Inducible defenses stay up late: temporal patterns of immune gene expression in *Tenebrio molitor*

Paul R Johnston, Olga Makarova, Jens Rolff

Evolutionary Biology, Institute for Biology, Free University Berlin, Berlin, Germany
Running title: Temporal patterns of immune gene expression
Abstract

The course of microbial infection in insects is shaped by a two-stage process of immune defense. Constitutive defenses, such as engulfment and melanization, act immediately and are followed by inducible defenses, archetypically the production of antimicrobial peptides (AMPs), which eliminate or suppress the remaining microbes. By applying RNAseq across a seven-day time-course we sought to characterize the long-lasting immune response to bacterial challenge in the mealworm beetle *Tenebrio molitor*, a model for both the biochemistry of insect immunity and persistent bacterial infection. By annotating a hybrid *de novo* assembly of RNAseq data, we were able to identify putative orthologs for the vast majority of components of the conserved insect immune system. Compared with *Tribolium castaneum*, the most closely related species with both a reference genome sequence and a manually-curated immune system annotation, the *T. molitor* immune gene count was lower with lineage-specific expansions of genes encoding serine proteases and their countervailing inhibitors accounting for the majority of the deficit. Quantitative mapping of RNAseq reads to the reference assembly showed that expression of genes with predicted functions in cellular immunity, wound healing, melanization and the production of reactive oxygen species was transiently induced immediately after immune challenge. In contrast, expression of genes encoding antimicrobial peptides or components of the toll signalling pathway and iron sequestration response remained elevated for at least 7 days. Numerous genes involved in metabolism and nutrient storage were repressed indicating a possible cost of immune induction. Strikingly, the
expression of almost all antibacterial peptides followed the same pattern of long-lasting induction, regardless of their spectra of activity, signaling possible interactive roles in vivo.

**Introduction**

Because of the importance of insects as models for vertebrate immunity (Lemaitre *et al.* 1996) and as vectors of disease (Enayati and Hemingway 2010), insect immune defenses have been studied in great detail (Rolff and Reynolds 2009; Kounatidis and Ligoxygakis 2012) and the interplay between constitutive and hence fast-acting immune responses and inducible defenses has been elucidated. Similar to vertebrates, insect immunity comprises a suite of constitutive responses such as phagocytotic engulfment, melanization and production of reactive oxygen as well as inducible components such as antimicrobial peptides (Rolff and Reynolds 2009; Kounatidis and Ligoxygakis 2012).

Insect immune systems and more generally invertebrate immune systems, however, are devoid of B- and T-cell mediated memory. Presumably this perceived lack of a memory mechanism explains why most studies on insect immune gene expression capture only up to 48hrs after infection.

Yet, many parasites, such as *Plasmodium* (Michel and Kafatos 2005) or microsporidia (Schwarz and Evans 2013) are present in the host for several days. It has been frequently reported that bacterial infections can persist in insect hosts for several days to even weeks. Persistent infections can also be beneficial. Mutualistic relationships with microbes are often established for the
lifetime of the host and interactions can be mediated by the insect immune system, for example by antimicrobial peptides such as coleoptericins (Login et al. 2011).

Independent of persistent infection, elevated antimicrobial responses in insects can be long-lasting. In the silk moth 9 days of elevated antimicrobial activity have been reported (Faye et al. 1975), 11 days in *Rhodnius prolixus* (Azambuja et al. 1986), 14 days in bumble bees (Korner and Schmid-Hempel 2004), 21 days in our model *Tenebrio molitor* (Haine, Pollitt, et al. 2008) and 44 days in dragonflies (Bulet et al. 1992). Hence the duration of the elevated antimicrobial response can be a significant part of total life span in many insects.

Upon infection insects utilize an array of recognition and effector systems adapted to bacterial, viral and eukaryotic pathogens. Recognition of bacterial infection has been intensively studied in *Drosophila melanogaster* and also *T. molitor* (Park et al. 2011) where Lysine-type peptidoglycan from Gram-positive bacteria and Diaminopimelic-type peptidoglycan from Gram-negative activate signaling via the Toll and IMD pathways respectively. After a breach of the cuticle, constitutive defences including phenoloxidase, some lysozymes and phagocytotic cells act quickly. Phagocytes are analogous to human macrophages, and recognize microbes using receptors and opsonins such as scavenger receptors, thio-ester proteins (TEPs) or the highly variable, alternatively-spliced Dscam (Cherry and Silverman 2006).

The insect equivalent to the liver, the fat body, is not only of great metabolic importance, but is pivotal in the production of inducible immune effectors
including antimicrobial peptides that follow constitutive responses over the course of an infection. The inducible antimicrobial defense responses are elicited by recognition of conserved MAMPs (Microbe Associated Molecular Patterns) by PGRPs and/or GNBPs which induce the Toll and IMD signal transduction cascades, complemented by the Jak/Stat and JNK pathways, and activate the NF-kappaB transcription factors relish, dorsal and dif which induce expression of antimicrobial peptides (Kounatidis and Ligoxygakis 2012). These pathways are conserved in many insects including disease vectors such as mosquitoes (Kafatos et al. 2009) and the ancient odonates (Johnston and Rolff 2013).

