The evolution of termite immunity

Inaugural-Dissertation

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by
Shulin He

2018

The doctorate studies were conducted under supervision of Prof. Dr. Dino P. McMahon from September 2014 to September 2018 in the Institute of Biology.

1st Reviewer: Prof. Dr. Dino P. McMahon

2nd Reviewer: Prof. Dr. Jens Rolff

date of defense: 25. 10. 2018

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Dr. Dino P. McMahon. It has been an honor to be his first PhD student. I appreciate his contribution of time, ideas and funding to my project and any related research. It is his patience, motivation and encouragement that helped me to finish this work. His broad knowledge and his excellent guidance to science helped me succeed my PhD.

I would like to express many thanks to the following persons: Thorben, Stefania, Alejandra, James for their professional and personal help and advice. I want to thank Thorben and James for their help to improve this thesis. With their company, the time I spent over the last four years has become invaluable memories.

I would like to sincerely thank the Rolff group at Institute of Biology that have contributed immensely to my personal and professional time during my PhD. I am grateful to Prof. Dr. Jens Rolff for his introduce, guidance, instructive advice, valuable professional comments, all of which improved my work and expanded my knowledge in science. His help and advice have excelled me in many other aspects. I am thankful to Dr. Paul R. Johnston for his great help, patient instructions and numerous discussions on bioinformatic analysis and in other ways. It is his guidance that made it possible for me to learn bioinformatics and analyze sequence data in my project. His advice on many other aspects also improved me and gave me courage to keep on. I would like to thank Renate Radek who provided important higher termite species which made a great improvement for my work. I also appreciate other members from that group: Alex, Baydaa, Bin, Christin, Guozhi, Miko, Regina, Sabine for their professional and personal advice and help during my PhD. I also would like to thank the following persons: Sophie, Luis, Bea from the Armitage group for their useful advice and company.

My appreciation also goes to the colleagues of Bundesanstalt für Materialforschung und -prüfung where I did almost all my work over the last four years. I would like to thank Dr. Rüdiger Plarre for allowing me to use termites and for his help on improving my thesis. I want to thank Yvonne for providing much help on catching the termites for studies. I also would like to thank Annett, Ronald and Thomas Dimke for their help on performing experiments, collecting termites and other personal matters. I want to thank Kerstin for her help on the improvement of this thesis and many other personal matters. I would like to thank all other colleagues whose names are not listed here but gave me personal or professional help. It is a great honor to meet and work with them. The valuable time and experience with them will always be remembered.

I would like to thank Prof. Dr. Zongqing Wang and Prof. Dr. Yanli Che from Southwest University, China, who provided much help on sampling subsocial cockroaches. They were

very kind host, had valuable expertise on subsocial cockroaches, and provided students with space for performing experiment. It is a great pleasure to me to know them and collaborate with them. I also would like to thank Susan Mbedi and other members from Berlin Center for Genomics in Biodiversity Research for their professional help and host for all transcriptome sequencing. In addition, I want to acknowledge Benno and Christoph for their help and expertise in proteomics. I also want to thank Jan Šobotník for providing important samples of higher termites.

I am grateful to many Chinese friends. My time in last four years was made enjoyable in large part also due to the time we spent together. It is a great pleasure to meet and have fun with them. Their company and support were valuable experiences that will be deeply remembered.

I would like to thank my family for their enormous encouragement and support. Most of all, I want to thank my girlfriend, who stands by my side with her patience, encouragement and support. It was them who gave me courage to have this valuable experience and to finish this work.

I want to acknowledge the Chinese Scholarship Council that provided me a four year PhD scholarship.

At last, I also would like to thank a lot of people whose names are not listed here for any kind of help they provided during my PhD. Words are not enough to express my gratitude. Thank you all.

Table of contents

List of abbreviations

AMP	Antimicrobial peptide
ATG	Autophagy protein
ATP	Adenosine triphosphate
BP	Biological process
BUSCO	Benchmarking universal single-copy orthologs
CAT	Catalase
CI	Confidence interval
CLIP	Clip-Domain Serine protease
CTL	C-type lectin
Dif	Dorsal-related immunity factor
DRS	Drosomycin
emPAI	The exponentially modified protein abundance index
ESS	Effective sample size
FAD	Flavin adenine dinucleotide
FADD	Fas-associated protein with death domain
FAS	Fatty acid synthase
FREP	Fibrinogen Related protein
GNBPs	Gram-negative binding proteins
GO	Gene ontology
GPX	Glutathione peroxidase
HPX	Heme peroxidase
IAP	Inhibitor of apoptosis
IKK	Iκb kinase
IMD	Immune deficiency
IκB	Inhibitor of nuclear factor κb
JAK	Janus kinase
JNK-ip	C-Jun amino-terminal kinase- interacting protein
LC-MS/MS	Liquid chromatography–mass spectrometry/ mass spectrometry
LIPR	Pancreatic lipase-related protein
LPSBP	Hemolymph lipopolysaccharide-binding protein
LYS	Lysozyme
MF	Molecular function
ML	MD2-like receptor
MRCA	The most recent common ancestor
MOT	Monocarboxylate transporter
mya	million years ago
MyD88	Myeloid differentiation primary response 88
ORF	Open reading frames
PCR	Polymerase chain reaction
PGRP	Peptidoglycan receptor protein
PGRP-SC2	PGRP-short chain
PPO	Polyphenol oxidase
PRR	Pattern recognition receptor
qPCR	quantitative PCR

SCR	Scavenger receptor
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SPZ	Spaetzle
SRPN	Serpin
Stam	Signal-transducing adapter molecules
STAT	Signal transducers and activators of transcription
TAB	TAK binding protein
TAK	Transforming growth factor-activated kinase
TEP	Thioester-containing proteins
TLR	Toll like receptor
TPX	Thioredoxin peroxidase

General introduction

1.1 Biology of termites

Termites, a type of social insect, are one of the most successful insects in the world. They live in groups of hundreds to millions of individuals, which leads to vast ecosystem-dominating life forms (Oster and Wilson 1978). With the considerable ecological importance, termites can compose up to 95% of insect biomass in tropical underground ecosystems (Watt *et al*. 1997) and 21% of the total invertebrate biomass in rainforest epiphytes (Ellwood and Foster 2004). They function as decomposers of dead organic matters in tropical and subtropical regions (Bignell and Eggleton 2000) due to their ability to digest lignocellulose with their symbionts that include bacteria and/or protists (Ohkuma 2003; Brune 2014).

Termites are sometimes referred to as “white ants” because its extreme phenotypical resemblance to ants, although they are not close relatives. Termites are diploid, hemimetabolous social insects that evolved from cockroaches (Inward *et al*. 2007a; Korb 2007, 2008), while ants are haplodiploid, holometabolous insects that evolved from wasps and are close relatives of bees (Thorne and Traniello 2003; Howard and Thorne 2010). The termites develop in incomplete metamorphosis from eggs, via larvae to different castes (Korb and Hartfelder 2008). The individuals in a termite colony are genetically closely related as normally a pair of reproductives are responsible for breeding, except for cases with multiple pairs of reproductives in a colony.

1.1.1 Phylogeny of termites

There are in total around 3000 living termite species, all of which are eusocial. The existing termites are classified into nine families: Mastotermitidae, Hodotermitidea, Archotermopsidae, Stolotermitidae, Kalotermitidae, Stylotermitidae, Rhinotermitidae, Serritermitidae and Termitidae (Engel *et al*. 2009). There are two suprafamilial termite lineages, the Euisoptera and the Neoisoptera (Engel *et al*. 2009; Cameron *et al*. 2012). The former is composed by termite species except Mastotermitidae and the latter is composed by Stylotermitidae, Rhinotermitidae, Serritermitidae and Termitidae (Engel *et al*. 2009).

Depending on the presence of protists in the hind gut, termite species are traditionally classified into two groups: lower termites (protists and bacteria) and higher termites (only bacteria) (Krishna and Weesner 1969; Krishna and Weesner 1970). The lower termites include termite species except the family of Termitidae that is composed all the higher termites. Around 70% of all termite species are composed by higher termites.

Termites are a sister group of subsocial wood-feeding cockroaches (Figure 1.1), the Cryptocercudiae, and nested in the cockroach order Blattodea based on phylogenetic analysis of gene markers and mitochondrial genomes (Lo *et al*. 2000; Inward *et al*. 2007a; Inward *et al*.

2007b; Legendre *et al*. 2008; Engel *et al*. 2009; Cameron *et al*. 2012; Bourguignon *et al*. 2015; Djernæs *et al*. 2015; Legendre *et al*. 2015). During the last two decades, the termite phylogeny has been vigorously investigated. These studies have adopted morphological data or multiple genes from nuclear or mitochondrial to resolve the termite phylogeny, especially the lower termite families Hodotermitidea, Archotermopsidae, Stolotermitidae. However, with the development of next generation sequencing technology, there is currently no comprehensive phylogenetic analysis of the termites using phylogenomic data.

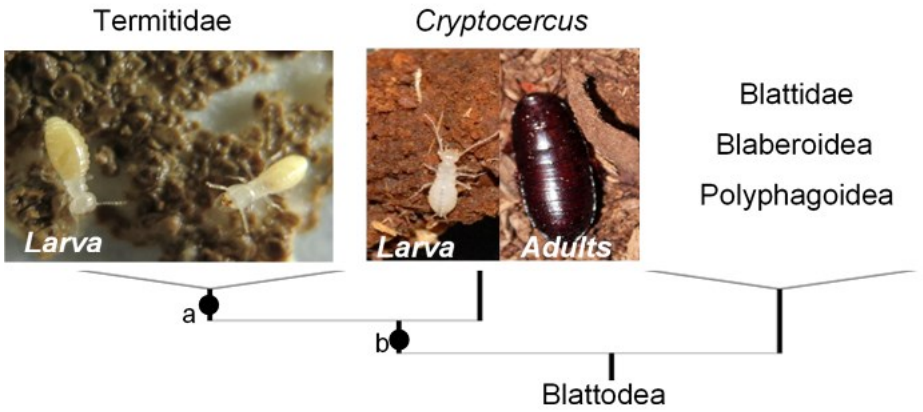


Figure 1.1 The simplified phylogeny of termite and *Cryptocercus* (Inward *et al*. 2007a). The picture of termites is from *Neotermes castaneus* and the pictures of *Cryptocercus* are from *Cryptocercus pudacoensis*. a, b represent the two important evolutionary events mentioned in text.

It has been reported that the termites have diverged from cryptocercid roaches in the Late Jurassic based on fossil records, which predates the origins of ants and bees by around 35 million years (Engel *et al*. 2009). This indicates that termites are probably the oldest eusocial animals (Engel *et al*. 2009). In addition, the most abundant termite family, the Termitidae, diversified during the Miocene (Engel *et al*. 2009).

1.1.2 Termites as social insects

Alongside sexual reproduction and multicellularity, eusociality is considered one of the major transitions in evolution (Szathmáry and Smith 1995), which mostly occurs in insects, the Hymenoptera (ants, bees and wasps) and termites. In both groups, the evolution of a reproductively altruistic caste was critical, as it facilitated the evolution of advanced division of labour and the emergence of sophisticated caste structures.

During the evolution of termites, there are two important evolutionary transitions (Figure 1.1). The first is the transition of solitary cockroaches to wood-feeding subsocial cockroaches. The prime social characters evolved and shared by *Cryptocercus* and termites: 1) unique flagellates, 2) biparental care, and 3) proctodeal trophallaxis (Inward *et al*. 2007a; Nalepa 2010). The second is the transition of subsocial cockroaches to social termites. The true social characters have evolved during this transition: 1) true sterile castes-soldier, 2) overlapping

generation, and 3) division of labour (Inward *et al*. 2007a; Nalepa 2010). The evolution of soldiers, a sterile caste in termites, is of particular importance as it represents the point of no-return in social evolution (Boomsma and Gawne 2018). This path of evolved sterile caste is different from Hymenoptera (Tian and Zhou 2014), where the first sterile caste to evolve was the worker. However, the appearance of true workers is a further transition in termites, which has been considered as multiple origin (Inward *et al*. 2007b; Legendre *et al*. 2008).

Termites have different castes within a colony which is a reflection of division of labour in social evolution, including workers, soldiers and reproductives (Figure 1.2). Workers are the most abundant individuals in a colony. In some lower termite species including Archotermopsidae, Stolotermitidae, Kalotermitidae, Prorhinotermitinae, the true workers are missing and the workers are called “false worker” or “pseudogates” as they can further develop into either reproductives or soldiers (Korb and Hartfelder 2008). In Mastotermitidae, Hodotermitidea, Rhinotermitinae and Serritermitidae, the true worker caste presents as in

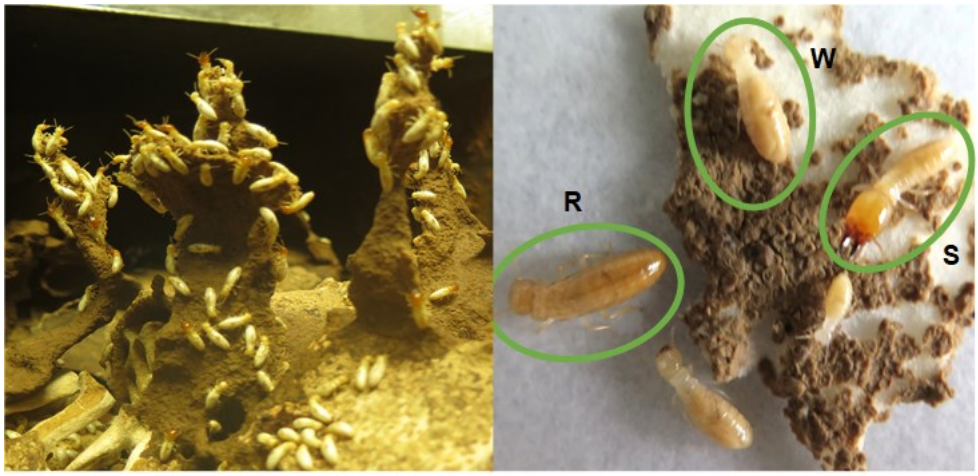


Figure 1.2 The left picture is from a colony of *Mastotermes darwiniensis*. The right picture is from *Neotermes castaneus* with different castes. R: Reproductives (a neotenic reproductives in picture); W: worker (“false worker” in this species); S: soldier.

higher termites (Inward *et al*. 2007b; Legendre *et al*. 2008). The soldier caste makes up 5-20% of a typical insect colony, and is the only true sterile caste that presents across all termite species except for a few species that underwent a secondary loss of the sterile soldier caste (Bourguignon *et al*. 2016a). The reproductives are normally the least abundant individuals in a termite colony and can be categorized into primary reproductives or neotenic reproductives (Korb and Hartfelder 2008). The primary reproductives are alates that shed their wings after the tandem flight and establish a new colony, while the neotenic reproductives are replacements of dead primary reproductives and developed from the origin colony where they live in (Korb and Hartfelder 2008).

The developmental pathways in termites differ between families (Korb and Hartfelder 2008; Roisin and Korb 2010; Korb *et al*. 2015). Depending on the presence of true workers, the development can be categorized as linear (Figure 1.3) or bifurcated (Roisin and Korb 2010) in lower termites and higher termites. In the linear development, the species have totipotent immature stages that can develop into caste options and possible with regressive moulting (Korb and Hartfelder 2008). In bifurcated development, workers and soldiers diverge from the nymphs and cannot subsequently develop into alates (Korb and Hartfelder 2008; Roisin and Korb 2010; Korb *et al*. 2015).

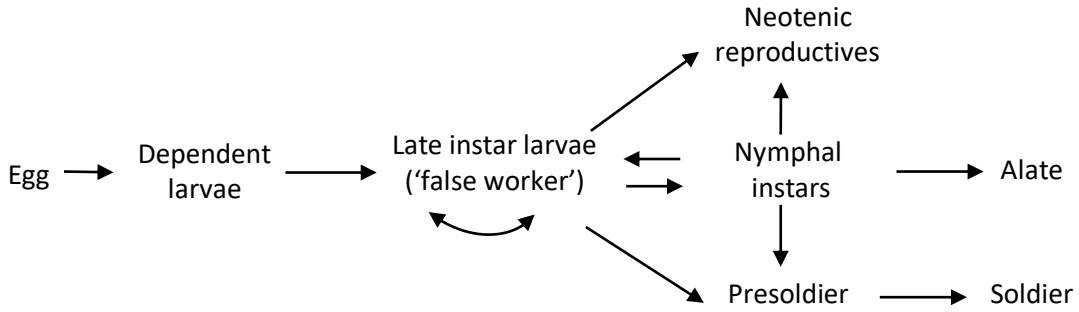


Figure 1.3 A representative linear developmental path of lower termites (except Mastotermitidea, Hodotermitidea, Rhinotermitinae and Serritermitidae) (Judith Korb 2008).

Different castes are responsible for different tasks in the colony. The task specialization in the castes is associated with multiple morphological, physiological and behavioral adaptations (Hölldobler and Wilson 2009; Tian and Zhou 2014; Bourguignon *et al*. 2016b; Engel *et al*. 2016; Kaji *et al*. 2016; Robson and Traniello 2016). Workers (where present) typically carry out the majority of housekeeping tasks such as brood care and foraging. Soldiers (where present) display explicit morphological and behavioral specializations adapted for defence (Šobotník *et al*. 2010; de Roode and Lefèvre 2012; Tian and Zhou 2014; Bourguignon *et al*. 2016b; Kaji *et al*. 2016). The reproductives are responsible for the production of eggs to guarantee the reproduction of the colony.

1.2 **Immunity in social insects**

The elaborate division of labour in social insects lead to their success in the eco-system. However, this does not come without costs. The genetically closed individuals and high population density within the colony are perfect environment for the propagation of parasites and pathogens (Alexander 1974; Schmid-Hempel 1998). But, termites have evolved a sophisticated immune system to counteract these drawbacks (Rosengaus *et al*. 1999b; Traniello *et al*. 2002; Cremer *et al*. 2007; Bulmer *et al*. 2009; Cremer *et al*. 2018). There are

two levels of immune defence in termites as other social insects: individual immunity and social immunity.

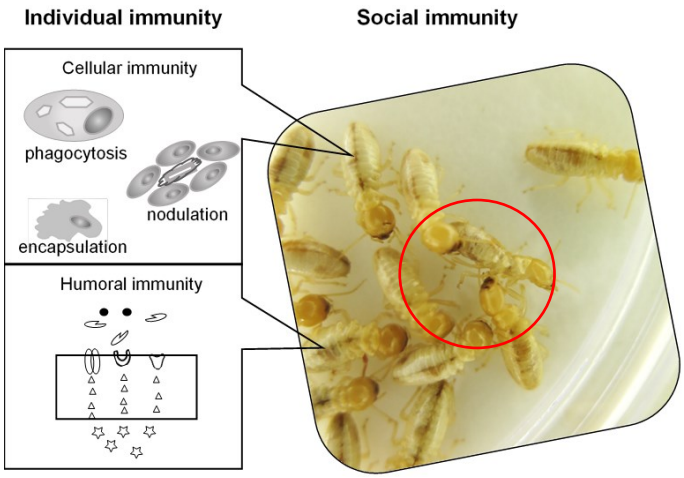


Figure 1.4 An illustration of the two levels immunity in termites, including individual immunity (cellular immunity and humoral immunity) and social immunity (an example of allogrooming in red circle representing a type of social immunity).

1.2.1 Individual immunity

The insect immune system has been widely studied in *Drosophila* and *Tenebrio*, which includes both cellular and humoral immunity. Cellular immunity comprises phagocytosis, encapsulation and nodulation, which are mediated by various types of hemocytes, including granular cells, crystal cells, oenocytoids and plasmatocytes (Lavine and Strand 2002). Humoral immunity is composed of three main immune pathways, Toll, immune deficiency (IMD), and Janus kinase/signal transducers and activators of transcription (JAK-STAT), and a melanisation process.

Insect innate immune molecules occur as three broad types: receptors, signaling components and effectors (Viljakainen 2015; Hillyer 2016). Following infection, pattern recognition receptors bind to microorganisms, which leads to the induction of three principal signaling pathways responsible for the regulation of the insect humoral immune response, known as the Toll, IMD, and JAK-STAT pathways. These canonical pathways are responsible for, amongst other effects, the synthesis of antimicrobial peptides such as defensins and attacins (Hillyer 2016). Many of the functions of genes involved in these pathways derive from a considerable body of research carried out in *Drosophila* and to a lesser extent, other insects. In flies we understand that the Toll pathway responds largely to fungi and gram-positive bacteria, and is mediated by peptidoglycan receptor proteins (PGRPs), gram-negative binding proteins (GNBPs), serine protease cascades, Toll-receptors, Myeloid differentiation primary response 88 (MyD88), Tube, Pelle, and Dorsal-related immunity factor (Dif)/Dorsal transcription factors (Valanne *et al*. 2011). The IMD pathway mainly responds to gram-

negative bacteria, and is comprised of PGRPs, Imd, Fas-associated protein with death domain (FADD), a caspase Dredd, Transforming growth factor-activated kinase 1 (TAK1)-binding protein (TAB), TAK, IκB (inhibitor of nuclear factor κB) kinase (IKK), and Relish (Myllymäki *et al*. 2014). Conversely, the JAK-STAT pathway is thought to regulate inflammation and stress responses. It is principally composed of Cytokines, Domeless, Hopscotch, and Signal transducers and activators of transcription (STAT) (Agaisse and Perrimon 2004). The melanisation process is initiated by the recognition of receptors mostly pattern recognition receptors, mediated by a cascade of serine protease and activated phenoloxidases which are the rate-limiting enzymes in the process of melanogenesis (Nakhleh *et al*. 2017). This process is toxic against a wide range of parasites, bacteria and fungi as well as some virus.

As an individual insect, the members in a termite colony have a full immune system like other insects. From previous genome studies, it has been shown that termites and cockroaches have full repertoire of immune genes (Terrapon *et al*. 2014; Korb *et al*. 2015; Li *et al*. 2018). In addition, a defensin-like class of antimicrobial peptides-the termicins- has been firstly identified in termites (Da Silva *et al*. 2003), which possess antifungal activity. But individual immunity has lack of fully understand in termite castes or in their relatives, subsocial cockroaches.

1.2.2 Social immunity

Apart from the individual immune system in the members of a colony, a collective immunity in the colony level has been found in social insects, termed as “social immunity” (Cremer *et al*. 2007). These mechanisms encompass a range of behaviours that reduce parasites by barring, burying or even cannibalizing infected individuals (Cremer *et al*. 2007) or communicating the presence of pathogens to other nestmates (Rosengaus *et al*. 1998b; Rosengaus *et al*. 1999a). It can also extend to hygienic behaviours such as mutual grooming (de Roode and Lefèvre 2012; Konrad *et al*. 2012), and the collection (de Roode and Lefèvre 2012; Konrad *et al*. 2015) or synthesis of antimicrobial compounds that reduce infectiousness and disease susceptibility (Bulmer *et al*. 2009). It also refers to socially-mediated immunization (Rosengaus *et al*. 1998b; Rosengaus and Traniello 2001; Hughes *et al*. 2002; Traniello *et al*. 2002; Konrad *et al*. 2012), whereby prophylactic transfer of molecular effectors (Hamilton *et al*. 2011) or low dose pathogens (Hughes *et al*. 2002; Hamilton *et al*. 2011; Konrad *et al*. 2012) lead to protection of susceptible nestmates against infection.

Apart from the size effect of groups, the caste formation seems also important to social immunity. It has been shown that social thrips and termite soldiers have dual roles in physical defence and antimicrobial protection (Turnbull *et al*. 2012; Mitaka *et al*. 2017b). In addition, the variety of castes can boost the protection of immunity in groups (Gao *et al*. 2012). This

effect could be mediated by the cuticle hydrocarbons of infected individuals (McAfee *et al*. 2017) and odorant proteins (Qiu and Cheng 2017). This protection can related to social behaviours (Pull *et al*. 2018) or physiological changes of nestmates (Hernández López *et al*. 2017).

1.2.3 Immunity in cockroaches

To reveal the evolution of immunity in termites, it is necessary to clearly understand the immunity of their ancestors-cockroaches, and especially their sister group, *Cryptocercus*. Many cockroaches are highly successful detritivores as well as being renowned domestic pests found across the globe (Bell *et al*. 2007). Frequent exposure to a rich antigenic environment should be associated with effective strategies to limit pathogen infection (Mayer *et al*. 2016). However, cockroach immunity has been ignored for a long time until recently the genomes of *Blattela germanica* and *Periplaneta americana* were sequenced (Harrison *et al*. 2018a; Li *et al*. 2018). Expansions of specific immune gene families have been reported in these two cockroaches, particularly of receptors GNBP and PGRP as well Toll-receptors in Toll immune pathway and hemolymph lipopolysaccharide-binding proteins (LPSBPs) (Harrison *et al*. 2018a; Li *et al*. 2018). This expansion seems to relate their adaptation to antigenic environment.

1.2.4 Evolution of immunity in social insects

As called social cockroaches, the evolve of molecular immunity in termites is very interesting, which could possibly help to understand the eusociality in social insects. In bees, it has been shown that a depauperated immune repertoire precedes the evolution of eusociality (Barribeau *et al*. 2015). In addition, there are positive selections in many immune related genes, including members of Toll and JAK-STAT pathways and serine protease inhibitors in both social and solitary bees (Viljakainen *et al*. 2009; Barribeau *et al*. 2015). In termites, positive selection has also been detected in termicin, GNBPs and Relish in *Nasutitermes* (Bulmer and Crozier 2004; Bulmer and Crozier 2005) as well as in termicin in *Reticulitermes* (Bulmer *et al*. 2010). However, how the termite immunity evolved during the evolution of eusociality is remained to be explored.

1.3 Aim of the thesis

The overarching aim of this thesis is to understand the evolution of immunity in termites in the following aspects: 1) the individual immunity in termite ancestors, cockroaches, 2) how the termite molecular immune system evolved during the transition of eusociality, 3) does the social immunity depend on caste formation in termites.

1.4 **Description of project**

In my study, I used transcriptome analysis to explore the evolution of immunity in termites. In order to explore immunity in termites, the immune genes and immune response of one of their ancestral cockroaches, *Blatta orientalis*, was firstly investigated. Secondly, I studied the evolution of immunity in a broad way by detecting the expansion and contraction of immune gene families based on a better constructed phylogenetic tree using transcriptomics. In addition, I compared the immune response among castes in a lower termite species, *Neotermes castaneus*, along with a comparison to a subsocial cockroach *Cryptocercus meridanus* and a solitary cockroach *B. orientalis*. Thirdly, to understand the high level of group immunity, I studied the social immune function of a sterile caste -soldier- in a basal termite species, *Mastotermes darwiniensis*.

In **Chapter I**, I challenged cockroach adults by injection with a mixture of heat-killed microbes (*Bacillus thuringiensis, Pseudomonas entomophila, Saccharomyces cerevisiae*) to stimulate an immune response. The immune genes in *B. orientalis* were identified and the immune response was analysed by transcriptomics. We found that *B. orientalis* has an expansion of receptors GNBP, PGRP and hemolymph LPS-binding proteins (LPSBP). This expansion also has been reported in other cockroaches, *P. americana* and *B. germanica*. After immune challenge, we found a broad immune response in *B. orientalis*, which may indicate an adaptation of antigenic environment in cockroaches.

In the first part of **Chapter II**, I constructed a phylogeny of termite species across five important families based on available transcriptomic and genomic data. The results confirm the location of termites as a sister group of *Cryptocercus*. The most recent common ancestor of both dated back to the lower Jurassic and diverged from Blattidae in the upper Triassic. In addition, the immune related genes from 47 gene families were identified across 18 species of termites and cockroaches in order to explore the expansion and contraction of immune genes. We found there is a putative loss of the drosomycin in the most recent common ancestor of *Cryptocercus* and termite species. In addition, we observed rapid changes in the diversity of immune gene families, especially notable contractions in effectors (catalase and thioredoxin peroxidase) and receptors (C-type lectin), during the origin and subsequent diversification of the major termite lineages.

Subsequently, the immune response of termite castes in a lower termite species, *N. castaneus*, was investigated in the second part of **Chapter II**. Different castes showed different immune responses after challenged with a mixture of heat-killed bacteria. Soldiers and reproductives showed a broader immune response than workers. Then, I compared the immune response of castes to the subsocial cockroach, *C. meridianus*, and the solitary

cockroach, *B. orientalis*. The cockroaches showed broad immune response whereas the immune response in termites varies in castes. These results indicate that the immune response in termites may have been shaped by the evolution of eusociality in two ways: contraction of immune gene families and the differentiated immune response.

In **Chapter III**, I studied the social immune function of soldiers in *M. darwiniensis*. Even though soldiers are unable to engage in grooming behaviour, it was found that the presence of soldiers significantly improves the survival of nestmates following entomopathogenic infection. I found that the oral secretions produced by soldiers are sufficient to protect nestmates against infection, and the secretions have potent inhibitory activity against a broad spectrum of microbes. Furthermore, I demonstrated the copious exocrine oral secretions produced by soldiers contain a high concentration of proteins involved in digestion, chemical biosynthesis, and immunity. These findings indicate that termites are likely to have evolved a sterile soldier caste with important functions not only in colony defence but also in social immunity.

In conclusion, the above mentioned results support that the termite immunity system is likely related to their eusociality. Along with the robust immune response in cockroaches, this also hints that the different immune response in termite castes is possibly related to the division of labour in termites. This is further supported by the result that social immunity at the group level is not only the effect of group size but also the formation of castes.

Chapter I

An expanded repertoire of immune genes in the cockroach *Blatta orientalis* as revealed by *de novo* transcriptome analysis

2.1 **Abstract**

The animal immune system acts as a key interface between hosts and microbes, yet little is known about immunity in a large majority of animal lineages. We address this by investigating immunity in the oriental cockroach (*Blatta orientalis*), a worldwide urban pest. The rich antigenic environment in which cosmopolitan cockroaches live makes them particularly interesting targets for research in immunity. Using a *de novo* transcriptome approach, we identify a full repertoire of insect immune genes, including all members of the canonical Toll, Immune Deficiency and JAK-STAT pathways. We report a high diversity of hemolymph lipopolysaccharide-binding proteins, which are C-type Lectins, as well an expanded set of genes involved in the Toll pathway. Following experimental immune challenge, we find that *B. orientalis* responds by inducing a broad immune response as well as shifting resources away from processes involved in transport and localization and towards immune defense. These results indicate that cockroaches possess effective and potentially long-lasting protection against infection, key traits for thriving in a rich antigenic environment. In addition to generating valuable insight into an ecologically and societally relevant insect, our study provides essential data for research into the evolution of insect immunity.

Keywords: cockroach, immune response, Toll, hemolymph lipopolysaccharide-binding protein

2.2 **Introduction**

Many cockroaches are highly successful detritivores as well as being renowned domestic pests found across the globe (Bell *et al*. 2007). Urban-dwelling cockroaches are adapted to antigen-rich surroundings due to frequent exposure to environmental microbes. Such cockroaches pose a substantial public health concern as vectors of emerging infectious diseases and as causes of allergies such as asthma (Pomés *et al*. 2017). The US Food and Drug Administration recognizes four common worldwide cockroach pest species: *Blattella germanica* (German cockroach), *Blatta orientalis* (Oriental cockroach), *Periplaneta americana* (American cockroach), and *Supella longipalpa* (Brown-banded cockroach). Many of the characteristics associated with these globally invasive pests represent attractive targets for research, including for studies into toxicology, chemical metabolism and communication (Li *et al*. 2018). Cockroaches also represent model organisms in social evolution (Lihoreau *et al*. 2012; Harrison *et al*. 2018b), behavioral ecology (Logue *et al*. 2009; Lihoreau and Rivault 2010), neurobiology (Booth *et al*. 2009), gut microbiota (Bertino-Grimaldi *et al*. 2013; Wada-Katsumata *et al*. 2015), as well as being a potential source of novel antimicrobial peptides for use in applied medicine (Lee *et al*. 2012; Kim *et al*. 2016; Mylonakis *et al*. 2016; Chowański *et al*. 2017).

Frequent exposure to a rich antigenic environment should be associated with effective strategies to limit pathogen infection (Mayer *et al*. 2016). Indeed, cockroaches employ both behavioral and physiological immune mechanisms to mitigate opportunistic infections. Cockroach behavioral immunity can include avoidance of dead infected conspecifics (Kaakeh *et al*. 1996), grooming (Bell *et al*. 2007), and even body temperature adjustments following immune-challenge (Bronstein and Conner 1984). In terms of physiological immunity, cockroaches possess robust innate mechanisms, including both cellular and humoral immune components. Following bacterial infection, cockroaches respond with cellular immunity, which can include phagocytosis and nodule-formation (Verrett *et al*. 1987; Rahmet-Alla and Rowley 1989; Kulshrestha and Pathak 1997). With respect to humoral immunity, many antimicrobial peptides have been identified from the american cockroach, *P. americana* (Kim *et al*. 2016) as well as several antibacterial and antifungal proteins, which have been characterized from the hemolymph (Jomori *et al*. 1990; Jomori and Natori 1991; Basseri *et al*. 2016; Arumugam *et al*. 2017). Interestingly, american cockroaches are thought to produce a two-phase immune response following infection (Faulhaber and Karp 1992) consisting of an initial short non-specific phase followed by a longer specific phase, possibly mediated by hemocytes (Ryan and Karp 1993) and/or proteins in hemolymph (Karp *et al*. 1994). However, until recently, the molecular mechanisms of cockroach immunity have remained poorly understood.

The insect immune system has been studied extensively in recent years, particularly in flies and beetles (Hoffmann 2003; Hoffmann and Reichhart 2002; Irving *et al*. 2001; Tauszig *et al*. 2000; Pham *et al*. 2007; Haine *et al*. 2008; Rolff and Reynolds 2009; Arefin *et al*. 2014; Buchon *et al*. 2014; Milutinović *et al*. 2016; Johnston *et al*. 2014; Duneau *et al*. 2017; Zanchi *et al*. 2017). Insect innate immune molecules occur as three broad types (not withstanding exceptions): receptors, signaling components and effectors (Viljakainen 2015; Hillyer 2016). Following infection, pattern recognition receptors bind to microorganisms, which leads to the induction of three principal signaling pathways responsible for the regulation of the insect humoral immune response, known as the Toll, Immune Deficiency (IMD) and Janus kinase/signal transducers and activators of transcription (JAK-STAT) pathways. These canonical pathways are responsible for, amongst other effects, the synthesis of antimicrobial peptides such as defensins and attacins (Hillyer 2016). Many of the functions of genes involved in these pathways derive from a considerable body of research carried out in Drosophila and to a lesser extent, other insects. In flies we understand that the Toll pathway responds largely to fungi and gram-positive bacteria, and is mediated by peptidoglycan receptor proteins (PGRPs), gram-negative binding proteins (GNBPs), serine protease cascades, Toll-receptors, Myeloid differentiation primary response 88 (MyD88), Tube, Pelle and Dorsal-related immunity factor (Dif)/Dorsal transcription factors (Valanne *et al*. 2011). The IMD pathway mainly responds to gram-negative bacteria, and is comprised of PGRPs, IMD, Fas-associated protein with death domain (FADD), Dredd, Transforming growth factor-activated kinase 1 (TAK1)-binding protein (TAB), TAK, IκB (inhibitor of nuclear factor κB) kinase (IKK) and Relish (Myllymäki *et al*. 2014). Conversely, the JAK-STAT pathway is thought to regulate inflammation and stress responses. It is principally composed of Cytokines, Domeless, Hopscotch and Signal transducers and activators of transcription (STAT) (Agaisse and Perrimon 2004). Last but not least, melanization plays a key role in insect immunity and is mediated by Pattern Recognition Receptors (PRR), serine proteinase cascades and phenoloxidase (Cerenius *et al*. 2008; González-Santoyo and Córdoba-Aguilar 2012).

Two recently published cockroach genomes, *B. germanica* and *P. americana* (Harrison *et al*. 2018b; Li *et al*. 2018) in addition to some transcriptomic studies (Zhou *et al*. 2014; Chen *et al*. 2015) indicate that these cockroaches possess a full repertoire of canonical insect immune pathways (Li *et al*. 2018). But next to nothing is known about the Oriental Cockroach, *B. orientalis*, a major yet neglected common cockroach pest species. Here, we carry out a systematic transcriptomic survey of *B. orientalis* immunity by analyzing differential gene expression following immune challenge. We show that *B. orientalis* possesses an extensive range of immune genes, including major expansions of immune families as well as a strong

immune response to immune challenge. Our study contributes much needed insight into a highly relevant but until recently overlooked group of insects.

2.3 **Material and Methods**

Insect culture

The adults of *B. orientalis* were provided by the German Environment Agency, Umwelt Bundes Amt and kept at 26 °C, 75% relative humidity in the dark. They were fed with *ad libitum* access to food (77.0 % dog biscuit powder, 19.2 % oat flakes and 3.8 % brewer's yeast) supplemented with apples and carrots, which were replaced weekly. We collected ootheca from adults at the same day to set up our experiment. Following hatching from ootheca, individual juveniles were kept separately in boxes in the same conditions as above, until the adult stage. Adults were immune challenged within 1-2 weeks after the final molt.

Microorganisms preparation

Pseudomonas entomophila (DSM 28517^T, Gram-negative), *Bacillus thuringiensis* (DSM 2046^T, Gram-positive) and *Saccharomyces cerevisiae* (DSM 1333^T) were used to raise a broad immune response in challenged cockroaches. *P. entomophila* and *B. thuringiensis* were purchased from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) and were stored at -70 °C in the Bundesanstalt für Materialforschung und -prüfung (BAM) prior to use. *S. cerevisiae* was available via the BAM microorganism collection (https://agw3.bam.de/biomikrosearch/searchRefOrg). *P. entomophila* and *B. thuringiensis* were activated overnight before being inoculated for growth at 28 °C and 30 °C in nutrient broth (recipe following to DSMZ instruction), respectively. *S. cerevisiae* was activated at 25 °C in universal yeast medium and grown for 36 hours. All cultures were washed twice with Ringers' solution, heat-killed at 95 °C for 10 min and mixed equal amount to form a cocktail with a final concentration of $5*10^8$ ml^{-1}.

Immune challenge

Adult cockroaches were weighed and injected with $5*10^6$ equivalent of cells per gram of the prepared microbial cocktail between 5^{th} and 6^{th} ventral abdominal sternites after being swabbed with 96% ethanol. Control adults were injected with the same amount of Ringer' solution adjusted by weight. We collected two replicates of four independent biological individuals for both the control and infected groups. After injection, cockroaches were kept in 55 mm diameter cups individually supplied with fresh water for 24 h before being frozen with liquid nitrogen. Samples were stored at -70 °C until RNA extraction.

RNA isolation and purification

Whole insects were used for total RNA isolation. Each individual was cut into 4-6 pieces with sterile scissors. For RNA extraction, each piece was suspended in pre-cooled Trizol (Thermo Fisher Scientific), and homogenized with a 5-mm steel bead (Qiagen) using a FastPrep®-24 homogenizer (MP Biomedicals) twice at 4 m/s for 15 s. Recovery of RNA was followed according to manufacturer's instructions for Trizol (Thermo Fisher Scientific), with chloroform extraction and isopropanol precipitation, followed by re-dissolving RNA in storage solution (Ambion). RNA from extracted pieces were pooled for individual cockroach samples and subsequently incubated with 2 units of TurboDNase (Ambion) for 30 min at 37 °C and then purified using an RNAeasy Mini kit (Qiagen) following manufacturer's instructions. Quantity and quality of RNA were determined by Qubit and Bioanalyzer 2100.

