Comparative genomic analysis of the *Tribolium* immune system

Zhen Zou*, Jay D Evans†, Zhiqiang Lu*, Picheng Zhao*, Michael Williams‡, Niranji Sumathipala*, Charles Hetru§, Dan Hultmark‡ and Haobo Jiang*

Addresses: *Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA. †USDA-ARS Bee Research Laboratory, Beltsville, MD 20705, USA. §Umeå Centre for Molecular Pathogenesis, Umeå University, Umeå S-901 87, Sweden. ‡Institut Biol Moléc Cell, CNRS, Strasbourg 67084, France.

* These authors contributed equally to this work.

Correspondence: Haobo Jiang. Email: haobo.jiang@okstate.edu

© 2007 Zou et al.; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract**

**Background:** *Tribolium castaneum* is a species of Coleoptera, the largest and most diverse order of all eukaryotes. Components of the innate immune system are hardly known in this insect, which is in a key phylogenetic position to inform us about genetic innovations accompanying the evolution of holometabolous insects. We have annotated immunity-related genes and compared them with homologous molecules from other species.

**Results:** Around 300 candidate defense proteins are identified based on sequence similarity to homologs known to participate in immune responses. In most cases, paralog counts are lower than those of *Drosophila melanogaster* or *Anopheles gambiae* but are substantially higher than those of *Apis mellifera*. The genome contains probable orthologs for nearly all members of the Toll, IMD, and JAK/STAT pathways. While total numbers of the clip-domain serine proteinases are approximately equal in the fly (29), mosquito (32) and beetle (30), lineage-specific expansion of the family is discovered in all three species. Sixteen of the thirty-one serpin genes form a large cluster in a 50 kb region that resulted from extensive gene duplications. Among the nine Toll-like proteins, four are orthologous to *Drosophila* Toll. The presence of scavenger receptors and other related proteins indicates a role of cellular responses in the entire system. The structures of some antimicrobial peptides drastically differ from those in other orders of insects.

**Conclusion:** A framework of information on *Tribolium* immunity is established, which may serve as a stepping stone for future genetic analyses of defense responses in a nondrosophilene genetic model insect.

**Background**

*Tribolium* beetles harbor a range of natural pathogens and parasites, from bacteria to fungi, microsporidians and tape-worms [1,2]. There is good evidence for genetic variation in resistance to the tapeworm and a linked cost of resistance in terms of growth and reproduction [3]. Cross-generational transfer of immune traits [4] may occur in *Tenebrio molitor*, a close relative of *Tribolium castaneum*. RNA interference
experiments demonstrate that Tribolium laccase-2 is responsible for cuticle pigmentation and sclerotization [5]. While these observations are interesting, our knowledge of the genetic constituents of Tribolium immunity is almost blank at the cellular and molecular levels, in contrast to the vast amount of information regarding Drosophila melanogaster and Anopheles gambiae defense responses [6,7]. Given the high efficiency of RNA interference and powerful tools of molecular genetics [8], it is particularly appealing to use T. castaneum for the dissection of insect immune pathways. Acquired knowledge may be useful in controlling beetle pests that feed on crop plants or stored products.

In the broader field of beetle immunity, research has been focused mainly on two effector mechanisms, namely antimicrobial peptide synthesis and prophenoloxidase (proPO) activation [9]. Defensins, coleopterins, cecropin and antifungal peptides have been isolated from coleopteran insects and characterized biochemically [10-12]. A homolog of human NF-κB (Allomyrina dichotoma Rel A) up-regulates the transcription of a coleopterin gene [13]. Active phenoloxidase generates quinones for melanin formation, wound healing, and microbe killing. ProPO activation has been investigated in Holotrichia diomphalia [14-16]. ProPO activating factor 1 (Hd-PPAF1) cleaves proPO to generate active phenoloxidase in the presence of Hd-PPAF2, the precursor of which is activated by Hd-PPAF3 via limited proteolysis. While all these PPAFs contain an amino-terminal clip domain, PPAF2 (in contrast to PPAF1 or PPAF3) does not have catalytic activity since its carboxy-terminal serine proteinase-like domain lacks the active site serine. A 43 kDa inhibitor down-regulates the melanization response in H. diomphalia [17].

To date, components of the innate immune system are hardly known in T. castaneum and neither is it clear how they differ from homologous molecules in the honeybee, mosquito or fruitfly [6,7,18]. This lack of knowledge does not seem to reconcile with the critical phylogenetic position of this coleopteran species, which should inform us a lot about genetic variations in the evolution of holometabolous insects. Information regarding defense responses in T. castaneum, a member of the largest and most diverse order of eukaryotes, is highly desirable for the biological control of crop pests and disease vectors. Consequently, we have used its newly available genome assembly to annotate immunity-related genes and analyze their phylogenetic relationships with homologous sequences from other insects. In this comparative overview of the Tribolium defense system, we describe plausible immune pathway models and present information regarding the molecular evolution of innate immunity in holometabolous species.

Results and discussion

Overview of the Tribolium immune system

T. castaneum has a sizable repertoire of immune proteins predicted to participate in various humoral and cellular responses against wounding or infection (Additional data file 1). Like other insects [6,7,19], cuticle and epithelia lining its body surfaces, trachea and alimentary tract may serve as a physiochemical barrier and local molecular defense by producing antimicrobial peptides and reactive oxygen/nitrogen species (ROS/RNS). While this line of defense may block most pathogens, others enter the hemocoel where a coordinated acute-phase reaction could occur to immobilize and kill the opportunists. This reaction, including phagocytosis, encapsulation, coagulation and melanization, is probably mediated by hemocytes and molecules constitutively present in the circulation. These first responders may not only control minor infections but also call fat body and hematopoietic tissues for secondary responses if necessary. At the molecular level, the following events should take place in all insects, including the beetle: recognition of invading organisms by plasma proteins or cell surface receptors, extra- and intracellular signal transduction and modulation, transcriptional regulation of immunity-related genes, as well as controlled release of defense molecules.