Recent work suggested that the persistence of bacterial infections is shaped by a two-stage process of insect immune defenses (Schneider and Chambers 2008). Haine et al (Haine, Moret, et al. 2008) reported from an infection experiment in T. molitor that the vast majority of Staphylococcus aureus are cleared within 1 hr of injection, yet induced antimicrobial activity is only detected after 6 hours and peaks even later around day 4 (Haine, Moret, et al. 2008). Bacteria which survive the initial immune response are more resistant to host defenses upon re-infection (Haine, Moret, et al. 2008). These observations led to the suggestion that fast-acting constitutive immune responses for example, melanization, phagocytic engulfment and generation of reactive oxygen species clear the majority of the infection and that the main function of the inducible immune response is to “mop up” remaining bacteria and to control persistent infections (Haine, Moret, et al. 2008; Schneider and
Chambers 2008). This latter notion is based on the observation that elevated antimicrobial activity following challenge with living or dead bacteria can be observed for up to 21 days in *T. molitor* (Haine, Pollitt, *et al.* 2008). These observations are based on functional zone of clearance assays that measure the overall antimicrobial activity of cell-free haemolymph. Hence a molecular analysis is warranted that elucidates which components of the immune system are up-regulated over the time course of an infection, especially during the start of the expression and peak activity around days 3-5 post infection in *T. molitor* (Haine, Pollitt, *et al.* 2008).

*Tenebrio molitor* is an established model for the biochemistry of insect immunity (Park *et al.* 2011) despite the lack of a reference genome sequence. The biochemical activation of the Toll pathway has been elucidated in *T. molitor* (Roh *et al.* 2009), though the IMD pathway is not yet described (Chae *et al.* 2011). Several antimicrobial peptides have been characterized in detail, most recently Tenecin 4 that bears similarity to *Drosophila* attacins (Chae *et al.* 2011). Recent work in *T. molitor* has also highlighted the role of PGRP-SA in the detection of peptidoglycan as well as D-alanylation-meditated evasion of PGRP recognition by *S. aureus* (Kurokawa *et al.* 2011).

Here we present the first comprehensive RNAseq study of the temporal dynamics of an insect immune response, up to seven days post immune challenge, using the model insect *T. molitor*. Based on the observations of long-lasting inducible immunity against *S. aureus* in *T. molitor* (Haine, Moret, *et al.* 2008; Haine, Pollitt, *et al.* 2008), we quantified gene expression 6hrs, 1, 3, 5 and 7 days post immune challenge to gain a comprehensive insight into
the temporal dynamics of the insect immune system over the period of a
week. We show that genome-independent transcriptome analysis is effective,
not only for annotation of the immune system but also for revealing temporal
patterns of differential gene expression. The transcriptional dynamics of
immune challenge are characterized by a striking separation of transient and
long-lasting responses with the latter dominated by induction of a suite of
antimicrobial peptides.

We present 1) a reference transcriptome assembly derived from insects
challenged with both Gram-positive and Gram-negative bacteria, and utilizing
data from multiple sequencing platforms; 2) annotation of genes encoding
components of the *T. molitor* immune system and 3) quantitative RNAseq
analyses of the response to challenge with *S. aureus* which reveal transient
as well as long-lasting induction and repression of gene expression.

MATERIALS AND METHODS

Insect culturing

Final-instar *Tenebrio molitor* larvae and Progrub formulated diet were
purchased from a commercial supplier (Livefoods Direct, Sheffield, UK).
Larvae were reared *en masse* under a 12:12 h photoperiod at 25 °C with *ad
libitum* access to diet supplemented with apple. Pupae were collected daily
and females were maintained individually in grid boxes. Newly-eclosed
beetles were provided with diet and fresh apple was replaced daily. All
experimental treatments were performed seven days after adult eclosion.
Bacterial preparations

*Staphylococcus aureus* SH1000 and *Escherichia coli* K12 were grown overnight at 37°C in Mueller-Hinton broth and Luria broth, respectively. Two bacterial preparations were produced for immune challenge experiments: the first with a 1:1 combination of *S. aureus* and *E. coli* and the second with *S. aureus* alone. Cultures were washed twice with sterile PBS, heat-killed at 95°C for 30 min and stored in 1-ml aliquots at -80°C until further use.

Immune challenge experiments

Seven-day-old beetles received 5-µl intrahaemocoelic injections of heat-killed bacteria (approximately $10^6$ cells after Haine, Pollitt, *et al.* 2008 and Haine, Moret *et al.* 2008) between the second and third abdominal sternites that were first swabbed with 96% ethanol. Control beetles received injections of sterile PBS. Beetles were maintained on diet supplemented with 2-mm cubes of fresh apple.