De novo transcriptome sequencing

Four barcoded, non-normalized cDNA libraries were prepared using NEXTflex™ Rapid Directional mRNA-seq kit (Bioo Scientific) and represented two replicates from challenged and control treatments. Libraries were prepared according to manufacturer's instructions. Briefly, polyadenylated mRNA was enriched by poly-A beads from 10µg pools of total RNA by pooling equal quantities from 4 individuals for each replicate. First-strand and second-strand cDNA from each pool was synthesized, fragmented and barcoded with NEXTflex™ RNA-seq Barcode Adapters. The prepared libraries were sequenced on an Illumina NextSeq500/550 platform at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv).

Transcriptome assembly and annotation

Raw reads were trimmed to remove sequencing barcodes and cDNA synthesis adaptors, while reads shorter than 25 bp following trimming were discarded using Trimmomatic as incorporated inside Trinity (version 2.3.2) (Grabherr *et al*. 2011). FastQC was initially employed to assess sequencing quality. Pair-end reads from all libraries were assembled using Trinity with default k-mer length (25). The assembly quality was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO v2) with the Insect BUSCO set from orthoDB (version 9) (Simão *et al*. 2015) as well as by examining the representation of reads. The assembly was subjected to BLASTp against nr database from NCBI by Diamond (Buchfink *et al*. 2015) for acquiring the taxonomic composition of the best blast hits and gaining insight into the presence of other organisms in samples.

The assembly was annotated by following the guidelines of Trinotate (https://trinotate.github.io/). The proteins from the assembly were predicted by TransDecoder (version 3.0.1) (http://transdecoder.github.io). Homology searches, predictions and domain identifications were performed locally and subsequently integrated into database at an e-value

threshold of 1e-03. Briefly, nucleotide and predicted peptide sequences predicted by TransDecoder were used to query SwissProt with BLASTx and BLASTp, respectively. Protein domains, signal peptides, and transmembrane domains were determined by HMMER (v3.1b2) against the pfam database(Finn *et al*. 2011), SignalP 4.0(Petersen *et al*. 2011), and TmHMM 2.0 (Krogh *et al*. 2001), respectively.

Immune related proteins identification

To confirm the identity of predicted proteins, a complementary prediction method was employed to search for proteins with putative immune function. We employed HMMER to identify proteins using a domain-based search strategy. Then we complemented a HMMER search with a blast approach inside the trinotate suites. To quantify the presence of domains containing putative immune functions, we modified a previously published method (Sackton *et al*. 2017). Briefly, immune gene families from 31 species (available on https://github.com/ShulinHe/Blatta_orientalis) in the orthoDB database as well as Termicin and Transferrins from Uniprot (insects) were first downloaded. We built a set of HMM profile-curated alignments based on all protein families. The complete set of predicted proteins (> 60 amino acids in length) from transcriptomes were searched for matches against predicted immune-related HMMs using HMMER 3.1. Afterwards, the HMMER output was filtered by: excluding targets with E-values > 0.001 for the best domain, excluding targets with overall E-value greater than 10^{-5}, and assigning the targets that have multiple HMMs to best e-value HMM. The genes that have multiple immune predicted proteins from different isoforms was assigned to the protein that has the highest overall E-value HMM. The filtered HMMER output were then further selected using annotations from trinotate. Putative gene targets were selected when the HMMER output of their predicted proteins fitted their annotations of blastp and blastx in trinotate. Subsequently, targets were removed when their predicted proteins were shorter than 100 amino acids in families other than antimicrobial peptides. We adopted a conservative approach for accepting the identity of immune gene target. Firstly, because it is theoretically possible that different components from the same subcluster may represent spliced isoforms of a single gene, we aligned nucleotide sequences and corresponding predicted proteins from each subcluster against one other using MAFFT (Katoh *et al*. 2017) and excluded sequences that were variable in length but otherwise identical (this applied to 5 of 377 putative immune gene sequences). Secondly, to account for different fragments of the same gene potentially appearing in different subclusters of a single cluster (and being erroneously described as two separate genes), we ran an additional blastx search on all putative subcluster sequences. If more than one subcluster had an identical target in the top 10 entries of a DIAMOND blastx search (and overlapped by less than 9 amino acids – a value determined by the use of a 25 k-mer parameter during transcriptome assembly), only the

longest subcluster was retained (this applied to 13 of 372 putative immune gene sequences). These additional measures enabled us to more accurately differentiate between spliced isoforms or fragmented gene sequences and true paralogs. The identified hemolymph lipopolysaccharide-binding proteins (LPSBPs) were compared with LPSBPs annotated from *Z. nevadensis*, *B. germanica* and *Cryptotermes secundus* by building a gene tree from all sequences aligned to a reference LPSBP sequence from *P. americana* (Appendix I-A, Appendix I-B).

Transcript Abundance Estimation and Differential Expression Analysis

Transcript expression following treatment was estimated by Kallisto (Bray *et al*. 2016). To minimize the potential influence of transcripts from symbionts, including protist and potential bacterial contamination, we excluded gene expression data according to taxonomic analysis. Differential gene expression was analyzed using the R package DESeq2 (Love *et al*. 2014) with standard settings in conjunction with tximport (Soneson *et al*. 2015). We defined genes as being significantly differentially expressed when fold changes were larger than 2, with an adjusted p-value < 0.05. Differentially expressed genes were subject to Gene Ontology (GO) enrichment analysis, as performed by the R package goseq with an adjusted p-value cut-off of 0.05. The GOs were extracted from Trinotate annotations. After enrichment analysis, GO redundancy was reduced by using REVIGO (Supek *et al*. 2011).

Quantitative PCR

Total RNA from each individual for sequencing was used for quantitative PCR. cDNA was synthesized with M-MLV Reverse Transcriptase (Promega) using Random (Promega) and Oligo(dT)15 Primer (Promega) according to manufacturer's instructions. The genes and primer sequences used for quantitative PCR are listed in Appendix I-D. Relative expression of these genes was determined using SensiFAST™ SYBR Lo-ROX Kit (Bioline) following three-step cycling. A standard curve of pooled, five-times serially diluted cDNA was run for the chosen genes. RPL22 (ribosomal protein 22) was used as a reference gene. Fold-change calculations were performed by using the Pfaffl method (Pfaffl 2001) and a Mann–Whitney U test was employed to compare gene expression between treatment and control groups using R v.3.2.3 (Team 2016). Data are presented as means ±SE.

Data availability

Appendix contains two figures of LPSBPs, a phylogenetic tree of LPSBPs (Appendix I-A) and an alignment of LPSBPs (Appendix I-B), a figure of fold changes of the genes in three immune pathways (Appendix I-C), a table of primer information for Quantitative PCR (Appendix I-D) and a table of fold changes of immune genes in the Toll pathway for 3 different species

(Appendix I-E). Appendix I-F contains details of identified immune related genes. Appendix I-G contains output of Gene Ontology analysis of differentially expressed genes. Sequence data are available at NCBI SRA under the accession number: SRP150731. Full code and scripts to perform the analyses in this study are made available at https://github.com/ShulinHe/Blatta_orientalis.

2.4 Results

2.4.1 Transcriptome statistics

In total, 151.4 million RNA-seq raw reads were generated from all libraries. Depending on the library, approximately 0.4 % of the reads were excluded after trimming and quality control, leaving 150.8 million reads available for subsequent *de novo* transcriptome assembly.

Table 2.1 Number of identified immune related genes for each family.

Family name	No. of genes	Family name	No. of genes
AMPs		**Receptors**	
Attacin	2	GNBP	9
Holotricin	2	PGRP	15
Drosomycin	1	Toll_receptor	11
Defensin	2	Spaetzle	7
Termicin	2	Fibinogen Related protein	4
Canonical immune		Galectin	5
Catalase	7	C-type Lectins (Hemolymph lipopolysaccharide-binding proteins [LPSBPs])	54 (46)
Transferrin	3		
Lysozyme	12		
Peroxiredoxin	7	MD2-Like Receptors	7
PPO	1	Thioester-Containing Proteins	4
Hemocyanin	1	**Other**	
Glutathione peroxidase	2	Apoptotic protease-activating factor	1
Peroxidase	16	Inhibitor of apoptosis	2
		Caspase	7
Pathways		Autophagy protein	19
Toll_pathway members	12	Scavenger receptor	17
IMD_pathway members	10	Clip-Domain Serine protease	103
JAK_STAT members	4	Serpin	23

The assembly contained 475,977 transcripts clustered into 400,034 contigs with E90N50 of 1151bp. The BUSCO analysis identified 97.3% complete orthologs (58.9% single-copy orthologs and 38.4% duplicated orthologs), 2.0% fragmented orthologs, and 8% missing orthologs. The assembly represented 94.24% of reads after mapping by bowtie2. We found the blastp results of the assembly when run against the nr database to be composed as follows: 23.7% *Blattella* (cockroach), 22.9% *Cryptotermes* (termite), 13.7% *Zootermopsis* (termite), 2.6% *Nilaparvata* (planthopper), 1.7% *Myzus* (aphid), 1.5% *Centruriodes* (scorpion), and 33.8% other. We used Trinotate to annotate our assembly and, in total, 21.9% of the transcripts (104,396 of 475,977) were annotated by trinotate suites.

2.4.2 Immune related gene identification

We used an HMM-based approach to identify predicted proteins with homology to previously characterized immune related gene families from 31 insect species. We found 372 immune genes in total from our assembly, including conserved Toll, IMD, and JAK-STAT pathways members as well as canonical receptors and effectors (Table 2.1; Appendix I-F). In these identified immune genes, 51.61% (192) consisted of complete open reading frames (ORF),38.00% (141) of 5' prime partial ORFs, 2.15% (8) of 3' prime partial ORFs and 8.33% (31) of internal ORFs.

2.4.3 Gene ontology enrichment analysis following immune challenge

After removing bacterial and protist transcripts, 99.7% of the total transcripts (472,826) were subjected to differential gene expression analysis. Of the 394,960 "genes" in *B. orientalis* with detectable expression in our analysis, 562 (FDR<0.05) were upregulated following immune

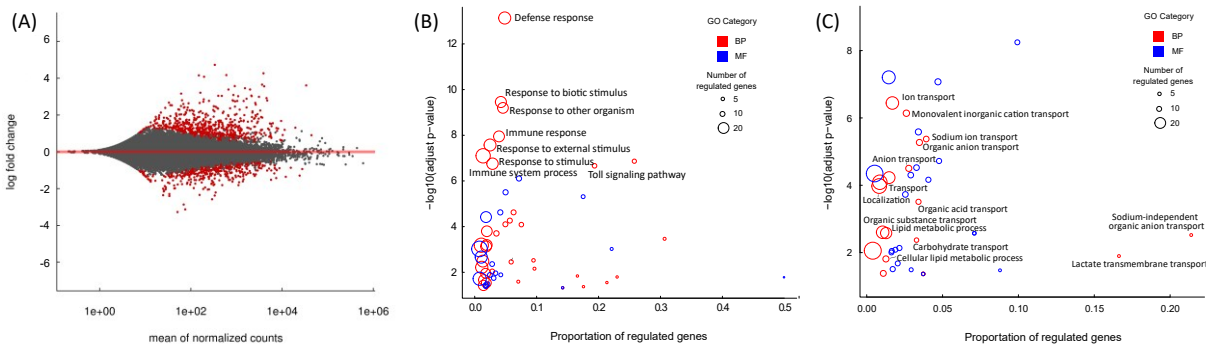


Figure 2.1 A) MA plot of expressed and differentially expressed genes marked in grey and red respectively. Differential expression analysis was performed by DEseq2. B) Plot of enriched GO categories in the immune-challenge group all relate to "Biological process" (BP, in red) and "Molecular function" (MF, in blue), except a single GO term (GO:0042943, Molecular Function, D-amino acid transmembrane transporter activity, adjusted p-value: 0.043, 2 genes upregulated [of 3 in total]). C)Plot of enriched GO categories in the control group all relate to BP in red and MF in blue. GO analysis was performed by goseq script in Trinity software and reduced redundancy by REVIGO.

challenge while 380 genes (FDR<0.05) were downregulated, representing 0.14% and 0.09% of expressed genes, respectively (Figure 2.1). Of the upregulated and downregulated genes, 87.3% (491) and 69.2% (263) are significantly differentially regulated compared to the control treatment. This reduced set of differentially expressed genes was used for GO clustering to uncover broad changes occurring in cockroaches following immune challenge.

As expected, genes upregulated by immune challenge are enriched for GO terms relating to immunity and stimulus response. Additionally, the upregulated genes were enriched in GO terms relating to bacterial structure degradation as well as in biological process GO terms that are suggestive of a coordinated protein synthesis, including “protein processing”, “regulation of cytokine production” and “proteolysis” (Figure 2.1, Appendix I-G). In contrast, genes downregulated by infection are enriched for GO terms that were related to transport, localization, and lipid metabolic process (Appendix I-G). These patterns indicate a physiological shift in cockroaches from transport and lipid metabolic to immune defence and stimulus response.

2.4.4 Immune gene regulation after infection

Of the differentially expressed genes, 42 were annotated as immune related genes, including 29 induced (5.91% of total differentially upregulated genes) and 13 repressed genes (4.94% of differentially downregulated genes). The differentially regulated immune related genes after

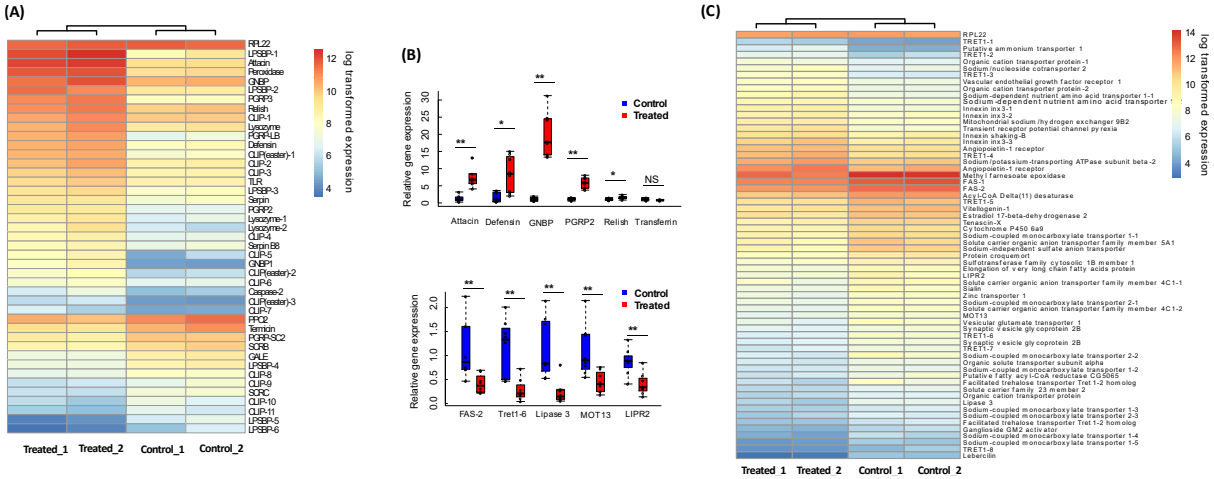


Figure 2.2 A) Heatmap of differentially expressed immune genes from DESeq2 analysis. B) qRT-PCR of attacin, defensin, GNBP, PGRP2, relish (upregulated) and transferrin (unchanged), Fatty acid synthase-2(FAS-2), Facilitated trehalose transporter Tret1 (Tret1-6), Lipase 3, Monocarboxylate transporter 13(MOT13), Pancreatic lipase-related protein 2(LIPR2) (downregulated) using RPL22 as a reference gene. Significance level comparisons: **, p<0.001; *, p<0.05; NS, not significant. Transferrin was not differentially expressed in the DESeq2 analysis (or qPCR) and so is not represented in panel A. C) Heatmap of differentially expressed transport and lipid metabolism related genes from DESeq2 analysis.

infection represented 11.29% of the total immune related genes that were identified (including oxidases and autophagy related genes, as well as C-type lectins, which are not included in the GO term “immune response” from the Trinotate annotation.). Of these genes, 24 were with complete ORFs, 16 were 5’ prime partial ORFs and 2 were internal ORFs. Upregulated immune related genes included antimicrobial peptides (attacin and defensin), recognition factors (3 hemolymph lipopolysaccharide-binding proteins [LPSBPs], 2 GNBPs,3 PGRPs), and signaling pathways components (1 caspase-2, 10 serine proteases, 2 serpins, 1 Relish and 1 Tolls) as well as 3 lysozymes and 1 peroxidase. Downregulated immune genes included 4 serine proteases, 3 LPSBPs, a Galectin-8, a PGRP-SC2, a Phenoloxidase 2 and a termicin. The expression of these immune genes is shown as a heatmap in Figure 2.2. We confirmed a subset of the expressed genes (5 upregulated, 5 downregulated, 1 no change) by quantitative PCR (Figure 2.2).

2.5 **Discussion**

We analyzed the immune repertoire and response of *B. orientalis* to a general immune challenge to gain greater insight into the molecular basis of immunity in this highly successful cosmopolitan pest species. Using a *de novo* approach, we assembled a transcriptome with high completeness, enabling us to identify 372 immune-related genes based on orthoDB and *Z. nevadensis* immune ortholog group predictions. We detected a broad response to immune challenge involving a number of established immune pathways, and this broad response was associated with significant shifts away from energy storage and cellular transport processes.

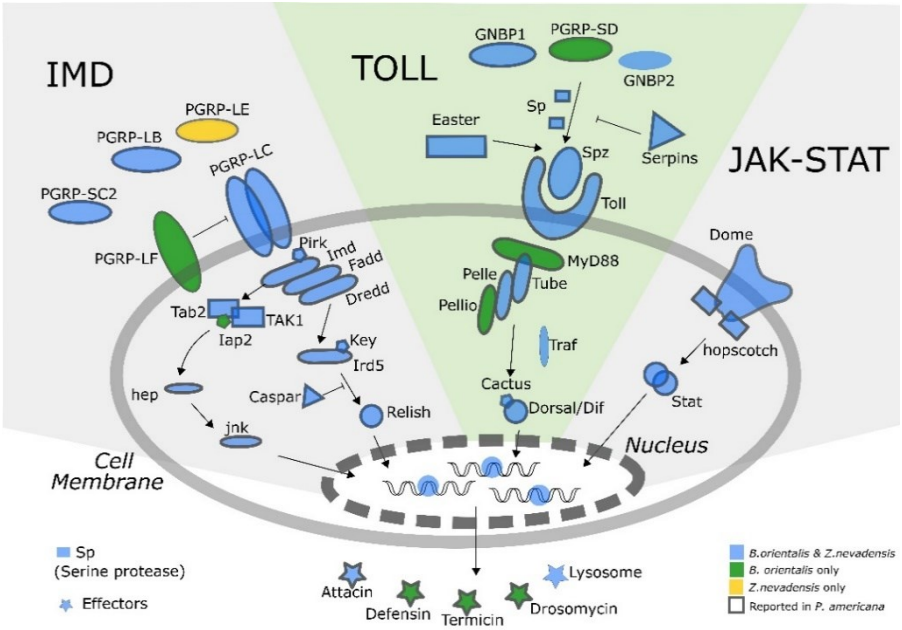


Figure 2.3 Schematic representation of members of the three main immune pathways (IMD, TOLL and JAK-STAT) in *B. orientalis*, as compared with *Z. nevadensis* and *P. americana*. Reported genes with a gray border indicate that these genes are also described in *P. americana*. Immune gene information combines data from the present study with two others (Terrapon *et al*. 2014; Li *et al*. 2018).

In comparison to other well studied insects, we find that *B. orientalis* possesses a conserved repertoire of immune genes, corroborating findings from two other cockroach species, *B. germanica and P. americana* (Dziarski 2004; Jeong *et al*. 2014; Zhou *et al*. 2014; Li *et al*. 2018). Components of entire pathways including Toll, IMD and JAK-STAT were identified (Figure 2.3), which is in contrast to some other insects such as the pea aphid, *Acyrthosiphon pisum* (Gerardo *et al*. 2010). Interestingly, we found a relatively expanded Toll pathway in *B. orientalis*, including 9 GNBPs, 11 Tolls and 7 spaetzles (Figure 2.4). This pattern of expansion also applies to *P. americana* and *B. germanica*, but not to the termite *Zootermopsis nevadensis* or to more distantly related insects such as *Tribolium castaneum* (Zou *et al*. 2007). This indicates a possible localized expansion in the cockroaches.

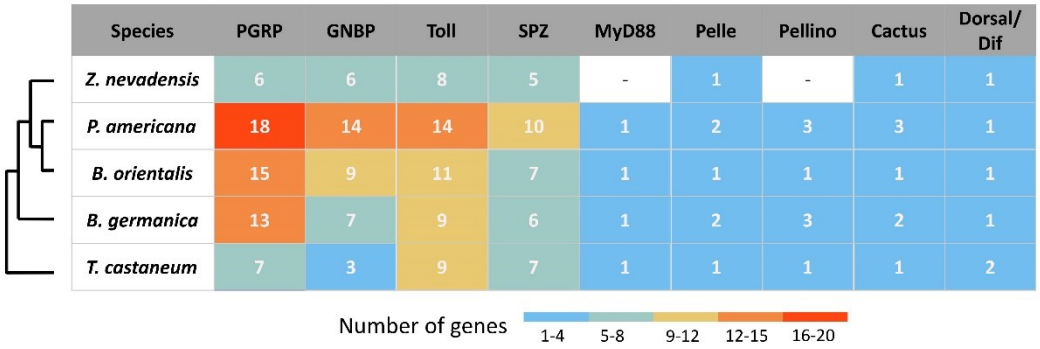


Figure 2.4 Number of predicted PGRP and Toll pathway genes. The cladogram is based on established insect relationships (Misof *et al*. 2014). Gene numbers derive from this and three other studies (Zou *et al*. 2007; Terrapon *et al*. 2014; Li *et al*. 2018). Box colors represent number of genes determined per family. White = not detected.

We identified 46 putative Hemolymph lipopolysaccharide-binding proteins (LPSBPs) in *B. orientalis* in addition to 8 other C-type Lectins (CTLs) (Table 1, Appendix I-F). Such a high diversity of CTLs has not been reported in any other insect until a recent report of 86 LPSBPs in *B. germanica* (Harrison *et al*. 2018a), although some holometabolan insect lineages (Diptera, Lepidoptera) reportedly possess moderately high species-specific expansions of CTL genes (Xia *et al*. 2017). We confirmed the identity and evolutionary divergence of cockroach LPSBPs by comparing our *B. orientalis* predicted protein sequences (N=46) against annotated LPSBPs from *B. germanica* (N=37); *Z. nevadensis* (N=39) and *Cryptotermes secundus* (N=24) (Appendix I-A, Appendix I-B). These data indicate the presence of a conserved expansion of diverse LPSBPs in cockroaches and termites. As a form of C-type Lectin, LPS-binding proteins may function as opsonins by binding surface molecules of invading microorganisms (Jomori *et al*. 1990; Jomori and Natori 1991; Jomori and Natori 1992). A C-type Lectin from the hemolymph of the cockroach, *P. americana*, has also been shown to possess phenoloxidase activity (Chen *et al*. 1995; Arumugam *et al*. 2017). Clearly, much greater research is required to understand the precise functions of these effectors in cockroaches, which may also include roles in nodule formation, melanization, encapsulation as well as microbiome regulation (Xia

et al. 2017). Such a high diversity of lipopolysaccharide binding proteins in *B. orientalis* points towards a strong immune effector presence in cockroach hemolymph, yet another indicator of this cockroach's ability to thrive in a rich microbial environment. Hemolymph LPS-binding proteins have also been implicated in the acute non-specific phase of the cockroach immune response (Jomori and Natori 1991) and we suspect that they could also feature in a more specific second phase of cockroach immunity (Faulhaber and Karp 1992), although this remains speculative. We also identified 15 PGRP proteins, similar to the 18 PGRPs found in *P. americana*, but more than the 13 and 6 PGRPs detected in *B. germanica* (Li *et al*. 2018) and the termite *Z. nevadensis* (Terrapon *et al*. 2014) respectively. This expansion of PGRP and Hemolymph LPS-binding proteins might explain the relatively specific (Faulhaber and Karp 1992) and strong antimicrobial response (Li *et al*. 2018) of cockroaches towards gram-negative bacteria. Such an effective response coupled with the need to identify effective antimicrobials against gram-negative bacteria could make these insects promising targets for novel antimicrobial compounds (Kim *et al*. 2016).

Antimicrobial peptides play a crucial role in the insect humoral immune response. We identified the classical antimicrobial peptides, attacin and defensin as well as five other defensin-like peptides: 2 termicins, 1 drosomycin and 2 holotricins. Attacin is a glycine-rich protein mainly possessing antibacterial activity against Gram-negative bacteria by binding the bacterial outer membrane and inhibiting protein synthesis (Carlsson *et al*. 1991; Carlsson *et al*. 1998). Defensin is a cysteine-rich peptide possessing antibacterial activity against Gram-positive bacteria by forming bactericidal channels in the outer membrane (Cociancich *et al*. 1993; Maget-Dana and Ptak 1997). The total number of antimicrobial peptides in our study was similar to the number identified in *P. americana* (11 AMPs) but more than the number reported in *B. germanica* (6 AMPs, although see (Harrison *et al*. 2018a) which unexpectedly reports 10 copies of drosomycin) and *Z. nevadensis* (2 AMPs) (Terrapon *et al*. 2014; Li *et al*. 2018). This AMP diversity could provide an additional layer of protection, potentially contributing to the diphasic immune response previously described in *P. americana*. Evidence for a diphasic response has also been found in *Tenebrio* beetles, which possess an expanded set of Tenecin AMPs that remain activated for a long period following infection (Johnston *et al*. 2014). In cockroaches and termites, the AMP Termicin, which was first identified in *Pseudacanthotermes spiniger* (Bulmer *et al*. 2012; Liu *et al*. 2016), shares structural similarities with defensin (Da Silva *et al*. 2003) and shows strong antifungal activity (Lamberty *et al*. 2001). Drosomycin is another antifungal antimicrobial peptide and it is regulated by the Toll pathway in Drosophila (Zhang and Zhu 2009). An abundance of antifungal AMPs suggests strong selection for defence against pathogenic fungi in cockroaches: traits that could well have been crucial during the subsequent expansion of the soil and substrate-dwelling termites.

After being challenged by a mixture of microbes including gram-negative and gram-positive bacteria and a yeast, cockroaches responded by regulating a number of relevant immune pathway components, including molecules involved in recognition and signaling as well as effector molecules (Appendix I-D). In general, GO-terms pointed to a significant enrichment of upregulated genes involved in host-immune defence and bacterial cell wall degradation, as well as upregulation of serine proteases and serine protease inhibitors. By contrast, downregulated genes were significantly enriched in functions relating to transport (both biological process and molecular function categories) as well as nutrient-reservoir activity, indicating a shift away from energy storage and cell-transport processes and towards immunity. Surprisingly, except two lipid metabolic related GO terms, we did not detect enrichment of genes directly involved in other metabolic activity, suggesting that cockroaches possess and utilize abundant energy reserves during infection.

Of the differentially regulated immune genes, we identified two antimicrobial peptides: attacin and defensin. Attacin and defencin may function together to regulate mixed infections. Alternatively, they may act synergistically by targeting components of bacterial cells (Baeder *et al*. 2016; Yu *et al*. 2016). In addition, we found that three hemolymph LPS-binding proteins were induced, which as described above are C lectin-related proteins that are thought to function as opsonins (Jomori and Natori 1992). Along with other detected canonical effectors such as lysozymes, the induction of these antimicrobial proteins indicate that cockroaches possess a broad response to infection. Induced proteins also included pattern recognition receptors (GNBPs, PGRPs), Toll receptors, Relish, serine proteases as well as serpins, demonstrating that *B. orientalis* engages both the Toll and IMD pathway to regulate antimicrobial protein and peptide expression. These findings show that cockroaches, like other insects, possess a full capacity to respond to infection (an example of toll pathway members in Appendix I-E), beginning with microbial recognition and ending with microbial elimination, supporting results reported previously for *P. americana* (Li *et al*. 2018). Interestingly, Termicin, which plays a an important antifungal role in the eusocial termites (Lamberty *et al*. 2001; Da Silva *et al*. 2003) was downregulated in cockroaches following immune challenge. This protein harbors a CSαβ structure, much like defensin, in addition to an amidated C-terminal, possibly explaining its primary function against fungi (Lamberty *et al*. 2001; Yi *et al*. 2014). The downregulation of this gene might be the result of the specific nature of the microbial mixture used to challenge the cockroaches. On the other hand, the cockroach immune response has been reported to last for over 14 days (Faulhaber and Karp 1992), indicating that further mechanistic studies over a longer time frame are required to understand the complete temporal dynamics of cockroach immunity.

To conclude, we find that *B. orientalis* possesses significant immune gene expansions including a high diversity of effector proteins, an enriched Toll pathway, and a broad response to immune challenge. Such a powerful armory is likely to provide effective and potentially long-lasting protection against infection: key traits for thriving in rich antigenic environments. In addition to generating valuable insight into an ecologically and societally-relevant group of insects, our study provides essential data for comparative research exploring the evolution of insect immunity.

2.6 **Acknowledgements**

We thank Jens Rolff for useful advice on this Chapter, Susan Mbedi (BeGenDiv) for technical support in transcriptome sequencing. Sequencing was performed at the Berlin Center for Genomics in Biodiversity Research. The High-Performance-Computing Facilities at ZEDAT (FU-Berlin) are also acknowledged for providing access to computational resources.

Chapter II

Evolution of termite immunity during the transition of eusociality

3.1 **Abstract**

As a major group of social insects, termites are an important target for the evolution of eusociality. However, termite immunity and knowledge relating to its evolution are unclear. In this study, we employed transcriptomics to study the evolution of individual immunity in termites. Firstly, we constructed a comprehensive phylogeny of termites and cockroaches based on phylogenomic data. Secondly, we explored the evolution of termite immune system by detecting the contraction and expansion of immune gene families in 18 species of termite and cockroach across a gradient of eusociality. Finally, we compared immune responses of a social termite, *Neotermes castaneus* with a solitary cockroach, *Blatta orientalis* and a subsocial cockroach, *Cryptocercus meridianus*. As a result, we found that the evolution of eusociality in termites can be dated to the lower Jurassic. In addition, we observed rapid changes in the diversity of immune gene families, especially notable contractions in effectors (catalase and thioredoxin peroxidase) and receptors (C-type lectin), during the origin and subsequent diversification of the major termite lineages. Furthermore, different immune responses were detected between termite castes, which may be a consequence of division of labor in termites. Interestingly, the immune response of the subsocial *C. meridianus* was similar to the response observed in the solitary cockroach *B. orientalis*. These results suggest that the molecular immune system in termites has been modulated by the evolution of eusociality. These findings provide important sights into the evolution of the immune system in a major social insects group, increasing needed knowledge concerning the key evolutionary event of eusociality.

Keywords: phylogeny, subsocial, contraction and expansion, caste, C-type lectin

3.2 **Introduction**

The origin of eusociality is considered to be one of the major evolutionary transitions (Szathmáry and Smith 1995). It occurs mostly in social insects, which live in groups of hundreds to millions of individuals. The hallmark of eusociality is the appearance of a permanently sterile caste, which in social insects can be achieved in two ways: via the evolution of a worker caste or the evolution of a solider caste (Tian and Zhou 2014). The former applied to social insects in Hymenoptera (ants, bees and wasps) and the latter applied to social termites. Compared with the well-studied Hymenoptera, termites are a key model for the study of the evolution of eusociality in the social societies where the soldier caste was the first sterile caste to evolve.

Termites are hemimetabolous diploid insects, which in contrast to the holometabolous haplodiploid Hymenoptera (Korb 2008). They are a sister group to *Cryptocercus,* a subsocial wood-feeding cockroach genus that lives in family groups (Inward *et al*. 2007a). Therefore, termites are also called as “social cockroach”. Evolved from a solitary cockroach ancestor, these lineages represent an interesting transition between solitary, subsocial and truly social groups.

During the evolution of eusociality, the formation of a social system with a permanently sterile caste represents a crucial point of no-return transition (Szathmáry and Smith 1995; Boomsma and Gawne 2018). In termites, the soldier is a sterile caste that presents in all species except a secondary evolutionary loss in a few higher termites (Inward *et al*. 2007b; Bourguignon *et al*. 2016a). Apart from that, true workers, a secondarily evolved sterile caste, can be found in all higher termite species and some lower termite species. Other lower termites that lack the sterile worker caste have a majority of false-workers (“pseudogates”) in colonies, which have the ability to develop either into soldiers or reproductives. In addition to sterile castes, termites have a reproductive caste: primary reproductives and/or neotenic reproductives. Primary reproductives consist of queens and kings that found the colony after a dispersal flight. They are winged and represent a terminal developmental stage. Neotenic reproductives, mostly known from lower termites, are replacement queens/kings that develop from the natal colony (Myles 1999; Korb and Hartfelder 2008). They also represent a terminal developmental stage with neotenic morphological features, such as aptery and a weakly sclerotized cuticle. In possessing a suite of divergent morphological and behavioral adaptions, different castes in termite colonies are specialized to perform different tasks, for example, soldiers for defense, (false) workers for foraging and reproductives for reproduction (Legendre *et al*. 2008; Tian and Zhou 2014; Engel *et al*. 2016). An effective system of differentially

specialized castes is thought to be one of the main reason for the raise of social insects, including termites, as ecosystem-dominating life forms (Oster and Wilson 1978).

An important adaption of eusociality in social insects is effective immune mechanism against easy spread of disease/pathogens in a high population density colony of genetical close-related members (Alexander 1974; Schmid-Hempel 1998). The immune system in social insects is composed of individual immunity and social immunity. As a social colony is constituted by individuals, each member would possess individual immune system, as is the case in other solitary insects. Individual immunity has been studied especially in flies and beetles (Hoffmann 2003; Hoffmann and Reichhart 2002; Irving *et al*. 2001; Tauszig *et al*. 2000; Pham *et al*. 2007; Haine *et al*. 2008; Rolff and Reynolds 2009; Arefin *et al*. 2014; Buchon *et al*. 2014; Milutinović *et al*. 2016; Johnston *et al*. 2014; Duneau *et al*. 2017; Zanchi *et al*. 2017). It includes three immune pathways: immune deficiency (IMD), Toll, and Janus kinase (JAK)-signal transducer and activator of transcription (STAT). These immune pathways are constituted by pattern recognize proteins, signaling components and effectors. Social immunity is a collective immune protection found in social insects, and is thought to operate mainly at the colony level (Cremer *et al*. 2007; Cremer and Sixt 2009; Cotter and Kilner 2010; Cremer *et al*. 2018). With cooperation of individuals in a colony, social immunity includes various types of social behavior, like allogrooming, to prevent infection (Cremer *et al*. 2007; Cremer and Sixt 2009; Cotter and Kilner 2010; Cremer *et al*. 2018). Consequently, individuals in a colony contribute to both levels of immunity. However, individual immunity of different castes in termites remains unclear. Furthermore, how individual immunity of termites evolved during the transition to eusociality is unknown.

In social insects, it has been reported that the expression of some genes, including some immune genes, is caste biased (Scharf *et al*. 2003; Mitaka *et al*. 2016; Jones *et al*. 2017; Mitaka *et al*. 2017a). Caste has been shown to significant impact on the expression of a number of immune genes in *Coptotermes formosanus* (Husseneder and Simms 2014). Therefore, we hypothesized that immune response in termites is differentiated by caste and relative weaker than subsocial wood roaches and solitary cockroaches because of specialized functions of castes in a social colony. According to genomic studies, the canonical insect immune gene families have been shown to be fully represented in termites (Terrapon *et al*. 2014; Korb *et al*. 2015). However, the social bees have instead shown to possess a depauperate immune repertoire (Evans *et al*. 2006), although this contraction in immune genes was later shown to have predated the evolution of eusociality (Barribeau *et al*. 2015). We also predicted that immune gene families would be fully represented and unlinked to transition of eusociality in termites as that in social bees.

In our study, we employed *de novo* transcriptome to study the evolution of individual immunity in termites across a gradient of eusociality. Firstly, we constructed a comprehensive phylogeny of termites and cockroaches based on currently available transcriptomic data sets. Secondly, we predicted the number of members in 47 immune gene families from 18 termite and cockroach species to explore the evolution of the immune system during the eusociality of termites. At last, we compared the immune response of a social termite, *Neotermes castaneus,* a solitary cockroach, *Blatta orientalis* and a subsocial cockroach, *Cryptocercus meridianus*.

3.3 **Materials and Methods**

Insects and microorganisms

Solitary cockroaches, *B. orientalis* and *B. germanica*, were kept at 26 °C, 75% relative humidity with full dark. They were fed with mixed dog food *ad libitum* and supplied with apples and carrots. Two subsocial wood roaches, *C. meridianus* and *C. pudacuoensis*, were collected in China. Larvae and different castes from 9 termite species were extracted from colonies that were kept in the Federal Institute of Materials Research and Testing (BAM), Berlin, Germany. Termite colonies were fed regularly with pre-decayed birch wood or dry grass. Seven species of higher termites were collected from China and Cameroon. The details of sampled insects are listed in Appendix II-A. A Gram-negative bacterium (*Pseudomonas entomophila*, DSM 28517^T), a Gram-positive bacterium (*Bacillus thuringiensis*, DSM 2046^T) and a yeast (*Saccharomyces cerevisiae*, DSM 1333^T) were stored in BAM and cultivated for use in subsequent immune challenge experiments.

Sample collection

P. entomophila and *B. thuringiensis* were grown at 28 °C and 30 °C in nutrient broth, respectively. *S. cerevisiae* were grown for 36 h in universal yeast medium. All cultures were washed twice with Ringers' solution, mixed equal mount to form a cocktail with a final concentration of $5*10^8$ CFU/ml. The suspension was heat-killed at 95 °C for 10 min before injection or pricking.

For *de novo* RNAseq assembly, all experimental insects (except wood roaches collected from China) were frozen in liquid nitrogen immediately after collecting from colony. Regarding species collected from China, they were taken back to laboratory, immersed in RNAlater or frozen in liquid nitrogen. In addition, to stimulate an immune response, experimental cockroach adults were weighed and swabbed with ethanol before injection with the equivalent of $5*10^6$ cells per gram prepared cocktail bacteria. Experimental cockroach larvae and all termites were pricked by using a sterile needle which was contaminated with prepared heat-

killed microbial suspension. Challenged insects (except wood roaches immersed in RNAlater) were frozen in liquid nitrogen at 24 h after challenge. All collected samples were preserved at -70 °C for RNA extraction. Each treatment and group had four replicates.

For quantification of gene expression by RNAseq, wood roaches and three termite castes of *Neotermes castaneus* were weighed and injected with the equivalent of $5*10^6$ cell per gram prepared cocktail bacteria. Each treatment had 16 replicates of each termite caste and 8 replicates for wood roaches. The control groups were injected the equivalent volume of Ringer's solution. After injection, individuals were kept separately under the same condition as mentioned previously. The termites were frozen in liquid nitrogen at 24 h after immune challenge and the wood roaches were immersed in RNAlater before stored in freezer prior to transportation. All sampled insects were preserved in -70 °C until RNA extraction.

Total RNA extraction and de novo transcriptome sequence

Whole insects were used for total RNA isolation. The termites and larvae of cockroach for *de novo* RNAseq assembly were pooled by treatment and caste for RNA extraction. The rest sample were extracted individually. For cockroaches, each Individual was separated into 4-6 parts for RNA extractions before total RNA was pooled together. Samples were suspended in pre-cooled Trizol (Thermo Fisher Scientific), and homogenized with a 5-mm steal bead (Qiagen) using a homogenizer (MP Biomedicals) twice at 2 M/s for 10 s. RNA was isolated according to the manufacturer's instructions with chloroform extraction and isopropanol precipitation, and dissolved in RNA storage solution (Ambion). Subsequently, the total RNA was incubated with 2 units of TurboDNase (Ambion) for 30 min at 37 °C and purified using an RNAeasy Mini kit (Qiagen) according to the manufacturer's instructions. Quantity and quality of RNA were determined by Qubit and Bioanalyzer 2100.

