minutes. The strips were then briefly placed in SDS running buffer and placed on a 10% polyacrylamide separating gel (Laemmli 1970), and then overlaid with agarose. The 2-D gel was then run for 40 minutes at 200 V and subsequently stained with Coomasie blue.

Image Analysis
The two gels representing samples from treated and untreated beetles were digitized using a GS800 calibrated densitometer (Bio-Rad), and the resulting images were analyzed using the PDQuest 2-D Analysis Software Version 6.2 (Bio-Rad). Spot detection was carried out with a sensitivity setting of 102.21, using the treated sample gel as the master gel. 464 and 374 spots were respectively detected in the treated and untreated sample gel images. The analysis software was used to select spots for identification by mass spectrometry. Many spots within the M_r range 10–100 kDa and the pI range 4–7, based on whether they were down- or up-regulated, were selected.

Spots were initially matched using the automated matching function of the program. Extended matching was then performed using the classical matching function. Both these methods involved matching spots based on the position of landmark spots. Manual spot matching and analysis was then performed across the gel images. This resulted in a final spot count of 135 spots in the master gel image, and 114 and 84 in the treated and the untreated images respectively.

Normalization was performed automatically by the program and was based on the total quantity in all valid spots option. This method assumed that changes in density average-out across the gels being analyzed.

The normalization formula used is as follows:

$$\text{Normalized spot quantity} = \frac{\text{Raw spot quantity} \times \text{Scaling factor}}{\text{Normalization factor (total quantity in all valid spots)}}$$

The scaling factor used was 10^6 parts per million.

Statistical analysis
Three biological replicates of haemolymph extracted from immune-challenged and unchallenged adult beetles were analyzed using 2D-PAGE and the PDQuest software. The means of density values from these biological replicates were used to construct the graphical representations of the optical density values for each spot. Two-sample t-tests were performed to test for statistically significant differences between the challenged and unchallenged groups for each spot at the 95% confidence interval. Two t-tests were computed, one assuming equal group variances and the other assuming different group variances. An F-test for the null hypothesis that the immune-challenged and unchallenged populations have the same variance was also performed. The results of these tests were presented as p- and F-values. All statistical tests were performed using the Statistix 8 software package (www.statistix.com).

Matrix laser desorption-time-of-flight mass spectrometry
Selected spots were manually excised from the gels and placed in a 5% acetic acid solution. After tryptic in-gel digestion following the method of Rabilloud (Rabilloud et al. 2001), the proteins were analyzed by mass spectroscopy either the matrix laser desorption-time-of-flight or the NanoLC-MS/MS (Bflex III Brüker Daltonics, a CapLC capillary LC system (Waters, www.waters.com)) coupled to a hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spec-

trometer (Q-TOF Micro, Waters). The sample (5 μl) was first concentrated and cleaned in a C18 PepMap precolumn cartridge (LC Packings/Dionex, www.dionex.com) and then separated on-line by the analytical reversed-phase capillary column (Pepmap C18, 75 μm inner diameter, 15 cm length; LC Packings) under a 200 nL min^{-1} flow rate. The gradient profile used consisted of a linear gradient from 97% A (97.9% H$_2$O; 2% ACN, 0.1% (v/v) HCOOH) to 95% B (98% ACN, 1.9% H$_2$O, 0.1% (v/v) HCOOH) in 45 minutes, followed by a linear gradient to 95% B in 3 minutes. The spray system (liquid junction) was used at 3.6 kV. Mass data acquisitions were piloted by MassLynx 4.0 software (Waters). NanoLC-MS/MS data were collected by data-dependent scanning, that is, automated MS to MS/MS switching. Fragmentation was performed using argon as the collision gas, and with a collision energy profile optimized for various mass ranges of ion precursors. Four ion precursors were allowed to be fragmented at a time. Mass data collected during a NanoLC-MS/MS analysis were processed and then submitted to *de novo* sequencing. Peak lists were generated by ProteinLynx 2.1 software (Waters) with internal lockspray calibration (leucine enkephalin at 556.2771 m/z). There was no smoothing. Fragmentation spectra were loaded onto the Peptide Sequencing software (BioLynx, Waters) and the sequences were processed manually before being submitted to a BLAST search without any taxonomy restriction.

Searching the *Euoniticellus intermedius* database for immune genes and other bioinformatics tools
The *E. intermedius* database contains a partial transcriptome of the untreated adult beetle (Khanyile et al. 2008). Sequences of innate immune response proteins were obtained from the UniprotKB/Swiss-Prot database using gene ontology identity number and species name. Protein sequences were selected from the human, *D. melanogaster* and *T. castaneum* databases. These included 600 human, 6000 *Drosophila* and 12 *Tribolium* sequences, and were deposited in a MySQL local database. Using TBLASTX, the local database was compared with the *E. intermedius* unigene database at a cut-off e-value of 1e-5.

Results

Euoniticellus intermedius life cycle in microbe-rich environment
The entire life cycle of *E. intermedius* occurs in cow dung (Figure 1A). Their adaptation to a habitat that is rich in different types of microbes, including human pathogens, suggests that dung beetles possess robust host-defense mechanisms. Extensive studies on the development of coleopterans have been conducted on *T. castaneum*, a short germ insect, but little is known about the developmental biology of *E. intermedius*.

The embryo develops in the brood ball apparently without physical contact with the interior surface. It is anchored on its posterior end ostensibly by the sticky black material termed the "maternal gift," as it is placed there by the mother (Figure 1B). The role of the maternal gift has not been proven, but apart from acting as an anchor for the embryo, it could serve other purposes (Byrne et al. 2013). The features of the developing larvae are typical of other scarab beetles; with a dorsal expansion (or "hump") at the middle of the body and caudal flattening (Edmonds and Halffter 1978). This seems to be a physical adaptation to facilitate movement of the larvae in the confines of the brood ball as it rotates within the ball to feed on dung. Feeding seems to start at the second larval instar, as these larvae look white and free of dung before this

stage. At later stages, the gut system resembles a "bag of dung."

Euoniticellus intermedius immune response to fungal challenge

2-D gel electrophoresis of haemolymph from adult beetles before and after infection with the entomopathogenic *B. bassiana* revealed differential expression of certain genes in response to fungal infection (Figure 2). Differentially expressed genes were determined by the PDQuest software in three separate experiments.

Initially, matrix laser desorption-time-of-flight was used to identify samples after in-gel tryptic digestion. The mass list obtained at this stage was used to search the United States National Center for Biotechnology Information (NCBI) database (www.ncbi.nlm.nih.gov) without any taxonomy restrictions using the Mascot search engine. None of the proteins could be identified by this method. Nano LC-MS/MS analysis was then performed in order to obtain sequence data. Short sequence fragments were then determined by *de novo* sequencing. BLAST search was performed on the NCBI BLASTP 2.2.14 on all non-redundant GenBank CDS containing, at this time, 3695564 entries. For many of the positively identified proteins, the MS-BLAST search engine was employed using the short peptide fragments for a given protein sequence that was used as input data into the MS BLAST search engine (http://dove.embl-heidelberg.de/Blast2/msblast.html).

Proteins identified by MS-BLAST were then subjected to manual comparison with the *E. intermedius* database (http://flylab.wits.ac.za/EI/est2uni/blast.php) using TBLASTN (Table 1). In some cases, alternative peptides matched sequences in the same contig or singleton. For example, spots 2203 and 3304, which were up-regulated upon fungal infection, were predicted as serine proteases from the short amino acid sequences produced by *de novo* sequencing using Nano LC-MS/MS. Alternative peptides produced from these spots could be assigned to different parts of the same contig. Sequences of two peptide fragments associated with spot 2203 (with molecular weight of approximately 30 kDa) matched parts of the predicted CL8Contig1 protein sequence. Two of these fragments had 100% identity with peptide sequences in the CL8Contig1-encoded polypeptide, which is predicted to be a chymotrypsin-type serine protease like Psh or sphinx1/2. The peptide fragments obtained from spot 3304, at approximately 50 kDa, facilitated the identification of two contigs (CL20Contig1 and CL23Contig1) in the *E. intermedius* database using TBLASTN. These contigs are also predicted to encode a serine protease like Psh or sphinx. Correlation of these spots with Psh was consistent with response to fungal challenge as observed in *D. melanogaster* (Imler et al. 2004). Based on predicted molecular weights of *D. melanogaster* sphinx1/2 (25 kDa) and Psh (43–50 kDa, sphinx could be assigned to spot 2203 and Psh to spot 3304, but a definite identification will require full length sequencing and more biochemical analysis.

Bioinformatics analysis of immune genes in the *Euoniticellus intermedius* transcriptome database

Comparison of the *E. intermedius* transcriptome database with databases of holometabolous (*A. melifera*, *B. mori*, *D. melanogaster*, and *T. castaneum*) and hemimetabolous (*A. pisum*) insects showed that *E. intermedius* was more comparable to *T. castaneum* than it is to other insect species (Table 2). Approximately 37% of the *E. in-*

termedius transcripts matched orthologues in the *D. melanogaster* database, indicating that the *D. melanogaster* database may be inadequate for functional annotation of the dung beetle genome. This apparent divergence could be due to evolutionary factors, as the primary radiation of coleopterans and dipterans occurred approximately 284 m.y.a., and dipterans have been shown to exhibit an accelerated rate of protein evolution (Savard et al. 2006).

Functional annotation against the *D. melanogaster* database revealed proteins involved in a range of biological processes and molecular functions (Figure 3). Since this ontological analysis provided general information about response to stress and did not specify immune-related genes, targeted analysis was conducted using more databases, as described in materials and methods. Consequently, more immune genes were predicted through comparison with genes in the human, *T. castaneum*, and *D. melanogaster* databases (Table 3).