We first performed an immune-challenge experiment with a combination of *S. aureus* and *E. coli* in order to obtain a comprehensive reference of immune genes expressed in response to both gram-positive and gram-negative bacteria using 454 GS FLX titanium sequencing. Five individuals were collected for RNA isolation at 6h after challenge and then every 24 h for 7 days. For quantitative RNAseq analysis using Illumina HiSeq2000, a second challenge experiment was performed with *S. aureus* and 10 individuals were collected at 6h and 1, 3, 5 and 7 days after challenge. Additionally, five control
individuals were collected at each time point. This experiment was performed twice on consecutive weeks.

**RNA isolation**

Insects were decapitated with a sterile razor blade and the intestines and reproductive tract removed with sterile forceps. From each individual, hemolymph and fat body were combined, suspended in cold Trizol (Sigma) and homogenized with a 5-mm steel bead (Qiagen) using a TissueLyser (Qiagen) twice at 20 Hz for 10 s. RNA was recovered from the individual homogenates by chloroform extraction and isopropanol precipitation according to the manufacturers instructions and re-dissolved in RNA storage solution (Ambion). Samples were subsequently incubated with 2 units of TurboDNase (Ambion) for 30 min at 37°C and RNA was isolated using an Rneasy MinElute cleanup kit (Qiagen).

**454 Sequencing**

Full-length cDNA synthesis, GS FLX titanium library construction and sequencing on the GS FLX titanium platform were performed by GATC Biotech (Konstanz, Germany). Briefly, polyadenylated RNA was isolated from a pool constructed using 2 µg of total RNA from each individual in the combined *S. aureus* and *E. coli* immune challenge experiment. Full length cDNA was constructed according to a SMART protocol. First strand cDNA synthesis was primed with oligo(dT) followed by RNA hydrolysis and adaptor-primed second strand synthesis. Following hydroxyapatite normalization,
cDNA was coligated, nebulized and sequenced three times on a GS FLX instrument using titanium chemistry on 1/16, 1/4 and 1 full picotitre plate, respectively. Resulting sequence data are available from the NCBI sequence read archive (SRA) under BioSample accession SAMN02389790.

**Illumina sequencing**

The construction of 12 barcoded, non-normalised TruSeq cDNA libraries and sequencing on the HiSeq2000 platform were performed by GATC Biotech. Briefly, polyadenylated RNA was isolated from total RNA pools representing each replicated timepoint as described above (using *S. aureus*-challenged beetles only) except that normalization was not performed. Pools of RNA representing ten individuals from each replicated timepoint as well as two pools representing control individuals were created by combining equal quantities of total RNA. cDNA from each treatment was barcoded with TruSeq universal adapters, pooled and sequenced on a HiSeq2000 using 2 lanes of a single flow cell. Resulting sequence data are available from the NCBI SRA under BioSample accessions SAMN02389798-SAMN02389809.

**Hybrid transcriptome assembly and annotation**

Raw 454 and illumina reads were trimmed using cutadapt to remove sequencing barcodes and cDNA synthesis adaptors. Trimmed 454 reads were filtered by length to remove reads less than 50 bp. Illumina reads were combined into a single fastq file and normalized to a maximum of 20-fold coverage using k-mers of length 20 by khmer version 0.2 (Brown *et al*. 2012).
Paired-end reads were simulated from 454 reads and normalized using `simulate_illuminaPE_from_454ds.pl` and `normalize_by_kmer_coverage.pl`, respectively, from trinity assembler version r2013-02-25 (Grabherr et al. 2011; Haas et al. 2013). Both sets of digitally normalized reads were then combined and assembled using trinity. Trinity assembly generates components which each comprise a group of contig sequences that are inferred to represent alternative splice forms or closely related paralogs (Grabherr et al. 2011). To eliminate possible artifacts, sequences representing less than 1% of the per-component expression across all mapped RNAseq reads (see below) were discarded. Annotation was performed following the trinotate annotation suite guidelines. Homology searches and predictions were performed locally and used to populate an sqlite database with the trinotate perl wrapper from trinity assembler version r2013-02-25 at an e-value threshold of 1e-5. Briefly, peptide sequences were predicted from the assembly by trinity transdecoder and used to query SwissProt with BLAST. Protein domains, signal peptides and transmembrane domains were determined using HMMER (Finn et al. 2011), signalP (Petersen et al. 2011) and tmHMM (Krogh et al. 2001), respectively. Putative orthologs were predicted from reciprocal best BLAST hits with the *Tribolium castaneum* predicted proteome official gene set (http://beetlebase.org/) as described elsewhere (Johnston and Rolff 2013). The Insecta level of OrthoDB version 6 was downloaded and used to define both gene ontologies of the *T. castaneum* official gene set as well as ortholog relationships with other published insect genomes (Waterhouse et al. 2011). Antimicrobial peptide genes were identified by both reciprocal best BLAST hits
with *T. castaneum* AMPs and annotation by BLAST and HMMER described above.