Equal quantities of total RNA from each extraction were pooled together according to species in *de novo* RNAseq assembly. For quantification of gene expression by RNAseq, total RNA from 8 individuals (each termite caste) or 4 individuals (wood roaches) from the same treatment were pooled. The pools of total RNA were used for library preparation. Barcoded cDNA libraries were prepared using a NEXTflex™ Rapid Directional mRNA-seq kit (Bioo Scientific) according to the manufacturer's instructions. Briefly, polyadenylated mRNA was enriched using poly-A beads from total RNA and fragmentated. First and second-strand cDNA were synthesized and barcoded with NEXTflex™ RNA-seq Barcode Adapters. The libraries were sequenced on an Illumina NextSeq500/550 platform at Berlin Center for Genomics in Biodiversity Research (BeGenDiv).

Transcriptome assembly

The raw sequence reads were trimmed and filtered to remove barcodes, adapters, short reads (<25 bp) and reads containing low quality bases using trimmomatic, as incorporated in Trinity (version v2.5.1) (Grabherr *et al*. 2011; Haas *et al*. 2013). The retained reads were assembled by Trinity with default parameters (Kmer length: 25) for annotation and/or differential expression analysis. The assembly completeness was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO v2) with the Arthropod BUSCO set from orthoDB (version 9) (Simão *et al*. 2015). For the phylogenetic analysis, the trimmed reads were further filtered by Botwie2 (Langmead and Salzberg 2012) to remove rRNA and mitochondrial DNA with converted indices built from related sequences of cockroaches, termites and protists from NCBI. For those raw reads of Illumina sequence that were downloaded from SRA database (Appendix II-B), we used the same filter procedures to prepare the assemblies for phylogenetic analysis. For those raw reads of 454 sequence that were downloaded from SRA database were assembled using Newbler v2.7 (454 Life Sciences/ Roche).

Ortholog inference and matrix preparation

For phylogenetic analysis, the assemblies were subjected to ortholog prediction and matrix preparation. To prepare for orthology analysis, each assembly was filtered to retain only the most highly expressed isoforms of each gene. Quantification was performed using Kallisto (Bray *et al*. 2016) and isoforms were filtered using script in Trinity. The redundancy in each assembly was further reduced by CD-HIT-EST (Fu *et al*. 2012) with 95% similarity cut-off. The potential remained rRNA and mitochondrial sequence in assemblies were filtered again using Bowtie2 with the same Bowtie2 indices mentioned previously. Subsequently, the final assemblies were translated into protein by Transdecoder (version 5.0.1) with a minimum length of 60 amino acids. The translated protein sequences were used for ortholog analysis by OrthoFinder (version v2.0.0), which is an all-to-all and gene length balanced method to find ortholog groups and suitable for transcriptome data (Emms and Kelly 2015). For the ortholog analysis, we also included the official gene sets from *Zootermopsis nevadensis* (http://termitegenome.org/) and *Macrotermes natalensis* (http://gigadb.org/dataset/100057).

After ortholog prediction, the single ortholog groups that meet the following criteria were selected for matrix building. To mitigate the taxon representation bias per orthogroup, we selected orthogroups that include at least one representative of each of the following taxa: 1) *Mastotermes*, 2)*Zootermopsis*, 3)Kalotermitidea(*Kalotermes*, *Neotermes*, *Cryptotermes),* 4)*Hodotermposis*, 5)*Coptotermes*, 6)*Reticulitermes*, 7)*Prorhinotermes*. The longest sequence from each selected orthogroup was quired against the ncbi nr database using blast to check for bacterial and protist contamination. Subsequently, these orthogroups were aligned using

MAFFT (Katoh and Standley 2013) with the L-INS-i alignment algorithm. To reduce potential ambiguously aligned positions, each aligned orthogroup was masked by trimAl v1.2 (Capella-Gutiérrez *et al*. 2009) with the gappyout function. Subsequently, orthogroups were concatenated with Phyutility (Smith and Dunn 2008).

Phylogenetic analysis and molecular dating

We employed two different approaches to analyse our matrix: maximum likelihood with RAxML(v8.2.12) (Stamatakis 2014) and Bayesian inference with ExaBayes (v1.4.1) (Aberer *et al*. 2014). In RAxML analysis, 1000 rapid bootstrap replicates were calculated by employing the PROTGAMMAAUTO model. The parsimony random seed (-p) and bootstrap random seed (-x) were set to 12345. In ExaBayes analysis, two runs were performed and each with four chains. The starting seed (-s) was set to 258. Analyses were run until both runs had average standard deviation of split frequencies (asdsf) below 1% for at least 10^6 generations.

To estimate the divergence of time for termites, a molecular clock analysis was performed with PhyloBayes (v4.1) (Lartillot and Philippe 2004). The topology of the phylogenetic tree was constrained to the consensus tree obtained from ExaBayes. An uncorrelated relaxed clock model, uncorrelated gamma multipliers (-ugam), was applied in our analysis under birth death prior (-bd) with soft bounds (-sb). Four independent chains were run with 5 fossil calibration points. To avoid constraining numerous nodes based on the same fossil, each fossil was used to constrain only a single node and no maximum age was set except for the root node. The following age constraints were employed in this study: all cockroaches and Isoptera: 140-311 mya (representing the age of root) (Labandeira 1994), Cryptocercus and Isoptera:137-∞ (Engel *et al*. 2007a), Hodotermitidae and other Isoptera, excluding Mastotermes: 130-∞(Krishna *et al*. 2013), Kalotermitidae and Rhinotermitidae plus Termitidae: 110-∞ (Grimaldi *et al*. 2008), Rhinotermitinae: 44-∞ (Engel *et al*. 2007b). We assessed burn-in, convergence among runs, and run performance by examining parameter files with the program TRACER v1.6.0 (Suchard *et al*. 2018). Each chain was run over 10000 cycles, sampling posterior rates and dates with an initial burning of 20%. Posterior estimation of divergence time was computed from the chain with the highest ESS.

Transcriptome annotation and identification of Immune related proteins

Each assembly (except *Pericapritermes sp.,*due to low completeness) was queried against the NCBI nr database using the DIAMOND implementation of Blastx (Buchfink *et al*. 2015) and taxonomic classification of each query sequence was performed using the Lowest Common Ancestor algorithm. The assemblies were annotated by following the guidelines of Trinotate (https://trinotate.github.io/). The proteins of each assembly were predicted by using

TransDecoder (v5.2.0) (http://transdecoder.github.io) with a minimum length of 60 amino acids. Homology searches, predictions and domain identifications were performed locally and subsequently integrated into SQLite database at an e-value threshold of 1e-03. Briefly, assembled nucleotide and corresponding peptide sequences predicted by TransDecoder were used to query SwissProt with Blastx and Blastp, respectively. Protein domains, signal peptides, and transmembrane domains were determined by HMMER (v3.1b2)(Finn *et al*. 2011), SignalP v4.0(Petersen *et al*. 2011), and TmHMM v2.0(Krogh *et al*. 2001), respectively.

Immune related proteins were identified by searching predicted proteins for the presence of immune function containing domains and annotations from Trinotate. To quantify the presence of domains containing putative immune function, we first downloaded immune gene families from 31 species (available on https://github.com/ShulinHe/Blatta_orientalis) as well as Termicin and insect transferrins from Uniprot and then constructed a set of HMM profiles based on alignments of all protein families. The complete set of predicted proteins from each transcriptome were searched for matches to predict immune-related HMMs using HMMER. Afterwards, the HMMER output was filtered by: excluding targets with E-values > 0.001 for the best domain, excluding targets with overall E-value greater than 10-5, and assigning the targets that have multiple HMMs to best e-value HMM. The genes that have multiple immune predicted proteins from different isoforms was assigned to the protein that has the highest overall E-value HMM. The filtered HMMER output were then further selected using annotations from Trinotate. Putative gene targets were selected when the HMMER output of their predicted proteins fitted their annotations of Blastp, Blastx or Pfam in Trinotate. Subsequently, targets were removed when their predicted proteins were shorter than 100 amino acids in families other than antimicrobial peptides. We adopted a conservative approach for accepting the identity of immune gene target. Firstly, because it is theoretically possible that different components from the same subcluster may represent spliced isoforms of a single gene, we aligned nucleotide sequences and corresponding predicted proteins from each subcluster against one other using MAFFT (Katoh and Standley 2013) and excluded sequences that were variable in length but otherwise identical. Secondly, to account for different fragments of the same gene potentially appearing in different subclusters of a single cluster (and being erroneously described as two separate genes), we ran an additional blastx search on all putative subcluster sequences. If more than one subcluster had an identical target in the top 10 entries of a DIAMOND Blastx search (and overlapped by less than 9 amino acids – a value determined by the use of a 25 k-mer parameter during transcriptome assembly), only the longest subcluster was retained.

Immune gene family analysis

Based on the dated phylogeny, the expansion and contraction of immune gene families was predicted by CAFE 4.0 (-p 0.05) (De Bie *et al*. 2006), which is based on protein family size and topology of a phylogenetic tree. The annotated immune proteins of *Z. nevadensis* (Terrapon *et al*. 2014) were used for estimation of error model (-diff 5) as true dataset and the immune proteins from this study were inferred as prune dataset. Subsequently, the estimated error model was applied to all of the species in the whole dataset. The model of birth and death rate (lambda) was estimated with two different parameters in cockroaches and termites, respectively. The significance of the chosen model was determined by genfamily and lhtest commands in CAFE.

Transcript Abundance Estimation and Differential Expression Analysis

Transcript expression after immune challenge in *C. meridianus* and different *N. castaneus* castes was estimated using Kallisto (Bray *et al*. 2016). Differential gene expression was analysed using the R package DESeq2 (Love *et al*. 2014) with remove of the potential transcripts from symbionts, including protist and bacteria from taxonomy classification. In this study, we considered the genes as significantly differential expressed when fold change > 2 and adjusted p-value < 0.05. The differential expressed genes were subject to Gene Ontology (GO) enrichment analysis by the R package GOseq with an adjusted p-value cut-off at 0.05. The GOs were extracted from the Trinotate annotation. After GO enrichment analysis, the redundancy of enriched GOs was reduced by using REVIGO (Supek *et al*. 2011).

To compare the immune response in different castes, the number of differentially expressed genes in each immune protein family was estimated according to different castes in *N. castaneus*. Furthermore, the number of significant differentially expressed immune related genes was also compared between different castes with *C. meridianus* and *B. orientalis* in order to explore the relation of evolution of immune response and eusociality.

Quantitative PCR

Total RNA from each individual for sequencing was used for quantitative PCR. cDNA was synthesized with M-MLV Reverse Transcriptase (Promega) using Random (Promega) and Oligo(dT)15 Primer (Promega) according to manufacturer's instructions. The genes and primer sequences used for quantitative PCR are listed in Appendix II-D. Relative expression of these genes was determined using SensiFAST™ SYBR Lo-ROX Kit (Bioline) following three-step cycling. A standard curve of pooled, five-times serially diluted cDNA was run for the chosen genes. RPL22 (ribosomal protein 22) and RPL24 (ribosomal protein 24) were used as reference genes for *N. castaneus* and *C. meridianus*, respectively. Fold-change calculations

were performed by using the Pfaffl method (Pfaffl 2001) and a Mann–Whitney U test was employed to compare gene expression between treatment and control groups using R v.3.2.3 (Team 2016). Data are presented as means ±SE.

3.4 Results

3.4.1 Transcriptome and annotation statistics

In this study, we sequenced 15 termite transcriptomes, 2 *Cryptocercus* transcriptomes, and other 2 cockroach transcriptomes. After quality trimming, each library retained 98.92%- 99.83% of total reads survived for following assembling. Each assembly per species has 0.12- 0.21 million transcripts with 82.7%-97.7% complete BUSCOs (except 69.0% of completeness in *Pericapritermes sp.*, which only was used for phylogeny analysis) (Table 3.1).

Table 3.1 Details of sequenced species and corresponding assemblies in this study

Specie name	**Library Size** (No. of reads [Million])	**No. of assembled transcripts**	**BUSCO** (orthodb v9, insect, n=1066)
Blattella germanica	33.3	169296	C:91.6%[S:65.2%,D:26.4%],F:7.0%,M:1.4%
Blatta orientalis	30.2	177500	C:82.8%[S:59.6%,D:23.2%],F:13.4%,M:3.8%
Cryptocercus meridianus	32.3	142716	C:90.5%[S:56.4%,D:34.1%],F:7.5%,M:2.0%
Cryptocercus pudacoensis	30.4	117983	C:83.3%[S:50.8%,D:32.5%],F:13.1%,M:3.6%
Mastotermes darwiniensis	36.6	200400	C:89.5%[S:55.6%,D:33.9%],F:8.3%,M:2.2%
Neotermes castaneus	40.3	214244	C:97.0%[S:46.7%,D:50.3%],F:2.4%,M:0.6%
Kalotermes flavicollis	39.0	180046	C:96.9%[S:48.6%,D:48.3%],F:2.6%,M:0.5%
Zootermopsis nevadensis	42.4	196687	C:94.5%[S:47.2%,D:47.3%],F:5.1%,M:0.4%
Cryptotermes brevis	30.5	175760	C:86.2%[S:55.6%,D:30.6%],F:10.4%,M:3.4%
Coptotermes formosanus	22.3	141751	C:84.5%[S:53.3%,D:31.2%],F:10.9%,M:4.6%
Reticulitermes flavipes	32.9	168192	C:97.7%[S:50.6%,D:47.1%],F:1.7%,M:0.6%
Prorhinotermes inopiinatus	28.7	189751	C:86.0%[S:51.3%,D:34.7%],F:10.5%,M:3.5%
Macrotermes subhyalinus	33.7	137016	C:84.1%[S:53.8%,D:30.3%],F:11.3%,M:4.6%
Pericapritermes sp.	21.9	122403	C:69.0%[S:51.6%,D:17.4%],F:20.9%,M:10.1%
Indotermes sp.	27.6	136912	C:82.7%[S:58.8%,D:23.9%],F:12.0%,M:5.3%
Dicuspiditermes sp.	26.5	165729	C:89.7%[S:57.0%,D:32.7%],F:7.2%,M:3.1%
Globitermes sp.	23.2	146581	C:83.0%[S:52.5%,D:30.5%],F:12.7%,M:4.3%
Bulbitermes sp.	28.6	154438	C:87.5%[S:53.4%,D:34.1%],F:9.4%,M:3.1%
Promirotermes sp.	36.6	149335	C:86.4%[S:49.2%,D:37.2%],F:9.9%,M:3.7%

Note: C, complete BUSCOs; S, complete and single-copy BUSCOs; D, complete and duplicated BUSCOs; F, fragmented BUSCOs; M, missing BUSCOs

3.4.2 Phylogenetic analysis

In order to construct a comprehensive phylogeny of termites, we analyzed 35 transcriptomes and genomes, of which 2 termite genomes and 14 available raw data sets were used. Five families (Mastotermitidae, Archotermopsidae, Kalotermitidae, Rhinotermitidae and Termitidae) of termites have been covered and two cockroach family (Blaberidae, Ectobiidae) were used as outgroup. An amino acid data matrix with an average of 85.86% gene occupancy per species was assembled from predicted orthogroups. The resulting matrix comprises 118 orthogroups with 18230 amino acid positions and 13.16% missing data.

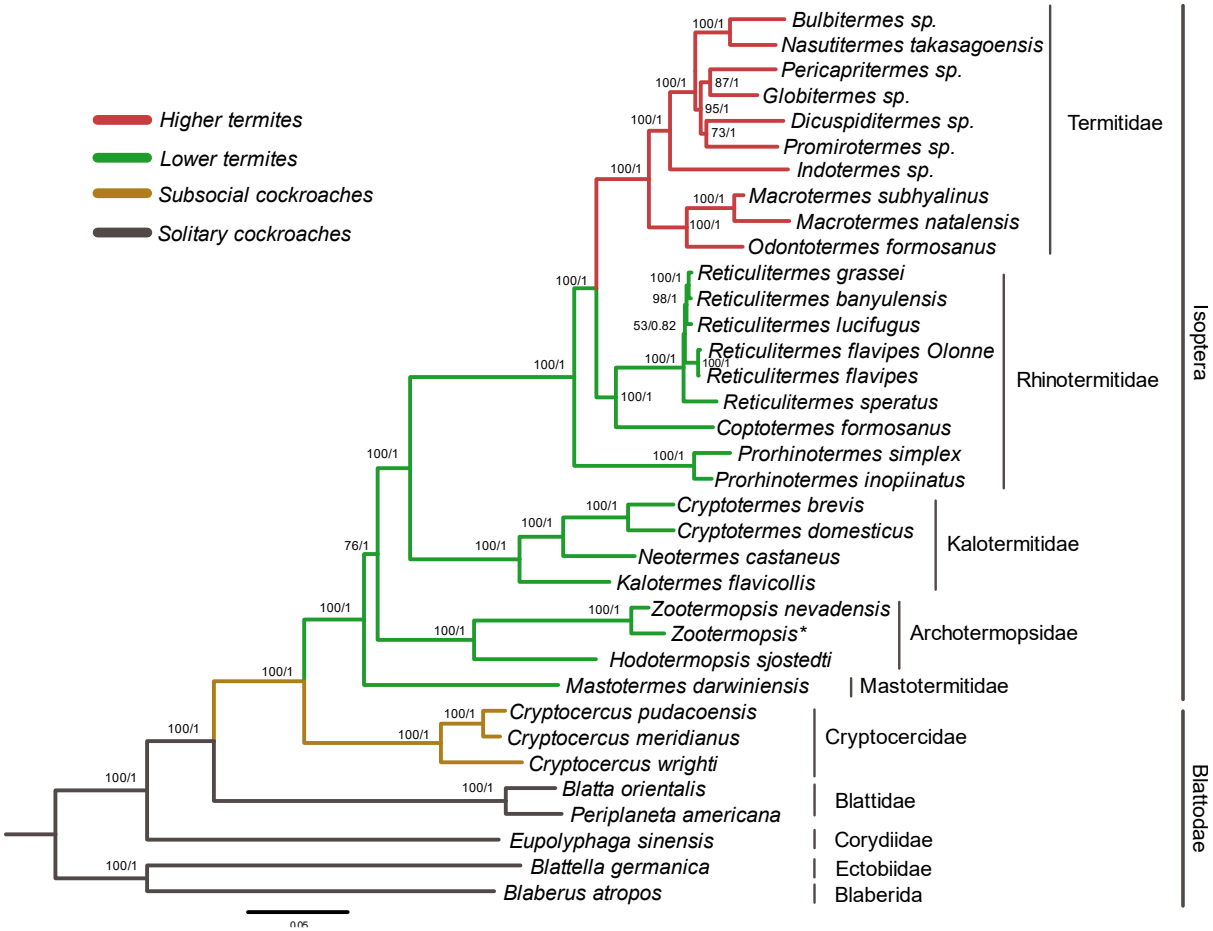


Figure 3.1 Phylogeny of termites based on RAxML and Exabayes. The number on each node represents support of boostrap values from RAxML/likelihood score from Exabayes. Different colors of lines indicate traditional classification of termites and cockroaches. *Zootermopsis*: Zootermopsis nevadensis nuttingi*.

The phylogenetic trees obtained from two different methods, RAxML and ExaBayes, have identical topologies (Figure 3.1). Cryptocercidae and Isoptera are sister groups and form a clade that is close related to Blattidae. Mastotermitidae is the basal family of termites and a sister group to all the others. Archotermopsidae is located between Mastotermitidae and Kalotermitidae. Kalotermitidae is a monophyletic group in the phylogeny. Rhinotermitidae is a paraphyletic group, comprised of the monophyletic Rhinotermitinea and Prorhinotermitinae. The Rhinotermitinea is comprised of *Coptotermes* and *Reticulitermes*. Termitidae is monophyletic and a sister group to Rhinotermitinae.

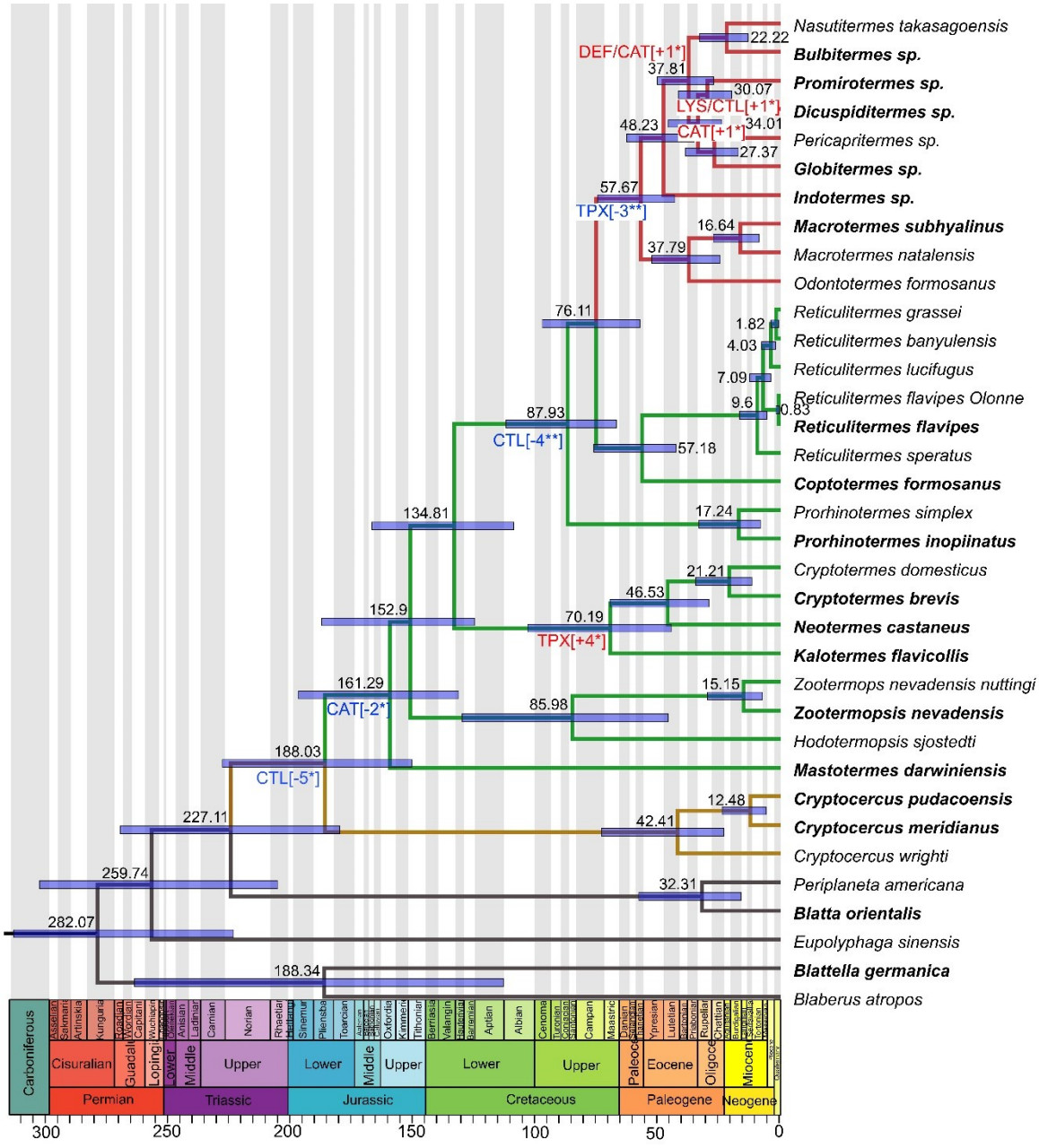


Figure 3.2 The fossil calibrated phylogenetic tree of termites from Phylobayes. The age of nodes is indicated with 95% confidence interval. The bold marked species were newly sequenced in this study and used for immune gene evolutionary analysis. The contraction and expansion of immune gene families of nodes were indicated in blue and red text, respectively. The number and * in [ ] indicated the number of change in that gene family and significance level (*: 0.05 and **: 0.01).

As illustrated in the time calibrated phylogenetic tree (Figure 3.2), most recent common ancestors (MRCA) of *Cryptocercus* and termites can be dated to the lower Jurassic, 188.785± 20.2835 (152.798-229.182, 95% confidence interval (CI)) million years ago (mya) and diverged from Blattidae in the upper Triassic, around 228.054±23.4771(182.986-272.735, 95% CI) mya. As the origin of sociality in termites, the root of termites is estimated to be 161.83±17.5812(132.681-199.622,95% CI) million years old from the upper Jurassic. The root

of higher termites, Termitidea, is estimated to be around 57.7964±8.20891(43.4321-75.9709, 95%CI) million years old from the upper Paleocene and diverged from lower termites around 76.5212±10.4448(58.7171-99.541, 95%CI) mya in upper Cretaceous.

3.4.3 Expansion and contraction of immune gene families

Immune related genes from 47 families were categorized as either receptor, effector or signaling component. Using a combined identification of hmmsearch and trinotate annotation, except a family of effector, drosomycin, that was lost in termites and wood roaches, all other gene families were represented in both cockroaches and termites (Figure 3.3).

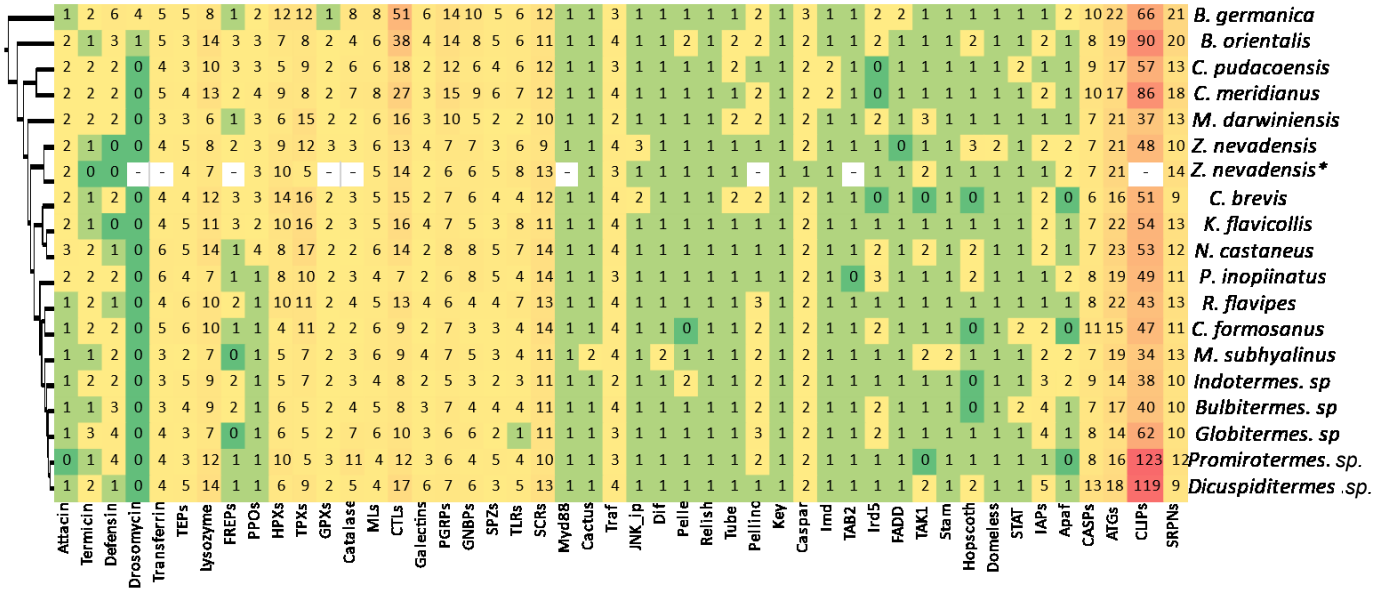


Figure 3.3 Predicted gene numbers in 47 immune gene families from 18 termite and cockroach species. *: the gene number of immune gene families from previous study (Terrapon *et al*. 2014). Blank represent not reported in previous study.

After applying an error estimation, we found the global evolution rate of immune gene families in cockroaches (birth/death rate[lambda]=0.0035) is lower than that of termites (lamda=0.0057). Different components of immune related genes have different evolutionary rates. In cockroaches, the evolutionary rate (lamda=0.0007) of effectors is much lower than that of signaling components and receptors, which is close to the global rate. However, three components have strikingly different evolutionary rates in termites. The signaling molecules have the highest evolutionary rate (lambda=0.0062). The evolutionary rates in effectors (lambda=0.0012) and effectors (lambda=0.0018) are close.

In effectors, we found that the thioredoxin peroxidase (TPX) gene family has undergone expansion in the root of monophyletic Kalotermitidae, while it had a contraction in the root of Termitidea. In addition, we found a contraction of catalase (CAT) in MRCA of all termites. Apart from that, CAT, lysozyme (LYS) and defensin also showed expansion in some nodes of higher termites (Figure 3.2). In the receptors, we found that C-type lectin (CTL) show

contraction during the evolution of social termites (Figure 3.2). It showed contractions in MRCA of subsocial wood roaches and social termites as well as in MRCA of Rhinotermitidae and Termitidae. We did not detect rapid change of signal molecules during the eusociality of termites.

3.4.4 Immune response in termite castes and cockroaches

In order to characterize immunity in termite castes, we compared immune responses of three castes from *N. castaneus*. After immune challenge, there were 67 genes significantly upregulated in workers, 219 in soldiers, and 477 in reproductive. There were 215 genes significantly downregulated in workers, 196 in soldiers and 760 in reproductive (Figure 3.4A/Figure 3.4B). Following gene ontology (GO) analysis, we observed a high number of enriched immune related GO terms from upregulated genes of soldiers (Figure 3.4C, Appendix II-E). In contrast, fewer enriched immune related GO terms was found in workers and reproductives (Figure 3.4C).

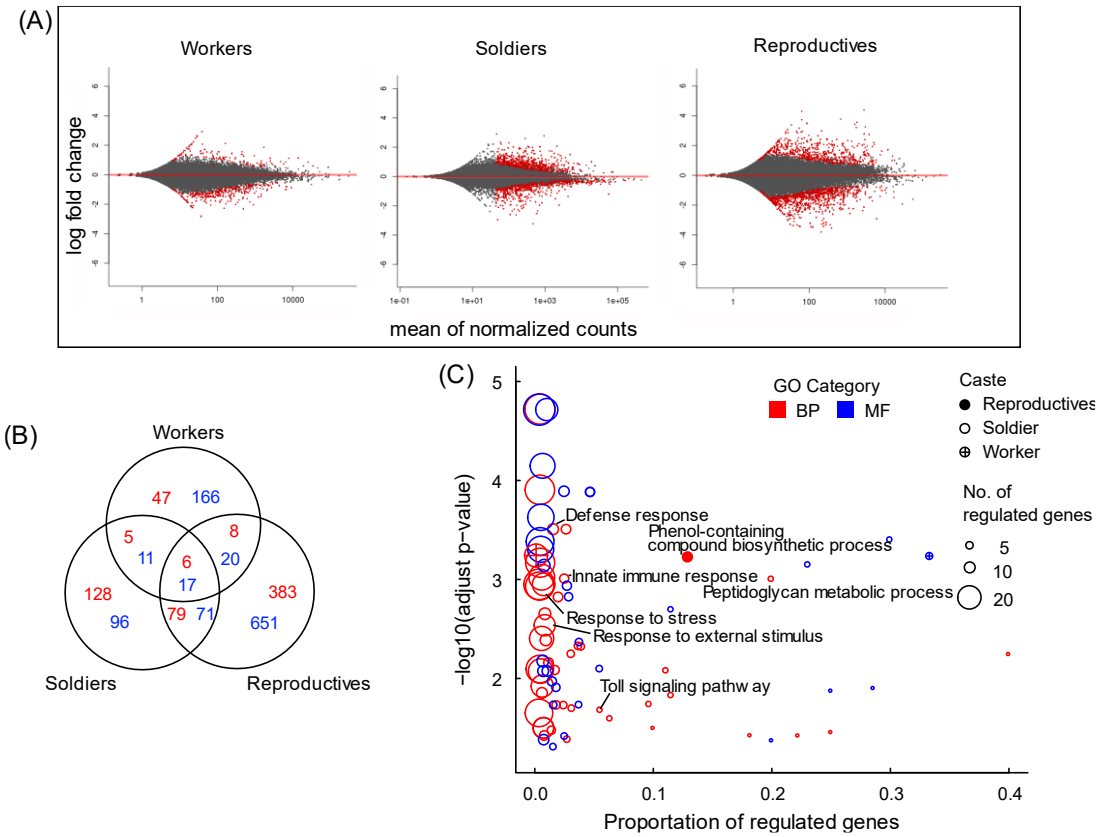


Figure 3.4 A) Ratio-average plot of gene expression in different castes. Red indicates differentially expressed genes. B) The number of significant differentially expressed genes after immune challenge in different castes. Red: upregulated, blue: downregulated. C) The significant enriched GO terms in categories of Biological process (BP) and Molecular Function (MF) from significant upregulated genes in treatment of different castes. Enriched GO terms were filtered by adjust *p*-value (<=0.05) and redundancy was reduced by REVIGO.

Expressions of immune related genes are categorized by castes according to the result of principle component analysis (Figure 3.5A) and the reproductives clearly had the highest expression of these genes (Figure 3.5B). After immune challenge, 5 immune related genes

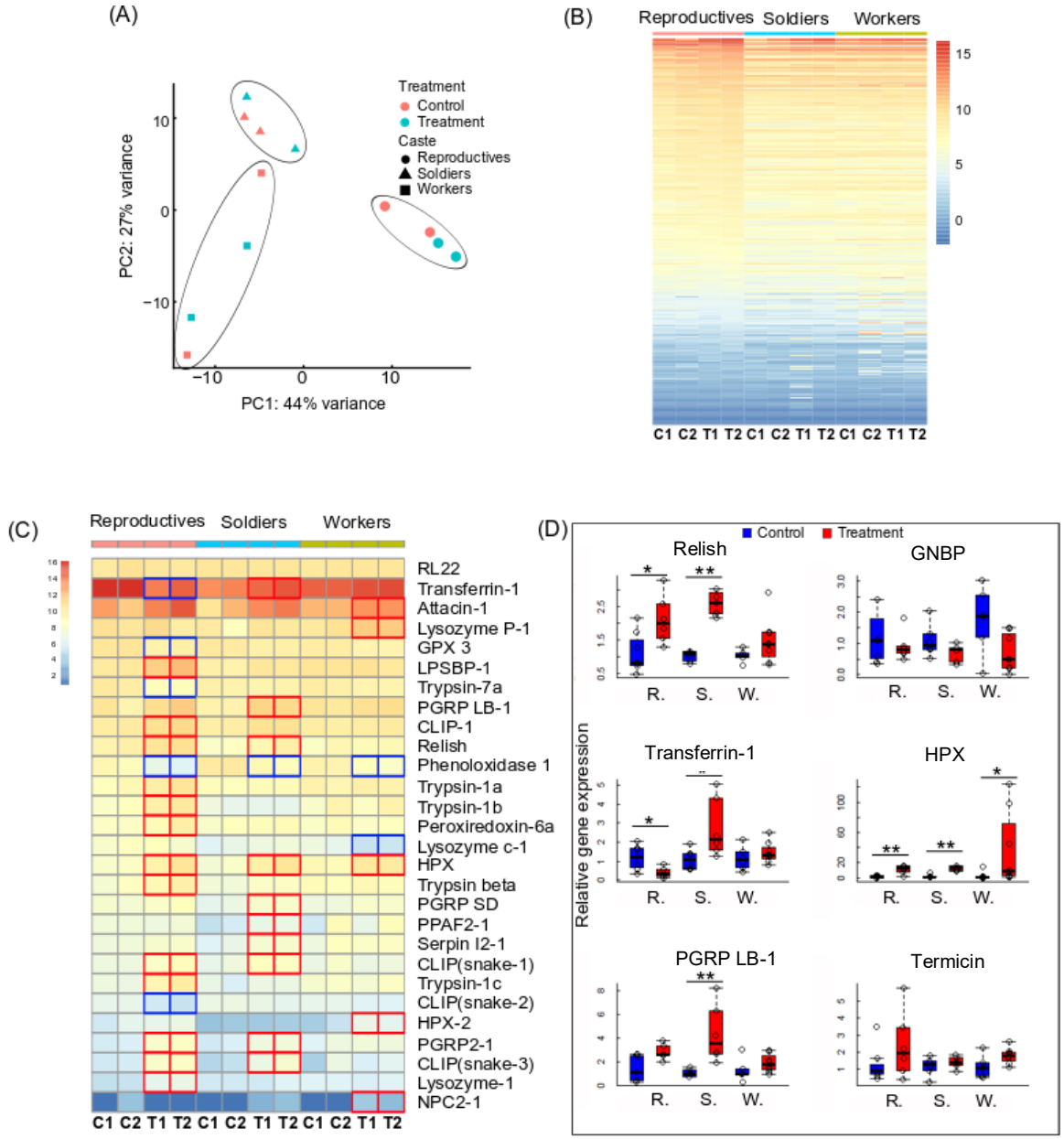


Figure 3.5 A) Principle component analyses of immune related genes from different castes of *N. castaneus*, red: control and light blue: treatment. B) The heatmap of expressed immune related genes in different castes. C) The significant differentially regulated immune related genes after immune challenge in different castes compared to control group. C: Control, T: Treatment. Red square: upregulated, blue square: downregulated. D) The qPCR of six immune related genes in different castes (each treatment and group has 6-8 individuals with two replicated of each). R.:Reproductives, S.: Soldiers, W.: workers. Significance level comparisons: **, $p<0.001$; *, $p<0.05$; NS, not significant. GNBP and termicin were not differentially expressed in the DESeq2 analysis (and qPCR) and so are not represented in panel c.

were significantly upregulated in workers, 10 in soldiers and 13 in reproductives (Figure 3.5C). The differential expression of part of these genes was confirmed by qPCR (Figure 3.5D).