Pathogen recognition

Peptidoglycan recognition proteins (PGRPs) serve as an important surveillance mechanism for microbial infection by binding to Lys- and diaminopimelate-type peptidoglycans of walled bacteria [20]. Some Drosophila PGRPs (for example, LC and SA) are responsible for cell-mediated or plasma-based pathogen recognition; others (that is, LB and SB) may hydrolyze peptidoglycans to turn on/off immune responses [21,22]. In T. castaneum, PGRP-LA, -LC and -LD contain a transmembrane segment; PGRP-SA and -SB are probably secreted; PGRP-LE (without a signal peptide or transmembrane region) may exist in cytoplasm or enter the plasma via a nonclassical secretory pathway. Bootstrap analysis and domain organization clearly indicate that Tribolium and Drosophila PGRP-LEs are orthologs - so far no PGRP-LE has been identified in Anopheles, Bombbyx or Apis. Other orthologous relationships (for example, TcPGRP-LC and AmPGRP-LC) are also supported by the phylogenetic analysis (Figure 1). The beetle and mosquito PGRP-LA genes encode two alternative splice forms (PGRP-LAa and -Lab). Like Drosophila and Anopheles, Tribolium PGRP-LA and -LC genes are next to each other in the same cluster. Most of the beetle PGRPs resulted from ancient family diversification that occurred before the emergence of holometabolous insects. In contrast, gene duplication occurred several times in the lineages of mosquito and fly (Figure 1).

Multiple sequence alignment suggests that β-1,3-glucan-recognition proteins (β GRPs) and Gram-negative binding proteins (GNBPs) are descendents of invertebrate β-1,3-glucanases [23]. Lacking one or more of the catalytic residues,
these homologous molecules do not possess any hydrolytic activity. They are widespread in arthropods and act in part to recognize microbial cell wall components such as β-1,3-glucan, lipoteichoic acid or lipopolysaccharide. We have identified three β GRPs in T. castaneum. Tc-β GRP1 and AgGNBP-B1 through -B5 are closely related and represent a young lineage, whereas Tc-β GRP2 and Tc-β GRP3 belong to an ancient group that arose before the radiation of holometabolous insects (Additional data file 2). Since Drosophila has no β GRP-B and Anopheles has five, the presence of a single gene (encoding Tc-β GRP1) in the beetle can be useful for elucidating function of this orthologous group. In addition to the glucanase-like domain, members of the second group contain an amino-terminal extension of about 100 residues. In Bombyx

Figure 1
Peptidoglycan recognition proteins. The amino acid sequences from eight Tribolium (Tc), thirteen Drosophila (Dm), nine Anopheles (Ag), and four Apis (Am) PGRPs are examined. The phylogenetic tree shows family expansion in Tribolium (shaded yellow), Anopheles (shaded pink) and Drosophila (shaded blue). TcPGRP-LA, -LC and -LD contain a transmembrane domain whereas TcPGRP-SA and -SB have a signal peptide for secretion. Pink arrowheads at nodes denote bootstrap values greater than 800 from 1,000 trials. The putative 1:1 or 1:1:1 orthologs are connected by green lines. TcPGRP-LB and -SB contain the key residues for an amidase activity.
and three EGF; CTL12 has six Ig and three FN3; CTL14 has all fibrinogen-related proteins (FREPs) contain a carboxy-terminal fibrinogen-like domain associated with different amino-terminal regions. In mammals, three classes of FREPs have been identified: ficolin, tenascins, and microfibril-associated proteins [32]. They take part in phagocytosis, wound repair, and cellular adhesion [33]. In invertebrates, FREPs are involved in cell-cell interaction, bacterial recognition, and antimicrobial responses [34-36]. The Tribolium genome contains seven FREP genes, which fall into three groups (Additional data file 5): the expansion of group I yielded four family members: Tc-FREP1 through 4. Sitting next to each other on chromosome 3, these beetle genes encode polypeptides most similar to angiopoietin-like proteins. During angiogenesis, the human plasma proteins interact with tyrosine kinase receptors (for example, Tie) and lead to wound repair and tissue regeneration [37]. In group II, Tc-FREP5 is orthologous to Dm-scabrous, which is required for Notch signaling during tissue differentiation [38]. Interestingly, Notch is also needed for proper differentiation of Drosophila hemocytes [39]. Group III includes Tc-FREP6, Tc-FREP7, Ag-FREPs and Dm-CG9593. No major expansion has occurred in the beetle or honeybee, in sharp contrast to the situations in the fly and mosquitoes - there are 61 FREP genes in the A. gambiae genome [29].

The extracellular signal transduction and modulation
Similar to the alternative and lectin pathways for activation of human complements, insect plasma factors play critical roles in pathogen detection, signal relaying/tuning, and execution mechanisms. Serine proteinases (SPs) and their noncatalytic homologs (SPHs) are actively involved in these processes. Some SPs are robust enzymes that hydrolyze dietary proteins; others are delicate and specific - they cleave a single peptide bond in the protein substrates. The latter interact among themselves and with pathogen recognition proteins to mediate local responses against nonself. The specificity of such molecular interactions could be enhanced by SPHs, adaptor proteins that lack proteolytic activity due to substitution of the catalytic triad residues. SPs and SPHs constitute one of the largest protein families in insects [29,41,42]. We have identified 103 SP genes and 65 SPH genes in the T. castaneum genome [29,41,42]. The alternative and lectin pathways for activation of human complements, insect plasma factors play critical roles in pathogen detection, signal relaying/tuning, and execution mechanisms. Serine proteinases (SPs) and their noncatalytic homologs (SPHs) are actively involved in these processes. Some SPs are robust enzymes that hydrolyze dietary proteins; others are delicate and specific - they cleave a single peptide bond in the protein substrates. The latter interact among themselves and with pathogen recognition proteins to mediate local responses against nonself. The specificity of such molecular interactions could be enhanced by SPHs, adaptor proteins that lack proteolytic activity due to substitution of the catalytic triad residues. SPs and SPHs constitute one of the largest protein families in insects [29,41,42]. We have identified 103 SP genes and 65 SPH genes in the T. castaneum genome [29,41,42].