**RNASEq analysis**

Trimmed illumina reads from each replicate were mapped to the reference assembly using RSEM (Li and Dewey 2011) and Bowtie (Langmead *et al.* 2009). Choice of methodology for analysis of differential gene expression was informed by a recent comparison of 10 RNaseq analysis methods (all of which are implemented in R) which utilized both real and simulated data and identified DESeq as the most conservative method with the lowest rates of type I error and false discovery and no method-specific signal (Soneson and Delorenzi 2013). Differential gene expression was determined using the R bioconductor package DESeq (Anders and Huber 2010) using the default sharing mode to estimate dispersions with false discovery rates at $p < 0.05$ following Benjamini and Hochberg procedures (Anders and Huber 2010). Transcripts with a minimum of 4-fold change in expression at $p < 0.05$ were extracted and clustered using the R package DIRECT (Fu *et al.* 2013) according to median centered log2 fragments per feature kilobase per million reads mapped (FPKM). Briefly, a one-parameter Dirichlet-process prior was used to induce a prior distribution and estimate cluster number. Partitions were sampled using a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) procedure. Resampling and relabelling were used to create an allocation probability matrix describing clusters of genes (Fu *et al.* 2013). Tests for over-representation of molecular function and biological process
Gene Ontology (GO) terms associated with lists of genes from DIRECT clusters were performed using the R package GOstats (Falcon and Gentleman 2007) using a hypergeometric test with a p-value cut-off of 0.01 and a non-redundant list of GOs associated with the reference assembly annotation as the gene universe (Falcon and Gentleman 2007).

**Quantitative PCR**

Total RNA was re-isolated from individual frozen Trizol homogenates as described above. From each biological replicate, three independent pools were created using 100 ng of total RNA from individual insects. cDNA was synthesized using a cDNA-Synthesis Kit H Plus (Peqlab). Relative gene expression was determined using a peqGOLD Hot Start-Mix kit (Peqlab) and a StepOne real-time thermocycler (Applied Biosystems) according to the manufacturers instructions. Relative expression was calculated using the comparative Ct method with the ribosomal protein gene RPL27a as the control (Chae et al. 2011).

cDNA cloning of antimicrobial peptide sequences

To verify the accuracy of the reference assembly, DNA sequences corresponding to the mature peptide of the five previously described *T. molitor* AMP genes were cloned. Ten μl of cDNA derived from each individual insect from the *S. aureus* challenge timecourse (synthesized as described above) was pooled and used as template for cloning. Primers were designed for regions with no sequence variation immediately upstream and downstream of
the predicted mature peptide sequence to allow for amplification of potential sequence variants. Primer sequences can be found in supplementary table 17. PCR was carried out in 25 μl reaction volume containing 12.5 μl Promega GoTaq 2x master mix, 10 μM primers and 2 μl cDNA template. PCR conditions were as follows: 95°C 2 min, 30 x (95°C 30 sec, 52°C 30 sec, 72°C 30 sec), 72°C 7 min, 4°C hold. Amplicons were purified using QIAquick PCR purification kit (Qiagen) and cloned into pGEM-T vector using pGEM-T Vector system kit (Promega). White colonies were screened by PCR with SP6 and T7 primers. PCR product purification and Sanger sequencing of positive clones with SP6 and T7 primers was performed by Macrogen Inc. (Seoul, Korea). Sequences were assembled using DNA baser and trimmed to the mature peptide using CLC sequence viewer 6. In total, 34 (+ 4 truncated at 5 or 3 prime end) clones were obtained for tenecin 1, 41 for tenecin 2, 31 (+12 truncated at 5 or 3 end) for tenecin 3, 36 for tenecin 4 and 41 for attacin. Cloned DNA sequences of mature AMPs can be found in Table S17.

**Results and discussion**

**Reference assembly**

In order to generate a comprehensive reference transcriptome, we performed a hybrid assembly of 454 GS FLX titanium sequencing reads from *Tenebrio molitor* adults challenged with a combination of gram-positive and gram-negative bacteria and illumina HiSeq2000 sequencing reads from insects challenged with the gram-positive bacterium *Staphylococcus aureus* only. Multiple trinity assemblies utilizing differently pre-processed combinations of
454 and illumina data were compared in order to identify the most comprehensive reference transcriptome for subsequent annotation and RNAseq quantitation (Table S1-S2). Assembly of digitally normalized illumina reads together with 76-bp paired-end reads simulated from the 454 data resulted in the greatest number of components, transcripts and putative orthologs of *Tribolium castaneum* genes as well as universal single-copy orthologs (Table S1, File S1). Filtering this assembly to retain transcripts representing at least 1% of the per-component expression produced a reference transcriptome of 44,516 components containing 90,956 sequences with an N50 of 1644 bp and N90 of 393 bp (File S2). The transcriptome assembly is primarily derived from hemocytes and fat-body tissue of adult females, yet only 20 of 112 existing *T. molitor* Sanger cDNA sequences failed to retrieve a near-identical reciprocal best blastn hit with the reference transcriptome assembly (Table S3). It is likely that these 20 sequences were not detected because they are expressed in a sex-, developmental stage- and/or tissue-specific manner (Table S4). For example, 7 of these sequences encode variants of a *T. molitor* antifreeze protein that is expressed in the midguts of over-wintering larvae (Graham et al. 2000) (Table S4), a tissue and developmental stage that was not represented in our sampling. Reciprocal blast analysis of 77,118 predicted peptides (File S3) versus the *T. castaneum* official gene set (OGS) proteome (16,645 proteins) identified 9,370 putative orthologs (Table S5). By assigning *T. molitor* sequences to ortholog groups based on *T. castaneum* reciprocal blast hits, 3,120 of 3,377 universal single-copy orthologs that are conserved across the arthropods were identified.
Despite the small deficit of universal single-copy orthologs and the observation that 11,758 genes in the *T. castaneum* OGS are predicted to possess orthologs in at least one other published insect genome outside of the order coleoptera (Waterhouse *et al.* 2011) it is clear that the reference transcriptome assembly is comprehensive.