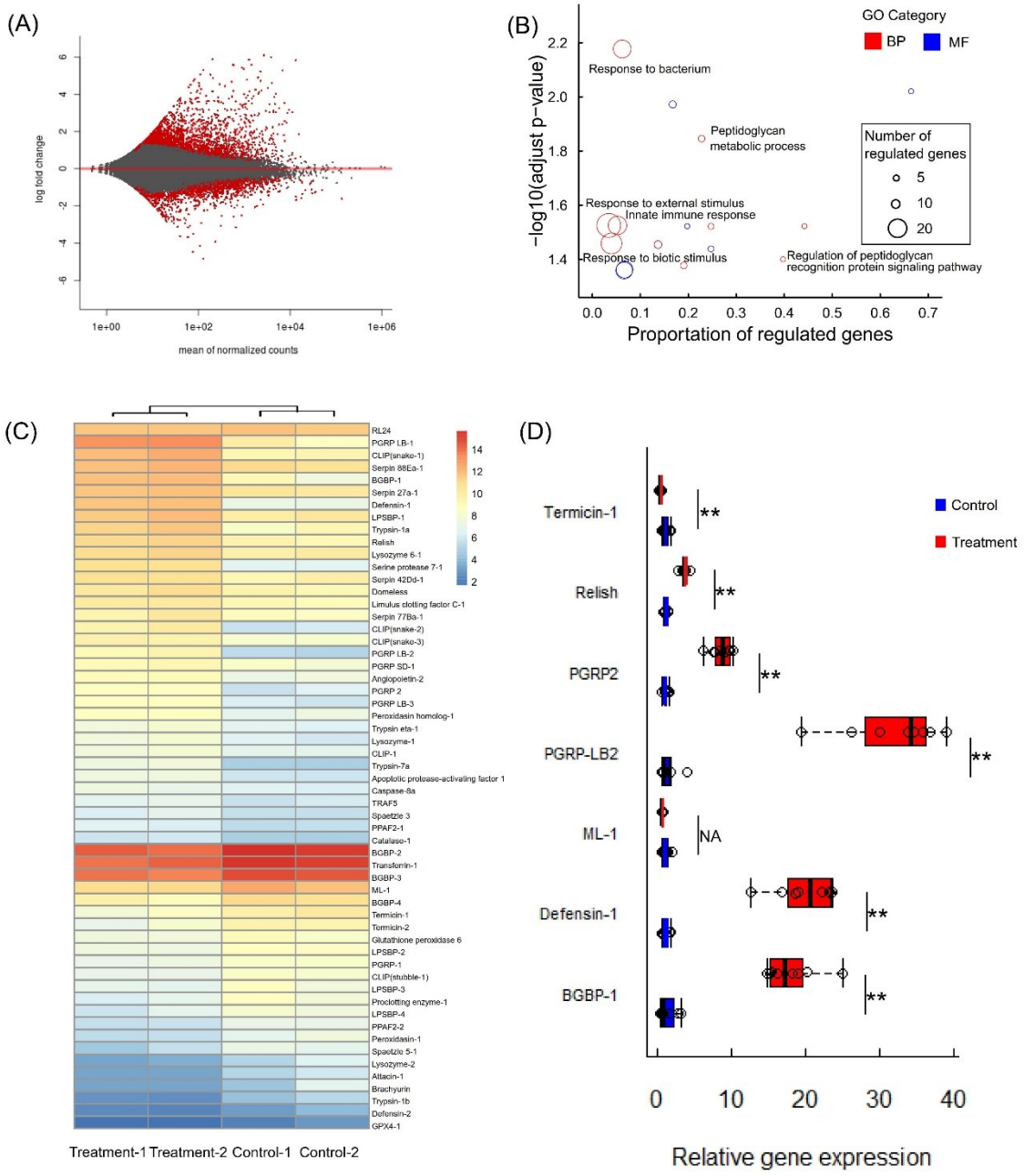


Figure 3.6 A) Ratio-average plot of expression genes in *C. meridianus*. red indicates differential expressed genes. B) The significant enriched GO terms in categories of Biological process (BP) and Molecular Function (MF) from significant upregulated genes in treatment. Enriched GO terms were filtered by adjusted p-value (<=0.05) and redundancy was reduced using REVIGO. C) Significant differentially regulated immune related genes after immune challenge in treatment compared to control group. D) Expression of seven immune related genes (each treatment and group have 6-8 individuals with two technical replicates of each) as measured by qPCR. Significance level comparisons: **, p<0.001; *, p<0.05; NS, not significant. ML-1 was not differentially expressed in the DESeq2 analysis (and qPCR) and so is not represented in panel c.

After immune challenge with heat-killed bacteria, 800 and 1507 genes were significantly downregulated and upregulated in the subsocial cockroach *C. meridianus*, respectively. The upregulated genes represent a robust immune response indicated by enriched immune related GO terms (Appendix II-E). In these significantly regulated genes, there are 34 upregulated and 23 downregulated immune related genes (Figure 3.6).

To explore the relationship between immune response and division of labour in termites, we quantified the number of immune-related genes which were differentially expressed in response to a common immune challenge in three termite castes, a subsocial cockroach and a solitary cockroach. We observed that the immune response of the two cockroach species is similarly broad with differential expression of receptors, signalling components, and effectors. Termite reproductives and soldiers displayed a similar but relatively weaker pattern of immune gene expression after challenge whereas differential expression in workers was limited to the effectors attacin, lysozyme and peroxidase as well as the ML receptor family (Figure 3.7).

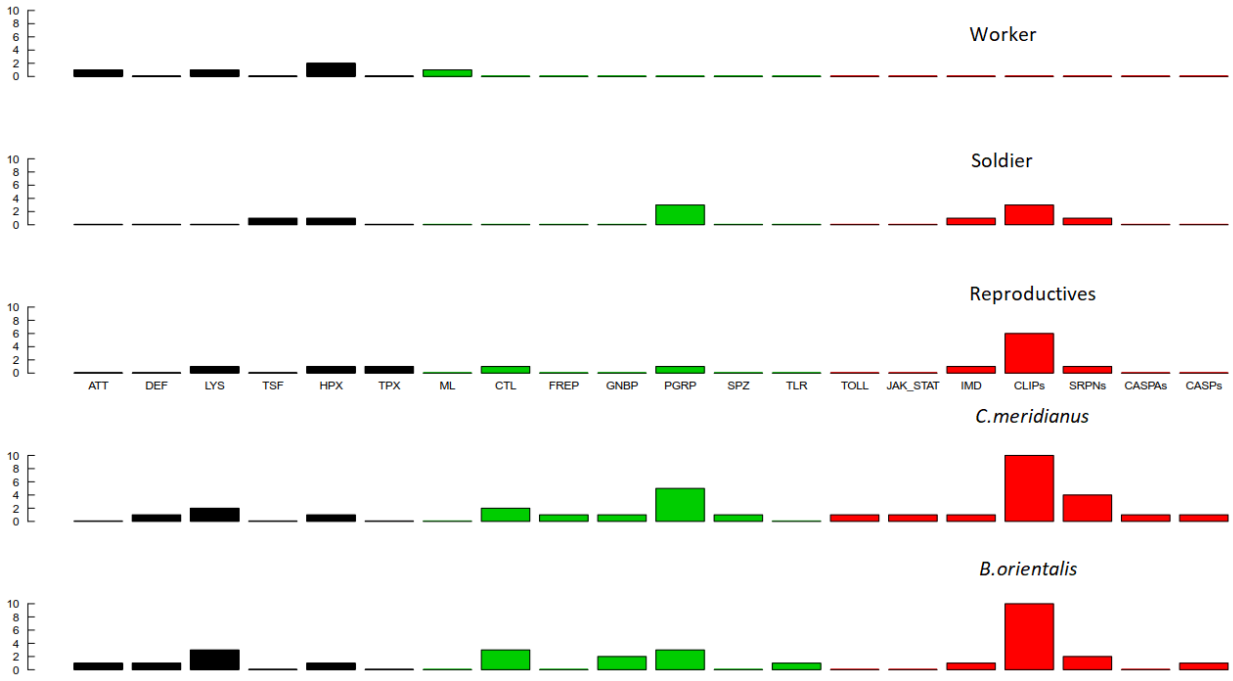


Figure 3.7 The number of significantly differentially expressed immune related genes in each gene families and three immune pathways (IMD, TOLL, JAK-STAT) of three castes (worker, soldier, reproductive in *N. castaneus*), *C. meridianus* and *B. orientalis*. Black: effectors, Green: receptors, Red: signling components.

3.5 **Discussion**

In this study, it is the first time that a number of transcriptomes from termites and cockroaches have been sequenced, especially for the difficulty to sample uncommon subsocial *Cryptocercus*. Firstly, a phylogenetic analysis of termites and cockroaches was performed

based on available transcriptomic data sets. It confirms the phylogenetic location of termites and shows that the root of termites can be dated to the lower Jurassic. Secondly, to characterize the immune systems in termites, we identified immune related genes of 47 families in termite and cockroach species followed by detecting the contraction and expansion of gene family during the evolution of eusociality. It shows that gene families of catalase, thioredoxin peroxidase and C-type lectins have undergone significant contractions during the origin and subsequent diversification of the major termite lineages. Subsequently, we compared the immune response of termite castes and cockroaches. We found different immune responses in termite castes which probably are related to division of labour, but also may reflect variation in the allocation of resources to individual immune defences among the sterile and non-sterile caste and potentially between immature and terminal stages of development.

As social insects, reproductive division of labour, especially the appearance of sterile caste, is a main character in termites. After immune challenge, we find a weaker immune response in workers, but a comparatively broader immune response in reproductives and soldiers. The observed weak immune response in workers may reflect a trade-off in individual immune system as they are the most expendable component of a colony's overall fitness. Workers are also responsible for the majority of daily tasks in a colony including social immunity (Rosengaus *et al*. 1998b), and individual immunity may receive comparatively lower investment by comparison. But, it is also possible that workers in lower termites don't have fully developed immune systems because they represent an immature stage, unlike reproductives and soldiers, which are terminal developmental stages (Korb and Hartfelder 2008). In contrast, a relatively robust induced immune response in soldiers may indicate the capacity of multiple defence roles in termite colony in addition to physical defence, which has been suggested in *Reticulitermes speratus* (Fuller 2007; Mitaka *et al*. 2017b). The relatively high colony-level cost of producing and maintaining soldiers may also contribute the consequence. Interestingly, a high overall expression of immune related genes in reproductives has been found despite potential trade-offs with reproduction (Calleri *et al*. 2007). Overall, different pattern of upregulated immune gene families and different enriched GO terms after immune challenge, as well as different expressions of total immune genes indicate that immune responses and immune investments are shaped by caste. This reflects a modulation of the individual immune system in insect societies following evolution of division of labour.

To characterize the change of immune system in the evolution of termites, a phylogenetic analysis in termites was performed based on available transcriptomes. The topology of the phylogenetic tree in this study is in line with previous studies that are based

on nuclear/mitochondrial gene markers or mitochondrial genome (Inward *et al*. 2007a; Inward *et al*. 2007b; Legendre *et al*. 2008; Engel *et al*. 2009; Cameron *et al*. 2012; Bourguignon *et al*. 2015). As has been shown, the sister groups of *Cryptocercus* and termites has been recognized (Inward *et al*. 2007a; Inward *et al*. 2007b; Legendre *et al*. 2008; Engel *et al*. 2009; Cameron *et al*. 2012; Bourguignon *et al*. 2015; Che *et al*. 2016; Bourguignon *et al*. 2017). The divergence of termites and *Cryptocercus* could be dated to the lower Jurassic, which is older than the origin of eusocial ants from the middle Jurassic (Moreau *et al*. 2006). In addition to the overlap of confidence interval, the estimated ages in this study are generally older than that in mitochondrial genome or phenotypic data (Engel *et al*. 2009; Bourguignon *et al*. 2015) but similar to a multiple-fossils calibration analysis (Ware *et al*. 2010).

Subsocial wood roaches are crucial to understanding the evolution of termites due to their evolutionary position (Klass *et al*. 2008). We compared the immune response of termite castes to a subsocial cockroach and a solitary cockroach. In terms of upregulated immune genes, soldiers but particularly reproductives showed similar patterns in inducing a relatively broad immune response compared to subsocial and solitary cockroaches. More studies of immune response of termite that possess true workers are needed to further understand this relationship. In addition, a similar pattern of response in cockroaches indicated that the transition from solitary to subsocial system did not significantly affect individual induced immunity, which is interesting since it also is detected that the contraction of certain immune gene family predated the divergence of *Cryptocercus* and termites. This raises the possibility that changes to the environment, diet, or even the gut microbiota were important drivers of immune gene contractions in the ancestor of termite and *Cryptocercus*.

However, both solitary and social bees have been reported to possess a depauperate immune repertoire (Barribeau *et al*. 2015), indicating a possible difference in the evolutionary route of immunity in bees and termites. For example, it has been demonstrated that rapid evolution of immune proteins in ants and bees (Viljakainen *et al*. 2009) may be due to relaxed selection constraint due to the evolution of eusociality (Harpur and Zayed 2013). However, it seems to be complicated in termites as an expansion of gene families in some clades of termites was also detected in my study. Furthermore, strong evidence exists to support expansion of genes in cockroaches compared to other non-social insects (Harrison *et al*. 2018a; Li *et al*. 2018), which would indicate the possible rapid expansion of genes in the ancestor of cockroach (Harrison *et al*. 2018a) followed by a partial reduction in termites.

A higher gain and loss rate of immune related gene families in termites does indicate that the appearance of a sterile caste system may have influenced the evolution of immune genes, especially in immune receptors and effectors. The evolutionary rate of signaling

components is lower than that of effectors and receptors suggesting the selection force of the former is not as strong as that of the latter groups that directly come into contact with microbes.

In these rapid changes of immune gene families, drosomycin had been lost in subsocial wood roaches and eusocial termites. The drosomycin was first identified in *Drosophila* as an antifungal peptide (Zhang and Zhu 2009). It is unclear whether this loss is caused by environmental change or the appearance of eusociality. But it is possible that novel pleiotropic antifungal functions of other molecules, such as GNBP2 (Bulmer *et al*. 2009; Bulmer *et al*. 2012), or synergistic function formed during this change (Velenovsky *et al*. 2016), eliminating the need for this additional antimicrobial peptide. Additionally, catalase, which repairs or prevents cell damage caused by oxidative stress (Finkel and Holbrook 2000), has undergone a contraction in the MRCA of termites followed by a re-expansion in some higher termite lineages. A contraction of TPX, a type of peroxidase known as peroxiredoxins (Radyuk *et al*. 2001), was also found in the MCRA of higher termites. Conversely, an expansion of this gene family was observed in the MRCA of Kalotermitidae. In addition to the expansion of antioxidants in cockroach (Harrison *et al*. 2018a), the rapid changes of these immune gene families indicate a particularly strong evolutionary correlation between antioxidant systems and termite eusociality or ecology. This could also be the reason for contraction of the C-type lectin gene family in the MRCA of *Cryptocercus* and termites as well as in the MRCA of Rhinotermitidae and Termitidae. The contraction of immune gene families during this transition could also possibly be an adaptation as a counterpart to social immunity, which has also been suggested in bees and ants (Harpur and Zayed 2013). These findings further indicate that the transition to sociality significantly shape the evolution of the termite immune system, in contrast to bees (Barribeau *et al*. 2015) and our previous hypothesis. This difference could be as a consequence of the different evolution paths of social system or as a consequence of major shift in the different living environment which were richly antigenic in cockroach ancestors, which have expanded set of some immune genes families (Chapter I)(Harrison *et al*. 2018a; Li *et al*. 2018).

In conclusion, we constructed a phylogenomic tree of termites and found the evolution of eusociality in termites could be dated to the lower Jurassic. In addition, it revealed different immune responses in termite castes, which could be the consequence of division of labour in termites. Furthermore, we found contraction of immune gene families during the evolution of termites, particularly in effectors and receptors. These indicate that the molecular immune system underwent significant modifications during the termite evolution.

3.6 **Acknowledgements**

We thank Michael T. Monaghan for advice on phylogenetic analysis, Susan Mbedi and Sarah Sparmann (BeGenDiv) for technical support in transcriptome sequencing, and Yvonne de Laval for providing help in collecting termites. The High-Performance-Computing Facilities at ZEDAT, FU-Berlin are also acknowledged for providing access to computing resources.

Chapter III

Termite soldiers contribute to social immunity by synthesizing potent oral secretions

This chapter was published in: *Insect Molecular Biology* (2018) 27(5), 564–576.

4.1 **Abstract**

The importance of soldiers to termite society defense has long been recognized, but the contribution of soldiers to other societal functions, such as colony immunity, is less well understood. We explore this issue by examining the role of soldiers in protecting nestmates against pathogen infection. Even though they are unable to engage in grooming behavior, we find that the presence of soldiers of the Darwin termite, *Mastotermes darwiniensis*, significantly improves the survival of nestmates following entomopathogenic infection. We also show that the copious exocrine oral secretions produced by Darwin termite soldiers contain a high concentration of proteins involved in digestion, chemical biosynthesis, and immunity. The oral secretions produced by soldiers are sufficient to protect nestmates against infection, and they have potent inhibitory activity against a broad spectrum of microbes. Our findings support the view that soldiers may play an important role in colony immunity, and broaden our understanding of the possible function of soldiers during the origin of soldier-first societies.

Keywords: external; social; immunity; soldier; antimicrobial; proteome

Summary

Alongside sexual reproduction and multicellularity, eusociality is considered one of the major transitions in evolution (Szathmary and Smith 1995). Eusociality has evolved most often among the insects, particularly the Hymenoptera (the ants, bees and wasps) and termites. The hallmark of social evolution in insects is the appearance of permanently sterile castes, which is reflected by reproductive division of labour. A notable feature of insect societies is the emergence of sophisticated immune adaptations at the individual and group level to control the spread of disease. However, the evolution of termite immunity remains poorly understood. In particular, information regarding molecular evolution of the canonical immune pathways, and how innate and induced immunity were shaped by the evolution of a sterile caste system, remain major gaps in knowledge.

A comparative approach in the study of the evolution of termite immunity requires robust knowledge of the immune system of the nearest non-social insect lineages: the cockroaches. To this end, the immunity of a cockroach, *Blatta orientalis*, was explored in **Chapter I**. Using *de novo* transcriptomes, a full repertoire of immune gene members was identified. Interestingly, expansions of immune gene families of receptors, including GNBP, PGRP and hemolymph LPS-binding protein (LPSBP) were identified. After immune challenging cockroaches with a mixture of heat-killed microbes (*Bacillus thuringiensis, Pseudomonas entomophila, Saccharomyces cerevisiae*), I was able to record a broad induced response in canonical immune pathways, pointing to the presence of effective and potentially long-lasting protection against infection, which is a key trait for organisms that thrive in a rich antigenic environment.

In the first part of **Chapter II**, I examined the evolution of immunity in termites by first reconstructing a termite phylogeny with 19 newly sequenced transcriptomes and 16 available genomic datasets. As a result, we confirmed termites as the sister group to the *Cryptocercus*, a subsocial cockroach genus, and located their most recent common ancestor (MRCA) to the lower Jurassic. An evolutionary analysis of immune related gene families was then performed based on 18 of the newly sequenced transcriptomes. A family of antimicrobial peptide, Drosomycin, was found to be lost in the ancestor to the subsocial wood roaches and all termites. A further analysis of two other classic effectors, catalase and thioredoxin peroxidase, revealed a rapid contraction of the former in the ancestor to all eusocial termite species and a rapid contraction of the latter in the root of Termitidae. In addition, a family of receptors, C-type lectins (CTLs), showed contraction in the MRCA of *Cryptocercus* and termites as well as in the root of the Rhinotermitidae. In addition, these contracted gene families underwent a subsequent re-expansion in some individual higher termite lineages. These results suggest a substantial re-modelling of the termite immune system during the evolution of eusociality.

This qualitative analysis focusing on major shifts in termite immunity was followed in the second part of **Chapter II** by a quantitative analysis of individual immunity across different castes of a representative lower termite, *Neotermes castaneus*. Gene expression changes were then compared with a subsocial wood roach, *Cryptocercus meridianus*, and the solitary cockroach, *B. orientalis*. Interestingly, I found evidence for higher investment into innate immunity in the reproductive termite caste as compared to sterile soldier caste members or false-workers. Furthermore, the induced immune response elicited in soldiers, but particularly in the reproductive caste mimicked the induced immune responses of *C. meridianus* and *B. orientalis* more closely than the response of false-workers. Additionally, the induced response to the same experimental immune challenge was remarkably similar between the subsocial *C. meridianus* and the solitary *B. orientalis*. From these results, I argue that the evolution of division of labor in termites was linked to the evolution of a fundamental change in individual immune defence between the sterile and non-sterile castes.

In **Chapter III**, I expand on the role of the sterile caste in eusociality and immunity by examining the function of soldiers in social immunity in the Darwin termite, *Mastotermes darwiniensis*. In this chapter, *M. darwiniensis* soldiers are shown to contribute significantly to the social immunity of the colony by increasing the survival of groups of workers, probably via the secretion of potent orally-derived antimicrobial substances. In a comprehensive proteomic analysis, I demonstrate that *M. darwiniensis* soldier oral secretions possess a rich array of immune related proteins and enzymes involved in the biosynthesis of cytotoxins such as benzoquinone. These findings shed new light on termite societies, indicating that termites are likely to have evolved a sterile soldier caste with important functions not only in colony defence but also in social immunity.

In this thesis I reveal how the termite immune system evolved during the transition to eusociality. I have established a robust foundation for the study of molecular immunity in termites and contributed new insights into the evolution of immunity in social animals in general. As the contraction and re-expansion of receptors and effectors in termites indicates, the function of a number of immune gene families should be examined in much greater detail. Furthermore, it will be particularly interesting to explore the individual immune (as well as general) responses of termite in a wider social context, particularly given the observed immune differences that were detected between the termite castes. Comparisons with immune adaptations in the Hymenoptera and other social animals would also be highly beneficial to understand commonalities and differences during this key evolutionary transition.

Zusammenfassung

Neben sexueller Reproduktion und Multizellularität wird Eusozialität als einer der größten Evolutionssprünge angesehen (Szathmary and Smith 1995). Eusozialität evolvierte am häufigsten bei Insekten, ins besonders bei Hymenopteren (Ameisen, Bienen und Wespen) und Isopteren (Termiten). Das Hauptkennzeichen der Evolution von Eusozialität bei Insekten ist das Aufkommen einer permanent sterilen Kaste, was durch reproduktive Arbeitsteilung widergespiegelt wird. Eine weitere bemerkenswerte Besonderheit von Insektengesellschaften artgleicher Individuen, ist die Entstehung von ausgefeilten Immunanpassungen auf individueller und auf Gruppenebene. Dadurch wird die Ausbreitung von Krankheiten verhindert. Die Evolution von Immunabwehr bei Termiten ist jedoch kaum verstanden. Vor allem die molekulare Evolution von kanonischen Immunsignalwegen und wie angeborene und induzierte Immunität durch die Evolution einer sterilen Kaste beeinflusst wurde, sind im Wesentlichen unverstanden.

Ein vergleichender Ansatz für Studien über die Evolution der Immunität bei Termiten erfordert solide Kenntnisse über das Immunsystem der nächsten nicht-sozialen Verwandten, den „Schaben“. Zu diesem Zweck wurde in **Kapitel I** das Immunsystem der von *Blatta orientalis* untersucht. Unter Zuhilfenahme von *de novo* Transkriptomanalysen wurde das volle Repertoire von Immungenen dieser Spezies identifiziert. Dadurch konnten Erweiterungen von Immungenfamilien von Rezeptoren wie GNBP, PGRP und dem Hämolymphe LPS-Bindeprotein (LPSBP) ausgemacht werden. Nachdem eine Immunantwort der Schaben mit durch Hitze abgetöteten Mikroben (*Bacillus thuringiensis, Pseudomonas entomophila, Saccharomyces cerevisiae*) induziert wurde, war ich dazu in der Lage als Antwort darauf ein großes Spektrum von induzierten kanonischen Immunsignalwegen zu dokumentieren. Dies deutet auf das Vorhandensein einer effektiven und langanhaltenden Krankheitsabwehr hin, welche ein wesentliches Merkmal von Organismen ist, die in reichen antigenen Umgebungen leben.

Im ersten Teil von **Kapitel II**, untersuchte ich die Evolution von Immunität bei Termiten indem ich zunächst eine Phylogenie mit 19 neu sequenzierten Transkriptomen und 16 bereits vorhandenen genomischen Datensätzen rekonstruierte. Als ein Ergebnis konnten dabei gezeigt werden, dass Termiten eine Schwestergruppe von *Cryptocercus*, welches eine subsoziale Schabengattung ist, darstellen. Außerdem verzeichnete ich jüngsten gemeinsamen Vorfahren (MRCA) im unteren Jura. Anschließend wurde eine evolutionäre Analyse von dem durch das Immunsystem zusammenhängenden Genfamilien basierend auf 18 der neuen Transkriptomsequenzen durchgeführt. Dabei stellte sich heraus, dass eine Familie von antimikrobiellen Peptiden, Dorsomycin, im Laufe der Evolution bei dem

Vorfahren der subsozialen Holzschaben und allen Termiten verloren gegangen ist. Eine weitere Analyse der anderen beiden klassischen Effektoren Katalase und Thioredoxin-Peroxidase konnte eine rapide Reduzierung des erstenren im Vorfahren aller eusozialen Termiten und eine rapide Reduzierung des letzteren in der Ursprung von Termitidae zeigen. Zusätzlich dazu zeigte die Rezeptorfamilie der C-Typ Lektine (CTLs) eine Reduzierung im MRCA von *Cryptocercus* und Termiten sowie ebenfalls im Ursprung der Rhinotermitidae. Interessanter Weise unterliefen diese reduzierten Genfamilien eine anschließende Rückexpansion in einigen individuellen Linien höherer Termiten. Diese Ergebnisse deuten auf eine substantielle Umbildung des Termitenimmunsystems während der Evolution von Eusozialität hin.

Dieser qualitativen Analyse fokussierend auf Evolutionssprüngen in der Immunität von Termiten folgte im zweiten Teil von **Kapitel II** eine quantitative Analyse von individueller Immunität anhand verschiedener Kasten einer repräsentativen niederen Termitenart, *Neotermes castaneus*. Änderungen in der Genexpression wurden daraufhin mit der subsozialen Holzschabe *Cryptocercus meridianus* und der solitären Schabe *B. orientalis* verglichen. Interessanter Weise fand ich Hinweise für ein höheres Investment in angeborene Immunität bei reproduktiven Termitenkasten im Vergleich zu sterilen Soldatkasten oder „falschen“ Arbeitern. Zusätzlich dazu imitiert die induzierte Immunantwort hervorgerufen in Soldaten und besonders in der reproduktiven Kaste die induzierte Immunantwort von *C. meridianus* and *B. orientalis* wesentlich ähnlicher/genauer als die von „falschen“ Arbeitern. Die angeborene Reaktion auf die gleiche Herausforderung des Immunsystems war bemerkenswerter Weise zwischen den subsozialen *C. meridianus* und den solitären *B. orientalis* sehr ähnlich. Anhand dieser Ergebnisse leite ich ab, dass die Evolution von Arbeitsteilung bei Termiten mit der Evolution von fundamentalen Änderungen in der individuellen Immunantwort zwischen sterilen und nicht-sterilen Kasten verknüpft wurde.

In **Kapitel III** erweitere ich die Rolle der sterilen Kaste bezogen auf Eusozialität und Immunität durch Beleuchten der Funktion von Soldaten bei der sozialen Immunität anhand der Darwintermite *Mastotermes darwiniensis*. In diesem Kapitel wird gezeigt, dass Soldaten von M. darwiniensis signifikant zur sozialen Immunität der Kolonie beitragen. Dies geschieht wahrscheinlich durch Erhöhung der Überlebensfähigkeit der Arbeiter durch die Sekretion von wirkungsvollen oralen antimikrobiellen Substanzen bei Soldaten. In einer umfangreichen Proteomanalyse konnte ich zeigen, dass die oralen Sekrete der Soldaten von *M. darwiniensis* ein reichhaltiges Arsenal von mit dem Immunsystem im Zusammenhang stehenden Proteinen und Enzymen, die in der Biosynthese von Zytokinen wie z.B. Benzoquinon eine Rolle spielen, aufweisen. Diese Ergebnisse werfen ein neues Licht auf das Sozialleben von Termiten indem sie darauf hinweisen, dass Termiten wahrscheinlich

eine sterile Soldatenkaste nicht nur für die Kolonieverteidigung benötigen, sondern auch in der sozialen Immunität evolviert haben.

In dieser Dissertation zeige ich wie das Immunsystem von Termiten während des Überganges zur Eusozialität evolvierte. Ich habe ein solides Fundament für künftige Studien zur molekularen Immunität von Termiten gelegt und neue Einsichten in die Evolution von Immunität bei sozialen Tieren im Allgemeinen geliefert. Wie die Reduktion und erneute Expansion von Rezeptoren und Effektoren bei Termiten zeigen, sollte die Funktion etlicher Immungenfamilien künftig noch detaillierter untersucht werden. Des Weiteren wird es besonders interessant sein die individuelle (als auch die generelle) Immunantwort von Termiten in einem weiten sozialen Kontext zu erforschen. Dies wird besonders durch die beobachteten Unterschiede zwischen den Termitenkasten bekräftigt. Außerdem wären Vergleiche bezogen auf Immunanpassungen mit Hymenopteren und anderen sozialen Tieren sehr nützlich um Gemeinsamkeiten und Unterschiede während dieses Schlüsselevolutionssprunges besser verstehen zu können.

Bibliography

Aberer, A.J., K. Kobert, and A. Stamatakis, 2014 ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution* 31 (10):2553-2556.

Agaisse, H., and N. Perrimon, 2004 The roles of JAK/STAT signaling in Drosophila immune responses. *Immunological Reviews* 198 (1):72-82.

Alexander, R.D., 1974 The evolution of social behavior. *Annual Review of Ecology and Systematics* 5 (1):325-383.

Altincicek, B., E. Knorr, and A. Vilcinskas, 2008 Beetle immunity: Identification of immune-inducible genes from the model insect *Tribolium castaneum*. *Developmental & Comparative Immunology* 32 (5):585-595.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, 1990 Basic local alignment search tool. *Journal of Molecular Biology* 215 (3):403-410.

Arefin, B., L. Kucerova, P. Dobes, R. Markus, H. Strnad *et al.*, 2014 Genome-wide transcriptional analysis of *Drosophila* larvae infected by entomopathogenic nematodes shows involvement of complement, recognition and extracellular matrix proteins. *Journal of Innate Immunity* 6 (2):192-204.

Arumugam, G., B. Sreeramulu, R. Paulchamy, S. Thangavel, and J. Sundaram, 2017 Purification and functional characterization of lectin with phenoloxidase activity from the hemolymph of cockroach, *Periplaneta americana*. *Archives of Insect Biochemistry and Physiology*.

Baeder, D.Y., G. Yu, N. Hozé, J. Rolff, and R.R. Regoes, 2016 Antimicrobial combinations: Bliss independence and Loewe additivity derived from mechanistic multi-hit models. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1695):20150294.

Barribeau, S.M., B.M. Sadd, L. Du Plessis, M.J. Brown, S.D. Buechel *et al.*, 2015 A depauperate immune repertoire precedes evolution of sociality in bees. *Genome Biology* 16 (1):1.

Basseri, H.R., A. Dadi-Khoeni, R. Bakhtiari, M. Abolhassani, and R. Hajihosseini-Baghdadabadi, 2016 Isolation and Purification of an Antibacterial Protein from Immune Induced Haemolymph of American Cockroach, *Periplaneta americana*. *Journal of Arthropod-Borne Diseases* 10 (4):519.

Bell, W.J., L.M. Roth, and C.A. Nalepa, 2007 *Cockroaches: ecology, behavior, and natural history*. Baltimore: Johns Hopkins University Press.

Bertino-Grimaldi, D., M.N. Medeiros, R.P. Vieira, A.M. Cardoso, A.S. Turque *et al.*, 2013 Bacterial community composition shifts in the gut of *Periplaneta americana* fed on different lignocellulosic materials. *SpringerPlus* 2 (1):609.

Bignell, D.E., and P. Eggleton, 2000 Termites in ecosystems, pp. 363-387 in *Termites: evolution, sociality, symbioses, ecology*. Springer.

Boomsma, J.J., and R. Gawne, 2018 Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. *Biological Reviews* 93 (1):28-54.

Booth, D., B. Marie, P. Domenici, J.M. Blagburn, and J.P. Bacon, 2009 Transcriptional control of behavior: engrailed knock-out changes cockroach escape trajectories. *Journal of Neuroscience* 29 (22):7181-7190.

Bourguignon, T., N. Lo, S.L. Cameron, J. Šobotník, Y. Hayashi *et al.*, 2015 The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Molecular Biology and Evolution* 32 (2):406-421.

Bourguignon, T., R.H. Scheffrahn, Z.T. Nagy, G. Sonet, B. Host *et al.*, 2016a Towards a revision of the Neotropical soldierless termites (Isoptera: Termitidae): redescription of the genus Grigiotermes Mathews and description of five new genera. *Zoological Journal of the Linnean Society* 176 (1):15-35.

Bourguignon, T., J. Sobotnik, J. Brabcova, D. Sillam-Dusses, A. Bucek *et al.*, 2016b Molecular Mechanism of the Two-Component Suicidal Weapon of *Neocapritermes taracua* Old Workers. *Molecular Biology and Evolution* 33 (3):809-819.

Bourguignon, T., J. Šobotník, C. Dahlsjö, and Y. Roisin, 2016c The soldierless Apicotermitinae: insights into a poorly known and ecologically dominant tropical taxon. *Insectes Sociaux* 63 (1):39-50.

Bourguignon, T., Q. Tang, S.Y. Ho, F. Juna, Z. Wang *et al.*, 2017 Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Molecular Biology and Evolution* 35 (4):970-983.

Bray, N.L., H. Pimentel, P. Melsted, and L. Pachter, 2016 Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34 (5):525.

Bronstein, S., and W. Conner, 1984 Endotoxin-induced behavioural fever in the Madagascar cockroach, *Gromphadorhina portentosa*. *Journal of Insect Physiology* 30 (4):327-330.

Brune, A., 2014 Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* 12 (3):168.

Brune, A., and C. Dietrich, 2015 The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annual Review of Microbiology* 69:145-166.

Buchfink, B., C. Xie, and D.H. Huson, 2015 Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12 (1):59.

Buchon, N., N. Silverman, and S. Cherry, 2014 Immunity in *Drosophila melanogaster*—from microbial recognition to whole-organism physiology. *Nature Reviews Immunology* 14 (12):796.

Bulmer, M., D. Denier, J. Velenovsky, and C. Hamilton, 2012 A common antifungal defense strategy in *Cryptocercus* woodroaches and termites. *Insectes Sociaux* 59 (4):469-478.

Bulmer, M., F. Lay, and C. Hamilton, 2010 Adaptive evolution in subterranean termite antifungal peptides. *Insect Molecular Biology* 19 (5):669-674.

Bulmer, M.S., I. Bachelet, R. Raman, R.B. Rosengaus, and R. Sasisekharan, 2009 Targeting an antimicrobial effector function in insect immunity as a pest control strategy. *Proceedings of the National Academy of Sciences* 106 (31):12652-12657.

Bulmer, M.S., and R.H. Crozier, 2004 Duplication and diversifying selection among termite antifungal peptides. *Molecular Biology and Evolution* 21 (12):2256-2264.

Bulmer, M.S., and R.H. Crozier, 2005 Variation in positive selection in termite GNBPs and Relish. *Molecular Biology and Evolution* 23 (2):317-326.

Buttstedt, A., C.H. Ihling, M. Pietzsch, and R.F. Moritz, 2016 Royalactin is not a royal making of a queen. *Nature* 537 (7621):E10-E12.

Calleri, D.V., R.B. Rosengaus, and J.F. Traniello, 2007 Immunity and reproduction during colony foundation in the dampwood termite, *Zootermopsis angusticollis*. *Physiological Entomology* 32 (2):136-142.

Cameron, S.L., N. Lo, T. Bourguignon, G.J. Svenson, and T.A. Evans, 2012 A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): robust support for interfamilial relationships and molecular synapomorphies define major clades. *Molecular Phylogenetics and Evolution* 65 (1):163-173.

Cao, X., Y. He, Y. Hu, Y. Wang, Y.-R. Chen *et al.*, 2015 The immune signaling pathways of Manduca sexta. *Insect Biochemistry and Molecular Biology* 62:64-74.

Capella-Gutiérrez, S., J.M. Silla-Martínez, and T. Gabaldón, 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25 (15):1972-1973.

Carlsson, A., P. Engström, E.T. Palva, and H. Bennich, 1991 Attacin, an antibacterial protein from *Hyalophora cecropia,* inhibits synthesis of outer membrane proteins in *Escherichia coli* by interfering with omp gene transcription. *Infection and Immunity* 59 (9):3040-3045.

Carlsson, A., T. Nyström, H. de Cock, and H. Bennich, 1998 Attacin-an insect immune protein-binds LPS and triggers the specific inhibition of bacterial outer-membrane protein synthesis. *Microbiology* 144 (8):2179-2188.

Cerenius, L., B.L. Lee, and K. Söderhäll, 2008 The proPO-system: pros and cons for its role in invertebrate immunity. *Trends in Immunology* 29 (6):263-271.

Che, Y., D. Wang, Y. Shi, X. Du, Y. Zhao *et al.*, 2016 A global molecular phylogeny and timescale of evolution for *Cryptocercus* woodroaches. *Molecular Phylogenetics and Evolution*.

Chen, C., H. Durrant, R. Newton, and N. Ratcliffe, 1995 A study of novel lectins and their involvement in the activation of the prophenoloxidase system in *Blaberus discoidalis*. *Biochemical Journal* 310 (1):23-31.

Chen, W., Y.X. Liu, and G.F. Jiang, 2015 De novo Assembly and Characterization of the Testis Transcriptome and Development of EST-SSR Markers in the Cockroach *Periplaneta americana*. *Scientific Reports* 5:11144.

Chowański, S., Z. Adamski, J. Lubawy, P. Marciniak, J. Pacholska-Bogalska *et al.*, 2017 Insect peptides-perspectives in human diseases treatment. *Current Medicinal Chemistry*.

Cociancich, S., A. Ghazi, C. Hetru, J. Hoffmann, and L. Letellier, 1993 Insect defensin, an inducible antibacterial peptide, forms voltage-dependent channels in *Micrococcus luteus*. *Journal of Biological Chemistry* 268 (26):19239-19245.

Colbert, J.D., S.P. Matthews, G. Miller, and C. Watts, 2009 Diverse regulatory roles for lysosomal proteases in the immune response. *European Journal of Immunology* 39 (11):2955-2965.

Colinet, D., D. Cazes, M. Belghazi, J.-L. Gatti, and M. Poirié, 2011 Extracellular superoxide dismutase in insects characterization, function, and interspecific variation in parasitoid wasp venom. *Journal of Biological Chemistry* 286 (46):40110-40121.

Cotter, S., and R. Kilner, 2010 Personal immunity versus social immunity. *Behavioral Ecology* 21 (4):663-668.

Cox-Foster, D.L., and J.E. Stehr, 1994 Induction and localization of FAD-glucose dehydrogenase (GLD) during encapsulation of abiotic implants in *Manduca sexta* larvae. *Journal of Insect Physiology* 40 (3):235-249.

Cremer, S., S.A. Armitage, and P. Schmid-Hempel, 2007 Social immunity. *Current Biology* 17 (16):R693-R702.

Cremer, S., C.D. Pull, and M.A. Fürst, 2018 Social immunity: Emergence and evolution of colony-level disease protection. *Annual Review of Entomology* 63:105-123.

Cremer, S., and M. Sixt, 2009 Analogies in the evolution of individual and social immunity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1513):129-142.

Da Silva, P., L. Jouvensal, M. Lamberty, P. Bulet, A. Caille *et al.*, 2003 Solution structure of termicin, an antimicrobial peptide from the termite *Pseudacanthotermes spiniger*. *Protein Science* 12 (3):438-446.

De Bie, T., N. Cristianini, J.P. Demuth, and M.W. Hahn, 2006 CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22 (10):1269-1271.

De Gregorio, E., P.T. Spellman, G.M. Rubin, and B. Lemaitre, 2001 Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. *Proceedings of the National Academy of Sciences* 98 (22):12590-12595.

de Morais Guedes, S., R. Vitorino, R. Domingues, K. Tomer, A.F. Correia *et al.*, 2005 Proteomics of immune-challenged *Drosophila melanogaster* larvae hemolymph. *Biochemical and Biophysical Research Communications* 328 (1):106-115.

de Roode, J.C., and T. Lefèvre, 2012 Behavioral immunity in insects. *Insects* 3 (3):789-820.

Djernæs, M., K.-D. Klass, and P. Eggleton, 2015 Identifying possible sister groups of Cryptocercidae+ Isoptera: A combined molecular and morphological phylogeny of Dictyoptera. *Molecular Phylogenetics and Evolution* 84:284-303.

Drapeau, M.D., S. Albert, R. Kucharski, C. Prusko, and R. Maleszka, 2006 Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res* 16 (11):1385-1394.