All fibrinogen-related proteins (FREPs) contain a carboxy-terminal fibrinogen-like domain associated with different amino-terminal regions. In mammals, three classes of FREPs have been identified: ficolin, tenascins, and microfibril-associated proteins [32]. They take part in phagocytosis, wound repair, and cellular adhesion [33]. In invertebrates, FREPs are involved in cell-cell interaction, bacterial recognition, and antimicrobial responses [34-36]. The Tribolium genome contains seven FREP genes, which fall into three groups (Additional data file 5): the expansion of group I yielded four family members: Tc-FREP1 through 4. Sitting next to each other on chromosome 3, these beetle genes encode polypeptides most similar to angiopoietin-like proteins. During angiogenesis, the human plasma proteins interact with tyrosine kinase receptors (for example, Tie) and lead to wound repair and tissue regeneration [37]. In group II, Tc-FREP5 is orthologous to Dm-scabrous, which is required for Notch signaling during
four groups of SP-related genes may represent lineages derived from ancient evolutionary events since similar subfamilies also exist in Anopheles and Drosophila. Moreover, expansion of individual subfamilies must have occurred several times to account for the gene clusters observed in the Tribolium genome (Figure 2). Evidence for lineage-specific gene duplication and movement is also present in the mosquito and fly genomes [29,41]. Based on the results of genetic/biochemical analysis performed in other insects [14-16,19,44,45] and sequence similarity, we are able to predict the physiological functions for some Tribolium clip-domain SPs and SPHs during proPO activation and spätzle processing. For instance, Tc-SPH2, SPH3 or SPH4 (similar to Hdl-PPAF2) may serve as a cofactor for Tc-SP7, SP8 or SP10 (putative proPO activating proteins); Tc-SP44 or SP66 may function like Drosophila persephone [46]; Tc-SP136 or SP138 may activate spätzle precursors by limited proteolyis [44,45].

Most members of the serpin superfamily are irreversible inhibitors of SPs and, by forming covalent complexes with diffusing proteases, they ensure a transient, focused defense response [47]. There are totally 31 serpin genes in T. castaneum, more than that in D. melanogaster (28), A. gambiae (14) or A. mellifera (7). This number increase is mainly caused by a recent family explosion at a specific genomic location - we have identified a cluster of 16 serpin genes in a small region of 50 kilobases on chromosome 8. These closely related genes constitute a single clade in the phylogenetic tree (Figure 3). Sequence divergence, especially in the reactive site loop region, is anticipated to alleviate the selection pressure imposed by the SP family expansion (Figure 2). Exon duplication and alternative splicing, found in 4 of the 31 serpin genes, also generate sequence diversity and inhibitory selectivity.

Intracellular signal pathways and their regulation

Drosophila Toll is a transmembrane protein that binds spätzle and relays developmental and immune signals [48]. Resulting from ancient family expansion, a total of five spätzle homologs and eight Toll-like receptors are present in the fly. There are seven Tribolium genes coding for spätzle-like proteins, most of which have putative orthologs in Drosophila and Anopheles (Additional data file 7). Like their ligands, Toll-like proteins have also experienced major family expansion and sequence divergence. The receptors are separated into two clusters, with the fly and beetle Toll-9 located near the tree center (Figure 4). While Toll-6, -7, -8 and -10 from different insect species constitute tight orthologous groups in one cluster, lineage-specific gene duplications have given rise to Drosophila Toll-3 and -4, Anopheles Toll-1 and -5, as well as Tribolium Toll-1 through -4. Located on the same branch with Drosophila Toll, the four Tribolium receptors could play different yet complementary roles in the beetle defense and development. In addition, we have identified eight MD2-related genes in the beetle. Mammalian MD2, Toll-like receptor-4 and CD14 form a complex that recognizes lipopolysaccharides [49]. The Anopheles MD2-like receptor regulates the specificity of resistance against Plasmodium berghei [50].

Contrary to the ligand-receptor diversification, components of the intracellular pathway appear to be highly conserved in insects studied so far (Figure 5a). In Drosophila, multimerization of Toll receptors caused by spätzle binding leads to the association of dMyD88, Tube, Pelle, Pellino and dTRAF6 [51]. With 1:1 orthologs identified in the beetle (as well as the other insects with known genomes), we postulate that a similar protein complex also forms to phosphorylate a cactus-like molecule (Tc0203). The modified substrate protein then dissociates from its partner (Tc07697 or Tc0806), allowing the Rel transcription factors to translocate into the nucleus and activate effector genes (for example, antimicrobial peptides). Functional tests are required to verify the suggested roles of individual components during defense and development in the beetle.

The IMD pathway is critical for fighting certain Gram-negative bacteria in Drosophila. Upon recognition of diaminopimelate-peptidoglycan by PGRPs, the ‘danger’ signal is transduced into the cell through IMD (Figure 5b). IMD contains a death domain that recruits dFADD (dTAK1 activator) and Dredd (a caspase). Active dTAK1 is a protein kinase that triggers the JNK pathway (through Hep, Basket, Jra and Kay) and Relish phosphorylation (through Ird5 and Kenny). The presence of 1:1 orthologs in T. castaneum strongly suggests that IMD-mediated immunity is conserved in the beetle. Furthermore, the modulation of these pathways may also resemble each other - we have identified putative 1:1 orthologs of IAP2, Tab2 and caspar in the Tribolium genome (Figure 5b).

The transcription of Drosophila TEPs and some other immune molecules is under the control of the JAK-STAT pathway [52]. This pathway, triggered by a cytokine-like molecule, Upd3, promotes phagocytosis and participates in an antiviral response. Based on sequence similarity, we predict that the conserved signaling pathway in the beetle is composed of the orthologs of Dm-Domeless, Hopscotch and STAT92 (Figure 5c). However, we have not identified any ortholog of Dm-upd, upd2, or upd3, possibly due to high sequence variation in the cytokine-like proteins.