**Annotation of the *T. molitor* immune system**

Immune genes (Table S6) were defined as transcripts encoding i) putative orthologs of the manually curated *T. castaneum* immune system (Zou *et al.* 2007), ii) previously described components of the *T. molitor* immune system (Fig.1) iii) putative AMPs and iv) any predicted protein annotated with the biological process gene ontology term “immune response” (GO:0006955).

For the 390 proteins previously defined as components of the *T. castaneum* immune system (Zou *et al.* 2007) 213 putative orthologs were identified (Table S6), including the conserved signalling pathways Toll, IMD and JAK/STAT (Figure 1). Of the 177 remaining *T. castaneum* immune genes (Table S7) for which no *T. molitor* orthologs were found, the majority (124) were either serine proteases (SP), their non-catalytic homologs (SPH) or serine protease inhibitors (serpins). Together with SPHs, SPs regulate several aspects of the insect immune response including proteolytic activation of the pro-phenoloxidase zymogen responsible for melanization as well as activation of the Toll pathway which induces AMP synthesis (Figure 1) (Kounatidis and Ligoxygakis 2012). Lineage-specific expansions of SPs/SPHs are evident in many insect genomes including *Anopheles* (Christophides *et al.* 2002),
Drosophila (Ross et al. 2003) and Tribolium (Zou et al. 2007) where repeated rounds of SP/SPH expansion (Zou et al. 2007) may account for the relative paucity of 1:1 orthology between T. castanuem and T. molitor. A similar pattern is observed in serpins where a recent major amplification within a 50-kb region of T. castaneum chromosome 8 has produced a cluster of 16 closely-related serpins (Zou et al. 2007) for which we find no T. molitor orthologs (Table S7). A further 40 putative immune system genes were identified by BLAST, HMMER and/or immune response Gene Ontology terms (Table S6). In this way we identified members belonging to several functional classes which are under-represented in T. castaneum such as chitotriosidases and the croquemort scavenger receptor family, as well as the i-type lysozymes which are not annotated in T. castaneum. Differences in gene content between Tribolium and Tenebrio were also apparent within AMP gene families, which is expected given that AMPs are subject to rapid diversification with frequent duplication and turnover (Yang et al. 2011). Expansion and divergence is apparent in the T. molitor coleoptericin and attacin AMP families with both possessing more members than in most other coleopterans with the possible exception of the ladybird beetle Harmonia axyridis (Vilcinskas et al. 2013). In addition to the previously described tenecin 4 (Chae et al. 2011) and attacin C (Dobson et al. 2012), 2 attacins were identified that belong to divergent phylogenetic groups (Figure S1) along with their respective T. castaneum orthologs. In contrast, the coleoptericins formed species-specific groups (Figure S2) with two novel T. molitor coleoptercins clustering with the previously identified tenecin 2. A single novel cecropin was identified, the
fourth member of this family to be reported from the coleoptera, supporting
the notion that cecropins may be widespread in this order (Zou et al. 2007).
As in T. castaneum the cecropin possesses an atypical tyrosine-rich c-
terminal extension (data not shown). Strikingly, we could reliably annotate only
a single defensin, the previously identified tenecin 1 (Moon et al. 1994) which
belongs to a coleoptera-specific clade of defensins (Zou et al. 2007). This is in
contrast to T. castaneum which possesses 4 defensins, one of which belongs
to a clade of primitive defensins that is found in diverse arthropods (Zou et al.
2007). Two putative defensins were discarded due to low coverage (Table S8)
but may represent transcripts of real genes with little or no expression in our
target tissues. Defensin duplications are apparent in many insect species
including wasps (Gao and Zhu 2010), termites (Bulmer and Crozier 2004)
ants (Zhang and Zhu 2012), and mosquitoes (Dassanayake et al. 2007), and
we cannot exclude the possibility that transcripts from recently duplicated loci
may have been collapsed during assembly leading to an underestimation of
copy number throughout the assembly.