Drapeau, M.D., A. Radovic, P.J. Wittkopp, and A.D. Long, 2003 A gene necessary for normal male courtship, yellow, acts downstream of fruitless in the *Drosophila melanogaster* larval brain. *Developmental Neurobiology* 55 (1):53-72.

Duneau, D.F., H.C. Kondolf, J.H. Im, G.A. Ortiz, C. Chow *et al.*, 2017 The Toll pathway underlies host sexual dimorphism in resistance to both Gram-negative and Gram-positive bacteria in mated *Drosophila*. *BMC Biology* 15 (1):124.

Dziarski, R., 2004 Peptidoglycan recognition proteins (PGRPs). *Molecular Immunology* 40 (12):877-886.

Eggleton, P., 2001 Termites and trees: a review of recent advances in termite phylogenetics. *Insectes Sociaux* 48 (3):187-193.

Ellwood, M.D., and W.A. Foster, 2004 Doubling the estimate of invertebrate biomass in a rainforest canopy. *Nature* 429 (6991):549.

Emms, D.M., and S. Kelly, 2015 OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16 (1):157.

Engel, M.S., P. Barden, M.L. Riccio, and D.A. Grimaldi, 2016 Morphologically Specialized Termite Castes and Advanced Sociality in the Early Cretaceous. *Current Biology* 26 (4):522-530.

Engel, M.S., D.A. Grimaldi, and K. Krishna, 2007a *Primitive termites from the early Cretaceous of Asia (Isoptera)*: Staatliches Museum für Naturkunde.

Engel, M.S., D.A. Grimaldi, and K. Krishna, 2007b *A synopsis of Baltic amber termites (Isoptera)*: Staatliches Museum für Naturkunde.

Engel, M.S., D.A. Grimaldi, and K. Krishna, 2009 Termites (Isoptera): their phylogeny, classification, and rise to ecological dominance. *American Museum Novitates*:1-27.

Evans, J., K. Aronstein, Y.P. Chen, C. Hetru, J.L. Imler *et al.*, 2006 Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Molecular Biology* 15 (5):645-656.

Faulhaber, L., and R. Karp, 1992 A diphasic immune response against bacteria in the American cockroach. *Immunology* 75 (2):378.

Finkel, T., and N.J. Holbrook, 2000 Oxidants, oxidative stress and the biology of ageing. *Nature* 408 (6809):239.

Finn, R.D., J. Clements, and S.R. Eddy, 2011 HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39 (suppl_2):W29-W37.

Flatt, T., A. Heyland, F. Rus, E. Porpiglia, C. Sherlock *et al.*, 2008 Hormonal regulation of the humoral innate immune response in *Drosophila melanogaster*. *Journal of Experimental Biology* 211 (16):2712-2724.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li, 2012 CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28 (23):3150-3152.
Fuller, C.A., 2007 Fungistatic activity of freshly killed termite, *Nasutitermes acajutlae*, soldiers in the Caribbean. *Journal of Insect Science* 7 (1).
Gao, Q., M.J. Bidochka, and G.J. Thompson, 2012 Effect of group size and caste ratio on individual survivorship and social immunity in a subterranean termite. *Acta ethologica* 15 (1):55-63.
Gao, Q., and G. Thompson, 2015 Social context affects immune gene expression in a subterranean termite. *Insectes Sociaux* 62 (2):167-170.
Geib, S.M., T.R. Filley, P.G. Hatcher, K. Hoover, J.E. Carlson *et al.*, 2008 Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences* 105 (35):12932-12937.
Gerardo, N.M., B. Altincicek, C. Anselme, H. Atamian, S.M. Barribeau *et al.*, 2010 Immunity and other defenses in pea aphids, Acyrthosiphon pisum. *Genome Biology* 11 (2):R21.
González-Santoyo, I., and A. Córdoba-Aguilar, 2012 Phenoloxidase: a key component of the insect immune system. *Entomologia Experimentalis et Applicata* 142 (1):1-16.
Gouy, M., S. Guindon, and O. Gascuel, 2009 SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27 (2):221-224.
Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29 (7):644-652.
Grimaldi, D.A., M.S. Engel, and K. Krishna, 2008 The species of Isoptera (Insecta) from the early Cretaceous Crato Formation: a revision. *American Museum Novitates*:1-30.
Haas, B.J., A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood *et al.*, 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8 (8):1494.
Haine, E.R., Y. Moret, M.T. Siva-Jothy, and J. Rolff, 2008 Antimicrobial defense and persistent infection in insects. *Science* 322 (5905):1257-1259.
Hamilton, C., B.T. Lejeune, and R.B. Rosengaus, 2011 Trophallaxis and prophylaxis: social immunity in the carpenter ant *Camponotus pennsylvanicus*. *Biology Letters* 7 (1):89-92.
Happ, G.M., 1968 Quinone and hydrocarbon production in the defensive glands of Eleodes longicollis and *Tribolium castaneum* (Coleoptera, Tenebrionidae). *Journal of Insect Physiology* 14 (12):1821-1837.
Harpur, B.A., and A. Zayed, 2013 Accelerated evolution of innate immunity proteins in social insects: adaptive evolution or relaxed constraint? *Molecular Biology and Evolution* 30 (7):1665-1674.
Harrison, M.C., N. Arning, L.P. Kremer, G. Ylla, X. Belles *et al.*, 2018a Expansions of key protein families in the German cockroach highlight the molecular basis of its remarkable success as a global indoor pest. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*.
Harrison, M.C., E. Jongepier, H.M. Robertson, N. Arning, T. Bitard-Feildel *et al.*, 2018b Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution*:1.
Hernández López, J., U. Riessberger-Gallé, K. Crailsheim, and W. Schuehly, 2017 Cuticular hydrocarbon cues of immune-challenged workers elicit immune activation in honeybee queens. *Molecular Ecology* 26 (11):3062-3073.
Hillyer, J.F., 2016 Insect immunology and hematopoiesis. *Developmental & Comparative Immunology* 58:102-118.
Hoffmann, J.A., 2003 The immune response of *Drosophila*. *Nature* 426 (6962):33.
Hoffmann, J.A., and J.-M. Reichhart, 2002 Drosophila innate immunity: an evolutionary perspective. *Nature immunology* 3 (2):121.
Hojo, M., M. Morioka, T. Matsumoto, and T. Miura, 2005 Identification of soldier caste-specific protein in the frontal gland of nasute termite *Nasutitermes takasagoensis* (Isoptera: Termitidae). *Insect Biochemistry and Molecular Biology* 35 (4):347-354.
Hölldobler, B., and E.O. Wilson, 2009 *The superorganism: the beauty, elegance, and strangeness of insect societies*: WW Norton & Company.
Hopkins, T.L., and K.J. Kramer, 1992 Insect cuticle sclerotization. *Annual Review of Entomology* 37 (1):273-302.
Howard, K.J., and B.L. Thorne, 2010 Eusocial evolution in termites and Hymenoptera, pp. 97-132 in *Biology of termites: a modern synthesis*. Springer.
Huang, X., J.T. Warren, J. Buchanan, L.I. Gilbert, and M.P. Scott, 2007 *Drosophila* Niemann-Pick type C-2 genes control sterol homeostasis and steroid biosynthesis: a model of human neurodegenerative disease. *Development* 134 (20):3733-3742.
Hughes, W.O., J. Eilenberg, and J.J. Boomsma, 2002 Trade-offs in group living: transmission and

disease resistance in leaf-cutting ants. *Philosophical Transactions of the Royal Society B: Biological Sciences* 269 (1502):1811-1819.

Husseneder, C., and D.M. Simms, 2014 Effects of caste on the expression of genes associated with septic injury and xenobiotic exposure in the Formosan subterranean termite. *PloS One* 9 (8):e105582.

Inward, D., G. Beccaloni, and P. Eggleton, 2007a Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biology Letters* 3 (3):331-335.

Inward, D.J., A.P. Vogler, and P. Eggleton, 2007b A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Molecular Phylogenetics and Evolution* 44 (3):953-967.

Irving, P., L. Troxler, T.S. Heuer, M. Belvin, C. Kopczynski *et al.*, 2001 A genome-wide analysis of immune responses in Drosophila. *Proceedings of the National Academy of Sciences* 98 (26):15119-15124.

Ishihama, Y., Y. Oda, T. Tabata, T. Sato, T. Nagasu *et al.*, 2005 Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics* 4 (9):1265-1272.

Jeong, J.E., H.J. Hwang, H.S. Park, H.J. Cha, Y.S. Lee *et al.*, 2014 Analysis of German cockroach (*Blattella germanica*) expressed sequence tags. *Genes & Genomics* 37 (3):271-280.

Johnston, P.R., O. Makarova, and J. Rolff, 2014 Inducible defenses stay up late: temporal patterns of immune gene expression in Tenebrio molitor. *G3: Genes, Genomes, Genetics* 4 (6):947-955.

Jomori, T., T. Kubo, and S. Natori, 1990 Purification and characterization of lipopolysaccharide-binding protein from hemolymph of American cockroach *Periplaneta americana*. *The FEBS Journal* 190 (1):201-206.

Jomori, T., and S. Natori, 1991 Molecular cloning of cDNA for lipopolysaccharide-binding protein from the hemolymph of the American cockroach, *Periplaneta americana*. Similarity of the protein with animal lectins and its acute phase expression. *Journal of Biological Chemistry* 266 (20):13318-13323.

Jomori, T., and S. Natori, 1992 Function of the lipopolysaccharide-binding protein of *Periplaneta americana* as an opsonin. *FEBS Letters* 296 (3):283-286.

Jones, B.M., C.J. Kingwell, W.T. Wcislo, and G.E. Robinson, 2017 Caste-biased gene expression in a facultatively eusocial bee suggests a role for genetic accommodation in the evolution of eusociality. *Proceedings of the Royal Society B: Biological Sciences* 284 (1846).

Kaakeh, W., B.L. Reid, and G.W. Bennett, 1996 Horizontal transmission of the entomopathogenic fungus *Metarhizium anisopliae* (Imperfect fungi: Hyphomycetes) and hydramethylnon among German cockroaches (Dictyoptera: Blattellidae). *Journal of Entomological Science* 31 (4):378-390.

Kaji, T., J. Keiler, T. Bourguignon, and T. Miura, 2016 Functional transformation series and the evolutionary origin of novel forms: evidence from a remarkable termite defensive organ. *Evolution & Development* 18 (2):78-88.

Kamakura, M., 2011 Royalactin induces queen differentiation in honeybees. *Nature* 473 (7348):478-483.

Kambara, K., T. Nagae, W. Ohmura, and M. Azuma, 2014 Aquaporin water channel in the salivary glands of the Formosan subterranean termite Coptotermes formosanus is predominant in workers and absent in soldiers. *Physiological Entomology* 39 (2):94-102.

Karp, R.D., L.E. Duwel, L.M. Faulhaber, and M.A. Harju, 1994 Evolution of adaptive immunity: inducible responses in the American cockroach. *Annals of the New York Academy of Sciences* 712 (1):82-91.

Katoh, K., J. Rozewicki, and K.D. Yamada, 2017 MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*.

Katoh, K., and D.M. Standley, 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30 (4):772-780.

Keller, L., 1998 Queen lifespan and colony characteristics in ants and termites. *Insectes Sociaux* 45 (3):235-246.

Kim, I.-W., J.H. Lee, S. Subramaniyam, E.-Y. Yun, I. Kim *et al.*, 2016 De novo transcriptome analysis and detection of antimicrobial peptides of the American cockroach *periplaneta Americana* (Linnaeus). *PloS One* 11 (5):e0155304.

Klass, K.-D., C. Nalepa, and N. Lo, 2008 Wood-feeding cockroaches as models for termite evolution (Insecta: Dictyoptera): *Cryptocercus* vs. *Parasphaeria boleiriana*. *Molecular Phylogenetics and Evolution* 46 (3):809-817.

Konrad, M., A.V. Grasse, S. Tragust, and S. Cremer, 2015 Anti-pathogen protection versus survival costs mediated by an ectosymbiont in an ant host, pp. 20141976 in *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society.

Konrad, M., M.L. Vyleta, F.J. Theis, M. Stock, S. Tragust *et al.*, 2012 Social transfer of pathogenic fungus promotes active immunisation in ant colonies. *PLoS Biology* 10 (4):e1001300.

Korb, J., 2007 Termites. *Current Biology* 17 (23):R995-R999.

Korb, J., 2008 Termites, hemimetabolous diploid white ants? *Frontiers in zoology* 5 (1):15.

Korb, J., 2015 Chapter four-juvenile hormone: a central regulator of termite caste polyphenism. *Advances in Insect Physiology* 48:131-161.

Korb, J., and X. Belles, 2017 Juvenile hormone and hemimetabolan eusociality: a comparison of cockroaches with termites. *Current Opinion in Insect Science*.

Korb, J., and K. Hartfelder, 2008 Life history and development-a framework for understanding developmental plasticity in lower termites. *Biological Reviews* 83 (3):295-313.

Korb, J., M. Poulsen, H. Hu, C. Li, J.J. Boomsma *et al.*, 2015 A genomic comparison of two termites with different social complexity. *Frontiers in Genetics* 6:9.

Krishna, K., D.A. Grimaldi, V. Krishna, and M.S. Engel, 2013 Treatise on the Isoptera of the world.(Bulletin of the American Museum of Natural History, no. 377).

Krishna, K., and F. Weesner, 1969 *Biology of termites. Volume I*. London: Academic Press.

Krishna, K., and F. Weesner, 1970 *Biology of termites. Volume II*. London: Academic Press.

Krogh, A., B. Larsson, G. Von Heijne, and E.L. Sonnhammer, 2001 Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* 305 (3):567-580.

Kulshrestha, V., and S. Pathak, 1997 Aspergillosis in German cockroach *Blattella germanica* (L.)(Blattoidea: blattellidae). *Mycopathologia* 139 (2):75-78.

Labandeira, C.C., 1994 A compendium of fossil insect families.

Lamberty, M., D. Zachary, R. Lanot, C. Bordereau, A. Robert *et al.*, 2001 Insect immunity constitutive expression of a cysteine-rich antifungal and a linear antibacterial peptide in a termite insect. *Journal of Biological Chemistry* 276 (6):4085-4092.

Langmead, B., and S.L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9 (4):357.

Lartillot, N., and H. Philippe, 2004 A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* 21 (6):1095-1109.

Lavine, M., and M. Strand, 2002 Insect hemocytes and their role in immunity. *Insect Biochemistry and Molecular Biology* 32 (10):1295-1309.

Lawson, S.P., L.T. Sigle, A.L. Lind, A.W. Legan, J.N. Mezzanotte *et al.*, 2017 An alternative pathway to eusociality: Exploring the molecular and functional basis of fortress defense. *Evolution* 71 (8):1986-1998.

LeBoeuf, A.C., P. Waridel, C.S. Brent, A.N. Gonçalves, L. Menin *et al.*, 2016 Oral transfer of chemical cues, growth proteins and hormones in social insects. *elife* 5:e20375.

Lee, S., R. Siddiqui, and N.A. Khan, 2012 Animals living in polluted environments are potential source of antimicrobials against infectious agents. *Pathogens and Global Health* 106 (4):218-223.

Legendre, F., A. Nel, G.J. Svenson, T. Robillard, R. Pellens *et al.*, 2015 Phylogeny of Dictyoptera: dating the origin of cockroaches, praying mantises and termites with molecular data and controlled fossil evidence. *PloS One* 10 (7):e0130127.

Legendre, F., M.F. Whiting, C. Bordereau, E.M. Cancello, T.A. Evans *et al.*, 2008 The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors. *Molecular Phylogenetics and Evolution* 48 (2):615-627.

Li, S., S. Zhu, Q. Jia, D. Yuan, C. Ren *et al.*, 2018 The genomic and functional landscapes of developmental plasticity in the American cockroach. *Nature Communications* 9 (1):1008.

Lihoreau, M., J. Costa, and C. Rivault, 2012 The social biology of domiciliary cockroaches: colony structure, kin recognition and collective decisions. *Insectes Sociaux* 59 (4):445-452.

Lihoreau, M., and C. Rivault, 2010 German cockroach males maximize their inclusive fitness by avoiding mating with kin. *Animal Behaviour* 80 (2):303-309.

Liu, Z., K. Yuan, R. Zhang, X. Ren, X. Liu *et al.*, 2016 Cloning and purification of the first termicin-like peptide from the cockroach *Eupolyphaga sinensis*. *Journal of Venomous Animals and Toxins including Tropical Diseases* 22 (1):5.

Lo, N., G. Tokuda, H. Watanabe, H. Rose, M. Slaytor *et al.*, 2000 Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Current Biology* 10 (13):801-804.

Logue, D.M., S. Mishra, D. McCaffrey, D. Ball, and W.H. Cade, 2009 A behavioral syndrome linking courtship behavior toward males and females predicts reproductive success from a single mating in the hissing cockroach, *Gromphadorhina portentosa*. *Behavioral Ecology* 20 (4):781-788.

Love, M.I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15 (12):550.

Maget-Dana, R., and M. Ptak, 1997 Penetration of the insect defensin A into phospholipid monolayers and formation of defensin A-lipid complexes. *Biophysical Journal* 73 (5):2527-2533.

Mayer, A., T. Mora, O. Rivoire, and A.M. Walczak, 2016 Diversity of immune strategies explained by adaptation to pathogen statistics. *Proceedings of the National Academy of Sciences* 113 (31):8630-8635.

McAfee, A., T.F. Collins, L.L. Madilao, and L.J. Foster, 2017 Odorant cues linked to social immunity induce lateralized antenna stimulation in honey bees (*Apis mellifera L.*). *Scientific Reports* 7:46171.

Meunier, J., 2015 Social immunity and the evolution of group living in insects. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1669):20140102.

Milutinović, B., R. Peuß, K. Ferro, and J. Kurtz, 2016 Immune priming in arthropods: an update focusing on the red flour beetle. *Zoology* 119 (4):254-261.

Misof, B., S. Liu, K. Meusemann, R.S. Peters, A. Donath *et al.*, 2014 Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346 (6210):763-767.

Mitaka, Y., K. Kobayashi, and K. Matsuura, 2017a Caste-, sex-, and age-dependent expression of immune-related genes in a Japanese subterranean termite, *Reticulitermes speratus*. *PloS One* 12 (4):e0175417.

Mitaka, Y., K. Kobayashi, A. Mikheyev, M.M. Tin, Y. Watanabe *et al.*, 2016 Caste-Specific and Sex-Specific Expression of Chemoreceptor Genes in a Termite. *PloS One* 11 (1):e0146125.

Mitaka, Y., N. Mori, and K. Matsuura, 2017b Multi-functional roles of a soldier-specific volatile as a worker arrestant, primer pheromone and an antimicrobial agent in a termite. *Philosophical Transactions of the Royal Society B: Biological Sciences* 284 (1859).

Moore, B., 1968 Studies on the chemical composition and function of the cephalic gland secretion in Australian termites. *Journal of Insect Physiology* 14 (1):33-39.

Moreau, C.S., C.D. Bell, R. Vila, S.B. Archibald, and N.E. Pierce, 2006 Phylogeny of the ants: diversification in the age of angiosperms. *Science* 312 (5770):101-104.

Morella, N.M., and B. Koskella, 2017 the value of a comparative Approach to Understand the complex interplay between Microbiota and Host immunity. *Frontiers in Immunology* 8:1114.

Myles, T.G., 1999 Review of secondary reproduction in termites (Insecta: Isoptera). *Sociobiology* 33:1-91.

Myllymäki, H., S. Valanne, and M. Rämet, 2014 The Drosophila imd signaling pathway. *The Journal of Immunology* 192 (8):3455-3462.

Mylonakis, E., L. Podsiadlowski, M. Muhammed, and A. Vilcinskas, 2016 Diversity, evolution and medical applications of insect antimicrobial peptides. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1695):20150290.

Nakhleh, J., L. El Moussawi, and M.A. Osta, 2017 The melanization response in insect immunity, pp. 83-109 in *Advances in Insect Physiology*. Elsevier.

Nalepa, C.A., 2010 Altricial development in wood-feeding cockroaches: the key antecedent of termite eusociality, pp. 69-95 in *Biology of termites: A modern synthesis*. Springer.

Napping, A.J., 1995 Superoxide anion generation in Irosophila during melanotic encapsulation of parasites. *European Journal of Cell Biology* 68:450-456.

Negulescu, H., Y. Guo, T.P. Garner, O.Y. Goodwin, G. Henderson *et al.*, 2015 A kazal-type serine protease inhibitor from the defense gland secretion of the subterranean termite *Coptotermes formosanus Shiraki*. *PloS One* 10 (5):e0125376.

Oba, Y., N. Yoshida, S. Kanie, M. Ojika, and S. Inouye, 2013 Biosynthesis of Firefly Luciferin in Adult Lantern: Decarboxylation of ʟ-Cysteine is a Key Step for Benzothiazole Ring Formation in Firefly Luciferin Synthesis. *PloS One* 8 (12):e84023.

Ohkuma, M., 2003 Termite symbiotic systems: efficient bio-recycling of lignocellulose. *Applied Microbiology and Biotechnology* 61 (1):1-9.

Oster, G.F., and E.O. Wilson, 1978 *Caste and ecology in the social insects*: Princeton University Press.

Otti, O., S. Tragust, and H. Feldhaar, 2014 Unifying external and internal immune defences. *Trends Ecology and Evolution* 29 (11):625-634.

Palma-Onetto, V., K. Hošková, B. Křížková, R. Krejčířová, J. Pflegerová *et al.*, 2018 The labral gland in termite soldiers. *Biological Journal of the Linnean Society* 123 (3):535-544.

Palmer, W.J., A. Duarte, M. Schrader, J.P. Day, R. Kilner *et al.*, 2016 A gene associated with social immunity in the burying beetle Nicrophorus vespilloides. *Philosophical Transactions of the Royal Society B: Biological Sciences* 283 (1823):20152733.

Petersen, T.N., S. Brunak, G. von Heijne, and H. Nielsen, 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8 (10):785-786.

Pfaffl, M.W., 2001 A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research* 29 (9):e45-e45.

Pham, L.N., M.S. Dionne, M. Shirasu-Hiza, and D.S. Schneider, 2007 A specific primed immune response in Drosophila is dependent on phagocytes. *PLoS Pathogens* 3 (3):e26.

Pomés, A., G.A. Mueller, T.A. Randall, M.D. Chapman, and L.K. Arruda, 2017 New Insights into Cockroach Allergens. *Current Allergy and Asthma Reports* 17 (4):25.

Prestwich, G., 1984 Defense mechanisms of termites. *Annual Review of Entomology* 29 (1):201-232.

Prestwich, G.D., 1979 Chemical defense by termite soldiers. *Journal of Chemical Ecology* 5 (3):459-480.

Prestwich, G.D., R.W. Jones, and M.S. Collins, 1981 Terpene biosynthesis by nasute termite soldiers (Isoptera: Nasutitermitinae). *Insect Biochemistry* 11 (3):331-336.

Pull, C.D., L.V. Ugelvig, F. Wiesenhofer, A.V. Grasse, S. Tragust *et al.*, 2018 Destructive disinfection of infected brood prevents systemic disease spread in ant colonies. *elife* 7.

Qiu, H.-L., and D.-F. Cheng, 2017 A Chemosensory Protein Gene Si-CSP1 Associated With Necrophoric Behavior in Red Imported Fire Ants (Hymenoptera: Formicidae). *Journal of Economic Entomology* 110 (3):1284-1290.

Radyuk, S.N., V.I. Klichko, B. Spinola, R.S. Sohal, and W.C. Orr, 2001 The peroxiredoxin gene family in Drosophila melanogaster. *Free Radical Biology and Medicine* 31 (9):1090-1100.

Rahmet-Alla, M., and A.F. Rowley, 1989 Studies on the cellular defense reactions of the Madeira cockroach, Leucophaea maderae: nodule formation in response to injected bacteria. *Journal of Invertebrate Pathology* 54 (2):200-207.

Reinhard, J., M.J. Lacey, F. Ibarra, F.C. Schroeder, M. Kaib *et al.*, 2002 Hydroquinone: a general phagostimulating pheromone in termites. *Journal of Chemical Ecology* 28 (1):1-14.

Robson, S.K., and J.F. Traniello, 2016 Division of labor in complex societies: a new age of conceptual expansion and integrative analysis. *Behavioral Ecology and Sociobiology* 70 (7):995-998.

Roisin, Y., and J. Korb, 2010 Social organisation and the status of workers in termites, pp. 133-164 in *Biology of termites: A modern synthesis*. Springer.

Rolff, J., and S. Reynolds, 2009 *Insect infection and immunity: evolution, ecology, and mechanisms*: Oxford university press.

Rosengaus, R., C. Jordan, M. Lefebvre, and J. Traniello, 1999a Pathogen alarm behavior in a termite: a new form of communication in social insects. *Naturwissenschaften* 86 (11):544-548.

Rosengaus, R.B., M.L. Lefebvre, and J.F. Traniello, 2000 Inhibition of fungal spore germination by *Nasutitermes*: evidence for a possible antiseptic role of soldier defensive secretions. *Journal of Chemical Ecology* 26 (1):21-39.

Rosengaus, R.B., A.B. Maxmen, L.E. Coates, and J.F. Traniello, 1998a Disease resistance: a benefit of sociality in the dampwood termite *Zootermopsis angusticollis* (Isoptera: Termopsidae). *Behavioral Ecology and Sociobiology* 44 (2):125-134.

Rosengaus, R.B., A.B. Maxmen, L.E. Coates, and J.F.A. Traniello, 1998b Disease resistance: a benefit of sociality in the dampwood termite *Zootermopsis angusticollis* (Isoptera : Termopsidae). *Behavioral Ecology and Sociobiology* 44 (2):125-134.

Rosengaus, R.B., K.F. Schultheis, A. Yalonetskaya, M.S. Bulmer, W.S. DuComb *et al.*, 2014 Symbiont-derived β-1, 3-glucanases in a social insect: mutualism beyond nutrition. *Frontiers in Microbiology* 5.

Rosengaus, R.B., and J.F. Traniello, 2001 Disease susceptibility and the adaptive nature of colony demography in the dampwood termite Zootermopsis angusticollis. *Behavioral Ecology and Sociobiology* 50 (6):546-556.

Rosengaus, R.B., J.F. Traniello, T. Chen, J.J. Brown, and R.D. Karp, 1999b Immunity in a social insect. *Naturwissenschaften* 86 (12):588-591.

Ryan, N.A., and R.D. Karp, 1993 Stimulation of hemocyte proliferation in the American cockroach (*Periplaneta americana*) by the injection of Enterobacter cloacae. *Journal of Insect Physiology* 39 (7):601-608.

Sabree, Z.L., C.Y. Huang, G. Arakawa, G. Tokuda, N. Lo *et al.*, 2012 Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite Mastotermes darwiniensis. *Applied and environmental microbiology* 78 (1):204-210.

Sackton, T.B., B.P. Lazzaro, and A.G. Clark, 2017 Rapid expansion of immune-related gene families in

the house fly, *Musca domestica*. *Molecular Biology and Evolution* 34 (4):857-872.
Scharf, M.E., D. Wu-Scharf, B.R. Pittendrigh, and G.W. Bennett, 2003 Caste- and development-associated gene expression in a lower termite. *Genome Biology* 4 (10).
Schmid-Hempel, P., 1998 *Parasites in social insects*: Princeton University Press.
Shi, X.-Z., X. Zhong, and X.-Q. Yu, 2012 Drosophila melanogaster NPC2 proteins bind bacterial cell wall components and may function in immune signal pathways. *Insect Biochemistry and Molecular Biology* 42 (8):545-556.
Sillam-Dusses, D., J. Krasulova, V. Vrkoslav, J. Pytelkova, J. Cvacka *et al.*, 2012 Comparative study of the labial gland secretion in termites (Isoptera). *PloS One* 7 (10):e46431.
Simão, F.A., R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, and E.M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19):3210-3212.
Smith, S.A., and C.W. Dunn, 2008 Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24 (5):715-716.
Šobotník, J., A. Jirošová, and R. Hanus, 2010 Chemical warfare in termites. *Journal of Insect Physiology* 56 (9):1012-1021.
Soneson, C., M.I. Love, and M.D. Robinson, 2015 Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4.
Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9):1312-1313.
Stroeymeyt, N., B. Casillas-Pérez, and S. Cremer, 2014 Organisational immunity in social insects. *Current Opinion in Insect Science* 5:1-15.
Suchard, M.A., P. Lemey, G. Baele, D.L. Ayres, A.J. Drummond *et al.*, 2018 Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* 4 (1):vey016.
Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One* 6 (7):e21800.
Szathmáry, E., and J.M. Smith, 1995 The major evolutionary transitions. *Nature* 374 (6519):227-232.
Tarver, M.R., E.A. Schmelz, J.R. Rocca, and M.E. Scharf, 2009 Effects of soldier-derived terpenes on soldier caste differentiation in the termite Reticulitermes flavipes. *Journal of Chemical Ecology* 35 (2):256-264.
Tauszig, S., E. Jouanguy, J.A. Hoffmann, and J.-L. Imler, 2000 Toll-related receptors and the control of antimicrobial peptide expression in Drosophila. *Proceedings of the National Academy of Sciences* 97 (19):10520-10525.
Team, R.C., 2016 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014.
Terrapon, N., C. Li, H.M. Robertson, L. Ji, X. Meng *et al.*, 2014 Molecular traces of alternative social organization in a termite genome. *Nature Communications* 5:3636.
Thorne, B.L., N.L. Breisch, and M.L. Muscedere, 2003 Evolution of eusociality and the soldier caste in termites: influence of intraspecific competition and accelerated inheritance. *Proceedings of the National Academy of Sciences* 100 (22):12808-12813.
Thorne, B.L., and J.F. Traniello, 2003 Comparative social biology of basal taxa of ants and termites. *Annual Review of Entomology* 48 (1):283-306.
Tian, H., S.B. Vinson, and C.J. Coates, 2004 Differential gene expression between alate and dealate queens in the red imported fire ant, *Solenopsis invicta Buren* (Hymenoptera: Formicidae). *Insect Biochemistry and Molecular Biology* 34 (9):937-949.
Tian, L., and X. Zhou, 2014 The soldiers in societies: defense, regulation, and evolution. *International journal of biological sciences* 10 (3):296.
Traniello, J.F., R.B. Rosengaus, and K. Savoie, 2002 The development of immunity in a social insect: evidence for the group facilitation of disease resistance. *Proceedings of the National Academy of Sciences* 99 (10):6838-6842.
Tunaz, H., Y. Park, K. Büyükgüzel, J.C. Bedick, A. Nor Aliza *et al.*, 2003 Eicosanoids in insect immunity: bacterial infection stimulates hemocytic phospholipase A2 activity in tobacco hornworms. *Archives of Insect Biochemistry and Physiology* 52 (1):1-6.
Tunaz, H., and D.W. Stanley, 2004 Phospholipase A2 in salivary glands isolated from tobacco hornworms, Manduca sexta. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 139 (1):27-33.
Turnbull, C., H. Caravan, T. Chapman, D. Nipperess, S. Dennison *et al.*, 2012 Antifungal activity in thrips soldiers suggests a dual role for this caste. *Biol Lett*:rsbl20120184.
Valanne, S., J.-H. Wang, and M. Rämet, 2011 The Drosophila toll signaling pathway. *The Journal of Immunology* 186 (2):649-656.

Vannette, R.L., A. Mohamed, and B.R. Johnson, 2015 Forager bees (*Apis mellifera*) highly express immune and detoxification genes in tissues associated with nectar processing. *Scientific Reports* 5:16224.

Velenovsky, J.F., J. Kalisch, and M.S. Bulmer, 2016 Selective sweeps in *Cryptocercus* woodroach antifungal proteins. *Genetica* 144 (5):547-552.

Verrett, J.M., K.B. Green, L.M. Gamble, and F.C. Crochen, 1987 A hemocoelic Candida parasite of the American cockroach (Dictyoptera: Blattidae). *Journal of Economic Entomology* 80 (6):1205-1212.

Viljakainen, L., 2015 Evolutionary genetics of insect innate immunity. *Briefings in Functional Genomics* 14 (6):407-412.

Viljakainen, L., J.D. Evans, M. Hasselmann, O. Rueppell, S. Tingek *et al.*, 2009 Rapid evolution of immune proteins in social insects. *Molecular Biology and Evolution* 26 (8):1791-1801.

Vizcaíno, J.A., A. Csordas, N. Del-Toro, J.A. Dianes, J. Griss *et al.*, 2015 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research* 44 (D1):D447-D456.

Wada-Katsumata, A., L. Zurek, G. Nalyanya, W.L. Roelofs, A. Zhang *et al.*, 2015 Gut bacteria mediate aggregation in the German cockroach. *Proceedings of the National Academy of Sciences* 112 (51):15678-15683.

Wang, L., B. Elliott, X. Jin, L. Zeng, and J. Chen, 2015 Antimicrobial properties of nest volatiles in red imported fire ants, *Solenopsis invicta* (hymenoptera: formicidae). *The Science of Nature* 102 (11-12):66.

Ware, J.L., D.A. Grimaldi, and M.S. Engel, 2010 The effects of fossil placement and calibration on divergence times and rates: an example from the termites (Insecta: Isoptera). *Arthropod Structure & Development* 39 (2-3):204-219.

Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T.H. Richardson *et al.*, 2007 Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450 (7169):560.

Watt, A.D., N.E. Stork, P. Eggleton, D. Srivastava, B. Bolton *et al.*, 1997 Impact of forest loss and regeneration on insect abundance and diversity. *Forests and insects*:273-286.

Wittkopp, P.J., K. Vaccaro, and S.B. Carroll, 2002 Evolution of yellow gene regulation and pigmentation in Drosophila. *Current Biology* 12 (18):1547-1556.

Wojda, I., 2016 Temperature stress and insect immunity. *Journal of Thermal Biology*.

Xia, X., M. You, X.-J. Rao, and X.-Q. Yu, 2017 Insect C-type lectins in innate immunity. *Developmental & Comparative Immunology*.

Yi, H.-Y., M. Chowdhury, Y.-D. Huang, and X.-Q. Yu, 2014 Insect antimicrobial peptides and their applications. *Applied Microbiology and Biotechnology* 98 (13):5807-5822.

Yu, G., D.Y. Baeder, R.R. Regoes, and J. Rolff, 2016 Combination effects of antimicrobial peptides. *Antimicrobial Agents and Chemotherapy* 60 (3):1717-1724.

Zanchi, C., P.R. Johnston, and J. Rolff, 2017 Evolution of defence cocktails: antimicrobial peptide combinations reduce mortality and persistent infection. *Molecular Ecology*.

Zhang, Z.T., and S.Y. Zhu, 2009 Drosomycin, an essential component of antifungal defence in Drosophila. *Insect Molecular Biology* 18 (5):549-556.

Zhou, X., K. Qian, Y. Tong, J.J. Zhu, X. Qiu *et al.*, 2014 De novo transcriptome of the hemimetabolous German cockroach (Blattella germanica). *PloS One* 9 (9):e106932.

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 (8):R177.

Publications

Shulin He, Paul R. Johnston, Benno Kuropka, Sophie Lokatis, Christoph Weise, Rudy Plarre, Hans-Jörg Kunte, Dino P. McMahon. Termite soldiers contribute to social immunity by synthesizing potent oral secretions. *Insect Molecular Biology*. https://doi.org/10.1111/imb.12499

Shulin He, Lei Song, Zhaoying Qian, Fujun Hou, Yongjie Liu, Xianzong Wang, Zhangming Peng, Chengbo Sun, Xiaolin Liu. Molecular characterization of *LvAV* in response to white spot syndrome virus infection in the Pacific white shrimp (*Litopenaeus vannamei*). *Developmental & Comparative Immunology*. 2015 Jul 1;51(1): 48-55. https://doi.org/10.1016/j.dci.2015.02.020

Shulin He, Zhaoying Qian, Jing Yang, Xianzong Wang, Xiao Mi, Yongjie Liu, Fujun Hou, Qiao Liu, Xiaolin Liu. Molecular characterization of a p38 MAPK from *Litopenaeus vannamei* and its expression during the molt cycle and following pathogen infection, *Developmental & Comparative Immunology*, Volume 41(2):217-221, 2013, http://dx.doi.org/10.1016/j.dci.2013.05.010.