Using this approach, several molecules associated with the Toll signaling pathway were predicted. CL8Contig1, CL20Contig1, CL15Contig1, and CL23Contig1 were predicted as either Psh or sphinx1/2. Others were predicted to be Toll-related genes (CL426Contig1, CL673Contig1) and spaëtzle (CL283Contig1). Manual searches using *Drosophila* or *Tribolium* sequences revealed another putative spaëtzle-encoding contig (Cl47contig1), a singleton (007956_1645_0789_c_s) that is predicted to encode b-1,3-glucan-binding protein (GNBP3) and GNBP1 (CL652Contig1), and pattern recognition receptors that recognize fungi and bacteria. Since the *E. intermedius* database was created from transcripts obtained from uninfected adult beetles and was thus not en-

riched for immune-related genes, the detection of these Toll pathway genes suggests that they were expressed constitutively in the adult beetle.

Identity of the predicted serine proteases

Several proteases were identified by LC-MS/MS analysis combined with sequence similarity searches using the MASCOT search engine as a first layer search on public databases without taxonomy restrictions. Furthermore, the sequences were manually compared to the *E. intermedius* database using TBLASTN. Using this approach, sequences of peptide fragments obtained from spots 2203 (molecular weight ~30 kDa) and 3304 (molecular weight ~50 kDa) matched proteins encoded by CL8Contig1, CL15Contig1, and CL20Contig1. These spots represented proteins that were up-regulated by fungal infection, and the matching transcripts encoded proteins that are predicted to be trypsin-like serine proteases. During immune response, serine proteases (SPs) and the non-catalytic serine protease homologues (SPHs), usually with N-terminal CLIP-domains, are known to mediate extracellular signaling activated by bacterial and fungal pathogens. SPHs are non-catalytic serine proteases where the catalytic triad has undergone mutations. Their function is poorly understood, but in some instances they have been reported to act as cofactors to the catalytic proteases (Gupta et al. 2005). In *H. diomphalia*, the prophenoloxidase (proPO) activating factor (PPAF-II) acts as a cofactor for PPAF-I in a limited proteolysis reaction to activate proPO (Kim et al. 2002).

An alignment of protein sequences encoded by these contigs with Psh, *Holotricia* PPAF-II, and other clip-domain SPs and SPHs showed that the catalytic triad was conserved in all three *E. intermedius* predicted proteins, indicating that they may be SPs (Figure 4). The

CLIP-domain is found in serine proteases such as Grass, Psh, and the SPH PPAF-II, but not in SPHs such as sphinx and spheroid (Kellenberger et al. 2011). The CLIP-domain SPs are known to be involved in the final steps of the proteolytic cascades. This study relied on the catalytic domains for analysis since the predicted *E. intermedius* proteins under investigation here did not have the full N-terminal sequence. Phylogenetic analysis conducted using the catalytic domain sequences showed potential relatedness between the *E. intermedius* proteins and those of other insects (Figure 4). In the neighbor-joining tree, *Drosophila* Psh and the *Holotricia* PPAF-II were grouped together with CL23Contig1 and possibly CL20Contig1, while sphinx1/2 may be related to CL8Contig1. Interestingly, *Tribolium* PPAF-II is predicted to be a Psh-like cofactor for putative proPO activating proteinases called TcSP7, SP8, or SP10 (Zou et al. 2007). This may be evidence for the involvement of Psh-like proteases in the pro-PO pathway. It is notable that the catalytic triad was conserved in CL8Contig1 and mutated in sphinx, suggesting that CL8Conting1 belonged to the class of SPs.

Other differentially expressed proteins in the haemolymph

In addition to the serine proteases, other proteins affected by fungal infection included pattern recognition receptor molecules, such as the apolipophorin III (spots 7203 and 7210 and CL123Contig1 in *E. intermedius* database) and the imaginal disc derived growth factor. Apolipophorin III and the imaginal disc derived growth factor protein spots are expressed at low levels in untreated beetles and elevated after fungal challenge. The Gram-negative-binding protein and the spaëtzle processing enzyme were detected by manually searching the *E. intermedius* database by TBLASTN using the *Drosophila*

proteins. At least one Toll-like protein was predicted.

Peptide sequences obtained from spots 0101, 3004, 4002, and 9003 indicated possible matches with Pez when compared against the NCBI database. However, these sequences matched the odorant binding protein (OBP) family when compared to the *E. intermedius* database. The size range of these spots was consistent with that of OBPs and was too small for Pez. Protein spot 4002 appeared to be down-regulated, but statistical analysis indicated that the differences between protein expression in challenged and unchallenged beetles was not significant. Similarly, spot 9003 and 3004 appeared to be up-regulated, but statistical analysis indicated that the apparent differential expression was not significant. Protein in spot 0101 was, however, significantly up-regulated following fungal infection. The translated *E. intermedius* sequence (CL32Contig1) showed perfect matches with two alternative MS/MS peptide fragment sequences that were both identified as PBP/GOBP family when compared to the *E. intermedius* database (Figure 6). Interestingly, this transcript appeared to have two contiguous polyadenylation signals similar to those reported for human and horse growth hormones (Masuda et al. 1988; Ascacio-Martínez and Barrera-Saldaña 1994).

Spots 6410 and 5310 were predicted to be transferrins and gelsolin respectively. Other differentially expressed genes include the esterases and ATP synthase. Some of these proteins are usually associated with insect immune response, and others are known components of the haemolymph (Vierstraete et al. 2003; Karlsson et al. 2004).

Discussion

Extracellular signal amplification and trans-

duction during insect immune response is characterized by three categories of molecules that transduce the signal into the cell: the pattern recognition receptors, the SPs, and transmembrane receptors. Some components of the known extracellular signaling molecules were evident in *E. intermedius* based on proteomic analysis and the study of the adult transcriptome.

Pattern recognition receptors

In insects, fungal pathogens are recognized by constitutively active and inducible pattern recognition receptors called GNBPs. Signals are transduced via an SP cascade in the extracellular space and across the membrane via the Toll receptor. In this study, GNBPs were not identified by proteomic analysis but were predicted in adult transcriptome of *E. intermedius* during manual searches of the database using sequences of *Drosophila* homologues. GNBPs and the related b-1,3-glucan-recognition proteins have been reported in several coleopterans, such as *Tribolium*. Evolutionarily b-1,3-glucan-recognition proteins are related to GNBPs and recognize bacterial and fungal cell wall components (Zou et al. 2007). Two inducible GNBP transcripts were identified in the burying dung beetle, *Nicrophorous vespilloides*, by suppression subtractive hybridization. These GNBPs cluster together with other coleopteran GNBPs from *Tribolium* (accession code NP_001164284) and *Tenebrio* (accession code BAG14263.1) with high bootstrap values (Vogel et al. 2011; Vogel et al. 2011).

Apolipophorin III was initially believed to be simply a subunit of lipid transport protein in the hemolymph (Pennington and Wells 2002). Later it emerged that apolipophorin III is important for insect immunity, and that it acts as a pattern recognition receptor that responds to b-1,3-glucan pathogen associated molecular

patterns found in fungi (Whitten et al. 2004; Rizwan-ul-Haq et al. 2011). Apolipophorin III has been shown to bind to bacterial and fungal cell wall components (Leon et al. 2006a, b) and is released into the haemolymph of insect larvae treated with lipopolysaccharide (Vogel et al. 2011). Here, the predicted *E. intermedius* apolipophorin III protein spots (molecular weight range 22–25 kDA) are significantly induced by fungal infection (Figure 2, Table 1). The precise effect of apolipophorin III interaction with pathogens is not well understood. It is reported, however, that apolipophorin III has a direct involvement with pathogens (Zdybicka-Barabas and Cytryńska 2011). The imaginal disc derived growth factor, also up-regulated by fungal infection (spot 7301), is a putative recognition protein with a chitin-binding lectin domain (Levy et al. 2004).

Serine proteases

Invertebrates respond to bacterial or fungal infection by the activation of three pathways: the toll, the immune deficiency, and the proPO pathways. The toll pathway is activated by bacterial and fungal virulence factors via pattern recognition receptors such as GNBP3 and peptidoglycan recognition proteins. Alternatively, the toll pathway is activated by "danger signals" via a rather poorly-defined pathway characterized by an SP cascade involving Psh (Chamy et al. 2008). Psh is also implicated in the proPO pathway, as the *Manduca sexta* orthologous protein HP6 was found to induce proPO activation in plasma (An et al. 2009). These are distinct pathways at the extracellular level, but less so intracellularly, where signals are transduced via the NFkB and Relish transcription factors leading to expression of antimicrobial peptides. The activation of immune response by Gram-positive bacteria and fungi is mediated by cascades of SP and results in the activation of

spaëtzle, a cytokine-like molecule that acts as ligand for the transmembrane receptor Toll. Many of these proteases are synthesized as inactive precursor zymogens, which are cleaved by limited proteolysis to generate the downstream effector molecules. The proPO pathway leading to melanization is activated by peptidoglycan, b-1,3-glucan ,and by lipopolysaccharide via a protease cascade whose components are currently under investigation by many researchers (Lee et al. 1998; Jiang and Kanost 2000; Jiang et al. 2005; An et al. 2009). In this study, no clear evidence of molecules in this pathway was observed.

Activation by the major fungal cell wall component, b-1,3-glucan, is mediated via GNBP3, followed by a protease cascade that culminates in the activation of spaëtzle. A detailed study of this pathway in a coleopteran has been conducted on the larvae of *Tenebrio molitor* (Roh et al 2009), providing evidence for the existence of the Toll pathway activated by virulence factors. Earlier, the activation of the *Drosophila* Toll pathway by fungal and bacterial proteases (or so-called danger signals) was shown to be mediated by a parallel proteolytic cascade involving Psh (Chamy et al). The mechanism by which Psh is activated by fungal proteases and the identity of its substrate have so far not been clearly described. This pathway is known to be controlled by a serpin known as Necrotic, and all *necrotic* phenotypes are dependent on Psh, indicating a direct relationship between the two molecules (Ligoxygakis et al. 2002).