Execution mechanisms

Phenoloxidases are copper-containing enzymes involved in multiple steps of several immune responses against pathogens and parasites (that is, clot reinforcement, melanin formation, ROS/RNS generation, and microbe killing) [53]. Synthesized and released as an inactive zymogen, proPO requires a SP cascade for its cleavage activation. SPHs and serpins ensure that the proteolytic activation occurs locally and transiently in response to infection. We have identified three proPO genes in the Tribolium genome, designated proPO1, 2 and 3. Tc-proPO2 and proPO3 are 98.8% identical in nucleotide sequence and 99.6% identical in amino acid
sequence. In the aligned coding regions (2,052 nucleotides long), 21 of the 24 substitutions are synonymous, corresponding to 0.0102 changes/site. These two genes are 530 kb apart and their aligned intron regions are 88.5% identical. Using the relative rate of nucleotide substitutions derived from an analysis of *Drosophila* alcohol dehydrogenase genes [54], we estimate that *Tc*-proPO2 and *Tc*-proPO3 arose by gene duplication approximately 0.6 million years ago. The phylogenetic analysis suggests that such evolutionary events are sporadic for this family: the total numbers of proPO genes in different insect species did not change significantly, except for the malaria mosquito (Additional data file 8). Of the nine *Ag-*
proPO genes, eight arose from gene expansion that occurred early in the mosquito lineage [29], some of which encode phenoloxidases for melanization.

Local production of free radicals is a critical component of the acute-phase oxidative defense, involving nitric oxide synthase, NADPH oxidase, peroxidase, phenoloxidase and other enzymes [53,55]. Due to the cytotoxicity of ROS and RNS, their conversion and concentrations must be tightly regulated by superoxide dismutases (SODs), glutathione oxidases (GTXs), catalases, thioredoxins, thioredoxin reductases, melanin intermediates, and certain metal ions. Changes in the free radical levels by gene mutation or knock-down affect the fecundity and antimalarial response of the mosquito [56]. We have annotated some of these genes in *Tribolium*, including peroxidases, GTXs, SODs, peroxiredoxins (TPXs) and catalases. *T. castaneum* GTX1-GTX2 and TPX2-TPX6 gene pairs are results of recent gene duplications, whereas several orthologous relationships have been identified in the SOD and TPX families in the phylogenetic analysis (Additional data file 9).

Coleopteran species have been explored at the biochemical level for various antimicrobial peptides (AMPs) [57]. While defensins are present in all insects studied, coleoptericins are related to the attacin/diptericin family of glycine-rich anti-
bacterial peptides in lepidopteran and dipteran species [58].
Four defensin genes are detected in the Tribolium genome,
three of which are found in a branch containing only coleop-
teran insects (Figure 6). Tc-defensin4 is in a miscellaneous
group containing Odonata, Lepidoptera and Arachnida spe-
cies. Interestingly, defensins of three other coleopteran
insects are in the same branch with the hymenopteran ones.

Like the beetle defensins, coleopterins belong to two phylo-
genetic groups, with the same separation of species in each
group.

With the genome sequence available, we are able to use the
other AMP sequences to identify homologous genes that are
not specified in beetles. Cecropins were mostly identified in
moths and flies - there was only one report on cecropin from a coleopteran species, *Acalolepta luxuriosa* [11]. In *Tribolium*, we find a single close homolog of the *Acalolepta cecropin*, although a frame shift in a run of seven adenosines indicate that this is a pseudogene (Tc00499). Closely linked to Tc00499 on chromosome 2 are two genes that encode cecropin-related peptides of unusual structure, with proline- and tyrosine-rich carboxy-terminal extensions (Tc-cecropin2 and Tc00500). These observations indicate that cecropins may widely exist in beetles. Attacins were found only in lepidopteran and dipteran species. We have identified a cluster of three attacin genes (Tc07737-07739) on *Tribolium* chromosome 4. Although we failed to identify a Drosomycin homolog in the beetle, our search resulted in a low-score hit of a
A calolepta A1

**Figure 6**
Evolutionary relationships of the coleoptericins (left panel) and defensins (right panel). The alignment of mature antimicrobial peptide sequences is used to build the phylogenetic trees on which their genus names are indicated. The beetle coleoptericins and defensins are divided into two subgroups (shaded blue and pink), whereas the more primitive defensins (shaded grey) are found in many arthropod species. Note that the blue clades include *Acalolepta*, *Tribolium* and Zophobas whereas the pink clades both contain *Allomyrina* and *Holotrichia*. Pink arrowheads at nodes denote bootstrap values greater than 800 from 1,000 trials. This analysis uses sequences from the orders of Coleoptera (*Acalolepta*, *Allomyrina*, *Holotrichia*, *Oryctes*, *Protaetia*, *Rhinoceros*, *Tenebrio*, *Zophobas*, *Diptera* (*Aedes*, *Anopheles*, *Drosophila*, *Phormia*, *Sarcophaga*, *Stomoxys*), *Lepidoptera* (*Galleria*, *Heliothis*), *Hemiptera* (*Phyrrhocoris*), *Mycetophilidae* (*Acalolepta* (*Apis*, *Bombus*, *Formic*), *Neuroptera* (*Chrysopea*), *Ephemeroptera* (*Acalolepta* (*Tribolium*)), *Odonata* (*Acalolepta* (*Gallina*, *Heliothis*)) and *Scorpionidae* (*Androctonus*, *Leirus*).

cysteine-rich sequence. The corresponding gene (*Tc11324*) encodes a 104 residue polypeptide containing 2 whey acidic protein motifs. While mammalian proteins with this motif possess antibacterial activities [59], expression and biochemical analyses are needed to test if the *Tribolium* protein has a similar function. Due to the presence of species-specific AMPs and severe sequence diversity of these molecules, our homology-based search has probably missed some AMP genes. Should there be a thorough exploration by sequence similarity, biochemical separation and activity assays (not only against Gram-positive and Gram-negative bacteria, but also against yeasts and filamentous fungi), we expect the total number of AMPs (currently 12) in *T. castaneum* may approach that (20) in *D. melanogaster*. In addition to these, we have found a cluster of four lysozyme genes in the *Tribolium* genome (Additional data file 10). Similar but independent family growths have occurred in different insect groups, giving rise to thirteen such genes in *Drosophila*, eight in *Anopheles*, three in *Apis*, and four in *Tribolium*.