Jakob Trimpert, Nicole Groenke, Maria Jenckel, Shulin He, Dusan Kunec, Moriah L. Szpara, Stephen J. Spatz, Nikolaus Osterrieder, Dino P. McMahon. A phylogenomic analysis of Marek's disease virus reveals independent paths to virulence in Eurasia and North America. *Evolutionary applications*. 2017. https://doi.org/10.1111/eva.12515

Curriculum Vitae

Education

Freie Universität Berlin, Berlin, Germany

PhD	Evolution and Ecology	Sept.2014-present

Northwest A&F University, Yangling, China

Master of Science	Aquatic Biology	Jun. 2014
Bachelor of Agriculture	Animal Science	Jun. 2012

Conference talks and posters

XI European Congress of Entomology 2018, “Using comparative transcriptomics to understand the evolution of immunity in termites” (Talk)

V Central European Meeting IUSSI 2017, “Differentiation of immune response with castes in termites” (Talk)

Congress of the European Society for Evolutionary Biology 2017, “Defend and disinfect: a flexible role for soldiers in termite society” (poster)

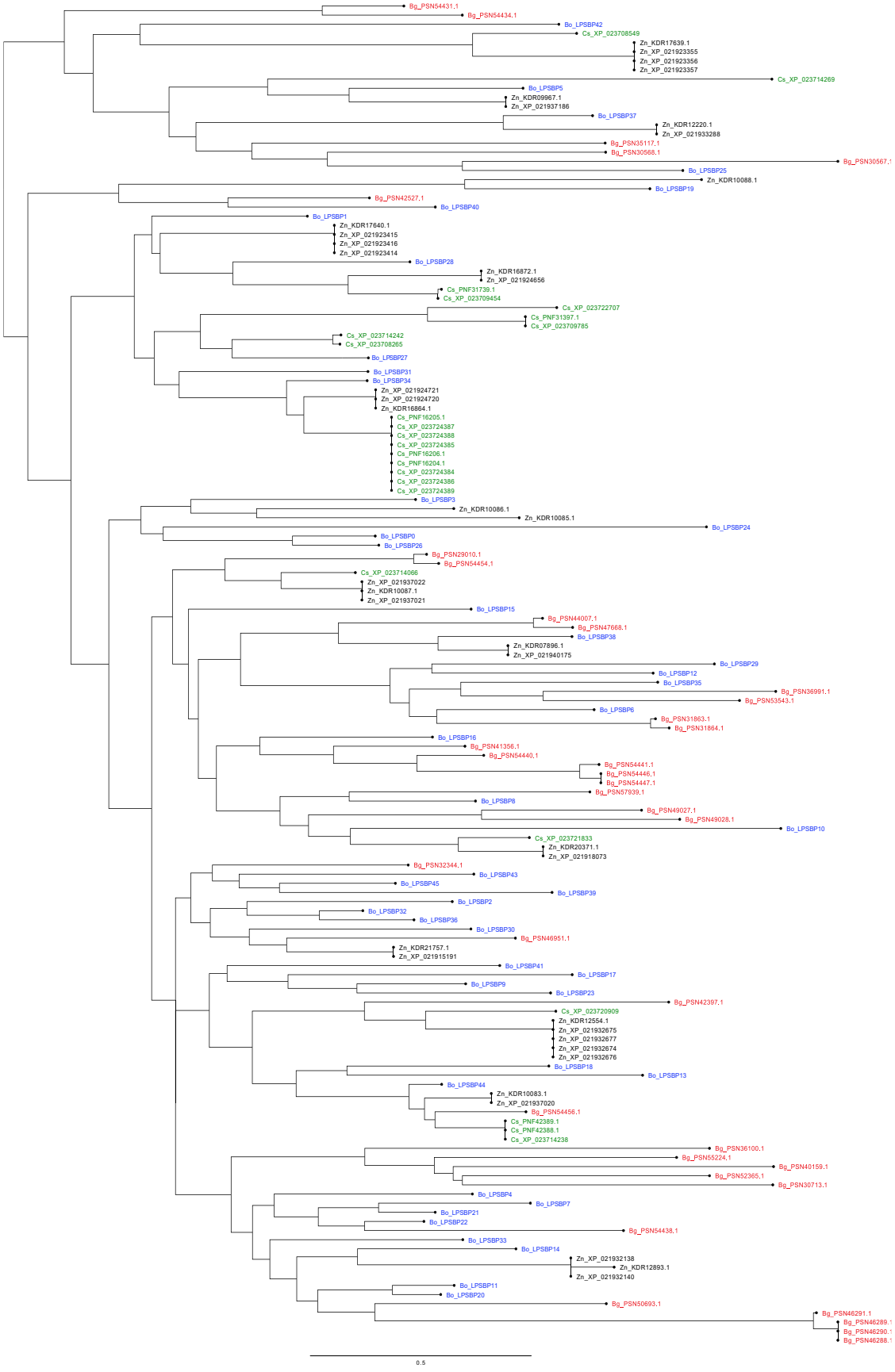
PhD student Workshop “Conflict and cooperation-bridging evolution, ecology and immunology” 2017, “The evolution of immunity during the transition to eusociality” (Talk)

6th European Congress of the IUSSI 2016, “A dual role for soldiers in a complex insect society” (Talk)

Appendix

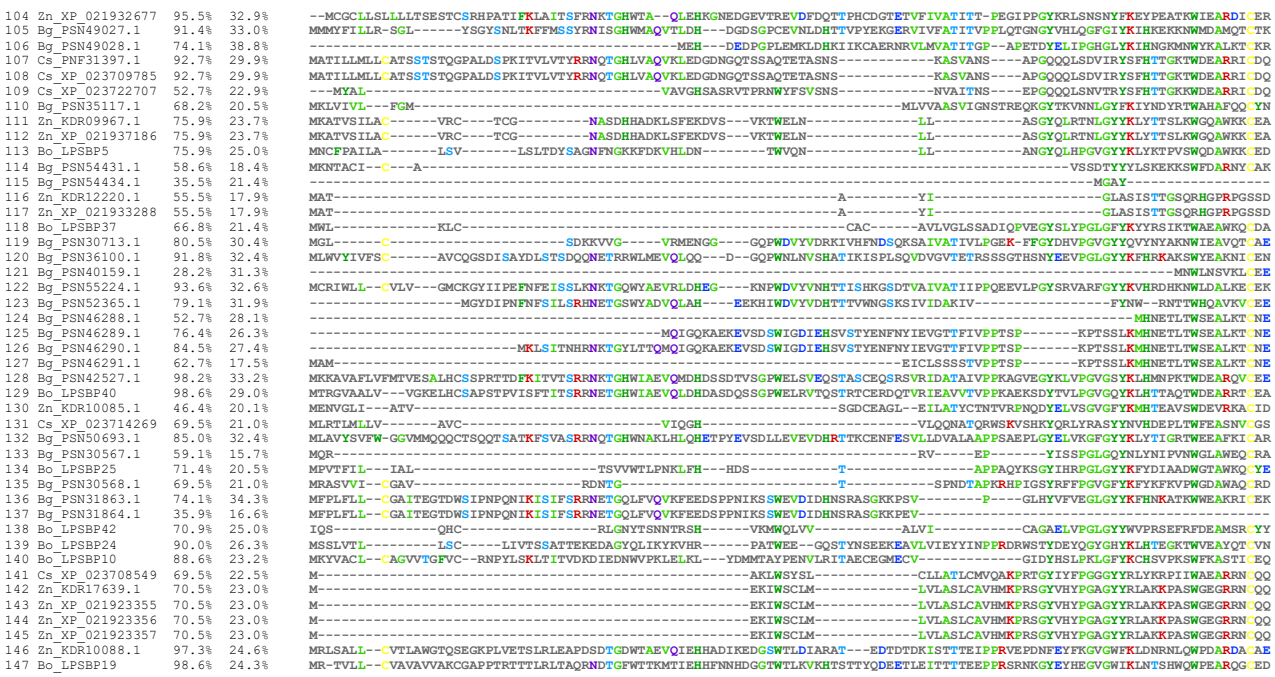
Appendix I-A

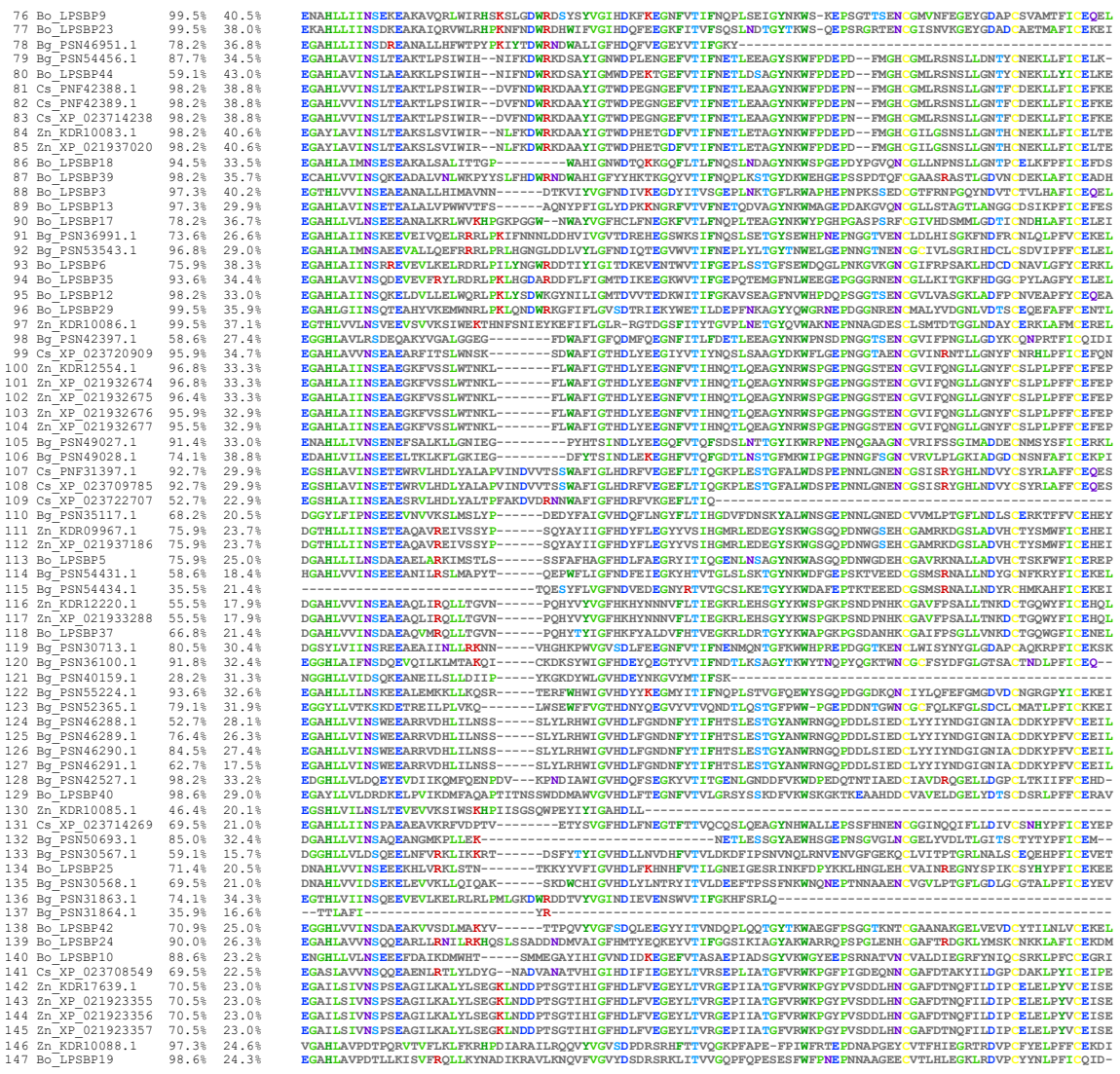
Phylogenetic reconstruction of putative LPSBPs from *B. orientalis, B. germanica, Z. nevadensis,* and *C. secundus*. Predicted protein sequence of *B. orientalis* from our study are named as follows: Bo_LPSBPXX. Sequences of the other four species were downloaded from NCBI, abbreviated as follows: Bg, *B. germanica*; Zn, *Z. nevadensis*; Cs, *C. secundus*. Protein IDs are as given in NCBI. The alignment was performed by MAFFT (Katoh and Standley 2013) and the tree was built by PhyML in SeaView (Gouy *et al*. 2009).



Appendix I-B

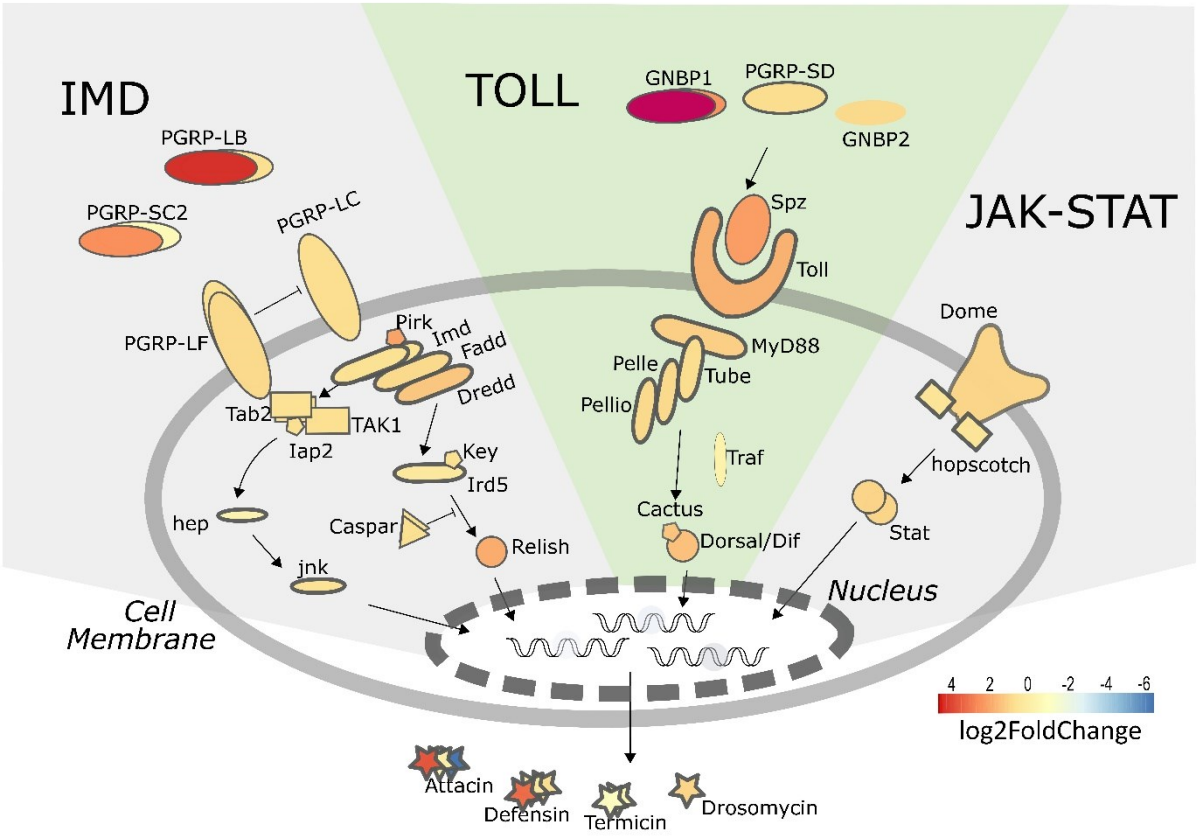
Alignment of putative LPSPBs from *B. orientalis, B. germanica, Z. nevadensis* and *C. secundus* against the reference sequence for *P. americana* (BAA00616.1), with gaps removed using trimAl (Capella-Gutiérrez *et al*. 2009). The region of the alignment containing the predicted C-lectin domain is indicated by dotted sections in the first row above the alignment.





Appendix I-C

Fold changes of the genes related to three main pathways. The overlap in gene families represents the fold changes of different genes in the same gene family, except Toll and Spaetzle family, whose fold changes have been indicated in following Appendix I-E.



Appendix I-D

Details on primers for quantitative PCR for **Chapter I**

Name	Primer	Temperature (°C)	Model Thermocycler
Relish	F:5'-TTCCTGGCTCTACCTGTC-3' R:5'-TTGCAGCTATACCGTCCT-3'	57	Stratagene Mx3005PTM
Attacin	F:5'-ACAGTGGTCGAAGGTGCT-3' R:5'-TTGGGATGAAGATGATTCTG-3'	57	
GNBP1	F:5'-TGGAAATTTGGCTCGTACCTC-3' R:5'-ACGTCTTGAACCCCATAACCT-3'	59	
Transferrin	F:5'-AACTACACGGACGTAATTGAGC-3' R:5'-ACATTTCTCCAGTTCCGTGTC-3'	59	
RPL22	F:5'-CAACAACTCTGAGCCAATC-3' R:5'-GTAAACTCCGACATTCCTT-3'	56[1]/57[2]	
Defensin	F:5'-TTAGCTGCTCCTCTGACA-3' R:5'-GTCTTCCTCTGCTGTGAC-3'	57	
PGRP2	F:5'-GCGGTTGGCACCAGATAG-3' R:5'-AGTTGCTTCGTGGCTTCA-3'	58	
FAS	F:5'-TGCTGGTAGCCCTATGGAA-3' R:5'-TCGTCTGGGAGTCAGTTGG-3'	57	
Tret1	F:5'-GCTGTGATCGGTCCTTGTA-3' R:5'-ATCCCATCGTGACTCCTCT-3'	57	Biorad CFX96 C1000
Lipase 3	F:5'-AGGACCCACGATGACCCAA-3' R:5'-TAACGGCGGACGGCTACTT-3'	57	
MOT13	F:5'-TTGGTGCTATCTTCGTCTT -3' R:5'-CCTAGTCCAGTGCCTTGTA-3'	57	
LIPR2	F:5'-CGCCCATGATTGCAGTAAA-3' R:5'-TCCATAACGACGGACGAAG-3'	57	

Appendix I-E

Comparison of fold changes of the genes in Toll pathway in *D. Oregon, B. orientalis, M. sexta*.

Gene/Gene Family	***Drosophila* Oregon^R (adult male)**(De Gregorio *et al*. 2001)		***B. orientalis(adults)***	***Manduca sexta* (naïve larvae)**(Cao *et al*. 2015)	
	Septic injury (24 hr)	Fungal infection (24 hr)	Our study (24 hr)	Fat body (24 hr)	Hemocytes (24 hr)
GNBP1	-	-	53.8/4.0	-	-
GNBP2	-	-	1.3	-	-
PGRP-SD	9.5	1.4	1.2	-	-
Spaetzle	1.8	1.5	0.5/3.3/1.1/0.9/1.0/1.4/0.6	1.4/1.5/4.1	3.9/2.2/3.6
Toll	2.3	1.3	0.9/0.6/0.9/1.3/2.5/2.3/1.2/0.8/1.1/1.1/0.8	2.5/0.7/2.5/2.5/2.9	6.2/0.8/6.2/6.4/0.9
MyD88	-	-	1.6	1.5	1.5
Traf6	-	-	1.0	-	-
Pelle	-	-	1.4	5.1	2.2
Cactus	3.7	2.1	1.8	9.2	1.8
Dif/Drosal	1.4/2.2	1.0	1.9	1.3	1.2
Tube	-	-	1.1	8.4	0.8
Pellino	-	-	1.4	2.3	1.3

Note: The multiple values in cells represent the fold changes of different genes in the same gene family.

Appendix I-F

Predicted immune-related genes in *B. orientalis*.

Subcomponent ID	Famliy Name	Gene Name	ORF type (Predicted protein)	Predicted protein length	Target ID (BLASTX,Trinotate)	E-value (BLASTX, Trinotate)
TRINITY_DN203797_c6_g1_i2	Apaf-caspas	Apoptotic protease-activating factor 1	complete	1399	APAF_MOUSE	2.1E-88
TRINITY_DN200927_c0_g1_i2	ATG12	Autophagy protein 12-like	complete	133	APG12_DROME	2.04E-42
TRINITY_DN210395_c7_g1_i1	ATG13	Autophagy-related protein 13 homolog	complete	404	ATG13_DROME	4.92E-73
TRINITY_DN207592_c4_g3_i4	ATG14	Beclin 1-associated autophagy-related key regulator	5prime_partial	495	BAKOR_HUMAN	4.44E-76
TRINITY_DN206239_c0_g1_i1	ATG14	UV radiation resistance associated protein	complete	880	UVRAG_MOUSE	5.11E-76
TRINITY_DN209538_c7_g1_i2	ATG18B	WD repeat domain phosphoinositide-interacting protein 2	complete	461	WIPI2_XENLA	0
TRINITY_DN209743_c1_g1_i1	ATG18B	WD repeat domain phosphoinositide-interacting protein 3	complete	345	WIPI3_XENLA	0
TRINITY_DN204622_c2_g1_i2	ATG18B	WD repeat domain phosphoinositide-interacting protein 4	complete	354	WIPI4_DANRE	1.69E-150
TRINITY_DN211938_c7_g1_i2	ATG2	Autophagy-related protein 2 homolog B	complete	2186	ATG2B_MOUSE	0
TRINITY_DN206859_c0_g1_i1	ATG3	Ubiquitin-like-conjugating enzyme ATG10	5prime_partial	128	ATG10_HUMAN	8.64E-33
TRINITY_DN203305_c6_g1_i2	ATG3	Ubiquitin-like-conjugating enzyme ATG3	complete	317	ATG3_BOVIN	9.78E-147
TRINITY_DN203283_c6_g1_i1	ATG4b	Cysteine protease ATG4D	complete	434	ATG4D_MOUSE	8.21E-117
TRINITY_DN207415_c8_g2_i1	ATG5	Autophagy protein 5	complete	265	ATG5_BOVIN	1.54E-108
TRINITY_DN199491_c0_g1_i1	ATG6	Beclin-1-like protein	complete	429	BECN1_DROME	1.05E-166
TRINITY_DN208632_c6_g1_i2	ATG7	Ubiquitin-like modifier-activating enzyme ATG7	complete	735	ATG7_MOUSE	0
TRINITY_DN208319_c2_g1_i1	ATG8	Gamma-aminobutyric acid receptor-associated protein	complete	118	GBRAP_RAT	1.13E-71
TRINITY_DN211417_c2_g1_i1	ATG9	Autophagy-related protein 9A	complete	814	ATG9A_HUMAN	0
TRINITY_DN198255_c0_g1_i2	Attacin	Attacin-A	complete	217	ATTA_DROME	3.5E-09
TRINITY_DN144643_c0_g1_i1	Attacin	Attacin-B	5prime_partial	139	ATTB_DROME	0.000457
TRINITY_DN207862_c0_g2_i1	Attacin	Holotricin-2	complete	120		
TRINITY_DN212656_c6_g1_i7	Cactus_Toll	NF-kappa-B inhibitor cactus	complete	448	CACT_DROME	8.97E-63
TRINITY_DN204960_c1_g1_i4	Caspar_IMD	FAS-associated factor 1	complete	670	FAF1_HUMAN	2.09E-166
TRINITY_DN204197_c0_g1_i1	Caspar_IMD	FAS-associated factor 2	complete	444	FAF2_XENTR	3.44E-140
TRINITY_DN200568_c1_g1_i1	CASPs	Caspase-1-1	complete	305	CASP1_DROME	5.9E-43
TRINITY_DN200683_c1_g1_i2	CASPs	Caspase-1-2	complete	491	CASP1_SPOFR	1.05E-57

(table continued on next page)

(table continued from previous page)

TRINITY_DN200765_c0_g1_i1	CASPs	Caspase-1-3	complete	368	CASP1_SPOFR	1.02E-41
TRINITY_DN202202_c0_g1_i2	CASPs	Caspase-1-4	complete	303	CASP1_SPOFR	1.6E-84
TRINITY_DN207137_c3_g1_i2	CASPs	Caspase-1-5	complete	468	CASP1_SPOFR	6.07E-82
TRINITY_DN211317_c5_g1_i3	CASPs	Caspase-1-6	complete	290	CASP1_SPOFR	6.46E-125
TRINITY_DN207884_c5_g1_i1	CASPs	Caspase-2	complete	426	CASP2_CHICK	9.97E-41
TRINITY_DN211335_c4_g1_i3	CASPs	Caspase-8	5prime_partial	648	CASP8_DROPS	2.03E-63
TRINITY_DN200371_c0_g1_i1	CATs	Catalase-1	complete	229	CATA_DROME	1.64E-113
TRINITY_DN206745_c0_g1_i1	CATs	Catalase-2	5prime_partial	546	CATA_RUGRU	0
TRINITY_DN209411_c5_g4_i1	CATs	Catalase-3	5prime_partial	163	CATA_BOVIN	6.07E-50
TRINITY_DN209411_c5_g5_i2	CATs	Catalase-4	5prime_partial	539	CATA_PIG	0
TRINITY_DN210101_c1_g1_i1	CATs	Catalase-5	complete	509	CATA_RUGRU	0
TRINITY_DN212150_c1_g1_i1	CATs	Catalase-6	complete	509	CATA_DROME	0
TRINITY_DN89736_c0_g1_i1	CATs	Catalase-7	5prime_partial	159	CATA_ASCSU	2.01E-27
TRINITY_DN200662_c1_g1_i1	CLIPs	Cationic trypsin-1	5prime_partial	271	TRY3_RAT	3.18E-49
TRINITY_DN209414_c10_g1_i2	CLIPs	Cationic trypsin-2	complete	296	TRY1_CANLF	1.72E-15
TRINITY_DN194388_c0_g1_i1	CLIPs	Chymotrypsin BI-1	complete	276	CTRB1_LITVA	2.37E-61
TRINITY_DN200811_c0_g1_i1	CLIPs	Chymotrypsin BI-2(CLIP-7)	5prime_partial	297	CTRB1_LITVA	5.21E-70
TRINITY_DN204403_c0_g1_i3	CLIPs	Chymotrypsin BI-3	complete	313	CTRB1_LITVA	7.72E-63
TRINITY_DN207574_c0_g1_i2	CLIPs	Chymotrypsin BI-4	5prime_partial	266	CTRB1_LITVA	2.8E-50
TRINITY_DN202663_c0_g1_i2	CLIPs	Chymotrypsin-1-1	internal	246	CTR1_SOLIN	1.94E-27
TRINITY_DN204325_c10_g3_i1	CLIPs	Chymotrypsin-1-2	complete	283	CTR1_SOLIN	1.37E-13
TRINITY_DN201014_c0_g1_i2	CLIPs	Chymotrypsin-2	5prime_partial	255	CTR2_VESCR	5.49E-47
TRINITY_DN202780_c0_g1_i1	CLIPs	Chymotrypsin-C	complete	267	CTRC_HUMAN	1.71E-25
TRINITY_DN212694_c2_g1_i3	CLIPs	Coagulation factor X	complete	289	FA10_CHICK	1.29E-43
TRINITY_DN205643_c7_g1_i3	CLIPs	Coagulation factor XII	complete	316	FA12_PIG	2.98E-44
TRINITY_DN198392_c0_g1_i1	CLIPs	Kallikrein-13(CLIP-11)	5prime_partial	332	KLK13_HUMAN	6.76E-18
TRINITY_DN198284_c0_g2_i1	CLIPs	Limulus clotting factor C	complete	291	LFC_CARRO	1.28E-47
TRINITY_DN207071_c1_g2_i1	CLIPs	Limulus clotting factor C(CLIP-3)	complete	620	LFC_CARRO	5.35E-39
TRINITY_DN177253_c0_g2_i1	CLIPs	Plasma kallikrein-1	5prime_partial	311	KLKB1_BOVIN	4.19E-32
TRINITY_DN204587_c3_g1_i2	CLIPs	Plasma kallikrein-2	5prime_partial	308	KLKB1_BOVIN	9.75E-54

(table continued on next page)

(table continued from previous page)

TRINITY_DN210028_c8_g1_i1	CLIPs	Plasma kallikrein-3	complete	309	KLKB1_HUMAN	3.93E-46
TRINITY_DN197303_c0_g2_i5	CLIPs	Proclotting enzyme-1	complete	211	PCE_TACTR	2.5E-26
TRINITY_DN201210_c0_g2_i1	CLIPs	Proclotting enzyme-2	complete	297	PCE_TACTR	1.49E-45
TRINITY_DN207232_c6_g2_i1	CLIPs	Proclotting enzyme-3	complete	306	PCE_TACTR	4.57E-45
TRINITY_DN43074_c1_g1_i1	CLIPs	Proclotting enzyme-4	internal	104	PCE_TACTR	3.69E-21
TRINITY_DN201196_c3_g1_i1	CLIPs	Proclotting enzyme(CLIP-10)	complete	328	PCE_TACTR	3.82E-46
TRINITY_DN203423_c3_g1_i1	CLIPs	Proclotting enzyme(CLIP-8)	5prime_partial	461	PCE_TACTR	3.03E-62
TRINITY_DN201352_c0_g1_i2	CLIPs	Putative serine protease 41	5prime_partial	574	PRS41_HUMAN	8.57E-34
TRINITY_DN200535_c0_g1_i4	CLIPs	Retinol dehydrogenase 14	complete	261	RDH14_HUMAN	6.05E-62
TRINITY_DN204701_c0_g1_i1	CLIPs	Serine protease 44	complete	371	PRS44_MOUSE	9.92E-42
TRINITY_DN199952_c0_g1_i1	CLIPs	Serine protease 48	complete	259	PRS48_HUMAN	3.85E-23
TRINITY_DN199209_c0_g2_i1	CLIPs	Serine protease easter-4	complete	355	EAST_DROME	1.14E-28
TRINITY_DN203131_c1_g1_i1	CLIPs	Serine protease easter-5	complete	360	EAST_DROME	7.42E-46
TRINITY_DN204331_c12_g1_i4	CLIPs	Serine protease easter-6	complete	308	EAST_DROME	3.52E-46
TRINITY_DN34868_c0_g1_i1	CLIPs	Serine protease easter-7	internal	146	EAST_DROME	3.65E-14
TRINITY_DN210614_c3_g1_i2	CLIPs	Serine protease easter-1	complete	418	EAST_DROME	4.34E-84
TRINITY_DN206030_c8_g1_i2	CLIPs	Serine protease easter-2	5prime_partial	399	EAST_DROME	4.8E-74
TRINITY_DN205038_c16_g1_i1	CLIPs	Serine protease easter-3	internal	531	EAST_DROME	3.83E-14
TRINITY_DN103011_c0_g1_i1	CLIPs	Serine protease hepsin	internal	102	HEPS_RAT	4.61E-09
TRINITY_DN191962_c0_g1_i1	CLIPs	Serine protease snake	complete	323	SNAK_DROME	3.87E-64
TRINITY_DN205149_c0_g1_i2	CLIPs	Serine protease snake(CLIP-2)	5prime_partial	392	SNAK_DROME	7.76E-70
TRINITY_DN203899_c0_g1_i1	CLIPs	Serine protease snake(CLIP-4)	complete	394	SNAK_DROME	1.33E-64
TRINITY_DN212041_c0_g2_i1	CLIPs	Serine protease snake(CLIP-5)	complete	352	SNAK_DROME	8.49E-54
TRINITY_DN202525_c0_g1_i1	CLIPs	Serine protease snake(CLIP-9)	5prime_partial	376	SNAK_DROME	1.48E-43
TRINITY_DN120666_c0_g1_i1	CLIPs	Serine proteinase stubble-1	5prime_partial	211	STUB_DROME	1.43E-16
TRINITY_DN180733_c0_g1_i1	CLIPs	Serine proteinase stubble-2	5prime_partial	307	STUB_DROME	4.25E-142
TRINITY_DN192813_c2_g1_i1	CLIPs	Serine proteinase stubble-3	internal	103	STUB_DROME	3.09E-16
TRINITY_DN199205_c2_g1_i1	CLIPs	Serine proteinase stubble-4	internal	130	STUB_DROME	3.76E-20
TRINITY_DN207404_c7_g1_i4	CLIPs	Serine proteinase stubble-5	complete	408	STUB_DROME	1.04E-38
TRINITY_DN211676_c1_g1_i1	CLIPs	Serine proteinase stubble-6	5prime_partial	396	STUB_DROME	2.37E-26

(table continued on next page)

(table continued from previous page)

TRINITY_DN205845_c4_g1_i1	CLIPs	Testisin	complete	271	TEST_MOUSE	1.23E-33
TRINITY_DN195636_c1_g1_i1	CLIPs	Transmembrane protease serine 11B-like protein	internal	144	TM11B_MOUSE	3.85E-12
TRINITY_DN213919_c0_g3_i1	CLIPs	Transmembrane protease serine 11G	complete	258	TM11G_RAT	1.94E-34
TRINITY_DN186830_c0_g1_i1	CLIPs	Transmembrane protease serine 3	complete	295	TMPS3_MOUSE	1.58E-30
TRINITY_DN202673_c1_g1_i1	CLIPs	Trypsin	complete	260	TRYP_PHACE	1.03E-38
TRINITY_DN205799_c16_g1_i1	CLIPs	Trypsin 3A1	complete	265	TRY3_AEDAE	5.19E-62
TRINITY_DN199291_c0_g1_i2	CLIPs	Trypsin 5G1	5prime_partial	253	TRY5_AEDAE	1.58E-21
TRINITY_DN146385_c0_g1_i1	CLIPs	Trypsin eta	3prime_partial	157	TRYU_DROER	2.23E-20
TRINITY_DN28487_c0_g2_i1	CLIPs	Trypsin eta	5prime_partial	150	TRYU_DROER	1.43E-24
TRINITY_DN116125_c0_g1_i1	CLIPs	Trypsin II-P29	5prime_partial	230	TRY3_CHICK	1.26E-47
TRINITY_DN198412_c0_g2_i1	CLIPs	Trypsin zeta	complete	263	TRYZ_DROME	5.59E-28
TRINITY_DN138339_c0_g1_i1	CLIPs	Trypsin-1-1	5prime_partial	282	TRYP_NEOBL	7.18E-43
TRINITY_DN165791_c2_g1_i1	CLIPs	Trypsin-1-2	internal	164	TRYP_ASTAS	1.74E-39
TRINITY_DN190257_c0_g1_i1	CLIPs	Trypsin-1-3	5prime_partial	263	TRYDG_DROME	1.46E-52
TRINITY_DN194806_c1_g1_i1	CLIPs	Trypsin-1-4	5prime_partial	159	TRY1_ANOGA	3.77E-41
TRINITY_DN201020_c2_g1_i1	CLIPs	Trypsin-1-5	5prime_partial	266	TRYP_ASTAS	6.01E-46
TRINITY_DN201073_c0_g1_i1	CLIPs	Trypsin-1-6	5prime_partial	264	TRY1_ANOGA	3.17E-69
TRINITY_DN201373_c1_g1_i1	CLIPs	Trypsin-1-7	5prime_partial	301	TRYP_ASTAS	1.68E-48
TRINITY_DN202628_c0_g1_i1	CLIPs	Trypsin-1-8	5prime_partial	314	TRY1_ANOGA	1.63E-59
TRINITY_DN202673_c0_g1_i1	CLIPs	Trypsin-1-9	5prime_partial	260	TRY1_ANOGA	1.03E-43
TRINITY_DN202753_c4_g1_i1	CLIPs	Trypsin-1-10	5prime_partial	176	TRYP_ASTAS	2.22E-31
TRINITY_DN203473_c0_g2_i1	CLIPs	Trypsin-1-11	5prime_partial	293	TRYP_ASTAS	1.19E-47
TRINITY_DN203701_c0_g1_i2	CLIPs	Trypsin-1-12	5prime_partial	273	TRY1_ANOGA	1.56E-54
TRINITY_DN204594_c1_g1_i3	CLIPs	Trypsin-1-13	5prime_partial	248	TRYP_ASTAS	3.24E-48
TRINITY_DN205704_c1_g1_i1	CLIPs	Trypsin-1-14	5prime_partial	260	TRY1_ANOGA	3.62E-61
TRINITY_DN210519_c2_g1_i1	CLIPs	Trypsin-1-15	complete	261	TRYP_ASTAS	2.25E-44
TRINITY_DN211373_c1_g1_i1	CLIPs	Trypsin-1-16	5prime_partial	262	TRY1_ANOGA	8.15E-60
TRINITY_DN229351_c0_g1_i1	CLIPs	Trypsin-1-17	5prime_partial	169	TRYP_ASTAS	3.7E-44
TRINITY_DN203274_c4_g3_i1	CLIPs	Trypsin-2-1	5prime_partial	125	TRY2_SALSA	1.39E-24
TRINITY_DN204349_c0_g1_i1	CLIPs	Trypsin-2-2	5prime_partial	287	TRY2_ANOGA	1.21E-42

(table continued on next page)

(table continued from previous page)

TRINITY_DN164285_c1_g1_i1	CLIPs	Trypsin-3-1	5prime_partial	256	TRY3_ANOGA	1.33E-45
TRINITY_DN198524_c0_g1_i1	CLIPs	Trypsin-3-2	5prime_partial	251	TRY3_ANOGA	1.09E-35
TRINITY_DN199137_c0_g1_i1	CLIPs	Trypsin-3-3	5prime_partial	269	TRY2_ANOGA	1.21E-52
TRINITY_DN202672_c0_g1_i2	CLIPs	Trypsin-3-4	complete	257	TRY3_ANOGA	1.68E-61
TRINITY_DN207975_c4_g4_i3	CLIPs	Trypsin-3-5	complete	261	TRY3_ANOGA	6.52E-54
TRINITY_DN198196_c0_g1_i1	CLIPs	Trypsin-4	complete	269	TRY4_ANOGA	6.74E-21
TRINITY_DN211048_c3_g1_i1	CLIPs	Trypsin-5(CLIP-1)	5prime_partial	261	TRY5_ANOGA	1.22E-41
TRINITY_DN168020_c0_g1_i1	CLIPs	Trypsin-7-1	internal	126	TRY7_ANOGA	2.07E-28
TRINITY_DN168098_c0_g1_i1	CLIPs	Trypsin-7-2	3prime_partial	250	TRY4_ANOGA	1.73E-48
TRINITY_DN197824_c0_g1_i1	CLIPs	Trypsin-7-3	5prime_partial	232	TRY7_ANOGA	2.86E-39
TRINITY_DN201441_c0_g1_i1	CLIPs	Trypsin-7-4	5prime_partial	268	TRY7_ANOGA	4.32E-61
TRINITY_DN201757_c0_g1_i1	CLIPs	Trypsin-7-5	5prime_partial	238	TRY7_ANOGA	2.23E-57
TRINITY_DN202314_c0_g1_i3	CLIPs	Trypsin-7-6	complete	265	TRY7_ANOGA	1.61E-26
TRINITY_DN205251_c0_g1_i1	CLIPs	Trypsin-7-7	5prime_partial	259	TRY1_ANOGA	3.49E-60
TRINITY_DN205378_c3_g1_i2	CLIPs	Trypsin-7-8	complete	261	TRY7_ANOGA	6.43E-58
TRINITY_DN205922_c0_g1_i1	CLIPs	Trypsin-7-9	complete	286	TRY7_ANOGA	7.29E-43
TRINITY_DN206189_c5_g1_i1	CLIPs	Trypsin-7-10	5prime_partial	140	TRY7_ANOGA	1.36E-39
TRINITY_DN209682_c5_g1_i1	CLIPs	Trypsin-7-11	5prime_partial	266	TRY7_ANOGA	1.34E-29
TRINITY_DN209701_c4_g2_i1	CLIPs	Trypsin-7-12	complete	259	TRY7_ANOGA	1.08E-43
TRINITY_DN211152_c0_g1_i3	CLIPs	Trypsin-7-13	complete	254	TRY7_ANOGA	1.06E-63
TRINITY_DN202673_c1_g2_i2	CLIPs	Trypsin-7(CLIP-6)	complete	260	TRY7_ANOGA	9.99E-37
TRINITY_DN198138_c1_g1_i2	CLIPs	Venom protease	5prime_partial	300	SP4_BOMPE	1.04E-42
TRINITY_DN202245_c0_g1_i1	CLIPs	Venom serine protease 34-1	5prime_partial	293	SP34_APIME	7.72E-67
TRINITY_DN208065_c2_g1_i1	CLIPs	Venom serine protease 34-2	complete	395	SP34_APIME	5.16E-74
TRINITY_DN210801_c1_g1_i3	CLIPs	Venom serine protease Bi-VSP-1	complete	378	VSP_BOMIG	1.73E-103
TRINITY_DN206428_c4_g1_i3	CLIPs	Venom serine protease Bi-VSP-2	5prime_partial	320	VSP_BOMIG	1.87E-72
TRINITY_DN210649_c2_g1_i1	CTLs	Collectin-12	internal	150	COL12_RAT	0.000000173
TRINITY_DN210649_c5_g1_i2	CTLs	C-type lectin domain family 4 member E	3prime_partial	253	MRC2_MOUSE	1.93E-09
TRINITY_DN213404_c0_g1_i1	CTLs	C-type lectin mannose-binding isoform	complete	194	LECM_OXYSU	1.34E-13
TRINITY_DN199675_c1_g1_i1	CTLs	C-type mannose receptor 2	3prime_partial	975	MRC2_HUMAN	7.27E-09

(table continued on next page)

(table continued from previous page)

TRINITY_DN207865_c5_g1_i3	CTLs	Galactose-specific lectin nattectin	complete	193	LECG_THANI	3.33E-08
TRINITY_DN137749_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-7	5prime_partial	140	LPSBP_PERAM	1.05E-39
TRINITY_DN143323_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-8	5prime_partial	117	LPSBP_PERAM	1.39E-33
TRINITY_DN167970_c3_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-9	5prime_partial	140	LPSBP_PERAM	5.44E-24
TRINITY_DN183662_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-10	complete	231	LPSBP_PERAM	8.06E-47
TRINITY_DN186074_c1_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-11	5prime_partial	139	LPSBP_PERAM	1.02E-33
TRINITY_DN191904_c1_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-12	5prime_partial	172	LPSBP_PERAM	6.15E-27
TRINITY_DN192635_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-13	complete	226	LPSBP_PERAM	3.13E-29
TRINITY_DN193512_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-14	5prime_partial	189	LPSBP_PERAM	1.46E-35
TRINITY_DN199055_c0_g1_i3	CTLs	Hemolymph lipopolysaccharide-binding protein-15	complete	244	LPSBP_PERAM	5.74E-49
TRINITY_DN200685_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-16	complete	218	LPSBP_PERAM	1.45E-21
TRINITY_DN200789_c0_g1_i2	CTLs	Hemolymph lipopolysaccharide-binding protein-17	5prime_partial	243	LPSBP_PERAM	1.17E-45
TRINITY_DN201843_c0_g1_i3	CTLs	Hemolymph lipopolysaccharide-binding protein-18	complete	228	LPSBP_PERAM	1.5E-32
TRINITY_DN203168_c1_g1_i2	CTLs	Hemolymph lipopolysaccharide-binding protein-19	5prime_partial	241	LPSBP_PERAM	3.87E-46
TRINITY_DN203647_c0_g1_i2	CTLs	Hemolymph lipopolysaccharide-binding protein-20	5prime_partial	201	LPSBP_PERAM	7.04E-43
TRINITY_DN203978_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-21	complete	226	LPSBP_PERAM	2.25E-38
TRINITY_DN204072_c2_g4_i2	CTLs	Hemolymph lipopolysaccharide-binding protein-22	5prime_partial	181	LPSBP_PERAM	2.47E-32
TRINITY_DN204436_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-23	complete	234	LPSBP_PERAM	3.98E-37
TRINITY_DN204569_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-24	5prime_partial	367	LPSBP_PERAM	4.12E-10
TRINITY_DN204627_c1_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-25	5prime_partial	255	LPSBP_PERAM	6.64E-51
TRINITY_DN204859_c1_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-26	internal	163	LPSBP_PERAM	1.49E-32
TRINITY_DN204859_c1_g2_i2	CTLs	Hemolymph lipopolysaccharide-binding protein-27	5prime_partial	240	LPSBP_PERAM	1.18E-43
TRINITY_DN205179_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-28	5prime_partial	260	LPSBP_PERAM	4.67E-43
TRINITY_DN205615_c0_g2_i4	CTLs	Hemolymph lipopolysaccharide-binding protein-29	complete	294	LPSBP_PERAM	7.05E-19
TRINITY_DN206020_c8_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-30	complete	230	LPSBP_PERAM	6.3E-57
TRINITY_DN206615_c15_g1_i4	CTLs	Hemolymph lipopolysaccharide-binding protein-31	complete	257	LPSBP_PERAM	1.15E-153
TRINITY_DN207869_c4_g1_i3	CTLs	Hemolymph lipopolysaccharide-binding protein-32	complete	235	LPSBP_PERAM	8.47E-37
TRINITY_DN207877_c7_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-33	complete	233	LPSBP_PERAM	4.6E-53
TRINITY_DN208497_c3_g2_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-34	complete	239	LPSBP_PERAM	8.64E-71
TRINITY_DN208586_c1_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-35	complete	223	LPSBP_PERAM	6.88E-54