A clear homologue of Psh has not been identified in coleopterans. Based on the phylogenetic tree, the *E. intermedius* protein encoded by CL23Contig1 may be the candidate orthologue of Psh. Protein spots 2203 and 3304, whose intensity increases significantly after fungal infections, yielded several fragments that matched parts of the sequences of proteins encoded by CL8Contig1 and CL20Contig1. It is noteworthy that a putative Necrotic transcript (CL111Contig1) was detected in the *E. intermedius* database. Necrotic has been identified as a protease inhibitor that directly regulates the action of Psh. It has been reported that *nec* phenotypes are pleiotropic, including constitutive activation of spaëtzle and spontaneous melanization (Ligoxygakis et al. 2002). This reinforces the possible involvement of Psh in both the toll and proPO pathways. In *T. molitor*, three serpins that negatively regulate the Toll and proPO pathways were identified. These serpins are specific for each of the three apical SPs in the Toll pathway, and at least two of them are inducible by Lys-type peptidoglycan and b-1,3-glucan (Jiang et al. 2009).

Conceivably, the identity of these putative Psh transcripts can be confirmed by a combination of biochemical and genetic experiments involving *psh* and *nec* mutants. In *Tribolium*, 30 SPs and 18 SPHs were predicted, and a putative Psh orthologue named Tc-Sp66 was identified on the basis of induction by Gram-positive bacteria and fungi (Zou et al. 2007). One orthologue was predicted to be Psh based on microarray evidence showing that its expression is induced only after fungal and Gram-positive bacterial challenge.

This study provides fair evidence for the presence of putative SPs that are up-regulated upon fungal infection. Better identification of these SPs requires cloning of the full-length transcripts as well as genetic and structural analyses of the genes and their respective proteins.

Other proteins

In addition to the identification of peptides that match the putative pattern recognition

receptors, trypsin-like SPs, and toll-like proteins, other proteins known to be involved in various aspects of host defense were putatively identified. These include transferrins, gelsolin, the OBP, and at least one protein tyrosine phosphatase. Transferrins are iron-binding proteins that regulate extracellular iron levels by sequestering iron atoms, thus preventing oxidative stress. They are particularly abundant in hematophagus insects such as Tsetse flies (Strickler-Dinglasan et al. 2006). Transferrin is often up-regulated after microbial infection and is part of the insect immune response, probably functioning as an antibiotic agent against pathogens (Lee et al. 2006).

Gelsolin is an actin-binding protein and is therefore a structural component of the cytoskeleton. It is found in two splice forms, cytoplasmic and secreted, in invertebrates such as *Drosophila* and is involved in cytoskeletal organization and biogenesis (Stella et al. 1994). It is involved in amyloid formation in vertebrates and haemolymph clotting in insects (Karlsson et al. 2004).

Insect OBPs are carrier proteins for transportation of small hydrophobic molecules through the haemolymph to receptors located in different tissues. They are small proteins (12–14 kDa) encoded by a divergent gene family. A comprehensive study showing diversity of these proteins and similarities among insects including coleoptera (*T. castaneum*) was conducted. This study showed emergence of subfamilies that probably arose through gene duplications (Gong et al. 2009). The existence of different isoforms that arose through gene duplications has been demonstrated in *T. molitor* (Graham et al. 2003). In *Tribolium*, remarkable expansion of odorant receptors was noted (Consortium 2008). Although OBP sequences differ significantly,

they share high structural similarities, characterized by four conserved cysteines, which contribute to two disulfide bonds, a b-barrel and two a-helices (Rothemund et al. 1999; Graham et al. 2001; Graham and Davies 2002). There is some evidence of OBP involvement in insect immunity, but this is poorly defined (Levy et al. 2004). OBPs have been identified in normal haemolymph of *T. molitor* with no clear function except their potential to transport small hydrophobic molecules to tissues (Graham et al. 2003). In *Drosophila*, an OBP was found to be up-regulated upon fungal infection and repressed by bacterial infection, suggesting an important role for these in immune response (Levy et al. 2004). Three protein spots were predicted to be differentially expressed OBPs, but only one of these was up-regulated upon fungal infection. The presence of multiple spots predicted to be OBPs suggests existence of different forms of the protein. The significance of the contiguous polyadenylation signals is currently not clear.

Acknowledgments

References

Altincicek B, Knorr E, Vilcinskas A. 2008. Beetle immunity: Identification of immune-inducible genes from the model insect *Tribolium castaneum*. *Developmental and Comparative Immunology* 32: 585–595.

An C, Ishibashi J, Ragan EJ, Jiang H, Kanost MR. 2009. Functions of *Manduca sexta* hemolymph proteinases HP6 and HP8 in two

innate immune pathways. *Journal of Biological Chemistry* 284(29): 19716–19726.

Anduaga S. 2004. Impact of the activity of dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae) inhabiting pasture land in Durango, Mexico. *Environmental Entomology* 33: 1306–1312.

Ascacio-Martínez JA, Barrera-Saldaña HA. 1994. Sequence of a cDNA encoding horse growth hormone. *Gene* 143(2): 299–300.

Bertone MA, Green JT, Washburn SP, Poore MH, Watson DW. 2006. The contribution of tunneling dung beetles to pasture soil nutrition. *Online. Forage and Grazinglands* DOI:10.1094/FG-2006-0711-02-RS.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantitites of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248–254.

Byrne MJ, Watkins B, Bouwer G. 2013. Do dung beetle larvae need microbial symbionts from their parents to feed on dung? *Ecological Entomology* 38: 250–257.

Chamy LE, Leclerc V, Caldelari I, Reichhart J-M. 2008. Sensing of 'danger signals' and pathogen-associated molecular patterns defines binary signaling pathways 'upstream' of Toll. *Nature Immunology* 9(10): 1165–1170.

Consortium TGS. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452(7190): 949–955.

Edmonds WD, Halffter G. 1978. Taxonomic review of immature dung beetles of the subfamily Scarabaeinae (Coleoptera:

Scarabaeidae). *Systematic Entomology* 3(4): 307–331.

Erezyilmaz DF. 2006. Imperfect eggs and oviform nymphs: a history of ideas about the origins of insect metamorphosis. *Integrative and Comparative Biology* 46(6): 795–807.

Farrell BD. 1998. "Inordinate Fondness" explained: Why are there so many beetles? *Science* 281: 555–559.

Fincher G. 1975. Effects of dung beetle activity on the number of nematode parasites acquired by grazing cattle. *Journal of Parasitology* 61(4): 759–762.

Forment J, Gilabert F, Robles A, Conejero V, Nuez F, Blanca J. 2008. EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. *BMC Bioinformatics* 9(1): 5.

Gong D-P, Zhang H-J, Zhao P, Xia Q-Y, Xiang Z-H. 2009. The odorant binding protein gene family from the genome of Silkworm, *Bombyx mori*. *BMC Genomics* 10(1): 332.

Graham LA, Brewer D, Lajoie G, Davies PL. 2003. Characterization of a subfamily of beetle odorant-binding proteins found in hemolymph. *Molecular & Cellular Proteomics* 2(8): 541–549.

Graham LA, Davies PL. 2002. The odorant-binding proteins of *Drosophila melanogaster*: annotation and characterization of a divergent gene family. *Gene* 292(1–2): 43–55.

Graham LA, Tang W, Baust JG, Liou YC, Reid TS, Davies PL. 2001. Characterization and cloning of a *Tenebrio molitor* hemolymph protein with sequence similarity to insect

odorant-binding proteins. *Insect Biochemistry and Molecular Biology* 31(6–7): 691–702.

Gupta S, Wang Y, Jiang H. 2005. *Manduca sexta* prophenoloxidase (proPO) activation requires proPO-activating proteinase (PAP) and serine proteinase homologs (SPHs) simultaneously. *Insect Biochemistry and Molecular Biology* 35(3): 241–248.

Handel K, Grünfelder CG, Roth S, Sander K. (2000). *Tribolium* embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. *Development Genes and Evolution* 210(4): 167–179.

Hoffmann JA. 2004. Primitive immune systems. *Immunological Reviews* 198: 5–9.

Holter P, Scholtz CH. 2007. What do dung beetles eat? *Ecological Entomology* 32: 690–697.

Imler J-L, Ferrandon D, Royet J, Reichhart J-M, Hetru C, Hoffmann JA. 2004. Toll-dependent and Toll-independent immune responses in *Drosophila*. *Journal of Endotoxin Research* 10(4): 241–246.

Jiang H, Kanost MR. 2000. The clip-domain family of serine proteinases in arthropods. *Insect Biochemistry and Molecular Biology* 30(2): 95–105.

Jiang H, Wang Y, Gu Y, Guo X, Zou Z, Scholz F, Trenczek TE, Kanost MR. 2005. Molecular identification of a bevy of serine proteinases in Manduca sexta hemolymph. *Insect Biochemistry and Molecular Biology* 35(8): 931–943.

Jiang R, Kim E-H, Gong J-H, Kwon H-M, Kim C-H, Ryu K-H, Park J-W, Kurokawa K, Zhang J, Gubb D, Lee B-L. 2009. Three Pairs

of Protease-Serpin Complexes Cooperatively Regulate the Insect Innate Immune Responses. *Journal of Biological Chemistry* 284(51): 35652–35658.

Karlsson C, Korayem AM, Scherfer C, Loseva O, Dushay MS, Theopold U. 2004. Proteomic analysis of the *Drosophila* larval hemolymph clot. *Journal of Biological Chemistry* 279(50): 52033–52041.