Cellular responses (that is, phagocytosis, nodulation and encapsulation) play key roles in the insect innate immunity...
In the past few years, breakthroughs have been made in the molecular dissection of these processes [61]. *Drosophila* Peste, Eater, scavenger receptor (SR)-CI, Dscam, TEPs, and PGRP-SCA1 seem to be implicated in the phagocytosis. Multiple SR-B genes are present in the *Tribolium* (16), *Drosophila* (12) and *Anopheles* (16) genomes, indicative of important functions of the subfamily. A phylogenetic analysis of the SR-Bs (Figure 7) demonstrates that nearly half of the members arose from ancient gene duplication events - we can easily identify orthologs from different insect species. More recent family expansions in the mosquito [29] and beetle account for the other half of the subfamily. There are two SR-B gene clusters in the *Tribolium* genome, one of which (TcSR-B14, -B15 and -B16) is located in the same branch containing Dm-peste. In addition to SR-Bs, *Drosophila* Nimrods are also involved in cellular responses [62]. The plasmatocyte-specific NimC1 directly participates in the phagocytosis of bacteria. For *Tribolium*, all three subclasses are represented: NimA, NimB and NimC, just like in the fly, mosquito and bee. However, unlike the other insects, the syntenic relationship is broken up in the beetle NimC homologs: the two NimC paralogs (Tc02053 and Tc15258) are not closely linked to the NimA and NimB homologs (Tc11427 and Tc11428). In the other insects, the order of nimA, nimB and nimC genes is well conserved.

**Expression analysis**

One characteristic of the innate immune system is that some of its components are transcriptionally up-regulated after a microbial challenge. To acquire evidence that the genes we annotated are involved in defense responses, we have exposed the adult beetles to *E. coli*, *Mycrocoecus luteus*, *Candida albicans* or Saccharomyces cerevisiae cells and isolated total RNA from the control and treated insects for expression analysis. Real-time PCR experiments indicated that transcript levels of some genes dramatically changed (Figure 8). *TcPGRP-SA and TcPGRP-SB* mRNA became more abundant after the bacterial infection, whereas the increase was much less significant for *TcPGRP-LA, -LE, galecin1 or TEP-C* after the *C. albicans* or *M. luteus* treatment. Following the Gram-positive bacterial or fungal challenge, we detected some elevations in *Tc-cSP66*, serpin29 and serpin30 transcripts.

Transcriptional regulation is not limited to pattern recognition molecules or extracellular signal mediators/modulators: we detected differential expression of ligand and their receptors (for example, *Tc-spätzle1, Toll-1 through Toll-4, and IMD*). mRNA level changes for the latter genes were small except for IMD (Figure 8). Toll-3 and Toll-4 induction after the *C. albicans or M. luteus* challenge was apparent, although not as notable as IMD. The subtle changes in Toll-1 transcript levels were somewhat different from those of Toll-2, -3 and -4, indicating that there could be functional differences and overlaps in antimicrobial responses for these closely related receptors (Figure 4).

We have also examined genes whose products are plasma proteins directly involved in microbe immobilization or killing. The transcripts of *Tc-proPOs, lysozyme1 or lysozyme4* did not significantly change when compared with the controls, whereas those of *Tc-lysozyme2* and 3 increased remarkably (Figure 8). The most dramatic increase in mRNA levels occurred in the AMP group of effector molecules, including *Tc-attacin2*, cecropin3, coleopterixin1, defensin1, and defensin2.

Cluster analysis of the expression patterns has revealed several trends of the transcriptional control of these immune genes. Buffer injected and uninjured adults form one cluster with the lowest mRNA levels, whereas *E. coli* - and *S. cerevisiae* -treated insects have the next higher level of overall gene expression (Figure 8). The yeast-injected beetles, instead of grouping with *E. coli*-treated insects, are found in the same cluster with *C. albicans*-challenged adults. Interestingly, immune responses toward the opportunistic fungal pathogen are greater than those toward *S. cerevisiae*, an environmental non-pathogen present in the diet. The responses toward *M. luteus* and *C. albicans* were significantly stronger than those towards *E. coli*, implying that the Toll pathway triggered by the Gram-positive bacteria and filamentous fungi more effectively up-regulated target gene expression than the IMD pathway did, which may be activated by the Gram-negative bacterial infection (Figure 5).

**Conclusion**

Through this comparative genome analysis, we have provided evidence in the red flour beetle for the functional conservation of intracellular immune signaling pathways (Toll, IMD and JAK/SAT) and for the evolutionary diversification of over 20 families of proteins (for example, PGRPs, clip-domain proteins, serpins, Toll-related receptors, antimicrobial proteins and scavenger receptors) involved in different mechanisms of insect defense against infection. The observed differences in conservation are likely related to distinct needs for specific molecular interactions and changes in microorganisms encountered by the host insects. For instance, *Drosophila* Myd88, Tube, Pelle, Pellino and TIRAF, which form a macromolecular complex with the Toll/interleukin 1 receptor domain (Figure 5), have 1:1 orthologs in *Anopheles*, *Apis* and *Tribolium*. In contrast, family expansion and sequence divergence in the PGRP and AMP families are perhaps important for specific recognition and effective elimination of evolving pathogens.

The summary of putative immune gene counts, families and functions (Additional data file 11) suggests that *T. castaneum* has a more general defense than *A. gambiae* does. While this system is critical for the survival of this beetle, we are unclear whether or not it correlates with the prosperity of coleopteran insects. Drastic lineage-specific expansions seem sporadic and, in most cases, *Tribolium* paralog counts are lower than...
Figure 7
Phylogenetic analysis of class B SRs (SR-Bs). The aligned central parts, including the CD36 domain, of sixteen Tribolium (Tc), eight Drosophila (Dm), eight Anopheles (Ag) and three Aps (Am) SR-B sequences are used for building the unrooted tree (upper panel). For simplicity, the other members of class B SRs from Drosophila (seven) and Anopheles (four) are not included in this analysis. Lineage-specific expansion (shaded yellow) is confirmed in the complete tree that includes all SR-Bs from the four species. The expansion is consistent with their chromosomal locations (lower panel). Pink arrowheads indicate nodes with bootstrap values exceeding 800 (from 1,000 trials), whereas green bars connect the putative orthologs with 1:1, 1:1:1 or 1:1:1:1 relationship.
those of Anopheles or Drosophila (but are considerably higher than of Apis). The only exceptions are the clip-domain SP/SPH and serpin families: 48, 41 and 37 protease-related genes and 31, 14 and 28 inhibitor genes are present in the beetle, mosquito and flies, respectively. Because clip-domain SPs are often regulated by serpins, positive selection may have played a role in the converted evolution of both families and in the maintenance of homeostasis.