Gene expression
The temporal response to immune challenge with heat-killed S. aureus was
quantified by mapping approximately 8 million 100-bp Illumina reads per
replicated time-point to the reference assembly followed by pairwise
comparison of timepoints using DESeq (Anders and Huber 2010). DESeq
identified 1,050 components (of 44,516) as differentially expressed (DE)
across the time-course (Table S9) with a minimum of a 4-fold change in
expression at $p < 0.05$ after FDR correction. Bayesian clustering with a Dirichlet-process prior was used to cluster DE genes by their temporal expression and to estimate the number of clusters within the data following the procedure of Housden et al. (Housden et al. 2013) and Fu et al. (Fu et al. 2013). The process produced 27 clusters of DE genes reflecting temporal expression, magnitude of change and variance (Figure S3, Table S10, Table S18). Across the timecourse, differentially expressed immune effector genes belonged to six clusters displaying three general patterns: transient induction immediately after immune-challenge; long-lasting induction; or long-lasting repression (Figure 2a), which are discussed below.

**Transient response to immune challenge**

The transient response immediately following immune challenge comprised induction of microbe recognition, signal transduction and immune effector gene expression. This included genes encoding a previously described beta-1,3-glucan recognition protein (GNBP3) (Zhang et al. 2003) as well as the NF-kappaB transcription factor Relish, the ultimate target of the IMD signalling pathway, which is itself responsible for transcriptional activation of numerous immune genes (Kounatidis and Ligoxygakis 2012). A gene encoding phenoloxidase, the key enzyme in the melanization response, which produces cytotoxic melanin as well as oxidative intermediates with broad-spectrum antibacterial activity (Zhao et al. 2007), was also up-regulated. Increased expression of genes encoding heme peroxidase, which participates in an alternative melanogenesis pathway (Nappi and Christensen 2005), and
apolipoporphins, which facilitate beta-1,3-glucan pattern recognition and
phagocytosis (Whitten et al. 2004; Hanada et al. 2011) provide further
evidence of a transient cellular response. Induction of FAD-glucose
dehydrogenase, which enhances the encapsulation response by generating
superoxide anions (Cox-Foster and Stehr 1994), and dual oxidase (duox)
gene expression suggest a role for reactive oxygen-mediated killing in the
early phase of the immune response. Recent work demonstrates a role for
duox in activating the transcriptional response to wounding in Drosophila in
addition to its role as an immune effector (Juarez et al. 2011). We also
detected further transcriptional evidence for a transient wound healing
response including increased expression of genes encoding spectrin (scab)
and integrin (karst), which are required for purse-string wound closure in
Drosophila (Campos et al. 2010), as well as cohesin (fascin) and
myospheroid, which mediate the wound-migratory response of plasmatocytes
(Comber K, Huelsmann S, Evans I, Sánchez-Sánchez BJ, Chalmers A,
(epidermal Growth Factor Receptor) and shark, which are components of the
wound closure signalling pathway (Geiger et al. 2011), as well as a homolog
of djub which positively regulates epithelial proliferation via the hippo pathway
in Drosophila (Das Thakur et al. 2010), were also up-regulated. Supporting
our inference of up-regulation of genes involved in immunity and wound
closure, the biological process Gene Ontology (GO) terms “mucosal immune
response” (GO:0002385, $p = 0.003$), “immune response in organ or tissue”
(GO:0002251, $p = 0.003$), “immune system process” (GO:0002376, $p =
“dorsal closure” (GO:0007394, $p = 0.0028$) and “morphogenesis of an epithelial sheet” (GO:0002011, $p = 0.0025$) were over-represented in the transiently-induced gene clusters (Figure 2b, Table S11-12).

**Long-lasting response to immune challenge**

Numerous immune-related biological process GO terms were over-represented in long-lasting response clusters including “immune response” (GO:0006955, $p = 6.151663e-04$) and several child terms as well as “Toll signaling pathway” (GO:0008063, $p = 0.001$) and the molecular function GO terms “serine-type endopeptidase inhibitor activity” (GO:0004867, $p = 1.621524e-08$) and “serine-type endopeptidase activity” (GO:0004252, $p = 1.582472e-04$) (Figure 2b, Table S13-14). Lasting induction of genes encoding components of the Toll signaling pathway was apparent, including the previously described Gram-negative binding protein 1 (GNBP1), which initiates the proteolytic activation cascade that converges on the Toll ligand Spaetzle (Kounatidis and Ligoxygakis 2012) (Figure 1), and an ortholog of *Tribolium* Beta-1,3-glucan-binding protein 2. Six serine protease genes were up-regulated including the Spaetzle-processing enzyme (SPE) and SPE-activating enzyme (SAE) (orthologs of *Drosophila* easter and snake respectively), which are responsible for the final steps of Toll activation, as well as seven serine protease inhibitor genes. As in *Drosophila* (De Gregorio *et al.* 2001), expression of *Toll* itself was also induced following immune challenge. Strikingly, the only immune effector genes to show long-lasting induction were those encoding antibacterial peptides and the iron-
sequestering protein ferritin. In contrast to the marked induction of antibacterial peptide gene expression, we detected no change in expression of genes encoding the thaumatins or Tenecin 3, which are active against fungi and yeast (Kim et al. 1998; Altincicek et al. 2008), or the i- and c-type lysozymes. A total of eight AMP genes were up-regulated including four attacins (two of which were previously described (Dobson et al. 2012; Chae et al. 2011)), three coleoptericins (including tenecin 2 (Roh et al. 2009)), and the defensin tenecin 1 (Moon et al. 1994). To verify this result, fold induction of gene expression relative to procedural controls was determined by relative qPCR for a subset of AMP genes across the time-course (Figure 3).