Kellenberger C, Leone P, Coquet L, Jouenne T, Reichhart K-M, Roussel A. 2011. Structure-function analysis of Grass clip serine protease involved in *Drosophila* Toll pathway activation. *Journal of Biological Chemistry* 286(14): 12300–12307.

Khanyile LM, Hull R, Ntwasa M. 2008. Dung beetle database:comparison with other invertebrate transcriptomes. *Bioinformation* 3(4): 159–161.

Kim MS, Baek MJ, Lee MH, Park JW, Lee SY, Söderhäll K, Lee BL. 2002. A new easter-type serine protease cleaves a masquerade-like protein during prophenoloxidase activation in *Holotrichia diomphalia* larvae. *Journal of Biological Chemistry* 277(42): 39999–40004.

Labandeira CC, Phillips TL. 1996. A Carboniferous insect gall: insight into early ecologic history of the Holometabola. *Proceedings of the National Academy of Sciences of the United States of America* 93(16): 8470–8474.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259): 680–685.

Lastro E. 2006. *Dung Beetles (Coleoptera: Scarabaeidae and Geotrupidae) in North Carolina Pasture Ecosystem (Under the*

direction of D. Wes Watson). Master of Science, North Carolina State University.

Lee KS, Kima BY, Kimb HJ, Seob SJ, Yoonc HJ, Choia YS, Kimd I, Hand YS, Jee YH, Leef SM, Kima DH, Sohna HD, Jina BR. 2006. Transferrin inhibits stress-induced apoptosis in a beetle *Free Radical Biology and Medicine* 41: 1151–1161.

Lee SY, Kwon TH, Hyun JH, Choi JS, Kawabata S-I, Iwanaga S, Lee BL. 1998. In vitro activation of pro-phenol-oxidase by two kinds of pro-phenol-oxidase-activating factors isolated from hemolymph of coleopteran, *Holotrichia diomphalia* larvae. *European Journal of Biochemistry* 254(1): 50–57.

Leon LJ, Idangodage H, Wan C-PL, Weers PMM. (2006a). Apolipophorin III: Lipopolysaccharide binding requires helix bundle opening. *Biochemical and Biophysical Research Communications* 348(4): 1328–1333.

Leon LJ, Pratt CC, Vasquez LJ, Weers PMM. (2006b). Tyrosine fluorescence analysis of apolipophorin III–lipopolysaccharide interaction. *Archives of Biochemistry and Biophysics* 452(1): 38–45.

Levy F, Bulet P, Ehret-Sabatier L. 2004. Proteomic analysis of the systemic immune response of *Drosophila*. *Molecular & Cellular Proteomics* 3(2): 156–166.

Ligoxygakis P, Pelte N, Hoffmann JA, Reichhart J. 2002. Activation of *Drosophila* Toll during infection by a blood serine protease. *Science* 297: 114–116.

Liu PZ, Kaufman TC. 2005. Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. *Evolution & Development* 7(6): 629–646.

Masuda N, Watahiki M, Tanaka M, Yamakawa M, Shimizu K, Nagai K, Nakashima K. 1988. Molecular cloning of cDNA encoding 20 kDa variant human growth hormone and the alternative splicing mechanism. *Biochimica et Biophysica Acta* 949: 125–131.

Montes EDO, Halffter G. 1998. Invasion of Mexico by two dung beetles previously introduced into the United States. *Studies on Neotropical fauna and Environment* 33(1): 37–45.

Ntwasa M, Goto A, Kurata S. 2012. Coleopteran Antimicrobial Peptides: Prospects for Clinical Applications. *International Journal of Microbiology* doi: 10.1155/2012/101989

Park J-W, Kim C-H, Rui J, Park K-H, Ryu K-H, Chai J-H, Hwang H-O, Kurokawa K, Ha N-C, Söderhäll I, Söderhäll K, Lee BL. 2010. Beetle Immunity. In: Söderhäll K, Editor. *Invertebrate Immunity*. pp. 163-180. Springer US.

Pennington JE, Wells MA. 2002. Triacylglycerol-rich lipophorins are found in the dipteran infraorder Culicomorpha, not just in mosquitoes. *Journal of Insect Science* 2:15. Available online: http://www.insectscience.org/2.15/

Pomfret JC, Knell RJ. 2006. Immunity and the expression of a secondary sexual trait in a horned beetle. *Behavioral Ecology* 17(3): 466–472.

Rabilloud T, Strub JM, Luche S, van Dorsselaer A, Lunardi J. 2001. A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as

fluorescent stains for protein detection in gels. *Proteomics* 1(5): 699–704.

Rizwan-ul-Haq M, Gong L, Hu M, Luo J. 2011. Apolipophorin III and transmission electron microscopy as toxicity indicators for harmaline and tea saponin in *Spodoptera exigua* (Noctuidae: Lepidoptera). *Chemosphere* 85(6): 995–1001.

Rothemund S, Liou Y, Davies P, Krause E, Sönnichsen F. 1999. A new class of hexahelical insect proteins revealed as putative carriers of small hydrophobic ligands. *Structure* 15(11): 1325–1332.

Savard J, Tautz D, Lercher MJ. 2006. Genome-wide acceleration of protein evolution in flies (Diptera). *BMC Evolutionary Biology* 6: 7.

Stella MC, Schauerte H, Straub KL, Leptin M. 1994. Identification of secreted and cytosolic gelsolin in *Drosophila*. *Journal Cell Biology* 125(3): 607–616.

Strickler-Dinglasan PM, Guz N, Attardo G, Aksoy S. 2006. Molecular characterization of iron binding proteins from Glossina morsitans morsitans (Diptera: Glossinidae). *Insect Biochemistry and Molecular Biology* 36(12): 921–933.

Tautz D, Sommer RJ. 1995. Evolution of segmentation genes in insects. *Trends in Genetics* 11(1): 23–27.

Valanne S, Wang J-H, Rämet M. 2011. The *Drosophila* Toll signaling pathway. *Journal of Immunology* 186(2): 649–656.

Van der Zee M, da Fonseca RN, Roth S. 2008. TGFβ signaling in *Tribolium*: vertebrate-like components in a beetle. *Development Genes and Evolution* 218(3): 203–213.

Vierstraete E, Verleyen P, Baggerman G, D'Hertog W, Bergh GVD, Arckens L, Loof AD, Schoofs L. 2003. A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph. Proceedings of the *National Academy of Sciences of the United States of America* 101(2): 470–475.

Vogel H, Altincicek B, Glockner G, Vilcinskas A. 2011. A comprehensive transcriptome and immune-gene repertoire of the lepidopteran model host Galleria mellonella. *BMC Genomics* 12(1): 308.

Vogel H, Badapanda C, Vilcinskas A. 2011. Identification of immunity-related genes in the burying beetle Nicrophorus vespilloides by suppression subtractive hybridization. *Insect Molecular Biology* 20(6): 787–800.

Whitten MMA, Tew IF, Lee BL, Ratcliffe NA. 2004. A novel role for an insect apolipoprotein (apolipophorin III) in β-1,3-glucan pattern recognition and cellular encapsulation reactions. *Journal of Immunology* 172(4): 2177–2185.

Zdybicka-Barabas A, Cytryńska M. 2011. Involvement of apolipophorin III in antibacterial defense of *Galleria mellonella* larvae. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 158(1): 90–98.

Zou Z, Evans JD, Lu Z, Zhao P, Williams M, Sumathipala N, Hetru C, Hultmark D, Jianh G. 2007. Comparative genomic analysis of the *Tribolium* immune system. *Genome Biology* 8: R177.

Table 1. Up-regulated and down-regulated proteins.