This comparative analysis has also uncovered interesting genes and gene families for future research. For instance, the existence of a 1:1 ortholog of Drosophila PGRP-LE in Tribolium (but not in Anopheles or Apis) may allow us to test whether or not TcPGRP-LE has a similar function. It can be interesting to explore the molecular mechanisms and evolutionary pathways of the large serpin and SP gene clusters in the beetle. The presence of TcToll-1 through -4 and subtle changes in their mRNA levels after immune challenges call for detailed analysis of their transcriptional regulation and physiological functions. Of course, the proposed extracellular and intracellular signaling pathways need to be tested, even though we have confidence in their general structures. The possible AMP function of Tc1124, which contains two whey acidic protein motifs, needs to be established experimentally.

It is noteworthy that the functions of Tribolium immunity-related genes are mostly assumed based on sequence similarity to studied proteins in Drosophila or other insect species. Functional analyses using the strong reverse genetic techniques available in Tribolium are necessary to test the hypotheses. Nevertheless, the framework of information established in this work should help clarify immune functions in an important agricultural pest from the most diverse insect order and a species that can serve as a tractable model for an innate immune system more generally.

Materials and methods

Database search and sequence annotation

Known defense proteins from other insects were used as queries to perform BLASTP searches of Tcastaneum Glean Predictions (2005.10.11) [63]. Protein sequences with E-values lower than 0.1 were listed, and every 5th sequence was retrieved for use as a query for another round of search. Based on the combined lists, respective protein sequences were retrieved, compiled in the order of ascending E-values, and improved by two methods. Firstly, Tcastaneum ESTs (2005.9.20) at the same HGSC site were searched with the corresponding nucleotide sequences to identify possible cDNA clones. The EST sequences were assembled using CAP3 [64] and the resulting contigs were used in pairwise comparison [65] to validate the gene predictions. Secondly, retrieved protein sequences were analyzed by CDART [66], PROSITE [67], and SMART [68] to detect conserved domain structures required for specific functions. Necessary changes were made after each step to improve the original predictions. Chromo-

| Execution | Recognition | Signal transduction | Pathogenicity |
|-----------|-------------|---------------------|---------------|
| F G+ G+ | S.c | E.c | PBS | C.a | M.l |

**Figure 8**

Real-time PCR analysis of expression of Tribolium immunity-related genes in adults 24 h after injections of M. luteus (M.l.), E. coli (E.c.), C. albicans (C.a.), S. cerevisae (S.c.), or phosphate-buffered saline (PBS). Uninjured insects (-) were used as another negative control. With green, black and red colors representing low, intermediate and high transcript levels, respectively, relative mRNA abundances were used to cluster samples by average-linker clustering.

**Table 1**

| Gene        |
|-------------|
| defensin1   |
| defensin2   |
| defensin3   |
| defensin4   |
| proPO1      |
| proPO2/3    |
| lysozyme1   |
| lysozyme2   |
| lysozyme3   |
| lysozyme4   |
| attacin2    |
| cecropin3   |
| coleoptericin1 |
| defensin1   |
| defensin2   |
| defensin3   |
| defensin4   |
somal location and exon-intron boundaries for each annotated sequence were acquired from Genboree [69]. To locate orthologs not identified by BLASTP, Tribolium Genome Assembly 2.0 [70] was searched using TBLASTN. The hits detected were analyzed using multiple gene prediction tools Genescan and Genemark [71,72]. All curated sequences then were deposited in the annotation database [73] as a part of Tribolium Genome Assembly 2.0.

**Phylogenetic analyses**

Unless otherwise specified, full-length Tribolium sequences were aligned with their homologs from other insects, including *D. melanogaster, A. gambiae* and *A. mellifera*. The sequences were retrieved from NCBI [74], Flybase [75], or Ensembl [76]. Multiple sequence alignments were carried out using ClustalX [77] and Blossum series of weight matrices [78]. Phylogenetic trees were constructed based on algorithm of neighbor-joining using PHYLIP [79] or maximum-parsimony using PAUP [80]. The divergence time of *Tc*-proPO2 and proPO3 were calculated using the rate of 1.7 × 10^{-8} synonymous substitutions/nucleotide/year derived from the *Drosophila* species [54].

**Gene expression analysis**

To study pathogen-induced gene expression, adult red flour beetles (approximately 240 per group) were pricked at the ventral thorax with needles dipped in concentrated phosphate-buffered saline or the buffer containing concentrated live *E. coli, M. luteus, C. albicans* or *S. cerevisiae* cells. Uninjured and asexually injured insects were employed as controls. Total RNA samples were extracted from the control and challenged insects (approximately 160 per group) 24 h later, using Micro-to-mid RNA Purification System (Invitrogen, Carlsbad, CA, USA). After DNA removal, each RNA sample (1.0-3.4 μg), oligo(dT) (0.5 μg, 1 μl) and dNTPs (10 mM each, 1 μl) were mixed with diethyl pyrocarbonate-treated H$_2$O in a final volume of 12 μl, and denatured at 65°C for 5 minutes. First strand cDNA was synthesized for 50 minutes at 42°C using SuperScript Reverse Transcriptase (200 U/μl, 1 μl; Invitrogen) mixed with 5 × buffer (4 μl), 0.1 M dithiothreitol (2 μl), RNase OUT (40 U/μl, 1 μl; Invitrogen) and the denatured RNA sample (12 μl). Specific primer pairs were designed for a total of 35 immunity-related genes (Additional data file 12) using Primer 3 [81] with annealing temperatures of 59.5-60.5°C and expected product sizes of 80-150 bp. Each primer pair was located in adjacent exons flanking an intron. Real-time PCR was performed in parallel reactions on 96-well microtiter plates using Taq DNA polymerase (1 U; Roche Applied Sciences, Indianapolis, IN, USA), 1 × buffer, 1 mM dNTP mix, 2 mM MgCl$_2$, 0.2 μM primers, 1 × SYBR-Green I dye (Applied Biosystems, Foster City, CA, USA) and 10 nM fluorescein. Amplifications were enacted on an iCycler thermal cycler (Bio-Rad, Hercules, CA, USA) with a profile of 95°C for 5 minutes followed by 40 cycles of 94°C for 20 s, 60°C for 30 s, 72°C for 60 s and 78°C for 20 s [82]. SYBR green fluorescence was measured during the 78°C step in each cycle and the cycle numbers for each target and control gene were recorded when the fluorescence passed a predetermined threshold. Proper dissociation and correct size of the products were examined by melting curve analysis and agarose gel electrophoresis, respectively. The real-time PCR was repeated twice and, in each of the three experimental replicates, the transcripts were normalized relative to the levels of Tribolium ribosomal protein S3. Averaged transcript abundance values (Ct control - Ct target) were then compared across genes and samples using average-linking clustering (Cluster 3.0) and visualized using TreeView [83].