The attacin Tenecin 4 and the coleoptericin Tenecin 2 show little to no antibacterial activity towards S. aureus and other Gram-positive bacteria in vitro (Roh et al. 2009; Chae et al. 2011). Furthermore, given that attacins kill Gram-negative bacteria via a non-lytic mechanism which involves lipopolysaccharide binding and inhibition of outer membrane protein synthesis (Carlsson et al. 1998), it is likely that Attacin C and the two putative attacins identified here also possess negligible activity towards S. aureus. Long-lasting induction of almost all antibacterial peptides, regardless of their spectra of activity, is striking and may arise from a lack of specificity in the immune response. Alternatively antibacterial peptides may interact when expressed in combination or act in a sublethal, bacteriostatic manner.

Metabolic repression
A single immune gene encoding a scavenger receptor class B member was repressed following immune challenge. Dramatic down-regulation of numerous metabolic genes provides clear evidence of general metabolic repression. Of 81 over-represented biological process GO terms, 55 described genes with functions in metabolism or biosynthesis including glucose metabolism, as well as lipid and vitamin biosynthesis (Figure 2b, Table S15-16). Expression of the hemolymph storage protein hexamerin, which forms a nutrient reservoir in many insects, was also repressed. Repression of dispensable metabolic pathways as a cost of intense immune gene expression was proposed following a genome-wide analysis of the immune response in *Drosophila* (De Gregorio *et al.* 2001) Recent work demonstrated that activation of the Toll pathway (but not IMD) in *Drosophila* suppresses insulin signaling within the fat body and thereby reduces nutrient storage (DiAngelo *et al.* 2009). Our results suggest that a trade-off between metabolism and immunity may be a general phenomenon.

**Conclusions**

Our data show that a genome-independent comprehensive RNAseq annotation of most of the conserved insect immune system in *T. molitor* is feasible. In accordance with suggestions of a clear two-stage process of insect immunity, as proposed before (Haine, Moret, *et al.* 2008; Schneider and Chambers 2008), we found distinct temporal profiles with clear groups of immune-responsive genes. Notably, antimicrobial peptide genes were clearly persistently up-regulated, whilst genes involved in constitutive defense
responses show only transient up-regulation, presumably partly to replenish
zymogens such as Prophenoloxidase. This is consistent with the idea that
antimicrobial peptides are up-regulated for a long duration to ‘mop up’ and
control persistent infections (Haine, Moret, et al. 2008). Finally we show the
suppression of metabolic genes, consistent with the proposed physiological
costs of immune defenses (Schmid-Hempel 2011), that have often been
shown at the physiological level (Adamo et al. 2008; Freitak et al. 2003) but
not studied very much at the molecular level.

Acknowledgements
We thank C. Titus Brown for making khmer digital normalisation code freely
available under BSD licence. We also thank Brian Haas, Joshua Orvis and
the rest of the trinity contributors for responsive software development and
support. We are grateful to Jack Werren for help in identifying the putative
ortholog of the dome ligand unpaired, to Audrey Fu for troubleshooting the
bayesian clustering package DIRECT, and to Yannick Pauchet for advice on
methodology. The research was supported by ERC grant 260986 to JR.

Footnotes
Sequence data have been submitted to the NCBI SRA under BioSample
accessions SAMN02389790 and SAMN02389798-SAMN02389809.

References


Juarez, M. T., R. A. Patterson, E. Sandoval-Guillen, and W. McGinnis, 2011 Duox, Flotillin-2, and Src42A are required to activate or delimit the spread

Kafatos, F., R. Waterhouse, E. Zdobnov, and G. Christophides, 2009


Kounatidis, I., and P. Ligoxygakis, 2012 Drosophila as a model system to unravel the layers of innate immunity to infection. Open biology 2: 120075.


Michel, K., and F. C. Kafatos, 2005 Mosquito immunity against Plasmodium. Insect biochemistry and molecular biology 35: 677–89.


Schwarz, R. S., and J. D. Evans, 2013 Single and mixed-species trypanosome and microsporidia infections elicit distinct, ephemeral


Yang, W., T. Cheng, M. Ye, X. Deng, H. Yi et al., 2011 Functional divergence among silkworm antimicrobial peptide paralogs by the activities of recombinant proteins and the induced expression profiles. PloS one 6: e18109.