	NCBI Predicted protein	Accession Number	DB ID	Ion Precursor	Peptide	Matching peptide (DB database)	Prediction (DB database)	Relative spot density
101	Similar to Pez CG9493-PA	XP_392377.2	CL32Contig1	673.91 $^{2+}$	LDALGQECLK	IDAIGQECLK	PBP/GOBP family	Uninfected 0.01 ≤ Infected 0.01 ≤ 0.193 ≥ 0.01 $p = 0.013$ F = 0.3535
2203F	Aic2 trypsinogen	AAF71515.1	CL8Contig1 / CL8Contig3	1036.10 $^{2+}$	SNAACAAAYG	SNAACAAAYG	Trypsin-like serine protease	Uninfected 0.012 ≤ 0.044 ≥ 0.012
			CL8Contig1	760.49 $^{2+}$	PAYPGVYSSV	PAYPGVYSSV	similar to Trypsin 29F	Infected 0.01 ≤ 0.13 ≥ 0.01
				578.83 $^{2+}$	ADQSSFTLK		unidentified	
			CL15Contig1	560.34 $^{2+}$	TSYPGNLPDR	TSYPGNIPDR	trypsin-like serine	
			CL20Contig1	890.58 $^{2+}$	FTSPLQYTDTLQLPR		similar to trypsin-	
2203U	-	EAT41458.1	006991_1682_0418_c_s	763.43 $^{2+}$	VSSAGCASGA	VSSAGCASGA	trypsin-like serine protease	$p = 0.0054$ F = 0.4388
			CL20Contig1					
			CL32Contig1		TGGMNADGTFNK	TGGMNADGTFNK	PBP/GOBP family	
			-		LADVRAHL		unidentified	
3004	Similar to Pez CG9493-PA	XP_392377.2	CL32Contig1	673.91 $^{2+}$	LDALGQECLK	IDAIGQECLK	PBP/GOBP family	Uninfected 0.103 ≤ Infected 0.31 ≤ 0.77 ≥ 0.31 $p = 0.1481$ F = 0.1020
3304	Serine protease	EAT46512.1	-	664.40 $^{2+}$	VVLTAAHAVLQ			Uninfected 0.015 ≤ 0.026 ≥ 0.015
	-		009895_1663_3195_c_s	548.81 $^{2+}$	FSLCAGGEQK	FNLCAGG	unidentified	Infected 0.03 ≤ 0.136 ≥ 0.03
			CL339Contig1	733.96 $^{2+}$	TQGSPLVCLP	TPLICLP	similar to CG17633 (peptidase)	$p = 0.048$ F = 0.2865
3403	Chaoptin	EAT39239.1	no hit	626.86 $^{2+}$	YPTNALPSFDK	-		Uninfected 0.08 ≤ 0.381 ≥ 0.08
	Similar to xmas-2 CG32562-PA	XP_395462.3	No hit	986.02 $^{2+}$	LCAEETSATQ	-		
	-	-	-	536.82 $^{2+}$	SAASG	-	unidentified	Infected 0.003 ≤ 0.003 ≥ 0.003
	-	-	-	687.33 $^{2+}$	SLSEPESCLT	-	unidentified	
	-	-	--	626.86 $^{2+}$	YPTNALPSFDK	-	unidentified	$p = 0.0094$ F = 0.0017
	Similar to CG7509-PA (with leucine rich-repeats)	XP_975320.1	-	699.37 $^{2+}$	NGEYLNFG	-	unidentified	
4001	ENSANGP00000006676 (putative glycoprotein hormone rk-like receptor) with Leucine-rich repeats	XP_317111.1	CL673Contig1*	665.65 $^{2+}$	EASTLAEFR	-	Similar to Toll 6	Uninfected 0 ≤ 0 ≥ 0
	Similar to Pez CG9493-PA	XP_392377.2	CL32Contig1	673.88 $^{2+}$	LDALGQECLK	IDAIGQECLK	PBP/GOBP family	Infected 0.03 ≤ 0.91 ≥ 0.03, $p = 0.001$
4002U	GA19474-PA	EAL30835.1	CL32Contig1	737.95 $^{2+}$	QEALDALGQECLK	IDAIGQECLK	PBP/GOBP family	Uninfected 0.46 ≤ 3.26 ≥ 0.46 Infected 0.37 ≤ 2.03 ≥ 0.37 $p = 0.1012$ F = 0.2794
5102F	CAD	AAQ67196.1	-	669.85 2+	PAETAYLYQK	-	-	Uninfected 0.07 ≤ 0.453 ≥ 0.07
	ENSANGP00000001529	XP_311192.2	-	630.30 2+	QFDDGGDQHLK	-	-	Infected 0.094 ≤ 0.736 ≥ 0.094 $p = 0.0835$ F = 0.4337
5310	Gelsolin precursor	EAT42358.1	014446_1673_2874_c_s*	768.90 $^{2+}$	YAPGGVASGFNQVNR	-	Gelsolin	Uninfected 0.01 ≤ 0.14 ≥ 0.01 Infected 0.1 ≤ 0.54 ≥ 0.1 $p = 0.0125$ F = 0.0213
5312	Gelsolin precursor	EAT42358.1	014446_1673_2874_c_s*	768.89 $^{2+}$	YAPGGVASGFNQVNR	-	Gelsolin	Uninfected 0.07 ≤ 1.24 ≥ 0.07 Infected 0.05 ≤ 0.05 ≥ 0.05 $p = 0.003$ F = 0.3389
5406	ENSANGP00000002901	EAA07673.3	-	905.46 $^{2+}$	NSPSEPLNVGLLYR	-	-	Uninfected 0.56 ≤ 1.19 > 0.56 Infected 0 ≤ 0 ≥ 0 $p = 0.0001$

(*) indicates sequences that were not identified directly from searches with the short peptides but indirectly by using the predicted NCBI orthologues to search the *Euonoticellus* database. The graphs of the relative spot density values represent the means of three biological replicates. The error bars indicate the standard error of these means. A two-tailed Student's *t*-test was performed at the 95% confidence interval to establish if there was a significant difference between the levels of protein expression between treated and untreated beetles, as represented by their optical density. *p*-values less than 0.05 demonstrate a significant difference between the challenged and unchallenged groups for each spot. F-values higher than 0.05 indicate that the variance between challenged and unchallenged groups is not equal.

Table 1 continued.

6005	Hypothetical protein	EAT45043.1	-	646.85 $^{2+}$	LSQFDDDVLLK	-	-	Uninfected $0 \leq 0 \geq 0$
								Infected $0.12 \leq 0.35 \geq 0.12$
								$p = 0.0001$
6203	ENSANGP00000016415	XP_306996.2	011224_1655_1295_c_s	594.84 $^{2+}$	SQLETLAN	SQLETLAN	Cytochorme P450	Uninfected $0.75 \leq 1 \geq 0.75$
	Transient receptor potential channel 4	EAT4295.1	-	596.83 $^{2+}$	VGAESFDLAGVK	-		Infected $0.48 \leq 0.95 \geq 0.48$
								$p = 0.995$ F = 0.2873
6204	GA18404-PA	EAL33860.1	011178_1666_3502_c_s	557.32 2+	NQLLQLEQK	NQLLQLEQK	Cytochorme P450	Uninfected $0.26 \leq 0.5 \geq 0.26$
	CG10639-PA	NP_609923.2	-	715.88 2+	ADLDGNLLEDK	-	-	Infected $0.3 \leq 0.4 \geq 0.3$
								$p = 0.776$ F = 0.4353
6410	Similar to transferrin 2 CG10620-PA	XP_396618.1	-	704.85 $^{2+}$	LVEGDGDVAFVK	-	-	Uninfected $0.03 \leq 0.58 \geq 0.03$
	Transferrin	ABI31834.1	012952_1546_3581_c_s	562.79 $^{2+}$	HVTVPGNTDGK	IGIPGNPDGR	similar to CG31997	
	Transferrin	ABI31834.1	-	618.31 $^{2+}$	TQEEPEAEFR	-	-	Infected $0.08 \leq 0.94 \geq 0.08$
	Transferrin	ABI31834.1	008196_1603_1902_c_s	777.94 $^{2+}$	VLELSENNVAKPNK	ELSHNNVAFP	eIF-4a	$p = 0.0365$ F = 0.0882
	Transferrin	ABI31834.1	012044_1541_2861_c_s*	713.39 $^{2+}$	AGYNAPLYTLVK	-	Similar to transferrin	
7106	Putative carbohydrate binding glycosyltransferase	ZP_01357938.1	CL351Contig1	981.98 2+	SLLDSYPTAYTNVK	DTYPTAY	unidentified	Uninfected $0.02 \leq 0.04 \geq 0.02$ Infected $0.08 \leq 0.27 \geq 0.08$ $p - 0.0475$ F $- 0.8$
7203	-	-	CL123Contig1	595.33 $^{2+}$	RAESFDQLVK	AESFDQLVK	apolipophorin-III	Uninfected $0.04 \leq 0.78 \geq 0.04$ Infected $0.35 \leq 2.6 \geq 0.35$ $p = 0.0067$ F = 0.011
7210	ENSANGP00000021859	XP_313925.1	CL123Contig1	532.78 $^{2+}$	AGENVQAFTK	AGENVQAFTK	apolipophorin-III	Uninfected $0.05 \leq 0.06 \geq 0.05$
	GA18404-PA	EAL33860.1	011178_1666_3502_c_s	557.33 $^{2+}$	NQLLQLEQK	NQLLQLEQK	Apolipophorin III	Infected $0.061 \leq 0.38 \geq 0.061$ $p = 0.0167$ F = 0.4043
7301F	IDGF PROTEIN	NP_001036847.1	-	713.39 2+	GLLSYPEVCGSK	-	-	Uninfected $0.22 \leq 0.323 \geq 0.22$
	Oligopeptidase	EAT48748.1	012961_1674_0476_c_s	836.91 2+	NYPAPLYSL	-	similar to chitinase 5	Infected $0.01 \leq 0.95 \geq 0.01$
7301U	IDGF 4 CG1780-PA	XP_396769.2	015002_1611_0756_c_s	725.94 2+	TPGLLSYPEV	HYPLPLYSV	-	$p = 0.1045$ F = 0.0021
7303	Similar to general receptor for phosphoinositides 1 CG11628-PA	XP_001123194.1	-	620.33 2+	HWLETKPSNK	-	-	Uninfected $0.0294 \leq 0.0486 \geq 0.0294$
	IDGF 4 CG1780-PA	XP_396769.2	-	725.93 2+	TPGLLSYPEV	-	-	Infected $0.044 \leq 0.286 \geq 0.044$ $p = 0.0149$ F = 0.3044
8306	Chaoptin	EAT39239.1	-	626.87 $^{2+}$	YPTNALPSFDK	-	-	Uninfected $0.21 \leq 0.573 \geq 0.21$ Infected $0.0133 \leq 0.0133 \geq 0.0133$ $p = 0.05$ F = 0.004
8402	Similar to CG13886-PA	XP_394035.2	-	930.08 $^{2+}$	LLTQQAECL	-	-	Uninfected $0.17 \leq 0.34 \geq 0.17$ Infected $0.294 \leq 0.586 \geq 0.294$ $p = 0.5147$ F = 0.2498
9003	Similar to Pez CG9493-PA	XP_392377.2	CL32Contig1	673.88 $^{2+}$	LDALGQECLK	IDAIGQECLK	PBP/GOBP family	Uninfected $0.082 \leq 0.113 \geq 0.082$
	Conserved hypothetical protein	EAT39907.1	-	550.34 $^{2+}$	VSSATMSR	-	-	Infected $0.035 \leq 0.323 \geq 0.035$
	-	-	-	-	-	-	-	$p = 0.1506$ F = 0.3361

Table 2. Comparison of the dung beetle database with those of other invertebrates. The total number of unigenes in the database was 2662.