**Abbreviations**

β GRP, β-1,3-glucan-recognition protein; AMP, antimicrobial peptide; CTL, C-type lectin; FREP, fibrinogen-related protein; GNBPs, Gram-negative binding protein; GTX, glutathione oxidase; PGRP, peptidoglycan recognition protein; PPAF, proPO activating factor; proPO, prophenoloxidase; RNS, reactive nitrogen species; ROS, reactive oxygen species; SOD, superoxide dismutase; SP, serine proteinase; SPH, non-catalytic serine proteinase homolog; SR, scavenger receptor; TEP, thioester-containing protein; TPX, peroxiredoxin.

**Authors’ contributions**

Zhen Zou: study design; data collection, analysis and deposition; annotation of clip-domain SPs/SPHs, serpins, spätzles, SRs and others; Toll and Imd pathways. Jay Evans: RT-PCR analysis; GNBPs and PGRPs. Zhiqiang Lu: C-type lectins, galectins, TEPs and JAK/STAT pathway. Picheng Zhao: Toll-like receptors, caspases and ROS/RNS production. Michael Williams and Dan Hultmark: FREPs, Nimrods, PGRPs and cecropins. Charles Hetru and Niranji Sumathipala: antimicrobial peptides and lysozymes. Haobo Jiang: study design; data analysis and interpretation; annotation of clip-domain SPs/SPHs; manuscript writing.

**Additional data files**

The following additional data are available with the online version of this paper. Additional data file 1 is a table listing immunity-related genes in *T. castaneum*. Additional data file 2 is a figure showing sequence alignments of βGRPs and GNBPs. Additional data file 3 is a figure showing sequence alignments of CTLs. Additional data file 4 is a figure showing sequence alignments of galectins. Additional data file 5 is a figure showing sequence alignments of FREPs. Additional data file 6 is a figure showing sequence alignments of TEPs. Additional data file 7 is a figure showing sequence alignments of Spätzle-related proteins. Additional data file 8 is a figure showing sequence alignments of proPOs. Additional data file 9 is a figure showing sequences of GTX, SOD and TPX. Additional data file 10 is a figure showing sequence alignments of lysozymes. Additional data file 11 is a table listing functions, families, and counts of putative defense proteins from *D. mel*
References

1. Wade MJ, Chang NW: Increased male-fertility in Tribolium con-fusum beetles after infection with the intracellular parasite Wolbachia. Nature 1995, 373:72-74.

2. Blaser M, Schmid-Hempel P: Determinants of virulence for the parasite Nosema whitei in its host Tribolium castaneum. J Invertbr Pathol 2005, 89:251-257.

3. Zhong D, Pai A, Yan G: Costly resistance to parasitism: evidence from simultaneous quantitative trait loci mapping for resistance and fitness in Tribolium castaneum. Genetics 2005, 169:2127-2135.

4. Moret Y: "Trans-generational immune priming" specific enhancement of the antimicrobial immune response in the mealworm beetle, Tenebrio molitor. Proc Biol Sci. 2006, 273:1399-1404.

5. Arakane Y, Muthukrishnan S, Beeman RW, Kanost MR, Kramer KJ: Genetical Res 2005, 89:139-143.

6. Lemaître B, Hoffmann J: The host defense of Drosophila melanogaster. Annu Rev Immunol 2005, 23:697-743.

7. Christophides GK, Vlachou D, Kafatos FC: Comparative and functional genomics of the innate immune system in the malaria vector Anopheles gambiae. Immunol Rev 2004, 198:127-148.

8. Brown SJ, Denell RE, Beeman RW: Beetle around the genome. Cell 2005, 120:155-161.

9. Iwanaga S, Lee BL: Recent advances in the innate immunity of invertebrate animals. J Biochem Mol Biol 2005, 38:128-150.

10. Barbaut F, Landon C, Guenneugues M, Meyer JP, Schott V, Dimarco JL, Vovelle F: Solution structure of Ato-3: a new knotting-type anti-parallel collagenous peptide from the insect Acrocercus longimanus. Biochemistry 2003, 42:14434-14442.

11. Satoh A, Ueda K, Imanura M, Assumi S, Tabunoki H, Miura N, Watanabe A, Kitami M, Satoh R: Purification and cDNA cloning of a cecropin from the longicorn beetle, Acalolepta luxurians. Comparative Biochem Physiol B Biochem 2005, 142:317-323.

12. Bulet P, Cociancich S, Dimarco JL, Lambert J, Reichardt JM, Hoffmann D, Hetruc C, Hoffmann JA: Insect immunity. Isolation from a coleopteran insect of a novel inducible antibacterial peptide and novel members of the insect defensin family. J Biol Chem 1991, 266:24520-24525.

13. Sagisaka A, Tanaka H, Furukawa S, Yamakawa M: Characterization of a homologue of the Rel/NF-kappaB transcription factor from a beetle, Allomyrina dichotoma. Biochem Biophys Acta 2004, 1675:85-93.