(table continued on next page)

(table continued from previous page)

TRINITY_DN208704_c1_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-36	complete	232	LPSBP_PERAM	1.22E-45
TRINITY_DN209415_c7_g1_i5	CTLs	Hemolymph lipopolysaccharide-binding protein-37	complete	224	LPSBP_PERAM	4.79E-28
TRINITY_DN210009_c2_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-38	complete	227	LPSBP_PERAM	1.95E-49
TRINITY_DN210940_c11_g1_i3	CTLs	Hemolymph lipopolysaccharide-binding protein-39	complete	167	LPSBP_PERAM	2.91E-19
TRINITY_DN211010_c3_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-40	5prime_partial	152	LPSBP_PERAM	5.22E-24
TRINITY_DN212295_c3_g1_i3	CTLs	Hemolymph lipopolysaccharide-binding protein-41	complete	232	LPSBP_PERAM	5.11E-38
TRINITY_DN212999_c2_g3_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-42	complete	236	LPSBP_PERAM	2.6E-48
TRINITY_DN213121_c0_g3_i5	CTLs	Hemolymph lipopolysaccharide-binding protein-43	5prime_partial	317	LPSBP_PERAM	1.28E-17
TRINITY_DN213148_c7_g1_i2	CTLs	Hemolymph lipopolysaccharide-binding protein-44	5prime_partial	243	LPSBP_PERAM	8.17E-50
TRINITY_DN214096_c6_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-45	5prime_partial	135	LPSBP_PERAM	2.34E-23
TRINITY_DN277272_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-46	5prime_partial	242	LPSBP_PERAM	3.41E-47
TRINITY_DN209100_c1_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-1	complete	238	LPSBP_PERAM	9.94E-73
TRINITY_DN207716_c2_g2_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-2	complete	240	LPSBP_PERAM	8.63E-60
TRINITY_DN212540_c0_g1_i6	CTLs	Hemolymph lipopolysaccharide-binding protein-3	5prime_partial	317	LPSBP_PERAM	6.66E-39
TRINITY_DN205710_c2_g3_i3	CTLs	Hemolymph lipopolysaccharide-binding protein-4	complete	178	LPSBP_PERAM	1.71E-18
TRINITY_DN202299_c0_g2_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-5	complete	233	LPSBP_PERAM	6.16E-29
TRINITY_DN190586_c0_g1_i1	CTLs	Hemolymph lipopolysaccharide-binding protein-6	complete	184	LPSBP_PERAM	4.65E-27
TRINITY_DN200869_c3_g1_i1	CTLs	L-selectin	5prime_partial	244	LYAM1_RAT	0.000326
TRINITY_DN201319_c0_g1_i1	CTLs	Snaclec agglucetin subunit alpha-1	internal	166	SLA1_DEIAC	0.0000166
TRINITY_DN210649_c10_g1_i1	CTLs	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	5prime_partial	1703	SVEP1_HUMAN	4.37E-13
TRINITY_DN199108_c0_g1_i1	DEFs	Defensin	5prime_partial	92	DEFI_ORYRH	1.2E-10
TRINITY_DN200357_c0_g1_i1	DEFs	Defensin-2	complete	73	DEFI_ORYRH	1.21E-10
TRINITY_DN138632_c0_g1_i1	DEFs	Holotricin-1	5prime_partial	90	DEF1_HOLDI	2.12E-10
TRINITY_DN203850_c0_g1_i2	destabilase	Lysozyme-3	5prime_partial	154	LYS_CRAGI	2.25E-09
TRINITY_DN210429_c5_g1_i4	destabilase	Lysozyme-4	5prime_partial	167	LYS_CRAGI	7.2E-09
TRINITY_DN187110_c0_g1_i1	destabilase	Lysozyme-7	5prime_partial	162	LYS_MERLU	0.00000242
TRINITY_DN199333_c0_g2_i1	destabilase	Lysozyme-8	5prime_partial	168	LYS_MERLU	7.07E-09
TRINITY_DN205389_c7_g4_i3	destabilase	Lysozyme-9	complete	148	LYS3_CRAVI	0.0000012
TRINITY_DN210346_c2_g2_i4	destabilase	Lysozyme-1	complete	156	LYS_OSTED	6.48E-12
TRINITY_DN207842_c0_g1_i5	Dif_Toll	Embryonic polarity protein dorsal	complete	795	DORS_DROME	5.94E-156

(table continued on next page)

(table continued from previous page)

TRINITY_DN210555_c5_g1_i5	Domeless_JAK-STAT	Cytokine receptor	complete	1027	DOME_DROME	5.64E-52
TRINITY_DN210445_c4_g2_i4	DRSs	Drosomycin	complete	67	DMYC_DROME	7.37E-20
TRINITY_DN207385_c5_g3_i3	Fadd_IMD	Fas-associated death domain protein	complete	229	FADD_DROME	1.2E-10
TRINITY_DN166725_c0_g2_i1	FREPs	Angiopoietin-related protein 1	5prime_partial	274	ANGL1_HUMAN	4.46E-49
TRINITY_DN29572_c1_g1_i1	FREPs	Protein scabrous	internal	312	SCA_DROME	3.4E-69
TRINITY_DN203196_c2_g1_i1	FREPs	Techylectin-5A	complete	726	TL5A_TACTR	6.55E-48
TRINITY_DN206797_c12_g1_i1	FREPs	Techylectin-5B	internal	101	TL5B_TACTR	4.46E-25
TRINITY_DN203975_c0_g1_i1	GALEs	32 kDa beta-galactoside-binding lectin-1	5prime_partial	396	LEG1_HAECO	1.08E-19
TRINITY_DN204225_c6_g1_i1	GALEs	32 kDa beta-galactoside-binding lectin-2	5prime_partial	327	LEG1_HAECO	1.28E-40
TRINITY_DN207109_c1_g1_i3	GALEs	32 kDa beta-galactoside-binding lectin-3	complete	509	LEG1_HAECO	1.71E-34
TRINITY_DN203081_c1_g1_i1	GALEs	Galectin-4-1	complete	301	LEG5_RAT	6.34E-29
TRINITY_DN205412_c1_g1_i1	GALEs	Galectin-4-2	complete	322	LEG4_MOUSE	4.66E-33
TRINITY_DN201583_c0_g1_i1	GNBP	Beta-1,3-glucan-binding protein	5prime_partial	363	BGBP_PENMO	2.96E-80
TRINITY_DN204546_c3_g3_i3	GNBP	Beta-1,3-glucan-binding protein	complete	352	BGBP_PENMO	2.45E-77
TRINITY_DN208082_c3_g2_i1	GNBP	Beta-1,3-glucan-binding protein	5prime_partial	395	BGBP_PENMO	8.61E-111
TRINITY_DN209559_c7_g2_i7	GNBP	Beta-1,3-glucan-binding protein	complete	353	BGBP_PENMO	1.5E-71
TRINITY_DN210026_c1_g1_i1	GNBP	Beta-1,3-glucan-binding protein	5prime_partial	209	BGBP_PENMO	2.12E-50
TRINITY_DN210026_c2_g1_i3	GNBP	Beta-1,3-glucan-binding protein	5prime_partial	370	BGBP_PENMO	2.51E-82
TRINITY_DN213231_c6_g1_i4	GNBP	Beta-1,3-glucan-binding protein	5prime_partial	384	BGBP_PENMO	1.33E-106
TRINITY_DN213231_c4_g1_i1	GNBP	Beta-1,3-glucan-binding protein (GNBP1)	internal	183	BGBP_PENMO	2.64E-45
TRINITY_DN209017_c1_g1_i4	GNBP	Beta-1,3-glucan-binding protein 1	complete	502	BGBP_BOMMO	9.38E-121
TRINITY_DN206442_c7_g2_i2	GPXs	Phospholipid hydroperoxide glutathione peroxidase-1	complete	196	GPX4_CALJA	6.95E-60
TRINITY_DN206811_c7_g2_i2	GPXs	Phospholipid hydroperoxide glutathione peroxidase-2	complete	170	GPX4_CALJA	1.52E-56
TRINITY_DN211448_c5_g3_i1	Hopscoth	Tyrosine-protein kinase hopscotch	complete	1117	JAK_DROME	2.4E-57
TRINITY_DN208562_c1_g1_i1	HPXs	Chorion peroxidase-1	internal	296	PERC_DROME	2.75E-56
TRINITY_DN212550_c1_g1_i3	HPXs	Chorion peroxidase-2	5prime_partial	986	PERO_DROME	3.95E-137
TRINITY_DN212846_c0_g2_i1	HPXs	Chorion peroxidase-3	3prime_partial	984	PERC_DROME	4.87E-114
TRINITY_DN79403_c0_g1_i1	HPXs	Chorion peroxidase-4	internal	109	PERC_DROME	7.61E-13
TRINITY_DN211353_c4_g2_i9	HPXs	Dual oxidase-1	complete	950	DUOX_DROME	2.47E-180

(table continued on next page)

(table continued from previous page)

TRINITY_DN240373_c0_g1_i1	HPXs	Dual oxidase-2	internal	176	DUOX_DROME	1.54E-99
TRINITY_DN27310_c0_g1_i1	HPXs	Dual oxidase-3	internal	113	DUOX_DROME	1.51E-48
TRINITY_DN165703_c0_g1_i1	HPXs	Myeloperoxidase	internal	210	PERM_MOUSE	8.27E-27
TRINITY_DN211655_c0_g1_i2	HPXs	Peroxidase-1	5prime_partial	718	PERO_DROME	0
TRINITY_DN213505_c6_g1_i3	HPXs	Peroxidase-2	complete	672	PERO_DROME	1.94E-98
TRINITY_DN177411_c0_g1_i1	HPXs	Peroxidase skpo-1	5prime_partial	314	SKPO1_CAEEL	7E-37
TRINITY_DN150155_c0_g1_i1	HPXs	Peroxidasin-1	internal	127	PXDN_XENTR	6.63E-36
TRINITY_DN200589_c0_g1_i1	HPXs	Peroxidasin-2	5prime_partial	532	PXDN_XENTR	4.61E-98
TRINITY_DN263380_c0_g1_i1	HPXs	Peroxidasin-3	internal	151	PXDN_DROME	5.15E-10
TRINITY_DN212828_c11_g2_i4	HPXs	Peroxidasin homolog	5prime_partial	1362	PXDN_MOUSE	0
TRINITY_DN151240_c0_g1_i1	HPXs	Thyroid peroxidase	internal	114	PERT_PIG	2.24E-25
TRINITY_DN210057_c3_g1_i2	IAPs	Death-associated inhibitor of apoptosis 1	complete	409	IAP_GVCPM	3.84E-67
TRINITY_DN205171_c1_g1_i1	IAPs	Death-associated inhibitor of apoptosis 2	complete	499	DIAP2_DROME	2.65E-73
TRINITY_DN183978_c0_g1_i1	Imd_IMD	Receptor-interacting serine/threonine-protein kinase 1-1	5prime_partial	655	RIPK1_MOUSE	0.00000406
TRINITY_DN202436_c1_g1_i1	Imd_IMD	Receptor-interacting serine/threonine-protein kinase 1-2	complete	252	RIPK1_MOUSE	9.46E-09
TRINITY_DN210495_c4_g1_i6	Ird5_IMD	Inhibitor of nuclear factor kappa-B kinase subunit alpha	complete	662	IKKA_XENLA	8.54E-122
TRINITY_DN211996_c0_g1_i5	JNK_ip_Toll	C-Jun-amino-terminal kinase-interacting protein 3	complete	1273	JIP3_HUMAN	0
TRINITY_DN204438_c0_g1_i2	Key_IMD	Optineurin	complete	358	OPTN_DANRE	6.59E-14
TRINITY_DN207725_c2_g1_i1	LYSs	Lysozyme-5	5prime_partial	155	LYS_GALME	3.04E-45
TRINITY_DN208075_c4_g1_i1	LYSs	Lysozyme-6	5prime_partial	153	LYS_BOMMO	9.6E-39
TRINITY_DN210486_c4_g1_i4	LYSs	Lysozyme c-1	complete	146	LYSC1_ANOGA	7.39E-41
TRINITY_DN205079_c0_g1_i1	LYSs	Lysozyme P	5prime_partial	221	LYSP_DROME	3.65E-26
TRINITY_DN211434_c0_g1_i3	LYSs	Lysozyme X	complete	137	LYSX_DROME	8.69E-21
TRINITY_DN209720_c5_g1_i2	LYSs	Lysozyme-2	complete	142	LYSC1_ANOGA	1.19E-39
TRINITY_DN196538_c0_g1_i1	MLs	Epididymal secretory protein E1-1	complete	148	NPC2_PANTR	2.73E-15
TRINITY_DN202244_c0_g2_i1	MLs	Epididymal secretory protein E1-2	5prime_partial	151	NPC2_PANTR	9.99E-13
TRINITY_DN263538_c0_g1_i1	MLs	Epididymal secretory protein E1-3	complete	147	NPC2_CANLF	6.69E-13
TRINITY_DN208537_c10_g1_i4	MLs	MD-2-related lipid-recognition protein	complete	160	ML1P_MANSE	1.43E-17
TRINITY_DN169438_c1_g1_i3	MLs	Protein NPC2 homolog-1	complete	102	NPC2_DROME	2.55E-24
TRINITY_DN206261_c0_g1_i2	MLs	Protein NPC2 homolog-2	complete	160	ES16_MANSE	2.37E-12

(table continued on next page)

(table continued from previous page)

TRINITY_DN208822_c2_g3_i1	MLs	Protein NPC2 homolog-3	complete	161	NPC2_DROME	1.1E-32
TRINITY_DN187224_c1_g2_i1	Myd88_Toll	Myeloid differentiation primary response protein MyD88	complete	410	MYD88_SALSA	2.43E-32
TRINITY_DN208835_c1_g2_i1	Pelle_Toll	Serine/threonine-protein kinase pelle	complete	798	KPEL_DROME	5.65E-65
TRINITY_DN199413_c3_g1_i2	Pellino-Toll	Protein pellino	5prime_partial	455	PELI_DROME	0
TRINITY_DN201679_c0_g1_i2	PepC54_ATG	Cysteine protease ATG4B	complete	373	ATG4B_DANRE	5.54E-120
TRINITY_DN265153_c0_g1_i1	PGRPs	Peptidoglycan recognition protein	5prime_partial	204	PGRP_BOMMO	1.24E-31
TRINITY_DN265153_c0_g2_i1	PGRPs	Peptidoglycan recognition protein	complete	263	PGRP_BOMMO	9.29E-31
TRINITY_DN212813_c2_g3_i4	PGRPs	Peptidoglycan recognition protein 1	5prime_partial	289	PGRP1_CAMDR	3.81E-41
TRINITY_DN183258_c0_g2_i3	PGRPs	Peptidoglycan recognition protein 3	complete	257	PGRP3_HUMAN	4.48E-35
TRINITY_DN208828_c7_g1_i1	PGRPs	Peptidoglycan recognition protein 3	5prime_partial	386	PGRP3_MOUSE	7.55E-61
TRINITY_DN206875_c0_g1_i1	PGRPs	Peptidoglycan-recognition protein 2	5prime_partial	205	PGRP2_HOLDI	1.12E-55
TRINITY_DN204473_c0_g1_i1	PGRPs	Peptidoglycan-recognition protein LB	complete	222	PGPLB_DROME	1.87E-48
TRINITY_DN211097_c8_g1_i1	PGRPs	Peptidoglycan-recognition protein LB	5prime_partial	262	PGPLB_DROME	4.32E-60
TRINITY_DN206097_c5_g1_i2	PGRPs	Peptidoglycan-recognition protein LF	complete	256	PGPLF_DROME	1.43E-37
TRINITY_DN208425_c4_g1_i1	PGRPs	Peptidoglycan-recognition protein LF	complete	246	PGPLF_DROME	1.48E-40
TRINITY_DN172177_c0_g1_i1	PGRPs	Peptidoglycan-recognition protein SB1	5prime_partial	171	PGSB1_DROME	9.04E-59
TRINITY_DN212786_c6_g2_i3	PGRPs	Peptidoglycan-recognition protein SB1	complete	140	PGSB1_DROME	8.06E-37
TRINITY_DN206082_c6_g1_i2	PGRPs	Peptidoglycan-recognition protein SC2	5prime_partial	292	PGSC2_DROME	6.07E-35
TRINITY_DN209777_c13_g2_i1	PGRPs	Peptidoglycan-recognition protein SC2	complete	205	PGSC2_DROSI	3.32E-50
TRINITY_DN204133_c0_g1_i2	PGRPs	Peptidoglycan-recognition protein SD	5prime_partial	314	PGPSD_DROME	1.87E-44
TRINITY_DN181389_c2_g1_i1	PPOs	Hemocyanin A chain	3prime_partial	184	HCYA_PANIN	2.29E-52
TRINITY_DN214369_c3_g1_i3	PPOs	Phenoloxidase 2	complete	695	PPO2_DROME	0
TRINITY_DN212806_c6_g1_i2	RELs	Nuclear factor NF-kappa-B p110 subunit	complete	957	NFKB1_DROME	3.64E-91
TRINITY_DN201036_c0_g1_i3	SCRBs	Protein croquemort-1	complete	477	CRQ_DROME	3.78E-77
TRINITY_DN204974_c10_g3_i3	SCRBs	Protein croquemort-2	complete	518	CRQ_DROME	5.92E-86
TRINITY_DN208844_c0_g1_i2	SCRBs	Protein croquemort-3	complete	520	CRQ_DROME	9.34E-106
TRINITY_DN209233_c5_g1_i1	SCRBs	Protein croquemort-4	internal	286	CRQ_DROME	9.94E-43
TRINITY_DN209233_c8_g1_i2	SCRBs	Protein croquemort-5	complete	528	CRQ_DROME	4.76E-66
TRINITY_DN209425_c4_g1_i1	SCRBs	Protein croquemort-6	complete	534	CRQ_DROME	4.03E-82
TRINITY_DN209511_c5_g1_i2	SCRBs	Protein croquemort-7	complete	515	CRQ_DROME	1.14E-100

(table continued on next page)

(table continued from previous page)

TRINITY_DN122913_c2_g1_i1	SCRBs	Scavenger receptor class B member 1-1	internal	111	SCRB1_CRIGR	2.88E-20
TRINITY_DN204157_c1_g1_i1	SCRBs	Scavenger receptor class B member 1-2	internal	436	SCRB1_RAT	4.11E-52
TRINITY_DN205230_c2_g1_i1	SCRBs	Scavenger receptor class B member 1-3	complete	575	SCRB1_PIG	1.27E-66
TRINITY_DN206915_c7_g1_i1	SCRBs	Scavenger receptor class B member 1-4	complete	545	SCRB1_PIG	2.6E-89
TRINITY_DN211612_c0_g1_i2	SCRBs	Scavenger receptor class B member 1-5	complete	570	SCRB1_MOUSE	3.3E-69
TRINITY_DN212784_c3_g1_i1	SCRBs	Scavenger receptor class B member 1-6	5prime_partial	539	SCRB1_BOVIN	1.3E-71
TRINITY_DN206998_c9_g1_i2	SCRBs	Sensory neuron membrane protein 1-1	5prime_partial	544	SNMP1_APIME	2.98E-147
TRINITY_DN212608_c7_g1_i7	SCRBs	Sensory neuron membrane protein 1-2	5prime_partial	524	SNMP1_APIME	2.47E-141
TRINITY_DN146135_c0_g1_i1	SCRCs	MAM and LDL-receptor class A domain-containing protein 2-1	5prime_partial	420	MLRP2_ACRMI	8.52E-42
TRINITY_DN194647_c0_g1_i1	SCRCs	MAM and LDL-receptor class A domain-containing protein 2(SCRC)	5prime_partial	780	MLRP2_ACRMI	4.68E-32
TRINITY_DN168267_c0_g1_i2	SPZs_Toll	Protein spaetzle-1	5prime_partial	311	SPZ_DROME	0.000000354
TRINITY_DN171056_c0_g1_i2	SPZs_Toll	Protein spaetzle-2	complete	215	SPZ_DROME	1.43E-12
TRINITY_DN192862_c1_g1_i1	SPZs_Toll	Protein spaetzle-3	complete	200	SPZ_DROME	5.25E-17
TRINITY_DN194130_c4_g1_i1	SPZs_Toll	Protein spaetzle-4	internal	136	SPZ_DROME	0.000577
TRINITY_DN196312_c4_g1_i1	SPZs_Toll	Protein spaetzle-5	complete	249	SPZ_DROME	9.63E-22
TRINITY_DN207008_c0_g1_i2	SPZs_Toll	Protein spaetzle-6	complete	207	SPZ_DROME	1.66E-14
TRINITY_DN27141_c0_g1_i1	SPZs_Toll	Protein spaetzle-7	5prime_partial	197	SPZ_DROME	5.79E-12
TRINITY_DN212647_c8_g1_i1	SRPNs	Alaserpin	5prime_partial	418	SERA_MANSE	3.28E-46
TRINITY_DN170074_c0_g1_i1	SRPNs	Leukocyte elastase inhibitor-1	5prime_partial	401	ILEU_BOVIN	7.02E-60
TRINITY_DN206893_c3_g3_i1	SRPNs	Leukocyte elastase inhibitor-2	5prime_partial	450	ILEU_XENTR	5.54E-34
TRINITY_DN208688_c6_g1_i5	SRPNs	Leukocyte elastase inhibitor-3	complete	401	ILEU_BOVIN	1.07E-50
TRINITY_DN210154_c3_g1_i1	SRPNs	Leukocyte elastase inhibitor-4	complete	440	Y2678_METMA	3.09E-17
TRINITY_DN211522_c5_g1_i1	SRPNs	Leukocyte elastase inhibitor-5	complete	570	ILEU_BOVIN	4.92E-39
TRINITY_DN204925_c0_g1_i1	SRPNs	Leukocyte elastase inhibitor B	5prime_partial	339	ILEUB_MOUSE	2.38E-44
TRINITY_DN201407_c0_g2_i1	SRPNs	Leukocyte elastase inhibitor C-1	5prime_partial	415	ILEUC_MOUSE	3.56E-36
TRINITY_DN206969_c8_g1_i5	SRPNs	Leukocyte elastase inhibitor C-2	5prime_partial	404	ILEUC_MOUSE	8.35E-79
TRINITY_DN208569_c16_g1_i1	SRPNs	Leukocyte elastase inhibitor C-3	5prime_partial	414	ILEU_BOVIN	7.76E-63
TRINITY_DN196029_c0_g1_i1	SRPNs	Neuroserpin-1	complete	404	NEUS_HUMAN	3.53E-47
TRINITY_DN206043_c1_g1_i2	SRPNs	Neuroserpin-2	5prime_partial	426	NEUS_RAT	5.04E-27

(table continued on next page)

(table continued from previous page)

TRINITY_DN207644_c13_g2_i1	SRPNs	Neuroserpin-3	complete	618	NEUS_CHICK	3.56E-56
TRINITY_DN210179_c0_g1_i2	SRPNs	Serpin B11	5prime_partial	441	SPB11_MOUSE	7.84E-58
TRINITY_DN204212_c0_g2_i1	SRPNs	Serpin B3	complete	421	SPB3_HUMAN	2.15E-56
TRINITY_DN201927_c0_g1_i1	SRPNs	Serpin B4-1	complete	405	ILEU_XENTR	3.46E-50
TRINITY_DN203015_c0_g1_i1	SRPNs	Serpin B4-2	5prime_partial	587	SPB4_HUMAN	4.81E-32
TRINITY_DN179560_c2_g1_i1	SRPNs	Serpin B8-1	internal	197	SPB8_MOUSE	2.06E-13
TRINITY_DN203799_c11_g1_i1	SRPNs	Serpin B8-2	5prime_partial	418	SPB8_BOVIN	3.17E-77
TRINITY_DN206323_c1_g1_i2	SRPNs	Serpin B8-3	complete	412	SPB8_BOVIN	1.8E-46
TRINITY_DN213132_c5_g2_i2	SRPNs	Serpin B8-4	complete	403	Y2678_METMA	3.51E-47
TRINITY_DN185998_c0_g1_i1	SRPNs	Serpin B9	5prime_partial	409	SPB9_HUMAN	2.18E-54
TRINITY_DN201023_c0_g1_i1	SRPNs	Uncharacterized serpin-like protein MM_2675	5prime_partial	291	ACH2_BOMMO	2.26E-38
TRINITY_DN205021_c0_g1_i1	Stam_JAK-STAT	Signal transducing adapter molecule 1	complete	414	STAM1_HUMAN	1.36E-146
TRINITY_DN210050_c5_g1_i1	STAT_JAK-STAT	Signal transducer and activator of transcription 5A	complete	813	STA5B_PIG	0
TRINITY_DN207416_c6_g2_i1	TAB2_IMD	TGF-beta-activated kinase 1 and MAP3K7-binding protein 2	3prime_partial	109	TAB2_RAT	0.0000484
TRINITY_DN204112_c0_g2_i1	TAK1_IMD	Mitogen-activated protein kinase kinase kinase 7	complete	583	M3K7_BOVIN	4.01E-168
TRINITY_DN212116_c6_g1_i3	TEPs	Alpha-2-macroglobulin-like protein 1	complete	1771	A2ML1_HUMAN	0
TRINITY_DN203691_c2_g1_i5	TEPs	CD109 antigen-1	complete	1631	CD109_HUMAN	2.44E-127
TRINITY_DN205654_c2_g1_i2	TEPs	CD109 antigen-2	complete	1462	CD109_HUMAN	0
TRINITY_DN206930_c2_g1_i2	TEPs	CD109 antigen-3	5prime_partial	1270	CD109_HUMAN	6.21E-55
TRINITY_DN208822_c2_g4_i2	Termicin	Termicin-1	5prime_partial	81	TERN_PSEUS	0.000000012
TRINITY_DN238294_c0_g1_i1	Termicin	Termicin-2	5prime_partial	69	TERN_PSEUS	0.000158
TRINITY_DN191714_c1_g1_i1	TLR_Toll	Protein toll-1	complete	1414	TOLL_DROME	1.95E-45
TRINITY_DN197463_c1_g1_i1	TLR_Toll	Protein toll-2	complete	1414	TOLL_DROME	8.32E-41
TRINITY_DN201305_c0_g1_i1	TLR_Toll	Protein toll-3	5prime_partial	1147	TOLL_DROME	5.39E-67
TRINITY_DN210363_c1_g1_i3	TLR_Toll	Protein toll-4	3prime_partial	932	TOLL_DROME	0
TRINITY_DN216320_c0_g1_i1	TLR_Toll	Protein toll-5	5prime_partial	259	TOLL_DROME	4.3E-32
TRINITY_DN199910_c0_g1_i1	TLR_Toll	Toll-like receptor 13-1	complete	835	TLR2_CRIGR	2.42E-35
TRINITY_DN207591_c0_g1_i2	TLR_Toll	Toll-like receptor 13-2	complete	979	TLR1_HUMAN	8.55E-21
TRINITY_DN212988_c1_g1_i1	TLR_Toll	Toll-like receptor 13-3	complete	846	TLR13_MOUSE	5.35E-30

(table continued on next page)

(table continued from previous page)

TRINITY_DN179468_c2_g1_i1	TLR_Toll	Toll-like receptor 2-1	internal	123	TLR2_MACFA	8.05E-22
TRINITY_DN213010_c2_g1_i8	TLR_Toll	Toll-like receptor 2-2	complete	818	TLR2_HORSE	1.53E-40
TRINITY_DN187166_c1_g1_i1	TLR_Toll	Toll-like receptor 2 type-2	5prime_partial	231	TLR22_CHICK	4.22E-28
TRINITY_DN206575_c10_g1_i1	TPXs	Peroxiredoxin 1	complete	197	PRDX1_DROME	1.09E-102
TRINITY_DN200539_c2_g1_i1	TPXs	Peroxiredoxin-4	complete	248	PRDX4_MOUSE	4.7E-119
TRINITY_DN199262_c2_g1_i1	TPXs	Peroxiredoxin-6	complete	232	PRDX6_PIG	5.1E-72
TRINITY_DN202519_c7_g1_i1	TPXs	Peroxiredoxin-6	complete	220	PRDX6_PONAB	1.03E-72
TRINITY_DN202739_c2_g1_i1	TPXs	Peroxiredoxin-6	complete	221	PRDX6_CHICK	4.27E-101
TRINITY_DN209350_c3_g1_i2	TPXs	Peroxiredoxin-6	5prime_partial	234	PRDX6_CHICK	5.35E-77
TRINITY_DN207196_c0_g1_i3	TPXs	Thioredoxin-dependent peroxide reductase	complete	235	PRDX3_RAT	4.74E-97
TRINITY_DN204719_c8_g1_i1	Traf_Toll	TNF receptor-associated factor 1	complete	413	TRAF1_MOUSE	8.47E-10
TRINITY_DN191477_c0_g1_i1	Traf_Toll	TNF receptor-associated factor 2	5prime_partial	583	TRAF2_HUMAN	3.25E-22
TRINITY_DN209739_c7_g1_i1	Traf_Toll	TNF receptor-associated factor 4	internal	394	TRAF4_MOUSE	7.42E-147
TRINITY_DN209961_c8_g1_i3	Traf_Toll	TNF receptor-associated factor 6	complete	383	TRAF6_BOVIN	1.21E-28
TRINITY_DN206535_c9_g1_i1	Transferrin	Melanotransferrin	5prime_partial	809	TRFM_RABIT	1.42E-124
TRINITY_DN207471_c0_g1_i1	Transferrin	Transferrin	complete	762	TRF_BLADI	5.24E-59
TRINITY_DN210772_c5_g4_i1	Transferrin	Transferrin	5prime_partial	515	TRF_BLADI	0
TRINITY_DN202454_c1_g1_i1	Tube_Toll	Interleukin-1 receptor-associated kinase 4	complete	520	IRAK4_HUMAN	6.96E-62
TRINITY_DN201921_c0_g1_i1	ULK_ATG	Serine/threonine-protein kinase ULK3	complete	466	ULK3_XENLA	3.97E-144
TRINITY_DN207690_c9_g1_i4	ULK_ATG	Serine/threonine-protein kinase unc-51	complete	794	ULK1_HUMAN	1.71E-105

Appendix I-G

Enriched gene ontology terms in treatments that in BP and MF. GO analysis was performed by goseq script in Trinity software with a cut off of 0.05 at Over represented FDR and redundancy was reduced by REVIGO (**Chapter I**).

Category	numDEInCat	numInCat	Term	Ontology	Over represented FDR
Enriched GO terms in Treatment:					
GO:0006952	28	578	Defense response	BP	7.33E-14
GO:0009607	21	497	Response to biotic stimulus	BP	3.49E-10
GO:0051707	19	417	Response to other organism	BP	6.33E-10
GO:0006955	20	510	Immune response	BP	1.12E-08
GO:0009605	27	1092	Response to external stimulus	BP	2.67E-08
GO:0050896	52	3825	Response to stimulus	BP	7.92E-08
GO:0007311	7	27	Maternal specification of dorsal/ventral axis, oocyte, germ-line encoded	BP	1.28E-07
GO:0002376	23	801	Immune system process	BP	1.76E-07
GO:0008063	8	41	Toll signaling pathway	BP	2.03E-07
GO:0030414	11	150	Peptidase inhibitor activity	MF	7.21E-07
GO:0004252	11	213	Serine-type endopeptidase activity	MF	2.90E-06
GO:0061783	6	34	Peptidoglycan muralytic activity	MF	4.59E-06
GO:0017171	11	255	Serine hydrolase activity	MF	2.20E-05
GO:0016485	10	155	Protein processing	BP	2.20E-05
GO:0008233	20	1097	Peptidase activity	MF	3.81E-05
GO:0051604	10	171	Protein maturation	BP	5.05E-05
GO:0001817	10	195	Regulation of cytokine production	BP	7.32E-05
GO:0030203	8	104	Glycosaminoglycan metabolic process	BP	7.62E-05
GO:0051704	19	970	Multi-organism process	BP	1.59E-04
GO:0031347	12	325	Regulation of defense response	BP	1.82E-04
GO:0010496	4	13	Intercellular transport	BP	3.26E-04
GO:0006508	20	1036	Proteolysis	BP	6.38E-04
GO:1901564	50	4708	Organonitrogen compound metabolic process	BP	6.50E-04
GO:0055114	20	1085	Oxidation-reduction process	BP	7.39E-04
GO:0022829	4	18	Wide pore channel activity	MF	9.07E-04
GO:0003824	74	9583	Catalytic activity	MF	9.47E-04
GO:0046914	28	2663	Transition metal ion binding	MF	0.002153901
GO:1901888	5	52	Regulation of cell junction assembly	BP	0.002840631
GO:0009056	26	1974	Catabolic process	BP	0.003231654
GO:0034097	7	116	Response to cytokine	BP	0.003289328
GO:0030246	9	302	Carbohydrate binding	MF	0.00404917
GO:0048583	29	2562	Regulation of response to stimulus	BP	0.006074953
GO:0001935	4	41	Endothelial cell proliferation	BP	0.00662464

(table continued on next page)

(table continued from previous page)

GO:0030155	9	300	Regulation of cell adhesion	BP	0.008505093
GO:0004553	8	227	Hydrolase activity, hydrolyzing O-glycosyl compounds	MF	0.010196919
GO:1901135	15	851	Carbohydrate derivative metabolic process	BP	0.011009493
GO:0016705	9	343	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	0.012133032
GO:0005539	6	138	Glycosaminoglycan binding	MF	0.012162977
GO:0019835	3	18	Cytolysis	BP	0.013956908
GO:2000351	3	13	Regulation of endothelial cell apoptotic process	BP	0.015411537
GO:0004040	2	4	Amidase activity	MF	0.016135069
GO:0016798	8	248	Hydrolase activity, acting on glycosyl bonds	MF	0.016768618
GO:0016787	38	4755	Hydrolase activity	MF	0.019081783
GO:0007166	18	1221	Cell surface receptor signaling pathway	BP	0.02150251
GO:0032963	4	56	Collagen metabolic process	BP	0.02452051
GO:0007249	3	14	I-kappab kinase/NF-kappab signaling	BP	0.027081516
GO:0019752	15	800	Carboxylic acid metabolic process	BP	0.028989039
GO:0005506	8	357	Iron ion binding	MF	0.030760479
GO:0020037	8	406	Heme binding	MF	0.033910908
GO:0055085	17	1200	Transmembrane transport	BP	0.037414709
GO:0046906	8	412	Tetrapyrrole binding	MF	0.038234553
GO:0043552	3	17	Positive regulation of phosphatidylinositol 3-kinase activity	BP	0.041060775
GO:0042943	2	3	D-amino acid transmembrane transporter activity	MF	0.043505279
GO:0046274	3	21	Lignin catabolic process	BP	0.0458794
GO:0052716	3	21	Hydroquinone:oxygen oxidoreductase activity	MF	0.0458794
Enriched GO terms in Control:					
GO:0016491	59	1626	Oxidoreductase activity	MF	3.99E-15
GO:0003824	167	9563	Catalytic activity	MF	9.01E-12
GO:0005506	23	306	Iron ion binding	MF	6.27E-11
GO:0055114	44	1152	Oxidation-reduction process	BP	9.71E-11
GO:0003674	289	21156	Molecular_function	MF	1.06E-10
GO:0016705	21	325	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	1.28E-08
GO:0020037	20	337	Heme binding	MF	7.09E-08
GO:0046906	20	345	Tetrapyrrole binding	MF	9.96E-08
GO:0044281	57	2266	Small molecule metabolic process	BP	1.24E-07
GO:0030246	20	368	Carbohydrate binding	MF	2.31E-07
GO:0016798	20	392	Hydrolase activity, acting on glycosyl bonds	MF	2.92E-07

(table continued on next page)

(table continued from previous page)

GO:0006082	39	1256	Organic acid metabolic process	BP	5.79E-07
GO:0008152	185	12740	Metabolic process	BP	7.70E-07
GO:0004497	16	248	Monooxygenase activity	MF	9.99E-07
GO:0005975	29	847	Carbohydrate metabolic process	BP	1.51E-06
GO:1901606	13	139	Alpha-amino acid catabolic process	BP	6.00E-06
GO:0004553	17	363	Hydrolase activity, hydrolyzing O-glycosyl compounds	MF	1.07E-05
GO:1901135	33	1126	Carbohydrate derivative metabolic process	BP	1.46E-05
GO:0008483	8	49	Transaminase activity	MF	3.11E-05
GO:0016769	8	49	Transferase activity, transferring nitrogenous groups	MF	3.11E-05
GO:0048037	23	623	Cofactor binding	MF	4.73E-05
GO:1901071	12	181	Glucosamine-containing compound metabolic process	BP	8.09E-05
GO:0005488	212	16517	Binding	MF	1.11E-04
GO:0006040	12	197	Amino sugar metabolic process	BP	1.79E-04
GO:0006022	14	265	Aminoglycan metabolic process	BP	2.28E-04
GO:0006629	35	1414	Lipid metabolic process	BP	2.90E-04
GO:1901605	15	305	Alpha-amino acid metabolic process	BP	6.41E-04
GO:0032787	20	566	Monocarboxylic acid metabolic process	BP	6.61E-04
GO:0009056	48	2326	Catabolic process	BP	9.42E-04
GO:0008061	10	160	Chitin binding	MF	9.42E-04
GO:0044255	28	1042	Cellular lipid metabolic process	BP	9.42E-04
GO:1901136	12	234	Carbohydrate derivative catabolic process	BP	0.001623593
GO:0001871	6	51	Pattern binding	MF	0.00178526
GO:0030247	6	51	Polysaccharide binding	MF	0.00178526
GO:0044706	8	97	Multi-multicellular organism process	BP	0.00180708
GO:0043167	127	9278	Ion binding	MF	0.002078887
GO:0000272	8	137	Polysaccharide catabolic process	BP	0.002919611
GO:0046692	6	56	Sperm competition	BP	0.003248247
GO:0006536	7	64	Glutamate metabolic process	BP	0.003407849
GO:0030170	8	96	Pyridoxal phosphate binding	MF	0.003407849
GO:0070279	8	97	Vitamin B6 binding	MF	0.00345296
GO:0022891	31	1444	Substrate-specific transmembrane transporter activity	MF	0.004905966
GO:0008810	4	36	Cellulase activity	MF	0.0067782
GO:0019842	11	246	Vitamin binding	MF	0.008802566
GO:1901566	26	1223	Organonitrogen compound biosynthetic process	BP	0.009738231
GO:0016614	15	501	Oxidoreductase activity, acting on CH-OH group of donors	MF	0.009738231
GO:0019695	3	10	Choline metabolic process	BP	0.011417138
GO:0005215	37	1943	Transporter activity	MF	0.011417138
GO:0051384	8	124	Response to glucocorticoid	BP	0.011732937