Database	Number of unigenes	%
Acyrthosiphon_pisum	899	34
Apis_melifera	998	38
Bombyx_mori	951	36
Drosophila_melanogaster	995	37
Tribolium_castaneum	1297	49
All databases	1334	50

Table 3. Predicted immune genes found in *Euonoticellus intermedius* transcriptome database.

| Flylabgenomebase ID	*Tribolium* orthologue	Drosophila orthologue	*Drosophila nucleotide*	HUMAN orthologue	EST2Uni
003936_1548_1151_c_s	NA	Dorsal interacting protein 4	NA	Ubiquitin-conjugating enzyme E2 N	CG7656-PC
004657_1617_2913_c_s	NA	Dorsal interacting protein 4	NA	Ubiquitin-conjugating enzyme E2 N	UbcD6-PA
005837_1622_2607_c_s	NA	Ras-related protein Rac2	NA	NA	Rho1-PA
006324_1633_0830_c_s	NA	Defense protein l(2)34Fc	NA	NA	NA
006345_1608_3088_c_s	NA	GK20406	NA	NA	CG6860-PB
007008_1624_1394_c_s	Toll-like	NA	NA	NA	NA
007466_1614_3158_c_s	NA	Ras-related protein Ral-a	NA	NA	Ras64B-PA
007472_1671_2184_c_s	NA	Tyrosine-protein kinase hopscotch	hop	NA	na
007472_1671_2184_c_s	NA	Tyrosine-protein kinase hopscotch	Tak1	NA	na
008422_1673_2668_c_s	NA	NA	Thor	NA	NA
008655_1605_1151_c_s	NA	Deformed epidermal autoregulatory factor 1	NA	NA	Deaf1-PB
008815_1595_2123_c_s	NA	Tyrosine-protein kinase hopscotch	NA	Ribosomal protein S6 kinase alpha-1	Src42A-PA
009315_1547_2493_c_s	NA	Mitogen-activated protein kinase 14B	p38b	Cyclin-dependent kinase 1	Cdk9-PA
009933_1651_3213_c_s	NA	Dual oxidase	NA	NA	Pxd-PA
010643_1552_3233_c_s	NA	Ras-related protein Ral-a	NA	NA	ltd-PC
010966_1592_1457_c_s	NA	NA	dnr1	NA	NA
CL111Contig1	NA	NA	nec	NA	Spn4-PG
CL138Contig1	NA	Dual oxidase	NA	NA	TpnC47D-PA
CL147Contig1	NA	Mitogen-activated phosphatase 3,	NA	Dual specificity protein phosphatase 3	Mkp-PB
CL15Contig1	NA	Sphinx1 or Sphinx2 or persephone	NA	Complement C1r subcomponent-like protein	CG5255-PA
CL20Contig1	NA	Sphinx1 or persephone	NA	Complement C1r subcomponent-like protein	Jon65Aiv-PA
CL23Contig1	NA	Sphinx1 or persephone	NA	Complement C1r subcomponent-like protein	CG10472-PA
CL283Contig1	NA	Protein spaetzle	NA	NA	NA
CL426Contig1	Toll-7-like protein	Toll-6 or Toll-9	18w	NA	NA
CL444Contig1	NA	Protein croquemort	crq	NA	NA
CL45Contig1	NA	40S ribosomal protein S6	RpS6	NA	RpS6-PB
CL515Contig1	NA	Inhibitor of nuclear factor kappa-B kinase subunit	Mpk2	Cyclin-dependent kinase 1	NA
CL515Contig1	NA	Mekk1, isoform B	Mpk2	Cyclin-dependent kinase 1	NA
CL673Contig1	Toll-6	18 wheeler/Protein toll	18w	Toll-like receptor 3	Als-PA