Zou, Z., J. D. Evans, Z. Lu, P. Zhao, M. Williams et al., 2007 Comparative genomic analysis of the Tribolium immune system. Genome biology 8: R177.

Figure legends

Figure 1. Comparison of immune genes and pathways annotated in Tenebrio molitor and Tribolium castaneum. Gene products are organized by pathway and cellular location after Obbard et al. (Obbard et al. 2009).
Green indicates genes which are annotated in both organisms. Yellow and blue indicate genes which have only been annotated in *T. castaneum* or *T. molitor* respectively. Blue lines highlight genes which have been previously described in *T. molitor*.

**Figure 2. Contrasting profiles of differential gene expression following immune challenge.** (A) Six clusters of differentially expressed genes showing three temporal profiles: transient induction; long-lasting induction or long-lasting repression. Vertical axes represent median centered log2 FPKM whereas horizontal axes represent days post immune challenge. Colored lines depict the median profile for each cluster. (B) Significantly Over-represented Gene Ontology in each temporal profile. *, \( p < 0.05 \); **, \( p < 0.01 \); ***, \( p < 0.001 \).

**Figure 3. Quantification of antimicrobial peptide gene expression by relative qPCR.** Error bars show standard deviation across 3 biological replicates of pools of 8-10 individual insects.

**Supporting Information**

**Figure S1.** Maximum likelihood phylogenetic tree showing relationships among beetle attacins. Protein sequences were aligned using MUSCLE and Gblocks. Trees were constructed using PhyML and TreeDyn using phylogeny.fr webserver. Red arrows indicate a confidence index > 0.8.
Figure S2. Maximum likelihood phylogenetic tree showing relationships among beetle coleoptericins. Protein sequences were aligned using MUSCLE and Gblocks. Trees were constructed usin PhyML and TreeDyn using phylogeny.fr webserver. Red arrows indicate a confidence index > 0.8.

Figure S3. Gene expression clusters produced by the R package DIRECT. Vertical axes represent median centered log2 FPKM whereas horizontal axes represent days post immune challenge. Colored lines depict the median profile for each cluster.

Table S1. Comparison of trinity assemblies utilising illumina and/or 454 reads.

<table>
<thead>
<tr>
<th>Assembly</th>
<th>sim PE, simulated paired-end; USCO, universal single-copy ortholog.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newbler</td>
<td>*assembly newbler included for reference. Newbler does not compute components and the number of isogroups is reported instead. Orthologs, the number of putative orthologs between the assembly and the <em>Tribolium castaneum</em> proteome. In each assembly Illumina data was first digitally normalized.</td>
</tr>
</tbody>
</table>

Table S2. Summary statistics for each of the data sets used in de novo assembly. sim PE, simulated paired-end; diginorm, digitally normalized. *all assemblies utilized digitally normalized illumina data. The properties of the original data are included here for reference. **refers to number of read pairs.

*a*fastq file, *b*fasta file.
Table S3. Reciprocal best blastn hits between assembled contigs and 89 previously published Tenebrio molitor gene sequences from GenBank.

Table S4. GenBank accession numbers and metadata for 20 previously published Tenebrio molitor gene sequences which do not retrieve a reciprocal best blastn hit from assembled contigs.

Table S5. Reciprocal best blastp hits between predicted proteins derived from assembled contigs and the Tribolium castaneum predicted proteome official gene set (http://beetlebase.org/).

Table S6. Details of annotated Tenebrio molitor contigs defined as immune genes.

Table S7. Details of Tribolium castaneum immune genes for which no putative Tenebrio molitor ortholog could be identified.

Table S8. Details of two putative defensins which were discarded due to poor read support.

Table S9. Genes which were differentially expressed at one or more timepoints in the timecourse.
Table S10. Details of cluster allocation for each differentially expressed gene.

Table S11. Results of a hypergeometric test for over-representation of biological process gene ontology terms associated with the transiently-induced temporal profile.

Table S12. Results of a hypergeometric test for over-representation of molecular function gene ontology terms associated with the transiently-induced temporal profile.

Table S13. Results of a hypergeometric test for over-representation of biological process gene ontology terms associated with the long-lasting induction temporal profile.

Table S14. Results of a hypergeometric test for over-representation of molecular function gene ontology terms associated with the long-lasting induction temporal profile.

Table S15. Results of a hypergeometric test for over-representation of biological process gene ontology terms associated with the long-lasting repression temporal profile.
Table S16. Results of a hypergeometric test for over-representation of molecular function gene ontology terms associated with the long-lasting repression temporal profile.

Table S17. Primer sequences used to clone AMP sequences from cDNA.

Table S18. Combined details of cluster allocation for each differentially-expressed gene together with annotations.

File S1. Full fasta format output of trinity reference assembly.

File S2. Reference assembly after filtering to remove sequences representing less than 1% of the per-component expression across all mapped RNAseq reads.

File S3. Full annotation report for 77,118 predicted peptides derived from the reference assembly.