(table continued on next page)

(table continued from previous page)

GO:0004609	3	8	Phosphatidylserine decarboxylase activity	MF	0.015463808
GO:0015144	8	152	Carbohydrate transmembrane transporter activity	MF	0.018149402
GO:0046394	13	366	Carboxylic acid biosynthetic process	BP	0.018149402
GO:0006103	5	37	2-oxoglutarate metabolic process	BP	0.018633909
GO:0050662	15	495	Coenzyme binding	MF	0.019302754
GO:0055085	28	1254	Transmembrane transport	BP	0.019404858
GO:0009636	7	111	Response to toxic substance	BP	0.02439381
GO:0071704	150	11794	Organic substance metabolic process	BP	0.02439381
GO:0045471	7	103	Response to ethanol	BP	0.026388349
GO:0044283	16	574	Small molecule biosynthetic process	BP	0.027272418
GO:0047801	3	10	L-cysteine:2-oxoglutarate aminotransferase activity	MF	0.027882472
GO:0006811	28	1327	Ion transport	BP	0.027882472
GO:0008643	8	163	Carbohydrate transport	BP	0.029397212
GO:1901564	84	5736	Organonitrogen compound metabolic process	BP	0.030147324
GO:0043434	9	187	Response to peptide hormone	BP	0.032128636
GO:0016717	4	26	Oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water	MF	0.033488597
GO:0009167	9	208	Purine ribonucleoside monophosphate metabolic process	BP	0.033581904
GO:0005976	8	190	Polysaccharide metabolic process	BP	0.033581904
GO:0043168	60	3848	Anion binding	MF	0.033735578
GO:0015766	6	95	Disaccharide transport	BP	0.035306738
GO:0015772	6	95	Oligosaccharide transport	BP	0.035306738
GO:0042947	6	95	Glucoside transmembrane transporter activity	MF	0.035306738
GO:0030239	5	72	Myofibril assembly	BP	0.0362929
GO:0046434	7	128	Organophosphate catabolic process	BP	0.044155005
GO:0006532	3	12	Aspartate biosynthetic process	BP	0.04633753

Appendix II-A

Details of sample in **Chapter II**

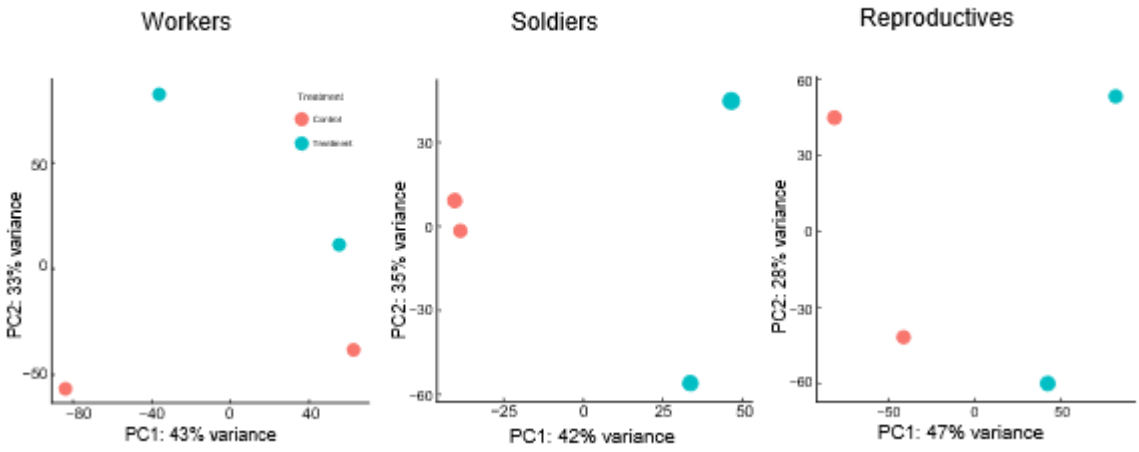
Species name	Sample (castes/categories)	Sample location	Experimental purpose
Blattella germanica	Larvae, Adults	In laboratory	*De novo* RNAseq Assembly
Blatta orientalis	Larvae, Adults	In laboratory	*De novo* RNAseq Assembly
Cryptocercus meridianus	Larvae, Adults	Yunshanping (27'14'N, 100'23'E, 3.250km),Yulongxueshan, Lijiang, Yunnan, China	*De novo* RNAseq Assembly, quantification of gene expression by RNAseq
Cryptocercus pudacoensis	Adults	Pudacuo (27'79'N,99'55'E,3.313km), Shangri-la, Diqing, Yunnan, China	*De novo* RNAseq Assembly
Mastotermes darwiniensis	Larvae, Workers, Soldiers	BAM	*De novo* RNAseq Assembly
Neotermes castaneus	Larvae, Soldiers, False-Workers, Neotenics	BAM	*De novo* RNAseq Assembly, quantification of gene expression by RNAseq
Kalotermes flavicollis	Larvae, Soldiers, False-Workers, Primary Reproductive, Nymph	BAM	*De novo* RNAseq Assembly
Cryptotermes brevis	Larvae, Soldiers, False-Workers, Primary Reproductive, Nymph	BAM	*De novo* RNAseq Assembly
Coptotermes formosanus	Larvae, Soldiers, Workers, Neotenics, Nymph	BAM	*De novo* RNAseq Assembly
Reticulitermes flavipes	Larvae, Soldiers, Workers, Neotenics, Nymph	BAM	*De novo* RNAseq Assembly
Prorhinotermes inopiinatus	Larvae, False-Workers, Soldiers, Nymph	BAM	*De novo* RNAseq Assembly
Macrotermes subhyalinus	Larvae, Big Workers, Small Workers, Big Soldiers, Small Soldiers	BAM	*De novo* RNAseq Assembly
Zootermopsis nevadasis	Larvae, False-Workers, Soldiers	BAM	*De novo* RNAseq Assembly
Pericapritermes sp.	Workers, Soldiers	China (N21.60213°, E101.58827°)	*De novo* RNAseq Assembly
Indotermes sp.	Worker, Soldier	China (N21.61799°, E101.58134°)	*De novo* RNAseq Assembly
Dicuspiditermes sp.	Worker	China (N21.61799°, E101.58134°)	*De novo* RNAseq Assembly
Globitermes sp.	Worker, Soldier, Nymph	China (N21.96151°, E101.20104°)	*De novo* RNAseq Assembly
Bulbitermes sp.	Worker, Soldier	China (N21.96151°, E101.20104°)	*De novo* RNAseq Assembly
Promirotermes sp.	Worker, Soldier	Camarron (N03.39228°, E011.47251°)	*De novo* RNAseq Assembly

Appendix II-B

Information of additional genomic and transcriptomic data sets for **Chapter II**

Species name	SRA Accession ID	Assemble program
Blaberus atropos	SRR921572	TRINITY v2.5.1
Eupolyphaga sinensis	SRR1184454, SRR1184455	
Periplaneta americana	SRR2994649, SRR2994650	
Cryptocercus wrighti	SRR921587	
Cryptotermes domesticus	SRR2039534	
Odontotermes formosanus	SRR528715	
Prorhinotermes simplex	SRR921637	
Reticulitermes banyulensis	SRR5253660	
Reticulitermes flavipes Olonne	SRR1325100	
Reticulitermes grassei	SRR13251[02-10]	
Reticulitermes lucifugus	SRR1325112, SRR1325111	
Hodotermopsis sjostedti	DRR013045	Newbler v2.7
Reticulitermes speratus	DRR013046	
Nasutitermes takasagoensis	DRR013047	
Zootermopsis nevadensis nuttingi	Official Gene Set OGSv2.2	
Macrotermes natalensis	Mnat_gene_v1.2.pep.fa	

Appendix II-C



Principle component analysis of gene expression after immune challenge in workers, soldiers, reproductives of *Neotermes castaneus*.

Appendix II-D

Details on primers for quantitative PCR for **Chapter II**

Name	Primer	Temperature (°C)	Species
Relish	F:5'-CAGTACAAGGCAAACCCTC-3' R:5'-TCATCTTCATCGTCGTCA-3'	57	*Neotermes canstaneus*
GNBP	F:5'-GCTCCAGGTAACGGCTTCGA-3' R:5'-ACCTTGCCAATAACTTCGT-3'	56.5	
Transferrin-1	F:5'-CAACAACTTCGCCTTCCTC-3' R:5'-TGCCCAGATCACCATTAGC -3'	61.5	
HPX	F:5'-CATGCCGTCTTTCCTACAC-3' R:5'-CTTCCGACCTTCGTTACCT-3'	59.5	
PGRP LB-1	F:5'-TGATTCTCATGGCCGCTTC-3' R:5'-ACATCGTAACCCGAGAGCAG-3'	61	
Termicin	F:5'-GGCACTGACTTCCATAACG-3' R:5'-GAGGGAGAACCTGGGCTAC-3'	56.5	
RL22	F:5'-AACGTCCATTATGTTGTCCT-3' R:5'-CAGCAACATATAAGGGCCAA-3'	56.5	
Relish	F:5'- CTTCAGCAATGGACCTCT -3' R:5'- GTCGCATTCTCAAGTCAG-3'	56.5	*Cryptocercus meridianus*
Termicin-1	F:5'- CTACCATCAACGCTATCA-3' R:5'- CTTGCGATGAATAATGTC-3'	56.5	
PGRP2	F:5'- GAGCGGAAGATGGTTGTC -3' R:5'- AGTTGCAGGCTGGAGTTA-3'	56.5	
PGRP-LB2	F:5'- GATGACGAACGGAACTGG-3' R:5'- GCTATTGTGACACGGGATG-3'	56.5	
ML1	F:5'- AACCGTCAAATTAAGGCAAC -3' R:5'- ACTCTATGTCCAATACCGTGA -3'	56.5	
BGBP-1	F:5'-TAGCAGTGGGTGGAGTAAA-3' R:5'- GAAGCCCGAGGTGAAATA-3'	58	
Defensin-1	F:5'- CAACAAACGCACTCTTCA-3' R:5'- ATTGCCAGCATCACTCAC -3'	56.5	
RL24	F:5'-CCGTGATCCGTATACCGTTG-3' R:5'-CCTCTTCATCAAGTGCGACGA-3'	56.5	

(table continued from previous page)

Appendix II-E

Enriched gene ontology terms in treatments that in BP and MF. GO analysis was performed by goseq script in Trinity software with a cut off of 0.05 at Over represented FDR and redundancy was reduced by REVIGO.

Category	nDIC	nIC	Term	Ontology	Over represented FDR
Enriched GO terms in Treatment group (Reproductives, *N. castaneus*):					
GO:0046189	6	46	Phenol-containing compound biosynthetic process	BP	0.0005730754059
Enriched GO terms in Treatment group (Workers, *N.castaneus*):					
GO:0004666	2	6	Prostaglandin-endoperoxide synthase activity	MF	0.0309841324376
GO:0008150	19	16810	Biological_process	BP	0.0309841324376
Enriched GO terms in Treatment group (Soldiers, *N. castaneus*):					
GO:0003824	44	7914	Catalytic activity	MF	0
GO:0008152	47	10688	Metabolic process	BP	1.90282E-05
GO:0005488	55	14344	Binding	MF	1.92225E-05
GO:0046914	17	1679	Transition metal ion binding	MF	1.92225E-05
GO:0016787	24	3650	Hydrolase activity	MF	7.11855E-05
GO:0071704	43	10028	Organic substance metabolic process	BP	0.00012424
GO:0008509	9	340	Anion transmembrane transporter activity	MF	0.00012424
GO:0005342	7	145	Organic acid transmembrane transporter activity	MF	0.000126496
GO:0046943	7	145	Carboxylic acid transmembrane transporter activity	MF	0.000126496
GO:0043169	29	5642	Cation binding	MF	0.000235537
GO:0043900	8	283	Regulation of multi-organism process	BP	0.000299331
GO:0006952	10	577	Defense response	BP	0.000299331
GO:0008745	3	10	N-acetylmuramoyl-L-alanine amidase activity	MF	0.000387795
GO:0043167	35	7772	Ion binding	MF	0.00041371
GO:0046872	28	5600	Metal ion binding	MF	0.000495519
GO:0065007	41	9012	Biological regulation	BP	0.000677044
GO:0042834	3	13	Peptidoglycan binding	MF	0.000692581
GO:0008270	13	1316	Zinc ion binding	MF	0.000692581
GO:0032502	27	4574	Developmental process	BP	0.000949345
GO:0045087	7	264	Innate immune response	BP	0.000949345
GO:0000270	3	15	Peptidoglycan metabolic process	BP	0.00096726
GO:0015291	7	243	Secondary active transmembrane transporter activity	MF	0.001119966
GO:0006950	18	2391	Response to stress	BP	0.001119966
GO:0009987	50	13574	Cellular process	BP	0.001119966
GO:0006807	37	9003	Nitrogen compound metabolic process	BP	0.001165683
GO:0052689	6	199	Carboxylic ester hydrolase activity	MF	0.00144739
GO:0006820	8	372	Anion transport	BP	0.00144739
GO:0061783	3	26	Peptidoglycan muralytic activity	MF	0.001966415
GO:0009605	12	1127	Response to external stimulus	BP	0.002135709
GO:0022414	15	1795	Reproductive process	BP	0.002900205

(table continued on next page)

(table continued from previous page)

GO:0006811	11	971	Ion transport	BP	0.003941574
GO:0050896	22	3814	Response to stimulus	BP	0.003941574
GO:0016810	5	129	Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	MF	0.004164143
GO:0002831	5	132	Regulation of response to biotic stimulus	BP	0.004550899
GO:0043902	5	124	Positive regulation of multi-organism process	BP	0.004613623
GO:0097164	5	157	Ammonium ion metabolic process	BP	0.005469845
GO:0061058	2	5	Regulation of peptidoglycan recognition protein signaling pathway	BP	0.005583491
GO:0005215	12	1357	Transporter activity	MF	0.0064028
GO:0009607	8	594	Response to biotic stimulus	BP	0.006658767
GO:0016811	4	72	Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	MF	0.007772816
GO:1901615	7	378	Organic hydroxy compound metabolic process	BP	0.007942822
GO:0050794	34	7958	Regulation of cellular process	BP	0.008024555
GO:0015804	3	27	Neutral amino acid transport	BP	0.008129793
GO:0016491	10	1110	Oxidoreductase activity	MF	0.008129793
GO:0030234	9	726	Enzyme regulator activity	MF	0.008129793
GO:1901564	23	4658	Organonitrogen compound metabolic process	BP	0.008419406
GO:0022804	7	434	Active transmembrane transporter activity	MF	0.010360296
GO:0007165	17	2692	Signal transduction	BP	0.011925102
GO:0000977	6	308	RNA polymerase II regulatory region sequence-specific DNA binding	MF	0.011925102
GO:0016714	2	7	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen….	MF	0.012304929
GO:0042943	2	8	D-amino acid transmembrane transporter activity	MF	0.013087253
GO:0051704	10	1254	Multi-organism process	BP	0.013512264
GO:0048067	3	26	Cuticle pigmentation	BP	0.014471957
GO:0007310	3	31	Oocyte dorsal/ventral axis specification	BP	0.017790031
GO:0015711	6	306	Organic anion transport	BP	0.017915722
GO:0016705	5	293	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	MF	0.017965907
GO:0015294	4	105	Solute:cation symporter activity	MF	0.017965907
GO:0015849	5	196	Organic acid transport	BP	0.018036856
GO:0018958	4	125	Phenol-containing compound metabolic process	BP	0.019488882
GO:0042133	3	54	Neurotransmitter metabolic process	BP	0.020357753
GO:0044238	34	9560	Primary metabolic process	BP	0.022287666
GO:0008063	3	47	Toll signaling pathway	BP	0.024807258
GO:1900619	2	20	Acetate ester metabolic process	BP	0.031109398
GO:0023051	14	1908	Regulation of signaling	BP	0.031480321
GO:0010646	14	1921	Regulation of cell communication	BP	0.031636838
GO:0032101	6	384	Regulation of response to external stimulus	BP	0.032274609

(table continued on next page)

(table continued from previous page)

GO:0002804	2	8	Positive regulation of antifungal peptide production	BP	0.034318823
GO:0055114	8	813	Oxidation-reduction process	BP	0.036321166
GO:0042940	2	11	D-amino acid transport	BP	0.036892599
GO:2000274	2	9	Regulation of epithelial cell migration, open tracheal system	BP	0.037068619
GO:0015081	4	154	Sodium ion transmembrane transporter activity	MF	0.037276238
GO:0098772	9	964	Molecular function regulator	MF	0.040099742
GO:0016485	4	142	Protein processing	BP	0.040099742
GO:0005243	2	10	Gap junction channel activity	MF	0.041556868
GO:0004497	4	243	Monooxygenase activity	MF	0.047716375
Enriched GO terms in Treatment group (*C. meridianus*):					
GO:0009617	18	289	Response to bacterium	BP	0.006653221
GO:0000270	6	26	Peptidoglycan metabolic process	BP	0.014135439
GO:0040040	5	20	Thermosensory behavior	BP	0.029819445
GO:0045087	20	381	Innate immune response	BP	0.029819445
GO:0009605	48	1403	Response to external stimulus	BP	0.029819445
GO:0042416	4	9	Dopamine biosynthetic process	BP	0.029819445
GO:0009607	29	725	Response to biotic stimulus	BP	0.034760363
GO:0072348	8	57	Sulfur compound transport	BP	0.034760363
GO:0061058	4	10	Regulation of peptidoglycan recognition protein signaling pathway	BP	0.039499368
GO:0048060	6	31	Negative gravitaxis	BP	0.041597768
GO:0016714	4	6	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen….	MF	0.009466975
GO:0061783	7	41	Peptidoglycan muralytic activity	MF	0.010561336
GO:0004611	4	20	Phosphoenolpyruvate carboxykinase activity	MF	0.029819445
GO:0004613	4	20	Phosphoenolpyruvate carboxykinase (GTP) activity	MF	0.029819445
GO:1901682	8	57	Sulfur compound transmembrane transporter activity	MF	0.034760363
GO:0008745	5	20	N-acetylmuramoyl-L-alanine amidase activity	MF	0.036106415
GO:0046943	15	223	Carboxylic acid transmembrane transporter activity	MF	0.043397566
GO:0005342	15	226	Organic acid transmembrane transporter activity	MF	0.043737707

Note: numDEInCat: number of significant differentially expressed genes in corresponding category; numInCat: number of total genes in corresponding category that derived from trinotate. BP: Biological process, MF: molecular functions.

Appendix III-A

SDS-PAGE and Liquid Chromatography-Tandem Mass Spectrometry

100µl of the diluted secretion was mixed with 5×SDS sample buffer, boiled for 5 min at 95 °C and immediately put on ice. After centrifugation, 25µl of sample with buffer was loaded in 10 % SDS-PAGE gel and run at 110V for 3 h (Electrophoresis Power Supply EPS 301 Amersham Biosciences, Little Chalfont UK; Electrophoresis dock SE300 miniVE Hoefer, Inc., Holliston, MA USA). Following the separation, proteins were stained by Coomassie Brilliant Blue (Roti©-Blue) for 6 h and washed with ddH2O until the bands were clear. Subsequently, the Coomassie-stained gel lane was cut into 20 slices and proteins were in-gel digested with trypsin. In brief, gel slices were washed with 50% (v/v) acetonitrile in 50 mM ammonium bicarbonate, shrunk by dehydration in acetonitrile, and dried in a vacuum centrifuge. The dried gel pieces were incubated with 50ng trypsin (sequencing grade modified, Promega) in 25µL of 50 mM ammonium bicarbonate at 37 °C overnight. To extract the peptides, 25 µL of 0.5% (v/v) trifluoroacetic acid (TFA) in acetonitrile was added and the extract was dried under vacuum. Peptides were reconstituted in 10µL of 0.1% (v/v) TFA, 5% (v/v) acetonitrile and 6.5µL were analyzed by a reversed-phase capillary nano liquid chromatography system (Ultimate 3000, Thermo Scientific) connected to an Orbitrap Velos mass spectrometer (Thermo Scientific). The LC system was coupled to the mass spectrometer via a nanospray flex ion source equipped with a stainless steel emitter (Thermo Scientific). Samples were injected and concentrated on a trap column (PepMap100 C18, 3µm, 100 Å, 75µm i.d. × 2cm, Thermo Scientific) equilibrated with 0.05% TFA, 2% acetonitrile in water. After switching the trap column inline, LC separations were performed on a capillary column (PepMap100 C18, 2µm, 100 Å, 75µm i.d. × 25cm, Thermo Scientific) at an eluent flow rate of 300 nL/min using a linear gradient of 3–50% B in 50 min. Mobile phase A contained 0.1% formic acid in water, and mobile phase B contained 0.1% formic acid in acetonitrile. Mass spectra were acquired in a data-dependent mode utilizing a single MS survey scan with a resolution of 60,000 in the Orbitrap, and MS/MS scans of the 20 most intense precursor ions in the linear trap quadrupole. The MS survey range was m/z 350-1500. The dynamic exclusion time (for precursor ions) was set to 60 s and automatic gain control was set to 1×10^6 and 5,000 for Orbitrap-MS and LTQ-MS/MS scans, respectively.

Appendix III-B

Details of identified secretion proteins from soldier of *M. darwiensis*.

Description	Mascot Score	Molecular Weight [Da]	Num. of significant unique sequences	Sequence coverage [%]	emPAI	Protein ID	Protein Abberation	E-value	Ants	Bees
Maltase 2	72145	66619	16	0.86	112.74	O16099	MAL2_DROVI	0	1	
Glucose dehydrogenase [FAD- quinone]	26018	69640	36	0.6	23.58	P18172	DHGL_DROPS	0	1	1
Apolipoprotein	4054	24153	18	0.63	21.07	PF01442.15	Apolipoprotein	0.0027		
Protein yellow	8199	50819	29	0.66	20.19	Q9BI18	YELL_DROPS	3.54E-76	1	1
Glucosylceramidase	33167	60438	40	0.67	17.58	Q70KH2	GLCM_PIG	1.66E-132	1	1
L-ascorbate oxidase	33498	72917	28	0.66	13.2	P14133	ASO_CUCSA	9.19E-56		
Apolipoprotein	2003	30695	17	0.46	9.05	PF01442.15	Apolipoprotein	0.000000074		
Apolipoprotein	17939	25920	14	0.39	8.44	PF01442.15	Apolipoprotein	0.066		
Leukocyte elastase inhibitor C	4768	45561	12	0.43	8.12	Q5SV42	ILEUC_MOUSE	2.99E-68	1	1
Fasciclin-2	4601	85753	39	0.48	8.09	P22648	FAS2_SCHAM	0		
Venom allergen 3	843	21331	7	0.36	6.01	P35779	VA3_SOLRI	4.05E-46		
Glucosylceramidase	3024	60426	23	0.44	5.54	P17439	GLCM_MOUSE	1.44E-130	1	1
Venom allergen 3	1376	27403	9	0.51	5.2	P35778	VA3_SOLIN	7.28E-61		
Regucalcin	6257	37834	14	0.21	4.86	Q2PFX5	RGN_MACFA	9.25E-66	1	1
Multiple inositol polyphosphate phosphatase 1	3276	52017	9	0.3	4.01	Q5R890	MINP1_PONAB	1.25E-38		1
Lazarillo protein	513	21326	8	0.52	3.75	P49291	LAZA_SCHAM	0.00000181		
Trehalase	5351	65790	20	0.4	3.63	Q8MMG9	TREA_PIMHY	0	1	1
Polyubiquitin	637	11469	4	0.46	3.17	P23398	UBIQP_STRPU	1.69E-68		1
Alpha-amylase 1	4687	56544	16	0.42	3.1	Q23835	AMY1_DROAN	0	1	1
Actin- clone 403	2040	41827	3	0.35	3.06	P18603	ACT4_ARTSX	0	1	
Serpin B6	3349	45128	14	0.38	3.02	Q4R3G2	SPB6_MACFA	1.06E-53	1	1
Glucose dehydrogenase [FAD- quinone]	3828	68117	15	0.35	2.89	P18172	DHGL_DROPS	2.99E-157	1	1
Cathepsin L	751	38004	10	0.33	2.75	Q26636	CATL_SARPE	2.07E-168	1	
Haemolymph juvenile hormone binding protein (JHBP)	241	9270	2	0.27	2.68	PF06585.8	JHBP	3.2E-17	1	

(table continued on next page)

(table continued from previous page)

Glutaminyl-peptide cyclotransferase	643	39910	10	0.38	2.52	Q16769	QPCT_HUMAN	6.56E-101		
Lysosomal aspartic protease	1412	41553	12	0.34	2.36	Q03168	ASPP_AEDAE	0	1	
Peroxidase	4602	76890	20	0.38	2.33	Q01603	PERO_DROME	0		
Glutathione S-transferase 1-1	154	24465	7	0.32	2.3	P30108	GSTT1_DROYA	1.96E-109	1	
14-3-3 protein zeta	889	28099	6	0.28	2.28	Q2F637	1433Z_BOMMO	9.4E-179		
Peptidyl-prolyl cis-trans isomerase FKBP14	408	28617	7	0.32	2.21	Q5R941	FKB14_PONAB	1.85E-53		
Protein FAM151B	1261	32514	7	0.29	2.18	Q6UXP7	F151B_HUMAN	1.89E-47	1	
Serpin B6 (leukocyte elastase inhibitor-like)*	1259	48043	13	0.3	2.11	P35237	SPB6_HUMAN	9.04E-80	1	1
Histone H4	687	11374	3	0.31	1.95	Q28DR4	H4_XENTR	3.5E-67	1	
Uncharacterized serpin-like protein (serine protease inhibitor 88Ea-like)*	397	46926	12	0.37	1.92	Q8PTN8	Y2678_METMA	7.19E-65	1	1
Lysosomal aspartic protease	1280	43453	9	0.27	1.63	Q03168	ASPP_AEDAE	6.84E-48	1	
Esterase FE4	916	39074	8	0.29	1.62	P35502	ESTF_MYZPE	4.34E-31		1
Nucleobindin-2	618	65361	10	0.23	1.62	P81117	NUCB2_MOUSE	2.62E-78		
Protein lethal(2)essential for life	287	21483	5	0.39	1.62	P82147	L2EFL_DROME	4.29E-17		
Putative ferric-chelate reductase homolog (putative defense protein)*	586	17161	3	0.29	1.61	Q8MSU3(AFZ 78849.1)	FRRS1_DROME	0.000000365 (6e-49)		
15-hydroxyprostaglandin dehydrogenase[NAD(+)]	159	27603	5	0.24	1.47	Q3T0C2	PGDH_BOVIN	1.17E-51		
Angiotensin-converting enzyme	118	23941	5	0.24	1.38	P12821	ACE_HUMAN	1.15E-09		
Histone H2A	228	14805	3	0.17	1.31	P19178	H2A_PLADU	1.36E-80	1	
Apolipoprotein	803	24966	5	0.25	1.3	PF01442.15	Apolipoprotein	0.00028		
Unkown	432	35610	3	0.11	1.27	Q8NBR0	P5I13_HUMAN	0.000797		
Peroxiredoxin 1	98	21795	4	0.2	1.15	Q9V3P0	PRDX1_DROME	1.25E-109		
Dehydrogenase/reductase SDR family member 11	134	27652	5	0.16	1.13	Q3ZBV9	DHR11_BOVIN	2.44E-85		
Peptidylglycine alpha-hydroxylating monooxygenase	416	39101	5	0.15	1.12	O01404	PHM_DROME	1.22E-143		
Chitooligosaccharidolytic beta-N-acetylglucosaminidase	85	16582	3	0.17	1.11	P49010	HEXC_BOMMO	5.28E-33		
OV-16 antigen (protein D3-like)*	1292	28414	5	0.24	1.08	P31729(XP_0 21938545.1)	OV16_ONCVO	1.39E-63(9e-120)		
Protein NPC2 homolog	708	16879	3	0.17	1.08	Q9VQ62	NPC2_DROME	3.16E-16	1	
Retinal dehydrogenase 2	650	52324	9	0.21	1.06	Q62148	AL1A2_MOUSE	0		
Venom carboxylesterase-6	778	69679	10	0.22	0.94	B2D0J5	EST6_APIME	1.16E-107		1

(table continued on next page)

(table continued from previous page)

ADP-ribosylation factor 1	58	20675	3	0.18	0.83	P61210	ARF1_LOCMI	3.55E-134		
Myophilin	95	20917	3	0.17	0.82	Q24799	MYPH_ECHGR	3.22E-53		
Histone H2B.3	205	13852	2	0.13	0.81	P35069	H2B3_TIGCA	1.28E-81	1	
Glyceraldehyde-3-phosphate dehydogenase	151	35588	5	0.18	0.8	Q4U3L0	G3P_GLOMM	0		
Multiple inositol polyphosphatase 1	356	51112	6	0.18	0.78	Q5R890	MINP1_PONAB	3.86E-50		
ATP synthase subunit alpha- mitochondrial	1229	59431	8	0.14	0.76	P35381	ATPA_DROME	0		
Aspartic protease Bla g 2	837	38523	5	0.16	0.72	P54958	ASP2_BLAGE	1.96E-89		
Histone H3.3	86	15318	2	0.1	0.71	Q6P823	H33_XENTR	4.43E-87		
Fructose-bisphosphate aldolase	209	39660	5	0.18	0.7	P07764	ALF_DROME	0		
Arginine kinase	183	39810	5	0.18	0.69	P91798	KARG_SCHAM	0		
DE-cadherin	56	15698	2	0.18	0.69	Q24298	CADE_DROME	1.07E-56		
40S ribosomal protein S14	626	16153	2	0.16	0.66	C0HKA0/1	RS14A/B_DROME	1.23E-94		
Synaptic vesicle membrane protein	185	50503	6	0.14	0.65	Q9HCJ6	VAT1L_HUMAN	5.67E-158		
Multiple inositol polyphosphate phosphatase 1	210	16679	2	0.2	0.64	O35217	MINP1_RAT	0.000244		1
Calmodulin	75	16800	2	0.13	0.64	P62154	CALM_LOCMI	1.58E-104		
Unkown	255	25177	3	0.13	0.64					
Aquaporin AQPAn.G	133	26309	3	0.14	0.61	Q7PWV1	AQP_ANOGA	5.78E-117		
Pathogenesis-related protein 5	93	26409	3	0.21	0.6	P28493	PR5_ARATH	1.21E-66		
Superoxide dismutase [Cu-Zn]	129	18154	2	0.14	0.58	Q01137	SODC_SCHMA	1.79E-53		
Protein disulfide-isomerase	208	55509	6	0.14	0.57	P54399	PDI_DROME	0		
Lachesin	430	46390	5	0.16	0.57	Q26474	LACH_SCHAM	0		
Laccase-2	535	75994	8	0.1	0.56	Q8RYM9	LAC2_ORYSJ	3.09E-52		1
CD109 antigen	1937	162598	15	0.13	0.55	Q6YHK3	CD109_HUMAN	0		
Serpin B11(serine protease inhibitor 77Ba-like)*	88	19100	2	0.12	0.54	Q96P15	SPB11_HUMAN	6.2E-28		
Chitin binding Peritrophin-A domain	916	29969	3	0.09	0.52	PF01607.21	CBM_14	0.0000097		
Heat shock 70 kDa protein cognate 4	1495	70845	3	0.12	0.52	Q9U639	HSP7D_MANSE	0		
Elongation factor 1-alpha	1891	50546	5	0.12	0.51	P29520	EF1A_BOMMO	0.00000162		
Unkown	77	20473	2	0.12	0.5					
Ras-like protein 3	62	20805	2	0.13	0.49	P08645	RAS3_DROME	3.93E-116		

(table continued on next page)

(table continued from previous page)

Beta-1-3-glucan-binding protein 2(Gram-negative binding protein 2, GNBP2)*	492	42754	4	0.15	0.48	Q8N0N3(AAZ 08505.1)	BGBP_PENMO	3.18E-108(0.0)		
60S ribosomal protein L12	41	21969	2	0.08	0.46	P23358	RL12_RAT	3.24E-35		
Putative cysteine proteinase C(cathepsin L)*	1124	169304	15	0.13	0.45	Q9VN93	CPR1_DROME	1.1E-123		
Glutathione S-transferase	56	23479	2	0.09	0.42	O18598	GST1_BLAGE	4.05E-103		
Annexin B9	75	35840	3	0.1	0.42	P22464	ANXB9_DROME	2.14E-179		
Peptidyl-prolyl cis-trans isomerase	34	23621	2	0.07	0.42	P24367	PPIB_CHICK	1.29E-97		
40S ribosomal protein S5a	88	24284	2	0.08	0.41	Q24186	RS5A_DROME	6.94E-138		
GTP-binding nuclear protein Ran	78	24616	2	0.09	0.4	Q9VZ23	RAN_DROME	2.86E-149		
Chondroadherin	50	25408	2	0.1	0.39	O55226	CHAD_MOUSE	8.12E-22		
Pleckstrin homology domain-contain protein	26	25090	2	0.05	0.39	Q9HB20	PKHA3_HUMAN	2.03E-75		
40S ribosomal protein S3	111	26520	2	0.08	0.37	P62909	RS3_RAT	2.39E-156		
Apolipophorins	862	366654	27	0.09	0.36	Q9U943	APLP_LOCMI	0		
ATP synthase subunit beta- mitochondrial	761	56554	3	0.07	0.35	Q05825	ATPB_DROME	0		
CD9 antigen	342	28847	2	0.08	0.34	P40240	CD9_MOUSE	1.14E-43		
Lysosomal Pro-X carboxypeptidase	271	43129	3	0.12	0.34	Q2TA14	PCP_BOVIN	1.72E-149	1	1
Phosphoglycerate mutase 2	66	28855	2	0.09	0.34	Q32KV0	PGAM2_BOVIN	7.4E-127		
Lipase 3	92	44146	3	0.07	0.33	O46108	LIP3_DROME	1.55E-116		1
Angiotensin-converting enzyme	165	44465	3	0.14	0.33	Q10751	ACE_CHICK	8.74E-126		
Phospholipase A2	106	30214	2	0.09	0.32	Q7M4I6	PA2_BOMPE	5.36E-20	1	
Alpha-N-acetylglucosaminidase	102	47553	3	0.11	0.3	P54802	ANAG_HUMAN	4.14E-105		
Protein O-linked-mannose beta-1,2-N-acetylglucosaminyltransferase 1	53	47970	3	0.09	0.3	Q5RCB9	PMGT1_PONAB	1.12E-09		
Tubulin beta-1 chain	85	50185	2	0.08	0.29	O17449	TBB1_MANSE	0		
Beta-amyloid-like protein	206	82203	5	0.06	0.29	P14599	A4_DROME	4.4E-139		
Protein-tyrosine phosphatase receptor IA-2	56	33046	2	0.1	0.29	PF11548.5	Receptor_IA-2	1.2E-24		
Tropomyosin	109	32757	2	0.08	0.29	Q8T6L5	TPM_PERFU	0		
Protein 5NUC	268	65771	4	0.06	0.29	Q9XZ43	5NTD_LUTLO	7.84E-150		
Insulin-like growth factor-binding protein complex acid(leucine-rich repeat-containing protein 15-like)*	256	85833	5	0.12	0.28	O02833(XP_0 21915787.1)	ALS_PAPHA	2.2E-29(0.0)		
Eukaryotic initiation factor 4	249	50903	3	0.07	0.28	Q5SV42	ILEUC_MOUSE	0		

(table continued on next page)

(table continued from previous page)

Aldose reductase	64	35179	2	0.07	0.27	P15121	ALDR_HUMAN	2.62E-106		
Endoplasmic reticulum resident protein 44	125	34667	2	0.08	0.27	Q3T0L2	ERP44_BOVIN	6.3E-94		
Endoplasmic reticulum lectin 1	155	52861	3	0.05	0.27	Q8VEH8	ERLEC_MOUSE	2E-90		
Alpha-L-fucosidase	54	54883	3	0.06	0.26	C3YWU0	FUCO_BRAFL	0		
Cathepsin B	80	37373	2	0.09	0.25	P07688	CATB_BOVIN	6.2E-141	1	
Angiotensin-converting enzyme	114	76702	4	0.06	0.25	Q10714	ACE_DROME	0		
Angiotensin-converting enzyme	203	76339	4	0.06	0.25	Q10751	ACE_CHICK	0		
2-oxoglutarate and iron-dependent oxygenase domain-containing protein 3	123	36920	2	0.09	0.25	Q5XGE0	OGFD3_XENTR	1.47E-90		
La-related protein 7	36	40427	2	0.05	0.23	Q28G87	LARP7_XENTR	1.67E-12		
Venom serine protease 34	248	43202	2	0.09	0.21	Q8MQS8	SP34_APIME	1.81E-85	1	1
Enolase	77	47147	2	0.05	0.19	P15007	ENO_DROME	0		
26S protease regulatory subunit	42	49371	2	0.05	0.19	P48601	PRS4_DROME	0		
Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	25	48424	2	0.03	0.19	Q9Z2I9	SUCB1_MOUSE	0		
Cuticlin-1 (Unkown)*	150	77931	3	0.04	0.18	Q03755	CUT1_CAEEL	0.000661		
Chitinase-like protein EN03	92	50796	2	0.05	0.18	Q9GV28	IDGFL_BOMMO	1.51E-144		
Unkown	44	51331	2	0.04	0.18					
Cytoplasmic Fragile-X interacting family	37	52171	2	0.03	0.17	PF05994.8	FragX_IP	3.2E-56		
Myosin heavy chain- non-muscle	276	225098	8	0.04	0.16	Q99323	MYSN_DROME	0		
Mitochondrial-processing peptidase subunit alpha	30	60355	2	0.03	0.15	Q0P5M8	MPPA_BOVIN	0		
Alkaline phosphatase 4	33	59589	2	0.04	0.15	Q24238	APH4_DROME	1.89E-154		
MAM domain-containing glycosylphosphatidylinositol anchor protein 1(Unkown)	238	94678	3	0.04	0.14	Q8NFP4	MDGA1_HUMAN	0.00000617		
Heat shock 70 kDa protein cognate 5	175	75607	2	0.03	0.12	P29845	HSP7E_DROME	0		
Heat shock protein 83	496	83422	2	0.03	0.11	Q9BLC5	HSP83_BOMMO	0		

(table continued on next page)

(table continued from previous page)

Neuroglian	79	135908	3	0.03	0.1	P20241	NRG_DROME	0		
Aconitate hydratase- mitochondrial	197	87052	2	0.03	0.1	Q99798	ACON_HUMAN	0		
Elongation factor 2	44	94566	2	0.02	0.09	P13060	EF2_DROME	0		
Beta-mannosidase	95	101075	2	0.02	0.09	Q4FZV0	MANBA_RAT	0		
Lysosomal alpha-mannosidase	61	116266	2	0.02	0.08	Q60HE9	MA2B1_MACFA	0		
Serine/threonine-protein kinase	27	201400	2	0.01	0.04	Q5VT25	MRCKA_HUMAN	0		
Nesprin-1	32	687948	2	0	0.01	Q6ZWR6	SYNE1_MOUSE	6.61E-120		

*:Identifications were derived from NCBI