CL8Contig1	NA	Persephone/ Sphinx1 /Sphinx2	NA	Complement C1r subcomponent-like protein	Try29F-PA
Manually identified genes					
007956_1645_0789_c_s		Gram-negative bacteria binding protein 3			
CL47Contig1		Spaëtzle processing enzyme			
CL652Contig1		Gram-negative bacteria binding protein 1			
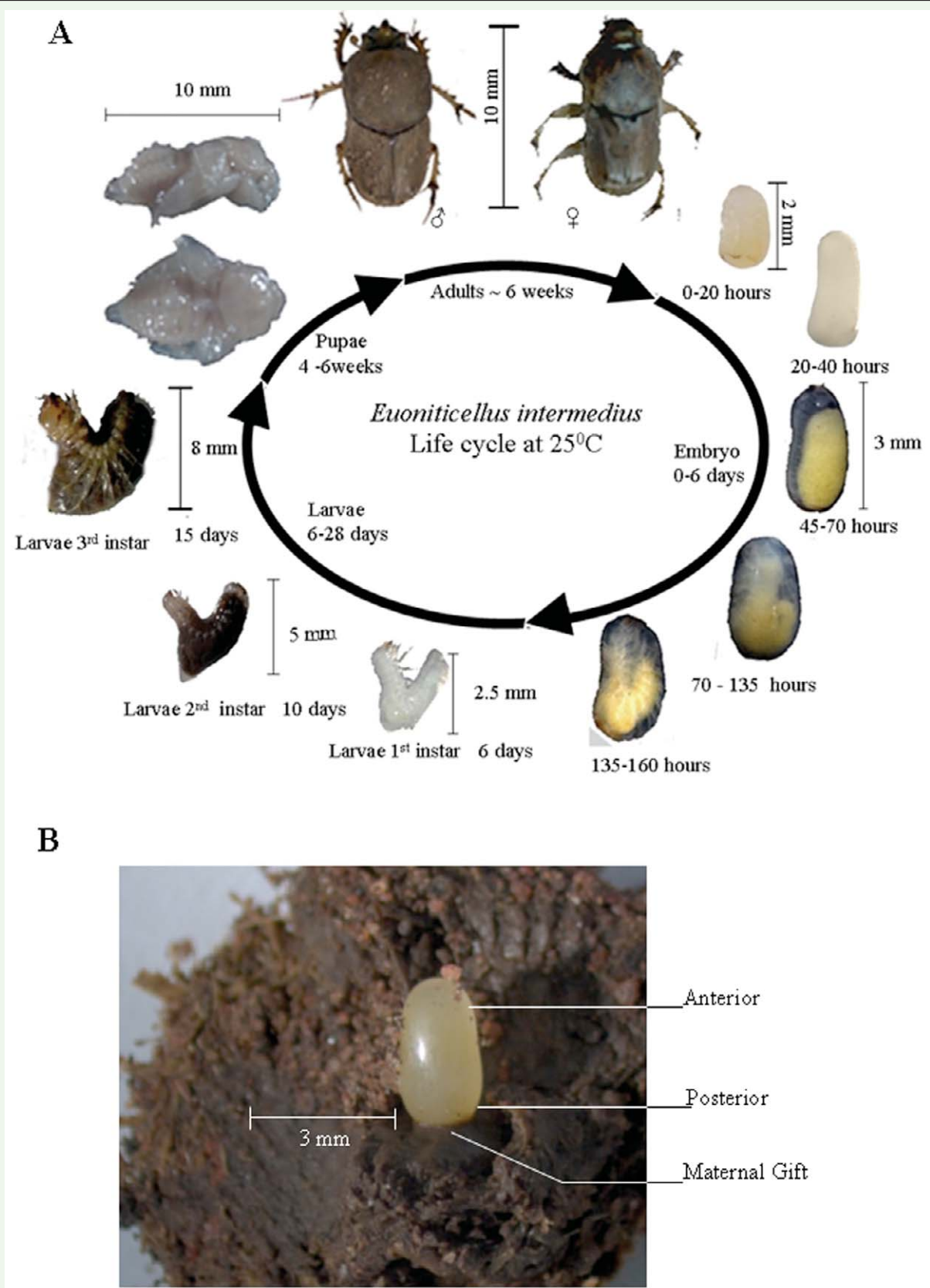
Figure 1. The life cycle of *Euonoticellus intermedius*. A. Developmental stages from embryo to adult. B. An embryo in a brood ball showing the maternal gift. High quality figures are available online.
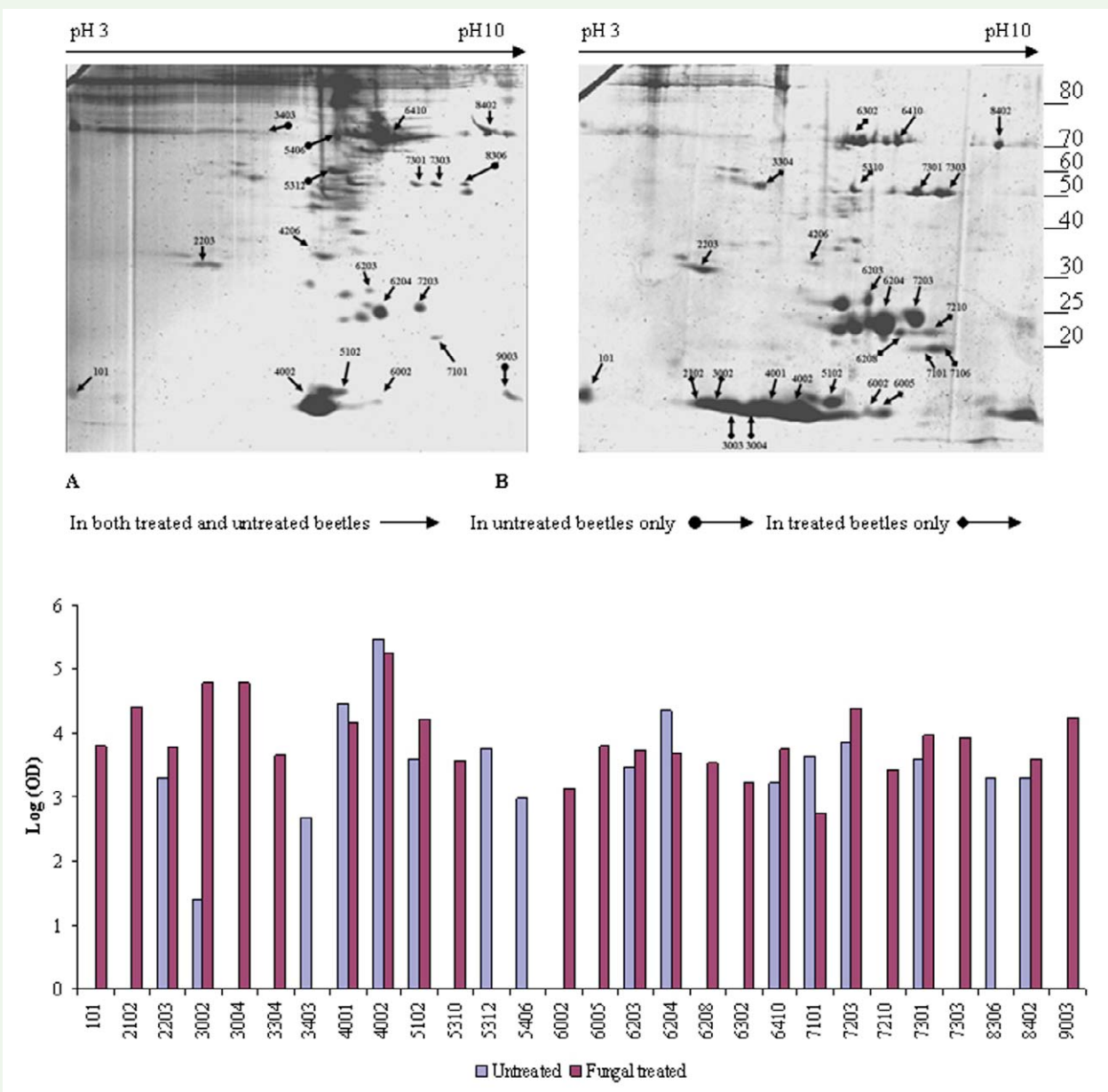
Figure 2. 2-D gel electrophoresis of *Euonoticellus intermedius* hemolymph proteins. The gel was stained with Coomasie blue. It shows protein spots in unchallenged (A) and in infected beetles (B). Proteins found in treated beetles only, in untreated beetles only, and in both treated and untreated beetles are shown. C. Optical density measurements (N= 3) of protein spots. High quality figures are available online.
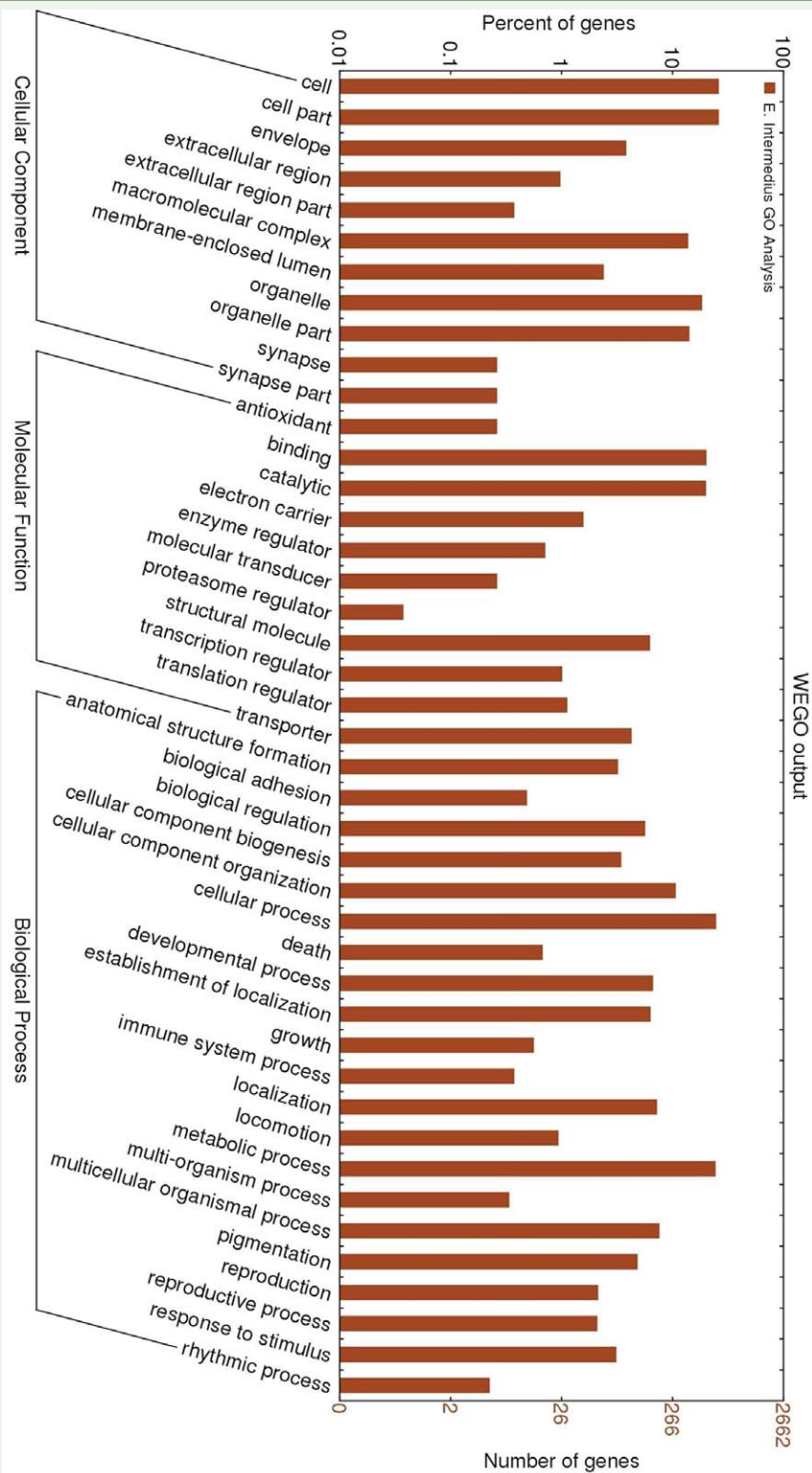
Figure 3. Gene Ontology Analysis of *Euonoticellus intermedius* genes. From the total of 2662 unigenes, 933 were functionally annotated by blast result using *Drosophila melanogaster*. A gene ontology study was then conducted on the 933 unigenes made up of 338 contigs and 595 singletons. The *D. melanogaster* gene ontology association file was used to annotate the unigenes with Gene Ontology terms, and results were stored in a local Mysql database. The gene ontology analysis is implemented by one of the modules of the EST2uni pipeline (Forment et al. 2008). The retrieval of all 2662 unigenes, left-joined by gene ontology terms, resulted in a WEGO (Web Gene Ontology) annotation plot. The Native Format file and WEGO were used for plotting GO annotation results as a histogram. The gene ontology terms have 1 to 7 levels of detail, 1 being most general and 7 most detailed/specific. The histogram shown here is of GO terms at level 2. High quality figures are available online.
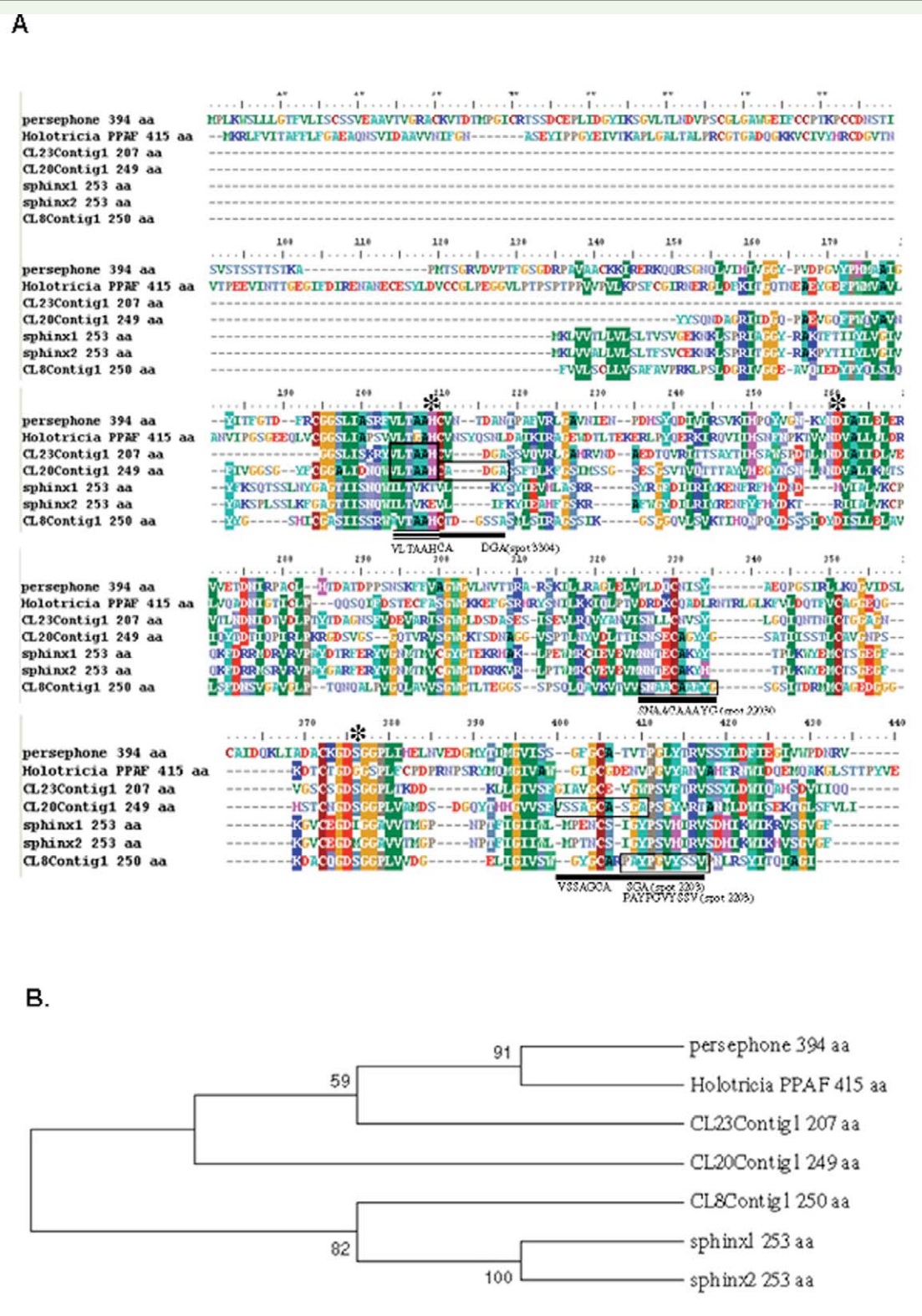