14. Kim MS, Baek MJ, Lee MH, Park JW, Lee SY, Soderhall K, Lee BL: A new easter-type serine proteinase cleaves a masquerade-like protein during prophenoloxidase activation in Holotrichia diomphalia larvae. J Biol Chem 2002, 277:39999-40004.

15. Kwon TH, Kim MS, Choi HW, Joo CH, Cho MY, Lee BL: A masquerade-like serine proteinase homologue is necessary for phenoloxidase activity in the coleopteran insect, Holotrichia diomphalia larvae. Eur J Biochem 2000, 267:6188-6196.

16. Lee SY, Kwon TH, Hyun JH, Choi JS, Kawabata SI, Iwanaga S, Lee BL: In vitro activation of pro-phenoloxidase by two kinds of pro-phenol-oxidase-activating factors isolated from hemolymph of coleopteran, Holotrichia diomphalia larvae. Eur J Biochem 1998, 254:50-57.

17. Zhao M, Soderhall I, Park JW, Ma YG, Osaki T, Ha NC, Wu CF, Soderhall K, Lee BL: A novel 43-kDa protein as a negative regulatory component of phenoloxidase-induced melanin synthesis. J Biol Chem 2005, 280:24744-24751.

18. Evans JD, Aronstein K, Chen YP, Hetruc C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D: Immune pathways and defence mechanisms in honey bees Apis mellifera. Insect Mol Biol 2006, 15:645-656.

19. Kanost MR, Jiang HB, Yu XQ: Insect immune responses of a lep- industrious insect, Manduca sexta. Immunity 2004, 19:97-105.

20. Steiner H: Peptidoglycan recognition proteins: on and off switches for innate immunity. Immunity 2004, 19:83-96.

21. Kim MS, Byun MJ, Oh BH: Crystal structure of peptidoglycan recognition protein-1B from Drosophila melanogaster. Nat Immunol 2003, 4:787-793.

22. Melloroth P, Karlsson J, Steiner H: A scavenger function for a Dro sophila peptidoglycan recognition protein. J Biol Chem 2003, 278:7059-7064.

23. Jiang HB, Ma CC, Lu QZ, Kanost MR: 1-3,1-3-Glucan recognition protein-2 (beta GRP-2) from Manduca sexta: an acute-phase protein that binds beta-1,3-glucan and lipoteichoic acid to aggregate fungi and bacteria and stimulate prophenoloxidase activation. Insect Biochem Mol Biol 2004, 34:89-100.

24. Ochiai M, Ashida M, Moret Y, Arakane Y, Brown SJ, Denell RE, Beeman RW, Kanost MR, Blose C, Blaser M, Schmid-Hempel P, Sassa T, Blass C, Brey PT, Collins FH, Dass S, Dimarco JL, Domazet M, Geier A, Gotoh K, Grunow S, Hoffman D, Hoffmann J, Hoffmann JA: Insect immunity. Isolation from a coleopteran insect of a novel inducible antibacterial peptide and novel members of the insect defensin family. J Biol Chem 1999, 274:1399-1404.

25. Koizumi N, Imanura M, Kodanari T, Yako K, Iwahana H, Sato R: The lipopolysaccharide-binding protein participating in hematocyte nodule formation in the silkworm Bombyx mori is a novel member of the C-type lectin superfamily with two different tandem carbohydrate-recognition domains. FEBS Lett 1999, 443:139-143.

26. Yu XQ, Kanost MR: Immulectin-2, a lipopolysaccharide-specific lectin from an insect, Manduca sexta, is induced in response to gram-negative bacteria. J Biol Chem 2000, 275:37373-37381.

27. Christophides GK, Zdobnov E, Bariellas-Mury C, Blandin S, Yu XQ, Kanost MR: Comparative and functional genomics of the innate immune system in the malaria vector Anopheles gambiae. Immunol Rev 2004, 198:127-148.

28. Brown SJ, Denell RE, Beeman RW: Beetle around the genome. Cell 2005, 120:155-161.

29. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Bluss C, Brey PT, Collins FH, Danielli A, Dimoupolous G, et al: Immunity-related gene and protein families in Anopheles gambiae. Science 2002, 298:159-163.

30. Tanji T, Ohashi-Kobayashi A, Natori S: Participation of a galacose-specific C-type lectin in Drosophila immunity. Biochem J 2006, 396:127-138.

31. Kamwashii S, Ramalho-Orrigao M, Pham VM, Kumar S, Lawyer PG, Turco SJ, Barillas-Mury C, Sacks DL, Valenzuela JG: A role for insect galectins in parasite survival. Cell 2004, 119:329-341.

32. Fujita T, Endo Y, Nonaka M: Primitive complement system - recognition and activation. Mol Immunol 2004, 41:103-111.

33. Litman GW, Cannon JP, Dishaw LJ: Reconstructing immune phy logeny: New perspectives. Nat Rev Immunol 2005, 5:866-879.

34. Gokudan S, Mutu T, Ueno H, Aoki K, Kawaihara T, Seki N, Mizunoe Y, Wai SN, Iwanaga S, Kawabata S: Horsehoe crab acetyl group-recognizing lectins involved in innate immunity are structurally related to fibrinogen. Proc Natl Acad Sci USA 1999, 96:10086-10091.

35. Schroder HC, Ushijima H, Krasa S, Gamulin V, Thakur NL, Diehl-Seifert B, Muller IM, Muller-WEG: Emergence and disappearance of an immune molecule, an antimicrobial lectin, in basal Metazoa - A tachylectin-related protein in the sponge Subites domuncula. J Biol Chem 2003, 278:32810-32817.

36. Wang X, Rceheula TA, Fuchs JF, Hillyer JF, Chen CC, Christensen BM: A novel lectin with a fibrinogen-like domain and its potential involvement in the innate immune response of Arimigeres subellatus against bacteria. Insect Mol Biol 2004, 13:273-282.

37. Morisita T, Kudota Y, Iwanto T, Tada T, Oike Y, Angelpoietins and angiopoietin-like proteins in angiogenesis. Endothelium J 2005, 12:71-79.

38. Mok LP, Qin TL, Bardot B, LeComte M, Homayouni A, Ahimou F, Wesley C: Delta activity independent of its activity as a ligand.