Figure 4. Sequence alignment of known proteases with predicted database polypeptides. A. *Drosophila* Persephone, sphinx proteins and *Holotricia diomphalia* PPAF were aligned with putative *Euonoticellus intermedius* SP domain proteins. The alignment was performed using the Clustal W program. The position of the catalytic triad is marked with asterisks. The triad is conserved in all three *E. intermedius* proteins. Peptide sequences obtained from the mass spectrometric analysis of the protein spots 2203 (single line) and 3304 (three lines) are aligned with one of the three *E. intermedius* sequences as shown. B. Phylogenetic tree of the abovementioned proteins. The tree was constructed using the neighbour joining method and the MEGA 5 software with 500 bootstrap replications. High quality figures are available online.
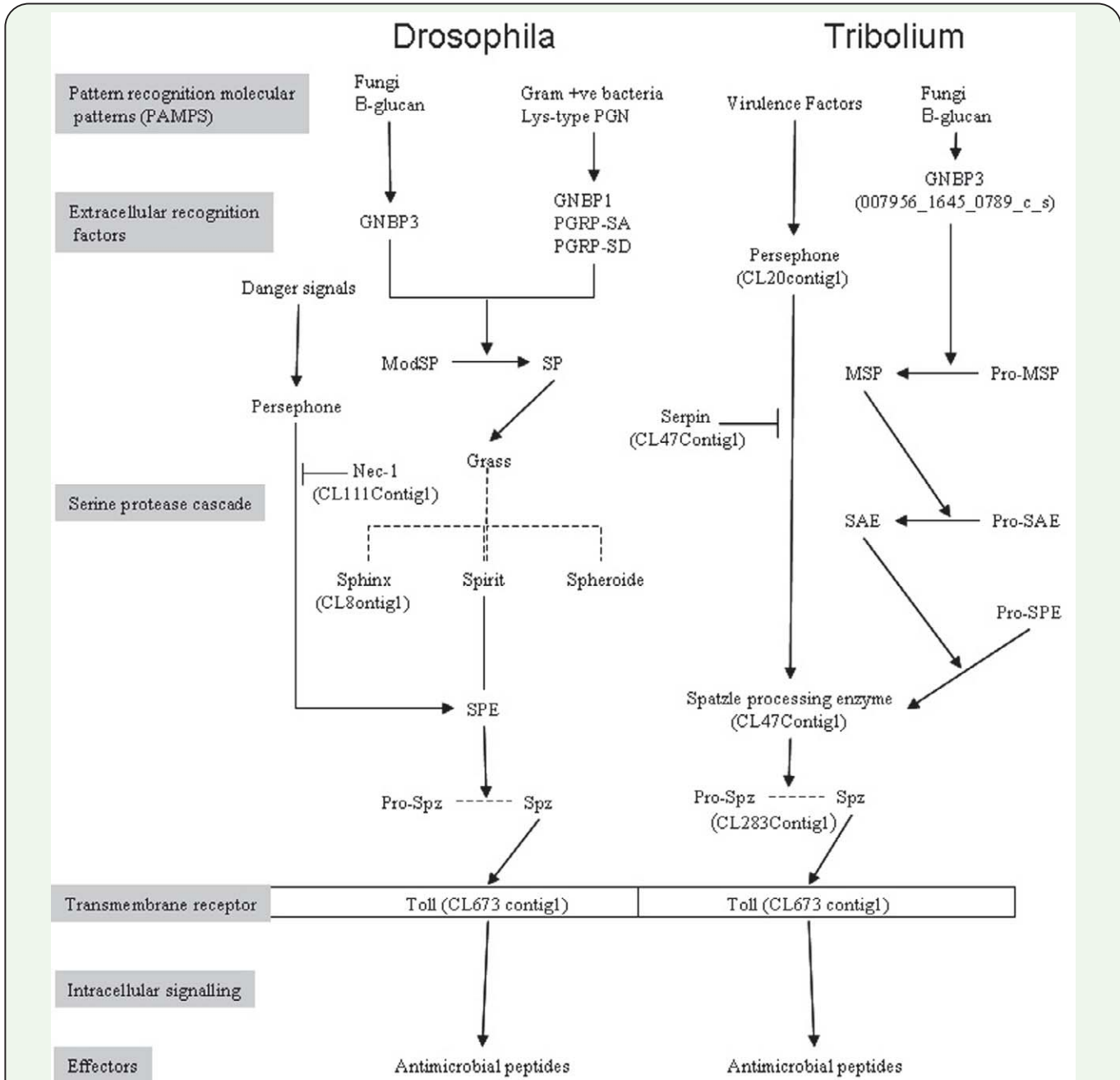
Figure 5. Schematic diagram of extracellular immune signaling in *Drosophila* and *Tribolium*. Pathogens are recognized by pattern recognition patterns (PRRs) based on pathogen associated molecular patterns (PAMPs) present on the pathogen. PAMPS may be Lys-type or DAP-type peptidoglycans. The pathway can also be activated by endogenous factors produced in live organisms (danger signals or virulence factors). The Toll pathway is activated when Gram-positive bacteria are sensed by the peptidoglycan recognition proteins (PGRP-SA). Fungal PAMPS are sensed by the glucan binding protein 3 (GNBP3). These PRRs are secreted proteins, and when they form complexes with appropriate PAMPS, they initiate the serine protease cascades in the hemolymph, culminating in the cleavage of pro-spaëtzle, converting it to active spaëtzle, the endogenous ligand of the Toll receptor. In *Drosophila*, a parallel protease cascade is activated by danger signals such as fungal and bacterial proteases and activates the Toll pathway via the trypsin-like serine protease known as Persephone and spaëtzle. Binding of spaëtzle to Toll causes a conformational change and subsequent recruitment of intracellular signaling molecules. Intracellular signal transduction is mediated by a phosphorylation cascade that culminates in the release of NF-kB-like transcription factors from a complex with cactus. These transcription factors activate expression of genes that encode for antimicrobial peptides and other effectors. High quality figures are available online.
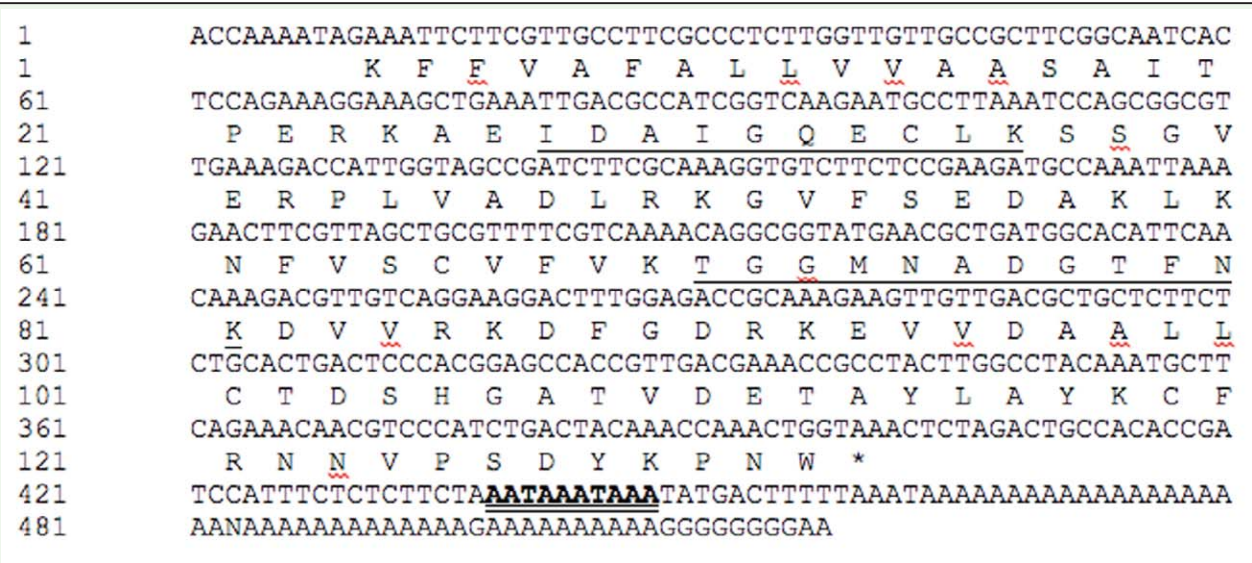
Figure 6. Translation of CL32Contig1. Peptide fragment sequenced obtained by MS/MS are underlined. Putative overlapping tandem polyadenylation signals in bold and double underline. High quality figures are available online.